

# OTeN

OBSERVATOIRE DES TERRITOIRES Numériques

■ Données publiques : clés et repères

■ 2011



L'ouverture des données publiques nourrit aujourd'hui de nombreux projets portés par l'Union européenne, l'Etat, des collectivités, notamment des grandes villes et des communautés urbaines. Cette problématique interpelle de longue date l'administration qu'elle soit d'Etat ou territoriale. Elle est pour les collectivités et les acteurs des territoires une exigence de transparence de l'action publique et suscite le développement de nouveaux services numériques au profit du développement économique, du développement social et culturel et de l'aménagement du territoire. L'ouverture des données publiques ne doit surtout pas être considérée comme un effet de mode. **Il s'agit d'une transformation profonde et durable de l'usage des données, de l'émergence des services numériques au service de l'action publique, des citoyens et des entreprises (les tiers utilisateurs).**

C'est en se plaçant à l'échelle régionale que ce document entend expliciter les concepts, les méthodes et les outils nécessaires pour construire une démarche de «libération des données publiques».

#### Celle-ci s'articulera autour de 4 idées clés :

- 1 - La mise à disposition d'informations utiles et réutilisables s'inscrit dans une action politique et non dans une simple approche technique ; elle accompagne une stratégie d'ensemble de la gestion de données et du rôle de l'acteur public dans ce domaine.
- 2 - L'utilité de l'information et des données est étroitement liée aux dynamiques de décloisonnement et de transversalité : la politique de données et de services s'inscrit dans des coopérations régionales, interrégionales, nationales, qui doivent être organisées et maîtrisées.
- 3 - La mise à disposition des données ne peut s'exonérer de choix en ce qui concerne le service public et la garantie d'un accès égal et pérenne aux données considérées comme essentielles.

4 - La diffusion, le partage de données et les coopérations et mutualisations régionales sont un nouvel aspect important de l'aménagement numérique du territoire.

C'est dans ce cadre général que doivent se poser les questions juridiques, économiques et techniques sur la base de l'évaluation réaliste des investissements et des coûts, du respect des lois et règlements, des politiques de réutilisation des données et du choix des licences.

Rappelant l'abondante documentation récemment accumulée sur l'ouverture des données publiques, l'étude s'oriente vers des propositions de moyen et long terme, établies à partir d'observations et d'entretiens avec des responsables des Conseils régionaux, des Communautés Urbaines, des groupements constitués sur des domaines liés à l'information géographique, des entreprises et des experts.

Le problème tel qu'il se pose aux décideurs est complexe et multiforme du fait de la diversité des données, des modalités de leur extraction et de leur mise à disposition. Il s'agit d'un phénomène nouveau. Hors les grands éditeurs publics spécialisés (JO, INSEE etc.), la diffusion d'information n'était pas du ressort de l'acteur public et n'était pas incluse dans le service public. Dès lors, de multiples questions se posent au niveau institutionnel, juridique, économique, technique auxquelles l'étude se propose de répondre..

L'année 2011 a été riche en matière de réutilisation des données publiques, de nombreux débats ont eu lieu, de nouveaux textes ont été promulgués, de nouvelles structures se sont mises en place. C'est un domaine en évolution rapide et l'essentiel va être, pour tous les acteurs, d'assurer une veille tout en s'engageant dans des démarches dont il faut évaluer la pertinence et l'opportunité.

C'est pour y contribuer que la présente étude a été lancée.

Alain Bénateau

Président de l'Observatoire des Territoires Numériques

## Remerciements

L'OTeN remercie tous ceux qui, membres et non membres, partenaires ou non de l'OTeN, organismes publics, associations et sociétés, ont bien voulu rencontrer les consultants et fournir documents et informations.

La liste des organismes et des personnes suit.

Par ailleurs, l'OTeN salue particulièrement les travaux menés par la Fondation Internet Nouvelle Génération (FING), l'Agence Aquitaine Europe Communication (AEC) et par le Groupement Français de l'Industrie de l'Information (GFII) notamment. Une grande partie des analyses et informations sont partagées avec ces associations, dont l'OTeN est partenaire, et qui ont mis en place des groupes de travail spécialisés et mènent des actions de sensibilisation, de formation et d'accompagnement de projets remarquables.

Enfin, l'étude a aussi utilisé les principaux sites où s'organise le débat sur la mise à disposition et la réutilisation des données publiques. Ces sites sont référencés dans le texte et leur liste figure en annexe.

De nombreux responsables ont accepté de consacrer du temps à des entretiens et des échanges avec l'équipe. Qu'ils en soient ici remerciés.

Remarque : quelques personnes n'ont pas été rencontrées mais interviewées par téléphone.

### Personnes rencontrées pour l'étude, dans les Régions

#### **Aquitaine**

Michel Eimer, *Conseil régional, - délégué régional TIC*

Eugénie Michardière, *Conseil régional - chargée de mission « données publiques »*

#### ▣ **Auvergne**

Philippe Wimart-Rousseau, *Conseil régional - Directeur du cabinet du Président*

Frédéric Deneux, *Centre régional auvergnat d'information géographique (CRAIG) – Directeur*

Sébastien Gaillac - CRAIG - *Chargé de mission Aménagement numérique et SIG*

#### **Bretagne**

Thomas Renault, *Conseil régional - Chef du service du développement numérique*

Michel Briand, *Brest Métropole océane - Vice Président*

Xavier Crouan, *Rennes Métropole - Directeur Information et Innovation Numérique*

Céline Faivre, *e-Mégalis - Directrice générale*

#### **Champagne-Ardenne**

Philippe Ricard, *Conseil général de l'Aube - Directeur des Systèmes d'information*

#### **Corse**

Eric Ferrari, *Collectivité Territoriale de Corse - Chef de la Mission des Technologies de l'Information pour la Corse*

#### **Ile de France**

Yannick Landais, *ARTESI - Directeur*

#### **Languedoc-Roussillon**

Jean-Marie Bourgogne, *Ville de Montpellier - Programme Montpellier Territoire numérique – directeur du programme*

Marie Brackers de Hugo, *Montpellier Territoire numérique - économiste, évaluation des politiques publiques*

Gilles Marc, *Montpellier Territoire numérique - community manager, innovations horizontales*

Jérémy Valentin, *Montpellier Territoire numérique - géographe*

- Midi-Pyrénées**
- Michael Combes, Grand Toulouse - Direction des Systèmes d'Information - Directeur Adjoint
- Sandrine Mathon, Mairie de Toulouse et Grand Toulouse - correspondante informatique et libertés
- Christine Gennaro, ARDESI - chargée de mission Internet public et citoyen
- Laurent Plainecassagne, Association pyrénéenne d'économie montagnarde (APEM) - Directeur

#### **Pays de la Loire**

Arnaud du Crest, Conseil Régional - Directeur de la Mission Observatoire

#### **Picardie**

Emeline Gasnier, Conseil régional - Mission des services au public, chef de projet Développement des services numériques

Cédric Ansard, Conseil régional - Direction de la prospective, de la programmation et de l'évaluation - responsable du département SIG et évaluation

Fabien Delanoy, Conseil régional - Responsable étude et développement informatique

#### **Provence Alpes-Côte d'Azur**

Stéphane Martayan, Conseil régional - Direction de l'Economie Régionale, de l'Innovation et de l'Enseignement Supérieur – chef de la Mission TIC

Sophie Tendeiro, Conseil régional - Chef de projet Services numériques innovants et mutualisation des données publiques

#### **Réunion**

Joëlle Nonet, Conseil régional - Chef du bureau du tourisme

#### **Rhône Alpes**

Cédric Lambert, Conseil régional - Direction des transports et des TIC, chargé de mission TIC

Isabelle Soriano, Conseil régional - Direction des politiques territoriales - responsable du service développement numérique

#### ▣ **GFII : Groupes « Données publiques » et « Données publiques des collectivités »**

Ruth Martinez, Déléguée générale du GFII

#### **Animateurs**

Bernard Marx, GFII, membre du COEPIA

Denis Berthault, LexisNexis - Directeur du développement des contenus en ligne

#### **Avec l'expertise de :**

Maître Jean Martin, avocat, expert auprès de la Commission européenne, professeur à Paris Dauphine

...et les nombreux membres du groupe de travail.

#### **FING**

Daniel Kaplan, Délégué général

Amandine Brugière, Chef de Projet « Partage et réutilisation des données publiques »

Thierry Marcou, Chef de Projet « CitéLabo »

Charles Népote, Chef de Projet « Partage et réutilisation des données publiques »

#### **Industriels et sociétés de services**

##### **Aceris numérique**

François Lamotte, Président

Martin de Mijolla, Directeur associé, co-fondateur

Bruno Perrin, Co-fondateur

Berger-Levrault

Guy Beaudet, Directeur de la stratégie

##### **Furet Company**

- ▣ Cédric Levret, Fondateur et directeur

**GeoAd**

Paul Desjonquière, Co-fondateur et directeur associé

**Logica**

Stéphane Batiot, Directeur associé en charge du secteur public

**Mensia**

Camille Depley, Consultante en charge de missions avec les collectivités publiques

**SOPRA**

Jean-Jacques Barbier, Directeur de projet auprès de la Mairie de Paris

**Institutionnels nationaux**

**Assemblée des Chambres Françaises de Commerce et d'Industrie**

Ludovic Bour, Responsable du département Innovation et Intelligence Economique

**Assemblée des Départements de France**

Jean – Pierre Quignaux, Responsable Aménagement numérique

**Délégation à l'information légale et administrative**

Régis Pérol, Responsable du Pôle veille et expertise

**Institut géographique national**

Alain Chaumet, Chargé de mission partenariats Géoportail

▣ **A propos de l'OTeN**

L'Observatoire des Territoires Numériques a pour objectif de favoriser le développement durable des territoires dans la société en réseau. C'est un lieu privilégié d'échanges et d'information sur les technologies et usages du numérique et sur les enjeux de la société de l'information pour les territoires.

L'OTeN est une association qui s'adresse directement ou indirectement à tous les acteurs des territoires : régions, départements, communes et intercommunalités. Son Conseil d'Administration regroupe les principales associations d'élus ou de collectivités.

IRIS (Initiatives Régionales Innovations et Stratégies) est un projet d'intelligence collective animé par l'OTeN et porté par des régions pilotes en matière de développement numérique, avec le soutien de la Caisse des dépôts et consignations et du réseau des CCI.

La base de données de projets numériques innovants d'IRIS comprend plus de 2000 monographies qualifiées et détaillées, régulièrement mises à jour, complétées par des études régionales et thématiques.

IRIS constitue un outil de projection au service des politiques numériques territoriales.

[www.oten.fr](http://www.oten.fr)



# 1 Données publiques : un problème ancien, une situation nouvelle

## 1.1 Prolégomènes

En charge du développement économique et de l'aménagement numérique du territoire, les conseils régionaux et leurs partenaires engagés dans le développement économique et l'aménagement des territoires sont confrontés à un nouveau challenge celui de l'ouverture des données publiques. Ils sont ainsi conduits à se positionner sur ce thème en relation avec les autres niveaux de collectivités territoriales en sachant que la notion de donnée publique pour être homogène doit agréger plusieurs niveaux administratifs et échelons territoriaux.

### 1.1.1 Explosion d'un mouvement « open data » (mise à disposition des données publiques)

Le mouvement « open data » trouve ses exemples emblématiques dans les pays anglo-saxons<sup>1</sup> même si l'environnement juridique, la définition même du service public, sont profondément différents avec la France – ce qui ne veut pas dire qu'il ne faille pas observer avec intérêt les « data.gov » divers, mais aussi qu'il n'est pas possible de transposer sans précaution ces expériences.

Cependant, ce mouvement vers la mise à disposition des données publiques traduit un mouvement de fond qui bouleversera à terme la vision du service public.

<sup>1</sup> Même dans d'autres développements, cette référence est loin d'être toujours pertinente. Sans nier le dynamisme des Etats-Unis, leur rôle dans le développement et même l'appropriation des techniques est très souvent surestimé par un courant de pensée systématiquement fasciné.

### ■ Des facteurs profonds

Les facteurs principaux des évolutions en cours se trouvent, en effet, dans des faits de société et plus particulièrement dans l'exigence croissante des citoyens d'accès à l'information et dans son souci d'avoir une meilleure transparence de l'action publique.

Cette montée en puissance du citoyen, que les auteurs anglo-saxons appellent son "empowerment" est incontestable dans tous les domaines – que ce soit pour les affaires publiques ou pour les activités personnelles, la consommation etc. Elle se traduit aussi par de nouvelles formes d'organisation, d'association, auxquelles les Technologies de l'Information et de la Communication fournissent de puissants moyens pour l'intervention et le débat, avec comme levier l'Internet, le Web 2.0, les réseaux sociaux ... Les services publics, doivent s'adapter à cette demande sociale de transparence et de communication.

Technologiquement l'évolution rapide des terminaux numériques connectés, ordinateurs, téléphones mobiles, tablettes et des services associés irriguent désormais une part de plus en plus importante de la population et investissent tous les secteurs d'activité (professionnels, personnels..).

Ainsi l'appropriation des nouveaux outils et des nouveaux services par le public est parallèlement de plus en plus rapide<sup>2</sup>. En outre, la diffusion des outils mobiles se fait d'abord dans le public, et non dans les entreprises, même s'ils sont utilisés dans toutes les activités : ce sont des objets personnels multi-usages.

Il faut noter une corrélation entre la problématique des données publiques et celle de l'environnement et du développement durable. Cela s'observe

<sup>2</sup> Pas seulement chez les jeunes. On constate une montée rapide des usages chez les retraités et les personnes âgées; la pénétration de la micro-informatique dans les entreprises puis dans les ménages est déjà ancienne et a maintenant touché un grand nombre de personnes; par ailleurs, les tablettes, les photos de famille sont des applications de masse.

dans beaucoup des demandes et des services qui naissent en 2011, dans le mouvement open data et dans les orientations et actions publiques, au premier rang desquelles l'Agenda 21 et la directive européenne Inspire (cf. «géographie» dans le chapitre Données). L'évolution vers un développement durable suppose en effet une prise de conscience et une prise de contrôle par de nombreux acteurs de terrain, une information fine sur les territoires et une circulation accrue des données.

#### ■ Les acteurs publics producteurs et détenteurs d'information

Aujourd'hui en France l'acteur public est le principal collecteur et producteur de données et de documents. C'est vers lui que se tournent les citoyens, les groupes et associations, les entreprises afin qu'il leur fournisse les informations dont il peut disposer. Si on ajoute à l'Etat et aux collectivités l'ensemble des organismes qui dépendent d'eux ou qui assurent en délégation des missions de service public - ce qui est le cœur de la définition de la donnée publique -, il ne fait pas de doute que cette masse d'informations si elle est valorisée constitue un élément susceptible de susciter l'innovation, l'apparition de nouveaux services, de nouveaux projets de développement et d'attractivité pour les territoires.

Dans de nombreux domaines, les données publiques sont les seules disponibles et sont riches de valeurs : géographie, environnement, météorologie, enseignement, formation, vie culturelle, population, institutions, entreprises et activité économique, commerces, droit et réglementation, marchés publics et privés dans de nombreux cas, vie de la cité, transports, travaux, infrastructures, système de santé et système social, tourisme, sciences et techniques...

Si les conditions d'accessibilité et de diffusion sont très diverses pour toutes ces informations, désormais les questions de l'accès puis de la réutilisation sont posées et ne peuvent plus être ignorées par le «service public» quel qu'il soit.

#### ■ L'espoir d'une dynamique vertueuse

Le mouvement actuel "open data" affirme l'impérieuse nécessité de rapprocher deux univers :

- celui des données et documents de tous types recueillis ou produits par les acteurs publics,
- celui de nouveaux acteurs de l'innovation qui peuvent développer de multiples applications utiles au citoyen.

Ces deux univers peuvent construire une dynamique vertueuse entre les collectivités qui se libèrent de l'élaboration de services numériques de plus en plus diversifiés pour se consacrer à la production de données et de jeunes sociétés ou mouvements citoyens qui ont une approche directe et innovante des outils informatiques actuels, du Web2, des réseaux sociaux et qui oseront développer des applications sur un corpus de données accessibles et fiables.

L'ouverture des données est donc susceptible, pour ses promoteurs, de permettre non seulement des applications très utiles pour tous mais en même temps de favoriser l'émergence d'un tissu d'entreprises, dont certaines pourront devenir des acteurs importants de l'économie numérique.

Ainsi, le mouvement de l'opendata est animé par deux moteurs : le désir de transparence démocratique et la possibilité de développer un secteur économique d'information. Si cette approche est naturelle dans la culture des Etats-Unis, il faudra en France vaincre quelques réticences et convaincre sur l'opportunité de développer ce type de partenariat public privé.

#### 1.1.2 Un phénomène et une réflexion déjà anciens

[cf. les § traitant des structures mises en place par l'Etat et des questions juridiques]

Les interrogations que se posent aujourd'hui les collectivités ont fait l'objet de réflexions et de propositions depuis longtemps au niveau des administrations centrales : la diffusion des données publiques est une vieille question, même si la généralisation non seulement des systèmes de consultation mais aussi des outils de production de données et documents et la numérisation des procédures et enquêtes lui donne une dimension nouvelle.

Dans les années 80, le problème des données publiques était celui des relations entre quelques grands organismes publics, un nombre limité d'utilisateurs accédant aux banques de données et un nombre limité d'offres. On parlait alors de «pétrole gris» pour les banques de données.

La situation évoluait cependant rapidement avec la télématique, autrement dit le Minitel. En ces temps, la France était en avance sur la diffusion électronique de l'information. Cela explique d'ailleurs le grand nombre de rapports qui ont été produits sur le sujet : d'une part en raison de l'importance du secteur public en France et du pouvoir régalié de l'Etat, d'autre part en raison de l'accès d'un public beaucoup plus large. Il y avait aussi une motivation très pratique : le système de paiement de la télématique faisait apparaître en même temps enjeux économiques et nouvelles sociétés.<sup>3</sup> Internet puis le mobile ont provoqué un nouveau saut.

Les premiers débats sont donc apparus avec les banques de données des grands acteurs nationaux producteurs d'information en raison de leur mission de service public : IGN, INPI, INSEE, JO, Météo essentiellement. Pour les autres administrations, le problème posé était plutôt de leur faire prendre conscience des gisements de données qu'elles détenaient. Les batailles sur les licences, les problèmes d'exclusivité, de concession etc. datent des années 80. Il suffit d'ailleurs de constater le nombre impressionnant de

<sup>3</sup> On a oublié cette époque. L'auteur du présent document, en mission aux Etats-Unis, rencontra des acteurs comme les banques qui enviaient notre système de banque en ligne...

rapports sur le sujet à partir de 1997. L'accélération s'est effectivement faite avec la progression d'Internet mais elle avait déjà été nettement amorcée en France par les services Minitel. Successivement, l'Etat confiait le dossier à la Commission d'accès aux documents administratifs (CADA – créée en 1978 pour l'accès des citoyens aux documents publics), puis créait l'Agence du patrimoine immatériel de l'Etat (APIE) en 2007, le Conseil d'orientation de l'édition publique et de l'information administrative (COEPIA) en 2010 et mettait en place en 2011 la mission Etalab chargée de piloter l'ensemble des administrations sur la question des données publiques et de créer un répertoire unique interministériel<sup>4</sup>.

L'évolution législative s'est elle-même accélérée avec la Directive européenne 2003/98 du 17 novembre 2003 concernant la réutilisation des informations du secteur public, transposée en droit français par l'Ordonnance n° 2005-650 du 6 juin 2005 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques et le décret n°2005-1755 du 30 décembre 2005. Ensuite, la Directive 2007/2/CE du 14 mars 2007 dite «INSPIRE» a joué un rôle très important. Transposée en droit français par l'Ordonnance n° 2010-1232 du 21 octobre 2010 portant diverses dispositions d'adaptation au droit de l'Union européenne en matière d'environnement, elle impose et organise la création d'infrastructures de données géographiques pour toutes les activités et informations ayant un impact sur l'environnement. *(Les aspects légaux et les textes sont abordés au Chapitre 3).*

### 1.1.3 Open data et open government

Les partisans des "open data" commencent d'ailleurs à parler d'"open government"

<sup>4</sup> S'il s'impose aux administrations centrales et déconcentrées, le portail peut accueillir, si elles le souhaitent, des références et des liens vers les données des collectivités territoriales.



## ■ *Démocratie et responsabilités dans une société de l'information*

L'ouverture, la mise à disposition et plus largement une politique de circulation des données constituent un nouveau domaine d'intervention pour les collectivités, correspondant à l'entrée dans la "société de l'information" et plus concrètement à la disponibilité de réseaux et à leur appropriation générale. En même temps que d'une opportunité économique, il s'agit d'une nouvelle exigence démocratique. Elle introduit un nouveau domaine de relations entre les citoyens, les acteurs économiques et les acteurs publics. La mise à disposition de données fines, la possibilité de corriger et critiquer donnée à tous, la capacité et même la nécessité de s'appuyer sur ceux qui étaient encore hier des administrés pour produire eux-mêmes des données et interagir, entraînent de nouvelles responsabilités et nécessitent inévitablement une régulation. Mettre à disposition des données, dont certaines sont d'ailleurs directement issues des citoyens ou des différents acteurs du territoire, ne peut se faire n'importe comment. Contrairement à ce que des approches militantes parfois simplistes peuvent laisser penser, les élus politiques et les fonctionnaires ont le devoir de veiller à ce que la circulation des données, leur utilisation ne lèsent pas tel ou tel indûment, ne créent des inégalités, n'aillent à l'encontre des choix politiques démocratiquement déterminés.

## ■ *Les difficultés spécifiques des collectivités : l'évolution des tâches*

Pour les entreprises, l'époque est aux évolutions rapides des marchés et des techniques, ce qui contraint à la flexibilité des systèmes d'information. Inversement, l'image des services publics reste encore parfois un peu marquée par le conservatisme et la rigidité. Rien n'est plus faux, en tout cas pour les collectivités territoriales, même si, effectivement, certains formulaires et certaines organisations peinent à évoluer.

En effet, toutes les collectivités sont confrontées à l'évolution très rapide

et surtout brutale de leurs compétences et de leurs missions, qui changent – et s'accroissent – en fonction des textes et des décisions politiques. C'est ainsi que le transfert de l'action sociale aux départements, l'intervention inévitable dans la gestion de la dépendance bientôt, introduisent de nouveaux mécanismes et génèrent de grandes masses de données. Il en a été de même avec l'aménagement numérique ou avec l'évolution des transports.

### 1.1.4 Des principes incontournables

#### ■ *Des choix politiques avant les débats techniques et économiques*

L'important, en dernier ressort, ce sont les usages et donc les utilisateurs et les services. L'époque est technophile et les outils informatiques tendent à modifier les comportements. Il n'en reste pas moins que ce n'est pas l'outil qui dicte une politique. L'important pour un élu ou un administrateur, c'est que les évolutions techniques servent une politique.

La question centrale est la demande d'information par les citoyens et par les entreprises, avivée par le support technique. En ce sens, le mouvement des données ouvertes correspond bien à une exigence politique importante. Le succès et les controverses autour de Wikileaks montrent qu'il y a un désir de transparence, risques et excès possible, incertitude dans les réponses politiques et sociales. Les débats sont parfois caricaturaux et il faut éviter de tomber systématiquement. Aussi importants soient les horaires de bus, ils n'ont pas grand chose à voir avec les courriers d'ambassades à propos d'une guerre...

Par ailleurs, il faut aussi éviter de centrer le débat initial sur les modalités de la diffusion des informations et en particulier sur les questions de gratuité et de rémunération. Les aspects économiques sont importants mais ils sont dépendants des décisions sur le fond. Les budgets en jeu ne sont pas

tels que la question du financement prenne le pas sur les enjeux sociaux et économiques de la diffusion des données et sur les choix politiques de principe sur la place du service public et le rôle de la collectivité.

*L'important est de mener une réflexion sur le service public et d'y intégrer une réflexion sur l'information.*

*C'est de distinguer, non en fonction d'opportunités, mais sur des principes pérennes établis par les politiques, ce qui doit être assuré par la collectivité, ce qui doit être assuré mais peut être délégué, ce qui peut être diffusé et utilisé par les citoyens et les entreprises mais contrôlé, ce qui est d'usage totalement libre.*

#### ■ *La construction des conditions techniques*

Devant des évolutions techniques toujours rapides – et qui vont continuer de l'être, il faut considérer des invariants, bien connus des spécialistes de l'information et des systèmes d'information.

Du point de vue technique, rien ne peut se faire sans que les infrastructures soient en place (aussi bien les câbles et autres réseaux que, pour la circulation des données, les serveurs, les logiciels de base, les référentiels communs) Or, l'évolution des infrastructures est bien plus lente que celle des terminaux, comme le savent bien les collectivités territoriales responsables de l'aménagement numérique du territoire.

L'élément critique dans l'ouverture des données publiques, ce sont les données. La mise à disposition, et même la simple diffusion, confrontent le gestionnaire non seulement aux questions de formats mais à celles ayant trait à la constitution même des bases, à leur liaison à d'autres, aux interfaces etc. Une base de données est un objet complexe dont la tenue à jour doit être organisée et suivie. Le temps de définition et de constitution des bases est tout autre que celui de renouvellement des outils de consultation.

*[cf., sur les infrastructures et les données, les § sur les répertoires et sur la technique]*

#### □ 1.1.5 Le point de vue des territoires et la démarche

*[Ce § introduit aux chapitres relatifs à l'économie, au rôle des Régions et à la mise en œuvre]*

##### ■ *Une nouvelle étape de l'aménagement numérique du territoire*

Les TIC sont maintenant un élément essentiel de l'économie – directement par la création d'activités et surtout par leur impact pour l'ensemble des acteurs économiques et des citoyens.

Les collectivités, et notamment les Conseils régionaux, l'ont très rapidement compris et elles ont travaillé, et souvent bataillé, pour l'aménagement numérique du territoire, désormais outil essentiel et inséparable de l'aménagement du territoire.

Or, après l'aménagement numérique des "tuyaux", la deuxième étape est celle des services. Elle nécessite des infrastructures matérielles et logicielles complémentaires mais elle suppose surtout d'accompagner des initiatives multiples en veillant à ce que les services utilisant les données publiques s'inscrivent dans une politique générale en direction de la population et des entreprises.

##### ■ *Articuler les visions des différentes collectivités et acteurs publics*

La problématique renouvelée des données publiques est actuellement au centre de développements importants dans deux domaines :

- L'ouverture des données des villes et des communautés urbaines (Bordeaux, Brest, Nantes, Montpellier, Paris, Rennes)<sup>5</sup>
- La création de portails d'information géographique régionaux (Aquitaine, Auvergne, Bretagne et bientôt d'autres), associant Etat, collectivités et autres institutions régionales.

<sup>5</sup> Ordre alphabétique... (les pionniers ayant été les Bretons)

- Élément de l'aménagement du territoire, de l'aménagement numérique et du développement économique, l'utilisation – et la réutilisation – des données publiques interpelle directement les Régions. Or, c'est aussi à ce niveau régional que se placent de plus en plus des administrations déconcentrées et des structures économiques importantes. Les Agences régionales de santé ou la réorganisation des Chambres de commerce en sont les plus récents exemples. De même, la directive Inspire amène à la création ou au renforcement de structures de coopération au niveau régional entre administrations d'Etat et acteurs des territoires.

Par ailleurs, c'est dans le rapprochement de données issues de sources différentes que se trouvent les pOTeNtiels de valeur ajoutée. C'est particulièrement vrai pour le territoire régional où interviennent les différents niveaux de collectivités, les services déconcentrés de l'Etat, les organismes spécialisés déjà mis en place par les collectivités elles-mêmes. De plus en plus, aussi, l'égalité d'accès à des services sur le territoire va conduire à des mises en relation entre le niveau de la vie quotidienne, où dominent communes et communautés urbaines, le niveau de l'intervention sociale qui est celui des départements et le niveau régional. Cette coopération, parfois difficile, est indispensable s'agissant d'information, domaine où le décloisonnement est critique et où les intervenants des territoires doivent aussi travailler et négocier avec l'Etat et ses représentants.

Ainsi, dans de nombreux domaines, l'action des autorités publiques n'est plus celle d'un régulateur / contrôleur ou d'un acteur exclusif mais celle d'un coordinateur / animateur. Qu'il s'agisse de transport, de tourisme, d'action sociale etc., il s'agit de stimuler la naissance d'écosystèmes et de les réguler.

Le caractère local et régional incite à la production de services beaucoup plus riches directement exploitables par les citoyens et par les différentes professions et donc plus visibles d'un point de vue pratique et d'un point de vue politique que les grands services nationaux essentiellement

professionnels. Cependant, il sera toujours important de distinguer les services de proximité et les services d'échelle plus large et notamment régionale d'une part, les services pour le public et les services professionnels d'autre part.

#### ■ *Oser des initiatives et raison garder*

La progression des outils numériques, depuis qu'elle a commencé, s'effectue entre deux dangers : d'une part la prudence excessive, avec la sous-estimation du rythme de développement des techniques et de leur appropriation sociale, d'autre part l'enthousiasme naïf et l'effet "bulle". Pour des élus et des collectivités, les risques sont soit du retard par rapport à de nouveaux développements alors assurés par d'autres, soit de la déception.

Les exemples abondent malheureusement des dégâts coûteux de l'effet bulle. Même avant l'informatique, il y a eu l'enthousiasme pour le câble citoyen. Ce furent aussi les villes misant sur de grands programmes télématiques, suivies par 36 unités de production de micro-ordinateurs, oubliées pour passer à l'ère des villes dédiées au très haut débit fixe ou mobile. Il ne s'agit pas de critiquer a posteriori tous ces efforts, ni de considérer que seuls les attentistes technophobes ont finalement raison, mais de tirer une leçon: les techniques d'information jouent leur rôle mais ne peuvent informer que sur ce qui existe et développer des activités nouvelles que dans un terreau propice. A elles seules, elles ne peuvent créer partout de nouvelles Silicon Valley. Elles doivent être intégrées dans l'économie d'ensemble, dont elles sont un outil important.

Pour favoriser le développement de nouveaux services, au dessus des infrastructures, il faut s'appuyer sur les dynamiques propres du territoire.

Il faut aussi rappeler qu'il n'y a pas de «free lunch» (pour reprendre une expression anglo-saxonne) : la gestion de données et de services nécessite des investissements et surtout des coûts de fonctionnement – que ceux-

ci soient assurés par le producteur de données ou par d'autres. Il s'agit en outre de s'engager dans une voie où des demandes nouvelles apparaîtront, où de nouveaux supports techniques exigeront des adaptations : c'est un processus qui commence.

Enfin, toute diffusion de données présente des risques, qu'il faut mesurer. Ici aussi, il s'agit de choix politiques dont les effets doivent être suivis.

## 1.2 Définitions : des données à la réutilisation

Les débats sur la réutilisation et ses aspects financiers ne doivent pas masquer le fond de la question: il s'agit de données et non de documents divers; il s'agit d'ouverture, non de publication ou d'accès; il s'agit enfin de données publiques, c'est-à-dire de données collectées ou produites dans le cadre d'une mission de service public. On voit immédiatement là que c'est bien le cadre d'exercice et le périmètre du service public qui sont en cause.

*[Les questions abordées dans ce § sont principalement développées dans le chapitre «Données» et dans le chapitre «Réutilisation et aspects juridiques»]*

### 1.2.1 Données

Il s'agit surtout de données, et pas seulement de documents, même si une différenciation stricte est difficile à mettre en place et même si les textes juridiques relatifs aux données généralisent en fait ce qui avait été envisagé ou réalisé pour les documents. Certes, un tableau statistique publié indépendamment, ou un formulaire de déclaration fiscale sont des documents qui contiennent essentiellement des données. Cependant, le document est un outil de présentation, dont la forme, l'auteur, le processus de circulation sont spécifiques. De leur côté, les données représentent des ensembles structurés, ordonnés ou ordonnables d'informations, pas seulement des chiffres mais des séries qui peuvent être des suites

d'évènements ou de relevés organisés selon un cadre précis, ou des ensembles de descripteurs graphiques et de chiffres pour les systèmes d'information géographique. Souvent, des documents sont en pratique des containers, des bases de données sur papier ou numérisées (par exemple les registres d'état civil). La différence est particulièrement importante en termes d'usage et d'accessibilité. Des données "temps réel" sur les horaires de transport ne se traitent pas comme des relevés de délibérations.

Dans le mouvement actuel, c'est bien d'abord des données structurées qu'il est question.

Or, au delà des données publiques et de leur réutilisation externe, c'est de l'importance de la gestion de ces données que les services publics, comme les entreprises, prennent progressivement conscience.

La collecte systématique d'informations à l'occasion de procédures administratives dans le secteur public, ou d'activités techniques et commerciales dans le secteur privé, a produit des silos renfermant des données que les autres services ignoraient ou ignorent encore et que des partenaires, administrés, clients, pourraient eux aussi exploiter utilement, souvent même dans l'intérêt du «producteur» des données. La construction de systèmes d'information a permis la circulation des données dans de multiples applications, avec une cohérence aussi bien pour la gestion que pour les relations extérieures. Les bases de données centrales ou communicantes, la gestion des données critiques sont devenues les points forts des systèmes. Au demeurant, cette évolution est loin d'être systématique et achevée dans le secteur privé.

En pratique, c'est souvent la relation avec des acteurs extérieurs qui a stimulé le mouvement vers une rationalisation de la gestion des données internes. Elle est ainsi nécessaire dans les entreprises pour parvenir en même temps à une gestion coordonnée des sous-traitances, des ressources internes et des circuits de distribution, jusqu'à la relation finale avec le client pour le suivi

des biens d'équipements et en général de la clientèle.

Des motivations et des mécanismes similaires vont se développer pour les données manipulées par et pour les services publics.

*Les responsables informatiques le disent : ouvrir les données à l'extérieur, c'est d'abord les organiser à l'intérieur.*

Il faut de toutes façons aussi les partager. Cette évolution correspond bien au rôle croissant des données dans un monde complexe, et des acteurs privés commencent eux aussi d'envisager un partage des données.

Cette démarche est d'ailleurs illustrée par des sociétés délégataires de services publics, par exemple pour l'organisation du cycle de l'eau ou les nouveaux réseaux intelligents de distribution d'électricité. Ceux-ci supposent un maillage très fin du territoire et l'utilisation de masses impressionnantes de données, que seul le partage avec de nombreux partenaires permet à terme d'envisager – avec les utilisateurs eux-mêmes, les collectivités, les opérateurs institutionnels. Ici encore, la rencontre entre intervention des usagers, ouverture et partage des données et développement durable est manifeste (cf. infra «Données»).

### 1.2.3 Données publiques : définition et enjeu

Les textes qui définissent les données publiques souffrent de la difficulté de définir des catégories précises et de distinguer clairement, entre autres, données et documents. Les ambiguïtés se retrouvent tout au long des débats et des travaux en cours. Ce sont les documents qui restent la référence dans les textes. Ce qui n'était pas fondamentalement gênant en dehors du monde numérique – car les citoyens entre autres ne demandaient l'accès qu'aux documents – le devient pour réguler l'ouverture des systèmes informatiques.

La définition de base figure dans la Loi CADA de 1978 :

Sont considérés comme documents administratifs, au sens des chapitres I, III et IV du présent titre, quels que soient leur date, leur lieu

de conservation, leur forme et leur support, les documents produits ou reçus, dans le cadre de leur mission de service public, par l'Etat, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission. Constituent de tels documents notamment les dossiers, rapports, études, comptes rendus, procès-verbaux, statistiques, directives, instructions, circulaires, notes et réponses ministérielles, correspondances, avis, prévisions et décisions.

Dans son rapport de 1999, Dieudonné Mandelkern<sup>6</sup> précisait :

“Données publiques” : «données collectées ou produites, dans le cadre de sa mission, par un service public sur fonds publics.», au sens de la circulaire du 14 février 1994 relative à la diffusion des données publiques. À cette définition, l'Atelier a ajouté la condition d'avoir vocation à être diffusée, afin de tenir compte des secrets légalement protégés. Il faut noter que cette définition s'applique quelle que soit la modalité d'exercice du service public : régie, EPA ou EPIC, délégation de service public.

On note ici deux évolutions, qui feront parfois débat. D'une part la référence au financement sur fonds publics, d'autre part la vocation à être diffusée.

La Loi CADA procédait en excluant les documents qui ne pouvaient être diffusés.

Enfin, la directive européenne de 2003, devant prendre en compte la diversité des législations, fixe ce qui sera la référence pour les différents pays. La transposition française, qui a étendu sur certains points la directive, ne s'en écarte évidemment pas.

<sup>6</sup> Président de l'Atelier «Des moyens nouveaux au service de la diffusion des données publiques.» dans le cadre de la Mission «L'État et les nouvelles technologies de l'information et de la communication.» - Rapport au Premier ministre.

La présente directive fixe un ensemble minimal de règles concernant la réutilisation et les moyens pratiques destinés à faciliter la réutilisation de documents existants détenus par des organismes du secteur public des États membres.

[...]

La présente directive ne s'applique pas :

- aux documents dont la fourniture est une activité qui ne relève pas de la mission de service public dévolue aux organismes du secteur public concernés en vertu de la loi ou d'autres règles contraignantes en vigueur dans l'État membre ou, en l'absence de telles règles, en vertu des pratiques administratives courantes dans l'État membre concerné;

... (Suivent les exclusions pour autres motifs comme la protection de la vie privée – cf. chapitre Données)

La directive précise :

Aux fins de la présente directive, on entend par "organismes du secteur public", l'État, les collectivités territoriales, les organismes de droit public et les associations formées par une ou plusieurs de ces collectivités ou un ou plusieurs de ces organismes de droit public.

Comme le montrent les hésitations et oppositions qui ont surgi et continuent de surgir, les frontières sont difficiles à tracer. Il est probable que le cœur du problème n'est pas dans un affinage très précis de la notion, mais dans une réflexion plus large sur le service public et les informations associées. Les préoccupations concrètes de terrain auront de plus en plus tendance à glisser vers des réflexions plus larges sur les données.

#### 1.2.4 Accès, publication, réutilisation

Trois notions doivent être distinguées. Elles sont rappelées ici, de façon

résumée, non rigoureuse du point de vue juridique (cf. le chapitre Réutilisation et questions juridiques et les documents de base, dont l'étude d'Aquitaine Europe Communication) :

- L'accessibilité est décidée par l'autorité publique responsable, à partir de la Loi et des règlements relatifs aux documents en cause. Elle est contrôlée par la Commission d'accès aux documents administratifs (CADA). Elle permet à toute personne ou organisme d'obtenir la communication de documents, indépendamment de l'usage qui en sera fait, si ces documents ne contiennent pas des informations non communicables (cf. les réserves dans le chapitre Données).

- La publication (par l'acteur public) peut se faire dans le cadre de procédures administratives (affichage de décisions, d'avis etc. sur un panneau), sur papier, sur journal d'annonces public ou privé (annonces de marchés...), soit sous forme électronique, en particulier sur le Web.

- La réutilisation est un usage particulier, par lequel un acteur, quelle que soit sa nature, publie à son tour des informations obtenues auprès de l'acteur public, dans un but commercial ou non. Il les intègre éventuellement à d'autres données et à des services mais ceci n'intervient pas pour qualifier la réutilisation.

Ces trois notions sont liées mais, si l'accessibilité conditionne la réutilisation, celle-ci peut être soumise à d'autres contraintes. La publication implique l'accessibilité, l'inverse n'étant pas vrai (notamment une personne peut avoir accès à des pièces la concernant mais non publiables). La publication sur le Web implique en pratique la réutilisation, qui, en France, est un droit opposable – c'est une extension de la directive, qui soumet la réutilisation à des règles mais ne l'impose pas. Cependant, il peut y avoir des limites, par exemple si certaines informations figurant sur le site sont grevées de droits d'auteur.

Les problèmes posés par la réutilisation du point de vue juridique sont

traités dans un chapitre particulier. Ils s'articulent autour de trois questions principales :

- les conditions juridiques de l'exercice du droit de réutilisation, inscrites dans les licences ;
- l'existence et le mode de détermination de redevances ;
- le périmètre des organismes qui doivent satisfaire aux demandes de réutilisation.

### 1.3 Une vue d'ensemble d'un paysage mouvementé

#### 1.3.1 Le mouvement open data

La nouvelle vague de la réutilisation des données publiques est portée par un courant qui appartient au mouvement plus large de liberté de circulation des objets immatériels : libre et open source pour le logiciel, mouvement et licences des Creative Commons pour les œuvres de l'esprit et la création, mouvement des archives ouvertes (Open Access) pour le libre accès aux articles scientifiques, open data pour les données - ces domaines se recoupant.

Ce mouvement est né des capacités d'échange et de regroupement par communautés sur le Web. Il a d'ores et déjà rencontré de très grands succès : part du libre en croissance très rapide dans les applications, développement lui aussi très rapide des Creative Commons, adoption du modèle des archives ouvertes par la plupart des universités et institutions scientifiques.

Pour les open data, des pôles associatifs se sont constitués en France, appuyés au départ sur les expériences et réalisations des Etats-Unis et du Royaume Uni.

- **Libertic** (<http://libertic.wordpress.com>), association nantaise, promeut « l'ouverture des données publiques, l'e-démocratie, le gouvernement 2.0 et d'accompagner notre territoire dans le développement et l'utilisation d'applications numériques d'utilité publique ». Libertic entend associer l'ouverture des données à l'économie sociale et solidaire.

- **Regards citoyens** (<http://www.regardscitoyens.org>) agit dans le domaine de la transparence du fonctionnement des institutions démocratiques.

Le pôle le plus important, où se retrouvent la plupart des acteurs, est la Fondation Internet Nouvelle Génération (FING – <http://fing.org/>). Elle mène réflexion et action soutenues, notamment dans le programme CitéLabos et dans l'accompagnement de groupes de travail dans plusieurs régions. (cf. dans la bibliographie le premier guide pratique de la FING sur le sujet et voir notamment les interventions du responsable du thème, Charles Népote).

De nombreux autres groupes, parfois informels, se constituent autour d'un thème ou comme membres français des groupes internationaux. Ainsi, OWNI travaille sur le "datajournalism" (<http://owni.fr>). La forte communauté mondiale OpenStreetMap, qui rassemble tous ceux qui veulent créer des cartes libres à partir de sources elles-mêmes libres et de saisie d'information par les citoyens, n'est pas constituée formellement en association mais son site annonce en juillet 2011 qu'elle cherche à se développer en rejoignant le chapitre français de la Fondation Geospatiale Open Source, OSGeo-fr (<http://www.openstreetmap.fr> - <http://osgeo.globe.org/drupal/node/3>)

Une caractéristique de l'ensemble du mouvement open data est que, à quelques exceptions près, ses axes de réflexion et ses motivations fortes s'organisent autour des thèmes de la transparence et de l'initiative citoyennes et de la dynamique des jeunes pousses ou des associations et enfin de la gratuité de l'accès aux informations pour consultation comme pour réutilisation. Cependant, la FING veille à maintenir un équilibre entre ces tendances et celles qui se font jour dans les grandes entreprises

intéressées par le mouvement et susceptibles d'y participer. Comme on le verra, leur nombre croît rapidement, elles aussi percevant l'importance de la circulation des données et l'impact économique pOTeNtiel de l'évolution en cours. (cf. *infra Données puis Economie*).

### 1.3.2 Les acteurs de l'information

A côté de la nouvelle vague de jeunes pousses surfant sur les applications mobiles et locales, un grand nombre d'acteurs, entreprises, groupements d'intérêt et institutions publiques, s'organisent, certains depuis les années 80, autour de la diffusion d'information électronique et en particulier d'informations d'origine publique.

Les sources principales : les administrations et organismes qui assurent le service public d'information : Direction de l'information légale et administrative (DILA, fusion du JO et de la Documentation française), IGN, INPI, INSEE, grands instituts de recherche publics – INRA, INRIA, INSERM...<sup>7</sup> S'y ajoutent quelques directions des administrations centrales directement productrices, à commencer par celles du ministère de l'Economie pour le cadastre ou les comptes des collectivités.

Un secteur d'activité s'est constitué autour de la circulation des données professionnelles, avec les Chambres de commerce, des bibliothèques, les prestataires informatiques et télécom et, pour la réutilisation, les éditeurs d'information professionnelle et fournisseurs de service, français et internationaux. Ce dernier groupe comprend des entreprises internationales comme de petits éditeurs spécialisés (cf. *la liste des membres du GFII sur le site gfii.fr*). Pour l'information grand public, le paysage a évolué rapidement avec le passage du Minitel au Web, à partir du groupe constitué par les

<sup>7</sup> Rappel : ils sont incités à entrer dans le mouvement des données publiques et de la gratuité mais ne sont pas soumis aux textes promulgués récemment sur le sujet, comme ne le sont ni les Etablissements publics industriels et commerciaux (que plusieurs sont) ni les Entreprises publiques.

acteurs nés avec la télématique.

Des associations représentent les différentes catégories d'intervenants – producteurs, prestataires de services, réutilisateurs. Elles se sont constituées sur un thème ou autour d'un support et d'un public.

- Le GFII - Groupement français de l'industrie de l'information (<http://www.gfii.fr>), rassemble les grands organismes publics et les sociétés de l'information électronique professionnelle, notamment dans les domaines du droit, de l'économie, de la propriété industrielle, de l'information scientifique et technique. Il a constitué un groupe de travail spécial sur les données publiques. Il a réuni sur le sujet un groupe de travail commun avec les autres associations citées ci-dessous. Le groupe données publiques comprend un sous-groupe sur les données des collectivités

- L'ACSEL - Association de l'économie numérique (<http://www.associationeconomie numerique.fr/>), regroupe des entreprises, grandes ou ex start-up ayant déjà bien progressé, travaillant principalement dans et autour du commerce électronique

- L'AFIGEO - Association française pour l'information géographique - (<http://www.afigeo.asso.fr/>) regroupe des collectivités et les acteurs de l'information géographique

- La FIGEC est la Fédération nationale de l'information d'entreprise et de la gestion de créances – (<http://www.figec.com/>)

- La FING (cf. *ci-dessus*) rassemble aussi de grandes entreprises, notamment fournisseurs d'information parce que gérant des services aux collectivités et des services publics ou agissant dans le domaine des TIC

- Le GESTE Groupement des éditeurs de services en ligne – GESTE – (<http://www.geste.fr/>) regroupe plutôt les sociétés ayant d'abord créé des services en télématique vers le grand public et maintenant présentes sur le Web, les mobiles etc.



• Le SPDG - Syndicat professionnel de la géomatique – (<http://www.spdg-geomatique.org/site/exindex.htm>) est le syndicat des spécialistes des outils informatiques de l'information géographique.

*[De multiples interrelations existent entre ces associations, et de nombreux membres sont communs]*

Il faut aussi prendre en compte, et de plus en plus, les acteurs de l'information traditionnelle, papier ou audio-visuel, qui se sont tous engagés dans l'information électronique sur le Web et les mobiles, les éditeurs de lettres d'information électronique, les spécialistes de l'information scientifique et technique y compris publics et les éditeurs d'ouvrages réunis dans le Syndicat national de l'édition (SNE).

Deux groupes en particulier jouent ou vont jouer un rôle dans le domaine et participent déjà aux réflexions car ils sont très directement intéressés à la diffusion de l'information publique et de l'information locale en particulier :

la presse quotidienne régionale avec le SNPQR, (<http://www.pqr.fr>)

la presse spécialisée (<http://www.fnps.fr/>)

### 1.3.3 L'Etat régulateur et les institutions chargées d'orienter et réguler la réutilisation des données publiques

Le principal fournisseur actuel d'information publique est aussi celui qui la régule, en même temps qu'il s'efforce d'unifier et organiser les pratiques des administrations.

Les organismes en charge ont été rappelés et seront rencontrés dans les différents chapitres du rapport :

**CADA** : Commission d'accès aux documents administratifs. Objet: veiller à l'accessibilité des documents publics communicables pour les personnes privées (citoyens, entreprises...) et arbitrer les différents sur le sujet.

□ Rattachement : Autorité administrative indépendante. Champ d'action : toute la sphère publique.

**APIE** : Agence du patrimoine immatériel de l'Etat. Objet: sensibiliser les administrations à la valorisation de ce patrimoine et en systématiser les règles (brevets, licences, fréquences, marques, savoir-faire publics, bases de données, droits d'accès ou images publiques). Rattachement : ministère de l'Economie. Champ d'action : administration d'Etat mais possibilité d'appuyer les collectivités.

**CNIL** : Commission nationale informatique et libertés. Objet : protection de la vie privée. Rattachement : Parlement. Champ d'action : toute la société.

**COEPIA** : Conseil d'orientation de l'édition publique et de l'information administrative. Objet: donner des avis sur l'édition administrative, la diffusion directe par l'administration et l'insertion dans le marché de l'information et de l'édition. Rattachement: Commission indépendante auprès du Premier ministre, appuyée administrativement par la DILA. Elle rassemble des représentants de l'édition publique et privée sur tous supports. Champ d'action : édition publique.

**Etalab** : Mission. Objet : mettre en place le répertoire des données publiques conformément à la transposition de la Directive et orienter la politique de réutilisation. Rattachement : Premier ministre. Champ d'action : administrations d'Etat mais possibilité d'appuyer les collectivités.

**CNN** : Conseil national du numérique (instance consultative formée d'expert d'Internet, créateurs d'entreprises). Objet : tous les aspects de l'économie numérique. Rattachement : Gouvernement, moyens mis à disposition par le ministre de l'Economie numérique. Champ d'action: l'économie. (Le CNN vient d'ouvrir un chantier données publiques en constatant le développement de conflits comme celui opposant la RATP à CMM – CheckMyMetro – cf. infra).

### 1.3.4 Comprendre les difficultés et les conflits

#### ■ *Un rôle non inscrit dans les missions*

Il est normal que des difficultés apparaissent quand il s'agit pour des acteurs publics de se placer dans un rôle de producteur / éditeur de données, capable de mettre à disposition sur les réseaux des informations de qualité. Ce n'est pas, en tout cas ce n'était pas, leur mission et cela entraîne inévitablement des charges humaines et financières, mais aussi des problèmes parfois complexes d'organisation et souvent des relations difficiles avec des partenaires extérieurs utilisant déjà les données, associés dans des services ou souhaitant en développer. C'est au demeurant exactement la même chose pour toute entreprise ou organisation dont ce n'est pas la vocation première d'être diffuseur d'information – et c'est en fait encore plus difficile dans beaucoup de cas. Seuls les acteurs dont la fonction était de longue date la diffusion de données sont préparés – même s'ils posent immédiatement les questions de coûts et de responsabilités pour la réutilisation. Ils sont d'ailleurs, pour l'Etat, exclus de l'application directe des textes (DILA, INSEE, IGN, INPI etc.).

#### ■ *Des oppositions apparues dès le début*

Des oppositions parfois très vives et prolongées ont marqué le développement du marché de l'information et donc de la réutilisation des données publiques. Il est intéressant de le rappeler parce que les mêmes types de débats vont ressurgir ou le font déjà. Il vaudrait mieux éviter les pertes d'énergie et de temps et gérer les conflits avant qu'ils dégénèrent ou s'éternisent.

Les débats ont commencé à propos de la rediffusion de données de l'INSEE par un grand acteur américain de l'information économique (Wharton). Le débat était aussi sur la tarification : il fallait, selon certains, donner

- l'information aux organismes publics et associatifs et la faire payer aux sociétés commerciales.

D'autres débats importants se sont ensuite développés, liés à la propriété légitime des données, à des tarifs de réutilisation jugés prédateurs, à des exclusivités. A titre d'exemple :

- Demandes réitérées à la Météorologie nationale, qui a toujours opposé le fait que l'information météo est essentielle à la sécurité des aéronefs.
- Opposition prolongée entre les Greffes des tribunaux de commerce et l'INPI pour la publication des bilans de sociétés (remis aux uns, conservés par l'autre), conduisant à deux services en concurrence et se disputant une exclusivité.
- Résistance du JO à la réutilisation, donc à la communication de données numériques, qui contraignait les éditeurs juridiques à une saisie en Asie, puis résistance à la publication sur le Web en mode texte, donc réutilisable.
- Plaintes renouvelées et générales sur les tarifs pratiqués par l'IGN et les freins mis à la réutilisation.
- Accès coûteux au cadastre pour les collectivités.

#### ■ *La culture de service public en France*

Ces positions « réactionnaires » doivent être comprises et replacées dans une culture d'organisme et plus largement dans la culture des services publics. Alors que le travail des administrations n'incluait pas en général la publication d'informations et encore moins la remise de données organisées à des opérateurs commerciaux, plusieurs organismes avaient été créés pour assurer cette mission, dans la tradition d'excellence du secteur public et de monopole d'Etat et de responsabilité qui était pratiquement consensuelle en France – et reste largement admise. En outre, les personnels de ces

organismes, et en général les agents publics, craignaient que leur travail ne soit plus reconnu et valorisé. En tout état de cause, les organismes en cause diffusaient de façon payante leurs propres produits - d'abord éditions papier - et construisaient avec beaucoup de difficultés un modèle économique réduisant leur poids dans le budget général et leur donnant plus d'autonomie.

Modifier l'équilibre ne peut se faire sans difficulté. Cette réalité doit aussi être prise en compte dans les développements actuels : les nouvelles méthodes de diffusion peuvent léser des acteurs qui n'ont pas démérité et qui se voient mis en difficulté, ou le craignent.

Plus généralement, un facteur culturel présent dans la société française intervient, qui gêne beaucoup d'opérations de réutilisation. Comme le montre une étude Pierre Audoin Conseil pour Logica (publication de juin 2011), les spécificités françaises doivent être prises en compte lorsque les politiques d'open data sont lancées en s'inspirant trop fortement de modèles anglo-saxons. Qu'il s'agisse des agents publics ou de la population, l'attachement au service public tel qu'il est défini en France est fort. Dès lors, une certaine méfiance s'exprime face à une intervention trop forte du secteur privé dans son fonctionnement. L'opposition entre commercial et non commercial est aussi beaucoup plus marquée en France, s'agissant de réutilisation des informations publiques. L'essentiel va être ici de clairement définir ce qui est du domaine du service public et ce qui doit demeurer sous forte garantie de pérennité et d'accessibilité, y compris s'agissant d'information, et de l'expliquer. Ces difficultés apparaissent aujourd'hui sur plusieurs dossiers (ex : la généalogie - cf. infra).

#### ■ *Les contradictions de l'Etat*

Enfin, l'Etat lui-même est partagé entre le souci de promouvoir le développement de la société de l'information et donc de la réutilisation et

□ l'espoir de recevoir rapidement ce que le rapport Lorentz de 1998 appelait des "recettes de poche" - elles ne sont jamais très élevées pour un service administratif au regard de son budget. Cette opposition s'est même marquée dans des divergences entre l'APIE et Etalab. Elles se sont réduites mais l'APIE a été amenée à publier un communiqué affirmant que sa position et celle d'Etalab étaient en parfaite harmonie, et que la plupart des jeux de données seraient libres de redevances.

Comme c'était prévisible, les difficultés surgissent de nouveau lorsqu'une société commerciale et supposée rentable demande à réutiliser les données - relation différente de celle où une ville autorise des start-ups locales à développer des nouvelles applications sur mobile. L'exemple malheureux de la généalogie le rappelle en 2011 (cf. infra). Pourtant, il ne s'agit pas que d'argent mais d'organisation, de compréhension des mécanismes de l'économie numérique mais aussi de reconnaissance du travail des agents publics. Comme cela a été rappelé, il s'agit bien d'une redéfinition du périmètre et du contenu du service public ainsi que d'une redéfinition des relations entre les structures publiques et les entreprises.

#### ■ *Gratuité et démocratie*

A côté de ces oppositions «traditionnelles», un autre type de débat agite le monde de l'open data, entre partisans de l'ouverture et de la gratuité (qui permet à des citoyens et de petites structures de diffuser et réutiliser) et responsables méfiants des administrations territoriales. Là encore, il est simpliste de voir les évolutions en cours simplement comme une offensive de la "société civile" pour ouvrir les coffres-forts d'administrations conservatrices.

De façon sous-jacente, comme cela a été indiqué ci-dessus, une redéfinition de la démocratie est aussi en jeu : toute action des politiques et de l'administration doit pouvoir être évaluée, contrôlée et jugée par le peuple.

On peut comprendre que cette évolution pose question et qu'elle doit en tout état de cause être encadrée.

#### ■ Une question politique

Aussi, comme le confirment les promoteurs de programmes d'ouverture des données, la démarche doit d'entrée impliquer les politiques, et ne pas être menée en fonction de l'opposition entre des cultures "opendata" et "conservatrices". Au delà des principes imposés par la Loi, il y a à décider sur des critères politiques de la mise à disposition de tel jeu de données. Il existe bien des données dont la mise à disposition sans contrôle peut compromettre une politique (par exemple des responsables communaux se plaignent de ce que l'accès à une information immédiate sur les embouteillages peut compromettre la politique de circulation – mais il est possible de noter qu'un opérateur pourra à terme obtenir ce type de données). Dans d'autres cas le mésusage d'informations peut être dangereux pour le consensus social (ex des données sur les agressions et l'activité policière – succès aux Etats-Unis).

*à Il est important de noter que beaucoup de débats font une confusion entre accessibilité et réutilisation. Il reste du ressort de l'administration (et d'abord des élus) de décider si une information est accessible.*

#### ■ Essayer de progresser par la concertation pour résoudre les conflits

Le développement de nouvelles activités et des nouveaux fonctionnements suppose d'éviter la contrainte et encore plus le recours au tribunal, ce qui se produit parfois. La résolution de beaucoup de conflits peut être obtenue par une meilleure connaissance réciproque, par le recours à des instances de concertation, par des concessions des uns et des autres. Le GFII a toujours recommandé cette méthode et a pu constater que beaucoup de tensions pouvaient être ainsi réduites.

□ L'affaire en cours des relations entre les sociétés Notre Famille et Ancestry et les archivistes des Conseils généraux commence de s'orienter dans la voie de la conciliation : une première rencontre a eu lieu (cf. "Généalogie"). En même temps, le tribunal a tranché dans le débat avec un Conseil général hostile. Les deux voies existent mais mieux vaut progresser ensemble, quand c'est possible.

### 1.4 Perspective internationale

Le mouvement des open data est international et atteint de très nombreux pays : les Etats-Unis et le Royaume Uni, qui furent leaders, mais aussi l'Europe et des pays comme le Kenya ou la Thaïlande.

Cependant, comme cela sera indiqué dans le chapitre "Economie", il est très difficile, voire illusoire, de rechercher autre chose que des déclarations sur l'intérêt social et économique de l'ouverture et parfois quelques chiffres non comparables, compte tenu des différences entre les pays.

L'ouverture des données, avant l'internationalisation d'un mouvement, avait reposé sur des relations contractuelles entre les acteurs, souvent internationaux, des banques de données professionnelles et les différentes administrations. Ce type d'accord repose sur le droit local – qui interdit par exemple aux administrations américaines de percevoir des redevances sur des données dont la collecte et le traitement sont financés par l'impôt (sauf exceptions).

Désormais, un vaste ensemble d'Etats annonce vouloir stimuler ouverture, réutilisation et gratuité. Pour ses membres, l'OCDE a mis en place un groupe de travail et suit l'évolution de la réutilisation.

La Commission européenne a joué ici un rôle essentiel avec la directive de 2003. Elle poursuit son action, l'orientation open data étant un objectif de la Déclaration des ministres sur l'eGovernment (Malmö 2009).

Our public administrations should therefore :

[...]

Increase availability of public sector information for reuse. We will increase availability of public sector information for reuse, in accordance with the spirit of and the conditions established by Public Sector Information Directive 2003/98/EC. We will encourage the reuse of public data by third parties to develop enriched services that maximise the value for the public. New demand-led information products and services enabled by the reuse of public sector information will support the transition of Europe to a knowledge-based economy

Neelie Kroes, Commissaire en charge du Digital Agenda, annonce un impact économique impressionnant (*cf. Economie*) et la Commission annonce en juin 2011 un portail, et prend parti pour la gratuité.

Des programmes spécialisés visent aussi la mise à disposition des données publiques, comme le programme MED (L'Europe en Méditerranée), auquel participe la Région PACA. L'un des «work package» du programme vise à une plus grande participation des citoyens à l'élaboration et à la mise en œuvre des politiques publiques (information, débat, production d'information par les citoyens eux-mêmes).

Enfin, la Commission soutient des actions de sensibilisation et d'étude avec l'association PSI Alliance (<http://www.psialliance.eu/>) qui représente la communauté de la réutilisation.

Par ailleurs, en Europe, l'association SharePSI (<http://share-psi.eu/>) réunit le W3C, l'association de normalisation des opérateurs télécom ETSI, l'Open Knowledge Foundation et des centres de recherche et des groupements de différents pays (la FING en France).

En France, les promoteurs de l'open data ont signalé un retard dans l'ouverture des données. La Fondation IFRAP (Fondation pour la Recherche sur les Administrations et les Politiques Publiques) a analysé la position française à partir de l'étude de l'OCDE Panorama des administrations publiques 2011. La France est en 21<sup>ème</sup> position sur 34 pays participants. Ce classement est effectué en fonction du degré de proactivité des administrations publiques dans la livraison spontanée de documents administratifs et de données publiques. L'IFRAP explique :

- D'une part, il révèle que « si le gouvernement français publie de manière proactive des documents budgétaires sur les sites des ministères ou des opérateurs », il n'en est pas de même des rapports d'audit ou des listes de fonctionnaires et de leurs rémunérations. Pour ces deux catégories de documents, il n'existe aucune obligation légale, ni de quelconque pratique régulière en la matière. Mieux, il est toujours possible en France de refuser la publication d'un rapport. Une pratique sensiblement différente chez nos voisins anglo-saxons ou scandinaves.

- D'autre part, il n'existe pas d'obligation légale ou réglementaire pour rendre les jeux de données publiques accessibles au public. Cette seconde assertion relevée dans le rapport OCDE est plus critiquable car elle fait fi du décret du 26 mai 2011 et de la circulaire attenante, ainsi que de la mise en place de la mission Etalab d'installation du portail central gouvernemental [data.gov.fr](http://data.gov.fr).

[...]

La France n'est pas la seule démocratie mature à être épinglée : l'Allemagne n'a pas répondu au questionnaire, la Suède est 24<sup>ème</sup>, l'Irlande 29<sup>ème</sup> et le Luxembourg 30<sup>ème</sup>. En revanche, le classement montre que les pays les mieux placés sont les jeunes démocraties (Slovénie en 3<sup>ème</sup> position) ou luttant contre la corruption (Mexique et Italie respectivement en 2<sup>ème</sup> et 1<sup>ère</sup> position).

[...]

Source : <http://www.ifrap.org/Classement-international-2011-Open-Data-et-Open-Government,12204.html>

Deux points à relever dans cette analyse : d'une part il n'y a effectivement pas de contrainte sur l'accessibilité – les textes visent la réutilisation. Par ailleurs, l'essentiel du classement porte en fait sur le volet «transparence démocratique» de l'open data et non sur la réutilisation commerciale. Le travail se place dans la rubrique «open data – open government».

Comme cela sera vu dans le chapitre Economie, il est impossible de mesurer l'impact actuel de l'open data, encore en phase de lancement, où le fonctionnement et la pérennité des applications ne sont pas stabilisés. Même, le bilan de l'open data au Royaume Uni, établi par l'Ambassade de France, signale que peu des nouvelles applications mobiles ont maintenant fait leurs preuves.

En même temps, les coupes budgétaires aux Etats Unis menacent la survie du modèle qu'était Data.Gov, obligé de revoir toute son architecture technique et finalement son rôle, faisant reposer les frais sur les différentes agences.

Cependant, les aléas des débuts ne peuvent être interprétés sur une courte période. La Commission, comme la France, confirme son orientation de moyen et long terme et prépare la refonte de la directive de 2003. En particulier, la réutilisation deviendrait une obligation, comme c'est le cas en France. La Commission poursuit les études d'impact et elle fera des propositions pour la révision à partir de fin 2012 (au plus tôt).

### *Les premiers contours de la future directive sur les données publiques*

Source : GFII

Au cours de la conférence annuelle de l'association PSI Alliance (<http://www.psialliance.eu/>) qui s'est tenue à Bruxelles le 28 juin 2011, Grazyna Piesiewicz, la juriste responsable de la révision de la directive de 2003 et bras droit de Neelie Kroes, membre de la Commission européenne, responsable de la Stratégie numérique, a indiqué quelles pourraient être les grandes lignes de la révision.

Extension du champ d'application de la directive aux instituts de recherche, aux émissions publiques de TV et radio (broadcasting) et aux données culturelles

Nouveau principe général: toute donnée accessible est automatiquement réutilisable

Définition plus précise des principes généraux de tarification et de certaines notions (coûts marginaux), ainsi que des exceptions, ..

Renversement de la charge de la preuve: ce sera à l'administration de prouver qu'un réutilisateur n'a pas respecté les termes de la licence

Les données devront être diffusées dans un format réutilisable

Création (si besoin) d'une autorité régulatrice sur les PSI, particulièrement sur les aspects de tarification (voire la concurrence public-privé).

## 2 Données et applications

Ce chapitre traite des questions relatives à la gestion des données, du point de vue fonctionnel et non technique, ainsi que des divers gisements et des applications qui ont déjà pu être observées.

De nombreux types de données et applications sont évoqués, quelles que soient les collectivités ou organismes publics susceptibles de les gérer. En effet, l'étude se place dans une perspective régionale d'aménagement de la circulation de données, aspect de plus en plus important de l'aménagement numérique du territoire.

Enfin, plusieurs points de vue doivent être adoptés et il ne faut pas chercher de classement logique. Il y a obligatoirement des recouvrements et des démarches disjointes, notamment par types de données, par domaine métier et par types d'utilisateurs.

Un autre aspect essentiel est que les données peuvent interagir : c'est largement ce qui permet une valeur ajoutée extérieure mais c'est aussi fondamental pour le système interne de l'acteur public (les anglo-saxons parlent de "linked data").

Ainsi, la gestion des données peut a priori paraître d'une extrême complexité. Il est très important de ne pas s'y arrêter et de ne pas chercher à tout organiser rationnellement et parfaitement dès le départ. Au contraire, il s'agit d'associer des démarches et des applications simples, pragmatiques et certainement incomplètes à une vision de long terme et à des règles d'encadrement, c'est-à-dire de définir une heuristique.

### Remarque : production et propriété des données.

On distingue souvent le "producteur" de la donnée du "réutilisateur". Or, la donnée n'est pas produite par tel ou tel service – elle circule et elle est traitée et transformée. Le producteur d'information sur un citoyen ou une entreprise, c'est cette personne privée elle-même. L'acteur public est le gestionnaire des données regroupées, traitées.

Cette même problématique est sous-jacente à la notion, dangereuse et floue, de "propriété des données" qui sous-tend nombre de débats. Il est extrêmement difficile voire prétentieux de se croire propriétaire d'un objet immatériel. Le propriétaire est-il celui qui l'a payé (et alors, tout ce qui est payé par le contribuable doit ensuite être donné gratuitement à la communauté des contribuables, un des arguments forts de l'open data)? Ou bien est-ce celui qui a traité et mis en forme et a créé un objet virtuel gérant des objets virtuels, la base de données? En revanche, il faut parler de droits attachés, et c'est complexe parce que de multiples droits vont progressivement se créer au fur et à mesure de la circulation de la donnée. (cf. chapitre sur la réutilisation et les règles de droit)

### 2.1 Principes : Gestion de données

*La gestion des données n'est pas d'abord l'affaire des services informatiques. Ils en savent les conséquences mais c'est bien aux services fonctionnels, aux services et aux agents en charge des procédures, en relation avec les citoyens et les métiers de définir et contrôler les politiques en matière de données.*

### 2.1.1 Flux des données, cohérence et synergies

Le circuit des données existantes a été déterminé par les compétences des différentes collectivités et des acteurs publics et par les procédures correspondantes ainsi que par les activités courantes, le fonctionnement et les marchés. C'est pourquoi il y a en général de multiples silos de données et documents – que ce soit sous forme papier ou électronique, au sein d'une même entité mais aussi, et encore plus, entre les différentes entités. Il serait au demeurant erroné d'opposer de ce point de vue une administration réputée mal gérée à un secteur privé efficace. La situation est la même dans de très nombreuses entreprises, et les évolutions d'organisation et de périmètre y posent des problèmes similaires de cohérence.

Dès lors, et en dehors des projets d'ouverture de données, la garantie de long terme d'une pérennité des actions et des systèmes d'information, et donc aussi de tout partage externe, est d'entamer un travail de fond sur les données.

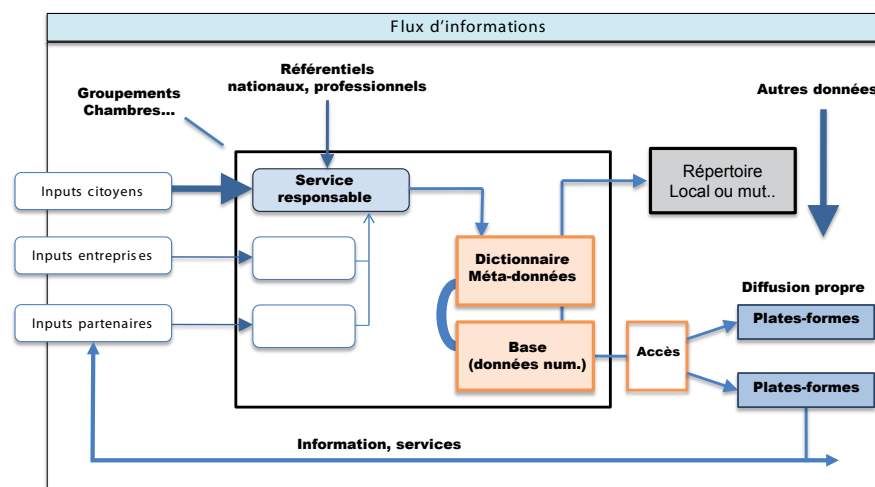
**Identification des flux** d'information et de données les plus importants.

**Définition précise des données** et de leur place dans les mécanismes d'utilisation et d'échanges internes et externes. Légitimement, différents services définissent et nomment différemment les données, lorsqu'elles ne sont pas strictement encadrées par une définition réglementaire – ce qui est heureusement assez fréquent dans les procédures publiques comme dans les indicateurs d'activité et les outils de contrôle (malheureusement, cette définition évolue souvent dans le temps, comme le savent bien les statisticiens). Des définitions précises doivent être établies en prenant en compte les partenaires extérieurs.

**Définition de métadonnées**, données caractérisant la donnée ou la série de données (y compris pour les documents et données non numérisés). Des normes – dictionnaires de métadonnées – existent dans les différents domaines, ou sont souvent en développement.

□ **Responsabilité dans la gestion d'une donnée.** La règle est qu'une donnée doit être produite ou validée par un seul responsable. Pour un acteur public, il s'agit d'un service métier, qui peut être en accord avec ses correspondants privés en charge d'un domaine (syndicat, chambre consulaire, observatoire etc.). *Remarque : un formulaire, une déclaration obligatoire peuvent contenir des données dépendant de plusieurs services.*

Au total, il faut connaître et valider, avant toute refonte des systèmes techniques, les circuits de production et contrôle des données. Comme cela sera rappelé ci-après, la communication et le partage de données – en open data ou autrement – sont des facteurs d'amélioration de leur qualité.



Il faut aussi analyser les données du point de vue juridique :

**Identifier les droits** qui pourraient, systématiquement ou partiellement, être attachés - droits d'auteur, droit à l'image, documents acquis sans vérifier les droits de propriété et de réutilisation, données nominatives, données



permettant une identification illicite de personne ou d'entreprise.

**Identifier enfin le caractère de donnée publique** et l'accessibilité à partir de l'origine des données et des différentes informations (ou prévenir éventuellement les difficultés s'il semble que la donnée puisse être demandée et si son caractère demeure incertain)

**Vérifier les relations contractuelles avec les sources intermédiaires** le cas échéant (DSP, organismes mutualisés etc.), voire essayer de les modifier s'il s'avère que rien n'a été prévu sur les possibilités d'utilisation des données.

Les usages sont un autre volet essentiel de l'analyse des données. Leur analyse fait aussi apparaître les manques et les compléments.

**Les usages** : Les questions doivent être posées des utilisations actuelles quelle qu'en soit l'objet et la forme – que ces usages concernent le papier, des données informatisées, des publications Web et de services déjà existants ; et qu'il s'agisse d'usages internes à l'institution, d'accès individuel par requête, appel téléphonique, site Web ; qu'il s'agisse enfin d'usages par des associations, des institutions, des entreprises, les services de l'Etat etc. Il faut connaître la nature des informations transmises et les objectifs, le volume des demandes ou accès, les difficultés rencontrées et les relations contractuelles. Il y a des priorités entre les différents gisements mais, à terme, un répertoire au moins interne des données devra préciser ces éléments.

**Manques, compléments, données liées** : L'analyse, à partir des demandes internes et externes, porte sur les manques, les difficultés dans le recueil, la production, la gestion et la mise à jour des données. Il est aussi intéressante de connaître, sur le territoire, les données proches, potentiellement complémentaires, gérées par d'autres acteurs (Collectivités, chambres, associations, entreprises) et par les administrations d'Etat centrales et déconcentrées.

## 2.1.2 Le volet informatique : les bases de données (rappel)

### ■ *Un passage à la gestion en base de données toujours important et toujours difficile*

Le premier chapitre a rappelé que l'effort principal sur les données, du point de vue du système d'information, était de passer à une organisation en base de données. Plus généralement, des efforts sont indispensables pour réaliser dans la pratique les objectifs décrits ci-dessus : relier les informations entre elles dans la collectivité, éliminer les incohérences, relier les données à des descripteurs, permettre aux différents services de la collectivité de fournir et retrouver les informations qui les concernent.

Ces tâches ont pris un retard certain dans la plupart des collectivités, mais l'Etat rencontre les mêmes difficultés (cf. les difficultés des projets Accord puis Chorus au ministère du Budget).

Les entreprises ne sont pas toutes plus avancées sur ce chemin, même si de nombreux chantiers sont ouverts. Il est intéressant de voir que c'est souvent le partage avec des acteurs extérieurs qui est déterminant dans cet effort.

En pratique, il s'agit de passer d'une gestion des applications internes à une gestion des données. C'est indispensable pour assurer la flexibilité du système et, justement, permettre le développement rapide d'applications à destination de l'extérieur. Un mouvement important dans les entreprises a en effet été depuis plusieurs années de passer à l'entreprise communicante, capable de partager des données avec ses distributeurs, ses partenaires, voire ses clients- en tout cas d'améliorer en théorie la relation client. Certes, il ne s'agit pas en entreprise d'ouvrir à des développeurs, mais il s'agit bien de gérer des données au delà de leur fonction initiale.

Rappel : les gestionnaires et les informaticiens connaissent le coût mais aussi l'intérêt de la maîtrise d'une base de données unifiée ou d'une

gestion coordonnée des bases : qualité, cohérence des informations, réactivité. A moyen terme, les coûts de gestion diminuent parce qu'il n'y a plus de multiples systèmes, les erreurs se réduisent, et surtout les métiers sont tenus d'échanger et les silos tendent à devenir au moins poreux.

■ *De l'analyse métier à la constitution des bases, ou de la base, de données<sup>8</sup>*

Le volet informatique, c'est-à-dire l'entrée des données dans un système d'information (et non dans des silos) est pratiquement très simple lorsque les analyses métier précédentes ont été menées et ont réellement impliqué en particulier les responsables internes de la collecte comme ceux des services utilisateurs. Il reste à transférer les informations dans un modèle des données et du système exprimé en termes de système d'information – tout au moins pour les données numériques.

**La numérisation** est évidemment préliminaire et dépend de l'analyse de besoins. Compte tenu de son coût toujours élevé, elle suppose une analyse coûts-avantages sur le moyen terme, des décisions de prise en charge directe, de partage, de sous-traitance (voire de partenariat privé – cf. BNF). Dans certains cas, il s'agira seulement de référencer des documents qui pourront ainsi être gérés par le système d'information et communiqués à l'extérieur sur demande. Néanmoins, si des décisions sont prises sur l'intérêt de la réutilisation par des acteurs extérieurs, la numérisation devient indispensable. Souvent, il s'agira de numériser l'acquis puis de chercher à numériser les documents à la source, proposant un service sur le Web.

**Etat de l'existant technique** : l'analyse d'un jeu de données précisera les

<sup>8</sup> Jusqu'ici, une base de données a été évoquée – dans certains cas plusieurs bases pourront être mises en place, qui devront communiquer. L'architecture finale des systèmes est l'affaire des informaticiens (il peut par exemple y avoir déport de bases et de services à des sites externes)

□ systèmes techniques actuels et les relations avec les éditeurs informatiques et les opérateurs de services (hébergement, traitements). Dans certains cas, les contrats avec des éditeurs peuvent rendre impossible – c'est-à-dire très coûteuse – les évolutions et notamment l'ouverture. Ces contraintes sont fréquentes avec les systèmes «propriétaires» (souvent tous les systèmes qui préexistaient aux normes actuelles centrées sur l'interopérabilité). Le choix de s'en libérer est coûteux mais à terme indispensable.

**Normes et accès aux données.** Pour les données numériques, l'analyse informatique portera aussi sur les formats, sur les interfaces proposés à l'extérieur pour l'entrée et pour l'exportation des données (cf. ch. «Technique»), sur les métadonnées de description attachées aux données. Il est évidemment préférable de profiter de toute refonte pour adopter, si ce n'est pas déjà fait, des formats, des interfaces de saisie, des métadonnées normalisés. Pour les interfaces, il s'agit de choix généraux de l'institution pour les échanges externes. Par ailleurs, si le format des données pose des problèmes, il faut privilégier la normalisation et la publication des métadonnées : c'est par elles que les correspondants, les réutilisateurs potentiels pourront identifier l'existence et la nature des gisements dans les répertoires (dont la constitution est une obligation). Or, les métadonnées font dans de nombreux domaines l'objet d'une standardisation ou d'une normalisation métier et informatique (géographie, archives, données financières.. et en général toutes les données qui ont vocation à être associées dans des ensembles plus larges ou présentées sur des portails). Il est au demeurant indispensable d'exercer une veille dans ces domaines (qui peut être exercée de façon mutualisée par les diverses plates-formes).

**Contraintes particulières.** L'ouverture à l'extérieur peut engendrer de nouvelles contraintes, qui devront avoir été analysées par les acteurs métier. C'est le cas si des garanties sont attachées aux données – par exemple pour des dates et des horaires : obligation de mise à jour, voire accès temps réel aux horaires dans le cas du transport.

### 2.1.3 Services et données

Il faut distinguer clairement les données et les services qui sont produits en les utilisant – dont les services d'information, qui supposent de les mettre en forme et de créer des interfaces d'accès.

Le problème est de distinguer la donnée "brute" de la donnée traitée pour être mise à disposition d'acteurs qui vont produire de nouveaux services à partir d'elle. Ces acteurs sont divers:

Services de la collectivité (ou de l'acteur public en général)

Services mutualisés offerts par des plates-formes (notamment pour l'information géographique)

Services produits par des réutilisateurs

Précisément, un des arguments forts pour l'ouverture et la réutilisation et pour la gratuité est que c'est seulement des acteurs privés, associations, start ups ou sociétés d'information plus établies, qui pourront développer une diversité de services à partir des données brutes. Ils pourront aussi les croiser avec d'autres et s'adresser à des publics hors du périmètre propre du gestionnaire de la donnée.

Des réflexions et des décisions importantes concernent le périmètre des services:

Quels services doivent être garantis à la population. Par définition, ces services doivent s'intégrer à la notion de service public et la collectivité doit veiller en particulier à l'égalité d'accès.

Quels services sont importants et peuvent être stimulés, mais hors d'une responsabilité forte de l'acteur public.

Quels services se développent dans le cadre de la liberté des citoyens et de la liberté du commerce.

□ *Rappel : les services d'information doivent être distingués de la même façon : ce n'est pas obligatoirement à l'acteur public de les fournir, en produisant des lettres d'information et en créant des sites Web. La tendance à le faire est forte et normale dans les structures publiques. Or, cette activité d'édition représente des charges, des choix pour les types d'accès etc. et fait souvent double emploi avec le travail d'entreprises de presse ou d'information. Là encore, les règlements et la décision politique doivent permettre de dire qui fait quoi et qui contrôle quoi.*

### 2.1.4 Service public et données - le périmètre de la donnée publique

Cette différence entre données et services se retrouve dans la question des périmètres respectifs du service public et de la donnée publique. Le statut des données peut être observé à partir des différents utilisateurs qui assurent des services en l'utilisant.

#### ■ *Publication par l'acteur public ou une entreprise sous-traitante*

Qu'il s'agisse de données acquises dans le cadre d'une mission de service public ou d'autres données, la publication est – cf. ci-dessus – un service, que la collectivité assure ou non.

*à Il faut distinguer la réutilisation (y compris pour publication) de la situation où l'acteur public délègue une mission de publication à une entité externe. Le statut des données reste celui qu'il a dans la collectivité, l'entité concernée étant un prestataire et ne devant pas se voir déléguer les questions de mise à disposition (en raison de l'interdiction d'exclusivité, car elle aurait alors un contrôle sur la diffusion).*

### ■ Délégation de service public

Les questions de la propriété des données et de la distinction entre données et services se retrouvent ici (elles seront évidemment présentes dans le chapitre sur la réutilisation).

Des problèmes se posent parce que la publication de données est un service particulier, qui en général n'a pas été précisé dans les contrats de délégation. Dès lors, les données sont la «propriété» du délégataire, sauf avis juridique différent, c'est-à-dire sauf si leur publication est essentielle pour le service. C'est pourquoi les horaires et trajets théoriques d'un service de transport sont souvent considérés comme tels mais pas les horaires en temps réel.

### ■ Groupement d'acteurs publics

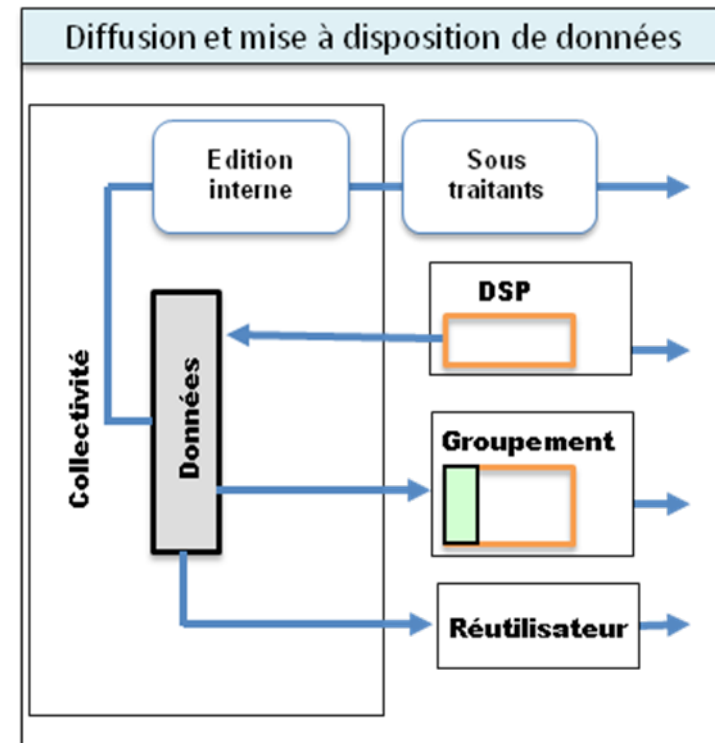
Les acteurs publics peuvent s'unir pour produire un service – plate-forme géographique, plate-forme d'information sur les transports, archivage à valeur légale etc. Les services mutualisés ainsi fournis ne rentrent pas en général dans la définition stricte du service public mais représentent bien un apport important. Les données qu'ils gèrent sont fournies par les différents membres.

Le problème qui se pose alors est celui de la mise à disposition des données pour d'autres services, qui seraient concurrents (même si ce n'est pas une concurrence financière, ce peut-être une concurrence d'usage et cela remet en cause l'investissement consenti). Ici – et ce sera plus difficile – les partenaires de ces groupements et, au delà, la Loi seront conduits à préciser le statut de ces données et des services qu'elles permettent.

### ■ Réutilisateurs

Les conditions de gestion des données, lorsqu'elles sont exploitées par le réutilisateur, dépend de la licence mise en œuvre (cf. chapitre spécial).

- En tout état de cause, les licences contrôlent l'usage des données mais la collectivité cesse d'être responsable.



Ainsi, par exemple, la Ville de Paris insiste sur le fait que son rôle se borne à mettre à disposition des données. La Ville se décharge de l'utilisation qui en est faite. Elle refuse de se positionner en tant que fournisseur de service, en suivant ainsi le concept définissant l'Open data. Au demeurant, les textes excluent pour l'essentiel de prendre en compte l'usage envisagé pour autoriser la réutilisation (point très discuté – cf. chapitre sur les aspects juridiques)

### 2.1.5 Données brutes, données traitées, données enrichies

La donnée fournie par un acteur n'est jamais vraiment "brute". Elle est réputée telle lorsqu'elle est mise à disposition des réutilisateurs, sous forme de fichiers (voire, en théorie, de piles de documents) tels qu'ils ont été constitués dans le cadre de la mission de service public qui exigeait leur collecte.

Cette situation sera de plus en plus rare.

En effet :

Comme cela a été indiqué ci-dessus, l'évolution à terme incontournable de la gestion d'une organisation est de constituer une base de données cohérentes, en apportant donc contrôles, indexations etc.

La mise à disposition, quel qu'en soit le bénéficiaire, suppose d'appliquer des filtres relatifs aux données nominatives, à la protection du secret des affaires, à la sécurité, aux droits de propriété intellectuelle. Après l'analyse des données, des traitements seront nécessaires, en particulier l'anonymisation pour ne pas divulguer d'informations sur les personnes. (*L'anonymisation et sa prise en charge sont traitées dans les chapitres suivants*).

Elle suppose aussi – et ce sera de plus en plus le cas, de mettre les données aux normes règlementaires et techniques pour permettre leur circulation.

Certains de ces traitements – en particulier l'ajout de métadonnées et la modification de formats – constituent un enrichissement des données.

Un principe important est sans doute – et surtout dans la phase actuelle – est de limiter les traitements autant que possible, en veillant surtout à contrôler les filtres règlementaires. C'est d'abord une raison pragmatique, pour ne pas retarder le démarrage d'un mouvement en prenant prétexte de lourdes tâches techniques. C'est aussi une règle de prudence, car les techniques et les usages évoluent rapidement.

□ L'argumentaire présenté par le GFII au COEPIA au groupe «Principes de tarification de la réutilisation des informations publiques» précise ce point.

### Position du GFII en vue de la rédaction des recommandations du COEPIA - Enrichissement des données

L'administration ne peut pas préjuger de la nature de la réutilisation envisagée et anticiper de façon générale les enrichissements des données qui pourraient être souhaités, tous les types de réutilisations répondant à des problématiques différentes tenant aux spécificités du marché et du produit envisagées.

Un enrichissement des données n'est possible que s'il répond à des contraintes légales et de sécurité des données (l'anonymisation par exemple) et s'il a fait l'objet d'une demande explicite des licenciés actuels ou futurs. C'est ce que prônent les sociétés réutilisatrices, entendues par la formation spécialisée du COEPIA en indiquant leur intérêt pour des données brutes afin de les adapter à leurs besoins spécifiques.

La notion d'enrichissement des informations publiques ne doit donc plus être présentée comme un facteur déterminant de la réutilisation de ces informations.

Ce qui implique que le coût d'absorption par le réutilisateur ne serait pas diminué et que la valeur marchande ne serait pas augmentée par un enrichissement qui ne serait pas adapté aux besoins du marché.

La valeur de l'information n'est pas définie par un enrichissement non souhaité et ce n'est pas parce que la valeur ajoutée par l'administration est très faible que le coût d'appropriation par le réutilisateur sera très élevé.

[...]

*La fourniture pour réutilisation, si elle n'est pas celle de données vraiment brutes, doit être, plutôt, aussi générique que possible et il est préférable de limiter les traitements, hors ceux nécessaires aux usages internes et à la bonne organisation des données et hors les contrôles relatifs au respect des lois et les traitements correspondants, s'il est important que les données soient rendues communicables. Les traitements de mise à disposition sont alors limités à une bonne information sur les données et à l'utilisation de formats standards simples (cf. infra Répertoires).*

### 2.1.6 Croisement des données : valeur ajoutée et freins

La réutilisation par des acteurs privés vers des marchés ciblés va permettre l'émergence de nouveaux services : la rencontre des données produit de la valeur ajoutée et la plupart des services pratiquent ainsi exploitent ainsi le "mash-up", popularisé par Google dont les cartes et informations sont gratuitement récupérables par les applications les plus diverses, et sont devenues de ce fait omniprésentes sur le Web, pour le plus grand bénéfice de Google.

S'agissant de données publiques, les difficultés sont grandes en raison de la diversité des statuts, des autorisations et des formats. La charge pour de petites sociétés serait excessive. Les premières applications de nouvelle génération sur le Web et les mobiles se contentent en général de combiner les données d'une ville avec une géolocalisation utilisant des données Google ou OpenStreetMap.

Parfois, le frein est le refus d'un acteur, ce qui empêche l'exploitation utile des données des autres. C'est ce dont se plaint par exemple Nestoria, incapable de relier des données immobilières à des données de transport en raison du refus de la RATP (cf. infra "Commerce").

### 2.1.7 Ne pas oublier les coûts ni les surestimer

Il est aussi faux de croire que l'ouverture des données n'engendrera pas de coûts internes que de surestimer ceux-ci. *Les éléments en seront évoqués dans le chapitre «Economie».* Qui peut croire en effet qu'une politique des données – qui peut éventuellement amener des gains importants à terme du point de vue de la gestion – peut se construire sans effort humain et financier. Comme cela a été rappelé ci-dessus, les coûts les plus notables tiennent à l'évolution de la gestion interne des données, quels que soient les utilisateurs, internes, partenaires et réutilisateurs.

## 2.2 Les données géographiques, outil et enjeu

*Les données géographiques représentent un domaine particulier parce qu'elles sont associées à de multiples applications et permettent de valoriser beaucoup d'autres jeux de données dont les informations sont localisables. Aujourd'hui, la géolocalisation sur les mobiles est l'aspect très médiatisé de l'utilisation des cartes, plans et objets géographiques, mais, pour les collectivités qui les gèrent, l'essentiel est la constitution de portails géographiques en partenariat et les conséquences de la directive INSPIRE.*

### 2.2.1 Les bases

#### ■ Définition

Le ministère de l'Ecologie définit ainsi les données géographiques :

On appelle donnée géographique" toute donnée faisant directement ou indirectement référence à un lieu spécifique ou une zone géographique. On doit comprendre que ce lieu ou cette zone peuvent être un point précis du territoire, une infrastructure linéaire telle qu'une route ou encore un périmètre donné: aire protégée, zone d'emplois, ville, etc.

On peut distinguer trois catégories de données géographiques :

- les référentiels géographiques (plans ou cartes, photographies aériennes ou satellitaires), qui servent surtout de fond de plan pour la visualisation des autres données ;
- les objets géographiques (bâtiments, routes, zones urbanisées, forêts, parcelles, limites de communes, etc.) ;
- les données proprement dites (par exemple la largeur d'une route, le nombre de logements, d'habitants ou d'emplois dans une zone, la population d'une commune...).

#### ■ INSPIRE

L'ordonnance du 21 octobre 2010 a transposé la directive "INSPIRE" du 14 mars 2007.

Cette directive vise à établir une infrastructure d'information géographique dans la Communauté européenne afin de permettre la mise en œuvre des politiques d'environnement.

Elle impose un cadre réglementaire pour la mise en place des moyens nécessaires à la production, au partage et à la réutilisation des données géographiques. Toutes les actions et informations ayant un impact sur l'environnement doivent pouvoir s'y relier. Pour cela, la Directive fixe des règles communes et des normes pour la mise en œuvre, les métadonnées, les structures et formats des données, les services accessibles, l'interopérabilité.

<http://inspire.jrc.ec.europa.eu/>

*cf. une présentation d'Inspire par le CIGAL (Coopération pour l'information géographique en Alsace).*

#### 2.2.2 Un débat historique

□ Depuis que des diffusions d'information numérique ont lieu, des récriminations sont exprimées contre l'Institut Géographique National, qui visent pour l'essentiel le niveau de prix et les conditions de licence des cartes et photos.

Il n'est pas utile de reprendre tous les éléments de ce débat sauf à indiquer qu'il évolue, en une lente progression, vers une concertation entre les différentes parties. La position de l'IGN a beaucoup évolué, puisque désormais les réutilisations de données comprenant des données IGN sont autorisées et gratuites pour les acteurs publics et les services non commerciaux.

Cependant, les contraintes mises par l'IGN ne sont pas acceptées par la communauté de l'open data. Aussi, comme ce fut le cas au départ au Royaume Uni, un nombre important de services se tournent vers OpenStreetMap, qui repose sur la mise à disposition de cartes ou photos libérées et la contribution de la population (*cf. chapitre 1 et sites*). Comme l'indiquait Michel Briand (Vice Pt de Brest Métropole océane) sur le site du Mouvement pour l'économie solidaire (<http://le-mes.org/>), *"La nouveauté 2010 c'est l'apport des collectivités qui versent dans Openstreetmap leur données géographiques: rues, cadastre, limites communales .. A Brest, Rennes, à la région Bretagne, à Paris.. des délibérations actent le fait que les données géographiques sont librement réutilisables."*

Une fois encore, il faut sans doute éviter les oppositions idéologiques, et essayer de contribuer à l'évolution des positions.

L'information géographique est coûteuse à obtenir et à gérer. L'IGN a incontestablement une compétence, des moyens humains et techniques, qui légitiment fortement ses travaux.

Son budget s'équilibre de longue date grâce à la vente de produits d'information – mais il faut savoir que les redevances représentaient en 2009 25M€ sur des ressources totales de 137M€ (dont seulement 77 M€ de subvention publique). Dans ce montant apparemment important,

une proportion importante provient des licences à Google et aux Pages jaunes.

Les collectivités, et les plates-formes coopératives (cf. infra) qui ne veulent pas accepter les conditions de l'IGN sont conduites à effectuer elles-mêmes des survols pour disposer de données orthophotographiques. C'est coûteux et il serait utile pour tous de réduire les doubles emplois et de pouvoir accroître la fréquence des survols car les informations évoluent rapidement – elles sont en toute rigueur immédiatement obsolètes. Les responsables de Centre soit passent des accords avec l'IGN soit craignent de ne pouvoir financer la maintenance des bases.

La gestion des données géographiques, la compétence sur les logiciels SIG (Systèmes d'information géographique) sont rares et il est impératif de les mutualiser.

Le modèle et les résultats d'OpenStreetMap sont impressionnants. Il est certainement important de s'appuyer sur les citoyens pour affiner les données géographiques. C'est même incontournable à terme (cf. infra la position de l'Association pour l'économie montagnarde – APEM)

Il est possible de considérer que les données géographiques de base, telles que celles produites par l'IGN sont un investissement national en infrastructure, mais ceci relève de la décision politique au niveau national – et ce sera difficile pour tout gouvernement compte tenu de la situation budgétaire.

Il est certain que, avec l'actuel déploiement des centres d'information géographique, en exclure les réutilisateurs commerciaux est assez incohérent alors que l'information géographique leur est souvent indispensable, mais que des entreprises en création ou des PME ne peuvent payer des redevances importantes. Des doublons pourraient intervenir.

Dans les conditions actuelles, les plates-formes d'information géographique sont un formidable gisement de données publiques régionales, et il va être de plus en plus incontournable de statuer sur les règles de réutilisation. Pour des réutilisateurs souhaitant travailler à une échelle pluri-régionale, la diversité ajoute complexité et coûts.

*Les données géographiques sont essentielles et stratégiques pour tous et singulièrement pour les collectivités en charge de l'aménagement du territoire, du développement économique et du développement durable. Il est important de réunir tous les acteurs au niveau régional et de participer au débat national sur le sujet. Il est sans doute utile de trouver des compromis pour les données géographiques et pour les règles que devraient suivre des plates-formes qui assurent des services d'intérêt national, placés sous l'empire d'une directive européenne et des lois françaises. Des redevances modestes pour un grand nombre d'utilisateurs et réutilisateurs peuvent par exemple assurer des ressources...*

La situation se débloque progressivement. La conférence "Rencontres des dynamiques régionales" tenue à Clermont-Ferrand en juin 2011 a été enrichissante de ce point de vue (organisée par l'AFIGEO - Association française pour l'information géographique, cf. Ch. 1 - avec l'IGN et le Centre régional auvergnat de l'information géographique).

### 2.2.3 Centres régionaux

La mutualisation s'est organisée au niveau territorial dans des plates-formes aux promoteurs et aux statuts variables. Ces plates-formes réunissent, dans des configurations diverses des administrations nationales déconcentrées, des collectivités, des agences d'urbanisme, des associations pour le développement des TIC, des associations travaillant sur l'environnement et le développement durable, des spécialistes de la géomatique etc. Elles



acquièrent des référentiels (cartes et photos) par marchés publics (l'IGN étant alors un prestataire particulier, notamment pour les vols photographiques) ou elles ont des liens structurels avec l'IGN.

Un document important a été publié par l'AFIGEO (cf. ci-dessus) : Catalogue des infrastructures de données géographiques françaises.

Des exemples :

**PIGMA**, Plate-forme de l'Information Géographique Mutualisée en Aquitaine

PIGMA : un rassemblement des acteurs (<http://cartogip.fr/>)

Source : <http://numerique.aquitaine.fr>

• Les élus régionaux, réunis en Commission permanente le 11 janvier 2010, ont décidé de signer la convention d'échange de données géographiques avec le Groupement d'Intérêt Public d'Aménagement du Territoire et Gestion des Risques (GIP ATGéRi) dans le cadre du projet PIGMA.

• Le G.I.P ATGéRi, constitué en 2005, réunit l'Etat, le Conseil régional (depuis le 23 novembre 2009), les services départementaux d'Incendie et Secours de la Dordogne, de la Gironde, des Landes, du Lot-et-Garonne et des Pyrénées Atlantiques, l'Association régionale de Défense des Forêts contre l'Incendie (ARDFCI), les Unions des Associations Syndicales Autorisée de Défense des Forêts contre l'Incendie et l'Office National de Forêts (ONF).

• Le projet PIGMA - pour « Plate-forme de l'Information Géographique Mutualisée en Aquitaine » - est spécifique à l'achat mutualisé de données géographiques, mises à jour annuellement associé à un volet web-mapping innovant. Ce programme bénéficie à l'ensemble des collectivités territoriales et notamment à la Délégation TIC dans le cadre du Programme cadre d'action du déploiement du très haut débit de la Région Aquitaine qui nécessite la collecte et le traitement des données

géo-référencées des infrastructures d'accueil (fourreaux, points hauts, fibre optique).

• Le GIP ATGéRi procède à l'achat mutualisé des droits d'utilisation des "couches basses" pour la sphère publique, parapublique et le secteur associatif dans le cadre d'un marché public. L'acquisition se concrétise par l'obtention d'une concession de droits d'utilisation illimités (hors utilisation commerciale) auprès du propriétaire des données permettant ainsi un nombre illimité d'utilisateurs.

#### Les objectifs de PIGMA:

L'ensemble de ces actions a pour principal objectif d'impulser une dynamique régionale de diffusion et d'utilisation de l'information géographique en Aquitaine mais également : de favoriser l'**interopérabilité entre les services**, d'impulser une **dynamique régionale** de partage autour d'outils d'aide à la décision en complément des actions locales, de générer une **économie d'argent public** par la mutualisation.

**Le Centre Régional Auvergnat de l'Information Géographique (CRAIG)** intègre quant à lui l'IGN comme membre. C'est un GIP constitué par l'Etat, le Conseil régional, les Conseils généraux, les Communautés d'agglomération et l'IGN. L'usage commercial est interdit.

**GéoBretagne** est géré par le syndicat mixte e-Mégalis, qui a d'abord été en charge du réseau haut débit régional. GéoBretagne regroupe 84 membres : le Conseil régional de Bretagne, les quatre conseils généraux de Bretagne, 10 communautés d'agglomération, 65 communautés de communes, quatre villes. Le Conseil régional a décidé en octobre 2010 de «placer la base de données géographiques d'intérêt général de la Région Bretagne et son contenu dans un cadre juridique permettant leur libre accès, libre usage et libre exploitation, gracieusement et sans contrepartie, via [geobretagne.fr](http://geobretagne.fr).»

Le **CRIGE PACA** (Centre régional de l'information géographique) a été « la première structure géomatique régionale créée sur le territoire français ». « C'est un centre de ressources en géomatique au service des organismes publics de la région. Depuis sa création fin 2002 par l'Etat et la Région, le principal objet du CRIGE consiste à développer les usages, la production et le partage d'information géographique entre les services publics ».

<http://www.crig-paca.org/>

En Rhône-Alpes, les services de l'Etat ont créé **PRODIGE**, Plate-forme Régionale pour Organiser et Diffuser l'Information Géographique de l'Etat en Rhône-Alpes. La plate-forme réunit tous les services, des établissements publics et des Services départementaux d'incendie et de secours<sup>9</sup>.

Un groupement d'acteurs privés : **l'Association pyrénéenne d'économie montagnarde**.

L'**APEM** regroupe les Chambres consulaires et d'autres acteurs privés (mais ayant certaines délégations de service public). Elle est inter-régionale et transfrontières (même, Andorre est dans le périmètre) et se construit dans un système maillé. Laurent Plainecassagne, directeur, considère qu'il faut se concentrer sur les ressources pertinentes, sans chercher à couvrir toute l'information. Les travaux sont menés autour des filières économiques – par exemple pour l'économie pastorale, avec des gens sur le terrain qui saisissent des informations avec une tablette.

L'**APEM** travaille en 2011 à un projet interrégional transfrontalier sur le changement climatique. Elle assure l'interopérabilité pour les acteurs du changement climatique (données d'enneigement, pluie, essences forestières). Il y a une entrée unique, un aiguillage qui va chercher les données dans les serveurs espagnol, français etc. et une seule requête utilisateur. Ce sera programmé en octobre 2011 et lancé par Martin Malvy (le Président du Conseil régional).

<sup>9</sup> La plate-forme utilise le logiciel open source Prodiges, que l'Etat promeut comme norme.

Tout ce qui est agrégé sur l'Observatoire est, pour l'**APEM**, une donnée publique mais c'est l'opérateur responsable de la saisie qui décide. Ainsi, les données pastorales obtenues par enquête du Commissariat de massif sont publiques, mais pas celles de la Chambre ou des pastoraux. Pour les données IGN, le Référentiel général à grande échelle - RGE – est donné aux collectivités et vendu aux privés.

## 2.3 Mobilité et transports

Les applications développées dans les villes pionnières de l'open data ont largement privilégié l'information sur les transports, et ce d'autant que ces applications sont destinées le plus souvent aux utilisateurs en mobilité.

Or, c'est dans l'intermodalité et la combinaison d'informations sur plusieurs territoires qu'il y a des besoins à satisfaire et que des réutilisateurs privés peuvent viser des publics ciblés, et c'est justement la création d'applications multi-sources par des opérateurs privés qui pose souvent problème. Pourtant, il est très important de pouvoir proposer des services alliant les réseaux aériens et ferrés, les réseaux de bus et cars, les locations de voitures, les taxis, les vélos en libre service et bientôt les voitures, l'auto-partage.

L'importance des transports dans l'aménagement du territoire, le poids des acteurs et des marchés ainsi que les nouveaux développements justifient un approfondissement des problèmes de gestion des données publiques, celle-ci étant préliminaire à des réutilisations encore très problématiques.

### 2.3.1 Multiplicité d'opérateurs et de statuts

- L'information transports est présente chez des opérateurs multiples, au sein d'un même territoire et entre les territoires. Le problème se pose donc de pouvoir combiner les données de multiples opérateurs, mais dépendant aussi de multiples Autorités organisatrices des transports (AOT), c'est-à-dire

de différents niveaux de collectivités.

- Les statuts des données varient avec la diversité des contrats, qui vont de la délégation de service public sous ses diverses formes au contrat de prestation de service de droit commun (pour les vélos en libre service, non inclus dans le service public de transport). Pour le covoiturage, il y a même le risque de divulguer des informations sur des personnes.

- Au sein d'un même service, la nature des données varie selon que les fournitures d'information à la collectivité ont été prévues ou non – en particulier dans une DSP. Il faut aussi distinguer le service public de transport et le service public d'information sur les transports. La pratique semble être que les horaires théoriques et les trajets et haltes font partie du service public parce qu'ils sont nécessaires à son exercice. Ce n'est en revanche pas le cas pour les horaires temps réel et les événements imprévus perturbant le service. Cependant, la RATP vient de rappeler que le plan du métro n'est pas réutilisable sans son autorisation (*cf. infra*).

- Les limites des territoires doivent pouvoir, par définition, être franchies (et donc les frontières des AOT). La question des services interrégionaux est posée.

### 2.3.2 Risques

- L'organisation des transports est un élément critique des politiques publiques, qu'il s'agisse de l'organisation au niveau de la Région, du département, d'une commune. Or, la diffusion d'informations sur la mobilité et les informations temps réel peuvent perturber les plans de circulation et la régulation. Sans aller jusqu'à évoquer les positions des radars, il faut rappeler que des communes se plaignent de l'affichage sur des GPS des embouteillages et de la proposition de trajets alternatifs, qui peuvent affecter des zones que les responsables voulaient préserver.

### 2.3.3 Les réticences des opérateurs

- Au niveau régional, la SNCF, qui a également une activité d'agence touristique, ne souhaite pas donner accès à toutes ses données d'horaires, ce qui permettrait à des opérateurs de déterminer des trajets intra ou interrégionaux combinant les horaires des bus avec ceux des TER et des trains SNCF. Elle ouvre ses données d'horaire théorique aux collectivités au travers d'un Guichet unique, moyennant paiement.
- La RATP, cela a déjà été noté, met aussi des limites strictes.

L'affaire "CheckMyMetro" n'est cependant pas tout à fait révélatrice d'une volonté de restriction. Service de la société LittleSphere, CheckMyMetro est une application collaborative qui permet aux voyageurs d'échanger des informations sur les événements, de repérer des chanteurs... ou des contrôleurs. La société a ajouté dans une nouvelle version un le plan et les horaires du métro, ce qui a entraîné une démarche de la RATP auprès d'Apple pour le retrait de l'application, la RATP considérant qu'elle a un droit d'auteur sur ces informations, accessibles sur son application payante mais provisoirement gratuite RATP Premium. Si l'argumentation est formellement fondée, il y a sans doute aussi là un moyen de bloquer CheckMyControleur. LittleSphere a retiré l'extension. L'affaire a déclenché l'intervention d'Étalab et même d'Eric Besson.

« Par exemple, la mise à disposition du plan du métro permettrait l'émergence d'un réseau social de ses usagers, qui pourraient s'informer sur les éventuels incidents, formuler des commentaires sur les artistes qui s'y produisent... Plus largement, la mise à disposition des plans et horaires des transports publics permettrait le développement de nouveaux services multimodaux, combinant les différents modes de transport : train, tramway, bus, voiture, vélo... »

*Eric Besson – 13/7/2011*

Pour l'instant, ChekMyMetro a lancé un concours pour la production d'un plan libre de droits.... L'affaire est révélatrice des imprécisions actuelles et des problèmes en suspens. L'utilisation de la notion de droit d'auteur est fondamentalement absurde, mais il est vrai, en revanche, que l'alerte sur les contrôles est une fonction nuisible. Le problème est qu'elle n'a rien à voir avec l'ouverture des données publiques. Elle ressort des problèmes que les TIC posent désormais.

- A Paris, les Velibs sont une prestation de JCDecaux, non un service public. La ville ne peut les intégrer à sa plate-forme de données ouvertes. Au début, JCDecaux a laissé des développeurs accéder à ses données mais a finalement bloqué ces accès en raison de la charge de ses serveurs. Il est encore possible d'obtenir des données et des applications existent, mais sans que ce soit officiel et contractuel.

### 2.3.4 Aspects juridiques

Bien que le chapitre 3 soit consacré aux aspects juridiques de la réutilisation, il est intéressant d'approfondir ici la question des informations de transport.

Une analyse a été menée en 2007 par la PREDIM (Plate-Forme de Recherche et d'Expérimentation pour le Développement de l'Information Multimodale) qui est un dispositif national mutualisé au bénéfice des autorités de transport (*cf. document*). Elle concluait à la nature administrative de l'information de transport, et donc à l'obligation de mettre à disposition et de permettre la réutilisation à titre commercial ou non.

En revanche, une argumentation développée depuis contredit cette analyse. Elle s'appuie sur le fait que l'information de transport est produite par les opérateurs de transports au titre de leur mission de service public à caractère industriel et commercial et se trouve donc exclue de la pleine application de la loi de 1978 ; cette exclusion est en effet prévue par l'ordonnance n° 2005-650 du 6 juin 2005.

On peut cependant relever dans la Loi LOPPSI2 (2011) une mission de l'information des usagers. En outre, le service de transport est souvent assuré dans le cadre d'une DSP. Un flou subsiste certainement, comme le ministre de l'Industrie l'a noté implicitement dans son intervention (*cf. ci-dessus*).

### 2.3.5 Structures et mise à disposition : des situations diverses

*Ce paragraphe et le suivant s'appuient en particulier sur des extraits du rapport du Centre d'Etudes Techniques de l'Équipement Méditerranée "Accès aux données décrivant l'offre TC - État des lieux - Janvier 2010 (cf. Documents).*

Un problème se pose en raison de l'enchevêtrement des AOT sur un même territoire (Région pour le ferroviaire, Départements pour le non urbain, Villes et Communautés d'agglomérations). Les autorités régionales ou départementales vont donc essayer de grouper les différentes AOT autour d'une plate-forme collaborative. Le syndicat mixte est en général la solution. Cette démarche, si elle permet une meilleure exploitation des données publiques, ne garantit pas leur réutilisation – elle peut parfois la freiner, le service mutualisé pouvant être réticent.

- L'offre de transport en commun des Alpes-Maritimes est répartie parmi huit autorités organisatrices de transport (AOT), associées dans le SYMITAM (Syndicat Mixte des Transports des Alpes- Maritimes, créé en 2005) pour créer le système d'information multimodale Céparou06. Elles y retrouvent le CG, des Communautés d'agglomération et Syndicats mixtes, le Conseil régional et des acteurs de Monaco.

- La Région Basse-Normandie est propriétaire des données d'offre TER et de la base constituée pour le fonctionnement du SIM ([commentjyvais.fr](http://commentjyvais.fr)) ; une charte a été signée avec les AOT pour la mise à disposition de leurs données. L'ensemble des données (lignes, arrêts, horaires) est diffusable sous forme de fichiers mais la stratégie de diffusion reste à établir.

- En Pays de la Loire Le SIM Destineo.fr centralise des données fines.

Accessibles par des développeurs, elles restent la propriété des parties signataires et la Région se porte garante des utilisations, dont elle contrôle les finalités. Les données ferroviaires sont gérées non par le Conseil régional mais par la SNCF. Elles sont accessibles par des web services mais pas par accès aux fichiers. La CADA a considéré qu'elles échappaient aux conditions de réutilisation en raison du caractère économique et commercial du service public délégué.

- Le Conseil Régional Rhône-Alpes est Maître d'Ouvrage du SIM multitud.org; les données comprennent la BD horaires centralisée avec la position des arrêts (mais pas les tracés des itinéraires). Les données brutes appartiennent à la SNCF, la Région n'a pas le droit de les transmettre à des tiers telles quelles. En revanche, une fois intégrées dans le SIM Multitud, elles appartiennent à la Région (ou aux partenaires) et peuvent être utilisées selon les accords que ceux-ci ont passé (par ex. pour réalisation de cartes représentant les temps d'accès aux principales gares par l'agence d'urbanisme de Lyon). Elles sont alors réutilisables<sup>10</sup> mais une stratégie reste à définir, qui doit réunir l'accord des 12 partenaires – actuellement les données sont retransmises avec l'autorisation du partenaire concerné. Le service juridique de la Région travaille sur ce dossier.

### 2.3.6 L'ouverture des données au service d'une politique des transports

Régions et départements visent d'une part l'obtention des données pour elles-mêmes et les autres AOT, d'autre part l'ouverture pour que des services à la population et aux entreprises puissent se développer, optimisant les flux de transport.

La stratégie d'ouverture des données du département de l'Isère est caractéristique de ce point de vue. Aux rencontres d'Autrans en janvier

<sup>10</sup> ... à disposition soit au format XML/trident, soit via des web services.

2011, le responsable des transports expliquait qu'il s'agissait de poursuivre l'amélioration des possibilités pour la population tout en n'augmentant pas un budget déjà très élevé. C'est ainsi que le département teste un service d'autopartage dynamique intégré avec l'information sur les autres moyens de transport. Le projet Itinisère centralise l'information des différents réseaux (lignes, arrêts, horaires théoriques et dynamiques), les événements, les temps de parcours. Le CG est propriétaire des données. Il a signé des conventions avec les AOT et avec la SNCF.

### 2.3.7 Action essentielle, concertation

Le CETE Méditerranée constate en 2010 "qu'aucune AO ne met à disposition ses données horaires de manière ouverte selon les termes d'une licence d'utilisation, ni a fortiori de manière libre sur l'internet ; les données sont fournies au cas par cas, en fonction des demandes." L'imperfection des données et le fait que les AO ne sont pas en général propriétaires expliquent cette situation. Comme on pouvait le craindre, les SIM, indispensables pour gérer les transports, sont réticents à ouvrir leurs données à d'autres services, surtout commerciaux. De leur côté, les élus sont souvent méfiants face à des opérateurs privés qu'ils craignent de voir prendre une place excessive dans la gestion des données (un réutilisateur notait qu'une fois les conditions techniques ou économiques précisées, il fallait encore convaincre les structures de gouvernance du Syndicat mixte).

Les transports et l'information sur les transports sont une composante essentielle pour l'aménagement du territoire mais aussi pour l'optimisation de leur activité par les citoyens et les entreprises. Sur le fond, ce sont des données publiques mais le flou juridique qui règne encore à leur sujet, la multiplicité des intervenants et des statuts gêne non seulement leur utilisation et réutilisation par et pour les différents publics mais aussi les collectivités elles-mêmes. Il est très important de prendre en compte les

questions d'information dès la signature des délégations de service public et en général des contrats. Comme pour toutes les données publiques, des licences type doivent être mises en place. Il est aussi très important de se coordonner (cf. le rôle des Régions) et d'impulser la réflexion nationale sur le sujet, en liaison avec l'Etat et avec les grands opérateurs.

## 2.4 Autres domaines et points de vue : la diversité des données publiques

Certains domaines sont déjà le lieu d'activités, de débats, de réalisations. D'autres sont cités pour mémoire.

### 2.4.1 Vie et gestion des collectivités et marchés publics

Seront regroupées ici trois types principaux d'informations :

- Vie politique et organisation de l'acteur public. C'est-à-dire ici l'organisation de la collectivité, l'activité des élus, les débats et délibérations.
- Au delà, c'est l'analyse du travail des élus qui est observable et analysable. Le site «Nosdeputes.fr», utilisant les données du Parlement, informe sur l'assiduité des députés mais aussi, fait-il remarquer aux critiques, sur leur travail.
- Les budgets et indicateurs de gestion
- Les appels d'offre et marchés publics
- C'est un domaine très important et en rapide développement.

D'une part, il est fondamental pour la vie politique et, singulièrement, pour les tenants de la transparence et de l'open government.

D'autre part, il rencontre un marché : celui des collectivités elles-mêmes et celui des fournisseurs et prestataires de services. Le chapitre «Economie»

présente les offres existantes et en développement dans ce domaine et indique les gains en gestion et en efficience qui peuvent être attendus.

Déjà, Rennes, après la première vague de données, a ouvert ses données budgétaires.

En Grande Bretagne, le site «Where does my money go» permet l'analyse et la comparaison du budget national et des budgets régionaux. Des sociétés envisagent des services permettant de comparer données et indicateurs de gestion (Berger-Levrault – *ch.Economie*).

L'analyse des marchés publics se développe. Elle est proposée en France par Transpacités (*cf. Economie*). Pour les marchés européens, Euroalert permet aussi de recevoir informations et alertes. A l'Open Data Challenge organisé avec la Commission européenne, Eva Vozarova, de la Fair-play Alliance, Slovaquie, a développé une application permettant de rendre plus transparentes les procédures de marchés publics.

**Euroalert** : <http://euroalert.net>

Société espagnole qui travaille depuis 10 ans avec la Commission. Elle lance un service d'information sur les programmes européens, les appels d'offre dans les différents pays, les marchés. Concept simple mais difficile à mettre en œuvre, selon la société car : structures des données non normée, beaucoup de données, langues différentes, de multiples sources. Rémunération par les alertes, par l'analyse et la veille, publicité, vente d'applications adaptées...

### 2.4.2 Citoyenneté

Ce sont notamment les données électorales. La préparation de primaires politiques a montré que l'obtention de listes électorales, mais aussi le mode d'utilisation sont en débat. Les préfets ont eu au départ des réactions très différentes face à la demande de communication.

### 2.4.3 Droit, activité réglementaire

Au niveau national, ce domaine a lui aussi été l'un des premiers et des plus symboliques, avec les services de bases juridiques qui ont bataillé pour avoir accès aux fichiers du JO, la naissance de LegiFrance puis de la DILA.

S'agissant de données régionales et locales, le gisement est constitué par les circulaires, arrêtés et réglementations pris dans les domaines de compétence des communes et des départements. Ces informations ressortent d'abord de la vie quotidienne et de la transparence de l'action publique. Elles peuvent être prises en compte par des portails régionaux ou départementaux sur les droits et démarches.

#### Saône et Loire

L'assemblée départementale du 18 juin a décidé que les données du Conseil général seraient rendues publiques. Les internautes pourront consulter gratuitement, entre autres, les cadastres, les circulaires, les règlements et de manière générale tous les documents produits ou utilisés par la collectivité.

Arnaud Montebourg inscrit cette action dans le cadre d'une politique open data de transparence. Les données et informations pourront être complétées par les citoyens, dans une démarche de crowdsourcing.

*Remarque. Il ne s'agit pas à ce stade de réutilisation mais de publication interactive.*

### 2.4.4 Population et vie sociale

Ce domaine comprend des statistiques, dont beaucoup sont produites par les niveaux national et régional de l'INSEE et des administrations centrales. Elles peuvent être enrichies par toutes les informations émanant ou pouvant émaner des collectivités et des multiples associations (sportives, culture, activités diverses...) sur les territoires et croisées avec les données d'activité, celles des professions etc.

Il faut prendre garde aux données nominatives. Sur un territoire restreint, des données permettent d'identifier une personne par recoupements, ce qui peut être illicite pour certaines informations.

Des expériences intéressantes : la publication par les citoyens eux-mêmes de documents sur les lieux et modes de vie (d'eux-mêmes et du passé) (*Un travail important est réalisé à Plozévet, en pays bigouden* - <http://plozevet.hypotheses.org>). On touche là au domaine culturel.

### 2.4.5 Généalogie

L'épisode Généalogie, dans le développement du marché de l'information et de l'open data, a concentré au cours de 2011 tous les problèmes qu'il eût mieux valu éviter. En même temps, le cas est exemplaire et de ce point de vue utile, même si le débat est, selon certains, traumatisant pour les Conseils généraux.

Rappel : il s'agit, pour trois sociétés, NotreFamille (service genealogie.com), Ancestry (société américaine) et Geneanet d'avoir accès aux données d'état civil remontant à plus de 75 ans – en excluant de relever des informations sur les personnes nées depuis 120 ans, de les numériser en texte à partir des images souvent déjà publiées par beaucoup de départements, de les indexer et de permettre des recherches nationales. Il y a un marché pour des services évolués de généalogie, mais encore faible en France (5M€ vs 50M€ au Royaume Uni). La généalogie intéresse de plus en plus les français (61% ont déjà fait une recherche généalogique).

Or, hormis quelques cas, une forte résistance s'est fait jour dans les Conseils généraux, animée par les archivistes. Se mêlent ici la crainte des archivistes que leur travail ne soit pas valorisé, que leur site perde des visiteurs et leur département des budgets mais aussi un facteur culturel général, qui conduit à ne pas apprécier qu'une société privée utilise ce travail pour faire un bénéfice. Il y a aussi le fait de privilégier des relations établies avec les

généalogistes et sociétés de généalogie, familiaux ou successoraux.

- Premier élément de débat : la nature des données, qualifiées de culturelles parce que les services d'archive des CG dépendent de la Direction des Archives de France au ministère de la Culture. Or, les données culturelles sont exclues actuellement du champ de la Directive, et la qualification de donnée culturelle est liée à l'organisme gestionnaire, non au contenu des données. Cependant, l'exception culturelle a été réduite en France et elle ne s'exerce que pour les œuvres de création, en pratique.

- Deux fois, la CADA a constaté qu'il n'y avait pas lieu de refuser l'accès aux données pour réutilisation, d'autant qu'elles sont publiées et donc au moins accessibles.

- Deuxième élément : la protection de la personne. Selon des déclarations de l'Association des archivistes de France, il s'agirait d'un fichage généralisé, permettant notamment de stigmatiser telle ou telle ascendance. La CNIL a donc été saisie. Elle a simplement produit une recommandation sur les précautions à prendre – ce d'autant que les informations sont déjà accessibles.

- En juillet, le tribunal administratif de Clermont enjoignait au département du Cantal de donner accès à ses fichiers.

**Jugement sur l'accès aux archives** – Source : Vivien Mann, sur Site collaboratif GFII

“Les partisans de l'open data s'opposaient aux archives départementales du Cantal. En s'appuyant sur l'article 213-2 du code du patrimoine qui permet l'accès de tous aux archives publiques au bout de 75 ans, le site genealogie.com demandait que les archivistes les laissent accéder aux données d'Etat civil des cahiers de recensement de 1831 à 1931. C'est à la suite du refus opposé au site que le tribunal administratif de Clermont-Ferrand a été saisi, avant qu'il enjoigne les archives départementales, par une décision du 5 juillet dernier, de délivrer les fameux registres.

La juridiction administrative précise dans son jugement que les archivistes qui refusaient jusqu'alors de communiquer à une personne morale les données personnelles de nos aïeux « ne disposent pas d'un pouvoir discrétionnaire leur permettant d'apprécier l'opportunité de faire droit ou non à une demande de réutilisation », avant d'ajouter que la demande des requérants « ne peut en l'espèce être regardée comme présentant un caractère abusif » et annule donc le refus du Cantal.

- Une autre stratégie a été menée : des tarifs de licence prohibitifs (3 centimes la vue par an – une page d'un registre) – NotreFamille devrait verser 10M par an de redevances. Il s'agit de tarifs prohibitifs, ce qui est illégal. Quoi qu'il en soit, les sociétés devront négocier les tarifs avec chacun des CG. De nouvelles barrières sont donc possibles (cf. Ch 3 – *Redevances*).

L'affaire de la généalogie a, selon l'ADF, traumatisé les Conseils généraux. Elle a illustré l'incompréhension entre une profession, celle des archivistes, et les demandes aussi bien de sociétés commerciales que des partisans de l'open data plutôt orientés vers les services non commerciaux aux citoyens. Elle s'est nourrie des illusions sur un pactole, sur la crainte, plus légitime, de se voir dévalorisé. Elle devrait évoluer progressivement vers une détente, l'Association des Archivistes de France ayant organisé une réunion avec les sociétés. Elle montre que l'ouverture des données est difficile, et que c'est bien en se rencontrant qu'on réduit les tensions.

Au demeurant, un nombre croissant de CG ont maintenant ouvert ces données et ont publié les conditions de réutilisation (ex. Allier, Aveyron, Bas-Rhin, Haute-Garonne, Isère, Sarthe...). En fait, le problème de la généalogie et des fichiers d'état civil aura eu un impact positif sur l'analyse des problèmes et la mise en place de pratiques dépassionnées.

#### 2.4.6 Vie quotidienne, villes

C'est d'abord sur ce type d'informations que le mouvement open data se



développe de la façon la plus visible (cf. Ch.1). Dans cette première étape, les applications concernent un territoire unique – ville ou communauté urbaine – et un seul domaine, où les données sont rapprochées de l'information géographique. L'innovativité est ici exceptionnelle, portée par des individus, des associations, des entreprises en création. Elle se révèle dans les concours d'innovation dont le modèle a été celui lancé par Rennes Métropole, dans les manifestations organisées notamment par la FING et ses correspondants dans de nombreuses villes (cf. Cantines, la Mêlée à Toulouse etc.), dans des concours européens comme l'Open Data Challenge (<http://opendatachallenge.org/>).

Des types de données et d'applications sont cités ici (la plupart se retrouvent dans des domaines spécifiques également abordés ici – environnement, transport, culture etc.. Le point commun est ici en général la mobilité et la nature des innovateurs, souvent des individus, de petites associations, des start-ups, le territoire enfin, limité au plus à une communauté d'agglomération. La plupart des applications ont été relevées à Brest, Bordeaux, Montpellier, Paris, Rennes, ce qui n'est pas indiqué ici (cf. les sites correspondants).

#### ■ Des applications...

- Services publics : listes, emplacements, horaires des établissements scolaires, crèches et halte-garderies, piscines etc. Aussi, dans le domaine culturel : bibliothèques (horaires, catalogues, statistiques), musées
- Urbanisme : cartes, autorisations d'urbanisme, plans d'urbanisme plans de voirie...
- Gestion de l'espace urbain, des services et des événements : arrêtés d'insalubrité, chantiers de voirie et autres, eau, bornes de tri...
- Citoyenneté : résultats d'élection, cartes des bureaux de vote, cartes électorales, décès, mariages, naissances, répertoire des prénoms déclarés

- • Mobilité : taxis, pistes cyclables, accidentologie, stationnement (places en temps réel dans les parcs), vélos (l'Open Data Challenge a récompensé Oliver O'Brien de l'University College London, qui a développé une application permettant de visualiser la disponibilité des vélos urbains à chaque station de location, dans 30 villes du monde)

- Environnement : arbres, parcs et jardins (détail de la flore)

- Sports et loisirs, plans de circuits pédestres

- Handicapés : espaces accessibles, transports (Rennes, avec Keolis, permet aux handicapés de savoir quels ascenseurs permettant d'atteindre les stations de métro sont en panne)

#### ■ Caractéristiques et difficultés

- Crowdsourcing. Ce type d'applications se prête particulièrement bien à la participation des usagers, qui vont pouvoir compléter, indiquer en temps réel les erreurs, les événements (cf. *infra* qualité des données et cf. le site anglais <http://www.fixmystreet.com/>)

- Croisement de données. De nouveaux services sont possibles, à condition de croiser les informations, et donc de disposer d'un ensemble de données réutilisables, émanant de divers opérateurs sur le territoire concerné. Des difficultés surgissent ainsi pour les transports (cf. *infra* le 5 spécial). Beaucoup d'applications seraient possibles en associant l'information géographique, l'information transports associant les différents moyens et les autres types de données.

Nestoria (cf. <http://www.nestoria.co.uk/> et <http://blog.nestoria.fr/>) permet de rechercher des propriétés immobilières en combinant des caractéristiques du logement à des données sur la population, les équipements et services publics, la sécurité, les temps de transport. Le site "Where can I live" - <http://www.where-can-i-live.com/londonproperty> - utilise cet outil. Disponible

dans des villes espagnoles, allemandes, italiennes, le service ne l'est pas en France – notamment parce qu'à Paris la RATP refuse de donner accès à la géolocalisation des stations de métro, de même que les sociétés d'auto-partage. Il est bien sûr possible de géolocaliser soi-même...

Cependant, les collectivités s'efforcent de rassembler des données de multiples sources, y compris dans les transports. On sait qu'à Rennes, Keolis, opérateur des transports, a été un initiateur en coopération avec Rennes Métropole.

Les données de Montpellier (cf. sites) comprennent ainsi des informations de la Chambre de commerce (flux d'actifs sortant pour travailler dans l'Hérault), INSEE, Association TELABOTANICA (inventaire botanique).

#### ■ Le problème des « données locales »

Alors que les services en développement dans beaucoup de villes concernent des informations de vie quotidienne, les démarches d'accès aux services publics dans les sites Web de données locales comme Proximités ont peiné à trouver leur financement. Ces sites proposent des informations pour chaque territoire, jusqu'au niveau fin de la commune – en associant information générale et adresses et localisations de la commune concernée à partir d'une technique d'association de données appelée co-marquage et initiée et gérée par la DILA. Ils fournissent des renseignements sur les transports, la vie locale, les travaux en cours, les démarches, les équipements et services publics, les associations, les sites touristiques etc..

Il s'agit de démarches systématiques "d'en haut" – elle ne bénéficient pas de la dynamique suscitée par les jeunes innovateurs et par le crowdsourcing. Cependant, précisément, il faut savoir d'une part si la dynamique actuelle peut perdurer en l'état et surtout si elle peut assurer une égalité de traitement et d'accès entre les différentes zones, entre les grandes agglomérations et

les autres territoires. Elle permet d'obtenir des informations multiples mais au travers d'applications différentes (ce qui est plus adapté au mobile qu'au micro-ordinateur) et risque en outre de générer des ergonomies et interfaces très différentes, ou de concentrer les services chez quelques entreprises qui s'imposeront finalement. (Sur ces différents risques, cf. en particulier «Rôle de la Région»).

Devant l'impossibilité de trouver un modèle économique et donc un prestataire intéressé, la CDC a transféré à la DILA la responsabilité de la Base d'orientation de Service Public Local (SPL).

Des régions ont particulièrement insisté sur cette démarche – l'Auvergne avec Proximités Auvergne, insistant sur l'attractivité du territoire, Midi-Pyrénées avec le SIP (Service d'information publique, privilégiant la facilitation des démarches administratives).

Compte tenu de la difficulté de maintenir ces services, il paraît un peu théorique de s'interroger sur les possibilités de réutilisation de leurs contenus, ou plutôt des fichiers de données correspondants. Pourtant, des demandes pourront être présentées si les services perdurent et se multiplient. Jusqu'ici, comme pour tous les sites collaboratifs, les données sont en fait sous le régime qui est le leur chez le service source (DILA, région, commune etc.).

#### 2.4.7 Criminalité et risques

On aura peut-être noté que le moteur Nestoria cité ci-dessus prend en compte, pour l'immobilier, la description de la population du quartier et la sécurité. Ce type d'information, fréquent dans le monde anglo-saxon, est soit refusé par une majorité des citoyens soit tout simplement illégal en France. Des débats sont assurés d'émerger, parce qu'il existe une demande et parce que des données existent. Ce sera bien un choix politique que de les mettre à disposition ou de ne pas les rendre accessibles (ce qui est une

solution extrême, les journalistes entre autres, et des associations citoyennes pouvant légitimement s'en émouvoir). La circulation de ces données peut stigmatiser un territoire, une population – et il faut rappeler que les critères de réutilisation n'incluent pas l'objectif (cf. le chapitre 3).

Aux Etats-Unis, sur le répertoire data.gov, l'exploitation des données criminelles aurait représenté 80% des réutilisations en 2008 et 2009 (Hubert Guillaud, Internet.Actu, 16/09/09).

Crime map, police.uk... permet de visualiser les statistiques des crimes et délits autour de son lieu de résidence. Le lancement de l'application fin janvier 2011 a propulsé police.uk, le site qui l'héberge, en tête des sites les plus visités au Royaume-Uni, avec plusieurs millions de visiteurs lors du lancement, avant qu'il ne retombe progressivement à une audience plus faible, avec 200 000 visiteurs uniques pour mars 2011 (cf. document «Rapport Science et Technologie au Royaume Uni» – Ambassade de France – mars-avril 2011)

#### 2.4.8 Entreprises et activité économique

Pour ce type d'information, il est important que les différentes collectivités, les observatoires et les acteurs économiques (Chambres consulaires et syndicats professionnels) se concertent.

- Statistiques, données globales, créées ou détenues par les services compétents des collectivités, de l'Etat et des organisations professionnelles – des directions régionales de l'INSEE aux syndicats professionnels et observatoires spécialisés. Ces données, et leurs rapprochements, intéressent les acteurs de l'information économique locale comme ceux travaillant au niveau national et international.

- Fichiers d'entreprises. C'est une richesse que détiennent souvent les Chambres consulaires – c'est un de leurs rôles que de les tenir à jour. Des syndicats professionnels possèdent des observations détaillées sur leur

secteur. Théoriquement, une bonne part de ces informations constitue des données publiques, puisque les Chambres ont mission de les tenir à jour et sont financées par une taxe parafiscale. Dans la pratique, il s'agit d'un de leurs principaux acquis et la réutilisation éventuelle de ces données doit être organisée, le cas échéant, avec elles comme avec les syndicats et les autres organisations professionnelles, notamment agricoles, concernés. L'objectif peut être fixé avec les professionnels pour initier des ; il se peut aussi qu'il faille répondre à des réutilisateurs potentiels<sup>11</sup>.

- Données approfondies sur des entreprises et des projets, détenues par les collectivités, en particulier les résultats d'appels à projets et les programmes d'aide lancés dans le cadre des Contrats de programme Etat Région ou d'autres. Il faut distinguer les informations confidentielles que l'entreprise a pu fournir des projets et de l'aide qui a été apportée par la collectivité. Les Régions sont ici directement concernées.

Pour toutes ces données, l'important est de maintenir le secret des affaires, ce qui vaut aussi pour les statistiques, qui ne peuvent être publiées ou exploitées si, sur un territoire restreint, elles permettent d'identifier des informations confidentielles pour une entreprise. Les pratiques des statisticiens sont ici à respecter.

#### 2.4.9 Information financière et comptable

Pour mémoire : il s'agit d'un marché important au niveau national et mondial, qui a suscité de très nombreuses difficultés (en raison du caractère rentable de l'activité), dont la plus remarquable, encore qu'étrange, a consisté en le différent récurrent entre les greffes des tribunaux de commerce et l'INPI pour les bases constituées à partir des bilans des entreprises.

<sup>11</sup> Au niveau national, le problème resurgit fréquemment avec les données des CCI – notamment pour l'annuaire des entreprises de France (6 millions d'adresses) que des sociétés d'information réclament.

## 2.4.10 Infrastructures et services associés

Les données essentielles, du ressort des différentes collectivités, sont :

- les données géographiques dont les réseaux routiers, les rues, les réseaux physiques de transport (trans et trams), les canaux, les pistes cyclables, les sentiers etc.
- les réseaux d'énergie, de télécommunications, d'eau et les services de collecte et traitement des déchets

Les questions relatives à la géographie et aux transports ont été évoquées ci-dessus. Pour l'essentiel, les difficultés proviennent de l'incertitude sur la propriété de l'information par les grands opérateurs qui sont maintenant des sociétés, engagées dans un monde de concurrence.

### ■ Télécommunications,

... qui sont dans une phase rapide de croissance et une concurrence très dure. La relation plus récente avec les collectivités a été souvent très difficile, depuis le maillage Minitel jusqu'aux couvertures Internet, haut débit et mobile. Aujourd'hui se pose le problème de l'accès à des cartes exploitables pour les réseaux et pour les antennes. Les résistances sont nombreuses, même si l'ARCEP contraint les opérateurs à fournir une information traitable, chaque fois qu'une collectivité le demande. L'information sur les zones blanches n'est pas aujourd'hui vraiment complète ; les citoyens et les entreprises peuvent être gênés mais pourraient contribuer à compléter les cartes. Ici encore, il s'agit de données très importantes pour l'aménagement numérique et pour l'économie et il semble bien qu'il faille clarifier et mieux réguler les relations entre collectivités et opérateurs (en liaison avec l'ARCEP).

Eric Besson, à l'occasion du lancement de l'appel à projets européen "Open data Proxima Mobile", le 13 juillet 2011, souhaitait que la réutilisation soit

- possible: *La mise à disposition de la carte des antennes relais de téléphonie mobile pourrait engendrer des services d'alerte à proximité de zones blanches, avant interruption de la connexion.*

*Il va sans dire que, si l'obtention des données est déjà difficile pour les collectivités, la réutilisation n'est pas encore à l'ordre du jour. On sait cependant que des acteurs extérieurs de grande taille comme Google ou Apple aujourd'hui pourraient recréer l'information, de même que, malgré les objurgations de beaucoup et les risques en matière de protection de la vie privée, ils ont commencé de repérer tous les points WiFi (Google, Apple), qui constituent une très importante infrastructure virtuelle.*

### ■ Eau, énergie

La gestion de la ressource en eau devient de plus en plus importante, au travers des DREAL, des Directions régionales de l'agriculture, des Agences de bassin. L'identification des données et de leurs statuts se fera de plus en plus à l'échelle régionale, les directions départementales de l'Etat ayant disparu.

En ce qui concerne l'énergie, c'est surtout dans le futur que des évolutions importantes vont se produire avec d'une part le recours aux énergies alternatives à production locale (éoliennes, solaire), qui a déjà commencé et, d'autre part et simultanément, le développement de réseaux intelligents. Comme le chapitre sur l'économie l'explique, les opérateurs de services indiquent que le partage des données avec les collectivités mais aussi avec les utilisateurs et tous les acteurs concernés sera essentiel pour optimiser la gestion (*intervention de Suez environnement à la rencontre FING Opendata-Citélabos de mai 2011*).

*Il faut suivre ces évolutions, qui seront sensibles à moyen terme, et commencer de fixer les relations avec ceux des opérateurs les plus ouverts à la coopération.*

## ■ 2.4.11 Environnement

Les données relatives à l'environnement doivent être rapprochées les unes des autres et disposer d'une infrastructure géographique commune, ce que réalisent les plates-formes régionales construites ou développées en accord avec Inspire (cf. supra). Les acteurs sont d'abord ici les collectivités, les observatoires régionaux, départementaux, locaux qui se constituent, les associations, les services de l'Etat, les centres de recherche universitaire, tous en général partenaires des Centres régionaux de l'information géographique ou appelés à le devenir.

De nouveaux gisements de données se constituent, répertoriant dans le détail les ressources, la flore etc. La coopération des associations et des citoyens est ici de plus en plus importante (cf. l'exemple de Telabotanica à Montpellier) et il y a beaucoup à faire avec les professionnels comme avec les scientifiques ou les promeneurs (cf. ci-dessus l'économie montagnarde et les activités pastorales – APEM).

Les Régions disposent en particulier d'informations importantes, dont celles relatives aux Parcs naturels.

Dans ce domaine, la coopération avec les régions et les pays voisins est souvent une nécessité et une opportunité (beaucoup de parcs sont transfrontaliers).

#### 2.4.12 Tourisme

Les données sur le tourisme sont de celles qui circulent le plus, même si elles posent de nombreux problèmes. Les enjeux sont en effet très importants en termes d'activité économique et d'attractivité du territoire.

##### ■ Des données complexes, un paysage animé

Les données émanent d'une multiplicité d'acteurs et sont gérées par les Offices du tourisme et les Comités départementaux et régionaux. Le tourisme

est un thème fort d'échange, de communication et de normalisation des informations – car elles doivent évidemment pouvoir être utilisées par des acteurs extérieurs. Il pose donc aussi de nombreux problèmes.

- L'association étroite entre des données générales et des informations commerciales ou à fort impact commercial est délicate. Les professionnels veulent en même temps diffuser leur information et maintenir un avantage. Ainsi, la publication comme la réutilisation ne doivent pas engager l'acteur public dans des problèmes de classement, en dehors des classements réglementaires en étoiles pour les hôtels, par exemple.

Il est difficile de diffuser une information neutre : ainsi, les hôteliers commencent à s'irriter de voir des chambres d'hôte en ville. En même temps, l'information sur cette nouvelle catégorie de prestation doit être placée sur le même plan que les autres. Paradoxalement, c'est un cas où le fait de transférer l'information à des réutilisateurs privés dégage la collectivité d'une responsabilité qui n'est pas la sienne, puisqu'il s'agit d'arbitrer une compétition entre deux types de prestations.

Il n'empêche que l'information doit être fournie et, en outre, publiée sur des sites publics. Il faut donc identifier les données librement accessibles sur les entreprises du tourisme (adresse, classement officiel etc.) et celles que les acteurs concernés acceptent d'intégrer à l'information publique.

Constitution d'acteurs coopératifs, les Offices et Comités de tourisme, dont les données ont souvent un statut flou, et qui sont par nature eux-mêmes dans une position ambiguë par rapport à la réutilisation de leurs données (droit, exploitation directe, commentaires et traitement des données, problèmes de concurrence)

Concurrence entre les territoires

- Lien entre les informations géographiques, les informations de vie quotidienne et d'environnement et les informations touristiques. Cela

pose des problèmes politiques, non techniques – cf. information sur la pollution des plages, voire statistiques sur le climat.

- Souhait de voir se développer des services locaux, innovants mais aussi nécessité de voir l'information diffusée le plus largement possible par de grands opérateurs.

- Relation difficile avec les partenaires des collectivités qui sont en même temps opérateurs commerciaux de tourisme (cas typique, la SNCF).

- Intervention de plus en plus massive de personnes qui donnent leur avis, donc utilisation du Web 2, mais aussi nécessité de contrôler ce qui se dit dans des sites importants, au cas où des informations seraient déformées, ou préjudiciables à tel ou tel intervenant, et où une collectivité ou un Office risquerait de se voir impliqué (cf. *infra qualité*).

- Relation très forte avec les données culturelles – sites touristiques, musées, événements (concerts, festivals) etc. et même plus généralement toute l'information sur la culture régionale, facteur d'attractivité (cf. *infra les problèmes de l'exception culturelle*). Certains textes, certaines images peuvent être grevés de droits : droit d'auteur, droit à l'image.

#### ■ Les systèmes d'information touristique régionaux

L'organisation de systèmes d'information touristique au niveau régional rassemble tous les intervenants avec deux objectifs principaux : assurer une synergie dans la promotion d'ensemble de la région, fournir des informations et des outils aux multiples partenaires. Ces systèmes sont donc de plus en plus un lieu d'échange et de communication de données. Ils permettent aux uns d'utiliser et intégrer certaines données des autres – aussi bien des listes de prestataires que des images etc.. Ils peuvent aussi mettre à disposition des données provenant d'organismes scientifiques et culturels. Ils doivent

donc définir des règles de mise à disposition, entre les partenaires. Ils peuvent aussi le faire en direction de réutilisateurs non membres mais cette évolution semble encore rare : comme dans tout système coopératif de ce type, il est difficile de s'accorder sur une licence générale de réutilisation.

#### Ex. Système d'information touristique aquitain - Sirtaqui

(<http://partenaires.tourisme-aquitaine.fr/>)

Le Sirtaqui définit ses données comme publiques et a choisi une licence Creative Commons (cf. "licences"). Les règles retenues :

- refuser que des modifications soient apportées aux photos ou vidéos,
- interdire leur réutilisation à des fins commerciales,
- imposer au réutilisateur de mettre ces œuvres à disposition du public sous ces mêmes options Creative Commons.<sup>12</sup>

Les usages autorisés : usage interne dans le cadre d'une activité économique, réutilisation commerciale en vue de l'élaboration par le Licencié d'un produit ou d'un service destiné à être mis à disposition de tiers à titre gratuit, réutilisation non commerciale. Une plate-forme de téléchargement en libre service est ouverte aux licenciés.

Ces règles signifient notamment que le réutilisateur ne peut pas, par exemple, reprendre une image pour en faire bénéficier seulement les hôtels membres d'une chaîne.

#### 2.4.13 Education - Recherche

Deux types de données sont accumulées dans ce vaste domaine : d'une part des contenus éducatifs, scientifiques, techniques (*repris ci-après pour certains*), d'autre part des données de gestion des établissements

■ 12 Licence CC By NC SA

et institutions ainsi que des données relatives à l'activité, impliquant les bénéficiaires / utilisateurs, dont l'exemple le plus visible est constitué par les résultats d'examen (donnée au demeurant publique, en tout cas de publication officielle obligatoire).

La confusion entre ces deux domaines vient de ce que, par défaut, une donnée est analysée à partir de l'organisme qui la gère.

Les différentes données sont donc susceptibles d'être grevées de droits d'auteur attachés et peuvent contenir des informations nominatives, dont certaines ne seraient pas accessibles.

Il s'agit en tout cas de gisements impressionnants, y compris dans les données de gestion et d'activité. Beaucoup de ces données (dont par exemple localisations, horaires, dates de concours etc.) sont gérées par le rectorat et peuvent être associées à des données locales (*Xavier Crouan, qui a organisé l'ouverture des données de Rennes Métropole, confirme qu'il n'y a eu aucune difficulté à obtenir pour réutilisation les données du rectorat*).

#### 2.4.14 Formation

La situation semble très diverse dans les régions pour un domaine à l'évidence très important et de la compétence des Conseils régionaux. Des entreprises se plaignent souvent de ne pas pouvoir identifier clairement les offres de stages, d'apprentissage, mais aussi les offres de formation. Une des raisons en est la multiplicité des intervenants, des statuts, des pratiques (un responsable régional parle de fouillis). Cependant des sites se mettent en place, comme <http://www.poitou-charentes-alternance.fr/> pour rapprocher offres et demandes de contrats d'apprentissage ou d'alternance.

Le filtrage de données publiques sera ici difficile. Des demandes postées sur un site par une personne peuvent se retrouver sur d'autres, que cette personne refuse de cautionner ainsi. Pourtant, il est tout aussi certain que des associations, sociétés, organismes privés pourraient jouer un rôle dans

la fluidité du monde de la formation et des offres et demandes. Le domaine est à explorer de ce point de vue.

#### 2.4.15 Information scientifique et technique

Les Universités et laboratoires publics entrent de plus en plus dans le mouvement de l'Open Access, un des volets de la démarche générale "Open". (*cf. sur l'Open Access les documents – en France voir l'archive national HAL*).

L'autre volet de l'IST est constitué par des informations émanant de sociétés, d'associations scientifiques, de groupements techniques professionnels et de l'édition scientifique. Les thèmes et les organisations varient d'une région à une autre et d'un territoire à un autre (par ex les informations sur les différentes activités agricoles). Au niveau national et international, de grandes bases concernent les brevets, d'autres les formules chimiques, les médicaments par exemple.

La Commission a commencé en juillet 2011 les consultations sur l'inclusion de l'information scientifique dans le champ de la réutilisation.

Dans un cadre régional, il semble important de vérifier avec les acteurs intéressés et éventuellement des acteurs extérieurs quels domaines pourraient bénéficier d'une mise en commun de données et d'une visibilité accrue. Les groupes de travail constitués autour de filières scientifiques ou économiques peuvent jouer ici un rôle.

#### 2.4.16 Données santé et sociales

##### ■ Un secteur hors compétence de tel ou tel ?

La référence aux données sanitaires et sociales provoque toujours une réaction de prudence, ces informations étant avant tout des données nominatives.

Par ailleurs, les données correspondantes apparaissent comme distinctes et gérées par des organismes différents, avec de solides frontières : la santé par l'ARS, le social par le Département.

Pourtant, les frontières ne tiennent pas et le monde sanitaire et social ne peut plus être isolé et fragmenté.

- Importance croissante du médico-social, prise en charge de la dépendance, soutien à des maisons médicales : toutes les collectivités sont impliquées

- Relation forte entre ces domaines et des problèmes d'équipement (par ex. la couverture numérique, les services de secours par les pompiers)

- L'importance de l'information est de plus en plus grande, et réclamée, par la population, par des professions, par des acteurs économiques (par ex pour délivrer des services non médicaux aux personnes dépendantes)

#### ■ *Partenariats et implication*

Les collectivités sont impliquées et consultées :

- Consultation pour le Projet régional de santé, le Plan régional de Santé publique, présence dans la Conférence régionale de la santé et de l'autonomie, les Conférences de territoire

- Gestion d'établissements et de services ou soutien par les CG et communes (CCAS, EHPAD, CLIC, Coordinations gérontologiques etc.)

- Gestion de l'APA et de la PCH par le CG

#### ■ *Des informations multiples*

Hors l'information sur les bénéficiaires (tant qu'elle n'est pas anonymisée) et l'information scientifique médicale, plusieurs types d'information sont du domaine public et peuvent être du ressort des collectivités :

- Professionnels et équipements médicaux, sanitaires, médico-sociaux mais aussi lieux d'assistance sociale (caractéristiques, localisation, horaires etc.)

- Statistiques sur les populations aidées, dépendantes etc.

- Données de santé publique

Certes, ces informations sont parfois délicates du point de vue du gestionnaire public. Les données de santé publique, notamment, peuvent inquiéter et être mal interprétées (elles peuvent aussi inquiéter parce qu'elles signalent un problème sur un territoire – c'est un débat politique et un enjeu démocratique).

#### ■ *Données sur l'assistance sociale, la dépendance*

Malgré les craintes aussi sur ce dossier, il est incontournable de définir et gérer les types de données dont vont avoir besoin les Conseils généraux et les intervenants en nombre croissant (associations et sociétés qui sont actives sur le domaine).

**L'APEM** (Association pyrénéenne d'économie montagnarde) a développé une approche très intéressante : une analyse isochrone qui concerne particulièrement les personnes dépendantes : utilisant les données géographiques, les informations sur le transport, les informations sur les équipements et services, elle permet de répondre à la question : « à quelle distance, en temps, est telle personne de tel service » ?

*Les collectivités seront de plus en plus impliquées, avec toute la société, dans le fonctionnement du système de santé, dans la prévention, la gestion de la dépendance comme dans les questions de santé publique et leur relation avec l'environnement. Des données publiques sont nécessaires, d'autres seront produites. Des acteurs anciens et nouveaux vont intervenir et des réutilisations, encore pratiquement inexistantes, interviendront.*



## 2.4.17 Données culturelles

### ■ Définition

La notion de donnée culturelle renvoie à une définition de la culture, qui ne sera pas discutée ici.

Légalement, en France, en tout cas dans l'application première de la Directive, la donnée culturelle est celle dont la responsabilité est confiée à un acteur identifié comme du monde de la culture. C'est ainsi que les archives d'état-civil, établies par les communes mais gérées par les services départementaux des archives ont pu être identifiées comme culturelles, ce service dépendant de la Direction des Archives de France.<sup>13</sup>

Progressivement cependant, un champ plus restreint se dessine : il ne serait pas normal que le budget d'un musée ne soit pas accessible aux citoyens (dans le cas où le musée est public), autant que celui d'une école.

L'ensemble de ces données est très vaste - monuments, poèmes, catalogues des bibliothèques, ouvrages numérisés, photos, collection des musées comme spectacles des théâtres..

Dans tout cet ensemble :

- la distinction évoquée ci-dessus reste parfois difficile à établir – si le budget d'un théâtre n'est pas, à l'évidence, une donnée culturelle, qu'en est-il du programme ? de l'inventaire sans commentaires des collections d'un musée? des catalogues des bibliothèques?

<sup>13</sup> Des réutilisateurs trouvent un point commun aux institutions culturelles: une résistance à échanger avec un monde extérieur, surtout commercial. L'expérience a montré, dans les musées français, des refus de confier ne fût-ce que quelques images et ne fût-ce que pour des sites universitaires. Ces cultures évoluent. Par ailleurs les musées pour les jeunes et les musées scientifiques auraient un comportement plus ouvert.

- beaucoup des données et documents sont grevés d'un ou plusieurs droits d'auteur et de droits de propriété intellectuelle (le droit d'auteur interdit de transformer l'objet sans autorisation de l'auteur ou de ses ayant-droit – le droit de propriété intellectuelle interdit de l'utiliser sans autorisation du propriétaire)<sup>14</sup>

- beaucoup d'objets sont grevés de droit à l'image (parfois exercés par des descendants de personnes photographiées par exemple)

Cependant, toutes ces données sont très importantes non seulement dans le secteur des activités culturelles mais pour l'attractivité des territoires, le tourisme, les recherches en sciences humaines et sociales, les banques d'images commerciales etc.

### ■ Conditions juridiques de la réutilisation

La directive européenne les excluait, la France a étendu : l'organisme peut décider de la réutilisation et de la tarification mais par défaut les données sont des données publiques réutilisables si accessibles.

Un débat oppose les tenants de l'exception culturelle et ceux qui veulent appliquer le droit commun.

- En 2009, un groupe de travail sur le patrimoine culturel numérisé, co-présidé par Bruno Ory-Lavollée, a remis à la ministre de la Culture des «Propositions pour une charte de la diffusion et de la réutilisation des données publiques culturelles numériques» (*f. document*). Il rappelait la loi CADA de 1978 : "les conditions de réutilisation des informations contenues dans des documents élaborés ou détenus par

<sup>14</sup> Certains cas peuvent ne pas avoir été prévus, par exemple pour les œuvres architecturales – il est ainsi impossible de réutiliser une photo de la pyramide du Louvre à des fins commerciales, par exemple pour une carte postale ou un livre, sans traiter avec l'architecte auteur (leoh ming Pei).

des établissements, organismes ou services culturels peuvent être, par dérogation, librement fixées par ces derniers." Le rapport présentait plusieurs propositions, notamment :

- Veiller à une signature sur les œuvres
- Associer des données géographiques aux données numérisées
- Favoriser la réutilisation par les éditeurs de sites Internet
- Adapter aux nouveaux usages - sites collaboratifs, réseaux sociaux
- Proposer une licence « clic » immédiate et la gratuité pour les utilisations non commerciales
- Tarifier aussi bas que possible pour les réutilisations commerciales

Le ministère de la culture confirmait en mars 2011 la spécificité des données culturelles (*article de Camille Domange <http://cblog.culture.fr/2011/03/01/la-valorisation-des-donnees-publiques-culturelles-moteur-dune-economie-numerique-de-la-culture> et document*): "La nature même de ces données ne peut, sans précaution, être assimilée au vaste ensemble des données publiques administratives assujetties au traitement de droit commun prévu par la loi du 17 juillet 1978. C'est la raison pour laquelle le législateur a instauré une dérogation au principe de libre réutilisation qui assure aux établissements, organismes ou services culturels détenteurs et/ou producteurs de données publiques la possibilité de déterminer le régime de réutilisation applicable. Les opérateurs culturels peuvent ainsi conserver une certaine latitude pour écarter ou limiter la réutilisation de certaines données sensibles (fonds d'archives de guerre ou données nominatives) ou données protégées par un droit de propriété littéraire et artistique ou industrielle."

Cependant, les arguments sont de plus en plus contestés. En fait, ceux qui portent sur les données nominatives, le droit d'auteur, les données sensibles sont généraux et ne fondent aucune spécificité par rapport aux autres données. (cf. article « De l'inutilité de l'exception culturelle en matière de

réutilisation des données publiques» – par «Calimaq» – mai 2011 – document et <http://scinfolex.wordpress.com/author/calimaq/>

La Commission envisage de réintégrer les données culturelles dans les conditions générales lors de la révision de la Directive. Elle a publié une étude « PSI re-use in the cultural sector » (mai 2011) (cf. document).

#### ■ Economie

Si les données culturelles sont pour l'instant souvent exclues de la réutilisation, ce n'est pas seulement en raison du conservatisme de responsables mais aussi en raison d'analyses économiques. C'est d'ailleurs l'objet de l'étude européenne : mesurer l'impact direct et indirect.

Séverin Naudet, directeur d'Etalab, expliquait qu'il y a un patrimoine fabuleux en France, et qu'il faut peut-être que la diffusion d'images et d'autres données contribue au financement.

Il faut donc balancer les impacts positifs et négatifs et considérer les types de réutilisation – ce qui n'est pas le cas ailleurs. Ainsi, un voyageur qui a besoin d'une image pour vanter un territoire devrait sans doute pouvoir l'obtenir et l'utiliser sans difficultés, alors qu'un publicitaire voulant illustrer un produit pourrait contribuer. Comme toujours, ce type de distinction sera souvent très difficile à mettre en œuvre et à contrôler.

Par ailleurs, l'argument du financement peut valoir pour de grands fonds et des monuments ou collections très connus, alors qu'un musée local a tout intérêt à atteindre une exposition maximale, ainsi que le territoire où il se trouve (l'exemple du musée Ingres à Montauban est cité dans l'article "L'exception culturelle se glisse dans l'ère du numérique" (*LaGazette.fr* – novembre 2011 – cf. document).

La situation est aussi particulière pour des organismes spécialisés en charge de la gestion de grands fonds d'archives et de leur réutilisation. C'est le cas de l'Institut national de l'audiovisuel (INA).

L'INA, rappelle l'étude européenne déjà citée, a comme ressource la plus importante la fourniture commerciale de documents : plus de 14M€ en 2010. Il offre 35000 heures de vidéo gratuitement sur son site public et 675000heures sur Inamédiapro à un marché professionnel : agences de publicité, chaînes TV, sites Internet...

#### ■ Evolution des procédures et des techniques

Le besoin de financement conduit désormais les institutions à rechercher des partenariats. Cela posera rapidement le problème de la réutilisation des données gérées avec les partenaires, mais traduit aussi une évolution vers la prise en compte du marché. L'appel à projets de Partenariat public privé lancé par la BNF pour la numérisation de collections est caractéristique : la Bibliothèque recherche un partenaire qui supportera la numérisation et pourra ensuite commercialiser des éléments.<sup>15</sup>

Un autre intervenant important et précieux est désormais le public. Cela s'est vérifié sur plusieurs fonds. L'exemple cité par l'article de LaGazette.fr (cf. ci-dessus) est celui de la La bibliothèque municipale de Toulouse qui, depuis juin 2008, diffuse le fonds « Trutat » (Collection de photos prises à la fin du 19<sup>e</sup> siècle par Eugène Trutat, directeur du Muséum d'histoire naturelle) sur Flickr, site de gestion et de partage de photos. L'objectif était d'enrichir la collection à partir de contributions extérieures permettant de localiser certains clichés pour lesquels la bibliothèque n'avait aucune mention de lieu. Or, des commentaires très intéressants ont été reçus.

<sup>15</sup> Google avait proposé de numériser mais il restait maître des usages – ici, il va s'agir d'un PPP. Néanmoins, il restera à savoir comment des contrats pourront être passés avec d'autres acteurs sur ces données. La National Gallery a eu une expérience avec Microsoft à l'époque des vidéodisques qui déléguait à Microsoft la gestion de la réutilisation, avec en pratique un monopole contre ses concurrents.

#### ■ Hack4Europe! Une exploration des possibilités.

(<http://data.gov.au/tag/libraryhack-2011/>)

*Hack4Europe! a été organisé par l'Europeana Foundation et ses partenaires, le Collections Trust, le Museu Picasso, le Poznan Supercomputing and Networking Centre et le Swedish National Heritage Board. Il s'agit d'une série de «hack days» (journées d'«impro informatique») qui se sont tenues à Londres, Barcelone, Poznan et Stockholm du 6 au 12 juin, et qui ont permis d'explorer, dans un environnement stimulant, les possibilités qu'offrent les données culturelles ouvertes pour favoriser le développement social et économique de l'Europe. Les 60 participants étaient pour l'essentiel des PME telles que des agences de développement web, des développeurs d'applications, des développeurs de logiciels et d'autres entreprises actives dans le domaine numérique. Se sont joints à eux non seulement des développeurs du secteur du patrimoine culturel, désireux de créer de nouvelles manières d'inciter le public à utiliser les ressources culturelles en ligne, mais aussi certains acteurs de grand calibre tels que le Google Technical Group et le Yahoo Research Group en Espagne.*

#### **Les lauréats de Hack4Europe!**

**Royaume-Uni :** Michael Selway de System Simulation Ltd., qui a développé une application pour obtenir de meilleurs résultats de recherche depuis Europeana à partir des appareils Android.

**Espagne :** Eduardo Graells et Luca Chiarandini de l'Universitat Pompeu Fabra / Yahoo! Research Barcelone, qui a créé un "Timebook" pour les personnages historiques. Son application intègre du contenu d'Europeana et de DBpedia et le présente sous une forme conviviale, avec pour chaque personnage, ses citations présentées sous forme de "posts", les personnes qui l'ont influencé – sous la forme d'«amis», ou encore des photos de ses œuvres.

**Pologne :** Jakub Jurkiewicz d'iTraff Technology. Sur la base du jeu de données d'Europeana, ce lauréat a développé une application capable

de traiter la photo de n'importe quelle peinture prise dans un musée et qui en donne, après quelques instants, une description dans toutes les langues de l'UE, et même à haute voix.

**Suède** : Martin Duveborg du Conseil national du patrimoine suédois, qui a développé pour Android un système pleinement fonctionnel de recherche d'Europeana basé sur la géolocalisation. Les utilisateurs peuvent prendre des photos et les associer à des objets existants d'Europeana. Via une fonction intégrée permettant de superposer de nouvelles photos aux images d'Europeana, l'application permet de visualiser la transition entre le passé et le présent. Les nouvelles photos sont envoyées avec la position GPS, ce qui permet à l'application de servir en outre d'outil de balisage géographique (geotagging).

■ *Un gisement à exploiter : exemple de l'opportunité identifiée en PACA*

*Le patrimoine culturel français est impressionnant et il est un atout important pour tous. Des initiatives régionales apparaissent possibles, en associant les institutions volontaires et les autres acteurs. En PACA, la Région et ses partenaires ont saisi l'opportunité « Marseille capitale européenne de la culture en 2013 pour initier le premier grand chantier régional d'ouverture et de réutilisation des données publiques » (cf. chapitre Mise en œuvre).*

## 2.5 Chaîne des données : qualité, validation, contrôles

### 2.5.1 Problématique

■ *Qualité*

La question de la qualité des données se retrouve au long de la chaîne et la modification de celle-ci, notamment en raison des nouveaux outils sur les réseaux, aura de toutes façons un impact.

□ Etant donné que beaucoup de débats, on va le voir, portent sur la qualité, la première question à se poser est de savoir de quoi il s'agit :

La qualité dépend aussi de celui qui observe la donnée. Par exemple, le positionnement géographique peut aller de quelques mètres à quelques dizaines de cm – précision souhaitée par exemple par les archéologues. Il en est de même pour la plupart des données.

La qualité d'un jeu de données vient aussi de son actualisation (par ex. les photos aériennes sont prises à une certaine fréquence; les horaires théoriques de transport doivent être actualisés aussi vite que possible et par définition, les horaires temps réel doivent être mis à jour en permanence (liaison dynamique aux opérateurs).

Elle est liée à la fiabilité de la saisie (qu'il s'agisse d'humains ou de capteurs, il y a toujours un taux d'erreur)

La qualité est appréciée en fonction de ce qui est promis, et il faut préciser le périmètre et le champ et les indiquer – y compris les variations – pour définir une exhaustivité.

La qualité peut aussi être technique : garantie de la mise à disposition sans interruption importante (garantie sur les serveurs – essentielle pour les réutilisateurs), stabilité des formats et des interfaces, respect ou non de normes pour les données et les métadonnées.

■ *Ouverture : les craintes*

Un des arguments pour ralentir ou refuser l'ouverture d'un jeu de données est qu'il faut assurer la qualité. La FING relève que, selon certains, les gens ont un peu peur qu'on voie leurs données, que la qualité ne soit pas bonne. C'est vu par les plus militants comme un prétexte ou un faux problème. La qualité n'est importante que par rapport à l'usage. Il faut démarrer et attendre le retour. "Si on pose d'abord la question de la qualité, on ne fait rien."

## ■ Validation des données mises à disposition

La question est posée de la validation des données. On verra que c'est une tâche difficile pour beaucoup de données provenant de saisies externes. La règle fondamentale énoncée au début de ce chapitre est valable : la qualité ne peut se valider qu'au plus près des acteurs concernés par la donnée et il faut qu'il y ait une seule filière et un seul service responsable pour chaque jeu de données.

### ■ Réutilisation, responsabilité, dénaturation, contrôle ?

C'est un débat récurrent et de fond : quelle est la responsabilité de l'acteur public qui a ouvert un jeu de données ? En théorie, elle est d'autant plus nulle que, pour une donnée déclarée réutilisable, tous les réutilisateurs sont traités également et qu'il n'y a pas lieu de demander pour quel usage la donnée est transférée.

Or, les textes indiquent que la donnée ne doit pas être "dénaturée". La difficulté est ici de préciser ce que peut être la dénaturation (*cf. infra*).

Aucun texte en tout cas ne prévoit que l'acteur public doive vérifier l'usage des données. On verra cependant qu'il a tout intérêt à le faire pour les améliorer, qu'il a intérêt à rester en contact avec les réutilisateurs, qu'il a, comme toute entreprise, parfois besoin de vérifier que les données ne sont pas utilisées de manière à gêner son action ou à commettre des actions illicites. Il n'y a là rien qui tienne à la réutilisation : il faut de toutes façons surveiller sur le Web, autant que possible, les sites qui sont susceptibles de léser l'acteur public. En tout état de cause, un contrôle systématique est évidemment impossible sur le Web.

Il faut rappeler que, si l'intention n'a pas à être annoncée (tout réutilisateur, de toutes façons, déclarera qu'il ne dénature pas et qu'il n'a pas d'intention criminelle), la justice permet toujours d'attaquer un site mensonger ou

□ illicite, qu'il utilise ou non les données du plaignant. Là encore, il n'y a rien de particulier : le dénigrement peut par exemple être attaqué.

### ■ Rappel : coûts

Il faut toujours le répéter : toute amélioration de la qualité passera au départ par un investissement ; en dehors des questions de réutilisation, elle permettra en général ensuite d'éviter des erreurs et permettra des améliorations de fonctionnement.

## 2.5.2 Impact de la réutilisation

Le principe est simple : s'il ne faut pas attendre une qualité parfois mythique et dépendant des usages, la réutilisation et en général la circulation des données va conduire à leur amélioration.

Les responsables informatiques insistent au demeurant sur l'importance d'améliorer, en tout état de cause, la qualité et la sécurité des données des collectivités. La réutilisation va souvent stimuler la démarche.

### *La rétroaction est essentielle*

Même si ce n'est pas obligatoire, il est très important d'avoir connaissance des utilisations et d'avoir un retour sur la mise à disposition. C'est d'abord avec les réutilisateurs que le contact doit être maintenu – ceci est toujours souligné par les sociétés de l'information, d'autant que les techniques et les données elles-mêmes évoluent. [Remarque : si des données sont erronées dans ce qui a été mis à disposition, ce sera aux réutilisateurs de se retourner vers l'acteur public – car il est certain que des garanties réciproques devront intervenir dans les licences et contrats.]

Ensuite, il est aussi très important d'avoir un retour des citoyens et des autres utilisateurs : c'est en dernière analyse pour eux que la collectivité a

ouvert ses données – qu’il s’agisse des citoyens du territoire ou, par exemple, d’usagers étrangers. Ceci étant, c’est d’abord sur le territoire que les retours seront connus – sauf si des marchés larges ont été atteints. Comme pour toute action, il est important de mesurer l’impact réel.

En termes d’amélioration des services publics et de l’administration, l’open data peut permettre à la collectivité de mieux comprendre des usages, des besoins des citoyens – même si c’est au travers de services produits par des entreprises réutilisatrices.

### 2.5.3 Le contrôle des mésusages : la dénaturation

Comme cela a été posé dans la problématique, la collectivité ou en général l’acteur public n’est en rien responsable des produits des réutilisateurs. Tant que c’est possible, il est, on l’a vu, intéressant pour elle de connaître les usages et de viser à une amélioration si elle le juge nécessaire, en concertation avec les réutilisateurs.

La notion de mésusage ne dépend pas de la réutilisation de telle ou telle donnée et surtout pas de la nature du site ou du service. Certains ont souhaité que des sites qui se révéleraient être sectaires, par exemple, ne puissent être acceptés comme réutilisateurs. Ce serait donner à l’acteur public considéré la possibilité de déterminer si un site est licite ou non. C’est là la vraie question, et c’est au tribunal de trancher, certainement pas à une collectivité, qui doit en revanche, comme tout citoyen, signaler les fautes qu’elle peut relever.

La question de la dénaturation est bien plus difficile, et elle est source de conflits potentiels. En effet, certaines licences (APIE) prévoient qu’il faut exiger du réutilisateur qu’il s’engage à ne pas dénaturer les données. Cela permettrait, paraît-il, de rassurer les services publics qui craignent d’ouvrir leurs données. Le problème est que personne ne peut définir a priori la dénaturation. En fait, la réutilisation n’a d’intérêt que si les données sont

au moins mises en relation avec d’autres, mais souvent aussi transformées, retraitées, commentées. Comme cela a été dit, une déformation volontaire (une valeur fausse glissée dans une série), un commentaire inacceptable, peuvent être attaqués – que la donnée soit une donnée réutilisée ou une autre. Au contraire, le danger serait grand de voir des élus ou des fonctionnaires chercher à contrôler les analyses faites à partir des données.

Si donc une donnée est jugée sensible et ne pouvant être interprétée que sous contrôle, il ne faut pas la déclarer accessible – ce qui est du ressort de l’acteur public. Cette décision pourra faire l’objet d’un recours à la CADA. Dans le cas où celle-ci concluerait que le refus de réutilisation n’est pas fondé, ce serait alors à la collectivité d’observer l’utilisation et, le cas échéant, de porter plainte si le véritable mésusage qu’elle redoutait se produit (mais sans se référer précisément à la licence, plutôt aux lois...).

Dans tous les cas, si les licences sont compatibles avec les licences internationales (on verra que c’est indispensable), elles ne peuvent inclure le concept flou de dénaturation.

### 2.5.4 Responsabilité ?

Pour autant, la collectivité ne peut espérer ni même souhaiter dégager entièrement sa responsabilité pour des données critiques, importantes pour les citoyens. C’est une autre raison pour laquelle elle ne peut se désintéresser des usages, surtout dans une période de transition. Il est certain que, si des erreurs importantes ou des actions illicites portaient sur un service et des données visibles réutilisées, il se trouverait des personnes pour se tourner vers elle et soit lui attribuer les erreurs, soit lui reprocher d’avoir laissé le réutilisateur incriminé accéder aux données.

C’est la responsabilité du politique et de l’Administration : les citoyens se tourneront vers eux et exigeront que de tels errements ne se reproduisent pas. D’ailleurs, si la RATP par exemple, interdit l’usage de son nom dans le

titre d'un service, c'est aussi parce qu'elle redoute qu'on lui en attribue les erreurs en même temps que la paternité. Beaucoup de services non essentiels qui seront assurés par des réutilisateurs seront attribués à l'acteur public.

Une fois encore, cette menace ne justifie pas de refuser l'entrée dans la société de l'information, où chacun connaît les difficultés de tout contrôle.

Des précautions peuvent cependant être prises.

Il faudrait que les sites et services des réutilisateurs indiquent d'une part que des données publiques sont utilisées, d'autre part que l'acteur public n'est pas responsable du contenu du site.

La seconde précaution est d'assurer la traçabilité de la donnée. Il ne faut pas voir là un contrôle sur les usages et notamment sur les réutilisations en cascade. Apposer un identifiant indélébile sur les documents est possible, voire souhaitable pour bien d'autres raisons que la réutilisation (c'est un « object identifier » très utile dans les systèmes d'information) ; certains souhaitent au moins un marquage unique pour identifier les données publiques. Or, ce qui est possible pour des documents est impossible, voire absurde, pour des données. Ce qui est possible, et souhaitable, c'est de tracer tous les transferts de données afin de pouvoir établir, si nécessaire, que la série de données (ou une donnée, ou un document) a été chargée à telle date, avec un contenu que la plate-forme de téléchargement a enregistré, par tel réutilisateur.

### 2.5.5 Evolution de la chaîne de l'information

La chaîne de l'information évolue : capacités de saisie décentralisée, contributions des citoyens, d'associations, d'entreprises. Elle suppose et permet de nouvelles formes de contrôle, dont celui qu'exercent les citoyens eux-mêmes.

#### ■ *Qualification*

Comme pour d'autres problèmes, la qualification des informations reçues et qui seront ensuite mises à disposition ne peut pas être considérée comme liée à la seule réutilisation : si ce sont des données qui intéressent la collectivité dans le cadre du service public, c'est bien le moins que de pouvoir les qualifier. La réutilisation vient seulement rappeler et renforcer cette exigence.

En tout état de cause, il ne s'agit pas de vouloir tout vérifier des déclarations des uns ou des autres, et surtout pas de porter de ce fait des jugements sur des informations qualitatives. Il suffira souvent de confronter des données vérifiables et sans ambiguïté (adresse, classement d'un hôtel, tarifs affichés, horaires théoriques de transport etc.) aux sources existant déjà – documents réglementaires, informations contractuelles.

Aller plus loin dépend du domaine : dans chacun d'eux, les institutions d'aide ou de mutualisation qui constituent les gisements de données publiques sont celles qui vont les mettre à disposition, ou avec qui les collectivités vont préciser les conditions de réutilisation. C'est en tout cas ces institutions qui doivent procéder aux validations. Ainsi, les comités et offices de tourisme participent à la validation des ressources et des offres. Ex : L'Agence de Développement Touristique du Département de l'Hérault " ([www.adt-herault.fr](http://www.adt-herault.fr)), "proche du terrain, participe et contribue à la qualification de l'offre labellisation, certification, démarche qualité dont le Plan Qualité Tourisme...)"

#### ■ *Intervention citoyenne, associations etc.*

Certains attendent beaucoup du crowdsourcing, qui fonctionne effectivement étonnamment bien dans de nombreux cas : c'est le cas d'Openstreetmap sur le modèle Wikipedia.

Néanmoins plusieurs problèmes se posent : retombée d'enthousiasme s'il apparaît en outre que la collaboration bénévole permet le développement de services payants, erreurs, informations mal intentionnées ou excessives (pour signaler par exemple des problèmes de voirie) etc.

Pourtant, c'est bien souvent les acteurs près du terrain qui peuvent fournir l'information correcte et à jour à la collectivité.

Plusieurs directions peuvent être indiquées.

D'abord, ce sont les personnes impliquées, autour d'un thème et d'un territoire précis, qui sont les plus à même, voire les seules à même, d'apporter à la fois une information fine et validée.

Aussi, les apports des professionnels vont ils être essentiels pour des services les concernant : c'est ce que pratique l'APEM (cf. géographie ci-dessus) pour produire et valider une information sur, par exemple, les activités pastorales. Pour atteindre à une information de qualité, le choix est ici de constituer un système maillé. On pourrait dire ici qu'en matière de participation à une information publique, on peut appliquer «*De chacun selon ses capacités, à chacun selon ses besoins*».

La même implication est vraie pour obtenir l'information sur la voirie ou sur les services publics auprès des habitants d'un quartier.

Les petites communautés, où chacun verra rapidement le résultat de ses apports sont aussi capables de développements remarquables. C'est ainsi que les habitants de Plouarzel (Finistère - 3000 habitants) ont réalisé eux-mêmes, au cours de carto-parties, les cartes de leur territoire sur OpenStreetMap.

D'autres approches visent à une validation des données, pour éviter les saisies fantaisistes ou inappropriées.

Constitution de communautés de personnes identifiées.

Contrôle et correction par des acteurs légitimes qui filtrent (agents des communautés de communes, association, service) – c'est ce que les responsables de la Communauté urbaine de Bordeaux appellent le crowdpatching.

Insertion dans des processus de signalement et fourniture de l'information par les agents, ce que les responsables de l'ouverture des données au Grand Toulouse appellent «agent sourcing». Les agents peuvent à leur tour être intermédiaires pour permettre les signalements de problèmes par la population.

**Grand Toulouse : agent sourcing.** Dotés d'un outil mobile, les agents qui patrouillent sur l'espace public signalent les problèmes – dépôts sauvages, stationnement gênant, occupation de l'espace public, détérioration de voirie, mobilier urbain, animaux. Ils sont formés sur ce qu'il est pertinent de déclarer et de ne pas déclarer. L'extension du système vers les déclarations des citoyens est possible. Actuellement, ils appellent "l'Office de la tranquillité" et un superviseur peut envoyer quelqu'un pour intervenir. La mise en relation des deux systèmes permettra de confronter les déclarations à l'observation de l'agent, qui sera automatiquement prévenu.

## 2.6 Répertoires

La constitution de répertoires des données publiques est une étape essentielle de l'ouverture des données et de la réutilisation mais aussi plus largement de la transparence de l'action publique. La constitution et le maintien de tels répertoires, qui peuvent être mutualisés – seule solution pour la plupart des acteurs publics - est une tâche techniquement complexe, si l'on veut obtenir un outil d'information efficace. (*Les aspects techniques seront évoqués dans le chapitre 5*)?

Il est indispensable de rappeler que le répertoire des données et documents



est une obligation légale, même si elle n'est guère remplie actuellement.

Par ailleurs, il faut rappeler l'importance pour l'actuel mouvement d'ouverture des données des sites répertoire que sont data.gov aux Etats-Unis et data.gov.uk au Royaume Uni.

### 2.6.1 Obligation légale

Le répertoire des données accessibles est une obligation légale depuis la "Loi CADA" (Loi n°78-753 du 17 juillet 1978 – article 17): *«Les administrations qui produisent ou détiennent des informations publiques tiennent à la disposition des usagers un répertoire des principaux documents dans lesquels ces informations figurent.»*

Il s'agit bien ici de l'accès des personnes privées aux documents, quel qu'en soit le support (essentiellement papier à cette époque).

L'Ordonnance du 6 juin 2005 rappelait l'obligation, qui a été précisée par le Décret n°2005-1755 du 30 décembre 2005: *«Le répertoire prévu à l'article 17 de la loi du 17 juillet 1978 susvisée précise, pour chacun des documents recensés, son titre exact, son objet, la date de sa création, les conditions de sa réutilisation et, le cas échéant, la date et l'objet de ses mises à jour. Lorsque l'autorité administrative dispose d'un site internet, elle rend le répertoire accessible en ligne».*

### 2.6.2 Un répertoire

#### ■ Structure

Un répertoire présente un ensemble de jeux de données, qui peuvent être classés selon plusieurs critères. Ceux-ci sont les métadonnées qui décrivent le jeu de données et permettent aux personnes qui consultent de trouver la

référence de données qui peuvent les intéresser. Ces métadonnées seront exploitées par des moteurs de recherche spécialisés pour "moissonner" les informations de plusieurs répertoires. C'est pourquoi, d'une part, il est impératif que les métadonnées d'un même répertoire soient cohérentes et c'est extrêmement souhaitable pour un ensemble de répertoires. La normalisation existe pour beaucoup de métadonnées et elle progressera (son absence ne doit pas être un frein).

Même si le site qui gère le répertoire peut simultanément héberger les données elles-mêmes, ce n'est pas lié. Etant donné que les répertoires seront le plus souvent mutualisés, ils fonctionneront comme des portails qui adresseront les visiteurs aux différents sites où se trouvent les données.

#### ■ Contenu : la description des jeux de données

En pratique, les constructions actuelles de répertoire ont un champ plus restreint que celui décrit par les textes car elles visent les données sur support numérique et réutilisables.

Néanmoins, et cela a été en particulier un souci du GFII depuis longtemps, il est très important de faire apparaître les gisements dormants, c'est-à-dire les informations qui peuvent être sur support papier ou numérique mais dont personne jusqu'ici ne connaît l'existence ou ne mesure l'importance.

Ainsi, le répertoire peut et même doit référer des informations non numériques ou des ensembles de documents contenant des données nominatives ou grevés de droits. Ces données ne sont pas réutilisables sans des investissements souvent notables (numérisation, anonymisation) mais ils doivent être signalés s'ils sont importants, des partenaires pouvant proposer des coopérations.

La loi impose, quelles que soient les données (cf. ci-dessus) les informations minimales suivantes :

- titre
- objet
- date de création
- conditions de réutilisation
- date et objet des mises à jour

Charles Népote, responsable du dossier à la FING, ajoute<sup>16</sup>:

- le sujet ou la catégorie, permettant ultérieurement d'effectuer des recherches plus efficaces
- la fréquence d'actualisation
- la liste des métadonnées ou le format des métadonnées
- le format physique
- le lieu de diffusion: une URL ou bien le moyen d'accéder à ces données
- les droits d'usage

Il propose en outre d'autres critères qui, « s'ils ne sont pas indispensables, facilitent la compréhension et l'usage des jeux de données » :

- la couverture spatiale ou temporelle
- la langue (qui sera utile pour le repérage par des étrangers et des catalogues internationaux)
- des mots-clés
- un échantillon (les réutilisateurs réunis dans le GFII et ayant déjà une expérience ancienne insistent sur l'importance de disposer de ce fichier test)

<sup>16</sup> ([http://www.reseaufing.org/pg/blog/openid\\_82/read/62628/comment-cataloguer-des-donnees-publiques-](http://www.reseaufing.org/pg/blog/openid_82/read/62628/comment-cataloguer-des-donnees-publiques-))

### 2.6.3 Objectif et mutualisation du répertoire

La construction des répertoires est à terme une tâche complexe puisque le développement des usages et l'abondance croissante des informations mises en ligne exigeront de plus en plus des outils sophistiqués pour trouver et trier les informations.

Cependant, dans une première étape, l'essentiel n'est pas dans la technique mais dans le mouvement consistant à identifier les données publiques susceptibles d'intéresser des utilisateurs et réutilisateurs, tâche importante pour un acteur et encore plus pour un groupe d'acteurs, et notamment ceux d'un territoire régional.

Comme cela a été souligné, la définition de métadonnées communes pour une communauté d'acteurs est déjà une étape importante et enrichissante. C'est un premier travail de mutualisation.

Au niveau d'un territoire régional, qui paraît privilégié pour cette mutualisation, l'objectif d'un répertoire est d'orienter vers les diverses sources sur le territoire disposant d'information sur un thème donné, ce qui va stimuler l'effort pour rapprocher les métadonnées.

Les agglomérations qui ont commencé d'ouvrir leurs données ont établi des catalogues sur leurs sites (*cf. annexe Sites*). Ces derniers renvoient aux données de la Communauté urbaine ou de la ville mais aussi à celles de partenaires.

Ainsi, pour Montpellier, "le catalogue des données rassemble l'ensemble des descriptions des données disponibles et publiées par les différents partenaires. Plusieurs partenaires participent en effet à son enrichissement. Il s'agit des institutions publiques locales mais aussi régionales ou nationales, d'acteurs non gouvernementaux comme des associations ou des entreprises privées dès lors que leur disponibilité peut engendrer un service avéré au public. Il est mis

à jour régulièrement au fur et à mesure de la production et la mise à disposition des données.”

#### 2.6.4 Répertoires nationaux

Trois développements existent au niveau national, très différents par leur objectif, leur contenu, leurs porteurs.

- Etalab, répertoire national officiel en construction en 2011,
- Nosdonnees.fr, site collaboratif lancé par Regards citoyens
- Data Publica, le plus avancé en 2011 et qui vise à une place de marchés des données, au delà à terme des données publiques.

##### ■ *Etalab : data.gouv.fr*

La mission Etalab est directement rattachée au Premier ministre et gérée par le Secrétariat général du gouvernement. Elle a comme objectif premier la création du portail national répertoire des données publiques pour les administrations d'Etat. Cependant, elle pourra accueillir les informations des collectivités si celles-ci le souhaitent.

Les objectifs et enjeux du portail, qui devrait être en ligne fin 2011 sont :

- Identifier et recenser
- Faciliter la réutilisation
- Valoriser les initiatives et les nouveaux usages permis par la réutilisation

Le portail offrira un moteur de recherche de données brutes et structurées et assurera un travail de qualification des données en plaçant un premier jeu de

□ métadonnées sur les jeux de données (producteur, date donnée, couverture géographique, période).

Au total, il s'agit de créer un «data store sémantique» des données de l'Etat. En outre, il sera ouvert aux interconnexions et en particulier interopérable avec les autres initiatives, notamment avec le futur portail européen.

##### ■ *Nosdonnees.fr*

Ce répertoire a été lancé par RegardsCitoyens, associations de personnes privées qui ont avant tout comme projet la transparence de la vie politique (l'association a notamment créé le site Nosdeputes.fr). Nosdonnees.fr propose un ensemble d'outils permettant notamment un travail collaboratif entre internautes pour repérer des jeux de données et les signaler dans un Wiki. Les outils de gestion du répertoire sont ceux qu'utilise data.gov.uk – ce sont les outils de l'Open Knowledge Foundation.

##### ■ *Data Publica*

<http://www.data-publica.com>

(voir «Economie» et «Technique»)

Le projet de cette société créée en 2010 est de créer une place de marché de données – et pas seulement, à terme, de données publiques, même si c'est par là que Data Publica commence. Les promoteurs pensent que les autres acteurs, en particulier les entreprises, vont aussi pouvoir et vouloir commercialiser et négocier des jeux de données. La ressource augmente rapidement – c'est ce que les américains appellent le «big data».

Fondée par François Bancelon, un ancien de l'INRIA, de l'Internet et des start ups ayant réussi, la société regroupe trois PMI et a été rejointe par l'INRIA,

avec lequel elle développe les outils avancés d'analyse qui permettront de trouver, dans une masse énorme, tous les jeux de données pouvant intéresser une recherche particulière. Une grande société de services, ATOS Worldline, a rejoint les fondateurs et des soutiens ont été apportés par les programmes publics.

Data Publica a rapidement inscrit dans son répertoire des jeux de données qui existaient dans les différents secteurs (par ex ceux de l'INSEE, ceux mis en accès libre par les villes etc.) en étendant son action déjà au delà de l'Etat et des collectivités puisque des données de l'assurance maladie sont présentes, par exemple. En fin d'année 2011, le site devrait offrir une plateforme d'achat.

Il est important de noter que Data Publica n'est pas un catalogue de données publiques ouvertes et gratuites mais un répertoire de jeux de données accessibles, quelle que soit leur origine et quel que soit leur statut (gratuits ou payants, sous tous types de licences et conditions). Par ailleurs, Data Publica collecte et héberge les données et offre des services de traitement et d'enrichissement. En juillet 2011, plus de 2600 jeux de données étaient répertoriés et documentés.

### 2.6.5 Un outil clé

Pour satisfaire à la Loi, il est important que les administrations publiques informent sur les données et informations qu'elles détiennent. Satisfaire à cette obligation peut consister en une liste de données placées quelque part sur le site Web de l'organisme. C'est déjà enrichissant pour ce dernier, qui se livre à un examen et à une brève analyse de ses fonds d'information.

Cependant, l'ambition et l'impact des répertoires, catalogues, annuaires est bien plus grande, en particulier dans cette phase de transition vers une communication et une réutilisation des données.

Pour une collectivité, et singulièrement un Conseil régional, la constitution du répertoire est aussi l'occasion d'une concertation avec des partenaires. Elle fait partie d'une démarche d'analyse des potentiels internes et coopératifs, d'une sensibilisation et d'une communication. Elle est une démarche initiale importante.

## 3 Réutilisation : points de vue et débats juridiques et réglementaires

La réutilisation de données est encore un concept flou, objet, on l'a vu, de nombreux débats et correspondant souvent à des pratiques différentes. Même si elles peuvent être précisées, les notions d'accessibilité des données, d'ouverture, de publication, de réutilisation se mêlent dans les textes et dans leur application. Une raison en est qu'il s'agit d'un phénomène nouveau pour la plupart des institutions. Certes, on l'a dit, les grandes bases d'information font depuis longtemps l'objet de transferts, de contrats entre des institutions essentiellement publiques et des sociétés commerciales, mais il s'agit d'échanges entre des acteurs dont une activité ou toute l'activité est la diffusion d'information. Une relation entre l'INPI et une banque de brevets peut être parfois difficile mais ces institutions savent se parler, même s'il s'agit de s'opposer. En revanche, les entreprises comme les administrations considéraient jusqu'ici leurs données comme leurs biens privés: de fait même la réutilisation interne était rare. L'entrée rapide dans ce que les Américains appellent le "big data" vient de la croissance exponentielle de la production de données et de l'utilisation massive d'outils d'information et d'analyse dans un monde en rapide évolution.

Le présent chapitre revisite les différentes situations et les aspects juridiques de la réutilisation tels qu'ils sont interprétés ou mis en avant par les uns et les autres. Il pose donc surtout beaucoup de questions. En tout état de cause, les difficultés juridiques existent mais elles ne sont pas bloquantes dans la situation actuelle, en grande partie parce que les enjeux économiques sont encore limités sur les nouveaux marchés. Il demeure aujourd'hui beaucoup d'incertitudes, qui ne pourront être levées que progressivement et dans la pratique.

- *Pour les analyses juridiques, il est essentiel de se reporter aux textes mais il sera également très utile de se reporter au Guide juridique et pratique publié par Aquitaine Europe Communication en décembre 2010.*

### 3.1 Des bases aux questions

(Dans ce § comme dans les suivants, on se reportera à la présentation générale du Chapitre 1, aux textes législatifs et réglementaires joints et au Guide AEC)

#### 3.1.1 L'évolution des textes

La base des textes qui régissent la réutilisation des données publiques est la loi CADA de 1978. La transposition en 2005 (ordonnance et décret) de la Directive européenne de 2003 a précisé et complété ce socle pour faciliter et réguler la réutilisation de fichiers de données numérisés et l'accès en ligne – même si, fondamentalement, et à juste titre, les textes ne différencient pas les supports. Or – et c'est de là sans doute que viennent en partie certains problèmes, la loi CADA visait l'accès des citoyens, des associations ou des entreprises à des documents, éventuellement des données, dans une perspective démocratique. Elle faisait de l'accès aux informations détenues par la puissance publique un droit fondamental. En revanche, la Directive de 2003 correspondait à un objectif économique : le développement de nouvelles chaînes de valeur et d'entreprises des TIC grâce à l'accès à un vaste gisement de données. Enfin, le mouvement open data inclut une dimension idéologique et politique, dans un environnement social et technique nouveau.

La Loi française permet à toute personne privée de demander l'accès à des documents et données détenus par l'administration, avec des réserves (protection de la vie privée, protection de l'entreprise, sécurité publique). La réutilisation a un champ plus limité puisqu'un citoyen ou une entreprise a

accès aux informations le concernant alors que la réutilisation concerne des ensembles de données (*cf. infra les filtres*). Cependant, la réutilisation est un droit pour toute donnée ou document accessible et c'est à l'administration de justifier son refus.

L'article 10 de la loi de 1978 prévoit que les administrations sont tenues de mettre les informations publiques à la disposition de toute personne (physique ou morale). Il précise également que les informations publiques peuvent être réutilisées à d'autres fins que celles de la mission de service public pour les besoins de laquelle les documents ont été produits ou reçus. La CADA, autorité administrative indépendante, peut être saisie en cas de refus de l'administration.

La transposition de 2005 et la circulaire du Premier ministre du 29 juin 2006 ont précisé que la réutilisation est un droit. L'ensemble des textes la constitue clairement comme un droit opposable. Cette position étend la directive européenne, qui impose des règles d'égalité de traitement, de non concurrence etc. si une administration décide d'ouvrir ses données, mais qui ne prévoit pas qu'on puisse l'y contraindre.

La montée du mouvement open data a conduit à de nouveaux textes, à une évolution dans l'interprétation des textes existants et à de nouvelles ambiguïtés. Jusque là, l'objectif économique était partagé entre l'administration et des industriels de l'information. Les discussions portaient sur les conditions d'accès et le mode de calcul des redevances. L'Etat affichait la volonté de rentabiliser lui aussi son activité, ce qu'il faisait déjà avec plusieurs organismes – en particulier l'INPI avec les brevets et l'information société, l'IGN etc. – et il voulait pouvoir obtenir des recettes avec les autres administrations. L'APIE établissait donc ses licences type dans cette perspective. (*cf. décret 2009-151 du 10 février 2009 relatif à la rémunération de certains services rendus par l'Etat consistant en une valorisation de son patrimoine immatériel*)

Le mouvement open data a pris de front cette stratégie. Les nouvelles tendances ont conduit le gouvernement à créer la mission Etalab par le décret du 21 février 2011 et le portail DataGouv par la circulaire du 26 mai. En outre, le décret n° 2011-577 du 26 mai 2011 et la circulaire encadrent les éventuelles redevances.

*Il est important de noter que ces derniers textes ne s'appliquent qu'à l'Etat et aux Etablissements publics administratifs et ne concernent donc pas les collectivités territoriales et autres structures publiques.*

*Au total, les textes dessinent un paysage plus ouvert que celui résultant de la Directive de 2003 – droit opposable de réutilisation, non exclusion a priori des données culturelles. C'est dans l'application concrète que gisent les difficultés.*

### 3.1.2 La situation difficile et parfois ambiguë de l'Etat, les limites incertaines des textes

L'évolution décrite ci-dessus a créé des incohérences entre les travaux de l'APIE (sensibilisation des administrations, information, licences type) et les orientations désormais données par Etalab. Beaucoup de questions se posent, reflétant les incertitudes qui subsistent dans les textes eux-mêmes. Les dissensions qui se font jour sur le terrain (Géographie généalogie, RATP etc.) révèlent un flou et l'inexistence d'une jurisprudence sur les situations les plus litigieuses.

- D'abord, l'application aux collectivités ne se fait que par les lois et le décret d'application de la Directive, mais les nouveaux textes qui limitent les possibilités en matière de redevances, ne les concernent pas.
- Des questions se posent quant aux nombreux organismes coopératifs associant divers types d'acteurs publics. (*cf. infra*)
- Les EPIC et les organismes privés en charge d'un service public

sont exclus de l'ordonnance de 2005 (transposition).

- Sont aussi exclues du droit de réutilisation les données produites dans le cadre d'une mission de service public à caractère industriel et commercial, y compris si elles sont produites par des administrations pour la part de leur activité effectuée selon les règles du commerce (elles peuvent être réutilisées mais cela dépend de la décision de l'administration concernée).

- Les entreprises publiques ne sont pas concernées.

Ainsi, en particulier, les textes ne s'appliquent pas en l'état aux services en charge de diffusion de données, qui relèvent tous d'une exclusion (DILA, Météo, IGN etc.).

### 3.1.3 Les règles de fond

La réutilisation des données publiques se développe dans un espace instable tant du point de vue des formes d'organisation sociale que de celui des possibilités techniques. Il est normal que subsistent de nombreuses incertitudes et il est certain que les réglementations vont encore évoluer – comme d'ailleurs la Directive européenne.

Aussi faut-il vérifier la conformité de projets aux principes fondamentaux posés par les textes, qui ont trait au droit des citoyens à l'information sur l'activité des structures publiques et à l'application des règles du commerce au monde de l'information numérique:

- L'accessibilité d'un jeu de données entraîne la possibilité d'exercer le droit de réutilisation, en respectant les exceptions (vie privée, secret industriel et commercial, sécurité publique essentiellement – soit en excluant les jeux de données correspondants, soit en les traitant, anonymisant etc.). Le droit à la réutilisation est encadré par des contraintes limitées, qui se réduiront encore (*cf. infra*).

- Même si les textes ne l'expriment pas actuellement pour les collectivités, la tendance à la gratuité est forte et les redevances devront être limitées (mais restent de la décision de chaque collectivité ou acteur public, avec des contraintes pour les administrations d'Etat).

- La collectivité doit traiter également tous les réutilisateurs (l'évaluation de cette égalité de traitement peut poser problème du point de vue d'éventuelles redevances). A fortiori, toute exclusivité est exclue sauf les cas où, pour une durée limitée, elle est «indispensable à l'exercice d'une mission de service public» (en pratique, cela peut concerner un délégataire de service public qui doit avoir accès à certaines données).

- L'ouverture et même la gestion de l'information publique exigent de :

- constituer des répertoires

- nommer une Personne responsable de l'accès aux documents administratifs - PRADA (selon les conditions de l'ordonnance de 2005)

- préparer et publier au moins une licence type

- appliquer une licence en cas de mise en place d'une redevance

Au delà, pour les différentes données et actions, il est essentiel de rechercher des accords avec des partenaires, d'éviter autant que possible les conflits et de rester en veille. C'est la pratique qui permettra de clarifier les zones floues et de définir des règles précises, qui dépendront vraisemblablement des domaines concernés.

### 3.1.4 Des débats vrais plutôt que des jeux de mots

Les pratiques sociales et les décisions politiques déterminent l'évolution du droit et toute autorité judiciaire apprécie les faits sur le fond. Aussi est-il coûteux et stérile de s'engager dans des débats en jouant sur des

■ interprétations des textes.

■ Il est important de le rappeler, le manque d'information sur les dossiers ou l'espoir de retarder des échéances pouvant conduire à ce genre d'erreur.

■ Quelques interprétations absurdes sont rappelées ici. Elles ont souvent surgi, en particulier, dans le débat sur la généalogie. Il est à espérer que des arguments plus solides soient échangés à l'avenir.

#### ■ *Qu'est ce qu'un acte de réutilisation ?*

Dans plusieurs affaires ou sur certains sites, on apprend que la consultation en ligne serait une réutilisation. Comme le demande un internaute : pourquoi ne pas demander une licence pour entrer dans une salle de lecture ?

#### ■ *La loi définit-elle un espace interdit aux réutilisateurs ?*

Les textes posent que «*La réutilisation des données est définie comme leur utilisation par d'autres personnes que les acteurs publics et pour d'autres fins que celles répondant à des missions de service public.*» Est-ce à dire que les activités du réutilisateur ne doivent pas se situer dans le même domaine que celle de l'acteur public ? . Cela n'est vrai que si les activités en cause sont dans un domaine interdit aux autres acteurs, ce qui n'a rien à voir avec la question de la réutilisation. Ce que veulent dire les définitions de la loi, c'est que les demandes de réutilisations destinées à assurer un service public, en général émanant d'acteurs eux-mêmes administratifs ou délégataires, ne rentrent pas dans le cadre. Grave confusion logique : ce n'est pas une restriction, c'est le fait que certains acteurs échappent à la contrainte.

#### ■ *La prise en compte de l'objectif autorise-t-elle des différences dans les redevances ?*

□ Dans plusieurs textes et décisions, on indique qu'il est possible de tenir compte de la valeur économique de la prestation pour le bénéficiaire. Certes, mais c'est parce que la loi fixe un maximum : il est loisible à l'acteur public de définir un niveau de redevance très au dessous du maximum. Simplement, ce niveau doit être le même pour tous les réutilisateurs, quelles que soient leurs intentions.

### 3.1.5 Perspective internationale

Les problèmes juridiques semblent se poser aujourd'hui au niveau local et national, même si les acteurs et les textes se réfèrent à la directive européenne et à des licences internationales.

Cependant, le niveau international doit être pris en compte et c'est pourquoi un consensus s'est fait sur la nécessité de définir des licences adaptées certes mais compatibles avec celles qui sont utilisées internationalement (*cf. infra Licences*).

L'intervention d'acteurs internationaux est déjà importante, le plus important actuellement étant sans doute Google. D'autres, les grandes sociétés d'information, gèrent très bien la diversité des règles entre les différents pays (pour le droit, l'information sociétés etc.). Pour des PMI, le problème de la diversité des réglementations et des licences est ou sera une difficulté importante.

L'importance d'une valorisation et d'une circulation fluide des données est à l'origine de la directive européenne. C'est lorsque le marché va se développer que les problèmes pourront surgir, notamment en raison des différences sur les notions de service public, d'entreprise publique et donc de données publiques.



## 3.2 Quoi ? Quelles données ?

### Les filtres par nature des données

La réutilisation est possible sous la contrainte de respect des règles générales de confidentialité concernant les citoyens et les entreprises ainsi que des règles relatives à la sécurité publique.

La divulgation d'informations qui enfreindraient ces règles entraînerait la responsabilité de l'acteur public, qui doit vérifier que les données mises à disposition passent bien au travers des filtres correspondants.

#### 3.2.1 Données nominatives, données personnelles protection de la vie privée

La gestion et donc la réutilisation des documents et fichiers contenant des données nominatives est soumise aux lois et règles gouvernant la protection de la vie privée, telles qu'elles sont appliquées et contrôlées par la CNIL.

Des jeux de données sont accessibles à la consultation par une personne privée mais ne sont pas réutilisables si cela permet la diffusion de données nominatives et leur exploitation par des personnes non autorisées – une personne pouvant consulter un document la concernant mais celui-ci pouvant se trouver hors du champ de la réutilisation.

Il ne faut pas confondre données nominatives et données personnelles. Celles-ci ont un périmètre beaucoup plus large (elles incluent les données qui sont en outre nominatives). Certaines données personnelles nominatives sont d'accès libre pour tous, tant qu'elles ne donnent pas sur la personne d'autres informations que celles qui sont légalement accessibles, comme son adresse. Les fichiers de données personnelles non nominatives quant à elles, peuvent permettre d'identifier immédiatement une personne (par exemple en incluant un adresse) et elles peuvent permettre de l'identifier par recoupements de plusieurs fichiers dans beaucoup de cas.

Le principe est que la diffusion d'une donnée relative à une personne ne peut s'effectuer qu'avec le consentement de celle-ci. Cependant, pour des types de documents bien définis, des autorisations validées par la CNIL pourront être données – une personne pouvant faire valoir un droit de retrait («opt out»). C'est notamment le cas pour les fichiers déjà publiés en ligne par l'acteur public.

Le respect des règles est contrôlé dans les administrations et les principales collectivités par le Correspondant à la protection des données à caractère personnel - correspondants informatique et libertés.

Cependant, comme vient de le montrer l'affaire de la généalogie, des incertitudes subsistent et l'argument d'atteinte à la vie privée peut être utilisé de façon parfois discutable.

Dans beaucoup de cas, la réutilisation supposera l'anonymisation des informations, lorsque cela ne retire pas d'intérêt aux fichiers. Le problème est qu'il s'agit d'une opération souvent coûteuse, qui pourra être prise en charge par le réutilisateur – avec des conséquences encore incertaines du point de vue de l'ouverture (*cf. infra*). L'autre problème, souligné par les spécialistes de la question, est que l'identification indirecte d'une personne, par recoupements de fichiers, sera de plus en plus possible. Encore une fois, il s'agit d'un problème général de l'informatique et d'Internet, qui va bien au delà de la question des données publiques.

#### 3.2.2 Propriété intellectuelle et droit d'auteur

La réutilisation maintient ou peut maintenir certains droits de propriété intellectuelle sur les données. Ceux-ci se décomposent en droits patrimoniaux et droits d'auteur. Par ailleurs il faut prendre garde à distinguer dans les applications des textes les droits qui grevaient déjà les documents

ou données reçus par l'acteur public dans le cours de sa mission et les droits qui sont générés à son profit.

Globalement, la propriété intellectuelle n'est pas cédée par l'administration, seul l'est un droit d'usage, qui peut en pratique être encadré ou laisser totalement libre celui-ci en fonction des contrats et licences.

#### ■ *Droits d'acteurs extérieurs grevant les données*

- La propriété intellectuelle grevant certains documents ou données dès leur acquisition par un acteur public concerne, par exemple, l'obtention par celui-ci de documents et de travaux auprès d'intervenants extérieurs – par communication, échange, achat. Or, dans beaucoup de cas, l'expérience montre que les conditions de réutilisation n'ont pas été précisées, au moins pour les documents déjà anciens<sup>17</sup>. La libération de données et documents suppose que la situation soit éclaircie (les éditeurs savent que la recherche de droits est souvent longue et coûteuse). Ces droits peuvent être multiples sur un même document (par ex. s'il y a des photos illustrant un article). Les choses sont en général mieux définies dans le cas des séries de données (par ex. séries statistiques).

- Le droit d'auteur, élément de la propriété intellectuelle est inaliénable pour sa partie «droit moral». Il interdit toute modification de l'œuvre concernée sans autorisation de l'auteur, même si les droits patrimoniaux, autre volet de la propriété intellectuelle, ont été cédés.

- Si les droits externes ne peuvent être identifiés, l'administration doit en prévenir le réutilisateur, qui doit indiquer un moyen pour les propriétaires de les faire valoir (pratique des éditeurs papier).

<sup>17</sup> Des projets scientifiques nationaux sont handicapés parce que des études, archivées dans un ministère, ne peuvent être incluses dans des bases documentaires accessibles.

#### ■ *Droits d'auteur des élus et agents*

Un des problèmes à régler pour les données publiques est d'ailleurs l'existence potentielle d'un droit d'auteur personnel pour les élus et les fonctionnaires – de nombreux textes ont précisé qu'il fallait que le document concerné n'ait pas été produit par l'agent public dans son travail courant, sans création intellectuelle particulière. En outre, les textes publiés officiellement ne peuvent être grevés d'un droit d'auteur. En revanche, un discours, un commentaire de textes, un cours professoral, ou même une compilation originale de textes, constituent une œuvre de l'esprit.

#### ■ *Droits générés par l'administration*

Les droits générés par l'administration elle-même ne concernent pas les documents administratifs et réglementaires mais les documents internes et les outils intellectuels utiles à l'exercice d'un service public et réutilisables : études, notes d'information, catalogue, nomenclatures, ensembles de mots clé etc... ainsi que les bases de données, pour leur architecture ou l'ensemble de leur contenu, les séries de données elles-mêmes et les transformations qu'a pu effectuer l'acteur public sur des séries qu'il n'a pas créées - regroupements statistiques, combinaison de données, traitements divers.

Or, le droit d'auteur, quand il ne s'agit pas d'écrits, d'images, d'œuvres audiovisuelles, est difficile à identifier. C'est un débat déjà ancien : une base de données, c'est-à-dire le rassemblement d'informations, peut elle être une œuvre ? La notion de création intellectuelle, le caractère d'originalité sont les critères que doit apprécier le juge si le problème lui est soumis.

Remarque : on peut s'interroger sur la revendication de la RATP qui considère que le plan du métro relève d'un droit d'auteur.

### 3.2.3 Confidentialité, secret industriel et commercial

Qu'il s'agisse d'entreprises ou d'autres formes d'organisation, des règlements fixent les données qui sont publiques. C'est le cas des informations déposées obligatoirement par les entreprises au greffe du tribunal de commerce.

Comme d'ailleurs la publication, la réutilisation de données ne doit pas divulguer d'autres informations, que pourtant les acteurs publics détiennent – par les déclarations fiscales, les dossiers d'implantation dans les communes, les procédures relatives à l'emploi, à l'environnement, les programmes d'aide etc.

En ce qui concerne les statistiques, les règles doivent être celles qu'applique l'INSEE, définies par la Loi 51-751 du 7 juin 1951 sur le secret statistique. Elles posent d'autant plus de problèmes que les données sont fines<sup>18</sup>.

### 3.2.4 Sécurité nationale, sécurité publique, sécurité des personnes

Qu'il ne faille pas permettre la réutilisation d'informations pouvant compromettre ces différentes sécurités peut paraître trivial. Le problème réside dans la définition de ces données. En outre, il semble qu'il devrait s'agir d'accessibilité et non de réutilisation, s'agissant de données sensibles.

Il y a bien des limites à fixer ainsi que des interprétations. Par exemple, au niveau national, un débat ancien a opposé la Météorologie nationale (qui est un EPA) à des sociétés qui demandaient à réutiliser ses informations. Or, la Météo souhaite garder un contrôle et commercialiser directement ses informations. Un premier argument était que ces dernières devaient être sous monopole parce qu'elles mettaient en jeu la sécurité des aéronefs (*le débat s'est déplacé – cf. infra «pour faire quoi»*).

<sup>18</sup> Pour l'INSEE, aucune case d'un tableau ne doit contenir de données concernant moins de trois entreprises; aucune case ne doit contenir de données où une entreprise représenterait plus de 85% du total.

□ Si les informations sur la Défense nationale semblent aisées à identifier, elles peuvent poser des problèmes, par exemple pour les données géographiques dans les zones de camp militaire.

Au niveau régional ou local, certaines informations doivent être maintenues confidentielles: par exemple, pour des activités de police, la localisation de certains contrôles (la carte des radars est un exemple parlant mais également celle de caméras de surveillance – la FING cite les emplacements de panneaux indiquant les points d'arrêt de convois de fonds).

Certaines contraintes peuvent être demandées par des professions ou activités particulières. Ainsi, les positions exactes des fouilles archéologiques ne sont pas indiquées, pour éviter les pillages et même simplement la dégradation des chantiers.

On a vu le succès aux Etats-Unis des bases sur la criminalité (*cf. Ch 2*): leur divulgation est-elle de nature à perturber l'ordre public ?

Dans la plupart des cas, le vrai problème est l'accessibilité de ces informations, qui relève de la décision politique.

## 3.3 La réutilisation : pour quoi faire ?

Beaucoup de demandes, même des interprétations des textes, tendent à placer des conditions strictes sur l'objectif du réutilisateur. De plus en plus, il paraît non seulement difficile mais dangereux de dépasser le cadre fixé par la loi et précisé par la pratique et la jurisprudence.

### 3.3.1 Une interprétation stricte des textes, sans extensions abusives

□ Ici encore, il ne faut pas confondre accessibilité et réutilisation. Si la diffusion d'une donnée pose problème, ce sera vrai aussi bien pour une publication qu'une réutilisation, voire un accès ponctuel.

Par ailleurs, une utilisation illicite d'informations, quelle que soit leur origine, doit être réprimée.

En revanche, il paraît difficile, et même dangereux, de demander à un agent public d'apprécier la nature du demandeur (secte, parti, association ou société peu fiable etc.) et, plus risqué encore, ses intentions.

Les textes indiquent que le réutilisateur doit faire état de ses objectifs mais il semble prudent de s'en tenir à cette déclaration, qui pourra être produite si des plaintes le nécessitaient. Evidemment, une utilisation déclarée illicite interdit de permettre la réutilisation. En revanche, aucun texte ne prévoit de la refuser au titre de la nature du demandeur. Il importe de s'en tenir au droit d'accès de tous.

Au cours du débat parlementaire sur la Loi LOPPSI 2, un amendement a été proposé qui introduisait un contrôle de moralité pour les réutilisateurs de données publiques sous licence. [...] Il a suscité une grande émotion, bien au delà des promoteurs de l'open data, et a été repoussé.

### 3.3.2 La question de la valeur ajoutée

Parmi les contraintes encore discutées, il y a celle qui impose aux réutilisateurs une valeur ajoutée.

Le consensus commence de se faire sur la difficulté et l'inutilité de cette approche. D'une part, il peut y avoir une bien meilleure présentation ergonomique, par exemple, et il sera difficile de démêler la valeur ajoutée (ou non ajoutée) si les données sont mêlées à d'autres dans un même service. D'autre part, on ne voit guère de raison de reprendre des données en l'état, sauf à s'attribuer le mérite de leur production – mais alors le réutilisateur ne respecte pas l'obligation de donner sa source, de tracer l'origine de la donnée publique. En tout état de cause, l'intérêt de la simple rediffusion est problématique.

### 3.3.3 Un objet différent de celui du service public initial

Les textes indiquent (Loi de 78 – article 10) : « *Les informations figurant dans des documents produits ou reçus par les administrations mentionnées à l'article 1<sup>er</sup>, quel que soit le support, peuvent être utilisées par toute personne qui le souhaite à d'autres fins que celles de la mission de service public pour les besoins de laquelle les documents ont été produits ou reçus. Les limites et conditions de cette réutilisation sont régies par le présent chapitre, même si ces informations ont été obtenues dans le cadre de l'exercice du droit d'accès aux documents administratifs régi par le chapitre I<sup>er</sup>.* »

La première partie de ce paragraphe signifie simplement que le réutilisateur n'est pas impliqué dans la mission de service public (d'autres textes régissent ce cas), ce que confirme la deuxième partie puisque les exceptions sont listées ensuite dans le même article – ce sont celles qui ont déjà été indiquées ici. Dire que les données peuvent être utilisées à d'autres fins que celles de la mission de service public ne signifie pas qu'il est interdit de les utiliser à cette même fin aussi. Autrement dit, si l'INSEE a mission d'informer sur l'économie, rien n'interdit au réutilisateur de viser le même objet – en fait c'est bien ce que font la plupart d'entre eux.

Cependant, certains essaient une autre interprétation pour des motifs économiques, sans grande raison qui plus est puisque les textes ont précisé après 1978 que les activités commerciales, et surtout existantes, donnaient en fait une grande autonomie aux services concernés pour décider de la réutilisation.

#### Site de Météo-France

Le site fournit une présentation très claire des données publiques en accès libre, avec le détail des fichiers et des conditions tarifaires. Il précise :

[https://public.meteofrance.com/public/accueil/presentation?document\\_id=21161&portlet\\_id=49077](https://public.meteofrance.com/public/accueil/presentation?document_id=21161&portlet_id=49077)

La réutilisation des informations publiques de Météo-France est soumise à la loi n° 78-753 du 17 juillet 1978 et au décret n° 2005-1755 du 30 décembre 2005. En application de la loi, la réutilisation n'est autorisée **que pour des fins différentes de la mission de service public pour laquelle les informations ont été produites ou reçues par Météo-France**. La redistribution en l'état des informations est une des missions de service public pour lesquelles les Informations ont été produites ou reçues par Météo-France et n'est donc pas autorisée.

### 3.3.4 Réutilisation commerciale – non commerciale

Une caractéristique des réutilisations fait l'objet de nombreuses réflexions et discussions : le caractère commercial ou non commercial, qui peut relever à la fois de l'intention déclarée et de la nature du réutilisateur (même si, dans la pratique des licences, c'est l'activité qui est seule identifiée). La distinction est très fréquente et les utilisations commerciales sont souvent exclues (y compris dans les licences Creative Commons "NC" cf. *infra Licences*).

#### ■ Une distinction parfois difficile à opérer

Cette distinction n'est pas forcément simple à mettre en œuvre, et il serait dangereux de s'appuyer d'abord sur la nature du réutilisateur : des associations peuvent avoir une activité lucrative par divers détours, des sociétés peuvent utiliser des jeux de données pour se valoriser mais en les publiant sur des sites d'information qu'ils créent ou sponsorisent. Un site qui est d'accès gratuit, ou un service, peuvent être entièrement financés par la publicité, qui n'existerait pas sans les données.

En pratique, il est cependant possible d'apprécier les diverses situations (mais encore faudrait-il dans beaucoup de cas justifier d'un refus ou d'une acceptation contestée).

#### ■ L'évolution de la tendance : distinguer selon les engagements

Alors que la tendance était jusqu'à récemment de distinguer strictement les applications commerciales, elle tend à s'inverser sous la pression du mouvement open data, ou tout au moins de sa majorité, qui considère que l'ouverture doit être sans conditions (cf. *Licences*).

Ce point posera certainement longtemps problème.

La différenciation est conforme à l'esprit de la loi si elle a pour conséquence de permettre ou non la réutilisation : il est possible de la réserver aux utilisations non commerciales (en pratique souvent en identifiant la nature du demandeur) – c'est un choix politique. En revanche, il sera de plus en plus difficile de justifier par là une différence de traitement entre les réutilisateurs, notamment pour le calcul des redevances.

Ce qui est possible, c'est de proposer des licences différentes aux réutilisateurs en fonction de leur statut et de leur intention déclarée, à condition que ces licences entraînent des obligations différentes. C'est ce que fait par exemple la Ville de Paris qui propose une licence SA (share alike) et traite différemment ceux qui ne veulent pas accepter cette clause. C'est aussi le cas pour les «licences clic» (cf. *infra*).

La différenciation se justifie parce qu'elle engage des responsabilités différentes. En revanche, le droit d'accessibilité rend impossible le refus de réutilisation commerciale (c'est bien le moins puisqu'un des objectifs actuels est le développement d'une activité économique).

Comme cela a été rappelé, la différenciation commercial – non commercial est particulièrement vive en France.

Certaines interventions dans le débat «Généalogie» ont un ton révélateur de ce point de vue : "Afin de nous "protéger" mutuellement, notamment contre une réutilisation commerciale des images provenant des AD\* il est IMPÉRATIF de citer la source en marquant cet ou ces actes que

vous allez mettre dans cette base.” (L’intervenant propose un marquage technique). \*AD : archives départementales

#### ■ Distinction par les usages : internes ou diffusion

Une différenciation est souvent introduite entre l’utilisation dans le cadre d’une activité commerciale mais en interne de l’entreprise réutilisatrice et l’utilisation commerciale où les données elles-mêmes, une fois éventuellement enrichies, sont commercialisées. Dans le premier cas, les données concourent à la réalisation d’un produit ou d’un service. Dans le second, elles en sont une composante : l’activité du réutilisateur est, au moins en partie, une activité de diffuseur.

Cette différenciation conduit en général à des responsabilités et à des redevances différentes. En effet, si une entreprise utilise les données en interne, les problèmes posés par la dénaturation des données ne se posent plus, ni le fait de devoir citer la source etc. Les données sont alors un moyen pour permettre à une activité économique de fonctionner. Dans certains cas, elles lui sont indispensables (par ex. des données géographiques).

### 3.4 Qui ? Quelles entités publiques sont ou peuvent être impliquées par les réglementations relatives à la réutilisation ?

La sphère publique est vaste et multiple. En outre, des entités nouvelles se créent constamment, qui associent entre eux des acteurs publics de statuts divers ou associent acteurs publics et privés.

Certes, dans un premier temps, la question de la réutilisation des données publiques se pose pour les cas les plus clairs et simples, une administration centrale, une Région, un Département, une Communauté urbaine mais, comme on l’a vu pour les données géographiques, les problèmes plus complexes vont commencer de se poser.

□ Pour un Conseil régional, qui a vocation à embrasser les différentes situations sur le territoire et à proposer des actions transversales, c’est à l’ensemble des acteurs qu’il conviendra de s’adresser.

Comme cela a été indiqué au Chapitre 1, c’est bien de la circulation des données publiques qu’il s’agit, mais il faut aussi considérer les acteurs proches dont les données sont en synergie et peuvent accroître la valeur ajoutée de la réutilisation.

Le point de départ est la Loi CADA : les documents administratifs sont ceux qui sont “produits ou reçus, dans le cadre de leur mission de service public, par l’Etat, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d’une telle mission.»

On l’a vu, les différents textes ont ensuite des périmètres distincts, incluant ou non les collectivités ainsi que les personnes de droit privé.

#### 3.4.1 Les autorités publiques

##### ■ La force du droit de réutilisation opposable

La tendance à l’ouverture est forte et correspond à une pression sociale ainsi qu’à une évolution des cultures. Les administrations et collectivités auront de plus en plus de difficultés à s’opposer à ce droit, dès lors qu’il est établi que les données satisfont sur le fond aux contraintes exposées ci-dessus. Le jugement du tribunal de Clermont-Ferrand contraignant le Département du Cantal à donner accès à ses données illustre cette évolution (cf. Ch. 2 Généalogie).

##### ■ Les activités commerciales des administrations

Les administrations d’Etat, les collectivités et les EPA ne sont pas concernés lorsqu’ils exercent une activité commerciale (cf. ci-dessus). Cette exclusion

pose les mêmes problèmes que ceux qui seront rencontrés avec les EPIC par exemple. Une partie des informations est effectivement protégeable par le secret commercial – situation malgré tout rare - mais peut-être pas l'ensemble.

#### ■ *L'intervention de contractants privés*

En fait, le principal problème va se poser avec la délégation de fonctions à l'extérieur – délégation de service public mais aussi simplement externalisation de fonctions. Si les problèmes posés par la DSP ont commencé d'être évoqués, les autres formes d'externalisation devront être considérées. En effet, la sous-traitance, par exemple, introduit un acteur privé dans le fonctionnement d'un service et cela peut, entre autres, obscurcir les analyses de gestion, les dépenses du sous-traitant ne pouvant être communiquées. Comme cela a été souligné, il est impératif de préciser clairement dans les contrats les droits relatifs aux informations sur le service. De même, sans qu'il s'agisse d'un service public, la décision de confier à une association des fonctions qui pouvaient être assurées en interne peut faire varier le périmètre de la donnée publique.

Les partisans de l'open government devraient donc se méfier d'un effet pervers classique, la fuite d'activités vers des acteurs moins exposés à la transparence.

L'acteur public se doit de prendre en compte l'ensemble des entités qui sont sous sa dépendance, soit pour une fonction donnée, soit globalement parce qu'il assure la majeure partie de leur financement. C'est d'ailleurs ce que font déjà des collectivités qui associent à leur démarche les structures dont elles sont le principal financeur ou au moins un partenaire incontournable.

#### ■ *Les établissements publics relevant des compétences des collectivités*

□ Selon leurs compétences et leur niveau, les collectivités sont impliquées en particulier dans la gestion des établissements de l'Education nationale et dans celle des hôpitaux et des autres structures sanitaires et sociales.

Au niveau régional mais aussi pour les autres niveaux de collectivité, il sera donc souvent nécessaire de partager des données avec le Rectorat ou avec l'ARS (relations très différentes compte tenu des compétences et compte tenu des statuts différents de ces deux acteurs). (cf. Ch. 2)

### 3.4.2 Activités industrielles et commerciales «publiques»

Trois types de structures publiques peuvent assurer des activités commerciales, et détenir – les règles et les opinions varieront – des données publiques : EPIC, GIP, Entreprises publiques. Dans le cas du GIP, les acteurs publics peuvent être associés à des acteurs privés minoritaires. La situation est la même pour les entreprises publiques, à ceci près que les moyens de contrôle sont pris en compte, ce qui peut autoriser une participation minoritaire en capital mais majoritaire en voix.

Ces trois types d'intervenants sont exclus du droit à la réutilisation et sont considérés comme des entreprises privées de ce point de vue. Une raison de fond est l'association aux partenaires privés et le fait que ces structures interviennent sur le marché.

Pourtant, des questions se posent, et de façon plus vive si ces entités assurent une mission de service public. Ici encore, il s'agit de choix politiques sur la définition du service public, sur l'introduction de la concurrence dans certaines activités, sur la transparence pour les citoyens. Même si le cœur du problème n'était pas là, l'affaire de la RATP a réveillé ce débat, beaucoup de personnes considérant que le plan du métro ne pouvait être une donnée non réutilisable.

### 3.4.3 DSP et missions de service public

Le débat est donc lancé sur les délégations de service public, et c'est souvent dans le domaine des transports que le problème se pose actuellement, en raison de l'importance des applications de mobilité.

La situation n'est pas claire. La directive n'aborde pas expressément le problème – ce d'autant qu'il serait difficile de produire une définition de la délégation de service public au niveau de l'Europe. La loi CADA incluait tout acteur public ou privé assurant une mission de service public. Cependant, à partir de la circulaire de 2006, les textes publics récents qui précisent les conditions de réutilisation ne concernent de toutes façons pas les entités ayant une activité industrielle et commerciale. En outre, comme cela a été rappelé précédemment, la mission ne comprend pas a priori un volet relatif à l'information.

C'est donc à la signature de tout contrat de délégation qu'il faut préciser le statut des données associées à la mission de service public – créées pour le permettre ou résultant de son fonctionnement.

En pratique, pour le transport, par exemple, il a souvent été considéré que les horaires théoriques, nécessaires au délégataire, étaient publics mais pas les horaires temps réel. On sait que les positions des délégataires varient, mais il s'agit de décisions commerciales par les entreprises concernées (*cf Keolis Ch. 2 – Transport*).

Certains organismes ont reçu leur mission de service public au niveau national. C'est le cas des chambres consulaires, et c'est pourquoi des demandes récurrentes ont été présentées pour obtenir les informations publiques, par exemple, de l'Annuaire des entreprises de France et, plus généralement, des fichiers détenus par les Chambres et créés dans le cadre de leurs missions de service public de connaissance des professionnels et entreprises. Ici aussi, la situation est peu claire – mais la pratique montre que

□ la coopération avec les Chambres fonctionne dans beaucoup de cas (*cf. aussi APEM dans le Ch. 2*)

### 3.5 Structures de mise en œuvre pour une gestion collective

Au cours du temps, les collectivités ont été conduites à créer des structures permettant la coopération entre elles, avec les acteurs publics du territoire, avec l'État, avec des partenaires privés etc. sur des domaines d'intérêt commun. Ces structures peuvent prendre de multiples formes. On retrouve ici des GIP, des Syndicats mixtes, des associations, etc.

Plusieurs de ces structures ont déjà été évoquées dans cette étude : Plates-formes d'information géographique, Autorités organisatrices des transports et leurs groupements, Plates-formes de services dématérialisés, Plates-formes d'information locale, Observatoires de l'environnement, Agences de développement, Agences d'urbanisme, Observatoires économiques régionaux...

Ces entités reçoivent des données de leurs membres et partenaires et les rediffusent. Elles sont des gestionnaires de données publiques. Elles sont régies par une convention, ou par des statuts, qui prévoient les conditions dans lesquelles elle reçoivent et retournent des données, et ce d'abord à leurs membres. Au delà, elles sont, actuellement, en général transparentes par rapport aux demandes de réutilisation, se retournant vers l'acteur concerné.

La diversité des partenaires pose des problèmes, surtout aux réutilisateurs. On sait que les débats peuvent être tendus lors de la constitution de structures de coopération, car il n'est pas possible d'associer des acteurs ayant opté pour choix économiques et stratégiques opposés dans le domaine des données publiques. Le cas emblématique des données IGN a été évoqué précédemment.



Quoi qu'il en soit, les structures de mutualisation seront des acteurs centraux pour beaucoup d'informations publiques. Elles sont créées pour faciliter leur gestion. Néanmoins, d'une part leur nature coopérative rend un peu plus difficile, à ce stade, les procédures de réutilisation. D'autre part elles sont parfois réticentes ou opposées à la mise à disposition des données qu'elles produisent ou contrôlent, d'autant que leur statut leur confère en général une exclusivité sur un domaine d'intervention.

En même temps, la visibilité des structures de coopération dans leur domaine, et leur action propre, les conduisent de plus en plus à prendre en main la question de la réutilisation, c'est-à-dire à coordonner leurs membres autour d'une position commune, que l'opérateur coopératif appliquera.

- C'est par exemple le cas du SIRTAQUI (Aquitaine) (cf. Ch 2. *Tourisme*) qui utilise une licence Creative Commons et n'applique pas de redevance mais signe une convention avec le réutilisateur.
- D'autres limitent la réutilisation aux membres – mais il peut y avoir des membres diffuseurs – c'est le cas du SITRA (Rhône-Alpes)

### 3.5.1 Entreprises

Le mouvement gagne les entreprises, dont celles qui travaillent en DSP (mais pas toutes : voir refus RATP, refus SNCF pour les approches multi-territoriales – interférence avec politique commerciale) (cf. Ch.2 *Infrastructures*).

## 3.6 Combien ? Redevances

La question des redevances reproduit de façon très amplifiée débats idéologiques et confusions, masquant parfois un problème trivial de ressources.

### 3.6.1 Les racines d'une question complexe

#### ■ *Un débat ancien*

Dès les premiers débats autour des données (INSEE vers 1985), la question de la redevance s'est posée. Les termes, usés, n'en bougeront plus pendant longtemps. Ils lient le plus souvent la redevance à la nature commerciale de la réutilisation (dont on a vu qu'elle était difficile à apprécier)

- Il n'est pas normal qu'un acteur privé profite d'un bien commun, payé par l'impôt, pour gagner de l'argent. Il faut que le réutilisateur privé retourne une partie de ses gains.
- Il n'est pas normal que ces données, dont la collecte a été financée par l'impôt, soient de nouveau payées lorsqu'un acteur veut les utiliser. Les entreprises aussi ont contribué par l'impôt. Elles paieront l'impôt sur leurs gains.
- Il faut distinguer les utilisateurs non commerciaux et les commerciaux et soit interdire la réutilisation commerciale soit exiger une redevance.
- Le développement d'une industrie de l'information passe par l'existence de sociétés commerciales – il faut leur fournir le matériau.

Ce débat existe pour tout choix de tarification d'une activité publique, voire d'un service public. Il s'agit de définir la part supportée par l'impôt et celle supportée par l'utilisateur. C'est un choix politique.

#### ■ *Une situation stabilisée, de nouvelles tensions*

Dans le domaine des grands acteurs de l'information, INSEE, DILA, Météo, IGN d'un côté, éditeurs de banques de données et autres industriels de l'information, la situation s'est peu à peu stabilisée et transformée en une relation commerciale presque banale.

Cependant, au fur et à mesure de l'extension de la diffusion à de nouveaux domaines et à de nouveaux acteurs, les mêmes questions sont débattues, et ce dans le contexte nouveau d'une situation économique difficile pour tous ainsi que d'une concurrence accrue et de plus en plus internationale dans le monde de l'information.

De son côté, le mouvement open data soutient assez largement la position de la gratuité systématique, et les tensions sont réapparues, en particulier avec l'IGN en ce qui concerne les plates-formes d'information géographique (cf. Ch. 2).

#### ■ La question du financement demeure posée

La situation économique et budgétaire limite les capacités immédiates d'investissement autant que les dépenses de fonctionnement des organismes publics. Certes, des gains sont à attendre de l'ouverture des données, mais ils sont souvent indirects et en tout cas ne se produisent qu'avec un décalage. Dès lors, il faut se demander, comme le fait Denis Berthault – LexisNexis, animateur du groupe du GFII – *“qui (du contribuable ou du réutilisateur) doit supporter les incontournables coûts de mise à disposition des données (conversion, anonymisation, chaîne de production) dans un format ouvert ? Et si l'Etat doit prendre ces coûts à sa charge, comment trouver le financement ?* La question vaut évidemment aussi pour les collectivités.

#### ■ La gratuité

Le mouvement open data privilégie la gratuité de la mise à disposition, arguant des coûts en réalité faibles de celle-ci, de gains importants à moyen et parfois court terme et de la nécessité de faire naître des applications innovantes portées par un tissu de PMI, de jeunes pousses, de jeunes passionnés, d'associations citoyennes. Tout montre d'ailleurs que ce tissu est

bien à même de faire apparaître des innovations multiples dont certaines correspondent à des besoins qu'il eût peut-être fallu à terme satisfaire, avec une dépense supérieure de l'acteur public.

Comme l'explique Daniel Kaplan, délégué général de la Fing, *“L'innovation ne viendra jamais des gros acteurs économiques traditionnellement réutilisateurs de données. Comme dans les autres secteurs, l'innovation vient des petits, qui seront rebutés par la moindre contrainte”*.

Ainsi, une redevance, même faible est une barrière à l'entrée. Or, il est impossible de consentir des conditions spéciales à tous ceux qui ne peuvent passer cette barrière, ce qui contredirait le principe d'égalité.

### 3.6.2 Principes et règles

#### ■ Ne pas faire de bénéfices : un principe réaffirmé ?

Le rapport Commerce électronique de 1998 le soulignait déjà : *«Pour l'ensemble des administrations et établissements, il serait souhaitable que le ministère de l'Economie veille à ce que les recettes de diffusion ne soient pas le critère de décision s'agissant de la diffusion des données.»*

L'ensemble des textes et des pratiques converge vers une absence de bénéfice, mais en prenant en compte des coûts d'investissement nécessités par la mise à disposition des données... et parfois plus.

La Directive de 2003 pose:

Lorsque des redevances sont prélevées, le total des recettes provenant de la fourniture et des autorisations de réutilisation de ces documents ne dépasse pas leur coût de collecte, de production, de reproduction et de diffusion, tout en permettant un retour sur investissement raisonnable. Les tarifs devraient être fixés en fonction des coûts pendant la période

comptable appropriée et calculés en concordance avec les principes comptables applicables aux organismes du secteur public concernés.

Modifiée par les décrets et ordonnances de la transposition puis l'ordonnance du 29 avril 2009, la loi CADA précise :

Pour l'établissement des redevances, l'administration qui a produit ou reçu les documents contenant des informations publiques susceptibles d'être réutilisées tient compte des coûts de mise à disposition des informations, notamment, le cas échéant, du coût d'un traitement permettant de les rendre anonymes.

L'administration peut aussi tenir compte des coûts de collecte et de production des informations et inclure dans l'assiette de la redevance une rémunération raisonnable de ses investissements comprenant, le cas échéant, une part au titre des droits de propriété intellectuelle. Dans ce cas, l'administration doit s'assurer que les redevances sont fixées de manière non discriminatoire et que leur produit total, évalué sur une période comptable appropriée en fonction de l'amortissement des investissements, ne dépasse pas le total formé, d'une part, des coûts de collecte, de production et de mise à disposition des informations et, d'autre part, le cas échéant, de la rémunération définie au présent alinéa.

Lorsque l'administration qui a produit ou reçu des documents contenant des informations publiques utilise ces informations dans le cadre d'activités commerciales, elle ne peut en facturer la réutilisation aux autres opérateurs à un coût supérieur à celui qu'elle s'impute, ni leur imposer des conditions moins favorables que celles qu'elle s'applique à elle-même.

- *Des textes favorables à la gratuité pour l'Etat et les EPA mais une application modeste*

□ Le décret n° 2011-577 du 26 mai 2011 soumet l'instauration d'une redevance à l'avis du Conseil d'orientation de l'édition publique et de l'information administrative (COEPIA).

Cependant, les redevances qui existaient avant le 1er juillet 2011 peuvent être maintenues à condition d'être inscrites dans une liste avant le 1er juillet 2012. A défaut, les titulaires de licences pourront utiliser les données gratuitement.

La tendance est non seulement à faire de la gratuité une formule possible pour l'administration d'Etat mais à en faire la solution par défaut. Même si l'application en est modeste, il reste que ces textes révèlent une tendance, que s'emploie au demeurant à stimuler et promouvoir la Mission Etalab.

Il y a eu là une rupture – en 2010, François Baroin, ministre du Budget, constatant que les redevances existantes rapportaient chaque année environ 50 millions d'euros à l'Etat, annonçait que les réutilisations pourraient être étendues à «*d'autres données comme les informations cadastrales mises en ligne par le ministère chargé du Budget, le prix de l'immobilier ou encore la liste des services publics de proximité ou leurs horaires d'ouverture*». Il envisageait un plan stratégique de valorisation du patrimoine numérique de l'Etat et déclarait : "*Cette politique de valorisation fondée sur une tarification équilibrée permettra d'entretenir un cercle vertueux. En effet, le produit des redevances permet d'investir dans l'amélioration de la qualité des données, dans leurs modes de diffusion et même dans la production de nouvelles données utiles à l'accomplissement des politiques publiques et à une meilleure information du citoyen.*"

*Même si les textes encadrant les réutilisations pour l'Etat ne s'appliquent pas aux collectivités, ils reflètent une tendance et ils seront certainement une référence en cas de dissensions et conflits. Il faut d'ailleurs rappeler que l'APIE comme Etalab, dont le domaine d'intervention est strictement limité à l'Etat, ont annoncé qu'ils étaient prêts à assister les collectivités qui le demandent,*

ce qui a été bien accueilli par des responsables de ces problèmes dans les collectivités.

Les applications sont modestes. En effet, les administrations peuvent ne pas modifier leurs pratiques actuelles. Cela choque certains mais c'est en fait inévitable : le passage à la gratuité, voire une réduction brutale de la redevance, handicaperait non seulement le service producteur mais aussi l'entreprise ou les entreprises qui, actuellement, ont atteint un modèle économique qui fonctionne et qui risqueraient de voir entrer des concurrents, petits mais aussi grands, attirés par la gratuité, sans avoir le temps d'adapter leur modèle et leur stratégie. Les textes prévoient d'ailleurs aussi des phases transitoires pour éteindre les exclusivités. Ce type de précaution devra au demeurant être pris pour tout programme.

Ceci étant, les administrations se sont empressées de s'inscrire et de publier des tarifs.

#### **Le service d'accès aux données comptables des collectivités locales proposé par la Direction générale des finances publiques.**

La DGFIP a annoncé un service payant. La redevance distingue une réutilisation non commerciale, une réutilisation interne (besoins internes du licencié) et plusieurs formes de réutilisation commerciale.

Les redevances varient, avec un barème assez complexe, de 500€ à 46200€ par an. Elles se composent d'une part fixe et d'une part variable de 20% du chiffre d'affaires hors taxes généré par la diffusion des informations.

Ce mécanisme confirme les craintes du GFII et des industriels de l'information face aux conséquences des difficultés budgétaires et au souci de rechercher des ressources. Il est sur le fond contraire à la Directive (cf. infra).

#### ■ Le débat sur la proportionnalité

De nombreux contrats ont pris en compte le chiffre d'affaires du réutilisateur. Des propositions ont tendu à ce que ce soit la base de la redevance. Des licences sont établies, qui suivent ce modèle, par des administrations – et non des moindres, comme le montre l'exemple de la DGFP.

Au delà de l'aspect comptable, il peut y avoir une justification à cette approche : comme on l'a vu, les redevances gênent ou bloquent les innovations de petites sociétés et l'égalité des conditions d'accès interdit de spécifier pour elles un traitement particulier. La solution pouvait être d'attacher un niveau de redevance à des classes de CA, ou de fixer la redevance en pourcentage. Ainsi peut se lever une barrière à l'entrée dont beaucoup disent qu'elle favorise évidemment les grands acteurs de l'édition – édition juridique, scientifique, bases financières.

Cependant, cette solution introduit un lien entre l'acteur privé et l'administration, qui devient intéressée à son activité et peut choisir des orientations en matière d'information en fonction des recettes, non des besoins du service public. Les relations privilégiées qui peuvent s'instaurer sont à tous points de vue dangereuses.

La relation entre la redevance et les recettes du réutilisateur pouvait paraître rationnelle, alors qu'elle a un côté absurde. Comme le note le GFII, dans le cas de location de lieux publics pour le tournage de films, nul ne songerait à ce que le tarif de location soit lié au chiffre d'affaires de la projection du film dans la salle.

Aussi, la Directive a posé le principe du lien entre les coûts et les redevances. Certes, il est possible de considérer que c'est seulement le maximum qui est ainsi fixé et que, au dessous, toute solution est possible. Ce n'est pas, sur le fond, ce que demande la Directive.

La CADA a précisé ce point très clairement:

**Avis CADA** : « Il est notamment prévu que l'administration saisie d'une demande en vue de la réutilisation d'informations publiques ne peut intégrer dans la fixation de la redevance un paramètre permettant de tenir compte des recettes que dégagera la réutilisation des informations» (avis 20062243, séance du 29/06/2006).

*On aura noté que la Direction générale des finances publiques, pour les données comptables des collectivités, a pris une position exactement opposée à cet avis de la CADA ainsi qu'à l'esprit de la Directive.*

Cet avis reflète bien un principe de fond, qui s'oppose à la prise en compte de l'activité du réutilisateur et peut même entrer en conflit avec la différenciation entre commercial et non commercial.

Une seule formule a jusqu'ici permis une variabilité de la redevance, la stratégie «bande passante» (appliquée en particulier par Paris), qui consiste à prendre en compte des niveaux de charge des serveurs, qui entraînent un supplément de redevance (*dans le cas de Paris, pour les réutilisateurs qui n'acceptent pas les conditions de la licence – cf. infra*).

*Un soutien à des PME, des jeunes innovateurs etc. est possible par les procédures normales de soutien aux entreprises innovantes, par des appels à projets comme celui de Rennes etc. Il faut seulement prendre garde – comme pour toutes ces actions en général – au droit de la concurrence.*

#### ■ Différenciation des réutilisateurs

Le débat sur l'usage commercial, évoqué précédemment (*cf. 1.3.4*) resurgit constamment pour la définition des redevances.

Cela oblige à rappeler que les distinctions entre les réutilisateurs ne

devraient être établies qu'à partir des coûts pour l'administration, et donc aussi à partir des engagements qu'elle prend. Les formulations devraient être plus précises qu'elles ne le sont en général : ce n'est pas l'usage commercial qui doit déterminer une licence et donc une redevance différente mais des contraintes de coût supérieures pour l'administration.

L'affaire de la généalogie est encore une illustration des possibilités de différenciation (contestées), ici entre l'usage interne et l'usage pour rediffusion par des acteurs ayant tous deux un but commercial.

La décision du Département du Bas Rhin fait le point : « *le seul cas où [le département] percevra une redevance sera la réutilisation commerciale avec diffusion publique d'images. La très grande majorité des généalogistes professionnels ne s'inscrivent pas dans ce cadre, puisque l'usage qui est fait des images, même s'il est en lui-même commercial, ainsi que l'indique expressément un conseil de la CADA qui avait été interrogée sur ce point, est en général interne et sans diffusion publique d'images. Cet usage est, et restera, gratuit.* ».

#### ■ La réutilisation en cascade

Au départ souvent interdite, elle apparaît comme inévitable alors qu'il s'agit de favoriser une économie de l'information mais aussi un écosystème. Le problème est maintenant de savoir quelles règles appliquer – par exemple, les licences "share alike" exigent que le réutilisateur partage à son tour les données retravaillées ou intégrant les données initiales.

Quoi qu'il en soit, l'autorisation et les conditions de ces sous-licences doivent être prévues dans la licence initiale (*cf. Licences*).

Du point de vue des redevances, certaines licences demandent que les redevances soient versées aussi par les réutilisateurs de deuxième niveau. Cette position affecterait notamment les réutilisateurs «ombrelle» qui

peuvent grouper les demandes d'un certain nombre de PME. Elle semble à exclure si on s'en tient au principe que le montant de la redevance est lié au coût de mise à disposition et non à l'utilisation.

#### ■ *Les acquis*

Ils se déduisent des textes et des décisions de la CADA.

- Pas de bénéfice sur le moyen terme, ce qui permet éventuellement de couvrir les coûts d'investissement consentis pour permettre l'accès aux données. Comme le souligne le Guide de la FING, les coûts de transfert de fichiers sont minimes mais le fait de laisser des acteurs extérieurs accéder à des bases de données et à des données temps réel suppose des investissements qui peuvent être élevés, dont l'amortissement peut donc être pris en compte. Il en est de même de coûts spécifiques à la réutilisation, dont le plus important est l'anonymisation, cité explicitement. Il reste à calculer et contrôler, ce qui n'est pas, on va le rappeler, une petite affaire.
- Les redevances ne peuvent dépasser le montant que l'acteur public évalue en comptabilité interne s'il utilise lui-même les données dans une activité commerciale.
- La base de toute redevance est le coût pour celui qui met à disposition. Elle n'est pas la recette du réutilisateur (*ce qui vient d'être rappelé avec le principe de non proportionnalité*), ni la valeur d'usage pour les publics qu'il vise. L'administration n'est pas partenaire d'une opération économique.
- Pas de différences entre les réutilisateurs, autres que celles que pourrait justifier la nature de leur demande et la charge spécifique qu'elle entraîne. Il est possible de distinguer les acteurs commerciaux et non commerciaux mais en termes d'accès global à la réutilisation et en fonction des choix de licences, pas en termes de montant de redevances (*cf. ci-dessus*).

- Mise en place d'une licence, d'application obligatoire s'il y a redevance.
- Transparence quant aux conditions de la réutilisation et au mode de calcul des redevances, qui doivent être publiés, ainsi que la licence type, sur le répertoire lui-même obligatoire.
- Pas de tarifs discriminatoires. Les textes insistent : la tarification doit être transparente et égale pour les réutilisateurs, et ses modalités de calcul ou son niveau ne doivent pas être destinés à privilégier ou exclure certains d'entre eux.
- Pas de «tarifs prédateurs» – trop élevés pour tous (*cf. infra*).

### 3.6.3 Un calcul difficile

#### ■ *Périmètre de la base de calcul*

Les textes parlent du coût de collecte et de production des informations. La pratique montre qu'il s'agit bien des coûts de collecte et de production qui sont entraînés par la réutilisation, non du coût d'une procédure ou d'une prestation objets du service public. La frontière n'est pas simple car, en même temps que se posent les questions de réutilisation, les procédures sont numérisées et les systèmes d'information se développent. La constitution d'une base de données est, on l'a vu, d'abord motivée par l'accroissement d'efficacité et d'efficience du service public. Les frais, même importants, pour l'ouvrir à des réutilisateurs, sont faibles par rapport à la refonte du système d'information.

#### ■ *Estimation des coûts*

Il est toujours difficile d'estimer des coûts marginaux, entraînés par un supplément d'activité – au delà des investissements informatiques, il peut y avoir des suppléments de charge pour les agents (*cf. Ch 4*). Les incertitudes sur le périmètre se cumulent avec les complexités d'une comptabilité analytique administrative.

Il faut distinguer les coûts d'investissement, qui ne sont supportés qu'initialement (et seront certainement renouvelés mais il faudra alors redéfinir les redevances). C'est en particulier le cas de la numérisation, qui peut être exigée par la seule réutilisation – mais elle eût peut-être effectuée sans cela (le problème sera progressivement moins aigu au fur et à mesure que les documents et séries de données sont créés par voie numérique, y compris les formulaires). Pour les investissements, leur prise en compte dans la redevance est délicate parce qu'elle dépend de la durée d'amortissement et de la durée de la réutilisation. La redevance peut donc varier dans le temps selon les solutions adoptées.

Le calcul est celui d'un maximum à ne pas dépasser. Sa complexité pose problème parce qu'il faut théoriquement permettre à des intervenants extérieurs, en cas de recours notamment, d'apprécier le bien fondé du niveau de redevance. C'est pourquoi les textes prévoient une transparence incluant les méthodes d'estimation.

#### ■ *Traitements en vue de la réutilisation, données brutes*

Il est possible de différencier les données nécessitant un traitement particulier, et spécialement l'anonymisation. Il faut cependant, on l'a vu dans la réflexion sur les données (Ch. 2), se garder de chercher à enrichir les données dans le seul but de la réutilisation, au delà des traitements incontournables. Comme l'a souligné le GFII, et comme cela est recommandé généralement (notamment par AEC), il est préférable de mettre à disposition des «données brutes». L'enrichissement, au delà de ce qui est nécessaire pour l'accès technique aux fichiers, est précisément l'affaire des réutilisateurs.

#### ■ *La prise en compte d'un total des redevances*

Le calcul des redevances doit se faire en prenant en compte un maximum total. Ce maximum concerne l'ensemble des réutilisateurs. Le problème ne se pose guère actuellement parce que les réutilisations sont rares ou gratuites,

en dehors des activités et établissements industriels et commerciaux et des grands fichiers déjà évoqués. Il y a cependant là une difficulté potentielle. Hormis la charge supplémentaire de serveurs, un nouveau licencié pour un même ensemble de données devrait faire baisser le montant unitaire des redevances si celles-ci ont atteint leur maximum théorique.

#### ■ *Tarifs prédateurs*

L'application de tarifs prédateurs est théoriquement impossible : de tels tarifs dépasseraient le maximum prévu par la Directive. Etant donné, malheureusement, les difficultés du calcul, et compte tenu des difficultés budgétaires, il est vraisemblable que les tarifs se fixeront en fonction du "consentement à payer" et surtout en fonction de recours auprès de la CADA et de la direction générale de la concurrence, de la consommation et de la répression des fraudes au ministère de l'Economie.

Contrairement à ce que certains peuvent penser, la fixation des tarifs n'est pas totalement libre pour les acteurs qui ne sont pas administrations d'Etat, EPA ou administrations culturelles – elle doit répondre à des règles générales mais aussi à l'interprétation des textes qui régissent la réutilisation.

La généalogie va encore servir de test pour cette nouvelle difficulté qu'est la fixation de tarifs. Les sociétés vont maintenant devoir négocier la réutilisation avec chaque Conseil général, et certains ont demandé jusqu'ici des tarifs manifestement destinés à ne pas être acceptés. Il est vraisemblable qu'un mécanisme et un niveau de redevances vont se standardiser et faire office de référence dans ce domaine, mais rien actuellement ne l'impose.

Or, la CADA, dans son avis n° 20111743 rendu le 26 mai 2011, a estimé que les dispositions de l'article 20 de la loi « ne lui donnent pas compétence pour déterminer le juste montant d'une redevance ou la durée adéquate de la licence de réutilisation, ni d'ailleurs, de façon générale, pour se prononcer sur la décision favorable par laquelle, l'administration

consent un droit à réutilisation. Elle considère, en revanche, qu'elles lui permettent d'émettre un avis dans les cas où les conditions mises par l'administration à la réutilisation sont à ce point contraignantes pour le réutilisateur potentiel qu'elles s'apparentent en réalité à une décision défavorable en matière de réutilisation d'informations publiques »

Elle a rappelé que la redevance pouvait inclure

- le coût de la collecte, de production et de mise à disposition des informations
- le coût éventuel d'anonymisation, si elle se révèle nécessaire.
- une rémunération raisonnable des investissements.

### 3.6.4 Analyses et orientations

L'APIE a publié en décembre 2010 un rapport réalisé par le Bureau d'Economie Théorique et Appliquée (BETA) de l'université de Strasbourg : "La valorisation des informations du secteur public : un modèle économique de tarification optimale".

Prenant en compte de nombreux paramètres, dont la dégradation des finances publiques, elle conclut que, pour assurer une large diffusion de certaines informations publiques, les producteurs publics peuvent avoir intérêt à investir dans le contenu, les formats et les modalités de mise à disposition pour créer une dynamique de la réutilisation. Une tarification raisonnable, permettant de couvrir les dépenses d'enrichissement peut alors être optimale. Elle assure un transfert de charge du contribuable vers le réutilisateur sans diminuer l'efficacité économique globale. Pour les usages non commerciaux, une tarification au coût marginal de mise à disposition est généralement optimale, le consentement à payer des réutilisateurs pour ce type d'usage étant généralement faible (*extrait d'un point de vue GFII sur l'étude*).

Ces conclusions vont à l'encontre des remarques précédemment rappelées du GFII (*Ch. 2*) sur le fait que le producteur ne peut préjuger de l'intérêt de tel ou tel enrichissement. Elles s'opposent donc au consensus de plus en plus partagé sur la mise à disposition de données brutes. Si les données sont bel et bien produites dans le cours de la mission de service public, elles forment un matériau assez aisément accessible. La conclusion du BETA revient à ce que l'administration crée une activité commerciale complémentaire, dont il faudrait prouver qu'elle aura vraiment un impact de marché, qu'elle serait plus rentable que ce que peuvent faire des réutilisateurs et que c'est en développant des activités commerciales que les administrations pourront réduire le déficit des comptes publics.

La position du GFII est rappelée brièvement ici. Dans sa contribution aux travaux du COEPIA sur le sujet, il propose de retenir plusieurs paramètres :

Identité de l'émetteur

- Régime juridique : administration centrale, collectivité territoriale, EPA,...
- Autonomie budgétaire (partielle ou totale)

Les coûts de mise à disposition s'ils sont distincts de ceux que supporte l'administration pour ses besoins propres et/ou leur mise en ligne sur un site public (ie. ouvert) ou fermé

- Disponibilité de ces contenus sur un site public ouvert ou un intranet
- Modification du format initial des données (ex pdf vers html)
- Structuration des données
- Indexation des données
- Modes de livraison (FTP, DVD,....)
- Nombre de licenciés (puisque le total des redevances ne doit pas dépasser les coûts)



- Volumétrie
  - Volumétrie et fréquence des mises à jour
  - Prise en compte des évolutions technologiques.
- Le contexte européen ou international
- Tarifs pratiqués dans les autres Etats membres, à qualité de données identique
  - Tarifs pratiqués par des revendeurs privés

### 3.7 Comment ? Licences

#### 3.7.1 Une exigence

La loi exige que tout acteur public mettant à disposition des données définisse une licence type de réutilisation et la publie sur son site Web s'il en a un.

Cette licence type sera obligatoirement le cadre d'application dès qu'une redevance est exigée. Des conditions particulières sont définies en fonction de la nature des données, et la licence doit permettre ces adaptations. En tout état de cause, une licence respectant la licence type doit être signée s'il y a redevance.

Il est souhaitable de distinguer des variantes pour des situations de réutilisation nettement différentes, et en particulier de définir des licences de réutilisation successive pour les données régulièrement mises à jour. L'APIE propose les deux types de licences sur son site.

On verra qu'il est toujours préférable de signer une licence, même pour une réutilisation gratuite, afin de préserver des garanties pour les deux parties. Néanmoins, si ce n'est pas le cas, il faut au moins se placer dans des

- conditions générales de réutilisation, qui doivent être annoncées. Le modèle de ces CGR, qui s'applique à l'Etat et à ses EPA, a été mis au point par l'APIE. Plusieurs collectivités l'ont choisi (Bordeaux, Montpellier, Rennes avec des adaptations).

Les principales caractéristiques des CGR sont d'abord le rappel de règles sur les données (notamment pour les données nominatives) et des contraintes sur la réutilisation, qui reprennent l'article 12 de la loi de 1978 : le réutilisateur s'engage à ne pas altérer les Informations publiques, à ne pas en dénaturer le sens et à indiquer leur source et la date de leur dernière mise à jour. Comme cela a déjà été indiqué (*cf. Ch. 2 – Données*), les contraintes relatives à l'altération et à la dénaturation devraient être précisées et strictement limitées, pour ne pas bloquer en fait l'intérêt de la réutilisation. Ce n'est pas le cas actuellement, et il est vraisemblable que l'expérience conduira à préciser de quoi il s'agit.

#### 3.7.2 Le monde du libre et ses licences

[Des tableaux comparatifs des différentes licences sont fournis en annexe – Source principale : <http://www.donneeslibres.fr/>]

Le souci de permettre une valorisation des gisements de données et d'être compatible avec des licences internationales (*cf. infra*) a conduit à s'orienter vers les différentes licences du libre.

Le monde des données ouvertes (pas seulement publiques) fait partie d'un courant culturel plus vaste qui se trouve aujourd'hui présent dans trois domaines :

- le logiciel libre, avec les licences GPL, BSD, CECILL (France)...
- la diffusion d'œuvres et la propriété intellectuelle avec les licences Creative Commons («CC»)



**Creative Commons** : trois caractéristiques optionnelles

**By** : - obligation d'attribution, c'est à dire obligation d'indiquer l'auteur

**SA** : share alike – obligation de licencier un document dérivé d'un document source dans les mêmes conditions que celles ayant permis son obtention (même licence, mêmes conditions financières éventuelles)

**NC** : exclusion des usages commerciaux

- l'ouverture de publications scientifiques avec l'Open Archive et son modèle d'accès ouvert international aux ouvrages, thèses etc.

Les différentes licences traitent de problèmes communs aux trois domaines: la définition et la protection de la propriété intellectuelle, la capacité à rediffuser et réutiliser, la capacité à étendre et retravailler, la possibilité d'intégrer les outils et documents correspondants dans des productions nouvelles et des services, commerciaux ou non. Pour l'essentiel, il s'agit de redéfinir le droit d'auteur et le copyright<sup>19</sup>.

Il faut en tout état de cause adapter les licences à la législation française. C'est ce qu'a fait le premier le ministère de la Justice en rédigeant une licence «Informations publiques» (IP) (<http://www.rip.justice.fr/>) adaptant la licence CC By SA. La mission Etalab et l'APIE travaillent à l'été 2011 sur ce dossier.

#### ■ Attribution et droit d'auteur

Or, la réutilisation des données publiques concerne surtout des données gérées dans des bases de données, pour lesquelles le droit d'auteur ne

<sup>19</sup> On parle parfois de copyleft pour insister sur le fait que, sans contredire le copyright, les licences libres modifient l'exercice des droits en décrivant les possibilités qu'a l'utilisateur de manipuler, transformer, redistribuer données ou documents.

□ s'applique pas, ou pas fréquemment (*cf. ci-dessus*). C'est l'ensemble de la base et non telle ou telle donnée qui peut être protégée dans certains cas. Quoi qu'il en soit, il semble bien qu'il faille pouvoir indiquer, pour un usage important, quel est l'auteur (ou l'institution auteur en général). C'est le principe d'attribution – mais son application est à l'évidence difficile dès lors que les données sont mêlées à d'autres.

En tout cas, les équipes qui mettent au point les bases et les gèrent, déjà sensibles au fait que d'autres peuvent en tirer un revenu, souhaiteront en général que leur travail soit reconnu et valorisé. Il est aussi légitime que les citoyens et les entreprises sachent que leurs contributions (fiscales mais aussi comme fournisseurs d'information le cas échéant) sont à l'origine de tel ou tel service.

#### ■ Share alike

L'autre principe en discussion est le share alike, issu des règles CC. Il indique que le réutilisateur doit à son tour ouvrir les données qu'il a retravaillées, dans les conditions qui lui ont été consenties par la source. Cette règle ne s'applique bien sûr pas si la réutilisation en cascade est interdite, ce qui est de moins en moins possible. En même temps, la contrainte de share alike est difficile à appliquer pour la constitution de services à partir de données publiques. Un acteur commercial ne voudra pas ouvrir des données enrichies à ses concurrents, en tout cas pas avec les mêmes conditions. Si la réutilisation est gratuite, il ne pourra développer son activité.

#### ■ Les bases de données

La licence CC0 exclut toutes les contraintes et elle peut donc s'adapter à la constitution d'une base de données. Cependant, le fait qu'elle n'inclut pas de contrainte d'attribution représente un manque que beaucoup considèrent comme handicapant.

C'est pour répondre à ces difficultés que la Ville de Paris a travaillé avec plusieurs associations du logiciel libre pour adapter l'Open Database License de l'Open Knowledge Foundation au droit français. La licence OdbL comprend l'attribution et le share alike.

Le reproche fait à cette démarche est que cette solution est pour l'instant minoritaire dans les autres pays et qu'elle inclut la contrainte de share alike, dont on a vu qu'elle bloque en pratique l'activité commerciale. A Paris, d'ailleurs, les réutilisateurs qui ne l'acceptent pas doivent passer par un contrat spécial assorti de redevances.

### 3.7.3 Quelle licence adopter ?

La diversité des licences est aujourd'hui très grande. Or, elle gêne de plus en plus tous les acteurs. Elle oblige les administrations et autres organismes publics à choisir, adapter, écrire des licences. Elle oblige les réutilisateurs désireux de s'adresser à plusieurs administrations à comprendre les différences, à s'adapter eux-mêmes à tous les cas et à jouir de droits différents sur le même type de données.

#### ■ *L'interopérabilité internationale*

Au delà, pour entrer dans le marché européen et mondial de l'information, au sens large (gratuit comme commercial), il faut une forte interopérabilité des licences pour permettre la réutilisation transfrontalière et permettre à des réutilisateurs étrangers d'intervenir (ce peut être très important, par exemple, pour l'attractivité d'un territoire). Il y aura toujours des spécificités nationales mais, pour l'essentiel, il faut un cadre national juridique précis et il faut que les divergences avec les autres pays d'Europe puis du monde soient à la fois clairement indiquées et limitées au strict nécessaire. L'interopérabilité européenne devrait être atteinte le plus vite possible, en

liaison mais si possible avant la révision de la Directive, prévue pour 2013 au plus tôt. Une licence type européenne s'impose – ou en tout cas un cadre commun pour l'écriture des licences si des domaines doivent être distingués. Cela paraît nécessaire au regard de l'ouverture d'un marché unique des TIC et à la mise en œuvre de la stratégie de croissance européenne (Europe 2020 qui prolonge la décennie de la stratégie de Lisbonne). La circulation et la valorisation de l'information en sont des outils importants.

#### ■ *Travaux et réflexions en cours*

Le chantier est ouvert, puisque beaucoup des licences françaises actuelles sont des adaptations de licences portées par des groupes internationaux du monde du libre, Creative Commons et ODBL. Comme on l'a vu, cependant, il est difficile de choisir puisqu'au niveau international non plus, il n'y a pas de règle unique, et il faudra de toutes façons adapter à la législation française et peut-être différencier les licences en fonction de grandes catégories de données.

Etant données les difficultés déjà signalées et la conscience qu'il faut en première étape converger vers une première version de licence type nationale, la tendance dominante en 2011 est de se tourner vers l'APIE dès qu'il y a des redevances. Néanmoins, les contraintes de non altération et non dénaturation paraissent inquiétantes à beaucoup, qui les considèrent comme floues et permettant des interprétations s'opposant, entre autres, à la coopération avec des projets fondamentaux comme Wikipedia et OpenStreetMap. Les licences CC paraissent alors préférables, si possible avec la contrainte d'attribution, et donc ni SA ni NC (*voir ci-dessus – la licence qui est sans aucune contrainte étant dite CCO mais il lui manque l'attribution*). Cependant, le choix du Non Commercial est assez fréquent – selon plusieurs intervenants dans les réunions, c'est une position plus fréquente en France que dans la plupart des autres pays (en fait il s'agit en général des pays anglo-saxons).

A l'été 2011, Etalab a créé un groupe de travail avec l'APIE pour produire une licence type nationale compatible avec les licences libres qui sera publiée sur le répertoire national data.gouv.fr. Une proposition devrait être présentée à l'automne.

#### ■ *Licence clic*

D'abord introduite par le portail britannique data.gov, la licence clic est maintenant de plus en plus présente en France. Elle permet une réutilisation non commerciale de données sans aucune formalité autre que le fait de signer un engagement sur le Web.

#### **Exemple : licence clic du Conseil général de l'Indre**

Le site propose notamment les données des archives. La licence clic pose les conditions suivantes (*extrait sans les commentaires*):

- Ne pas faire un usage commercial des données
- Ne pas les diffuser publiquement ou à des tiers, sauf dans le cadre d'un usage privé ou interne.
- Citer systématiquement, de façon visible, et quel que soit l'usage que vous souhaitez faire de l'information :
  - les auteurs des textes réutilisés s'il y en a (par exemple pour les fiches de recherche),
  - les Archives départementales de l'Indre comme lieu de conservation du document, et la cote des documents (qui est toujours indiquée sur les images du site).
- Ne modifier les données qu'avec le consentement écrit des Archives départementales de l'Indre.

A ces conditions, toute personne est libre d'utiliser les informations

et données mises en ligne sans avoir à contacter préalablement les Archives départementales de l'Indre.

Une interprétation large du concept, dont se plaignent les internautes, demande l'envoi de documents papier et une acceptation formelle par l'administration.

Or, cette licence a pour principal objectif de permettre des réutilisations ponctuelles et immédiates : illustration d'un document ou d'une page dans un réseau social, série statistique sur un cours, information sur un territoire pour le site ou un document d'une association ou pour un article Wikipedia etc. C'est d'ailleurs pour des utilisations de ce type que le rapport sur le patrimoine culturel (Ory Lavollée – Partager notre patrimoine culturel – 2009) en proposait le premier l'utilisation.

#### ■ *Plusieurs licences*

L'administration peut proposer plusieurs licences. Il semble qu'elles puissent être différenciées selon les critères suivants:

- Types de données («essentiels», «culturelles», etc., en fonction des politiques de diffusion et, pour les redevances, des coûts induits)
- Engagements demandés aux réutilisateurs (de la licence clic aux licences avec contraintes pour l'acteur public, par exemple en termes de délais).

Comme on l'a vu précédemment, les différenciations entre réutilisations commerciales et non commerciales ne peuvent se justifier sur le fond que si les engagements des deux parties sont différents, conduisant à des licences différentes.

Comme on l'a vu précédemment, le cas des réutilisations qui visent non à une forme quelconque de rediffusion mais à un usage interne par un acteur,

lui-même commercial ou non commercial, peut être traité par une licence spéciale : elle ne comprend pas les engagements découlant de la diffusion des données et l'entreprise – ou l'association – est responsable des usages qu'elle fait en interne. Il y a donc souvent trois types de licences – qui peuvent s'analyser en termes d'engagements croissants : libre sans contrainte et sans engagements (clic), avec des engagements limités, avec des engagements forts. Il serait préférable d'identifier ainsi les licences (et non en utilisant le terme de «commercial» pour désigner les licences avec engagement des parties et redevances, ce qui correspond à des cas fréquents mais peut conduire à des contre-sens lorsque le marché des données aura progressé).

Il est conforme au droit de la concurrence comme aux textes sur les données publiques que des droits et devoirs différents puissent être distingués dans des licences. En revanche, il paraît contraire à la loi qu'une même licence s'accompagne de redevances différentes – la veille juridique permettra de suivre les évolutions sur ce point.

### 3.7.4 Les éléments d'une licence : les conditions particulières

Les conditions particulières sont des cadres au sein duquel l'acteur public et le réutilisateur précisent les engagements mutuels. Elles devront être identiques pour les mêmes jeux de données quels que soient les réutilisateurs.

La liste d'éléments qui suit regroupe ceux qui figurent dans les conditions générales de l'APIE avec des recommandations du Guide AEC, du GFII et d'intervenants dans les débats. Elle peut servir de base de discussion. Il serait nécessaire de mettre au point progressivement une grille comprenant des éléments obligatoires et optionnels. Une telle grille se construira au fur et à mesure des opérations concrètes.

Le plan des rubriques ci-après est proche de celui de l'APIE.

#### **Nature et caractéristiques des informations**

- Dénomination des informations
- Description détaillée du contenu (avec un fichier test joint)
- Source et date de mise à disposition (pour les licences sans livraisons successives)
- Périodicité de la livraison des données en cas de livraisons successives (ou les événements déclenchants)
- Mode d'organisation et de présentation des informations

#### **Finalités de la réutilisation des informations**

- caractère commercial ou non
- réutilisations envisagées (il faut noter que les licences APIE ne prévoient pas cette rubrique, qui est utile – il ne s'agit pas de contrôler mais de connaître les usages prévus, en tout cas les premiers

#### **Caractéristiques de la licence**

- Durée (période, durée de validité)
- Conditions de cession de la licence, possibilité, ou non, de sous-licences de réutilisation et contraintes correspondantes
- Conditions de reconduction, si nécessaire, : tacite ou non.

#### **Droits de propriété intellectuelle de l'Administration**

- Nature des droits
- Conditions d'exploitation

#### **Modalités particulières de mise à disposition**

- Délais de première mise à disposition des informations
- Délais de mise à disposition des mises à jour

- Formats et modalités techniques de transmission des informations
- Eventuelles restrictions d'utilisation que l'acteur public demande et contraintes relatives à la modification et au traitement des données (référence à la dénaturation)

#### Éventuels systèmes de protection des données

**Interlocuteurs, lieu de livraison, modalités de réception, litiges et possibilités de règlement** (conditions commerciales générales)

#### Conditions financières

- Redevance
- Éléments de coût justifiant la tarification
- Délais et modalités de paiement

#### Résiliation

- Modalités
- Statut des données reçues en cas de résiliation

#### Manquements, difficultés

- Procédures et engagements en cas de changement important dans les contenus et formats des données, interruption de la fourniture
- Indemnités dues par l'administration en cas de manquement aux conditions ou résiliation

#### Références juridiques applicables

Les licences sont importantes et la mise au point de licences-type nationales, compatibles avec les licences internationales, est aussi nécessaire que les normes de base qui permettent à Internet d'être une plate-forme mondiale.

- Il est impératif, cependant, de rester souple et de veiller au fond du respect des engagements tout en sachant que de nouvelles possibilités conduiront peut-être à revoir les termes des contrats. Nous sommes passés en 30 ans des banques de données réservées à de grandes entreprises au Minitel puis au micro puis à Internet puis au mobile et enfin, en 2011, aux tablettes. Déjà s'annoncent de nouveaux développements et l'évolution est tous les jours plus rapide. On sait ce qu'il peut en coûter de manquer une évolution et de s'accrocher à des outils mais aussi à des textes dépassés. C'est pourquoi il faut surtout s'en tenir aux objectifs et aux engagements de fond, et maintenir des relations avec les réutilisateurs.

### 3.8 Relations avec les réutilisateurs et les partenaires industriels

#### 3.8.1 Le suivi des réutilisations

##### ■ *Nécessité de rigueur et de dialogue*

Les situations évolueront et de nouveaux problèmes, mais aussi de nouvelles opportunités, vont se présenter.

On l'a vu, les licences et les contrats vont devoir, comme les règles, être solides sur le fond et souples dans la forme, et il est important de maintenir des relations avec les utilisateurs – en excluant de ce souci, en général, les réutilisateurs ponctuels signataires de licences clic. Il est aussi important de signer une licence, même pour les réutilisations gratuites.

Il faut en effet fixer des responsabilités des deux parties. Par exemple, une jeune pousse (start-up) ayant un accès gratuit à des données pour une

application mobile doit être protégée contre un arrêt de la mise à disposition ou contre un changement brutal des formats et interfaces d'accès. Si l'application a rencontré un succès, ce sont aussi les utilisateurs qui seront lésés. Comme on l'a vu, même si ce n'est pas toujours formalisé, l'ouverture de données publiques crée des responsabilités, et ce auprès des citoyens eux-mêmes.

#### ■ *Des garanties réciproques raisonnables*

Il importe donc que les deux parties s'engagent réellement sur des garanties réciproques.

Celles-ci doivent être cependant limitées au nécessaires et avec des clauses de prudence. Dans le monde de l'information numérique, de multiples facteurs peuvent empêcher une des parties de satisfaire à des obligations trop strictes. Ainsi, il faut estimer, notamment avec les services informatiques, les limites en termes de charge, de délais, de capacité des accès temps réel etc.

#### ■ *Traçabilité*

L'intérêt de tous est la traçabilité des données, par l'enregistrement de l'historique des transactions, par l'inscription des sources, des producteurs, des responsables, des entités ayant validé les données le cas échéant. Ces tâches incombent au partenaire public mais il est souhaitable que chaque partie conserve des traces.

Dans le cas de contrats portant sur des données sensibles et avec un fort impact en cas d'erreur ou de retard, il est possible de se tourner vers un service de notariat informatique, qui pourra garder trace des transactions.

#### ■ *Dénaturation*

□ Ce concept flou a déjà été évoqué (*Ch. 2*) et il faut l'objet de beaucoup de critiques, car il est difficile de distinguer une dénaturation par erreur grave ou malveillance d'un traitement et d'un enrichissement des données. Il résulte de la loi CADA article 12 – obligation de loyauté. Une relation maintenue avec le réutilisateur doit permettre, si les deux parties sont loyales, de gérer les divergences. Dans les autres cas, on l'a vu, il faudra à terme compter sur les utilisateurs qui signaleront les problèmes (et se plaindront souvent auprès de l'administration fournisseur).

### 3.8.2 Prestations de service, partenariats et réutilisation

L'intervention de nombreux industriels dans la gestion et la circulation des données pose parfois et posera à terme la question de la frontière entre les diverses formes de relations.

L'exemple des propositions de Google aux bibliothèques montre comment le problème peut se poser : Google comme avant lui Microsoft pour les musées propose de prendre en charge le coût de numérisation et de pouvoir utiliser les données dans ses services, au demeurant gratuits. Le problème est que cela donne à une entreprise un contrôle sur toute l'opération et un monopole de fait, ou en tout cas à terme un rôle de concessionnaire pouvant céder à son tour des droits. Le problème s'est posé pour les musées dans les années 90. La BNF a choisi de refuser cette offre mais de lancer un appel pour un Partenariat Public Privé (PPP), où elle reste le maître d'ouvrage.

#### ■ *Traitement des données pour permettre leur utilisation*

Des traitements sont nécessaires pour mettre en ligne des données, qu'il s'agisse de publication par l'administration ou de mise à disposition de réutilisateurs.

Il sera toujours difficile de distinguer ce qui est lié à la réutilisation des données, surtout pour la numérisation, de plus en plus nécessaire à la gestion des données et documents. Il en est de même pour l'indexation, tâche souvent très lourde. Enfin, l'anonymisation ne se pose que si des publications ou des communications externes sont envisagées. A son sujet, le décret du 30 décembre 2005 a précisé que «*Lorsque la réutilisation n'est possible qu'après anonymisation des données à caractère personnel, l'autorité détentrice y procède sous réserve que cette opération n'entraîne pas des efforts disproportionnés*»

La question se pose donc de la prise en charge de ces tâches si elles représentent une charge importante, ce qui est très souvent le cas, et si la collectivité concernée ne peut pas les assurer seule, ou pas dans des délais suffisants, soit en interne soit en sous-traitance.

- **Prise en charge d'un traitement général par un réutilisateur, en général la numérisation**

Confrontée à une charge importante, l'administration va être conduite à rechercher ou accepter des accords avec des partenaires qui assureront ce travail en échange d'une formule avantageuse de réutilisation – en général le service administratif conserve le droit d'usage sur les données pour ses propres besoins; il n'y a que la réutilisation par d'autres acteurs qui pose problème.

Il n'est pas possible d'accorder une exclusivité à l'entreprise qui a réalisé la numérisation (ou aussi l'indexation). Il est aussi essentiel que la collectivité conserve le contrôle des demandes de réutilisation, qui ne peut être déléguée. Il faut donc prévoir, dans un contrat spécial, des versements de la collectivité à l'entreprise pour toute réutilisation, en prenant en compte les volumes concernés en particulier. Ces versements seront prélevés sur les redevances le cas échéant, celles-ci intégrant cet élément de coût. Pour des réutilisations gratuites, le contrat doit prévoir

les conditions d'éventuels paiements – qui seront équivalents à une charge de sous-traitance.

- **Prise en charge directe par un réutilisateur**

Ce cas correspond souvent à l'anonymisation, qu'une entreprise réutilisatrice peut effectuer elle-même pour ses seuls services. C'est pour elle et pour ses clients à un enrichissement des données. Ceci est d'autant plus fréquent que l'anonymisation est une tâche complexe, qui peut dépendre des croisements que le réutilisateur va opérer et des usages qu'il veut développer (cf. *anonymisation in Ch. 4*).

Le fait d'autoriser le réutilisateur à effectuer cette opération, et donc de lui donner un accès à des données non accessibles en l'état, correspond à un changement de nature du contrat et oblige à préciser très exactement les conditions de confidentialité de l'opération. Celle-ci doit être déclarée à la CNIL et être autorisée. Dans la plupart des cas, les personnes qui désapprouveraient le service résultant peuvent demander à ce que leurs données soient retirées (opt out).

*Remarque: la même situation peut se poser pour des travaux universitaires.*

- **Partenariat**

Un partenariat, notamment de type PPP doit être géré comme dans le cas de travaux supportés par le réutilisateur et réutilisables par d'autres (*ci-dessus*) si les données sont incontestablement publiques – c'est-à-dire acquises dans le cadre d'une mission de service public. Dans ce cas, le partenaire ne doit pas tirer de cette situation une exclusivité, ni se voir déléguer la gestion des réutilisateurs. Le PPP concerne avant tout un arrangement financier. Les autres réutilisateurs doivent se voir appliquer des redevances dont une partie devra vraisemblablement – selon le contrat – être partagée avec le partenaire.



Le partenaire privé acquiert à l'évidence des droits sur les données et il faut veiller à clairement identifier ceux-ci. Cela renvoie à l'existence de droits attachés aux données (*cf. ci-dessus droits de propriété intellectuelle*). En particulier, le partenaire peut détenir ou codétenir les droits de propriété intellectuelle sur l'architecture de la base de données.

Néanmoins, l'ensemble de ces droits ne devrait pas correspondre à un transfert de propriété, aliénation du bien public initial. Ils doivent être très précisément décrits dans le contrat et ils vont entraîner des contraintes pour les autres réutilisateurs – par exemple des limites quant au transfert de certaines extensions techniques des fichiers, mais ce n'est pas souhaitable. Dans ce cas, il faut prévoir dès le départ quels fichiers types pourront être proposés aux réutilisateurs.

## 4 Economie, gestion, entreprises

Les analyses économiques présentées à propos de la mise à disposition et de la réutilisation des données publiques souffrent de l'imprécision initiale de la question et des incertitudes fortes sur la mise en œuvre, et ce depuis le début du débat sur le sujet, au niveau de l'Etat.

Déjà, en 1998, lors des premiers travaux au ministère de l'Economie sur le commerce électronique, il apparaissait que l'Etat rachetait souvent, par subvention ou directement par le budget d'autres services, les informations qui engendraient quelques recettes de poche dans d'autres départements et établissements publics.

L'analyse économique est pratiquement impossible, en fait, alors qu'un système se met en place et qu'aucune organisation et aucune règle ne sont vraiment stabilisées. Il faut seulement poser une problématique, et distinguer :

- Les problèmes de gestion à court terme et le financement des actions
- L'impact direct attendu sur le producteur / éditeur d'information
- L'impact économique général

C'est par ce dernier volet que le sujet est abordé ici, car il est maintenant clair que l'essentiel est bien le développement économique lié à la circulation et à l'utilisation des réseaux et des données, et non l'amélioration de la situation budgétaire des acteurs publics. C'est l'objet de la politique européenne aussi bien que des orientations maintenant affirmées par l'Etat en France.

Le rapport 2010 du Conseil d'orientation de l'édition publique et de l'information administrative le confirme clairement :

« L'ensemble de la formation spécialisée [experts réunis par le COEPIA] s'est accordé pour considérer que l'objectif économique et social du développement de la réutilisation des informations publiques est essentiel et doit gouverner les actions mises en œuvre pour favoriser la réutilisation.»

Au demeurant, la question des redevances, si elle est incontournable et importante, tend à polariser des positions idéologiques autant qu'économiques, risque de bloquer des projets au départ et restreint la vision à des problèmes de court terme, sur des montants en général limités. L'exemple des institutions nationales montre qu'on parvient à terme à dépasser ce débat. Cette question sera abordée dans le chapitre consacré aux modalités de la réutilisation.

Le présent chapitre, après avoir abordé l'économie d'ensemble et la gestion de la collectivité, se place du point de vue des entreprises qui sont sur le marché ou y entrent.

## Impact économique

### 4.1 Confrontation à un nouvel élan de la dynamique TIC

#### 4.1.1 Raison garder...

Une démarche économique est certes indispensable. Encore faut-il qu'elle soit quelque peu raisonnable et ce n'est souvent pas le cas. A une époque où il est de bon ton d'affirmer l'importance de l'économie et le souci de la gestion, et alors que des contraintes économiques bien réelles pèsent sur tous, il était inévitable que de réelles opportunités sociales et techniques soient parées des vertus d'un nouvel eldorado. Cela a été le cas, il faut le rappeler, de tous les développements de l'informatique, des communications, et en

général des technologies nouvelles...jusqu'à la bulle suivante.

- Pour résumer, toute nouvelle avancée des TIC est toujours supposée :
- permettre, pour ceux qui la mettent en œuvre, le développement des activités en réduisant les coûts ;
- susciter dans un cercle vertueux le développement aussi bien des offreurs de techniques et de services que de l'ensemble de l'économie.

Les expériences précédentes ont rendu gestionnaires et responsables méfiants face à ces promesses, et le risque est qu'à une période d'enthousiasme succèdent déceptions et rejet.

#### 4.1.2 ... Mais avancer à temps

Pourtant, après un temps d'apprentissage et d'émergence des vraies créations de valeur sociale et économique, les usages se diffusent et des équilibres nouveaux apparaissent.

L'époque actuelle est incontestablement marquée par l'accroissement du rôle des techniques d'information et de la circulation rapide des données. Le méconnaître est s'exposer à des déboires, et ce d'autant que les histoires sociales et techniques s'accélèrent, et que personne ne repasse les plats dans ce qu'on appelle, peut-être un peu excessivement<sup>20</sup>, l'économie de l'information. Comme le dit Charles Népote (spécialiste des données publiques à la FING), si vous ne libérez pas vos données, d'autres le feront pour vous.

<sup>20</sup> Les responsables politiques et économiques savent que des problèmes très ardues se posent, qui n'ont pas grand chose à voir avec l'information, et que c'est encore plus vrai s'agissant de la situation internationale et de la planète. Les TIC peuvent apporter leur contribution mais leur influence ne peut être sensible qu'à moyen et long terme. Aujourd'hui encore, l'agriculture fait à peu près jeu égal avec les industries électroniques en Californie. En outre, le contenu en import de l'agriculture est bien plus faible...L'important est que l'agriculture californienne et ses exportations importantes utilisent les techniques informatiques (par ex pour le vin).

- La prise en compte trop tardive de l'importance de la micro-informatique ou d'Internet par les responsables, aussi bien politiques qu'universitaires et chefs d'entreprise, a déjà coûté des efforts de rattrapage très importants en France.

Comme cela a été précisé, l'affaire n'est pas ici seulement l'ouverture des données publiques, c'est la possible sous-estimation de l'importance de la production de données puis de leur circulation, qu'elles soient publiques ou privées.

Ainsi importe-t-il à la fois de ne pas s'illusionner et d'avancer sérieusement.

#### 4.1.3 Un investissement en avenir incertain

Il importe de rappeler...

- que les mesures économiques, maintenant contestées au niveau national (PIB vs Bien-être), sont d'autant moins aisées à établir et interpréter que l'observation se fait sur des territoires plus restreints,
- qu'il n'y a malheureusement pas de miracle en économie et que rien ne permet d'accélérer un développement de façon spectaculaire,
- qu'aucun barbier n'est gratuit, et que la mise en place de nouveaux systèmes rapporte aux uns mais coûte aux autres,
- que les techniques, les services, les acteurs du domaine de l'information sont en très rapide évolution,
- que l'économie dans son ensemble est actuellement dans une période d'incertitude et de crise, bien au delà de l'économie de l'information, et que les hypothèses sur l'avenir sont aussi risquées qu'indispensables.

Ainsi, mesurer l'impact de la libération des données publiques ne peut se faire qu'en avenir incertain, puisqu'il s'agit de mécanismes nouveaux, de modes d'appropriation sociale récents et encore largement en développement (avec

l'effet cumulé des nouveaux outils mobiles et de l'impact démographique, les générations n'ayant pas les mêmes comportements). Ce n'est pas seulement vrai pour les données mais pour tout l'aménagement numérique : il s'agit pour l'essentiel de suivre un mouvement d'ensemble et de donner aux acteurs du territoire les mêmes moyens que ceux qui se développent ailleurs, si possible avec les meilleurs atouts. Mesurer l'impact sur une brève période est aussi vain que cela l'eût été pour les premières chaussées automobiles : c'est toute l'organisation et la répartition des activités sur le territoire que l'automobile a modifiés – profitant aux uns et faisant pratiquement disparaître d'autres. Pourtant, ignorer l'automobile était la seule option impossible pour un aménageur.

Fondamentalement, il s'agit de décider d'un investissement, humain et financier, pour assurer un développement de moyen et long terme.

Au delà de ses impacts économiques directs ou indirects mesurables à terme, cet investissement des collectivités est au demeurant d'abord déterminé par l'apport de meilleurs services aux citoyens ainsi qu'aux activités sociales et économiques sur le territoire.

#### 4.1.4 La circulation d'information : un gain assuré à long terme

Aujourd'hui, après les "tuyaux", il est demandé à l'Etat et aux collectivités de porter une attention accrue aux données et à l'information qu'ils produisent ou contrôlent, parce qu'il s'agit d'un gisement impressionnant, sans doute plus de la moitié des informations appelées à circuler sur les réseaux de manière publique, gratuitement ou commercialement (par définition, l'information des entreprises et d'autres groupes privés ne peut circuler librement que si cela apporte un avantage dans un monde concurrentiel<sup>21</sup>).

21 Des entreprises l'ont fait. Ainsi, IBM a rendu publique sa très importante base de données sur les brevets, ce qui a semé beaucoup de perturbations dans le monde des offices spécialisés.

Pour les infrastructures, il apparaissait clairement que la puissance publique en charge de l'aménagement du territoire se devait d'assurer l'accès égal, autant que possible, comme cela a été le cas pour les transports. En revanche, la collecte et la diffusion ou la mise à disposition d'information était une activité de support pour des fonctions opérationnelles. Le problème du libre accès aux données, qui apparaissent, contrairement aux infrastructures, comme des «propriétés» de l'acteur public concerné n'était pas directement posé en termes de service public.

Il y a d'abord un pari optimiste : au delà des effets positifs autres pour les citoyens, la circulation améliorée de données concernant un territoire, sur ce territoire et au dehors, doit permettre à tous les acteurs économiques d'améliorer leur fonctionnement en étant mieux informés et elle doit ouvrir à tous de nouvelles opportunités.

L'importance économique de la circulation de l'information a été mise en évidence depuis longtemps, et elle ne date évidemment pas de l'informatique, qu'il s'agisse des innovations, de la monnaie, des informations commerciales et financières etc. Les TIC ont marqué une nouvelle étape dans une progression historique de la vitesse de circulation, qui a permis le passage d'un seuil qualitatif, avec les avantages et les risques afférents.

## 4.2 Economie du territoire

### 4.2.1 Les acteurs des réseaux et de l'informatique

Chacun peut relever que les premiers développements (à Brest, Rennes, Montpellier etc.) sont souvent assurés par des entreprises locales, et notamment par des start-ups. Les nouvelles applications sont pour elles un sérieux appoint, et ces entreprises sont visées par les appels à projets. Proches des préoccupations locales ou régionales, s'adressant au départ à la population locale, elles sont bien placées pour développer applications et

marchés correspondants. C'est particulièrement net pour les informations transport sur les mobiles et plus généralement tout ce qui est lié à la géolocalisation. La coopération avec l'Université permet la création de jeunes pousses et la formation de jeunes entrepreneurs.

Certaines entreprises, comme NewLC à Rennes, ont développé des applications suffisamment génériques pour pouvoir être ensuite proposées à d'autres villes (application transports sur les mobiles eoCity).

Il ne faut pas cependant se leurrer à ce sujet. La concentration va jouer, et elle est en général particulièrement rapide dans le monde des TIC. Certaines entreprises dépasseront le cadre local, d'autres resteront sur des niches et la plupart disparaîtront. Les grands de l'informatique et des réseaux – qui en rachèteront quelques unes – vont arriver. Il ne s'agit pas de le craindre mais de se préparer à des relations un peu plus conventionnelles et souvent plus difficiles qu'avec les PMI locales<sup>22</sup>.

L'essentiel du gain économique ne sera pas dans les entreprises des TIC mais dans l'usage de l'information. Le passé a montré qu'il ne fallait pas se concentrer trop sur l'outil technique – il y avait en 1986 dans toutes les grandes villes françaises ou presque une entreprise de production de micro-ordinateurs largement soutenue localement – au total 36 entreprises, dont pas une n'a survécu. Cette prolifération protégée par les autorités locales a conduit à un émiettement industriel catastrophique sur ce secteur hautement concurrentiel.

#### 4.2.2 Les autres acteurs économiques : des gains en améliorant et différenciant l'information

C'est sur l'ensemble des secteurs économiques que la circulation des données publiques va, en moyenne, avoir un impact positif. Celui-ci dépend

<sup>22</sup> On voit ainsi Google, Yahoo et Microsoft se grouper pour définir une norme «Schema.org» pour la mise à disposition des données publiques (cf *Technique*)

des politiques qui seront menées, et qui seront adaptées aux territoires. Comme cela a été vu, dans tous les domaines, les collectivités mais aussi les autres institutions régionales (Chambres consulaires notamment) et l'Etat disposent d'une importante masse d'informations. Elles porteront sur la forêt, l'aéronautique ou la pêche en fonction des caractéristiques régionales. Globalement, la mise à disposition d'informations doit aider les entreprises à la fois à mieux maîtriser leur production et leurs marchés et à se faire mieux connaître sur les marchés étrangers.

La réutilisation apporte ici l'un de ses bénéfiques annoncés : la capacité d'entreprises (ou d'associations et autres groupements) à s'adresser à une grande diversité de publics, à l'échelle nationale et internationale. Cela implique d'ailleurs qu'elle s'étende à des diffuseurs étrangers, ce qui va au départ être parfois difficile, en raison des différences juridiques, et va en choquer plus d'un : ces entreprises vont gagner de l'argent en vendant éventuellement des informations sur la région, alors qu'elles n'ont en rien contribué à leur production, si elles ne paient pas d'impôt en France, et que le retour fiscal sera souvent faible ou nul. C'est la règle dans le monde d'Internet, et il n'est pas possible à moyen terme d'envisager un autre fonctionnement.

Dans plusieurs domaines, les gains à attendre sont certains, et ce dans tous les territoires, mais également les problèmes soulevés par une information largement diffusée et moins contrôlée.

Attractivité du territoire. La mise à disposition d'informations sur le territoire va, en général, accroître son attractivité, tant pour les personnes que pour les entreprises qui veulent s'y installer ou travailler avec des acteurs locaux. C'est bien sûr un pari et une affaire de concurrence, y compris entre territoires, et des choix politiques seront faits pour orienter dans la mesure du possible les informations à disposition. En même temps, leur diffusion, à terme inévitable, conduira peut-être à améliorer certaines faiblesses. Il est vrai que cette visibilité peut inquiéter et poser des problèmes en termes

d'aménagement. Un représentant de la ville de Colmar craignait (lors de la réunion organisée par ePSI et la FING à Rennes) que la visibilité plus grande des territoires renforce ceux qui ont déjà des ressources plus importantes, c'est-à-dire les grandes agglomérations. Ce problème se pose de façon générale (cf. Aménagement).

Tourisme. A l'évidence, un secteur où l'action des Régions et des autres collectivités en termes de diffusion d'information et d'attractivité est déjà forte. Elle a conduit à des structures coopératives qui mutualisent l'information (cf. Tourisme).

Transports. Une information plus large, adaptée aux différents publics, permet certainement un meilleur usage et une meilleure coordination, donc des gains pour tous, y compris les opérateurs, ce qui a été compris par Kéolis à Rennes, mais pas par tous les acteurs. De nombreux problèmes se posent, propres au domaine, en raison du grand nombre d'intervenants, pour obtenir une vraie synergie (cf. Transport).

Formation, apprentissage. Un besoin constant des entreprises et des opportunités plus grandes pour les élèves, étudiants en recherche de stages et d'apprentissage : le secteur est complexe et l'amélioration de la mise en relation peut résulter d'une diffusion plus large, dans et hors de la Région.

Presse et information régionale. Ces acteurs peuvent par définition bénéficier de la mise à disposition et de la réutilisation, tant comme opérateurs que pour leur diffusion propre.

#### 4.2.3 Vision large et de long terme

Les gains doivent s'envisager dans la longue période, et en prenant en compte des impacts très indirects. L'exemple le plus net est peut-être ici donné par le transport – que ce soit les opérations urbaines comme Rennes ou la stratégie du département de l'Isère. Au delà même de l'impact sur les

□ dépenses budgétaires en matière de transport (cf. Transports, l'exemple de l'Isère), les programmes facilitent le recours aux transports en commun et à l'auto-partage par l'information que les nouveaux services donnent aux usagers et ils limitent d'autant le recours à l'automobile individuelle et les coûts qu'elle engendre comme le budget transports de la collectivité. Il s'agit bien là d'impact très indirect, correspondant à un choix politique général.

### 4.3 Economie de l'information : vue générale

De nombreuses estimations circulent sur le marché que devrait représenter la réutilisation des données publiques. Le chiffre de 27 milliards d'euros par an dans l'Union Européenne est avancé par la Commission [cf. site PSI de la Commission – liste de sites – et étude MEPSIR Réf.]. Dans le monde, une étude de l'OCDE, non publiée à l'été 2011, annoncerait un CA de 140 Mds \$. Ces gains correspondent à l'estimation de l'activité du secteur des sociétés d'information qui rediffusent les données publiques.

D'autres études ont tenté de cerner les impacts. Elles sont indiquées en fin du rapport et accessibles sur le site OTeN ou par redirection.

*L'une d'elles, très récente, estime l'impact direct en Espagne (étude Aporta Infomediary de juin 2011 – Cf. Document). Elle peut être citée en exemple de ce type d'approche.*

*Note de Denis Berthault – Lexis Nexis – GFII sur cette étude*

*Une étude officielle menée en Espagne évalue le poids économique des données publiques à un chiffre d'affaires annuel de 650 millions d'euros. Conduite de mars à mai 2011, cette étude s'appuie sur les 230 entreprises identifiées réutilisant des données publiques. Le CA directement relié aux données publiques s'établirait entre 550 et 650 million d'Euros par an, ce qui constituerait 35% à 40% percent du revenu total des dites entreprises. 5000 à 5500 personnes ont un job qui dépend directement de la réutilisation de données publiques.*

Pour l'essentiel, ces études, comme celle réalisée pour la Commission (MEPSIR) considèrent le chiffre d'affaires des sociétés qui réutilisent aujourd'hui des données publiques, c'est-à-dire les sociétés traditionnelles de l'information électronique et des banques de données travaillant dans l'information financière, presse, juridique, économique, scientifique et technique, brevets [cf. une analyse du secteur dans l'étude Serda - GFII de 2010 "Le marché français de l'information électronique professionnelle"]. Ces sociétés réutilisent principalement les données géographiques, les données juridiques et réglementaires, les informations démographiques et économiques. En France, l'étude Serda - GFII estimait le marché de l'information électronique professionnelle à 1,57 milliard d'euros en 2008, en croissance de +3,4% (PIB : 1933Mds€, en croissance de 2,5% sur 2007).

De telles estimations de l'impact de la réutilisation des données publiques sont évidemment à considérer avec la plus grande prudence, et les auteurs l'admettent. Il faut d'ailleurs rappeler que de nombreux services d'information sont gratuits et financés par la publicité. Jusqu'ici, cela n'empêche pas de les compter dans l'activité économique (c'est d'ailleurs pourquoi il est difficile de distinguer les services commerciaux et non commerciaux). Cependant, mieux vaudrait aussi rappeler que le plus grand fournisseur d'information dans le monde est Google, uniquement financé par la publicité, et en particulier par la publicité locale. Le marché publicitaire n'est pas extensible à l'infini, surtout en période de crise, comme le ressentent aujourd'hui les médias. Dès lors, s'il y a accroissement de l'activité, il est difficile de la localiser - c'est encore toute l'incertitude du monde Internet, qui est encore plus durement concurrentiel que le monde physique.

Enfin, s'il est vrai que le monde "traditionnel" des services d'information forme son offre à partir d'une grande quantité de données publiques, sa situation ne reflète pas ce qu'est le potentiel du développement de l'open data dans la phase qui s'ouvre actuellement. En effet, ce nouveau développement est principalement le fait, pour l'instant, de jeunes pousses exploitant

des données de proximité et de maillage fin du territoire, en direction des mobiles du grand public. Il y a déjà de très nombreux utilisateurs mais les principaux bénéficiaires sont les opérateurs, car - et chacun le reconnaît - le modèle économique n'est pas identifié pour ce type de services (comme l'a montré Proximités sur le Web).

## Gestion : coûts directs et avantages pour la collectivité productrice

Le débat sur la réutilisation, et en particulier sur les tarifs (cf Chapitre Réutilisation) ne peut être isolé de la conjoncture économique. Les services publics sont tous confrontés à des difficultés, qui exigent des choix politiques lourds. Il est peu probable que la diffusion des données publiques, si elle pèse trop sur les budgets, soit entièrement supportée par l'impôt, et ceci même si les avantages économiques ou de gestion induits sont supposés grands : les budgets sont annuels et les dépenses pèsent à court terme.

Il n'est que de voir à ce sujet que qu'explique Séverin Naudet, directeur de la Mission Etalab, en ce qui concerne les données culturelles - qui, en France, ne sont pas exclues systématiquement de la réutilisation, et même de la réutilisation gratuite.

« Certains modèles existants, lorsqu'ils se fondent sur de nécessaires équilibres, doivent être préservés. Le service public de la culture - auquel nous tenons tous et qui est une spécificité française dont nous pouvons être fiers - a un coût dont les ressources propres sont une part importante du financement »

## 4.4 Charges

La mise à disposition des données publiques pour réutilisation va inévitablement entraîner des coûts. L'énumération des dépenses possibles peut inquiéter mais elles demeureront au total limitées. Mieux vaut d'entrée les prévoir et surtout surveiller leur évolution.

### 4.4.1 Coûts directs

#### ■ *Charge accrue pour les personnels*

Cette charge correspond aux différents coûts qui sont indiqués dans les paragraphes suivants, tant pour la gestion du contenu des données que pour l'informatique.

Elle est au demeurant symétrique du gain en qualité qui sera – voir ci-après les avantages – une conséquence de la diffusion.

Un coût initial est la formation des agents qui seront en charge des données publiques dans les différents services et qui devront suivre les évolutions des réglementations en même temps que des flux internes.

Déjà, lors de la réunion de formation organisée à Bordeaux par l'APIE et la CADA en avril, les nombreuses Personnes responsables de l'accès aux documents administratifs (PRADA) s'inquiétaient de devoir valider des demandes croissantes et d'avoir à suivre demain les réutilisations.

#### ■ *Données*

Comme cela est précisé dans le chapitre consacré aux données, les coûts de la mise à disposition et de la réutilisation ne peuvent être isolés d'une démarche de meilleure gestion des données, à peu près partout indispensable – notamment en ce qui concerne la sécurité, comme le soulignent les

responsables de ce domaine [cf. *Chapitres Données et Techniques*]. La réutilisation impose seulement de la mettre en œuvre.

La chaîne de production puis d'édition et de mise à disposition des données comprend : numérisation (lorsque les données ou les documents sont sur support papier, ce qui est encore fréquent s'agissant de formulaires administratifs entre autres), indexation et production de métadonnées. Elle comprend ensuite éventuellement la mise aux normes pour passer de formats et métadonnées internes à des informations partageables.

La réutilisation impose en outre une analyse et des traitements supplémentaires pour veiller à respecter les obligations légales, ce qui peut être une tâche lourde pour certains jeux de données : protection de la vie privée, protection de la confidentialité pour les organisations et sociétés, recherche des droits de propriété intellectuelle.

Enfin, et quoi qu'en aient les partisans d'une totale irresponsabilité (de fait légale) de la collectivité par rapport au réutilisateur, il sera nécessaire d'effectuer un suivi et des contrôles sur les informations les plus importantes et les plus sensibles. Ici encore, ce n'est pas au fond la réutilisation des données publiques qui est seule en cause. Dans un monde où l'information tient une grande place, un suivi des données présentes sur Internet et pouvant handicaper un territoire devra être effectué. Ce problème est celui auxquels sont confrontés aujourd'hui aussi bien la police que les sociétés commerciales, pour détecter des sites illégaux ou des campagnes de dénigrement, entre autres.

#### ■ *Informatique et technique*

Comme pour le contenu des données, une part importante de la dépense informatique nécessitée par la réutilisation et le partage est en fait liée à une meilleure gestion des données. C'est le cas pour la constitution de bases centrales.



En revanche, la diffusion externe de l'information oblige à des dépenses spécifiques (cf. *Chapitre Techniques*) : constitution d'une plate-forme accessible en Web services, constitution du Répertoire (cf. *paragraphes correspondants*) ou participation à des plates-formes mutualisées. Il y aura aussi initialement des charges de mise à jour pour des données qui n'étaient pas temps réel (horaires...). Il y a enfin les traitements spéciaux nécessités par les obligations légales, et en premier lieu l'anonymisation, lorsqu'elle est possible, de jeux de données incluant des informations personnelles.

Par ailleurs, outre les coûts de maintenance et d'évolution inhérents à l'informatique et notamment à l'évolution des formats (coûts supportés par la collectivité ou un sous-traitant), il y aura une charge accrue pour les serveurs, si la réutilisation est un succès. C'est d'ailleurs pourquoi plusieurs des réalisations actuelles (Ville de Paris entre autres) ont prévu une indemnisation liée à cette charge.

#### 4.4.2 Des conséquences à terme

Progressivement, la circulation des données publiques et plus largement une politique régionale de développement de fourniture de données va s'auto-accélerer. Une fois le principe acquis, des demandes seront formulées, par la population, par les réutilisateurs, par d'autres entreprises, par les partenaires de la collectivité dans des opérations communes.

Pour chaque jeu de données en outre, l'utilisation, dans des services de la collectivité elle-même ou des réutilisateurs, va provoquer des demandes complémentaires mais aussi des réclamations, pour signaler des erreurs, voire un mésusage.

Par ailleurs, comme cela est indiqué et recommandé (*Chapitre Réutilisation*), il faut mettre en place un suivi de la relation avec les réutilisateurs et répondre à leurs demandes.

## 4.5 Avantages internes

Souvent difficilement mesurables, ils seront en général importants.

### 4.5.1 Meilleure gestion

L'ouverture des données publiques a un impact sur la gestion de la collectivité.

- Une meilleure gestion des données implique qu'il y ait plus d'échanges entre des services qui ont parfois fonctionné en silos. Plus généralement, le bénéfice interne de la cohérence accrue des données et de leur facilité d'accès peut être grand. Par ailleurs, les contrats qui devront être passés avec les réutilisateurs vont entraîner une plus grande qualité, des mises à jour plus régulières.
- En opposition aux accroissements de charge envisagés pour les personnels en raison de la diffusion d'information, il peut y avoir des allègements parce que certaines informations seront plus aisément accessibles – par exemple il peut y avoir moins de demandes de consultation des plans d'urbanisme.
- L'ouverture des informations sur les marchés publics permet une meilleure connaissance et analyse. En fait, cette ouverture est surtout intéressante collectivement : les entreprises peuvent communiquer de nombreuses informations et indicateurs sur les marchés (cf. *Transpacité et Euroalert*), voire sur les propres marchés de la collectivité cliente.

### 4.5.2 Eléments d'une politique de maîtrise budgétaire des dépenses

- Intégrée dans une stratégie, la politique de partage des données avec d'autres acteurs, y compris privés, peut être un élément de stabilisation

budgétaire. L'exemple en est donné par le Département de l'Isère, qui entend développer une meilleure gestion des transports (cf. *Transport*).

### 4.5.3 La transparence et ses effets

- Les partisans de l'opendata comme moyen d'assurer la transparence de l'action et du fonctionnement de la collectivité soulignent que cela permet à la population de contrôler les dépenses de la collectivité. Cependant, des responsables plus officiels partagent cette opinion. Ainsi, George Osborne, Chancelier de l'Echiquier britannique, déclarait en mai 2011 (trad. auteur):

*« Notre ambition est de devenir le premier acteur mondial dans le domaine de l'open data, et d'accélérer la révolution de la responsabilité induite par le développement d'internet. En effet, soyons clairs, les bénéfices en sont immenses. Pas uniquement du point de vue du repérage et gaspillage et de la baisse des dépenses, bien que les effets de la transparence dans ce domaine se fassent déjà ressentir pour les finances publiques. Mais, par dessus tout parce que les bénéfices économiques et sociaux de l'open data sont encore plus grands. »*

Ainsi, l'ouverture soumet les gaspillages et les dysfonctionnements à la vigilance des citoyens et de leurs associations. Il ne s'agit pas ici de prendre position sur cette vision de «l'open government» mais de signaler cet effet, qui peut au demeurant inquiéter les collectivités et peut aussi certainement conduire parfois à des excès.

## 4.6 Gains

### 4.6.1 Recettes : les redevances des réutilisateurs

*Cet aspect est traité dans le Chapitre spécial sur la réutilisation.*

Quoi qu'il en soit, il faut rappeler que l'activité ne doit certainement pas permettre de dégager des bénéfices, par rapport aux coûts engendrés. Il est donc erroné de parler, comme on le fait parfois, de "retour normal sur investissement". En effet, pour une entreprise, un tel retour inclut un bénéfice net. Ce n'est pas le cas pour les acteurs publics, à l'exception des établissements publics à caractère industriel et commercial, des services dont la mission est la diffusion de données (INSEE, IGN, Météo...), et des activités qui ne peuvent subsister sans ressources privées importantes (cf. *ci-dessus position de Séverin Naudet sur les données du service public de la culture*).

*Globalement, comme le soulignait le rapport Commerce électronique de Francis Lorentz en 1998, il ne faut pas que les recettes de diffusion soient le critère de décision s'agissant de la diffusion des données.*

### 4.6.2 Gains virtuels

Des calculs ont été présentés, qui évaluaient le coût de développement des applications par les réutilisateurs et en déduisaient des économies virtuelles. Ainsi, pour Rennes, après le concours pour la création d'applications, on a pu écrire que 17 applications avaient été créées, qui auraient coûté en moyenne 40 000 € à développer<sup>23</sup>.

Lors de l'Opendata Garage de juillet, Simon Chignard, animateur de Wiki Rennes, a souligné que ce type de raisonnement est un discours de communication (interne et externe). Il néglige les redondances entre applications et surtout rien ne permet d'affirmer, à court terme, lesquelles répondent à un besoin, lesquelles auraient été développées par la collectivité sans l'ouverture. La valeur obtenue est donc au mieux un maximum et ne

<sup>23</sup> Xavier Crouan a indiqué les dépenses directes : 50000€ pour le concours d'applications, 30000€ pour le site et la communication, 20000€ pour l'évaluation, co-financés avec les partenaires – Région, CdC, Kéolis, Ouest France, Rennes Métropole n'ayant déboursé que 20000 € [Site LiberTic]

traduit pas la valeur créée ni les dépenses économisées. En fait, ce calcul met en avant une économie virtuelle et non les bénéfices collectifs.

Cette approche est même dangereuse : le choix entre faire soi-même ou attendre que des acteurs privés fassent doit être guidé par une politique des services. La collectivité ne peut, avec un argument d'économie, abandonner systématiquement les services d'information.

#### 4.6.3 L'Etat mesure et s'interroge

En période de difficultés budgétaires, la vision annuelle tend à grossir l'intérêt de la redevance. S'agissant de la situation française, les redevances versées aux organismes en charge de la diffusion de l'information au sein de l'Etat (DILA, IGN, INPI, INSEE, Météo notamment) échappent aux textes limitant en pratique la redevance, comme échappent aussi les licences existantes des administrations et les établissements publics industriels et commerciaux.

Dans ces conditions, les montants ne sont pas négligeables. François Baroin, ministre du budget, s'en félicitait en 2010.

#### Redevances vers les administrations d'Etat (F. Baroin – 2010)

S'agissant du marché de l'information en France, il a été estimé à près de 3,7 milliards d'euros en 2007 (dont 1,6 milliard d'euros pour les données électroniques), dont 60 %, soit 2,2 milliards d'euros, sont issues de données publiques à la base (informations légales, financières, économiques, etc.).

En effet, les licences déjà octroyées qui concernent, à l'heure actuelle, essentiellement les secteurs traditionnels de l'information géographique, météorologique, économique et financière, rapportent chaque année environ 50 millions d'euros à l'État<sup>24</sup>. Elles pourraient être étendues à d'autres données comme les informations cadastrales

<sup>24</sup> Budget 2009 - Dépenses: 292.9 milliards d'euros – Recettes : 196.3 milliards d'euros

ou encore la liste des services publics de proximité ou leurs horaires d'ouverture...

Le ministre envisageait alors une valorisation plus forte du patrimoine immatériel, et l'Agence du patrimoine immatériel de l'Etat s'organisait pour faire croître le flux. L'orientation a été au moins discutée et nettement infléchie en 2011 avec la création de la mission ETALAB et l'accent mis sur la gratuité. La prise en compte des gains de gestion et surtout de l'impact économique de la réutilisation sont de plus en plus mis en avant.

## Entreprises

### 4.7 Les acteurs

Le secteur de l'open data concerne différents types d'industrie, de sociétés de services et conseils en informatique et Internet, et de PME du domaine des « TIC innovantes » :

- Opérateurs des télécoms mais aussi grands acteurs d'Internet
- Prestataires informatiques et télécom auprès des administrations
- Réutilisateurs créant des services, en général d'information (et développeurs d'applications pour ces réutilisateurs, le cas échéant)

#### 4.7.1 Infrastructure : internet et mobilité - téléphonie mobile

En premier lieu, les grands opérateurs de télécommunication et les grands acteurs du matériel et des infrastructures en téléphonie mobile sont motivés par l'open data et ont une présence bien visible. A titre d'exemple, lors de la séance d'ouverture de « Mobile Monday » promue par la Mairie de Paris le 6 juin, un opérateur national de télécommunication (Orange) et un équipementier en téléphonie mobile (Nokia) avaient une forte visibilité et étaient associés à la promotion de l'évènement.

L'intérêt pratique de ces sociétés est évident, et le rationnel est simple : utilisation des infrastructures de télécommunication pour accès à l'information en mobilité sur matériel de type smartphone, tablette, iPad, iPhone... Personne ne doute que l'open data représente une excellente opportunité de génération de trafic sur Internet.

Ces grandes sociétés ont créé des départements dont la mission consiste à favoriser la mise à disposition et l'utilisation des données publiques, en contractualisant et en supportant les acteurs proposant des services à valeur ajoutée en ce domaine – et en particulier les PME TIC innovantes.

#### 4.7.2 Développement et conseils : domaine de l'informatique et des SSII

Les sociétés de conseil en organisation, qu'elles soient grandes ou petites, sont concernées à au moins deux titres :

- L'open data leur est nécessaire pour la réalisation quotidienne de leur activité de missions d'étude et conseil.
- L'open data représente une nouvelle source intéressante de revenus :

Dans l'activité d'assistance et conseil à maîtrise d'ouvrage, pour aider l'administration en général et les collectivités territoriales en particulier à structurer leur approche dans le domaine de l'open data.

Dans l'activité de développement informatique et technique de portail d'accès ou de plateformes de service valorisant les données publiques.

Dans ce domaine, des sociétés spécialisées apparaissent, qui souvent vont s'allier aux acteurs historiques pour développer de nouvelles fonctionnalités. C'est ainsi que Microsoft arrive un peu tard avec sa plate-forme de gestion et mise à disposition «Open Government Data Initiative» (<http://ogdisdk.cloudapp.net/>), face à une start-up spécialisée, Socrata, qui est déjà

implantée jusqu'au Kenya (<http://www.socrata.com/>).

### Des exemples de Sociétés de services agissant dans le secteur des collectivités et des services publics

**Berger Levraut** ([www.berger-levraut.fr](http://www.berger-levraut.fr)) - entretien avec Guy Beudet, directeur de la stratégie)

Les données publiques constituent un domaine stratégique pour Berger Levraut, qui centre depuis longtemps son domaine d'activité sur les données de l'Etat et la dématérialisation des formulaires administratifs en particulier.

Axe : des comparaisons d'indicateurs de gestion (projet Performance)

La société démarche, recherche les données de qualité.

Le modèle : une partie gratuite pour les gestionnaires communaux et une partie privée - mon tableau de bord, mes indicateurs, lien avec mes données locales. On peut imaginer que la partie publique soit complètement publique.

Clients : les élus, pour présenter leurs données

Ici, c'est le prestataire qui engage la collectivité à libérer des données. Si d'autres interviennent, la société compte sur sa connaissance du domaine.

G. Beudet insiste sur le besoin de régulation et sur l'exigence de qualité des données.

A la différence d'autres sociétés, Berger-Levraut se positionne bien sur un service de fourniture d'information, mais à destination de l'ensemble des collectivités qui par ailleurs produisent les données.

**Logica** ([www.logica.fr](http://www.logica.fr)) - entretien avec Stéphane Batiot, directeur associé en charge du secteur public

Logica travaille sur de grands projets, comme la refonte de gestion du système d'archives nationales ou le Web sémantique pour le Centre Pompidou. La société les regarde du point de vue stratégie et conseil : mise en place de plateformes pour mise à disposition de données.

S. Batiot considère que, pour la valorisation, on doit envisager l'agrégation de données.

Il insiste sur le fait que l'administration doit renoncer à contrôler toute la chaîne et qu'il faut rester proche de l'utilisateur – ce que les sociétés commerciales peuvent souvent mieux faire.

**Sopra** ([www.sopragroup.com](http://www.sopragroup.com)) - entretien avec Jean-Jacques Barbier, responsable des projets « Open » - Open source et Open data.

Sopra travaille notamment avec la Mairie de Paris et développe là des outils qui seront réutilisables.

Les fournisseurs de logiciels et intégrateurs doivent maintenant concevoir leurs offres en fonction de l'importance croissante de la mise à disposition et des interfaces. Les sociétés devront participer à la création d'un écosystème, avec des outils de remontée d'information. Un axe de développement important : les données temps réel – et pas seulement pour les transports (ex: disponibilité d'un ouvrage en bibliothèque).

Le positionnement comme prestataire, aidant la Mairie à procéduraliser la fourniture de données publiques, est contradictoire avec une activité de production de services.

#### 4.7.3 Innovation «open data»

Un fait indéniable : l'open data suscite un fort intérêt dans le domaine d'Internet et du Web2 et pour l'accès aux données en mobilité. Il est frappant de voir le nombre des participants aux «Mobile Monday» organisés avec le soutien de la Mairie de Paris: on rencontre les porteurs de nombreuses PME innovantes ou «start-up», de jeunes créateurs d'entreprise...

Ces entrepreneurs proposent des fonctionnalités dans le domaine du jeu, du commerce, de la culture, en général du service en mobilité: identification de ressources de proximité accessibles par tout matériel portable, depuis le smartphone jusqu'à l'iPad ou le PC. Il s'agit indéniablement d'une opportunité de création d'emplois, de création de valeur par apparition de nouveaux services. Le consensus existe sur le fait que la France peut et doit jouer un rôle dans ce domaine et y favoriser la création d'entreprises innovantes. Cet objectif suppose une stratégie pérenne, et notamment que l'administration garantisse la stabilité dans la fourniture des données.

**Furet Company** ([www.furetcompany.com](http://www.furetcompany.com)) - entretien avec Cedric Levret, fondateur

Toute jeune entreprise, créée en 2011 – deux personnes et des stagiaires dans le domaine des jeux interactifs en mobilité valorisant le patrimoine. Les clients sont les offices de tourisme, les mairies, des parcs à thèmes, les Comités départementaux de tourisme, la RATP... Ils fournissent les données, les informations du patrimoine à valoriser. La société va ainsi combiner des données sur les musées, des fonds d'image, etc. avec la géolocalisation grâce aux fonds et outils Google.

Le modèle est ici la valorisation de données, qui ne sont en général pas ouvertes à d'autres. L'initiative vient de l'acteur public qui recherche de nouveaux canaux de valorisation.

Le modèle économique : vente de licence, hébergement, maintenance, conseil et intégration du site. Actuellement, pas de jeux payants.

C. Levret pointe deux problèmes : la difficulté de prise de décision pour les organismes de mutualisation, qui doivent passer par leur Conseil d'administration, présidé par un élu ; l'impossibilité de travailler dans le domaine de la culture, les conservateurs ne voulant pas de ce type de services (hormis musées pour les jeunes et musées des sciences).

Ici, il s'agit bien d'une politique sur les données : la collectivité confie des données brutes et elle laisse un acteur externe proposer un service innovant, qui va toucher un public particulier.

#### 4.7.4 Alliance de jeunes sociétés, d'une grande société et de la recherche : Data Publica

Le cas de Data Publica est particulier, mais il illustre une rencontre – il y en aura sans doute d'autres – entre différents types de sociétés, dont un poids lourd.

Data Publica ([www.data-publica.org](http://www.data-publica.org)) est un projet très ambitieux de répertoire mais aussi de place de marché pour toutes sortes de données, d'abord actuellement les données publiques mais à terme celles d'entreprises ou d'autres organismes. (cf. dans cette étude «Répertoires» et «Techniques»). Créée en janvier 2010, Data Publica réunissait Araok, société de conseil en commerce électronique, Nexedi, l'éditeur de l'ERP libre ERP5 et Talend, spécialiste de la gestion de fonds de données open source. Le groupe a été rejoint par l'INRIA et ATOS. Le modèle économique est ici celui de la place de marché : Data Publica offre des services de recherche et d'agrégation de données et gèrera des transactions. Data Publica a reçu plusieurs soutiens des programmes publics.

#### 4.7.5 Industrie de l'information

Le développement de nouvelles formes de réutilisation ne doit pas faire oublier l'existence d'un secteur de l'industrie de l'information déjà très actif et qui continue lui aussi à s'étendre sur des marchés en émergence : clientèle de particuliers et de très petites entreprises et, lui aussi, utilisations en mobilité. Ces acteurs, que l'on retrouve dans le Groupement français de l'industrie de l'information, où ils rencontrent les grands producteurs et éditeurs publics, ont été évoqués ci-dessus (cf. § «Economie de l'information»).

Ces acteurs se connaissent de longue date et maîtrisent assez bien leurs relations : après des conflits souvent longs et vifs, le marché a atteint la maturité et les divergences se règlent par la négociation. Pour ce marché, le cadre juridique semble suffisamment robuste, et les industriels de l'information, s'ils travaillent avec leurs partenaires et l'Etat à l'améliorer, ne réclament pas la levée de toutes les barrières. Cependant, ils ne sont pas contre la gratuité pour la réutilisation, même si elle permet l'entrée de nouveaux intervenants – en fait, ils travaillent surtout avec des organismes publics qui échappent actuellement à la pression vers la gratuité. Leur principal souhait est que, lorsque de nouveaux domaines s'ouvrent, les licences permettent la réutilisation commerciale.

En même temps, de nouveaux venus apparaissent, profitant – ou cherchant à profiter – de la mise à disposition de gisements de données publiques...

**Généalogie.com** (<http://www.genealogie.com/>) Ce service a été créé par NotreFamille.com.

Toussaint Roze, Président du groupe, le présente ainsi dans le Journal du Net du 6 juillet 2011 : "... projet innovant de valorisation des archives d'état-civil de plus de 120 ans. L'objectif : permettre à tous les Français de retrouver facilement leurs origines. Pour servir cet objectif, Genealogie.com doit agréger l'ensemble des données disponibles au niveau départemental

et transcrire de façon systématique ces documents manuscrits anciens afin de mettre en place des moteurs de recherche patronymiques.

(...)

*Même si les départements ont développé des programmes de numérisation, il manque les outils qui permettent de s'orienter simplement dans ces fonds et de les décrypter. C'est dans cet intervalle que des initiatives privées peuvent s'inscrire sans dénaturer l'objectif premier de toute politique culturelle : apporter à tous la connaissance d'un patrimoine et les moyens de l'apprécier."*

Ce projet a suscité de très fortes résistances de la part des archivistes des Conseils généraux, qui ont été "traumatisés". Le projet a illustré, de façon regrettable et conflictuelle, les difficultés de l'ouverture : refus des CG, incertitude sur le statut des données car elles sont gérées par un service qui dépend du ministère de la Culture, recours devant la CADA (qui a confirmé l'accessibilité et la possibilité de réutilisation), plainte à la CNIL par les adversaires etc. La question évolue lentement; l'Association des Archivistes de France a organisé un débat et l'heure est maintenant à la fixation des redevances.

**Ancestry.com**, société américaine concurrente, est associée à Genealogy.com dans cette tentative d'ouverture de données publiques.

**Transpacités** (<http://www.transpacites.fr/>)

*(Notes issues d'un entretien)*

Venu du monde des opérateurs privés dans les services publics locaux Plastic Omnium puis Sodexo puis SITA, Didier Gorrion a fondé Transpacités en 2010.

Le postulat ou l'intuition de départ : Alors que les grands groupes de services utilisent une approche marketing au sens large pour les entreprises, les relations avec les collectivités sont fondées sur le

relationnel local. De leur côté, les collectivités ont été longtemps dans une logique de copier-coller – on reprend l'existant ou on appelle des collègues.

Hypothèse : progressivement le marketing va arriver. ...Transpacités développe un moteur et une base de données sur les collectivités, sur les services publics locaux et se positionne en interface entre les collectivités, les opérateurs privés et, pourquoi pas, les usagers.

La base de données porte sur les 36 000 communes, les intercommunalités etc. Elle indique qui gère quoi – ( commune, collectivité de communes, Syndicat mixte..). Le socle est alimenté par les données du ministère de l'intérieur - en accès libre sur Internet. Elles sont récupérées par Transpacités; actuellement sans contrat.

Ces données peuvent intéresser les opérateurs privés. La concurrence arrive, des collectivités reprennent des services en régie (cf. Paris), ce qui est grave pour le prestataire. Pour chaque marché, la collectivité va savoir quels sont les marchés des communes pour tel ou tel offreur, les échéances etc. Tout ce qui est utile pour la négociation. Les services aident les groupes à préparer leur action commerciale – service Ciblation. Transpacités associe des informations issues d'enquêtes auprès des habitants et de panels de décideurs.

Le modèle économique est validé.

**Euroalert.net** se positionne sur un créneau du même domaine : information sur les appels d'offre en Europe – alertes, analyse etc. (cf. ch.2)

#### 4.8 Quelques pistes de réflexion sous l'angle industriel

Les entretiens et les rencontres avec des industriels font apparaître un certain nombre de traits forts, et aucune contradiction de principe entre les

approches ou les visions, qui autorisent à avancer plusieurs propositions.

#### 4.8.1 Du modèle économique

Tous les entretiens que les acteurs du terrain, depuis le «partner» d'une des plus puissantes SSII françaises jusqu'au jeune créateur de start-up, «geek» frais émoulu d'une grande école, convergent vers quelque chose qui est peut-être une évidence: on rencontre un fort scepticisme sur la valorisation de la donnée publique par le paiement à la consultation ou au service apporté qu'en ferait le grand public. Comme le faisait remarquer un chef de service d'une SSII, prestataire de la Mairie de Paris:

*« De manière générale je ne discerne pas de possibilité évidente de proposer des services payants en valorisation des données publiques. D'autant plus qu'il existe une forte contradiction entre la vision du donneur d'ordre (gratuité de mise à disposition) et la position qu'aurait la dite société d'exploitation du service: faire payer une information dont la source est gratuite; c'est-à-dire faire payer la valeur ajoutée à l'utilisateur? »*

Il faut noter que cette culture industrielle n'est pas celle des éditeurs d'information professionnelle, qui font payer la valeur ajoutée à l'utilisateur, même si, en général, ils versent des redevances pour la mise à disposition des données. Les applications développées par les collectivités et les start-ups sont d'abord destinées au grand public et non, pour le moment, aux PME du territoire par exemple.

Ceci posé, on constate que les grandes SSII ont décelé un nouveau secteur de développement informatique et de conseil pour:

- Les études confiées par l'administration (études préalables, assistance à maîtrise d'ouvrage, conseils sur le rassemblement des données, sur les besoins ressentis...)

- la création de portails ou de sites Internet, le développement d'outils coopératifs pour la mise en ligne directe ou la mise à disposition.

Par ailleurs, les sociétés d'études et conseil traditionnelles (hors informatique), travaillant pour le compte de l'administration, savent coopérer avec elle pour identifier et avoir accès à toutes données publiques dont l'utilisation est nécessaire au cours de la réalisation de la mission.

Enfin, pour les start-up innovantes, le payeur est un intermédiaire: des commerçants pour les applications liées au commerce de proximité; des municipalités ou des organismes associatifs, pour la promotion du territoire, pour l'environnement; des musées, des associations, pour l'accès à la culture...

Bien sûr, les start-up innovantes s'appuient sur une infrastructure Internet qui permettrait le paiement à la connexion pour le grand public, mais ce mode de paiement ne semble pas correspondre, on l'a vu, à l'approche de marché open data pour les segments grand public.

#### 4.8.2 Du rôle et des responsabilités de l'administration, entre données brutes et création de services

On a vu que dans les cas de missions d'étude ou déploiement de services assurés par les SSII ou les sociétés de conseil, l'administration elle-même assure le financement des missions, des développements, du fonctionnement des sites (qui dépendent alors directement d'elle).

Plusieurs interlocuteurs des collectivités territoriales avaient signalé cette ambiguïté ou anomalie par définition, l'administration a essentiellement un rôle de collecteur et de mise à disposition des données brutes. En pratique, elle va beaucoup plus loin, quelle qu'en soit la raison mise en forme des données et service d'accessibilité avec valorisation pour le grand public (la réalisation des portails lourds étant souvent confiée à des grandes SSII



spécialisées). Ainsi, dans les faits, les administrations étendent souvent la définition du service public pour englober la fourniture d'information sur le service, pour les secteurs qui sont au centre de leur mission – formation, éducation, domaine social. Il est possible de rappeler qu'il s'agit là de décision politique mais que le risque est que les administrations prennent trop en main, avec des partenaires, la production de services dans des domaines qui ne sont pas au cœur de leur mission et où des prestataires privés – sociétés ou associations, pourraient assurer des services efficaces et diversifiés.

Ainsi, des visions sous-jacentes différentes apparaissent entre les responsables de l'ouverture des données dans les collectivités et les sociétés de conseil ou de développement de plates-formes. Pour les premiers, la création de nouveaux services exige d'identifier les promoteurs, souvent les start-ups du privé. Pour les seconds, le modèle économique principal est celui dans lequel les collectivités assurent ou assureront également un rôle de maîtrise d'œuvre pour permettre l'accès aux données et à des services.

Ce point de la relation entre données brutes et fourniture de services a déjà été abordé dans le présent travail (*Chapitre «données»*). Certaines conséquences seront discutées dans les conclusions du présent chapitre.

#### 4.8.3 Des difficultés et complexités de la relation avec l'administration

Un autre point commun dans les points de vue des sociétés est une certaine méfiance par rapport à l'administration, même s'ils entretiennent de bonnes relations et s'en félicitent souvent. Ils déplorent :

- la complexité des relations : les difficultés tiennent notamment à la diversité des personnes qu'il faut rencontrer pour expliquer un projet puis mettre en place une relation claire et pérenne. Dans certains cas, il faut des aller-retour entre différents services et entre ceux-ci et les élus. Le problème se complique encore si plusieurs acteurs sont en cause.

Ces étapes sont lourdes pour de petites structures, les grandes étant en outre plus expérimentées dans le domaine.

- la méconnaissance par l'administration des besoins et comportements des usagers et en général des réalités commerciales.

#### 4.8.4 Des incertitudes juridiques

Les entreprises se plaignent des incertitudes qui subsistent, et du fait qu'il n'y ait pas une réglementation précise et une régulation. Le fait de ne pas savoir clairement ce qui est interdit présente des risques.

La CNIL inquiète aussi. En effet, elle peut être saisie par des citoyens et des associations et cela peut conduire à l'arrêt d'un service même s'il est finalement valide. L'évolution de la réglementation est aussi une source d'incertitude : ainsi certains craignent qu'une application puisse être interdite en raison d'identification de la présence en mobilité, par exemple.

Les événements de juillet 2011 – conflit entre la RATP et ChekMyMetro - montrent aussi que des surprises peuvent venir du droit d'auteur (même si la RATP n'est actuellement pas soumise au droit relatif aux données publiques et même si ce n'était qu'un prétexte (*cf. chapitres sur la problématique générale et sur les contraintes juridiques*)).

En d'autres termes, l'environnement réglementaire reste encore trop flou pour le développement et la prise de risque financière. La définition d'un cadre réglementaire est en tout état de cause une condition pour le développement économique.

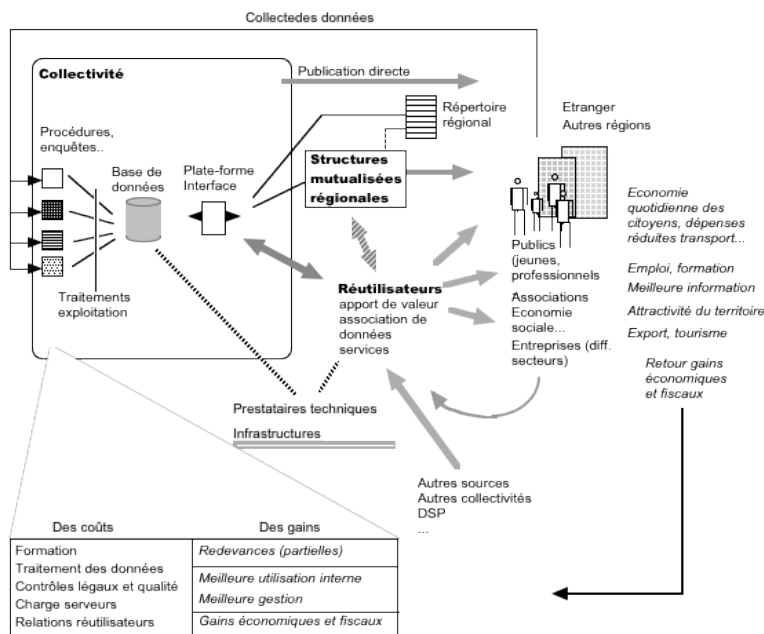
## Quelques conclusions et tendances

### 4.9 La mise en place d'un écosystème

La circulation des données ne naît pas en quelques mois ou années. Elle existe déjà, mais de façon limitée, par des circuits multiples, avec des modes de contrôle marqués par des moyens et des outils souvent dépassés. L'informatisation et les réseaux la poussent à un autre niveau et conduiront progressivement à un écosystème, où se réorganiseront les différents acteurs et où se redéfinira, à un niveau plus fin, le partage entre les différentes modalités de financement et de régulation économique.

Le schéma suivant rappelle quelques uns des éléments de cet écosystème.

#### Ecosystème de la circulation de données régionales



### 4.10 Une démarche raisonnée

Il faut éviter autant le conservatisme que les illusions et les déceptions. L'ouverture des données se fera progressivement. L'organisation des partenariats avec d'autres acteurs pour développer un véritable écosystème des informations régionales, au sein d'un écosystème national et international, va prendre du temps.

Il faut donc tracer une voie de développement, suivre les coûts, les usages, les gains.

Le fait de commencer une mise à disposition constitue un engagement. Pour des données complexes, normalisation, anonymisation, gestion de bases lourdes représentent une charge. Il faut donc prendre garde à ne pas s'engager sur un domaine particulier sans avoir étudié les coûts et le financement (surtout du fonctionnement). En effet, une fois un service mis en place (et même garanti par contrat aux utilisateurs – associations citoyennes, sociétés commerciales etc.), il faudra le maintenir.

### 4.11 Un domaine qui se diversifiera et où il faut fixer des règles et décider de ce qui est ou sera un service public

Certains aspects de l'ouverture des données publiques sont communs à tous les types de données – ce sont principalement les questions juridiques qui règlent l'accessibilité, la publication et la réutilisation.

En revanche, il sera bientôt nécessaire de distinguer les domaines et les marchés car les mécanismes économiques diffèrent à l'évidence beaucoup entre, par exemple, le tourisme, la formation, la publication de données de gestion des acteurs publics. D'ailleurs, l'opposition est déjà apparue entre les données à forte valeur marchande et les autres, qui concernent plutôt la vie quotidienne et qui sont associées à des modèles économiques incluant la subvention, la quote part versée par le distributeur pour les applications mobiles, la publicité locale.

La différenciation va continuer de se faire – peut-être différemment en fonction des choix politiques – entre les informations dont la diffusion est considérée comme critique et contrôlée par la collectivité responsable et les autres, financées par d'autres sources. Dans le premier cas, les données publiques sont ou seront pour partie directement mises en ligne sur des sites ou des services – gratuits - par la collectivité elle-même. Dans le second, les données brutes seront réutilisables

Cette évolution modifiera sans doute le paysage qui se dessine actuellement. La différenciation entre service public gratuit essentiel, service public délégué et services commerciaux se modifie avec l'entrée dans la société de l'information. Certains services aujourd'hui encore un peu marginaux, sur le Web et les mobiles vont rapidement devenir critiques. Dès lors, de nouvelles régulations et délégations devront apparaître, ce qui ramènera l'acteur public dans les prises de décision et dans le financement, ces services ne pouvant être dépendants de la publicité ou de services annexes payants ("de type – l'information sur les horaires ou sur les hôpitaux est gratuite tant que N... apporte de la pub"). Dans de nombreux cas, il est donc probable que des applications vont revenir au modèle de financement public, et donc aux appels d'offre ou à la complète intégration dans l'activité de l'acteur public concerné. Des régulations spéciales devront intervenir sur ce type de marchés, car la tendance à la concession exclusive pour un suivi de l'application sera grande. Dès lors, des sociétés récupéreront des monopoles pratiques – qui peuvent aussi apparaître pour des structures de type mutualisé ou associatif. Ces sociétés ne seraient pas forcément locales – et si elles le sont, elles risquent de ne pas pouvoir exister sur le marché national et mondial, avec une fragmentation entre petits acteurs TIC sur les territoires, ce qui est tout aussi dangereux.

Le risque le plus grand est que la collectivité se déssaisisse du contrôle réel de données essentielles et de leur diffusion, en raison de l'évolution des techniques et des coûts des migrations. C'est pourquoi il est si important

de réintégrer la mise à disposition et la circulation des données dans la politique générale sur un domaine.<sup>25</sup>

#### 4.12 Les incertitudes du secteur, les contradictions et la création de valeur

Comme l'a exprimé un intervenant à "l'Open data Garage" de juillet, « *Des gens dont la mission est le développement économique hésitent à mettre les données parce que les entreprises vont gagner de l'argent avec ! Mais c'est ce qu'on veut!* » Quelque opinion qu'on en ait, il y a effectivement un flou dans les objectifs s'agissant du développement économique.

L'étape actuelle est marquée par l'entrée de nouveaux acteurs – les collectivités comme producteurs, les start-ups comme réutilisateurs. Il est impossible de savoir quelle place ce nouveau segment de marché va occuper. Le rapport de l'ambassade de France sur l'open data au Royaume Uni [cf. document] indique que, dans ce pays en avance : « *Sur les applications, le bilan est encore faible. Il y a finalement peu d'applications qui ont passé le stade du simple buzz et qui se sont imposées comme véritable service au quotidien pour les citoyens. La libération de données locales et à forte valeur ajoutée pourra redonner un nouvel élan à la communauté de développeurs. La mise à jour régulière des données, une meilleure granularité, leur mise à disposition dans des formats ouverts et réutilisables par les machines favoriseront l'innovation.* »

En fait, comme cela a été indiqué dans la première partie de ce chapitre, les effets économiques sont avant tout attendus dans la meilleure information des usagers et des entreprises. Pour l'essentiel, les applications de vie quotidienne ont des impacts en termes de meilleure qualité de vie, de transparence, d'économies induites (transport...) et ne se traduisent que peu en termes directement financiers, ce qui ne sera pas le cas si de nouveaux

<sup>25</sup> Le Chapitre «Techniques» rappelle que le risque est particulièrement fort avec les architectures du Nuage, où toutes les applications sont quelque part sur le Web.

■ marchés professionnels se développent – mais ils n’ont guère été envisagés jusqu’ici.

■ Par ailleurs, il demeure des obstacles non négligeables, que le mouvement actuel, qui est local et concentré sur un producteur en général, n’aide pas à lever. En effet, dans la chaîne de valeur, c’est surtout l’agrégation de données qui peut provoquer un gain et faire apparaître des bénéfices pour l’utilisateur. L’exemple type : services associant données sur les logements, les transports, les services locaux, les prix de l’immobilier. Mais, précisément, il est très difficile d’agréger les données compte tenu des statuts différents des acteurs et des informations. L’exemple sans doute le plus net est donné par le transport : il est difficile de réunir les informations de différentes AOT et de différents opérateurs.

Un autre obstacle, au sein d’un territoire puis à l’échelon national, est le grand nombre des collectivités concernées, avec lesquelles les réutilisateurs doivent négocier pour obtenir les mêmes données. Or, ici aussi, les gains dans la production de valeur sont liés aux agrégations et aux comparaisons. C’est d’ailleurs précisément ce que visent des sociétés comme Berger-Levrault ou Transpacités. Cependant, il est très coûteux, en temps et en redevances, de s’adresser à un grand nombre de collectivités pour obtenir la matière première du service. L’exemple de la généalogie le montre, où les réutilisateurs doivent de toutes façons négocier avec tous les Conseils généraux. La promotion de règles communes et d’une licence commune serait un facteur important de développement du marché.

#### 4.13 Modèles économiques : difficultés, perspectives, intérêt du foisonnement des start-ups

L’opinion courante est que le modèle économique de la réutilisation des données publiques est encore à trouver. C’est bien sûr tout à fait faux s’agissant de l’information traditionnelle et de son organisation autour de quelques producteurs et éditeurs de données à forte valeur marchande. Les grands

■ éditeurs juridiques, ceux de l’information financière ou des brevets ont bien un tel modèle... Tel n’est pas le cas pour les sociétés qui commencent de réutiliser les données des collectivités sur des marchés locaux.

Or, sur Internet et les mobiles, ce n’est pas seulement pour la diffusion des données publiques que la détermination de modèles économiques reste difficile: elle l’est pour tous les acteurs, pris entre le payant, le gratuit subventionné par d’autres activités et le gratuit subventionné par la publicité. Jusqu’ici, seule une minorité de services, souvent le fait de grands éditeurs, s’est avérée rentable. Le mobile permet certes, comme l’a fait le Minitel avant lui, de développer des applications sur des marchés étroits, avec comme rémunération une quote part versée par le diffuseur (AppStore, Android Market...) mais s’il s’agit d’une application mobile et non d’un accès Web. De petites ou très petites sociétés peuvent vivre sur ce modèle, mais parce que peu de grands acteurs contrôlent la distribution.

Dans le domaine des données publiques plus précisément, un service pourtant apparemment très utile, Proximités, n’a pu être maintenu par la Caisse des Dépôts, car aucun marché et aucun offreur n’émergeait vraiment (la CdC n’a pas vocation à maintenir son intervention dans la durée sur ce type de service). Autre signe inquiétant, la crise est là, et le modèle des grands répertoires - un peu le symbole de l’ouverture des données - le Datagov américain, voit son budget drastiquement réduit et il est presque menacé de disparition depuis les élections de mi-mandat qui ont porté au Sénat une majorité désireuse de freiner les dépenses publiques.

C’est bien là le problème : les avantages globaux pour l’économie ne peuvent plus être obtenus grâce à une dépense continue et soutenue des acteurs publics, en tout cas pas exclusivement et pas facilement. Dès lors, il faut que des entreprises puissent investir sur une période assez longue, ou que des marchés solvables se développent rapidement. Sinon, seuls les réutilisateurs associatifs, centrés sur la transparence et l’information du citoyen, continueront à terme à réutiliser les données ouvertes.

Il faut aussi rappeler que, contrairement à une certaine vulgate, le monde d'Internet est foisonnant mais dur et terriblement concurrentiel. S'y côtoient des multitudes d'intervenants non commerciaux, fonctionnant grâce une économie de l'expression personnelle et de la coopération, dont le type emblématique est Wikipedia, et des intervenants commerciaux d'un nouveau genre, qui captent une part croissante de la ressource publicitaire, et dont le type emblématique est Google. Dans les deux cas, il s'agit d'acteurs mondiaux ayant atteint une très forte puissance de frappe en termes de confiance (pour les deux) et de capacité financière (pour Google et ses pareils).

Etant donné qu'une grande part des gains attendus de la réutilisation des données publiques, du point de vue économique, viendra du rapprochement de données, de la capacité à identifier les informations voisines, d'intervenir au delà de son territoire pour se faire connaître, il est probable que de grands acteurs vont être les opérateurs dominants à terme – la concentration est encore plus forte dans l'économie de l'information que dans l'économie "physique" - les investissements nécessaires pour croître tout en restant à la pointe des techniques et en s'adaptant aux différents marchés la stimulent. Ces acteurs ne seront peut-être pas ceux qui dominent actuellement, car les positions changent très rapidement dans le monde électronique.

Il est certain que quelques unes des start ups exploitant les données publiques se dégageront et connaîtront une forte croissance. Cela ne veut pas dire que les autres auront travaillé en vain. Comme ce fut le cas pour les autres segments de l'activité dans le domaine, l'essentiel est aussi la formation d'une nouvelle génération. Les créateurs et développeurs actuels connaîtront de nombreuses entreprises (s'ils sont jeunes<sup>26</sup>) et ils sont

<sup>26</sup> Ils ne sont pas tous très jeunes. On peut citer le fondateur de Data Publica, François Bancilhon, dont le parcours, après sa sortie des Mines en 1968 puis sa thèse et son travail à l'INRIA, a consisté à créer des sociétés de technologie ayant connu le succès, dont O2 (bases de données objet - 1991), Mandrake-Mandriva (éditeur Unix, 2002), Data Publica et Mobile Service Initiative (2010) (et d'autres).

indispensables à la présence de la France, et de leur région, dans l'économie mondiale qui se réorganise et où l'innovation d'une part, les outils de communication d'autre part, tiendront en tout état de cause une place essentielle. Aussi, l'effort commun des autorités publiques, des Universités et Centres de recherche et des acteurs économiques, au niveau régional comme national et européen, doit il être soutenu pour faire émerger aussi bien les talents que les outils techniques et pour soutenir le passage de la recherche à l'innovation.

## 5 Questions techniques

Les problèmes techniques sont à la fois essentiels et seconds. D'une part, les évolutions techniques sont très rapides et elles déstabilisent constamment les systèmes mis en place dans tous les domaines. Elles offrent de nouvelles opportunités ou entraînent de nouveaux risques et il est important de maintenir une veille et de ne jamais rester lié à une option dépassée. Pourtant, d'autre part, l'essentiel est de fixer des stratégies métier, des principes politiques pour la réutilisation dans les différents domaines et plus généralement pour la circulation de données, qui, quant à elles, sont bien plus stables que les supports.

La charge que représentent non seulement les outils mais surtout les personnes expertes capables de piloter les systèmes et de suivre les évolutions doit être mutualisée. C'est une des motivations fortes pour se grouper dans des plates-formes pour les diverses fonctions. Gérer un répertoire, s'adapter aux évolutions des systèmes d'information géographique, suivre les normes, être accessible aux différents outils..., tout cela exige de coopérer et de disposer de centres de compétence en même temps que d'outils communs.

Enfin, un danger de la complexité technique est qu'elle peut inciter à déléguer complètement des fonctions à des acteurs techniques – voire aux centres de compétence – sans pouvoir ensuite les faire évoluer. Les outils techniques peuvent être un piège, et parfois séduisant par les innovations ou le confort qu'ils apportent. La seule solution pour échapper à ce risque – qui en a frappé plus d'un dans tous les secteurs – est d'échanger avec d'autres utilisateurs et de veiller à ce que les outils soient ouverts et respectent des normes.

Le présent chapitre aborde des problèmes et des perspectives dont beaucoup ne sont pas immédiatement à l'ordre du jour, ou ne semblent pas l'être. Il est vrai qu'il ne faut pas chercher à aborder immédiatement des démarches et respecter des contraintes qui seront celles de demain, par exemple pour des normes ou pour les sémantiques. Il faut cependant se placer dans cette

perspective et prendre garde : dans le monde des TIC, demain arrive souvent très vite.

Aussi faut-il saisir les opportunités et commencer par des applications simples mais il faut définir un schéma directeur de développement. C'est au demeurant ce que font les acteurs qui ont commencé de mettre en place des politiques d'ouverture.

### 5.1 Préambule et mises en garde

#### 5.1.1 Technique ?

L'ensemble de ce chapitre n'a pas vocation, ni prétention, à traiter de la technique au sens informatique. Celle-ci est complexe, affaire de spécialistes, avec plusieurs types d'outils, de formats et d'écoles. L'important, c'est qu'il y ait un dialogue efficace entre les responsables fonctionnels et les directions des systèmes d'information. Il faut garder à l'esprit ce fait encourageant, étonnant mais flagrant, que le Web fonctionne, bien que des divergences opposent les grandes sociétés et les spécialistes sur des points importants. Tous préservent l'interopérabilité, qui est fondamentale. De nouvelles tentatives de définir des écosystèmes redevenus étanches existent avec le mobile, et on verra qu'il faut s'en soucier. Néanmoins, l'heure est bien à l'interopérabilité et celle-ci se construit sur les analyses fonctionnelles pérennes et communes et sur les consensus métier.

#### 5.1.2 Une vision commune, des domaines particuliers

L'organisation d'une ouverture des données vers l'extérieur comprend un certain nombre d'articulations incontournables, mais les éléments de base varient en fonction des types d'applications et de domaines. Ce sera vrai pour l'ensemble des problématiques techniques : elles partagent des

règles communes mais la mise en œuvre et les types de logiciels, entre autres, dépendent des domaines considérés. Ouvrir un fonds d'information géographique ne se fait pas de la même façon qu'ouvrir un jeu de statistiques ou un ensemble de documents budgétaires. En particulier, les outils de recherche et d'interrogation sont différents. Si une politique générale d'ouverture a été décidée, il faut s'attendre à ce que des données de différents domaines soient concernées, ce qui suppose une architecture générique, des logiciels capables de gérer de multiples formats de données et des outils pour les développeurs internes, et souvent pour les réutilisateurs afin qu'ils puissent manipuler les données.

### 5.1.3 Progressivité et articulation de systèmes d'information

Trois niveaux de systèmes peuvent être distingués.

- Le système de gestion de données de l'acteur public lui-même. Comme cela a été indiqué (Ch. 2 Données), son organisation voire sa refonte est souvent prioritaire. Des projets d'ouverture seront souvent menés à bien avant des améliorations programmées du système d'information car il ne faut pas attendre avant de lancer une opération opportune sous prétexte de lacunes dans les systèmes. Ces projets ponctuels seront ensuite réintégrés dans le système d'information.
- Les systèmes mis en place pour le partage entre un acteur public et les réutilisateurs de ses données. Ces systèmes sont ceux qui sont d'abord développés indépendamment d'autres logiques de partage (par des plates-formes mutualisées, sur des programmes communs multi-acteurs)
- Les systèmes plus larges de circulation de données multi-partenaires. Ils sont aujourd'hui mis en place dans les plates-formes partagées – géographie, e-services, tourisme, transports etc.

Progressivement, les systèmes s'intégreront.

### 5.1.4 Indépendance par rapport à la technique ?

L'indépendance entre l'organisation, l'activité et la technique n'est jamais complète dans le temps. De nouvelles possibilités vont modifier jusqu'à l'objet même des efforts entrepris. La transparence ou la réutilisation commerciale ne seraient pas les mêmes si les citoyens ne disposaient pas maintenant d'outils numériques et de moyens d'intervenir directement dans une collecte de données ou un débat. En même temps, on le verra (Ch. 6), l'acteur public doit veiller à ce que la détention et la maîtrise d'outils ne créent pas des inégalités d'accès insupportables.

Au delà de cet impact visible, le danger des outils techniques est qu'ils risquent de faire entrer dans un sous-système contrôlé par une entreprise particulière ou par un autre acteur. Le contrôle peut d'ailleurs provenir d'un opérateur puissant, national ou international, d'une petite société spécialisée, ou d'un groupe particulier de passionnés informatiques. Tous ces dangers sont malheureusement bien connus des responsables de systèmes d'information. La vigilance est donc de mise, qui consiste d'une part, pour les responsables, en un contact régulier avec leurs pairs ou collègues, d'autre part en une adhésion sourcilleuse aux normes ouvertes, contrôlables.

Cette vigilance est incontournable s'agissant de l'ouverture de données publiques. Dans les développements de ce chapitre, les architectures comme les formats apparaissent indépendants des questions juridiques, ce qu'ils sont effectivement... mais pas tout à fait. En effet, une entreprise comme une collectivité, un groupement etc. peuvent souhaiter ouvrir des données exclusivement à des partenaires – sous-traitants et distributeurs, autres administrations, organismes sous tutelle etc. Dans ce cas, il suffit de vérifier que cet éco-système restreint est cohérent avec l'éco-système d'outils techniques retenus. Il peut sembler qu'un format non ouvert soit longtemps efficace dans un tel environnement. C'est d'ailleurs un cas fréquent, un prestataire ayant développé un système sur mesure sans se soucier de respecter des normes.

En revanche, la démarche open data exige de se conformer à des normes et standards ouverts, définis par des communautés vastes et qui évolueront par consensus. Ainsi, le fait que certains acteurs ou certaines données puissent être exclus doit provenir d'une action volontaire et de l'utilisation d'outils de contrôle et de sécurité qui, eux aussi, respectent des normes.

*Certes, les développements «propriétaires» peuvent pendant un temps s'avérer utiles et même nécessitant des investissements moindres, pour un système d'information, fût-il partagé de manière restreinte. A terme, les coûts de tels systèmes explosent en raison des difficultés de suivi et de maintenance, de l'impossibilité de faire jouer une concurrence, de l'arrivée de générations de produits adaptés aux standards ouverts du Web et inutilisables par les systèmes fermés. La démarche d'ouverture des données publiques est en même temps un puissant moteur pour faire évoluer les systèmes d'information vers une flexibilité et une évolutivité dans le temps qui sont à terme incontournables.*

### 5.1.5 Des difficultés réelles

Pour ouvrir des données (même à des partenaires), il faut une maturité des systèmes d'information, au minimum des possibilités d'interfaçage suffisantes.

Comme cela a été rappelé au début du présent travail, ce premier niveau est loin d'être atteint dans de nombreuses administrations, de tous niveaux, mais aussi dans de nombreuses entreprises. Les administrations publiques sont peut-être un peu plus handicapées de ce point de vue, pour des raisons qui ont été indiquées :

Evolution des compétences et des charges, qui ont contraint à des extensions rapides non cohérentes des systèmes techniques, à des applications en silo etc.

Contraintes budgétaires.

Faiblesse de demandes de communication externe.

Par ailleurs, beaucoup de prestataires informatiques ont pendant longtemps privilégié le développement spécifique, qui rend l'utilisateur dépendant.

Ainsi, la mise à disposition de données va buter sur des difficultés de premier niveau.

Le groupe OWNI<sup>27</sup> a effectué un sondage sur 30 administrations, principalement des collectivités, en février 2011. La demande portait sur la communication du budget. OWNI relève que: «*Dans plus de la moitié des cas, il nous a été répondu qu'il était impossible d'obtenir un fichier Excel ou csv. Parfois, le système informatique est seul coupable. Les solutions propriétaires achetées à prix d'or auprès des grands noms de l'informatique d'entreprise n'ont pas été conçues pour être compatibles avec le monde extérieur. Pour les collectivités concernées, les prestataires réclament des sommes pouvant aller jusqu'à 100000€ pour ajouter une fonctionnalité d'export en csv.*»<sup>28</sup>

### 5.1.6 L'évolution vers un espace de données

L'ouverture des données vise, on l'a vu, des domaines différents et des utilisations différentes, de la localisation d'une détérioration d'une chaussée à l'analyse de millions de données ouvertes, par des outils de "business intelligence". Deux tendances nouvelles apparaissent, l'une liée à la mobilité, à un maillage très fin du territoire, où se croisent et se combinent les données – aujourd'hui celles d'une collectivité et de ses partenaires, demain celles de multiples acteurs locaux : l'espace de données sera un outil important de la vie de la cité. L'autre tendance est dans l'analyse de masses de données et de documents à l'échelle d'une région, d'un pays, de

<sup>27</sup> «Objet Web Non Identifié – owni.fr» Ce groupe défend l'open data et la liberté sur Internet. Son fondateur, Nicolas Voisin, a été nommé en avril au Conseil national du numérique

<sup>28</sup> csv : Comma separated value (valeurs séparées par des virgules). Remarque : Excel n'est pas un format ouvert mais, à tout le moins, il est si répandu qu'il est considéré comme un standard d'échange.



la planète. D'ores et déjà, des outils apparaissent, avec les projets des grands du Web, et en particulier de Google. Ils illustrent ce fait que la libération des données publiques a largement commencé.

**Google public data explorer** (<http://www.google.com/publicdata>) montre une application de l'ouverture des données dans l'espace du "big data", encore simple et encore en développement. Un entrepôt récupère des données statistiques libres sur la planète et permet au visiteur de trouver des séries par thèmes et localisation, de les visualiser à sa convenance, de les intégrer dans son site, les mises à jour pouvant être automatiques.

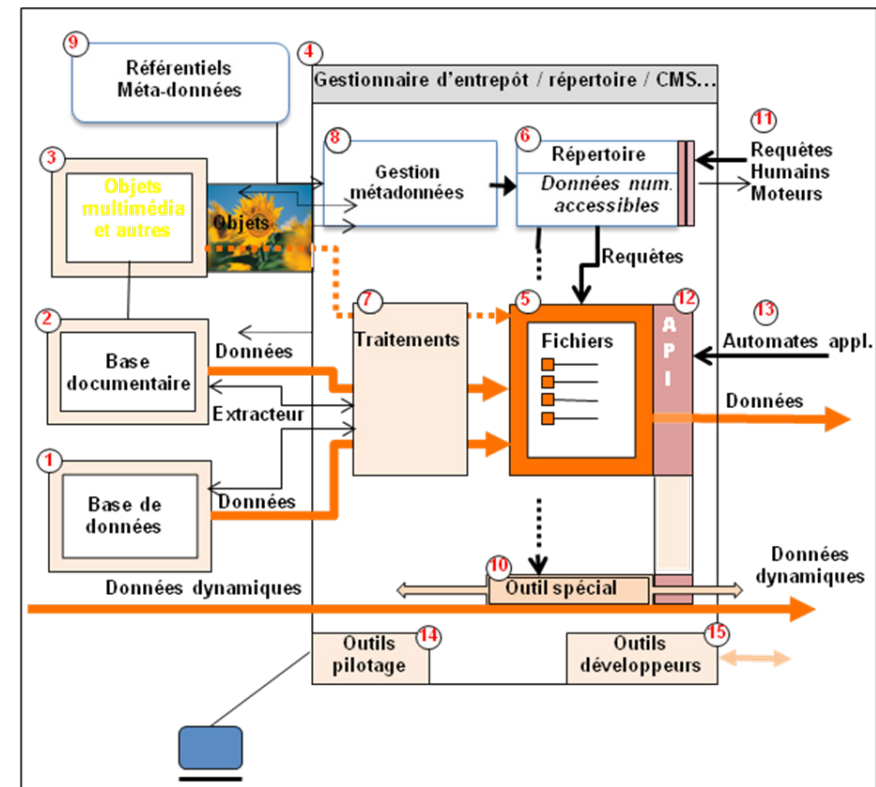
L'espace des données va certainement permettre le développement de places de marché au delà des seules données publiques. Cela implique une normalisation et, à cette condition, le développement de nouveaux acteurs du monde de l'information.

## 5.2 Architecture

### 5.2.1 Les éléments

Le schéma suivant présente les principales composantes des systèmes pour un acteur qui ouvre ses données – que ce soit d'ailleurs à des partenaires ou dans une perspective d'ouverture de données publiques. (Les paragraphes correspondent aux items numérotés sur le schéma).

.../...



Eléments du systèmes de mise à disposition de données

#### ■ 1 - Base de données

Outil clé pour pouvoir répondre à des demandes diverses sur les données, elle doit être flexible, car il faut pouvoir modifier les données, ajouter des items<sup>29</sup>. De ce fait, qu'elles soient locale ou hébergées, partagées, les bases de données doivent être protégées de modifications externes.

<sup>29</sup> Ainsi, Joelle Nonet, responsable du bureau Tourisme au Conseil régional de la Réunion souligne que la base centrale tourisme (cf. Ch2.) est flexible. Elle est modifiée et améliorée avec la prise en compte des usages et des demandes. La contrepartie est que le Conseil régional doit en conserver la maîtrise totale.

Remarque : pour les données de gestion et l'activité principale, les données seront en général dans une même base. Cependant, des bases importantes techniques seront distinctes : c'est en particulier le cas des bases de données cartographiques.

## ■ 2 - Base documentaire

A côté des bases de données, des systèmes de gestion de documents peuvent être également accédés. Les liens entre la démarche documentaire et la démarche des bases de données ne sont pas étudiés ici – des bases peuvent référer les documents – de nombreux outils existent qui associent les deux formes d'information. Il est important de noter que certaines bases documentaires, gérées par les archivistes, les bibliothécaires ou les documentalistes respectent déjà des systèmes de métadonnées avancés – en fait qui ont été à l'origine du concept et de ses développements sur le Web. De telles bases sont accédées au travers des caractéristiques des documents et d'ensembles de mots clé, souvent normalisés.

## ■ 3 - Objets multimédia et autres

Des gisements importants de données publiques sont des images, des photos, des vidéos. Ces objets volumineux justifient d'un traitement spécial – ils doivent être référés dans les différents systèmes comme le documentaire, les bases et l'entrepôt de données publiques mais ils ne seront en général pas reproduits dans ce dernier. Par exemple, pour les photos, une imagerie permet, selon les autorisations, d'appeler l'image en définition plus ou moins grande.

C'est un exemple d'objets autres que des données chiffrées ou des documents. D'autres ressources peuvent être accédées, en particulier les bases et outils d'information géographique et autres bases techniques complexes.

## ■ 4 - Gestionnaire d'entrepôt / répertoire / CMS

C'est en fait un ensemble d'outils, qu'on appelle souvent globalement l'entrepôt de données. Son rôle est de permettre de trouver les fichiers et de les accueillir, prêts à être chargés par un utilisateur. Il assure fondamentalement trois fonctions principales :

- récupérer les données depuis les systèmes de gestion
- les stocker dans un entrepôt
- les mettre à disposition pour les humains et les applications

Ces fonctions se décomposent en plusieurs éléments dont les principaux sont indiqués ci-après.

Il est souvent constitué autour d'un Content Management System (CMS – Système de gestion du contenu) comme Drupal, SPIP, Joomla, outils libres très présents dans les administrations pour construire des sites. Des outils de ce type, eux aussi dans le monde libre, visent particulièrement l'open data (CKAN, qui est au cœur du Datagouv britannique et de NosDonnées en France, Lutèce, développé par la Ville de Paris). Les éditeurs de logiciels proposent aussi, et de plus en plus, des offres assurant des ensembles très complets de fonctions (*par ex. Microsoft tente de revenir sur un marché qu'il avait négligé avec une offre test, en open source mais utilisant ses plateformes, l'Open Government Data Initiative*).

## ■ 5 - Fichiers (l'entrepôt au sens restreint)

Les fichiers sont placés dans un "datastore". Ce sont des ensembles de données brutes, référencés dans le répertoire par leur adresse. Le fichier individuel peut être téléchargé par un utilisateur humain ou une application.

## ■ 6 - Répertoire

Pièce essentielle et visible du dispositif, il permet la présentation des jeux de données (ou collections de documents) aux utilisateurs humains comme aux moteurs de recherche. Dans les deux cas, même si les interfaces sont différents, les fichiers seront trouvés parce qu'ils sont représentés dans le répertoire par les métadonnées, indiquant leur nom et leurs caractéristiques. Les outils de gestion du répertoire permettent l'association des métadonnées aux différents fichiers.

*Rappel : Le répertoire doit aussi référencer, pour un acteur, les données et documents publics qui ne sont pas numérisés, ainsi que ceux qui ne sont pas accessibles, ou accessibles sous conditions. En pratique, les premiers répertoires sont des répertoires de données ouvertes. Ils indiquent pour chaque fichier les éventuelles restrictions d'usage.*

## ■ 7 - Traitements

Un ensemble d'outils est chargé de sortir les informations de différentes bases de données et autres gestionnaires de ressources et de les transformer en des fichiers téléchargeables placés dans l'entrepôt. Le système comprend un extracteur qui dialogue avec les interfaces des ressources internes, avec des procédures de mise à jour. Les traitements vont comporter des mises au format si la base de données ou le gestionnaire de documents ne peut pas produire les formats téléchargeables demandés – les bases sont en général capables d'exporter des séries en formats csv, Excel etc. (cf. *Formats*), les bases de documents contiennent des objets souvent directement téléchargeables mais des conversions peuvent être apportées pour multiplier les possibilités. Au total, des manipulations automatiques sont fréquemment nécessaires, et cela devient lourd quand un format interne doit être converti (par ex. si des données statistiques ou géographiques ne suivent pas les normes).

□ Les outils de traitement devront souvent évoluer pour prendre en compte l'accroissement des types de données à traiter et les évolutions des formats.

## ■ 8 - Gestion des métadonnées

Les données qui décrivent les données. Elles caractérisent les jeux de données ou les collections de documents et ce sont elles qui permettent aux humains comme aux moteurs de recherche de les identifier, au travers du répertoire. Les outils de gestion des métadonnées vont permettre de les faire évoluer, de récupérer celles qui existent déjà dans le système (bases documentaires en particulier, types de données dans les bases), de définir celles qui s'appliquent à tel fichier ou ensemble de fichiers. Les outils de gestion du répertoire s'interfacent à la gestion des métadonnées pour associer les fichiers.

## ■ 9 - Référentiels de métadonnées

Tant en structure qu'en sémantique, les métadonnées doivent être autant que possible puisées dans des ressources partagées au sein d'un domaine. Ce n'est pas encore toujours le cas dans beaucoup de domaines mais les communautés de l'open data travaillent à la construction des référentiels (cf. *infra les normes*).

## ■ 10 - Outil spécial données dynamiques

Les données qui évoluent rapidement et qui doivent être mises à jour pour les utilisateurs en temps réel posent un problème particulier. Il faut qu'un outil, tout en maintenant une protection sécuritaire sur les données, puisse établir une connexion rapide entre les applications clientes des réutilisateurs externes et les systèmes internes qui gèrent ces données.

### ■ 11 - Requêtes au répertoire

- Requêtes humaines: le répertoire présente une interface et des outils de recherche qui permettent à l'utilisateur d'identifier des jeux de données qui l'intéressent. Si c'est le cas, l'utilisateur demande le fichier et le répertoire transmet la requête au système de gestion des fichiers qui effectue le téléchargement.

- Requêtes de moteurs de recherche. Les moteurs généralistes ne peuvent voir que les pages Web et certains objets liés. En outre, ils ne peuvent guère identifier correctement les jeux de données (cf. *infra Web sémantique*). Cependant, à côté de moteurs spécialisés, les grands moteurs commencent de mettre en place des outils. En effet, les moteurs doivent pouvoir «moissonner» les métadonnées, autrement dit récupérer les métadonnées telles que les répertoires, et d'autres gestionnaires d'information les présentent. Des normes sur les métadonnées et les interfaces permettent aux moteurs de fournir des réponses issues de multiples répertoires (cf. *infra Normes*).

### ■ 12 - API

"Application programming interfaces". Ce sont les mécanismes et les formats qui permettent à des applications externes de s'interfacer avec le système (d'un acteur ou d'une plate-forme) pour récupérer les données. Dans certains cas, les API permettent d'intégrer directement les données dans un site ou un service – c'est le mash up (exemple le plus caractéristique : les API qui permettent l'intégration des objets Google map dans les sites Web). Les API doivent être construites à partir des standards du Web (validés par le W3C) avec plusieurs familles dans la catégorie générale des Web services (qui décrivent la présentation des sites pour des automates et les protocoles de dialogue et d'échange de données).

### □ 13 - Requêtes des automates applicatifs à l'entrepôt

En effet, des automates vont s'interfacer avec l'entrepôt dès lors qu'il ne s'agit pas pour un réutilisateur d'un téléchargement ponctuel mais qu'il faut pouvoir effectuer des téléchargements périodiques et/ou intégrer les données dans un site, un service ou une application mobile.

Les requêtes des applications s'adressent directement aux serveurs. En fonction des situations et des accords, l'initiative du transfert de données peut être prise par l'entrepôt – pour les mises à jour prévues. Un fonctionnement précis doit être garanti pour les données temps réel.

### ■ 14 - Outils de pilotage

Les tâches principales sont l'intégration de jeux de données, l'association à des métadonnées, les évolutions des interfaces publics, le contrôle et le suivi de l'activité, la correction des erreurs et la résolution des problèmes posés aux visiteurs et aux réutilisateurs. L'interface de gestion est un élément important de concurrence entre les offreurs d'entrepôts et autres éditeurs de systèmes. L'idéal, souvent annoncé par eux, est que des non spécialistes puissent gérer l'entrepôt, modifier la caractérisation des fichiers par les métadonnées etc. Le but est que ce soient des personnes travaillant dans des services fonctionnels qui puissent le faire. Quoi qu'ils en aient, et même si les interfaces de gestion tendent à se simplifier, il faut que les tâches soient assurées par des personnes spécialement formées et travaillant régulièrement sur le système.

### ■ 15 - Outils pour les développeurs d'applications

Le système doit fournir une interface spéciale pour les développeurs d'applications réutilisatrices des données, leur indiquer les procédures,

les formats, les contraintes, leur permettre de connaître les évolutions etc. Il est aussi possible de placer à leur disposition des applications utiles au traitement de jeux particuliers de données – développées par l'acteur public ou par d'autres utilisateurs, constituant alors une communauté de type open source, où les développeurs peuvent échanger modules et information.

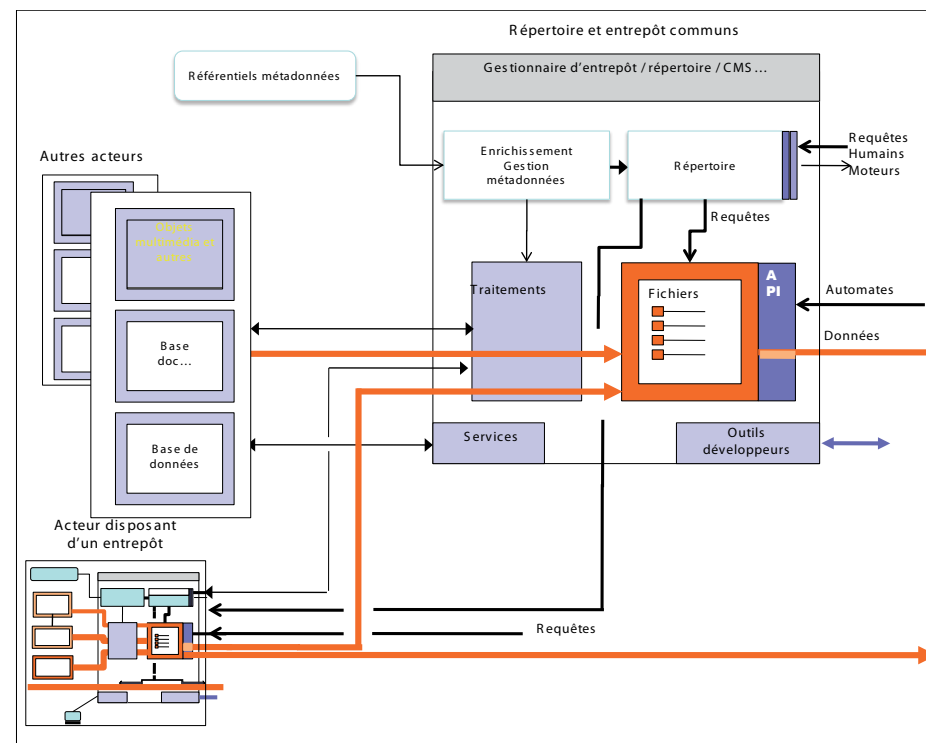
### 5.2.2 Plates-formes mutualisées

En 2011, il n'existe pas vraiment de plate-forme assurant complètement une ouverture mutualisée de données publiques.

Les plates-formes coopératives déjà abordées (géographie, tourisme, transports en particulier) gèrent souvent de telles données mais elles ont des autonomies très limitées pour mettre à disposition directement des données (cf. Ch. 3). Dans l'avenir cependant, des entrepôts se créeront qui regrouperont entre autres plusieurs collectivités. En même temps, les plates-formes spécialisées auront elles aussi à répondre à des demandes de plus en plus nombreuses et se verront déléguer la gestion de certains ensembles de données.

Le schéma ci-contre esquisse le fonctionnement d'une plate-forme. Il rappelle qu'il y a deux types d'acteurs : ceux qui disposent des moyens pour assurer eux-mêmes une mise à disposition et ceux qui la confient à la plate-forme. Il faut noter qu'un acteur peut conserver ses outils, ses relations avec des réutilisateurs etc. et reproduire tout ou partie de son répertoire et de son la plate-forme. De même, il pourra y avoir plusieurs plates-formes, l'important étant qu'elles soient interopérables. Le choix de différentes solutions peut se faire en fonction de critères d'organisation et de charge des systèmes.

En pratique, le répertoire va se comporter comme un portail, transmettant des requêtes vers l'entrepôt de la plate-forme ou vers des participants si les données ne sont pas hébergées sur le même site.



### Entrepôt mutualisé

## 5.3 Stratégies techniques de développement

### 5.3.1 Rappel - une seule logique de gestion de données : interopérabilité et sécurité

La gestion des données, interne ou à destination du système de mise à disposition, doit répondre aux mêmes exigences métier et techniques.

L'amélioration des données, l'identification de métadonnées et la réflexion sémantique associée, le choix des formats informatiques utilisés pour les échanges externes etc. doivent être en synergie avec les évolutions du système d'information et être mis à profit pour les circulations et réutilisations internes.

Les responsables informatiques des collectivités considèrent donc souvent qu'il serait utile d'appliquer en interne le Référentiel général d'interopérabilité défini jusqu'en 2011 par la DGME au ministère de l'Economie – maintenant la Direction interministérielle des systèmes d'information et de communication (DISIC). C'est un premier niveau d'interopérabilité, sur les protocoles de base et un ensemble de formats de fichiers (cf. Documents). Obligatoire pour les échanges externes des administrations, le RGI s'applique aux collectivités mais dans la mesure où son respect ne contrevient pas au principe de libre administration. En fait, il s'agit d'un guide de bonnes pratiques, même en ce qui concerne l'Etat.

En particulier, des DSI considèrent que beaucoup de collectivités ne maîtrisent pas la sécurité des systèmes d'information. Ils craignent que l'open data ne s'accompagne de risques et proposent d'ouvrir en même temps les chantiers de sécurité et diffusion. Il s'agit dans ce cas d'appliquer le Référentiel général de sécurité (RGS) associé au RGI. Ici encore, la politique open data pourrait au contraire avoir un effet d'amélioration de la sécurité et de la qualité.

### 5.3.2 Anonymisation

L'anonymisation est un traitement particulier des données qui, lui, est uniquement nécessité par l'ouverture, lorsque celle-ci est souhaitée mais ne peut se faire sans divulgation de données personnelles.

Le GFII et l'AFCDP (Association française des correspondants aux données personnelles) ont créé un groupe de travail spécial pour analyser les

problèmes liés à l'ouverture des données publiques et à l'anonymisation. Ils constatent que chaque cas est spécifique. Il dépend des objectifs fixés, de la nature des applications et de la structure des bases qui gèrent les données. Plus il y a d'informations sur une personne, par exemple, et plus l'inférence par recoupements est possible, ce qui va exiger des transformations, l'effacement de certaines caractéristiques. Les applications qui ne sont intéressantes qu'au niveau de cas individuels supposent de gérer les homonymies etc. Toute opération d'anonymisation doit donc être gérée comme un projet spécifique, allant de l'analyse du mode de collecte et définition des données concernées aux vérifications sur les fichiers anonymisés. L'opération peut être coûteuse, et il est difficile d'évaluer le budget correspondant.

Les spécialistes sont un peu décourageants quand ils affirment que les techniques et les possibilités d'identification par inférence progressent rapidement. L'expérience montre que, avec toutes les données qui sont aujourd'hui présentes sur le Web pour un nombre croissant de personnes, les recoupements permettent de ré-identifier de plus en plus facilement des données anonymisées.<sup>30</sup>

### 5.3.3 Utiliser et réutiliser sur tous supports

Dès aujourd'hui, le mobile a largement bouleversé des démarches qui avaient été centrées sur le Web, lequel avait déstabilisé le Minitel etc.

Des choix techniques en découlent, et des montants d'investissements, car l'interopérabilité présente bien des lacunes. Ainsi a-t-il semblé utile, après l'ouverture vers le Web, de penser à des mises en forme utiles pour les réutilisations sur iPhone, puis sur Android. Cependant, la question se pose toujours de choisir entre une application Web et des applications mobiles. De nombreux débats ont lieu de ce fait, comme celui sur les formats.

<sup>30</sup> cf. article de Paul Ohm - [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1450006](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006) - ou Documents – téléchargés grâce à la politique d'Open Archives

Cette évolution rapide et la diversité des supports finaux possibles conduisent à privilégier une analyse approfondie et une gestion des données indépendante de l'objectif de restitution, produisant des fichiers de données brutes caractérisés par des métadonnées normalisées. La généricité doit permettre les diffusions et les réutilisations multi-canaux. Les contraintes relatives aux différents systèmes de diffusion importants seront en général intégrées à leurs produits par les éditeurs d'entrepôts. La diversité des supports sera souvent assurée par la diversité des réutilisateurs.

### 5.3.4 Choix d'outils

Comme cela a été souligné, les logiciels ne doivent pas constituer un obstacle à l'interopérabilité. Comme l'enquête d'OWNI l'a encore montré (cf. ci-dessus «difficultés»), certains systèmes d'information sont incapables de transmettre, à tout le moins sans effort humain important, des formats même simples. En tout état de cause, l'évolution conduit à une élimination progressive de telles applications. Dans certains cas, la mise en place de plates-formes mutualisées, offrant des services pour la mise à disposition et le partage de données, conduit à la suppression d'applications hétérogènes chez de nombreux partenaires.<sup>31</sup>

Ainsi, le déport de fonctions sur une plate-forme introduit une dépendance. C'est pourquoi il faut prendre garde aux mécanismes de gouvernance de telles plates-formes.

Or, une possibilité actuelle importante, en informatique, est le remplacement de fonctions assurées par l'informatique interne ou par des prestataires locaux contrôlables (hébergement de données à l'extérieur, etc.) par des services situés dans le «nuage». En pratique, d'ailleurs, c'est un peu revenir à ce qu'était l'informatique dite en «service bureau» dans les années 80,

<sup>31</sup> Cela a été par exemple le cas avec la mise en place de la plate-forme de services pour les acteurs du tourisme réunionnais (Soubik).

avec en plus toutes les facilités et les interfaces du Web. Le problème est que cette solution introduit elle aussi une dépendance accrue par rapport au prestataire. Aujourd'hui, elle se place dans un cadre de protocoles normalisés – et c'est une condition essentielle que de vérifier que toutes les normes et tous les contrôles nécessaires sont respectés. Le risque serait que demain des ensembles d'acteurs soient entraînés dans des tensions entre grands offreurs. Cette solution présente donc aussi un risque. Elle est cependant tentante parce que ces offres assurent souvent des coûts inférieurs et, précisément, la garantie de respect des évolutions techniques, car elles sont le fait d'acteurs puissants qui participent aux travaux de normalisation.<sup>32</sup>

Il faut en tout cas porter une grande attention aux contrats de licence ou d'abonnement.

## 5.4 Normes

### 5.4.1 Une exigence, sans dogmatisme, liée aux systèmes ouverts et à la communication

#### ■ *Un monde en évolution*

La norme n'a pas toujours bonne réputation, notamment dans le monde des systèmes d'information, où sa mise au point, qui doit être consensuelle, semble souvent, et a souvent été en effet, trop longue par rapport au développement de l'informatique. En outre, une confusion règne entre norme et standard de marché, avec ce handicap supplémentaire que l'anglais, langue dominante actuelle du secteur TIC, ne connaît que le mot standard.

En outre, la normalisation paraît multiple et complexe, allant des matériels de communication aux langages et représentations du réel.

<sup>32</sup> Confronté à de graves réductions budgétaires, le répertoire américain data.gov va se placer dans le nuage.

Pourtant, contrairement à des idées reçues, même en informatique, il n'y a pas que les standards de marché (ceux de Microsoft notamment, mais Microsoft est aussi très actif dans la normalisation). Ainsi, des formats comme jpeg, , mpeg, mp3, SGML (dont tous les Xml dérivent) etc. sont des normes ISO officielles.

Par ailleurs, dans le monde d'Internet, des organisations internationales ont modernisé et accéléré le processus de normalisation – en accord avec le circuit ISO officiel : l'Internet Engineering Task Force, le W3C, OASIS pour les échanges en particulier. La caractéristique des nouveaux mécanismes est que les propositions de normes, présentées en général par des consortiums regroupant dans des proportions variables industriels et grands utilisateurs, doivent reposer sur des applications effectives. Ensuite, les points de vue sont rapprochés et la communauté retient les meilleures solutions, prenant souvent le meilleur des diverses propositions.

#### ■ Une parenté étroite avec les systèmes ouverts

Normalisation et communication sont étroitement liées. C'est parce que les gens veulent échanger des images qu'une norme publique s'est imposée. L'exemple des télécommunications est évident, mis en œuvre d'ailleurs par une organisation internationale (UIT).

En approfondissant, on comprend que normalisation et communication sont à leur tour très liées au concept d'ouverture et à ses déclinaisons dans les systèmes ouverts. Ainsi, un format standard mais non ouvert, émanant de grandes sociétés (tous les leaders de l'informatique en ont plusieurs) peut évoluer sous le seul contrôle de son propriétaire et cela déclenche régulièrement des problèmes de communication. En revanche, le format normalisé ouvert – ou un format créé par une communauté libre mis à disposition de tous – ne sera modifié que consensuellement et la disponibilité des informations sera totale, permettant à de multiples concurrents et utilisateurs d'adapter immédiatement leurs systèmes.

□ De même, les protocoles et structures ouvertes de communication – comme par exemple celles de l'EDI ou des Web services – posent comme principe l'indépendance avec les logiciels qui produisent les données, ce qui évite la création de zones non intercommunicantes.

C'est pourquoi le monde du libre est né réellement sur Internet. Tout y est en fait lié : logiciels open source, protocoles ouverts, standards de définition des structures de données etc. C'est un monde communicant et coopératif – ce qui ne veut pas dire que tout y soit gratuit. Il s'agit de la construction d'un écosystème. Un logiciel libre et ouvert permet plus facilement l'adaptation, par une communauté de développeurs, aux formats et langages libres développés par une autre.

Ceci étant, aucun dogmatisme ne doit être de mise dans ce domaine, si on n'est pas soi-même un acteur de l'informatique ou un passionné. En fait, les sociétés qui se sont établies sur le "propriétaire" savent s'adapter, à commencer par Microsoft, dont plusieurs initiatives dans le domaine respectent des licences libres et dont les logiciels supportent les formats ouverts. De même, Google, qui se construit depuis le début sur la communication, base-t-il ses applications sur des composants libres. On verra qu'avec Microsoft, Google et Yahoo!, des industriels féroce ment en compétition s'entendent pour proposer des standards ouverts pour l'open data (*mais sont contestés – cf. infra*).

*Remarque. Il faut toujours rester vigilants. Ainsi, la bataille très dure qui oppose actuellement Apple et Google sur le mobile a bien pour but de recréer des espaces fermés disjoints du Web et contrôlés. Des conséquences sont à tirer à chaque tentative de ce type.*

#### ■ Plusieurs niveaux dont un plus crucial

Les normes portent (très grossièrement) sur :

- les protocoles d'échange et les interfaces que présentent les sites et



applications pour se parler. Ce sont là des questions très techniques que les informaticiens maîtrisent – il faut seulement qu'ils suivent la réflexion nationale et internationale, avec plusieurs organismes donnant des avis autorisés, dont le ministère de l'Industrie, la DISIC, l'AFNOR, les instituts de recherche nationaux et régionaux etc.

- les formats de données et documents – mêmes remarques, avec une forte intervention possible des utilisateurs (la non lecture d'un format connu est notamment inacceptable)
- les structures sémantiques qui vont permettre de se comprendre et de nommer et caractériser d'abord les fichiers (métadonnées) puis d'analyser les données et documents eux-mêmes.

Ce dernier niveau est le plus riche et le plus crucial. Surtout, c'est celui qui doit être maîtrisé par des organisations métier et des responsables fonctionnels. Il suppose de s'accorder, comme cela s'est toujours fait dans les métiers, mais en précisant les langages communs dans des expressions rigoureuses, pour éviter les incompréhensions qui sont d'ailleurs aussi toujours fréquentes dans les échanges humains mais irrattrapables entre automates. Il faut donc définir des concepts, des périmètres, des conditions d'observation – l'exemple de l'emploi est patent, avec de multiples définitions pour les effectifs, les taux d'activité, les rémunérations etc. Il faut aussi préciser les mesures employées : les risques sont forts si un système émet des données en milliers d'euros et que l'autre comprend des millions, ou des g vs des mg...

*Les débats actuels sur les formats cèderont et cèdent déjà la place aux débats sur l'enregistrement de données «signifiantes» dans les pages Web et les objets et sur les vocabulaires normalisés. C'est la naissance du Web sémantique, et les données publiques en sont une composante importante.*

- *Faire simple et ouvert...*

□ La complexité de la normalisation peut rebuter. En fait, la tendance à l'ouverture est inhérente aux développements actuels des TIC et il faut laisser les techniciens travailler. Il faut choisir des solutions simples et s'informer dans les communautés d'utilisateurs, puis suivre les projets et s'inquiéter seulement des difficultés d'interopérabilité que signaleraient des réutilisateurs et des utilisateurs. La mutualisation dans des communautés est la meilleure voie pour aller vers des normes.

*Les développements qui suivent reprennent notamment les recommandations présentées par le groupe de travail du GFII sur les formats et présentées notamment au COEPIA (cf. documents).*

- ... tout en sachant les perspectives : le web sémantique

Après les applications ponctuelles, l'organisation dans un plus vaste espace exigera rapidement de mieux caractériser les informations. Les paragraphes qui suivent et évoquent le Web sémantique ne renvoient donc pas à un futur très éloigné.

Aujourd'hui, les données brutes sont publiées sans liens entre elles – il faudra bientôt s'y retrouver dans une masse de données, dont certaines se recouvrent, ou qui dépendent d'autres jeux. C'est le vrai enjeu des plates-formes et catalogues : au début, une liste avec quelques caractéristiques suffit pour les développeurs et les services tant qu'il n'y a pas de très nombreuses données. Ensuite, il faut à la fois une démarche métier et une rigueur de normalisation informatique (descripteurs, métadonnées, référentiels de tous types comme les adresses, les noms de lieux, les positions géographiques, les unités de mesure etc.).

Sinon, après un premier succès, les entrepôts et plates-formes commenceront à ressembler à des capharnaüms, voire des poubelles à données.

Il y aura en fait beaucoup d'intelligence humaine et informatique à mettre

dans ces plates-formes. Certes, il ne faut surtout pas vouloir commencer avec des outils de haute sophistication, mais il faut veiller et échanger.

*La nécessité d'une organisation plus poussée et plus précise des données, utilisant les techniques du Web sémantique, va apparaître rapidement dans les domaines qui sont déjà des enjeux de coopération et de visibilité internationale : environnement, tourisme, transport, ressources économiques, marchés publics entre autres. (D'autres domaines sont par nature déjà bien structurés, au moins au niveau national, notamment l'information juridique.)*

#### 5.4.2 Référentiels : nomenclatures et systèmes d'identification

Un certain nombre de nomenclatures et de systèmes d'identification sont essentiels pour toutes les données. Ils doivent faire l'objet d'une normalisation nationale et devraient au demeurant être systématiquement mis à disposition de tous gratuitement (*Recommandation du GFII, notamment*).

Il s'agit des nomenclatures et des identifiants que l'administration définit pour les produits, les entités économiques, les territoires. Des serveurs nationaux doivent permettre leur accès, l'identification des versions (les nomenclatures évoluent et toutes doivent indiquer l'entité responsable, le périmètre, le n° de version).

Notamment, les collectivités le savent : les systèmes d'adressage sont multiples et les découpages sont multiples – pompiers, INSEE, Poste... - et cela pose problème aux acteurs du territoire.

Ces nomenclatures sont définies par les différentes administrations responsables, qui attribuent également les n° d'identification (par exemple les ministères sociaux pour les identifiants des établissements et acteurs du système socio-sanitaire).

*Remarque : cette ouverture des référentiels et des identifiants renvoie au problème de la mise à disposition des grands fichiers de base, tels SIRENE pour*

*l'INSEE ou AEF pour les chambres de commerce, qui listent les entités et les adresses en les typant par les nomenclatures. Ces fichiers sont une ressource économique et parfois politique importante et les demandes de réutilisation gratuites se heurtent partout à des refus, ce qui est compréhensible. La gratuité d'accès à la nomenclature et la gratuité d'obtention d'un n° dans un fichier ne préjugent pas de la réutilisation d'ensemble ou partielle de celui-ci, mais il est certain que les débats reprendront régulièrement sur ce point.*

#### 5.4.3 Formats

Au delà des grands formats informatiques reconnus par tous, chaque domaine a ses propres formats, par exemple la géographie, les plans techniques etc. Heureusement, les domaines qui disposent de formats spécifiques sont organisés en communauté. Les débats se font souvent entre les communautés et les institutions responsables au niveau national et international.

##### ■ *Un minimum*

La recommandation minimale est que les fichiers de données soient dans un format réutilisable – cela veut dire que, y compris avec des traitements difficiles, un réutilisateur (ou tout autre partenaire) pourra les convertir, si besoin est, en des formats standards. On l'a vu, ce n'est pas toujours le cas.

En outre, tout dépend de l'utilisation : ainsi, des documents textuels en format image (pas en pdf) peuvent être utiles s'il s'agit de présenter des exemples de documents mais inexploitable si c'est le texte qui est recherché et si les images sont trop difficiles à analyser par des logiciels de reconnaissance. Il faut dans ce cas préciser que la possibilité de réutilisation est limitée.

*Une administration ne devrait pas non plus proposer des formats inadaptés alors qu'elle dispose des données dans un format utilisable. Il n'est ainsi pas*

sérieux de donner accès à des budgets seulement en pdf alors qu'au départ ils sont en xls (Excel) (cf. infra "le cas de PDF"). C'est pourtant ce que font plusieurs collectivités, comme l'a constaté OWNI dans son enquête. Le Journal Officiel avait tenté cette stratégie en plaçant en accès libre une édition en pdf et non en texte en 1997 (Une édition "pirate" en mode texte amena le JO à changer de position – le gouvernement ayant pris clairement parti pour l'ouverture). Des recours auraient vraisemblablement raison de ce type de comportement, qui revient de fait à bloquer la réutilisation ou à l'handicaper lourdement, sans justification.

Au delà, beaucoup d'industriels de l'information considèrent que l'administration ne doit pas attendre d'avoir fait évoluer ses systèmes vers des formats standards pour ouvrir ses données. Au mieux, cela retarderait la confrontation à la pratique, au pire ce serait souvent un alibi. Les sociétés – ou par exemple des plates-formes spécialisées – peuvent prendre en charge les conversions. Cela concerne bien sûr les sociétés commerciales qui ont investi dans le domaine et pensent pouvoir faire face à cette charge. C'est d'autant plus vrai si elles ont une vocation internationale car elles vont de toutes façons avoir à rapprocher les données des différents pays<sup>33</sup>.

En revanche, une contrainte est importante : ne pas changer de format sans laisser un délai d'adaptation aux réutilisateurs.

#### ■ Des formats ouverts

Ceci étant, dans toute la mesure du possible, l'administration doit proposer des formats ouverts (*Recommandation du GFII*), en tout cas des formats largement répandus et pour lesquels les communautés d'utilisateurs et de réutilisateur peuvent disposer d'un maximum d'outils de traitement et de présentation.

<sup>33</sup> Position exprimée par exemple par Euroalert, qui fournit information et services sur les marchés publics en Europe (cf. Ch 4).

□ Les formats ouverts et les standards de marché correspondent à cette définition. Comme on l'a vu, les standards dépendent d'un industriel – parfois d'un groupe contrôlé par un industriel largement dominant ou d'un consortium d'industriels puissants. C'est le cas des formats bureautiques de Microsoft ou des formats d'Adobe (pdf, flash). En revanche les formats ouverts sont libres de droits et contrôlés par des communautés très larges et les organismes reconnus de standardisation et normalisation. Ce sont entre autres html, odf et csv.

#### Le cas de PDF (Recommandation du GFII)

Sans être un format ouvert, pdf est de fait un standard de marché largement répandu et efficace. A son sujet, le GFII a présenté la recommandation suivante :

Des données source à côté des PDF.

PDF est un excellent format de données pour présenter un document, notamment au grand public.

Il est en revanche souvent un très mauvais format pour la réutilisation, car il oblige le réutilisateur à un exercice contre-productif de rétro-ingénierie.

Dans la majorité des cas, le fichier PDF texte résulte du passage en PDF de documents plus exploitables (word, excel, powerpoint, etc.). Donner accès aux données source qui ont servi à produire le PDF (quand ils sont disponibles) ne présente aucun coût supplémentaire pour l'éditeur.

Nous recommandons donc que dans le cas de publication d'un fichier PDF texte, le fichier source du PDF, s'il est disponible à l'éditeur, soit systématiquement publié en même temps que le fichier PDF.

- Privilégier la syntaxe XML et utiliser des formats d'échange existants ou contribuer à leur définition

Le GFII recommande de poursuivre le mouvement d'adoption d'XML, largement entamé dans les administrations d'Etat. XML est en fait, si on écarte les innombrables spécifications techniques, une manière hiérarchisée de baliser un texte pour le rendre assimilable par les automates – mais aussi compréhensible par des humains patients. Ainsi, un nom pourra être précédé par une balise <nom> et suivi par </nom>, lui-même enchassé par exemple dans un ensemble <abonné>...</abonné>. Aussi y a-t-il de nombreux "langages XML" pour les différents métiers. C'est une première étape importante de la sémantisation des données.

Le GFII propose aussi d'utiliser systématiquement des formats d'échange existants. En effet, pour les Echanges de données informatisés (EDI), des langages XML ont été définis dans de nombreuses communautés (transport, banque, commerce, comptabilité, tourisme, santé etc.). Ces langages servent aux échanges entre les systèmes d'acteurs différents au sein des communautés. Ils peuvent être utilisés chaque fois que c'est possible dans les entrepôts pour les données spécialisées. En effet, un fichier comptable en format xls, voire csv, est moins aisément utilisable que le fichier converti en format d'échange comptable – avec des balises normalisées qui indiquent la signification des différents postes. Pour les connaître, et éventuellement contribuer à leur évolution, il faut se tourner vers les communautés EDI<sup>34</sup> ou les administrations directement concernées (par exemple la direction du tourisme).

<sup>34</sup> Par exemple EDIFICAS pour les données comptables et financières, Edibuild France pour le bâtiment, GS1 France pour le commerce et la logistique etc. Il n'y a pas aujourd'hui d'organisation centrale – c'est l'UN CEFACT au niveau mondial ONU – on peut se renseigner auprès du ministre de l'Economie, le chef de la délégation française étant Jean-Pierre Henninot en 2011.

#### □ 5.4.4 Accessibilité

Pour mémoire : l'accessibilité des informations et même leur manipulation à distance sont un chantier clé pour tous les acteurs du Web. Aussi, de nombreux outils et méthodes différents sont proposés. Ils ont en tout cas deux caractéristiques communes : ils sont tous ouverts, publiés etc. et par ailleurs aucun ne vise à refermer un espace, même si des divergences existent. Personne ne conteste l'interopérabilité générale, au moins sur le Web. Les outils permettent d'implémenter des fonctions différentes.

Par exemple, Microsoft propose protocoles et outils "odata" pour accéder à distance aux données contenues dans les bases et les récupérer ou les mettre à jour (fonction par exemple très utile pour les plates-formes spécialisées ou mutualisées qui gèreront des entrepôts connectés à leurs membres). Or, il est annoncé qu'odata ressort de la "Open specifications promise", une sorte de garantie d'ouverture et de promesse que Microsoft ne fera pas jouer de brevet sur les implémentations du protocole dans des logiciels.

Un protocole est particulièrement important pour les archives scientifiques et en général toutes les collections de documents culturels au sens le plus large : Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), largement adopté désormais par les bibliothèques universitaires et la communauté croissante d'organismes et de chercheurs réunis dans l'Open Archive. Le protocole permet de lire des métadonnées présentées selon la norme de structure Dublin Core (cf. *infra les normes de structure pour les métadonnées*) et donc de moissonner les métadonnées de multiples sites d'archives pour répondre à une même requête d'un moteur de recherche.

On peut aussi citer un protocole important qui permet l'analyse multidimensionnelle de jeux de données sur le Web, Online Analytical Processing (OLAP). Des applications respectant ce protocole ou proches permettent à l'internaute de manipuler les données statistiques en ligne. Google expérimente en 2011 son Public Data Explorer construit sur ce principe et donnant accès à un entrepôt de données des instituts nationaux et

de grandes institutions internationales. Google exploite aussi ses ressources cartographiques pour la présentation des données, qu'il est possible d'intégrer dans un site de façon dynamique, avec les mises à jour automatiques. Ce type d'applications montre ce qu'il est possible de faire en utilisant les données publiques, ici avec la puissance informatique d'un Google.

Il faut bien sûr que des normes soient respectées pour les métadonnées et surtout pour la manière dont elles sont présentées vers l'extérieur, et c'est là que, en 2011, des problèmes se posent (*cf. infra*).

#### 5.4.5 Des référentiels de base pour les données et les métadonnées

Il sera rapidement utile de travailler à définir un identifiant pour les jeux de données publiques. De tels identifiants d'objets sont de plus en plus courants dans le monde des TIC et du Web. Leur origine remonte loin – sans doute à la bibliothèque d'Alexandrie, où un système de classification et d'identification avait été mis en place. Les bibliothécaires ont créé les numéros ISBN et les fiches Mark d'identification et de classification des ouvrages. On verra ci-dessous qu'ils ont à l'origine d'une fiche générale pour tout objet sur le Web. Dès lors, il faudrait que la hiérarchie des entrepôts et des fichiers fasse l'objet d'une identification non ambiguë et d'une fiche minimale, ce qui va exiger de développer une nomenclature de premier niveau indiquant le domaine et le type d'objet. Chaque fichier aurait donc son titre, son nom, sa localisation et un ensemble de descripteurs généraux.

Il serait important de confronter les dénominations retenues dans les différents répertoires de données publiques existants et d'organiser la réflexion sur le premier niveau commun à mettre en place. De ce point de vue, les nomenclatures du Géocatalogue de l'IGN représentent une ressource importante. Cette nomenclature identifie et relie les types d'objets qui peuvent être recherchés sur des cartes (*cf. infra*).

#### 5.4.6 Web sémantique

Si le Web sémantique est encore vu comme une promesse, le chantier est largement avancé et de nombreuses applications s'en approchent déjà, notamment dans les domaines où des moteurs spécialisés existent. En ce qui concerne les données publiques, le Web2 a apporté le crowdsourcing, le Web sémantique est une promesse d'accroître dans des proportions impressionnantes la valorisation de la ressource.

##### ■ *Objet*

Il s'agit de définir des ensembles d'information qui, placés sur les ressources du Web (sites et autres, identifiés par les Unified Resource Identifier, dont les URL sont un cas particulier), vont permettre de comprendre la nature de la ressource et le sens des informations ainsi que de la lier à d'autres (qui peuvent être des objets du Web ou des objets physiques). Il sera alors enfin possible de distinguer, par exemple, un site de vente d'un site scientifique sur le même domaine, de les classer dans des rubriques différentes, de les associer à des sites de même nature. Le Web sémantique traite donc de l'organisation et du contenu des métadonnées placées pour présenter une ressource – dans le cas des entrepôts de données publiques, celles qui vont décrire la nature de l'entrepôt et des fichiers qu'il présente.

Il importe donc de définir d'une part le classement des métadonnées dans des rubriques et une organisation hiérarchique, d'autre part de définir des vocabulaires et, plus profondément, des ontologies (c'est-à-dire la relation entre des concepts : par exemple une démarche aussi générique que possible pour décrire un site touristique ou un budget de fonctionnement).

Les métadonnées vont pouvoir différencier les types de fichiers par nature et type (par exemple analyses économiques sur un secteur vs liste d'entreprises du secteur). L'utilisation de langages normalisés et de métadonnées appartenant à des vocabulaires eux-mêmes normalisés va

permettre aux moteurs de retourner des réponses cohérentes issues des différents entrepôts.

En allant plus loin, le moteur pourra retourner des réponses synthétiques mariant les informations de multiples sources et fichiers. Il sera alors possible d'interroger par exemple un site régional d'informations sur l'eau, un site sur l'agriculture et le site administratif sur les réglementations correspondantes, ou de comparer des temps de trajets et des prix immobiliers renseignés par plusieurs territoires, ou encore de rapprocher les résultats des sites traitant de la protection d'une espèce particulière etc. Une autre approche sera de créer des sites de références, par exemple sur un territoire, en reprenant des informations issues de nombreux sites spécialisés.

Ce type de développement correspond au projet d'un "Web de données" (Linked data) auquel le W3C consacre un groupe important. L'application la plus spectaculaire à ce jour est la construction par Wikipedia d'un entrepôt Dbpedia qui rassemble, pour une même entité, des informations issues de plusieurs articles et d'autres sources sur le Web.

#### Sémantisation des données parisiennes. (<http://parisemantique.fr/>)

Un projet privé a entrepris de classer et caractériser les séries mises en ligne par la Ville de Paris, en construisant une ontologie, qui relie entre eux les différents concepts caractérisant les fichiers. Dès lors, il est possible de rechercher des informations non à partir de la liste présentée par le répertoire de la Ville mais à partir de requêtes sur une rue ou sur un type d'événement.

- *L'organisation des métadonnées, clé de la présentation future des répertoires et du Web sémantique*

La présentation des métadonnées sur un site doit se faire selon une structure et avec une logique de dialogue précis. Jusqu'à juin 2011, deux approches existaient.

La plus officielle sur le Web est un schéma XML de description d'un objet (site, page, fichier, collection d'objets) qui propose la structure **RDF** (Resource Description Framework). C'est une présentation riche, capable d'identifier un site, une page, un fichier dans la page en décrivant la hiérarchie et permettant la liaison à d'autres éléments par leur URI (Uniform Resource Identifier).

Une application particulière notable de cette structure est le **Dublin Core** des bibliothécaires qui liste des éléments de base pour décrire des objets physiques ou des sites en tant qu'objet publié. C'est à partir de cette structure ou en s'en inspirant qu'un descripteur commun de fichiers, envisagé ci-dessus, pourrait être construit.

#### Dublin Core : les éléments minimaux obligatoires

Title	Format
Creator	Identifier
Subject	Source
Description	Language
Publisher	Relation
Contributor	Coverage
Date	Rights
Type	

Une autre approche moins structurée que RDF et moins ambitieuse, les microformats, permet d'insérer des informations supplémentaires dans le code html d'une page. Des groupes ont défini de nombreux microformats pour préciser ce que contient un élément (liste, adresse, citation etc.). Un groupe de travail du W3C développe les standards pour inclure ces balises sémantiques dans HTML5.

Cependant, au printemps 2011, les trois leaders des moteurs et annuaires, Google, Microsoft et Yahoo! ont annoncé une méthode et une ontologie destinés, selon eux, à accélérer et simplifier cette recherche de sens sur le Web : **Schema.org.**, avec un format de présentation des métadonnées, "microdata". Ils déclarent viser les données structurées qui ne sont pas reconnaissables comme telles dans des pages Web. Les moteurs (ceux des partenaires ou des concurrents...) pourront identifier les balises spéciales que le propriétaire d'un site pourra récupérer sur leur site. L'ontologie, hiérarchique, distingue les différents objets possibles (œuvre créative, évènement, personne, lieu etc.) puis progresse par sous-groupes (pour les œuvres, il y a aussi bien la page Web, l'élément, un livre, une recette de cuisine etc.). L'influence de Google est évidemment très forte dans la proposition.

La réunion de ces concurrents, comme le débat qui s'en suit depuis juin, illustre l'importance des données publiques et en général du "big data". Pour les uns, Schema.org est un accélérateur vers le Web sémantique. Pour les autres, c'est un moyen pour les trois acteurs de prendre le contrôle d'un aspect de plus en plus important du Web, au détriment des groupes ouverts travaillant au sein du W3C et des autres instances de standardisation et normalisation. En effet, les fondateurs de Schema.org soulignent que, même si les moteurs pourront toujours comprendre les anciens systèmes de métadonnées, il est recommandé de ne conserver que schema.org pour ne pas semer de confusion.

Les spécialistes auront à suivre l'évolution de cette nouvelle divergence sur le Web.

#### 5.4.7 Des vocabulaires, des classifications par secteur et par métier

Que ce soit pour les données, les formats, les métadonnées, les méthodes,

un très grand nombre de standards et de normes sont développés par des communautés sectorielles ou métier. La diversité des données publiques exige donc rapidement de se relier à de telles communautés pour se conformer aux standards mais aussi les faire évoluer si nécessaire.

Il y a souvent une contradiction entre la volonté de suivre des standards internationaux d'une part, et de pleinement décrire les spécificités d'un territoire d'autre part. C'est déjà vrai pour la relation entre les normes et standards mondiaux et les situations des différents pays, notamment pour se conformer aux législations. La seule voie est de se grouper pour enrichir les normes nationales puis mondiales, pour définir des volets obligatoires et des volets optionnels afin de permettre l'adaptabilité de la norme. L'effort dans ce domaine en France a souvent été notoirement insuffisant, notamment dans les TIC.

Un exemple issu de la santé illustre la vigilance nécessaire. Une norme mondiale essentielle proposée par les américains comprenait obligatoirement l'origine ethnique... (et d'ailleurs listait les tribus indiennes) dans la liste «par défaut» de la spécification.

Les communautés métier capables de définir concrètement les standards existent dans de nombreux secteurs. Il faut prendre garde au fait que les mêmes mots peuvent avoir des sens différents pour plusieurs communautés mais aussi, plus profondément, que le même objet peut être vu avec des points de vue également très différents. Chaque métier et chaque domaine ont leurs outils de classification – par exemple les catégories administratives pour décrire un logement ne sont pas les mêmes que celles des architectes, des sociologues ou des historiens, mêmes s'ils partagent certains concepts : toute ontologie, tout vocabulaire ou nomenclature doit préciser quel est son périmètre et quelle entité la définit. S'ajoutent à ces problèmes les questions de langue. Tout cela peut paraître impossible à gérer. Pourtant, la vie sociale et les échanges, y compris internationaux, ont créé une grande quantité d'outils. L'exemple le plus manifeste est celui des douanes qui, par

nécessité, ont créé de longue date un vocabulaire commun pour décrire les produits et leurs caractéristiques – du point de vue d'un douanier.

L'importance des standards et normes a été réaffirmée par le GFII, qui recommande «*l'élaboration, l'adoption et la mise en œuvre de standards de métadonnées et données dans toutes les thématiques métiers (juridique, économique, SHS,...) Ces standards doivent être fondés sur les normes déjà largement admises comme DublinCore, ISO et AFNOR et s'inscrire dans le cadre du référentiel d'interopérabilité...Ces normes et standards doivent être établies par les représentants des filières métier, avec la participation des éditeurs publics concernés, et en prenant en compte la définition de normes similaires au niveau européen.*»

#### 5.4.8 Des normes dans les domaines de coopération et mutualisation

Plusieurs grands domaines de coopération sont engagés dans des démarches importantes de normalisation. Ils ne peuvent exister sans elle.

##### ■ *Remarques sur la géographie, chantier national et international essentiel*

Il ne s'agit pas ici d'entrer dans le monde des SIG et des normes de l'information géographique, domaine des géomaticiens, mais seulement d'en évoquer la complexité mais aussi les ressources que la normalisation apporte ici à l'ensemble des données publiques.

Les données géographiques ont fait l'objet depuis longtemps de travaux de normalisation, que la directive Inspire et la construction de plates-formes de données géolocalisées ont encore stimulés et étendus. Un vaste ensemble de normes est respecté par les spécialistes du domaine, avec une coordination en France par le Conseil national de l'information géographique, de l'IGN, du

□ BRGM, au niveau mondial par l'ISO et l'Open Geospatial Consortium. Chacun connaît par ailleurs les bases de Google, d'OpenStreetMap, de l'IGN et les outils permettant d'intégrer des ressources dans un site ou un service.

Pour ceux qui ne sont pas géomaticiens, c'est-à-dire notamment pour les responsables métier qui sont conduits à placer leurs données sur des cartes, l'important est la relation avec les normes utilisées pour les métadonnées géographiques, car celles-ci concernent non seulement les objets internes de l'information géographique (localisation, parcelles, altitudes etc.) mais aussi tous les objets qui peuvent être affichés sur les cartes (domaine santé, transports, réseaux d'énergie, espèces animales etc.). Dès lors, ces normes fournissent un point de départ important pour la gestion générale des métadonnées publiques. Les différents systèmes géographiques respectent la norme ISO 19115, puis, au niveau européen, la directive Inspire, dont les annexes détaillent les 34 thèmes pour lesquels il est demandé de respecter des normes. L'acteur français central est l'IGN avec son Géocatalogue.

D'autres problèmes importants sont traités par l'information géographique, notamment les toponymies.

La gestion de noms pour les localisations et les territoires doit répondre à des normes, compte tenu des diverses toponymies et de l'emboîtement des territoires. Elle doit aussi permettre, l'utilisation de plusieurs langues ou le respect des dénominations administratives comme des habitudes locales, par exemple. Pour qu'une question portant sur le Massif central, les Cévennes, le Lake district ou le pays de Gex, sur Aix la Chapelle ou Aachen retourne la carte ou la photo aérienne idoine, il faut se référer à un index toponymique (en outre des noms du passé, utilisés par les archéologues ou historiens, doivent aussi pouvoir être localisés). Le service mondial GeoNames contient plus de 10 millions de noms de lieux avec leurs caractéristiques et les liens. Il permet de lier nom, adresse, coordonnées etc. GeoNames est un gazetier coopératif ; c'est un service gratuit.



De son côté, l'IGN possède sur la France une base de données administratives officielle - BDAAdmin.

#### ■ *Statistiques et données économiques et sociales*

Pour comparer les données, pour les regrouper au sein d'un territoire et en général les rendre accessibles, des initiatives internationales avaient conduit d'abord à des échanges sur les réseaux en mode transfert EDI. Depuis 2001, l'ensemble de normes SDMX (Statistical Data and Metadata eXchange) est porté par plusieurs organismes internationaux : BCE, OCDE, Eurostat, FMI, ONU, Banque mondiale, BRI. Le reproche à ce type de normes est qu'elles sont complexes à mettre en œuvre. En pratique, la transmission la plus courante se fait par l'envoi de fichiers xls et ne permet pas la manipulation à distance. On a vu que des services proposent celle-ci (Google Public Data Explorer). Cependant, le développement des publications de données publiques et de leur réutilisation va stimuler une évolution des normes et l'adoption de plusieurs niveaux de standards (à suivre avec l'INSEE et les observatoires économiques en particulier).

Dans un premier temps, la première tâche pour les données économiques et sociales et les statistiques est de préciser clairement les définitions et les périmètres – ce n'est pas toujours le cas pour les données trouvées sur le Web.

#### ■ *Le tourisme et TourinFrance*

Le ministère du tourisme – maintenant les services en charge du tourisme au ministère de l'Économie – a mis au point avec les professionnels la norme TourinFrance.

TourinFrance permet de définir les conditions d'échange de données entre deux Systèmes d'Information Touristique (S.I.T). TourinFrance comporte

deux volets :

« Un Format Commun de Description des Données Touristiques qui rassemble treize bordereaux décrivant des objets touristiques (hôtels, musées, restauration...).

Un Format Commun d'Echange qui organise les conditions de l'échange entre deux bases de données hétérogènes et sans «lien fixe». »

## 5.5 Répertoires

Des répertoires apparaissent, sur les sites des administrations et organismes nationaux et sur ceux des collectivités et autres acteurs publics qui ont décidé d'ouvrir des données. Cependant, ces répertoires sont seulement la présentation des fichiers accessibles à la réutilisation, avec les conditions. Ils ne recensent pas les principaux jeux de données accessibles et qui représentent un vaste gisement. (*Les répertoires ont été commentés dans le Chapitre 2*).

La création de répertoires efficaces, lorsque beaucoup de jeux de données seront libérés par de multiples acteurs, supposera de créer des répertoires régionaux ou thématiques.

Un répertoire doit permettre d'identifier des jeux de données proches, parlant de la même chose. Il doit être capable de répondre à des requêtes sophistiquées sur l'ensemble des données. Il devra utiliser un moteur capable d'extraire les méta-données associées à des fichiers de sources diverses. C'est pourquoi la mutualisation est incontournable, et peut-être aussi des partenariats.

### 5.5.1 Trois répertoires nationaux

Le développement de répertoires nationaux est à l'ordre du jour. Ils assurent

la visibilité des jeux de données et en promeuvent l'utilisation. En l'absence de moteurs capables d'interroger efficacement de multiples entrepôts, ils permettent à l'utilisateur comme au réutilisateur de trouver les ensembles de données qui peuvent les intéresser.

Trois démarches principales en France visent des objectifs et des publics différents :

- Etalab est le portail national de l'administration d'Etat. Il pourra cependant accueillir des collectivités intéressées. Sa première mission est l'évangélisation des administrations et la promotion d'une cohérence dans les démarches.
- NosDonnées.fr, lancé par Regards Citoyens, vise à « *apporter plus de visibilité aux données publiques librement accessibles à chacun.* » C'est un projet axés sur la transparence et sur la collaboration : une «chasse au trésor» doit permettre d'informer le site sur les gisements de données et de proposer leur inscription dans le portail.
- Data Publica vise à devenir une place de marché pour les données en général et met l'accent sur la sémantique et le traitement des données. Aussi est-il intéressant d'indiquer quelques unes des fonctionnalités annoncées par Data Publica (*dont le développement et les objectifs sont indiqués au chapitre 4 – Economie*).

### 5.5.2 Data Publica

Data Publica héberge les jeux de données et son principe est de prendre en charge un maximum de fonctions, le producteur des données fournissant des fichiers bruts. Comme la société vise aussi le marché des PMI – en tant qu'offres ("éditeurs") et utilisateurs -, elle doit veiller à préserver une simplicité pour tous les partenaires. Les mises à jour des données évolutives sont automatiques.

La plate-forme stocke les données et les métadonnées fournies par le producteur ou créées par elle et elle les indexe à destination de son moteur interne. En même temps, des outils assurent automatiquement des traitements : mises au format, enrichissement par ajout de métadonnées, qualification. Les recherches doivent permettre de trouver tous les fichiers, quelle que soit leur origine, traitant d'un objet particulier : l'emploi, les platanes, un village... et produisant pour un utilisateur des réponses organisées en fonction de la requête. Enfin, Data Publica assurera le transfert de fichiers utilisables par l'entreprise connectée ou l'intégration des divers types de données au sein de services clients.

Le projet de Data Publica est techniquement très avancé. Il est porté par un ensemble d'acteurs qui entendent en faire un véritable outil du Web sémantique.

## 6 Le rôle de la Région et l'organisation

Le présent chapitre ne propose pas une démarche mais vise à resituer la Région par rapport à l'ouverture et à la réutilisation des données publiques.

La Région est considérée ici d'un double point de vue : comme elle-même productrice et éditrice de données d'une part, comme devant s'interroger sur son rôle dans le mouvement en relation avec les autres acteurs régionaux d'autre part. En effet, les gisements de données des collectivités et des acteurs publics vont jouer, on l'a vu, un rôle important dans l'aménagement numérique du territoire et dans le développement de son attractivité.

Le niveau régional est celui où se réorganisent aujourd'hui beaucoup de structures de l'Etat et c'est aussi à ce niveau que se mettent en place de nombreux observatoires et des plates-formes aussi essentielles que celles des centres d'information géographique.

Une coordination et une gouvernance seront en tout état de cause nécessaires au niveau régional. Il ne s'agit pas ici de prendre position sur la manière dont elle va se construire. La Région a certainement un rôle essentiel à y jouer.

### 6.1 Données de la Région et données régionales

#### 6.1.1 Position de la Région au centre de masses de données hétérogènes

##### ■ *Moyens et politique de coordination*

La Région a des moyens internes limités du point de vue technique. L'exercice des compétences donne souvent lieu en pratique à des délégations, à la création de structures coopératives ou à des sous-traitances et les

□ opérateurs sont parfois devenus incontournables. Les partenaires publics, les administrations déconcentrées, le rectorat, les observatoires contrôlent d'importants gisements de données.

Cette multiplicité d'acteurs est vue par certains comme une faiblesse. Elle peut au contraire être une opportunité. Elle s'inscrit dans un rôle de coordinateur et de pilote et les structures auxquelles sont déléguées des fonctions sont le plus souvent un lieu de rencontre et de mutualisation.

Une stratégie open data régionale mettra souvent en jeu plusieurs partenaires dès le départ. C'est autour de politiques qu'elle peut s'organiser.

Néanmoins, il faut aussi que la Région montre l'exemple sur les données qu'elle collecte, contrôle et produit elle-même directement. Elle va donc s'interroger sur ses gisements internes, que ses services utilisent ou que d'autres réclament.

Outre son action sur ses domaines propres de compétence et ses politiques, la Région peut être incitatrice, coordinatrice, support des infrastructures et équipements qui vont être nécessaires.

La Région joue un rôle central dans l'aménagement du territoire et l'aménagement numérique du territoire. Elle va devoir veiller à ce que la libération des données n'aggrave pas les inégalités entre les différentes zones, et notamment entre les grandes agglomérations et les autres territoires.

##### ■ *Une grande diversité de l'information régionale et de ses acteurs*

L'échelon régional est le lieu de la collecte et de la production de très nombreuses informations de base, en particulier dans le domaine des statistiques économiques et sociales, par de multiples organismes, dont la coordination est au mieux faible, au pire inexistante – même s'agissant de services de l'Etat. Le résultat est qu'il y a beaucoup de données disponibles,

certains déjà publiques, mais sans cohérence forte – y compris sur les définitions, les périmètres etc.

Comme Sylvain Maire, délégué de la FING pour assister la Région Aquitaine le soulignait sur le blog de la FING :

*“En effet, comme on peut le voir sur l’annuaire des données publiques françaises de Data Publica, la plus grande partie des données faisant références aux régions sont produite par d’autres organisme, tels que l’INSEE, le Service de l’observation et des statistiques (SOeS, <http://www.statistiques.developpement-durable.gouv.fr/>), etc. Le seul jeu de données dans l’annuaire de Data Publica provient du Conseil Régional Rhône Alpes avec la liste des comités et ligues de sport et loisirs dans la région Rhône Alpes (<http://www.data-publica.com/dataset/364>).*

Les chantiers qui vont s’ouvrir vont donc concerner un vaste ensemble d’acteurs régionaux, publics ou privés mais jouant un rôle dans les services publics ou en général dans les services à la population et aux entreprises : services déconcentrés et décentralisés de l’Etat, chambres consulaires et groupements professionnels, observatoires, plates-formes coopératives.

#### ■ *Pression faible et stratégie proactive : intérêt de prendre la main*

Il y a encore peu de demandes présentées aux régions, contrairement à ce qui fut le cas pour les grandes administrations et organismes nationaux. Les poussées pour l’open data viennent d’associations et d’acteurs militants et, dans de rares cas, de quelques sociétés (Transpacités, Généalogie) qui ont un modèle économique, ou de jeunes start up et petites sociétés qui visent des niches. Les nouvelles demandes sont d’abord tournées vers les grandes agglomérations.

La raison principale est celle qui vient d’être évoquée. La Région ne collecte ou publie directement que peu de données et l’axe principal de son activité

est la coordination et le développement.

Pourtant, l’exploitation de données régionales cohérentes constituerait certainement un apport important pour la région, notamment dans le domaine du développement, de l’aménagement du territoire et de son attractivité. C’est précisément en associant des données de divers producteurs que peuvent être développés services et informations utiles. Il va donc être indispensable d’identifier les domaines stratégiques et les données essentielles et de coordonner les acteurs régionaux autour d’objectifs. Au delà du travail prioritaire sur ses propres données, la Région sera conduite à prendre la main sur les dossiers prioritaires.

#### 6.1.2 Des axes et des principes

Quelques principes vont encadrer l’action. Ils dérivent de l’égalité d’accès aux services et de l’objectif d’aménagement du territoire régional.

##### ■ *Garantie du service public, accès des citoyens*

Comme cela a été souligné dans d’autres chapitres, la fourniture de services performants aux citoyens et aux entreprises par des acteurs privés pourrait conduire à l’affaiblissement du service public et à une inégalité d’accès des citoyens sur le territoire. Les différents acteurs concernés, notamment les collectivités, doivent se coordonner pour distinguer les missions de service public pour lesquelles la continuité et l’égalité d’accès doivent être garantis.

##### ■ *Aménagement du territoire et réduction des inégalités*

Le développement de l’open data dans les grandes agglomérations rappelle ce qui a pu se passer ou se passe encore, par exemple, pour le développement des réseaux de télécommunication, depuis le Minitel jusqu’au mobile et au

haut débit dans des périodes récentes. Le rôle de la Région va être essentiel pour éviter que les différences ne s'accroissent entre les territoires, d'abord entre les grandes agglomérations et les autres, mais aussi entre les différents territoires. C'est là une tâche politique difficile, voire un défi puisque les différents territoires sont aussi en concurrence, mais c'est précisément le rôle de la Région que de prendre en compte les différents intérêts.

La mutualisation des services techniques doit permettre à la fois de faire bénéficier tout le territoire des avancées des grandes agglomérations et de réduire des coûts que les collectivités modestes ne pourraient pas supporter.

*On peut cependant remarquer que la faiblesse des moyens est parfois compensée par le lien social plus fort susceptible d'exister au sein de petites communautés. Les carto-parties de Plouarzel (3000 ha) en sont une illustration (cf. Ch. 2). La mutualisation peut permettre parfois des transferts de compétences du petit vers le gros...*

#### ■ Normalisation

On l'a vu notamment dans le chapitre technique (Ch. 5), la normalisation des données et des protocoles est indispensable à terme. Elle est aussi coûteuse à court terme, et exige une veille au niveau national et international. La Région se doit d'en être le relais et le coordonnateur sur son territoire, et elle sera conduite à intervenir aux différents niveaux sur des domaines essentiels pour son territoire, en fonction des caractéristiques de celui-ci, en particulier dans les domaines agricole et environnemental.

#### 6.1.3 Les données des régions

##### ■ Des données dans les principaux domaines de compétence

Pour montrer l'exemple, le Conseil régional se doit d'identifier d'abord les données qu'il détient et pourrait ouvrir et qui, souvent – comme dans toutes

les structures – dorment après leur collecte et leur utilisation (formulaires divers, contrats etc.). Les gisements vont varier selon les régions.

Cependant, de grands domaines sont communs, même si les données elles-mêmes et leurs modalités de gestion vont varier. Ces domaines se recoupent souvent et sont en tout cas très liés.

- Importance de l'information géographique comme dénominateur commun
- Aménagement du territoire et aménagement numérique (en partenariat avec les Départements). Le rôle des Régions tend à se renforcer dans le haut débit et le très haut débit – avec, en ce qui concerne les données, des relations parfois difficiles avec les opérateurs
- Environnement, développement durable et biodiversité
- Attractivité régionale, tourisme
- Développement économique et emploi
- Transports (au moins ferroviaire mais la coordination avec les autres modes sera incontournable – cf. chapitres précédents, notamment 2)
- Lycées et apprentissage
- Formation professionnelle.
- Culture (musées régionaux, manifestations etc.)

##### ■ Des outils collectifs déjà en place

Les Régions mettent en place des schémas directeurs dans les domaines principaux de compétences – Schéma régional de développement économique ; Schéma régional de cohérence écologique ; Schéma régional de l'éducation et de la formation. Par ailleurs, la Région est associée à la préparation de schémas régionaux par des administrations et organismes d'Etat - comme en particulier l'Agence régionale de santé pour les Schémas

régionaux d'organisation des soins et d'organisation médico-sociale (L'Agence est un Etablissement public administratif).

Ces documents cadre n'incluent pas en général, actuellement, des précisions sur la gestion des données correspondantes – hormis l'exemple fondamental des données environnementales prises en charge conformément à la directive Inspire. Ils sont cependant déjà une source de données que les Régions peuvent utiliser et ouvrir et ils sont un cadre de coopération avec l'Etat et les autres partenaires régionaux, qui intégrera progressivement la gestion des données.

Comme cela a été vu dans les chapitres précédents, d'autres structures collectives sont des lieux de coordination pour la gestion des données publiques – géographie, tourisme, environnement, transport en particulier.

Enfin, les nombreux observatoires sont par définition détenteurs et diffuseurs de données. Ils font partie des acteurs importants pour chaque domaine d'action.

*Le problème, déjà abordé au chapitre 4 par ses aspects juridiques, est que la multiplicité des acteurs conduit à créer des plates-formes coopératives dans lesquelles il n'est guère possible à ce stade de gérer l'ouverture des données publiques. Ces plates-formes mettent en place des sites de publication mais jamais des espaces d'ouverture des fichiers de données. Ceux-ci ne leur appartiennent pas. Au contraire, les plates-formes se constituent comme diffuseurs des données publiques. Un acteur économique – mais aussi une association – qui souhaite fournir un service utilisant des données de plusieurs acteurs et de plusieurs régions est confronté à une multiplicité de démarches. Sur des domaines clé, les Régions auront là un rôle important de coordination interne et interrégionale.*

*Même si le statut et les mécanismes d'accès aux données directement collectées et gérées par les organismes régionaux coopératifs ne permettent pas à ceux-ci de développer une politique d'ouverture, il faut s'interroger sur la possibilité de*

*faire figurer les jeux de données dans les répertoires régionaux ou spécialisés qui vont se constituer – ces répertoires sont censés lister toutes les séries de données qui sont de nature publique, même si elles ne sont pas accessibles sans démarche particulière.*

#### ■ *Environnement : vers le développement durable*

Ce domaine devient un axe politique fort et les collectes de données se multiplient.

L'environnement a été le facteur de création ou de développement des plates-formes d'information géographique. L'émergence de la préoccupation environnementale a donné lieu à la création d'agences et d'observatoires régionaux. Depuis quelques années, dans la foulée du Grenelle de l'environnement, les collectivités lancent des programmes d'action conformes à l'Agenda 21 défini à Rio en 1992, appuyées sur des consultations de la population. En 1991, un décret du 19 juin précise "le contenu et les modalités d'élaboration du rapport sur le développement durable que les collectivités territoriales doivent réaliser (texte pris en application de l'article 255 de la Loi Grenelle2 dont les dispositions ont été introduites dans le Code général des collectivités territoriales).

Les projets Agenda 21 sont financés par l'Etat et la Région dans le cadre des contrats de projet, des contrats de pays ou des contrats d'agglomération et par des fonds européens.

Au total, par l'ensemble de ces actions, les Régions disposent d'ores et déjà de gisements d'information, par exemple sur les parcs naturels, directement ou, ici encore, au travers du Syndicat mixte qui gère le parc.

Au niveau de la région, l'ensemble des informations va devenir impressionnant, d'autant que de multiples associations contribuent à leur collecte. Il est donc probable qu'il faut saisir l'opportunité d'actions particulières pour ouvrir des données et pour coordonner des acteurs sur un même thème.

L'évolution actuelle tend à la création de sites multi-partenaires, dans lesquels interviennent les collectivités, les associations, les Agences diverses dont les Agences de bassin et la DREAL. Le problème, rappelé ci-dessus, est que ces sites ne peuvent pas aussi, en tout cas actuellement, être des portails de mise à disposition de données ouvertes – même, ils sont un argument pour ne communiquer les données que sous forme (technique et juridique) non réutilisable par des acteurs extérieurs. En outre, des rapprochements avec les autres régions peuvent être difficiles. Le potentiel est pourtant impressionnant, et des évolutions seront rapidement à envisager, au moins sur des thèmes consensuels.

Enfin, les synergies de données s'annoncent avec le passage du domaine environnement à celui, plus large, du développement durable. Comme cela a été exprimé à la réunion de l'Observatoire national des Agenda 21 locaux des 3 et 4 novembre 2010, la démarche est maintenant d'aborder la culture, la santé, l'emploi, le développement de l'économie sociale et solidaire, le lien intergénérationnel sous l'angle de leur articulation et leur intégration dans les Agenda 21 locaux.

**L'Agence régionale pour l'environnement de Midi-Pyrénées**, devenue Agence du développement durable en 2011, est l'une de ces structures de coopération porteuses d'une information publique très riche. (<http://www.arpe-mip.com>). Créée en 1991 par la Région avec les associations de protection de l'environnement, elle regroupe de nombreux acteurs. Elle a en charge l'Observatoire régional de l'énergie (créé par la Région et l'ADEME en 2003), elle a accompagné en 1995 la création du Parc naturel régional des Grands Causses. A partir de 2006, elle agit pour le développement et la reconnaissance des actions Agenda21. Elle est maintenant devenue Société publique locale, à côté de son activité associative.

L'exemple de <http://communes.bretagne-environnement.org> illustre à la fois le potentiel des données publiques sur l'environnement et le rôle de la Région. Le GIP Bretagne environnement associe l'Etat et la Région pour présenter les données relatives aux communes. Il "s'appuie sur les producteurs de données tels que la direction régionale de l'environnement, de l'aménagement et du logement (Dreal), les conseils généraux, l'Agence de l'eau, l'Ademe, le ministère de l'environnement, etc. mais aussi les associations comme Air Breizh ou Cohérence." La richesse du site est impressionnante mais les données sont protégées.

■ *Economie, formation, recherche : un ensemble important et fortement relié de données*

Les Régions mènent des politiques qui associent le développement économique, la formation et la recherche.

C'est pourquoi, par exemple, les Pays de la Loire substituent au Schéma régional du développement économique un "Schéma Régional de l'Economie et de l'Emploi durables qui intègre désormais les notions de formation continue et d'innovation».

Les Régions s'impliquent dans le développement de pôles de compétitivité et autres pôles d'innovation. Elles soutiennent les laboratoires de recherche et le développement des sociétés innovantes. Elles suivent les opérations engagées dans le cadre des Contrats de projets Etat Région. Les interventions se situent notamment dans le cadre de l'article 11 de la loi du 15 juillet 1982, qui leur confie la définition des pôles technologiques régionaux et des programmes pluriannuels, dans le cadre des CPER.

Les données et documents relatifs à ces actions représentent une source d'information importante sur les entreprises et sur la recherche. Leur exploitation par des acteurs français ou internationaux peut donner une plus grande visibilité aux laboratoires et entreprises. Leur ouverture suppose

cependant de prendre garde à respecter le secret industriel et commercial et les droits de propriété intellectuelle.

Le domaine de la formation professionnelle est à rapprocher de cet ensemble. Ici encore, des plates-formes se mettent en place, mais pas de services d'ouverture des données. Comme pour le développement économique, les actions sont menées avec les Chambres consulaires. Elles répondent notamment à un besoin crucial de rapprochement entre les besoins des entreprises et les offres de contrats d'apprentissage ou d'alternance (exemple: <http://www.poitou-charentes-alternance.fr/>)

#### ■ *Tourisme*

Les plates-formes régionales de tourisme ont notamment été évoquées au chapitre 4, en raison des problèmes particuliers qu'elles posent, à la charnière entre l'action collective pour l'attractivité du territoire et la promotion commerciale.

Précisément, comme le fait remarquer Joëlle Nonet, responsable du domaine au Conseil régional de La Réunion, la neutralité de la Région lui facilite la création de bases de données et de services utilisables par tous les intervenants. La Région est la seule à même de ne pas privilégier une zone, une filière, tout en laissant chacun, Office de tourisme, chaîne d'hôtels etc. exploiter au mieux, sur son site et dans ses offres, les données communes (cf. projet Soubik - <http://tic.regionreunion.com/-E-Tourisme/>)

#### ■ *Culture*

Le travail autour de données communes est un mécanisme d'apprentissage – et ce même si certaines données ne relèvent pas directement de la problématique classique des données publiques.

□ L'exemple de PACA, autour de «Marseille Capitale européenne de la culture» en 2013 est très intéressant (cf. Ch 7) : s'appuyer sur une opportunité consensuelle pour progresser, tout en maintenant par ailleurs la réflexion sur les différents types de données.

Ainsi, un objectif culturel commun – il n'en manque pas en France – lié à un événement, à une spécificité régionale etc. peut être un laboratoire pour ouvrir des données, associant les grandes agglomérations et les autres territoires, des collectivités, des institutions (y compris d'Etat) et des associations.

Par ailleurs, des chantiers importants sont déjà ouverts pour la mise en commun de ressources culturelles.

**La Banque numérique du savoir d'Aquitaine** a été créée par le Conseil général de la Gironde, dont le service des archives est le représentant français du projet européen EuropeanaLocal. Destinée au grand public, elle est soutenue par le Conseil régional. Elle donne accès aux ressources mises en ligne par les institutions culturelles (Services d'archives, musées, associations, bibliothèques, centres documentaires, éditeurs...). L'accès à la Banque et ses échanges avec les fournisseurs respectent le protocole de moissonnage de l'Open Archive (OAI – PMH) qui permet aux moteurs de récupérer les métadonnées des différents sites pour que l'internaute puisse obtenir un ensemble de documents intéressant sa recherche.

#### ■ *Responsabilité*

La question de la responsabilité reviendra longtemps dans les débats sur l'ouverture des données. Elle se pose particulièrement lorsqu'il s'agit pour un Conseil régional d'organiser, au sein d'une structure commune, une coopération avec de nombreux autres acteurs, dont les autres collectivités.



Comme on l'a déjà noté, même si, sur le plan formel, il n'y a pas de responsabilité à assumer, au delà de la précision des licences de réutilisation, il n'en reste pas moins qu'il peut sembler légitimement inquiétant de confier les fichiers de données brutes. Dans beaucoup de cas, les données source mêlent des informations non communicables aux informations accessibles. Pourtant, la publication sur un site Web implique que l'analyse a été faite pour certaines données. Il ne reste qu'à pouvoir fournir, si possible, des fichiers en format utilisable.

En tout état de cause, l'évolution sera progressive mais c'est bien au sein des organismes coopératifs que la question de l'ouverture de données publiques se posera principalement, au delà de la création de sites par ces organismes. Les Régions auront à coordonner les positions sur le sujet, ce qui est dans le prolongement naturel de leur action et de leur méthode.

#### 6.1.4 Action interrégionale, nationale, internationale

Un rôle particulier de la Région est de franchir la frontière régionale, voire nationale et de grouper les acteurs régionaux pour un ensemble d'actions.

- Coordination interrégionale, notamment avec les régions voisines
- Coordination face à l'Etat et, au côté des collectivités et autres structures concernées, négociation avec les intervenants nationaux comme l'IGN, l'INSEE etc.
- De même, coordination pour une position commune en face des grands opérateurs d'infrastructures – énergie, transports, télécom – avec une attention particulière aux zones limitrophes et aux opérations interrégionales. Par exemple, pour l'information transports, les frontières régionales ne doivent pas être infranchissables...
- Participation à des projets européens – voir par exemple la participation de la région PACA au programme MED – projets de

développement communs multi-régionaux sur une zone du rivage nord de la Méditerranée. (<http://www.programmemed.eu/>)

## 6.2 Aménagement numérique, relations entre les collectivités

### 6.2.1 Le problème des données concernant l'aménagement numérique lui-même

L'ouverture des données publiques joue un rôle dans l'aménagement numérique (*cf. Ch. 1*). Comme le notent les responsables des programmes d'ouverture, elle stimule le développement des jeunes pousses ou en général des industriels locaux des TIC. La mutualisation pour permettre à toutes les collectivités et aux autres acteurs publics de mettre en œuvre l'ouverture, la construction d'un répertoire, ou de répertoires régionaux, s'intègrent dans la politique d'aménagement numérique (*cf. infra*).

Cependant, il faut prendre garde à un paradoxe. En même temps que la circulation des données stimule l'aménagement, les évolutions des systèmes techniques tendent à gêner la transparence des informations sur les infrastructures et sur certaines données elles-mêmes.

En effet, la diversité des supports techniques et les concurrences entre opérateurs font cohabiter de multiples solutions en matière de télécommunications et de services. Dès lors, chaque acteur tend à considérer les données concernant ses réseaux et ses outils comme appartenant à son secret industriel. Par ailleurs, la connaissance des données peut et doit être de plus en plus fine : la régulation des réseaux nécessite de connaître les flux jusqu'à l'abonné et de disposer d'informations sur ses équipements. Ces informations sont tout aussi nécessaires à la maîtrise d'ouvrage ou lors d'appels d'offres, ou pour décider d'options en matière d'aménagement. En

cas de changement d'opérateur sur un secteur ou une technique, le capital informationnel, nécessaire pour assurer un service de qualité aux citoyens et aux entreprises, risque d'être perdu.

Ainsi, l'aménagement numérique, stimulé par le développement des nouveaux marchés de données, doit il de son côté être protégé contre une déperdition du contrôle par les acteurs publics et en particulier les collectivités.

Ce sera une fonction importante pour l'aménagement numérique du territoire et plus généralement l'aménagement que de rassembler les informations concernant infrastructures et usages et de veiller aussi à ce que des données acquises par des opérateurs ne soient pas utilisées de façon inappropriée ou dangereuse<sup>35</sup>.

*Remarque : les mêmes problèmes se posent pour d'autres infrastructures, toujours liés aux données et aux TIC – par exemple pour l'eau. On a vu précédemment (Ch. 5) que certains industriels pensaient qu'il fallait partager ces données.*

### 6.2.2 Mutualisation et réduction des disparités de moyens entre les territoires

La diffusion des données publiques s'intègre dans les stratégies de développement des services et industries du numérique et elle va s'appuyer sur les nombreuses actions collectives déjà coordonnées et financées par les Régions. La Région va donc jouer un rôle central dans la mutualisation.

Une réflexion commune doit permettre de définir des cadres pour les différents territoires – grandes agglomérations, zones urbaines, zones

<sup>35</sup> Google, qui n'est pas un opérateur (pas encore en France), s'est vu sanctionné par la CNIL pour avoir repéré l'implantation des routeurs WiFi, recueillant des données personnelles à l'insu des personnes concernées (mots de passe, courriels...)

rurales, zones de montagne etc. En même temps, les outils doivent fournir des normes pour que tous les territoires puissent s'inscrire dans des analyses communes et des services cohérents.

Plusieurs interventions, non exclusives, interviennent pour la mutualisation.

- Répertoire des données publiques. Comme cela a été noté ci-dessus, un tel répertoire peut indiquer les jeux de données accessibles et non accessibles. Ici aussi, il faut commencer par soi-même, mais il est immédiatement possible – et souhaitable – que les autres collectivités ou acteurs publics soient informés qu'ils peuvent y participer : soit en référençant des jeux de données sur le répertoire régional, soit en participant à une opération coopérative (cf. *Chapitre juridique les structures porteuses possibles*). De ce point de vue, il faut rappeler que les répertoires ne sont pas exclusifs : par exemple les communautés d'agglomération qui ont déjà créé les leurs peuvent aussi apparaître – directement ou par renvoi, question d'organisation sur le Web.
- Mutualisation de l'hébergement de jeux de données, en général sur le même site que le répertoire, qui devient un portail.
- Mise à disposition d'outils (SIG, banque de données urbaines...)

*La mutualisation est importante et il serait coûteux que trop de sites de référence se développent – ils ne pourraient de toutes façons pas survivre. Ceci étant, il ne faut pas non plus chercher l'unicité systématique d'un site de référence régional – il suffit que différents sites puissent dialoguer et respectent des normes communes. Cette remarque est aussi valable par domaines – il est possible que des groupes différents rassemblent chacun une partie des acteurs sur un thème. Il est aussi possible que des sites nationaux accueillent des informations régionales (c'est même souhaitable). Sur Internet, le contrôle de l'information est de plus en plus un mythe – l'important va être dans les usages et les services.*

Comme cela avait été évoqué dans le Chapitre 2, la Région, ou le groupement de collectivités qui établit une plate-forme de mutualisation peut s'appuyer sur l'article L14-25 du Code général des collectivités territoriales, à condition que cette extension à des services soit considérée comme concernant une infrastructure dans le cadre de l'aménagement numérique du territoire.

Il pourrait être pertinent d'inclure des plates-formes régionales de gestion de données et de signalement (répertoires) dans le Schéma directeur territorial d'aménagement numérique (SDTAN).

### 6.2.3 Fédération d'efforts et relations entre collectivités

Comme cela a été le cas pour les nombreuses plates-formes régionales, et en particulier la géographie, la mutualisation et plus généralement la politique régionale vont s'appuyer sur les acteurs qui ont déjà ouvert des chantiers. Ce serait une erreur d'opposer les développements menés dans les grandes agglomérations et les difficultés des autres niveaux de collectivité.

Compte tenu des situations actuelles, la coopération Région – grandes agglomérations sera souvent la base de la mutualisation. S'y ajoutent un certain nombre (encore limité) de Conseils généraux (cf. *infra Gironde*). Il faut partir des situations concrètes : préoccupations communes aux collectivités, initiatives des Conseils généraux sur les domaines de leurs compétences, où ils disposent directement de masses de données – domaine social en particulier, transports interurbains, infrastructures etc.

Le mouvement open data est en 2011 principalement actif dans les grandes agglomérations parce qu'il naît souvent autour de la mobilité dans la ville et s'appuie sur les riches données de proximité. La coopération doit permettre que ce mouvement tire les développements dans les autres territoires – avec des modalités différentes.

□ Dans de nombreux domaines, les coopérations entre collectivités se sont avérées fructueuses (ex des Espaces numériques de travail).

L'essentiel est que le plus grand nombre d'acteurs acceptent de coopérer, dans le respect de leurs compétences et, au départ, sur des projets d'intérêt commun.

*Il est indispensable qu'une gouvernance régionale se mette en place pour coordonner l'action des différentes collectivités et acteurs publics et piloter les projets communs. Une gouvernance régionale, mais pas une gouvernance de la Région.*

### 6.2.4 Une première expérience : Aquitaine et Gironde

La Région Aquitaine et le Département de la Gironde ont créé un premier portail de mise à disposition de données publiques (<http://datalocale.fr/>), encore au stade expérimental à l'été 2011.

Le projet est la construction d'un portail mutualisé de données locales. Les partenaires « font donc le choix d'une exploitation durable et valorisante de leurs ressources informationnelles, par la mise en place d'une plate-forme de publication mutualisée, en cherchant à maximiser leur ré-utilisabilité et leur accessibilité ».

Les données retenues concernent l'environnement, le transport, les données locales (équipements sportifs, déchets...)

### 6.2.5 Les plates-formes d'archivage

Cet autre aspect de la gestion des données publiques – documents et bases de données – est lié à la mise en disposition des données publiques. Ces acteurs gèreront pour les collectivités la mise à disposition des données archivées. L'archivage numérique mutualisé s'accompagne en outre d'un effort de normalisation sur les contenus et les méta-données.

Des centres de gestion de données sont en constitution, qui seront donc un opérateur de l'ouverture, en fonction des décisions des collectivités déposantes et des règles propres aux données. (Il est à souhaiter que la communicabilité des données ne soit pas différentes une fois qu'elles sont archivées (hors celles qui ne peuvent être communiquées qu'après un long délai – comme celles de l'état civil).

Ainsi, l'Agence landaise pour l'informatique (ALPI) est un syndicat mixte groupant plus de 500 collectivités, des établissements publics, une CCI, une Chambre des métiers. Elle développe le projet Archiland. (<http://www.alpi40.fr/article/archive/996/>)

### 6.3 Le rôle fédérateur par rapport à un vaste ensemble d'acteurs

#### 6.3.1 Le rôle fédérateur : collectivités, administrations déconcentrées, partenaires privés, réseaux d'infrastructure

Comme cela apparaît pour tous les types de données, de très nombreux acteurs collectent, produisent, diffusent des données sur les territoires de la Région, au delà des collectivités : directions et établissements publics d'Etat, acteurs et organismes délégataires de service public ou opérateurs sous contrat avec les collectivités, organisations privées, associations citoyennes.

C'est dans la coordination autour de projets régionaux que la Région va jouer un rôle essentiel.

La coordination de l'ensemble des collectivités pour négocier avec les acteurs régionaux, en particulier les administrations d'Etat et les opérateurs de réseaux, a déjà été évoquée. Il sera essentiel d'aller au delà et d'être proactif pour améliorer de façon significative l'information régionale. Comme on a pu le voir, c'est dans des projets relatifs à l'environnement, à la culture, à la

santé que des coopérations peuvent se développer, qui s'étendront ensuite à d'autres domaines.

Cela implique que, pour les actions communes déjà courantes, les services de la Région veillent à préciser avec leurs partenaires le statut des données qui seront produites. Cette nécessité, déjà exprimée pour les contrats de délégation de service public, s'étend à toutes les constructions communes, en particulier aux Schémas régionaux de toutes natures que la Région coordonne ou à la conception desquels elle est associée (comme pour l'organisation sanitaire et médico-sociale) : schémas d'aménagement du territoire, de développement économique, de cohérence écologique, de formation professionnelle, d'organisation des transports, d'aménagement numérique...

#### 6.3.2 Les chambres consulaires

Etablissements publics d'Etat, ayant des missions de service public, les chambres sont un lieu important de collecte d'information économique. Elles disposent de fichiers d'entreprises, qu'elles collectent dans le cadre de leur mission de service public (mission financée par la taxe parafiscale qui est leur ressource). La question des fichiers est sensible, et il faut tenir compte des intérêts de chacun – le problème est d'ailleurs régulièrement posé au niveau national, des sociétés privées souhaitant pouvoir les réutiliser comme une donnée publique (notamment pour l'annuaire des entreprises de France).

Les chambres mènent souvent des actions d'information spécifiques dans le domaine économique et il apparaît essentiel qu'au moins un premier niveau d'interopérabilité entre les données des uns et des autres soit assuré au niveau régional.

Des exemples :

**L'Assemblée pyrénéenne d'économie montagnarde (APEM)** est une organisation "interconsulaire, interdépartementale, interrégionale,

internationale puisqu'elle concerne le massif des Pyrénées". Elle regroupe les Chambres consulaires et d'autres acteurs privés (mais ayant certaines délégations de service public). L'APEM travaille sur les enjeux des différents territoires en les spatialisant (cf. ci-dessus ch. 2 et 3). Chaque organisme membre détermine le statut des données auxquelles il contribue. L'APEM coopère étroitement avec les plateformes géographiques d'Aquitaine et Midi-Pyrénées.

#### **BASECO Bourgogne**

[http://www.basecobourgogne.fr/crci\\_b/index.asp](http://www.basecobourgogne.fr/crci_b/index.asp)

Soutenues par un financement du Conseil régional et du Conseil général de la Nièvre, les C(R)CI de Bourgogne se sont groupées pour développer et optimiser l'information économique et territoriale en Bourgogne, en créant Baseco Bourgogne. Des partenariats ont été signés avec le Rectorat, l'ANPE, la Préfecture, le TPG et l'ensemble des services déconcentrés de l'Etat

Ce pôle d'information est essentiel et c'est à son niveau que des problèmes d'accès aux données publiques se poseront.

#### **Midi-Pyrénées Economie**

<http://www.e-midipyrenees.net/>

La CRCI, avec le soutien de l'Etat, du Conseil Régional, de l'Union Européenne, s'est engagée avec les 10 CCI territoriales dans la création d'un portail sur la thématique entreprise/économie.

Dans de nombreuses régions, les Chambres de commerce en particulier développent ainsi des portails d'information économique. Les exemples ci-dessus les concernent, mais des coopérations existent avec les Chambres d'agriculture et les Chambres des métiers, elles aussi collectrices et productrices d'informations accessibles sur les portails régionaux et départementaux.

□ Une question se posera sur le statut des données et documents (cf *Chapitre sur les questions juridiques de la réutilisation*). En effet, à la différence de la tenue à jour des fichiers consulaires, les données en cause ne sont pas clairement collectées dans le cadre d'une mission de service public. Cependant, et même si ce n'est pas la base de la définition, le statut des Chambres et le fait qu'elles soient financées par une taxe parafiscale sont souvent utilisés comme argument pour un droit à réutilisation. En fait, l'essentiel est d'une part dans la position des Chambres sur l'intérêt pour l'économie régionale de la réutilisation, d'autre part dans d'éventuelles conditions qui seraient mises par les collectivités au financement des systèmes ainsi développés.

Au niveau régional, une politique par rapport aux données économiques doit être définie avec les Chambres.

#### **6.3.3 Les observatoires**

Un nombre croissant d'observatoires se consacrent dans les régions à un secteur ou une fonction particulière. Ils assurent une concentration et des rapprochements de données.

**L'Observatoire régional économique et social des Pays de la Loire** a lui-même repéré cette multiplicité d'intervenants.

(<http://ores.paysdelaloire.fr/1006-les-observatoires-en-region.htm>)

"La fonction d'observation en région est assurée par différents organismes, généralement spécialisés dans un secteur, un territoire ou une fonction.

Les services de l'Etat produisent et diffusent les données relatives à leur champ d'intervention (transport et construction pour la DREAL, commerce extérieur pour la DRCE...), les Chambres consulaires publient des données sur leurs ressortissants, les observatoires sectoriels sur leur

domaine (emploi et formation pour le Carif-Oref.<sup>36</sup>..), les observatoires territoriaux sur leur territoire en fonction de la collectivité locale dont ils dépendent. L'Insee occupe une place spécifique par son rôle de maître d'ouvrage du recensement de la population, la gestion de l'état-civil, du répertoire des entreprises, des nombreuses enquêtes menées chaque année, etc.

Il y a au total une trentaine d'organismes assurant une fonction d'observation de ce type dans la région.

L'Ores est un observatoire territorial multi-secteurs au niveau de la région, les autres observatoires territoriaux couvrant des territoires infra régionaux.

Le groupe de travail des observatoires en région a été mis en place à la création de l'Ores et propose aux observatoires, trois à quatre fois par an, des rencontres sur des thèmes communs, comme les indicateurs de développement durable, les indicateurs synthétiques, l'exploitation des données du recensement (en collaboration avec l'Insee), etc.”

**Les observatoires sont par nature des collecteurs de données.** L'Ores indique ainsi que la collecte des données externes se fait en relation avec plus de 35 organismes...services extérieurs de l'État, établissements publics tels l'Insee, la Banque de France, l'Ademe, les observatoires sectoriels (Emploi-formation, Santé, Transports, Bâtiment, Tourisme), les Chambres consulaires, des organismes de gestion de la formation et le Pôle emploi, EDF, et de nombreux organismes techniques spécialisés.

Les Régions jouent pratiquement toujours un rôle important, compte tenu de leurs compétence d'aménagement du territoire, de développement économique et de formation. Les Conseils régionaux sont fondateurs, ou

36 Centre d'Animation, de Ressources et d'Information sur la Formation - Observatoire Régional Emploi Formation

financeurs, ou partenaires. Les statuts varient, dépendant de la nature des partenaires – ainsi, en Poitou-Charentes, l'Observatoire de l'environnement est une association, qui intègre elle-même plusieurs associations. En revanche, l'Institut atlantique d'aménagement du territoire (IAAT), centre de ressources régional en information territoriale, est une régie à personnalité morale et à autonomie financière créée par délibération du Conseil régional, qui assure un service public d'information, avec un très grand nombre de partenaires.

#### 6.3.4 Une grande masse de données et le risque de silos

Au total :

- - Des domaines fondamentaux et qui relèvent essentiellement des compétences des Conseils régionaux donnent lieu à la création de sites et portails d'information – c'est le cas pour l'aménagement du territoire et l'environnement (observatoires de l'environnement, agences de l'eau...), l'emploi et la formation avec les sites des Carif-Oref, les observatoires économiques, les observatoires du tourisme, de la santé, des transports, ceux consacrés à un secteur important pour la région, les observatoires des agences d'urbanisme et des comités d'expansion sur des territoires etc.
- - Une masse impressionnante de données sont collectées et circulent au sein des territoires régionaux. C'est bien sur ces données que se jouera une politique de diffusion mais aussi de réutilisation.

Or, des risques existent.

Les remarques de l'ORES Pays de la Loire montrent que les acteurs de l'information ressentent le besoin de se rapprocher pour traiter de problèmes transverses – il faut rappeler que les données en général, et les données publiques en particulier, gagnent à être rapprochées et que de nombreuses

synergies sont possibles. Or, la fragmentation entre de nombreux acteurs peut avoir des effets négatifs, avec en particulier la possibilité de périmètres différents, de définitions différentes, de non interopérabilité technique mais aussi sémantique et pratique.

Cependant, le fait que, comme les plates-formes régionales (géographie, services, archives), les observatoires se constituent en rassemblant des données de nombreux acteurs est un signe positif sur la capacité à partager et à interopérer. Pour aller plus loin, il convient toujours d'aborder la réutilisation possible à partir d'un projet commun, d'un événement, d'une opportunité.

La question des données publiques se posera aussi parce que des acteurs nationaux ou internationaux souhaiteront intégrer les données dans une approche plus large. Il est certainement de l'intérêt des Régions d'éviter que ne se développent des silos de données trop hermétiques, tant à l'intérieur que dans la perspective d'une visibilité nationale et internationale.

## 6.4 Services et responsables dans la collectivité, ou avec la collectivité

### 6.4.1 L'organisation dans le Conseil régional

#### ■ *Le Conseil, les élus*

En premier lieu, comme pour toute action de la collectivité, une politique générale de gestion et de réutilisation des données doit être définie par le Conseil régional. Pour cela, il est certainement important qu'un élu en soit responsable, puisse interpréter en termes politiques les opportunités et difficultés qui se présentent et proposer au Conseil une heuristique et une stratégie.

□ Cependant, comme cela a pu être noté, le développement concret d'espaces d'ouverture des données va se faire par domaines et secteurs plutôt qu'abstraitement. C'est pourquoi de nombreux dossiers vont être préparés en commun avec les conseillers en charge de l'économie, des transports, de la formation, de l'aménagement numérique etc. Néanmoins, il est effectivement utile qu'un conseiller assure une cohérence stratégique et maîtrise en particulier les principales questions juridiques et politiques que pose la réutilisation.

Quelle que soit l'orientation adoptée par le Conseil régional, elle devrait d'abord s'exprimer à propos des données directement contrôlées par lui. Sur ce plan, le Conseil régional se trouve dans la même situation que les agglomérations qui ont abordé la question des données publiques dans une délibération. Une action particulièrement marquante peut faire l'objet d'une délibération : cela a été le cas en Aquitaine, où le Président de Région a proposé au Conseil la signature d'une convention avec la FING pour mener des expériences d'ouverture des données publiques (délibération du 18 octobre 2010). Dans d'autres cas, le Président de Région a présenté une proposition au Conseil économique, social et environnemental : c'est le cas en Bretagne où la proposition a été incluse dans le volet aménagement numérique. La libération des données publiques y figure parmi 10 axes de travail (Très haut débit, espace de connaissance, GéoBretagne...). Comme on le verra (Ch 7), la Région PACA, qui aborde les données publiques par le projet Marseille-Provence 2013 (capitale de la culture) envisage d'élargir la démarche et de proposer une délibération sur la politique de libération des données publiques en fin d'année 2011..

*Plus encore que dans les villes, où la question de l'information locale avait largement été débattue, il est utile de disposer d'un premier dossier avant de proposer une formulation stratégique en Conseil.*

## ■ Les services responsables

La libération des données publiques est un problème transverse, qui va concerner à un moment ou à un autre tous les services de la collectivité. Cependant, il est indispensable qu'une personne coordonne la démarche, nourrisse la stratégie d'ouverture et veille à la cohérence des approches dans les différents services et avec les organismes liés au Conseil régional et les partenaires extérieurs.

Sur le fond, l'important est que ce responsable ait des qualités de négociation et de compréhension des difficultés et des résistances ou des attentes des uns et des autres, puisque tout dossier de ce type va impliquer plusieurs services. Il doit aussi être bien placé pour dialoguer avec des acteurs divers, de l'administration d'Etat, du monde associatif, des entreprises - association citoyenne, start-up innovante, opérateur de télécom etc.

Aussi, l'essentiel est la personne, au delà du service dans lequel il se trouve. Ceci étant, du point de vue stratégique, la libération des données ne doit pas apparaître comme avant tout un problème de droit ou de format de données, mais comme un facteur de développement économique et d'innovation. Aussi, le responsable coordonnateur gagnera à se trouver dans un service compétent sur l'économie ou l'aménagement du territoire.

Dans de nombreux cas, cependant, dans les collectivités qui ont commencé de mettre en œuvre une politique, le responsable est situé à la DSI. En même temps, il constitue de toutes façons une équipe pluri-disciplinaire et travaille avec les différents services. Le fait de le situer au sein d'un service économie ou aménagement est avant tout un signal interne et externe pour insister sur l'impact réel de l'ouverture et marquer l'importance du dossier.

### Quelques situations

En **Aquitaine**, le responsable en charge, Michel Einer, est délégué régional TIC.

En **Bretagne**, Thomas Renault est chef de service du développement numérique.

En **PACA**, Stéphane Martayan est responsable "Economie et innovation numérique" au sein de la direction Economie et Innovation.

L'entrepôt de données publiques de **Rennes Métropole** et de la Ville de Rennes est géré par le service Innovation Numérique de la Direction générale de l'information et de la Communication de Rennes Métropole (le programme d'ouverture était dirigé par Xavier Crouan, à l'époque directeur adjoint Information et Innovation Numérique)

Le chef du projet **Montpellier Territoire numérique**, Jean-Marie Bourgogne, est situé dans la DSI mais il a pris soin de constituer une équipe pluridisciplinaire et son action est transverse.

Pour **Brest Métropole Océane**, la libération des données s'appuie sur le service "Internet et expression multimédia" au sein de la Direction de la proximité de la Ville de Brest. Il est intéressant de noter son rôle : "Le service a pour mission de soutenir et mettre en œuvre la politique de la collectivité dans le domaine de l'appropriation sociale des technologies de l'information et de la communication (TIC) et de l'expression multimédia des habitants. Cette politique qui répond à des enjeux de cohésion sociale et de solidarité territoriale vise à : développer l'appropriation des usages de l'internet et du multimédia, favoriser une équité d'accès aux TIC, développer le lien social, développer l'expression multimédia, accompagner des usages innovants, créer un contexte favorable à l'émergence de projets multimédia et de coopération."

## ■ Une coopération indispensable au sein de la collectivité

On verra dans la mise en œuvre qu'il est important de créer une cellule spécifique pour conduire les projets – il est alors essentiel d'y intégrer les



responsables métier concernés, qui seront responsables des projets relatifs à leurs domaines.

Cependant, les questions relatives aux données et aux données publiques font intervenir en tout état de cause plusieurs services et plusieurs personnes ressource clé dans l'organisation, qui peuvent constituer un réseau de compétences (le responsable «données publiques» pouvant être au demeurant l'un d'eux):

- Personne responsable de l'accès aux documents administratifs (PRADA)
- Correspondant informatique et libertés
- Représentants des services...
  - juridiques – Ce sera souvent la PRADA elle-même, qui appartient dans beaucoup de cas à ce service (dans 11 régions sur 18 ayant nommé une PRADA fin 2010, les autres se trouvant dans les secrétariats généraux, à la direction générale, dans les services documentation),
  - documentation,
  - archives,
  - DSI,
  - Communication,
  - Aménagement numérique,
  - Développement économique.

*Il faut veiller à inclure une compétence en utilisation de l'information géographique – sans doute la personne en relation avec la plate-forme régionale.*

#### 6.4.2 Au delà de la collectivité : appui sur des structures de mutualisation

La libération des données d'une collectivité doit impliquer les acteurs internes. Dès lors que l'ouverture va concerner des ensembles d'acteurs, que des outils mutualisés (dont un répertoire) vont être mis en place, il sera intéressant de s'appuyer sur une structure spécifique régionale – au sein de laquelle le Conseil régional joue un rôle important et à laquelle il apporte un soutien financier (ou une participation) mais avec les autres acteurs de la Région – il s'agit de mettre en place non seulement une gouvernance régionale sur ce dossier (mais pas une gouvernance du Conseil) mais aussi un maître d'œuvre.

Cet acteur peut être une association dédiée au développement numérique de la Région, comme Aquitaine Europe Communication, l'ARDESI en Midi-Pyrénées ou l'ARTESI en Ile de France. Ces associations, liées au Conseil régional, assurent la promotion des TIC. Elles peuvent regrouper (par exemple l'Agence Régionale des Technologies et de la Société de l'Information Ile-de-France) les collectivités, le Conseil économique et social régional, des acteurs d'Etat, des organismes socioprofessionnels, des associations, des entreprises publiques ou privées. D'autres possibilités existent – notamment un Syndicat mixte comme, en Bretagne e-Mégalis. On a vu (Ch. 3) que plusieurs structures étaient envisageables pour porter la mutualisation, à commencer par ce qui sera souvent le premier objet commun, le répertoire.

## 7 Mise en œuvre

*Ce chapitre reprend beaucoup d'éléments déjà identifiés pour les replacer dans une démarche d'ouverture des données publiques. A côté d'une tendance générale, les spécificités régionales et les choix politiques orientent, ici comme ailleurs, cette démarche.*

*Les différents domaines et types de données ne sont pas évoqués ici : ils ont été décrits dans les précédents chapitres, en particulier le chapitre 6 consacré au rôle de la région.*

### 7.1 Rappel de principes

L'ouverture des données est une tendance de fond, qui correspond à l'appropriation par la population des outils de réseaux et aussi bien à la circulation d'information dans les groupes sociaux et les associations citoyennes qu'à la nécessaire ouverture des entreprises vers leurs clients et leurs partenaires. Pour autant, les formes que cette ouverture prend, les modèles économiques qu'elle va appuyer ou susciter sont loin d'être définis aujourd'hui. Les débats qu'elle suscite actuellement seront dépassés et oubliés demain.

Dès lors, il est important que les démarches pratiques prennent en compte, au delà des enthousiasmes, des craintes et des modes, les tendances de fond de l'organisation des données et de leur insertion dans les politiques, et en particulier :

- les politiques d'aménagement et de développement économique,
- le développement d'une transparence démocratique responsable.

- Au delà de démarches de mise en œuvre qui doivent être adaptables et prendre en compte les spécificités régionales et les opportunités, les décisions d'ouverture doivent être intégrées aux politiques suivies dans les domaines d'intervention des collectivités et en particulier de la Région, et doivent être expliquées.

Il faut toujours en revenir à la mission de service public – qui peut évoluer –, et à la gestion des données. L'important, ce ne sont pas les principes sur les « open data », ce n'est pas non plus de rechercher d'emblée une grande qualité technique, c'est d'utiliser le mouvement actuel pour développer le service public, veiller à l'égalité d'accès des citoyens et à l'aménagement du territoire.

*Il faut donc oser avancer – les erreurs dans la démarche ne sont pas graves s'il n'y a pas d'erreur de fond sur la politique ou l'impact de l'ouverture.*

*Les Régions ne sont pas soumises à des demandes pressantes. Elles sont au demeurant engagées dans des démarches coopératives de gestion de données (géographie, données locales, environnement etc.). Cependant, il est important qu'elles s'emparent rapidement du dossier des données ouvertes et y prennent des initiatives. Elles vont avoir un rôle central à jouer dans la coordination et la mutualisation des collectivités sur le territoire et elles doivent d'abord maîtriser le problème autour de leurs propres données pour pouvoir coordonner des politiques au niveau régional.*

*Les répertoires de données sont une obligation légale et il est impératif, pour tous les acteurs, qu'ils ne se multiplient pas de façon non coordonnée sur le territoire. C'est par là que la Région peut commencer d'animer un mouvement.*

*Il faut aussi, on le verra, se saisir d'opportunités : l'important est de commencer à coopérer et à gérer des données partagées, autour d'un projet commun.*

Enfin, la réactivité est essentielle.

Les réflexions et actions des prochaines années vont être menées dans un espace extrêmement mouvant, du point de vue des pratiques comme des techniques. Les termes du débat évoluent rapidement. Les démarches actuelles sont datées : elles correspondent à une phase d'entrée dans le monde des données numériques potentiellement accessibles à tous.

A court terme, il faut pouvoir réagir. A tout moment, une demande peut être présentée, qu'il sera difficile de retarder : l'exemple de la généalogie et des réactions diverses des Conseils généraux est là pour le rappeler. Il faut évidemment essayer d'éviter que ce type de situation se reproduise, car cela aurait des effets négatifs sur l'ensemble de la gestion de l'information régionale.

*C'est pourquoi il convient de ne pas attendre les difficultés mais de définir une heuristique et de nommer rapidement l'équipe de personnes en charge.*

*De ce point de vue, la création d'un forum d'échange qui pourra réunir les élus intéressés comme les équipes en charge et leurs partenaires invités doit permettre une veille sur les enjeux, les opportunités, les risques.*

## 7.2 Organisation générale du domaine

Le schéma ci-dessous rappelle les principaux ensembles de problèmes, qui sont en même temps les principaux éléments d'une démarche. Ces éléments se retrouvent au travers des différents processus et des étapes qui sont décrits ensuite dans ce chapitre.

Le schéma isole des domaines logiques, et partiellement seulement chronologiques. Pour des actions importantes, il faut parfois parcourir le même trajet, revoir par exemple le choix de licence, ou la gestion de certaines données, ou les orientations de communication.

Certaines actions sont de long terme (mais à commencer le plus vite possible), d'autres d'opportunité, à mener en parallèle etc.

Le déroulement sera indiqué dans les paragraphes suivants.

Les blocs identifiables sont :

### ■ [1] Le pilotage d'ensemble

- Direction politique (élaboration d'une position, suivi du dossier et des impacts)
- Organisation des services : mise en place des responsabilités,
- Accompagnement interne.
- Stratégie, budget, agenda – pour la politique de moyen terme à mener, après délibération du Conseil régional

### ■ [2] Le travail systématique sur les données

- Données : de l'analyse interne à l'analyse des sources et des usages et aux modifications dans la gestion et la collecte. Il s'agit là d'une tâche de moyen et long terme, mais avec des phases d'actions à court terme en fonction des demandes et des opportunités, pour un ensemble particulier de données.

Il faut prêter attention au fait que la démarche "données" est aussi bien valable en interne que pour des données régionales gérées avec des partenaires externes (lien avec le troisième ensemble)

### ■ [3] Partenariats, vision régionale et politique

Recherche et suivi des collaborations avec des partenaires, en fonction des décisions politiques sur l'ensemble de la région et avec une vision de long terme.

■ [4] Répertoire et communication

- Répertoire : l'un des premiers chantiers (lié à tous les autres – il faut au moins avoir décidé de libérer quelques jeux de données)
- Communication – l'ouverture est toujours objet et facteur de communication.

■ [5] Emergence et lancement de chaque action

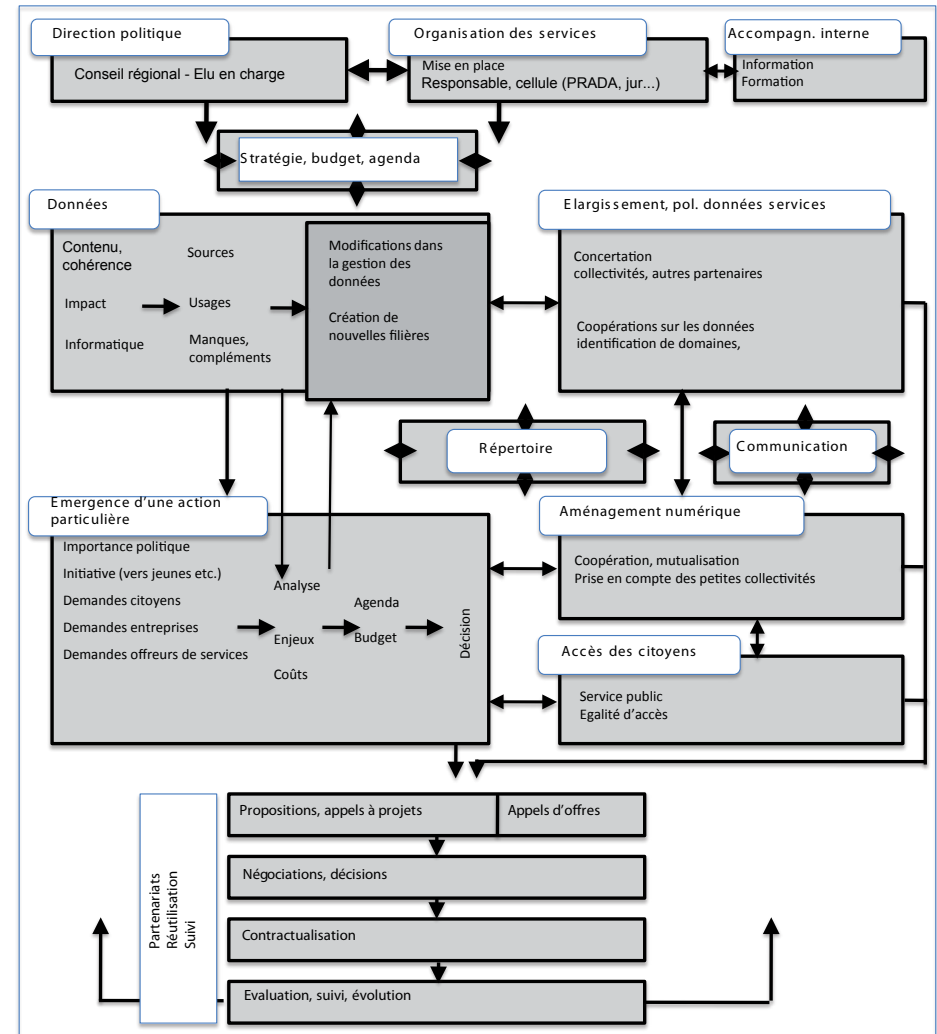
Quelles que soient les stratégies de long terme, et en fonction d'opportunités ou de demandes externes, il faut pouvoir maintenir le cap tout en répondant à l'actualité.

■ [6] Montage, évaluation, suivi

- Propositions, appels à projets, appels d'offres si des investissements sont nécessaires
- Négociations, décisions
- Contractualisation. Comme on l'a vu précédemment, elle est importante pour toute réutilisation (même si une procédure simplifiée peut être mise en œuvre)
- Evaluation, évolution, suivi.

.../...

□ Eléments de la démarche générale d'ouverture



La démarche qui est présentée ci-après concerne le lancement de l'ouverture des données publiques dans une région où cela n'est pas encore fait. Cependant, les éléments du processus seront à parcourir de nouveau, avec l'omission de certaines phases, comme l'ouverture d'un nouveau domaine important. Dans ce cas, par exemple, la cellule en charge s'adjoindra éventuellement un responsable métier et la délibération du Conseil, au lieu d'être générale, portera sur le domaine considéré.

Les processus se chevauchent souvent (cf. Schéma «Enchaînement des processus»).

### 7.3 PROCESSUS A1 - Du lancement du travail à la délibération et au plan de communication

- Désignation d'un élu porteur du dossier.
- Désignation d'un responsable dans les services et constitution de la cellule en charge

*(Le responsable est de préférence dans le secteur du développement économique ou de l'aménagement du territoire mais le principal critère est la motivation et la compétence)*

Il est choisi au sein de la cellule, qui comprend des représentants de : développement économique, aménagement, DSI, communication, service juridique, documentation. En sont membres la PRADA et le CIL (qui peuvent ou non être l'un des représentants précédents). Un des membres doit être celui qui assure la liaison avec la plate-forme d'information géographique.

- Phase de consultation et d'étude

...sur les principes, sur les points clé dans la région, sur les partenaires.

- Rencontres et concertations avec les élus et les services en charge des différents domaines métier et avec les services fonctionnels, en particulier l'informatique.

Rencontres et concertations avec les délégataires de services publics.

Rencontres et concertation avec les partenaires existants déjà éventuellement en matière de données publiques (Géographie, AOT, tourisme...) et avec les collectivités qui auraient déjà amorcé une ouverture ou auraient été confrontées à des demandes (Conseils généraux et généalogie).

En outre, la cellule met en place une veille sur le domaine et s'intègre aux différents sites et groupes de discussion sur le mode foisonnant de l'opendata. Elle échange en particulier avec les autres régions.

- Note d'orientation

La cellule complète une étude rapide du dossier :

- Bases de données et de documents accessibles.
- Problèmes, difficultés, opportunités déjà identifiées.
- Demandes d'accès ou de réutilisation déjà formulées
- Coopérations existantes sur les données (géographie, services publics, économie, transports etc.) – groupements, collectivités, délégataires.
- Ouverture éventuelle pratiquée dans les autres collectivités de la Région
- Opportunités liées à des événements ou domaines particuliers (ce qui va déterminer des partenaires, y compris en dehors du champ opendata proprement dit) - exemples : événement culturel, évolution sur l'organisation d'un parc naturel, partenariat dans un projet international, mise en place d'un grand programme d'équipement, etc.

Une note d'orientation est remise à la Direction des services et aux élus. Elle signale les partenariats existants. Elle permet à tous d'aborder le sujet. Elle est accompagnée d'informations de base pour les personnes peu informées du phénomène opendata. Elle esquisse des pistes.

#### ■ *Choix politiques initiaux*

La cellule doit aussi présenter aux élus les choix politiques initiaux sur des points clé :

- Type de domaines à aborder, notamment décisions sur budgets et activités de la collectivité, sur les transports, le tourisme (*cf. chapitre 6 sur les données régionales*)
- Type de partenariats à rechercher, en fonction de la nature et de la qualité des relations...
- Orientations quant aux relations avec les sources nationales (IGN, Finances etc.)
- Orientation quant à la gratuité, à la réutilisation par des associations citoyennes, par des sociétés commerciales (en expliquant ce que sont aujourd'hui les contraintes et obligations)
- Licences

#### ■ *Premier programme*

Après discussion avec les élus, la cellule prépare un plan et des estimations budgétaires pour la première phase. Le plan comporte des options claires de principe sur les questions qui se poseront immédiatement, notamment la forme des licences. Il propose en particulier :

- Données internes identifiées qui pourront initier la démarche. Dans la première phase, les données susceptibles de poser manifestement des

problèmes juridiques et techniques ne sont pas retenues – problèmes de protection de la vie privée, de protection du secret commercial, de droit d'auteur. En revanche, des jeux de données non clairement identifiés dans les données publiques, comme les données culturelles, ou les données mises à disposition par un partenaire mais non publiques au sens strict peuvent être incluses (ex des vélos STAR à Rennes).

- Investissement sur des données stratégiques (qui seront incontournables à terme mais ne seront sans doute pas dans le premier ensemble)
- Licences
- Programmes pouvant être lancés rapidement avec des partenaires
- Plate-forme de présentation des données / Répertoire, avec possibilité d'associer dès l'abord des partenaires (collectivités, administrations déconcentrées, organismes publics, chambres, institutions, transporteurs etc.. mais sans rigidifier : le partenariat ne doit pas ralentir l'ensemble du processus)
- Axes de communication (options)

#### ■ *Information et formation des personnels*

En parallèle avec le mécanisme conduisant à une décision politique, la cellule complète l'information du personnel dans les différents services.

Il est important que la politique d'ouverture n'apparaisse pas, et ne soit pas, une ingérence dans le travail de services déjà chargés et qui pourraient se placer – c'est fréquemment le cas – dans une position de résistance – avec souvent des raisons qui sont l'incomplétude ou la qualité médiocre des données, ou la charge de travail – ils signalent ainsi des problèmes réels et il faut parvenir à un consensus sur la démarche. Il s'agit donc non seulement d'informer mais d'écouter, afin d'éviter une opposition caricaturée entre

des «conservateurs trouvant de bons prétextes pour fermer leurs données» et des « militants de l'open data décidés à ouvrir de gré ou de force ». L'essentiel est bien de prendre en compte les difficultés réelles des services et de permettre au contraire la valorisation de leur travail.

Cette tâche de conduite du changement doit aussi être menée avec les élus – un dossier d'information doit être constitué.

Pour tous, il faut informer sur les causes et les modalités d'un processus d'ouverture qui est nouveau et par de nombreux points s'oppose aux pratiques et à la culture des services. Il faut d'ailleurs expliquer et respecter les obligations réglementaires (comme pour la publication des marchés publics), mais aussi l'intérêt de la libération des données.

Enfin, il est important de fournir des explications et si nécessaire des formations aux personnes qui vont mettre en œuvre différents processus. Lors de réunions sur le sujet, des plaintes se sont exprimées, par exemple des bibliothécaires pour la mise à disposition des données – dans les agglomérations déjà engagées et qui ont libéré des données des catalogues. Dans d'autres cas, ce sont des personnels de la collectivité, ou d'organismes associés, qui doivent être informés, impliqués, voire formés (par exemple si des données relatives aux parcs naturels sont ouvertes, ce doit être accompagné d'une action d'information et d'un suivi avec les personnes concernées). Les personnels peuvent aussi jouer un rôle dans la validation des données (voir par ex l'agent sourcing du Grand Toulouse – Ch. 2 et annexe).

Les membres de la cellule peuvent initier l'information en organisant des réunions de service. Cependant, il est utile d'organiser, à ce stade, un séminaire de formation au problème de l'ouverture des données publiques, qui pourra s'adresser à tout le personnel concerné, même s'il s'intéressera plus spécialement aux services représentés dans la cellule – documentation, informatique, service juridique, communication, développement économique, aménagement du territoire.

□ Pour les services métier, à ce stade, ils auront été vus dans la phase de préparation du programme. Ceci étant, un tel séminaire sera utile pour tous: par exemple pour les services en charge du budget, qui seront confrontés à des demandes. Il est possible, selon la situation et les manifestations d'intérêt, qu'il faille organiser plus d'une réunion.

#### ■ *Préparation de la décision politique*

Si des partenariats importants existent (entre autres sur l'information géographique, le tourisme, les transports) ou si des collectivités ont déjà entamé une démarche d'ouverture, il peut être important de se concerter à nouveau avec les partenaires avant de finaliser la proposition soumise à la délibération du Conseil régional.

Les options sont discutées avec les élus.

#### ■ *ETAPE et POINT DE CONTROLE : Délibération de principe du Conseil régional*

#### ■ *Elargissement politique*

Communication et débat au CESER.

#### ■ *Plan de communication*

Elaboré par la Direction de la communication, il différencie les cibles : grand public à partir notamment de la presse régionale et des sites d'information pour les différents publics; acteurs économiques; sociétés des TIC et start ups (par des relais spécialisés comme les agences de développement de l'usage des TIC).

## 7.4 PROCESSUS A2 – Réaction à des demandes externes et opportunités nouvelles

Il serait préférable que ce processus ne soit pas activé. En effet, les demandes externes comme les opportunités doivent être prises en compte dans le cadre des processus normaux de lancement puis de développement de la politique d'ouverture.

Cependant, en pratique, il sera souvent nécessaire de prévoir un mécanisme de réactivité.

### ■ Irruption de demandes et d'opportunités

Dans toute la mesure du possible, la demande, ou l'opportunité nouvelle doit être rapidement intégrée au processus contrôlé et voté. Elle peut cependant exiger une réponse rapide, n'entrant pas dans le programme envisagé mais devant être prise en compte, soit en raison d'obligation juridique soit en raison d'un intérêt particulier.

Les événements de ce type risquent de se multiplier avec l'intérêt général pour les données. Des demandes seront présentées par des associations citoyennes comme par des réutilisateurs commerciaux. Des partenaires peuvent aussi proposer des projets communs offrant une opportunité.

### ■ Analyse par la cellule données publiques

La cellule examine la demande ou le projet (qui peut être proposé par un partenaire, par exemple). Elle évalue les difficultés et les coûts.

En fonction de cette évaluation, elle peut élaborer une proposition qui doit respecter les règles de la diffusion de données publiques.

Il est essentiel d'établir une coordination avec les demandeurs / proposant,

□ pour mettre en particulier au point les questions de droits et tarification – notamment s'il y a des travaux spéciaux à effectuer.

La cellule soumet à la DGS une note d'orientation.

#### • Action

Il est préférable, il faut le répéter, d'accélérer autant que possible le processus A pour éviter des difficultés et se placer dans le cadre de la politique générale. Cependant, une action peut être possible même si aucune délibération n'a encore eu lieu, surtout s'il s'agit d'une demande légale. Il faut alors calquer les contrats éventuels au plus près des règles courantes et des licences nationales (Etablissements en ayant publié une).

Il faut à tout prix éviter de s'engager dans des batailles juridiques coûteuses pour tous et néfastes à l'ensemble du processus. Il faut au contraire utiliser l'action en cause pour le stimuler. S'il s'agit d'une demande et si elle rentre dans le cadre de la Loi, il n'y a plus qu'à en négocier, si nécessaire, les conditions.

Sur le plan technique, il faut minimiser les coûts et les risques et contractualiser avec le réutilisateur, ou les réutilisateurs.

### ■ ETAPE et POINT DE CONTROLE : signature d'un contrat

Avec le réutilisateur ou avec les partenaires qui ont proposé l'opération.

### ■ Suivi

Toute action devra être réintégrée dans les processus normaux de suivi. Tant que ce n'est pas le cas, un indicateur sera défini et une méthode de suivi sera prévue avec le ou les réutilisateurs.



## 7.5 PROCESSUS B1 : Travail interne : les données

Ce travail doit être commencé bien avant la délibération du Conseil. Il ne faut pas, en effet, un intervalle de temps important entre la délibération et l'ouverture des premiers jeux de données. Il commence après la phase "note d'orientation".

### ■ Application du programme en interne : les données

Définition fine de la liste des données de première vague. Si possible avec des données émanant des différents partenaires, mais aussi en invitant des acteurs à s'associer, en fonction des relations existantes – important pour cela de passer par les directions métier, qui doivent de toutes façons être associées en continu pour les données qui les concernent.

### ■ Description des données

Enclenchement du travail sur le dictionnaire des données : sa tenue, les méta-données retenues pour la description (cf. référentiel IGN, voir les autres acteurs des données publiques).

Effectuer ces choix avec les services et avec les partenaires, et en les liant aux travaux en cours dans d'autres collectivités et au niveau national (Etablab en particulier).

### ■ Analyse technique : les données

Analyse avec les services et la DSI, détaillant pour chaque jeu de données :

- le circuit de collecte et les responsabilités,
- les problèmes et la fréquence de mise à jour
- les risques éventuels à surveiller

- le processus technique d'accès
- le ou les formats accessibles
- l'identification du jeu par des méta-données

### ■ Mise en forme

Vérification de la qualité des données

Validation ou évolution des formats

Association à des méta-données

### ■ Les données des partenaires

Contacts et concertation avec les partenaires éventuels sur ces points. Il faut en effet s'accorder sur les méta-données, sur les licences (même si plusieurs sont possibles) et autant que possible sur les formats.

(Voir ci-dessous une structure plus permanente de concertation)

### ■ ETAPE et POINT DE CONTROLE :

*Remise d'un rapport et engagement sur les données*

Normalement, ce rapport est remis peu après la délibération du Conseil.

## 7.6 PROCESSUS B2 : Répertoire de la région et lancement public du programme

### ■ Choix d'architecture et d'organisation

Il peut y avoir un répertoire de la Région ou, d'entrée, le lancement d'un répertoire régional, auquel peuvent participer d'autres acteurs régionaux.

Ce sera d'ailleurs le cas dès le lancement d'un programme vers les petites collectivités.

Il est préférable de s'orienter vers le lancement d'un répertoire régional. Dans ce cas, les différents acteurs auront le choix d'être simplement référencés avec envoi vers leur propre répertoire (qui peut d'ailleurs techniquement être hébergé sur le même site) soit d'être intégrés ensemble dans un répertoire commun.

Dans le cas où des répertoires existent déjà sur la région, il faut se concerter avec leurs responsables – ce peut être des agglomérations, des acteurs de l'environnement etc.

#### ■ Partenaires

L'identification de partenaires initiaux est liée au choix des données à libérer et doit être menée avec les services métier.

Pour les délégataires de service public, le service juridique devra examiner les contrats pour savoir le statut des données. En tout état de cause, une concertation est indispensable.

Par ailleurs, les organismes, associations, que la Région finance sont au premier chef associés au répertoire. Cependant, il faut comprendre que, parfois, l'ouverture leur posera des problèmes spécifiques, notamment s'ils diffusaient leurs données, même gratuitement. Le nouveau système peut les priver d'une visibilité, ou plutôt en général le leur faire craindre. S'ils n'assurent pas une mission de service public, leurs données n'entrent pas automatiquement dans l'obligation de mise à disposition, mais pour autant elles sont à considérer. La concertation sera aussi organisée avec ces acteurs.

#### ■ Mode de gestion et structure porteuse

□ S'agissant d'un répertoire collaboratif, il faut choisir une structure capable d'assurer la maîtrise d'ouvrage (pas forcément la maîtrise d'œuvre), qui peut être une association spécialisée (promotion des TIC par exemple), un syndicat mixte (lui aussi dédié aux TIC, ou à l'archivage ou d'autres possibilités en fonction de l'existant et des opportunités).

Il est possible que cette structure ait déjà un statut et un fonctionnement tels qu'elle peut accueillir les différents partenaires. Pourtant, il ne sera pas facile d'élargir à un trop grand nombre de communes, par exemple, et impossible d'élargir à des acteurs privés qui voudraient s'associer – ce qui est souhaitable. Il faut donc, avec les services juridiques, aborder la question d'un organe de concertation qui pourrait être rattaché au CESER.

En tout état de cause, il faut préparer des formules de contractualisation entre la structure porteuse et les différents acteurs associés.

Rappel : les licences types et si possible une licence « clic » conforme aux définitions nationales seront relatives aux données publiques. Les données n'entrant pas dans ce cadre et placées par des partenaires, surtout privés, doivent figurer avec leur statut de réutilisation, placé par le propriétaire. Si elles ne sont pas libres, c'est vers ce propriétaire que le réutilisateur potentiel sera renvoyé.

#### ■ Contenu

Une étude doit spécifier le contenu du répertoire (*cf. chapitres sur ce sujet*).

Il serait très intéressant que le répertoire comprenne non seulement les jeux de données accessibles mais aussi ceux qui le seront à terme. L'enregistrement de demandes les concernant présente certes un risque de pression mais une réponse programmée avec un agenda prévisible peut aussi être un élément de transparence. Il faut rappeler que les textes obligent, en principe, à publier tout jeu de données, même non numérique,

susceptible de pouvoir être ouvert. L'identification de tous les jeux est sans doute quelque peu utopique mais il serait utile d'indiquer, par exemple, ceux qui sont en cours de numérisation, ou qui sont en création, ou qui vont être ouverts.

Remarque: certains sites de répertoire présentent les applications utilisant leurs données publiques. C'est sans doute normal dans une période de lancement et où la plupart des applications sont le fait de jeunes pousses régionales – qui plus est ayant été primées lors d'un appel à projets, ou ayant bénéficié d'aides comme entreprises innovantes. A terme, il y aurait là une confusion et un risque, si des précautions ne sont pas prises. Lorsque des réutilisateurs plus nombreux ou extérieurs à la région interviendront, des concurrents non référencés pourraient s'émouvoir. Par ailleurs, la présentation sur le site aurait – quelles que soient les précautions – valeur de garantie de la part de la collectivité. Il est en revanche possible de fournir la liste des sociétés ayant acquis la licence ou signé un contrat et de celles ayant été primées dans les appels à projet, ou s'étant au moins présentées : ce sont des données publiques... Il faut seulement veiller à une stricte égalité de traitement.

#### ■ *Information dans les services*

Rappel : information interne nécessaire à chaque étape.

#### ■ *ETAPE et POINT DE CONTROLE : remise d'un rapport, décision de lancement*

#### ■ *Développement*

A partir des choix d'organisation et avec la DSI, analyse technique et économique. Choix d'un outil logiciel et d'un hébergement puis d'une procédure pour la constitution pratique.

□ S'agissant d'un site et d'outils qui seront ouverts à des acteurs très divers, un marché public semble s'imposer, qui peut être passé par la Région ou par un Syndicat mixte (voir avec les services juridiques).

Le développement comprend donc les phases habituelles d'écriture du cahier des charges, d'appel d'offres, de sélection, de contractualisation.

#### ■ *ETAPE et POINT DE CONTROLE : Décision d'ouverture après réception*

#### ■ *Ouverture, communication, référencement*

L'ouverture du répertoire doit s'accompagner d'une communication :

- vers le public pour l'informer
- vers les acteurs des TIC, les acteurs économiques, les associations citoyennes, les associations professionnelles etc.

Une réunion avec les professionnels peut être organisée, avec les partenaires de l'opération.

Sur le Web, il faut veiller au référencement du répertoire dans les moteurs.

#### ■ *Action spéciale vers des réutilisateurs*

Dans les agglomérations qui ont ouvert des données, la pratique a souvent été d'organiser un concours, ou un appel à projets pour l'utilisation des données mises en ligne. Même si l'effet médiatique ne sera plus le même, il reste très intéressant de stimuler des jeunes pousses, voire des développeurs individuels ou des groupes d'étudiants, mais aussi des sociétés locales des TIC ou des associations citoyennes.

La cellule proposera un plan de ce point de vue, en fonction de la situation locale et des données ouvertes, en liaison avec les services en charge de l'aménagement numérique, avec la DSI et avec les services qui travaillent avec le Rectorat, selon les options retenues.

Un événement pourra être organisé pour présenter les applications ainsi développées.

#### ■ *Relations dans le réseau des répertoires*

Il faut établir des relations avec les répertoires existants : ceux de la région, ceux d'autres régions ou de domaines spécialisés, les répertoires nationaux et peut-être aussi des répertoires internationaux (voir en particulier les répertoires possibles de régions voisines, françaises ou étrangères).

#### ■ *Outils de suivi*

Définition d'indicateurs et de leur diffusion, pour suivre les consultations, les demandes de données, les problèmes (normalement prévu dans le cahier des charges, à compléter par une procédure interne).

## 7.7 PROCESSUS B3 : Travail d'approfondissement sur les données

*Ce processus concerne les données de la Région. Il se reproduira chaque fois que des ensembles importants de données seront susceptibles d'être ouverts.*

#### ■ *Identification de données stratégiques*

Elle a pu être menée au cours du processus A mais il se peut que d'autres facteurs, politiques ou révélés par les premiers développements et demandes, conduisent à identifier des jeux de données stratégiques mais dont l'ouverture représente une charge non négligeable de travail et éventuellement une dépense externe.

Les données sont classées en fonction de leur importance, du type de

difficulté et des coûts associés, des délais envisageables pour la mise en ligne : numérisation, anonymisation, libération de droits de propriété intellectuelle.

Un rapport présentant les enjeux, les choix, les coûts, les délais est remis aux élus.

#### ■ *Mise à niveau du système d'information et amélioration de données*

La DSI reprendra son propre programme de développement du système d'information. Elle proposera, en fonction de l'analyse d'ensemble du problème, d'éventuelles modifications de trajectoire, ou des modifications importantes du système pour le rendre capable de mieux supporter les demandes d'ouverture et pour pouvoir proposer des formats normalisés, le cas échéant – refonte des bases de données, outils de conversion etc. Elle présentera donc, si c'est nécessaire, un document d'évolution du SI, avec les budgets et agendas correspondants.

#### ■ *ETAPE et POINT DE CONTROLE : Décision, éventuellement délibération*

#### ■ *Développements*

Pour des jeux de données complexes, il peut y avoir nécessité de passer des marchés, notamment pour l'anonymisation ou le respect des secrets d'affaires.

Certains jeux peuvent aussi nécessiter la définition de règles particulières pour la réutilisation, une analyse technique, une analyse juridique et la mise au point d'un contrat (qui devrait devenir un contrat type). Ce sera notamment le cas pour les données dynamiques.

## 7.8 PROCESSUS C : Stimuler et coordonner le développement sur l'ensemble de la région

Ce processus ne peut être initié qu'après la fin des processus A, afin de travailler sur des exemples concrets.

### ■ *Coordination avec les autres acteurs déjà engagés*

Cette action a été menée au cours du processus A mais, comme cela a été indiqué, les partenariats comme les actions d'autres intervenants ne doivent pas ralentir celui-ci.

En revanche, il est toujours impératif de poursuivre et de développer la coordination avec les collectivités et les groupements qui, sur le territoire, ont déjà pu s'engager dans un processus d'ouverture, ou projettent de le faire.

### ■ *Information, concertation avec les autres collectivités*

Il s'agit pour la Région de veiller à réduire les disparités entre les différents territoires, et en particulier de s'adresser aux collectivités qui n'ont pas les ressources humaines, techniques ou financières pour mener seules une politique d'ouverture, jusqu'aux petites communes ou communautés de communes.

La cellule en charge va mener là un travail d'étude mais aussi d'information et de promotion.

- Etude de la situation, enquête auprès de collectivités pour cerner les besoins, le potentiel, les types de données à privilégier.
- Réunion d'information organisée par la Région, associant l'ensemble des collectivités, y compris le cas échéant celles déjà engagées.

### ■ *Programme de services et de synergie pour les petites collectivités*

Elaboration de propositions et d'un programme :

- Ensembles de données communes
- Identification des cas particuliers liés à des territoires (par exemple liés à l'agriculture, à d'autres activités productives, au tourisme, à des caractéristiques environnementales)
- Proposition d'une méthode de travail sur ces ensembles de données
- Définition de services d'assistance : des services en mode «Nuage» sur un serveur doivent permettre à la petite collectivité de gérer et publier ses données et de les placer dans le répertoire.
- Analyse des problèmes de définition des données du point de vue sémantique – ici encore, il faut veiller à l'harmonisation des langages et définitions, et cela va être difficile s'il s'agit par exemple de données liées à un territoire particulier et à un métier particulier. En effet, il faut à la fois permettre à la spécificité du territoire de s'exprimer et permettre que les données concernées puissent être rapprochées d'autres données du même type dans la région mais aussi dans d'autres régions ou pays.
- Analyse juridique pour définir les mécanismes permettant l'adhésion des collectivités intéressées, les droits et obligations, la contractualisation avec l'entité (syndicat..) qui gère la diffusion des données et le répertoire.

### ■ *ETAPE et POINT DE CONTROLE : Délibération du Conseil régional*

La proposition de lancement d'un programme pour l'ensemble des collectivités donnera lieu à une délibération.

## ■ *Lancement*

- Information vers les collectivités et vers le public
- Publication d'un document d'information et article sur le site – communication sur les différents supports destinés aux communes.
- «Campagne de recrutement» avec inscription des collectivités dans le programme. Signature d'une convention type précisant droits et obligations, éventuellement impacts financiers, possibilité de retrait etc.

## ■ *Développement*

Ces services reviennent à adjoindre au répertoire des outils accessibles sur le Web pour les collectivités adhérentes.

Le développement de ces outils peut exiger un marché, ainsi que leur mise à disposition, qui peut se faire de façon transparente sur un autre serveur que celui d'hébergement. Il faut pourtant essayer de ne pas multiplier les interlocuteurs. Le problème doit être étudié par le service juridique, en fonction des choix faits pour la gestion et l'hébergement.

## ■ *Ouverture à d'autres partenaires*

En parallèle ou en seconde étape, suivant la charge de travail et les difficultés rencontrées, il faut s'adresser aux acteurs régionaux qui sont susceptibles d'être intéressés par l'ouverture mais n'ont pas rejoint d'emblée le programme régional initial. Cet élargissement peut se faire en concertation avec le CESER.

Il s'agit en effet de coordonner l'action de la Région avec celle d'acteurs comme les chambres consulaires, des associations professionnelles, des

- organismes culturels. L'effort sera continu et il reviendra aux services métier comme aux élus de stimuler la synergie, la coopération, voire l'intégration dans des services communs, en particulier le répertoire régional.

Comme cela a été indiqué ci-dessus (B2), pour ces acteurs, quelques données relèvent du droit des données publiques mais d'autres non. Dans la plupart des cas, l'acceptation de placer les données sur le répertoire ne préjuge pas de leur statut, qui doit être indiqué. Les décisions de réutilisation sont le fait de chaque propriétaire de données.

## 7.9 PROCESSUS D : Relations avec les réutilisateurs

*Ce processus démarre dès qu'il y a des réutilisateurs...*

### ■ *Identification de premiers réutilisateurs*

Dès le lancement du processus d'ouverture, il peut y avoir des réutilisateurs connus, à partir des demandes d'associations ou de sociétés qui n'ont pas pu être satisfaites. Ces acteurs seront consultés par la cellule données publiques dès les premières étapes.

Ensuite, s'agissant des premières vagues de données, il peut être important, en fonction des données retenues, d'alerter différents types d'acteurs (ce sera également vrai pour des programmes ultérieurs sur des données plus longues à mettre en place).

Enfin, comme on l'a vu ci-dessus (A1), le lancement initial peut s'accompagner d'une action vers les développeurs et acteurs des TIC locaux.

### ■ *Suivi de la relation*

On verra qu'il faut suivre avec précision les réutilisations.

Il faut inclure dans les différentes licences des annexes proposant des méthodes communes de suivi. Il faut aussi prévoir des rencontres régulières qui peuvent être individuelles avec des réutilisateurs importants ou pour régler des problèmes particuliers.

En tout état de cause, les relations doivent être prévues, mais différentes selon les types de réutilisateurs :

- Jeunes pousses, associations locales etc. : une rencontre générale annuelle (ou semestrielle si le phénomène devient important) permettra une concertation avec les responsables de l'ouverture au niveau régional – la cellule et l'organisme porteur du répertoire.
- Utilisateurs industriels et commerciaux régionaux, nationaux ou internationaux. Des rencontres seront organisées, où ils seront conviés annuellement à débattre des problèmes rencontrés. Cela suppose, même s'il s'agit d'un acteur mondial, de connaître toujours le nom du responsable de la réutilisation des données chez cet acteur.

## 7.10 PROCESSUS E : Suivis et évaluations

### ■ *Politique d'évaluation : définition*

La politique d'évaluation sera définie dès l'ouverture des premières données. Elle est très importante mais ne doit pas être menée de façon rigide : elle doit apporter des éléments d'interprétation, car les projets se développent en environnement incertain.

La démarche est différente pour les utilisations intéressant la collectivité et pour les autres. Il existe en effet plusieurs types de services de ce point de vue :

- Les services qui sont demandés par des entreprises réutilisatrices

travaillant sur un marché non régional et n'intéressant pas la collectivité, en tout cas pas directement – l'exemple type en est la généalogie, mais ce pourrait être aussi des services d'information juridique et réglementaire, par exemple. La collectivité n'est concernée que par le risque de données mal retranscrites ou sciemment déformées. Le suivi porte ici essentiellement sur l'impact interne.

- Les services destinés à un public régional
- Les services destinés à un public plus large mais qui ont un impact direct sur la visibilité de la Région (culture, tourisme, économie etc.)
- Les services développés par des sociétés locales, quel que soit le public visé – le suivi est utile en particulier si les sociétés concernées bénéficient d'un soutien de la Région .

L'effet de nouveauté contraint à observer les services et leur utilisation dans la durée : dans la période qui suit l'ouverture de données, l'utilisation peut être lente à démarrer, ou au contraire relever de la curiosité initiale.

Les responsables devront être particulièrement attentifs aux applications de processus nouveaux – comme le "crowdsourcing", par lequel les citoyens eux-mêmes peuvent fournir en temps réel des données. Les informations recueillies conduiront souvent à des évolutions supportées par la collectivité et les réutilisateurs.

### ■ *Définition d'indicateurs et de méthodes*

La cellule doit définir des indicateurs sur l'usage, sur l'économie du système, sur les impacts.

L'évaluation se fait à deux niveaux : celui des objectifs qui avaient été annoncés et celui du fonctionnement pratique de la réutilisation.

Sur ce deuxième point, en interne à la collectivité, il peut s'agir d'indicateurs

sur le nombre de connexions à un répertoire et de consultation des fiches d'information sur les données, du nombre des réutilisateurs et de la pérennité de leur service, d'indicateurs de suivi des fonctionnements informatiques – fréquence des mises à jour, charge des serveurs, incidents et délais. Si des recettes ont été prévues, leur suivi sera évidemment disponible. Enfin, l'estimation des coûts doit faire l'objet d'une réévaluation régulière : accroissement des charges pour les personnels ou de charges externes (liées par exemple à un hébergement) – il faut donc une connaissance précise des dépenses lourdes, par exemple dues à l'anonymisation. Symétriquement, les gains de fonctionnement qui avaient été supposés seront à suivre (baisse de demandes d'information par exemple).

Surtout pour les applications intéressant la Région, soit parce qu'il s'agit de services à la population ou aux entreprises, soit parce qu'il s'agit de services apportant une visibilité externe, des enquêtes pourront être menées auprès des réutilisateurs – il serait bon que les contrats prévoient le retour d'indicateurs d'utilisation. Par ailleurs, et singulièrement pour les services intéressant la population ou les entreprises régionales, des enquêtes pourront être menées. On est alors dans un premier niveau d'évaluation des objectifs autant que du fonctionnement.

La mesure d'une action par rapport aux objectifs finaux est cependant toujours difficile, et d'abord parce que les objectifs de l'ouverture sont liés à des stratégies globales : aménagement, économie, environnement, qualité de vie... Elle l'est aussi parce que l'action devrait être confrontée à d'autres possibilités, à ce qui se serait passé si d'autres formes d'action, de diffusion des données, avaient été retenues.

Il faut en tout cas pouvoir repérer, pour un groupe de données, quels sont les utilisateurs au moins potentiels. En effet, c'est souvent avec eux que l'évaluation peut être menée, s'agissant de politiques régionales. L'évaluation va donc scinder les jeux de données en groupes a priori, mais il peut y avoir des usages non prévus et c'est si possible avec les réutilisateurs

que ces usages pourront être identifiés.

Il s'agit surtout de recueillir des éléments d'information et de réflexion.

L'évaluation s'appuiera aussi sur le signalement d'erreurs par les utilisateurs – signalements qui prendront parfois la forme de plaintes, les utilisateurs attribuant le service à la collectivité (cf. Ch. 3 *La question de la responsabilité*). En tout cas, et surtout pour les applications régionales, il est pertinent de prendre en compte ce «crowdsourcing de contrôle».

#### ■ *Recueil d'information, méthodes*

- Mise en place d'outils de mesure sur le répertoire et éventuellement des sites associés.
- Enquêtes auprès des publics et acteurs concernés (concernés par les données ou visés par les réutilisateurs), notamment :
  - Pour mesurer le soutien au secteur régional des TIC ou de l'information : développement des entreprises qui ont été clairement intéressées par la réutilisation, voire qui se sont créées avec elle.
  - Pour évaluer des services destinés aux PME : entretiens avec des patrons de ces entreprises.
  - Pour évaluer des services à la population : réunions avec des personnes actives dans les domaines concernés, sondages sur la notoriété des applications.
- Enquêtes d'usages non ciblées (car les usages ne sont pas toujours ceux qu'on attend)
- Enquêtes qualitatives auprès des fournisseurs de données, des différents partenaires, des réutilisateurs.



- Enquête dans les services
- Confrontation et comparaison avec d'autres régions, territoires, pays

### ■ Réunions

La cellule devrait organiser une réunion interne annuelle pour entendre les positions des personnels concernés, mesurer les difficultés rencontrées, les charges entraînées ou au contraire les allègements dus à la diffusion d'un certain type de données.

Un atelier du même type devrait être organisé au niveau de l'entité coopérative gérant le répertoire régional, avec tous les partenaires – mais en veillant à impliquer les différents personnels concernés et non simplement des décideurs.

### ■ Bilan

- Remise d'un rapport périodique (semestriel) sur l'utilisation des données et sur les usages,
- Remise d'un rapport annuel sur l'ensemble du bilan et en particulier sur l'économie du système.

### ■ Bilan politique et réorientations éventuelles

Le programme doit être validé ou amendé annuellement – ou plus rapidement si des événements externes l'exigent.

	Note	Délibération																		
A1 Démarrage à plan com.																				
A2 Demandes aléatoires																				
B1 Données																				
B2 Répertoire et lancement																				
B3 Approfondissement																				
C Ensemble de la région																				
D Relations réutilisateurs																				
E Suivi, évaluation																				

## 7.11 Articulation

### Vue d'ensemble des processus

(sans préjuger de la durée – le schéma suppose un processus très rapide pour démarrer si on considère les colonnes comme correspondant à un mois)

## 8 Perspectives

Le mouvement pour la diffusion et la réutilisation des données publiques est une tendance lourde. Celle-ci suscite des espoirs et des craintes souvent excessifs en raison de son apparente nouveauté – il s’agit plutôt d’extension et d’accélération – et parce qu’elle est liée à la vague des nouveaux outils de communication. Elle est bien cependant le début d’une nouvelle démarche. Pour les collectivités, elle s’accompagne de relations nouvelles avec les citoyens comme avec les acteurs économiques, à moins qu’elle ne les accompagne.

Le dossier est important et il ne peut faire l’objet de simples opérations de communication, sans doute intéressantes, ni de réponses au coup par coup à des demandes et des obligations nouvelles. Par ailleurs, il faut sans doute, alors que des passions s’expriment, adopter une attitude pragmatique, savoir ramener la question à une dimension raisonnable. La libération des données publiques est un élément d’une politique économique et de la vie démocratique, mais elle n’est sans doute pas l’amorce d’une révolution des pratiques sociales ou d’un nouvel eldorado économique, non plus d’ailleurs qu’une redoutable menace pour le service public. Ce type de débat passionnel relève de la jeunesse relative du phénomène.

La gestion et la réutilisation des données publiques ont un objectif simple: fournir aux citoyens et à l’économie des ressources qui paraissent de plus en plus porteuses de valeur, maintenant que les moyens de les diffuser et de les exploiter se sont développés et sont appropriés par la société. En même temps, il s’agit de s’engager dans des pratiques nouvelles et de nombreux problèmes se posent. Les chapitres de cette étude les ont abordés, et un échange doit se poursuivre entre tous ceux qui sont concernés.

Le présent chapitre rappelle quelques points clés et tente aussi de replacer les problèmes dans une perspective de moyen et long terme, parce que

- la nouveauté de la situation tend à concentrer l’attention sur des aspects techniques déterminants dans cette période de transition.

La politique d’information et de services exploitant les données publiques, tant vers la population que vers les acteurs économiques, ne peut être une politique de court terme. Il est indispensable d’engager une réflexion et une stratégie de moyen et long terme: il s’agit d’une nouvelle étape d’aménagement numérique, avec une différence importante par rapport aux précédentes : elle se développe dans un espace virtuel où les données sont produites sur le territoire mais où leur impact et leur usage peuvent en dépasser largement les limites et avoir une portée nationale et internationale.

Dans cette perspective, l’essentiel n’est pas le débat, au demeurant nécessaire et incontournable, sur l’ouverture des données publiques en général mais une politique des données impliquant des ensembles d’acteurs, au service des objectifs politiques définis pour le territoire. Cette politique est à décliner selon les domaines et les choix correspondants. La directive européenne Inspire et sa transcription française ont permis le lancement de ce qui est le premier grand chantier de gestion de données partagées, avec les plates-formes d’information géographique et les groupements qui les portent.

Dans le même temps, il est important, à court terme, d’être prêt à répondre aux sollicitations extérieures et donc de fixer des règles, tout en se conformant aux obligations légales et réglementaires.

Pour un Conseil régional, il faut aussi, en même temps, engager une analyse de ses propres ressources tout en approfondissant les relations avec les partenaires produisant ou détenant des données d’intérêt public, en renforçant les coopérations existantes et en lançant des actions communes. C’est ainsi que se mettront en place progressivement des réseaux de gestion de données de la Région.

## 8.1 Des certitudes de moyen terme

### 8.1.1 Le flot de données

La production de données et d'information ne va pas cesser de croître. Les mécanismes sont en place – de la production de textes, images, sons, documents de tous types par de multiples acteurs à la dématérialisation des échanges et des procédures, qu'elles soient administratives, commerciales, associatives. Le problème n'est plus l'existence de données numériques accessibles mais leur gestion. C'est aussi de savoir trouver ce que l'on cherche, d'être soi-même visible ou caché et de protéger les informations confidentielles. Une masse croissante de données va circuler et les intermédiaires capables de les organiser et de les valoriser ou utiliser vont jouer un rôle de plus en plus important.

En outre, la notion de données publiques tendra à s'estomper – non que l'importance du service public sera mise en cause mais parce que, de plus en plus, il faudra des arguments solides pour refuser de mettre à disposition des données – on le voit aujourd'hui avec les tendances qui apparaissent même dans les entreprises, et ce sera particulièrement fort pour les acteurs publics.

### 8.1.2 Economie de l'information

Les données publiques, et plus largement les données, sont considérées comme une ressource importante de l'économie de l'information.

Or, une confusion s'est installée sur la question de l'information, qui n'est pas vraiment levée. Parler d'économie de l'information signifie pour certains que l'information serait une nouvelle marchandise et que son passage au numérique permettrait d'en tirer de nouvelles et substantielles recettes. Or, si cette expression galvaudée a un sens, il ne s'agit pas de cela mais

de développer une économie où la qualité de l'information, l'efficacité de sa mise à disposition et l'efficience de son usage permettent un essor de l'ensemble des secteurs économiques, favorise l'innovation et l'initiative ainsi qu'en général une meilleure efficacité de tous les acteurs. Dans cet ensemble, les informations et documents numériques eux-mêmes peuvent être ou non des marchandises, mais c'est en tout état de cause secondaire.

### 8.1.3 Des batailles juridiques et sécuritaires temporaires

Tout nouveau processus important – et la mise en réseau d'une planète en est un – bouscule les organisations et les règles. Il est normal que des batailles juridiques se déclenchent autour de beaucoup de problèmes. Il est normal qu'il faille établir de nouvelles règles pour garantir la sécurité des uns et des autres. Il est également normal de réguler le processus de transformation pour éviter de léser gravement tel ou tel. Pourtant, l'essentiel n'est pas là : les débats permettent de définir des règles et surtout des voies d'évolution, mais c'est la société et l'économie dans leur ensemble qui changent et se réorganisent et souvent le débat perdra toute signification à terme.

Un exemple actuel peut être rappelé. Il peut paraître choquant de considérer que tout le débat, et les tensions fortes, sur la question de la musique et des œuvres de l'esprit, est secondaire, mais c'est bien le cas – d'ailleurs ce débat perd de sa force. Ce qui est vraiment en jeu est une nouvelle organisation de la chaîne sociale et économique qui relie les créateurs à des lecteurs / auditeurs / spectateurs.

Pour les données publiques, de même qu'il a fallu trouver des solutions juridiques et économiques pour les infrastructures (et ce n'est pas fini), de même, les règles, la répartition des charges entre la collectivité, les prestataires, les utilisateurs s'établiront progressivement. L'important est que les débats ne bloquent certains acteurs, qui seraient alors handicapés.

#### 8.1.4 Quels acteurs utiliseront les données territoriales

En ce qui concerne la diffusion d'information – en général gratuite et financée indirectement, il est vraisemblable que l'essentiel des données d'un territoire sera réutilisé par des associations et autres groupements sans but commercial mais voués à la promotion d'un domaine (environnement), aux intérêts d'une partie de la population, à l'action politique etc.

Pour les services, de nombreux seront payants, à des tarifs faibles, caractéristiques des offres sur le Web et le mobile, mais beaucoup aussi seront gratuits, financés par la publicité locale ou par du sponsoring.

En revanche, des acteurs diffusant ou commercialisant des données sur le marché national ou international vont apparaître – pour l'instant, ils ne sont encore que peu nombreux face aux collectivités mais leur pression va s'affirmer. Il s'agira de spécialistes agissant sur certains marchés (comme par exemple les marchés publics) ou des opérateurs généralistes. Ceux-ci pourront être directement liés au marché des données qui commence d'émerger – avec en France Data Publica. Le plus souvent, ils seront des réutilisateurs directs commerciaux, l'archétype actuel étant Google, dont le modèle économique nécessite qu'il dispose de masses énormes de données fines tout en visant uniquement, actuellement, les recettes publicitaires.

#### 8.1.5 Paiements et redevances

Quelles que soient les recettes de poche que la réutilisation puisse procurer à une collectivité, la question du paiement n'est pas d'abord économique, et surtout pas budgétaire : les redevances ne rapportent jamais beaucoup à une administration ou à une collectivité (qui n'est pas un Institut de brevet ou l'IGN). Les critères sont ailleurs – importance politique, impact économique, intérêt pour les citoyens, transparence.

Quoi qu'il en soit, et comme la pratique s'en répand, les redevances doivent

être aussi faibles que possible, dans les cas où des frais supplémentaires sont entraînés par la mise à disposition. Elles visent à rendre l'opération financièrement possible, et non financièrement rentable. Progressivement, la "neutralité du Net" s'imposera ici (comme elle le doit d'ailleurs du point de vue de la Loi) : égalité de traitement pour tous les réutilisateurs (et donc une start up locale et Google ou Apple), pas de relation avec les recettes du réutilisateur etc.

Il ne faut pas que des tarifs excessifs dissuadent les réutilisateurs, gênant les petits acteurs, inhibant l'émergence de services ou la visibilité d'un territoire à l'extérieur.

La tendance en ce sens est de plus en plus manifeste.

## 8.2 Des principes, pas des passions

### 8.2.1 Vigilance sur des services essentiels

Comme cela a été largement explicité par de nombreux acteurs de la libération des données publiques, et repris dans les chapitres précédents, des risques existent. Les deux principaux sont :

- le retrait de réutilisateurs rendant un service, d'information ou autre, jugé essentiel, avec donc dégradation du service public, au moins temporaire et toujours coûteuse à réparer ;
- l'accroissement d'inégalités entre des territoires ou des groupes de population, soit parce que certains territoires ne pourraient mettre à disposition des données, soit parce que des opérateurs jugeraient peu intéressant de cibler les besoins de certains groupes. Ces « zones blanches » de données devraient alors faire l'objet de services assurés par la collectivité (ou sans doute des prestataires).

La préservation du service public pour tous est évidemment au cœur des politiques de données qu'il s'agit de mettre en place.

Un risque inverse existe, qui est de stimuler le développement de nouveaux services sans prendre en compte l'existence d'acteurs les assurant partiellement aujourd'hui. Le pragmatisme et la prudence sont ici de règle.

### 8.2.2 Pragmatisme et réalisme

Les premières mises en ligne de fichiers ne représentent pas une dépense importante, ni une tâche très complexe pour la direction des systèmes d'information. Pour autant, il en sera tout autrement par la suite, face à l'accroissement de la demande, tant en nombre d'utilisations différentes qu'en diversification des jeux de données et aux contraintes et contrôles afférents. Cette activité nécessitera en tout état de cause des dépenses et surtout des tâches accrues pour les personnels (PRADA, informatique, départements métier etc.). La collecte des données elles-même pourra ensuite peser lourd si elle doit être améliorée et développée. Les données géographiques en donnent une première et forte illustration. On a vu cependant que l'ouverture peut entraîner des économies. Il n'empêche qu'un schéma de moyen terme pour les systèmes informatiques, pour les charges des services et pour les budgets externes doit être établi et tenu à jour. Tout n'est pas possible et il ne faut pas engager d'actions qui devraient ensuite être arrêtées. Ce réalisme doit intervenir, sans passion, dans les discussions sur les tarifs et sur la création de données. L'investissement sur les données se justifie mais doit être maîtrisé et raisonné.

### 8.2.3 Une politique des données

L'aménagement numérique du territoire, dans cette étape de développement du partage de données, suppose d'identifier, au delà des données propres à

la collectivité, les principaux gisements existants dans les domaines jugés prioritaires, ainsi que les éventuels manques, ressentis par les collectivités comme par les entreprises, les associations, les citoyens.

Cette réflexion doit être menée avec l'ensemble des partenaires régionaux. Il est vraisemblable que, après une décision du Conseil régional relative à la politique d'ouverture des données, le Conseil économique, social, environnemental pourra être saisi.

Les principaux domaines prioritaires sont aujourd'hui identifiés (encore qu'il y ait des spécificités régionales) : environnement, développement économique, enseignement et formation, transports, tourisme. La question de la géographie est essentielle, et liée à la plupart des domaines – on a vu que les expériences des diverses régions vont permettre d'aller progressivement – et en relation avec l'acteur national qu'est l'IGN – vers des solutions, même s'il y a encore des divergences fortes.

Des domaines méritent une attention particulière :

- Transparence de l'action de la collectivité (délibérations, budgets, marchés, action économique...)
- Les infrastructures – télécom, énergie en particulier, pour lesquels de nombreux problèmes se posent, où une coordination interrégionale est indispensable ainsi que le groupement des collectivités face à des entreprises puissantes. Le problème est directement lié au développement des infrastructures, dont il n'est qu'un aspect – et bien connu des collectivités.
- La culture, qui ne répond pas actuellement aux mêmes règles que les autres types de données – mais ce ne devrait plus être bientôt le cas. Or, les données culturelles vont représenter des enjeux de plus en plus importants, sur tous les supports et en provenance de nombreux acteurs, y compris des citoyens eux-mêmes.

- La santé : le traitement de ces données suscite encore des interrogations, parce qu'elles proviennent d'un monde très régulé et protégé et supposent pour beaucoup d'entre elles une anonymisation difficile. Pourtant, elles sont essentielles pour la santé publique et pour la maîtrise de nombreux projets d'aménagement de l'espace et elles intéressent beaucoup d'intervenants. La presse s'en empare d'ailleurs déjà, puisque beaucoup sont communicables.

#### 8.2.4 Qui fait quoi pour des données essentielles : les réflexes et les risques

Face à un type d'information ou de service considéré comme important, la question du développement direct d'un site (ou sur un site existant) ou d'une application se posera toujours. Ce sera d'autant plus vrai que la diffusion de données sur le domaine considéré aura pris un caractère de service public, ce qui ne manquera pas d'arriver dans l'avenir.

Pour des activités au cœur de l'action de la collectivité et/ou de sa communication, la publication directe d'éléments (au moins généraux) s'imposera certainement. Cela n'empêchera d'ailleurs pas de les ouvrir, et ce gratuitement, à charge pour les autres diffuseurs, surtout commerciaux, de faire mieux. Cependant, hormis pour les communications centrales sur la politique de la collectivité, les réutilisateurs devraient pouvoir mieux cibler vers des cibles diverses.

Comme cela a été rappelé ci-dessus, le risque dans ce domaine est que des réutilisateurs, constatant la non rentabilité de leurs opérations, se retirent, ce qui conduira la collectivité à reprendre le service. Le cas des données locales est à méditer, mais il s'agissait sans doute de services en avance sur l'époque car l'usage du Web et des mobiles n'était pas encore suffisamment banalisé.

#### 8.2.5 De la diffusion des données issues de l'action publique à l'action publique pour créer des données

Dans le moyen et long terme, des procédures se développeront pour faire face au cas où un acteur public veut qu'un certain type de données se diffuse, qui n'étaient pas collectées jusqu'ici dans le cadre du service public. En effet, jusqu'ici, la diffusion de données n'était une mission que pour un nombre limité d'organismes d'Etat (INSEE, IGN..).

Ce n'est plus le cas dès lors que les données deviennent une composante stratégique de certaines actions. Le problème a d'ailleurs commencé de se poser avec les données géographiques, comme cela a été signalé.


Dès lors, des marchés pourront être passés par une collectivité, ou plus souvent par un groupement. Ils peuvent concerner tout le processus, ou une partie (traitement, hébergement, diffusion). Si ces données sont considérées comme publiques – et il serait paradoxal qu'elles ne le soient pas, elles seront d'accès libre et sans doute gratuit. Dès lors, le prestataire de service technique, même s'il est lui-même réutilisateur, ne pourra jouir d'aucun avantage ou exclusivité (*problème évoqué dans l'étude*).

### 8.3 Quelques points forts de la démarche

Il n'y a pas de démarche unique précise. Même pour le répertoire, qui est une obligation, la situation locale peut permettre une solution collective ou au contraire un lancement rapide par le Conseil régional seul en première étape. Les éléments qui suivent rappellent seulement quelques points clé.

Les points principaux dont évoqués ici, beaucoup correspondant à des attitudes, des réflexions, des tâches simultanés.

- Y aller plutôt que d'autres y aillent, qui n'auraient pas les préoccupations



correspondant à la politique régionale, notamment d'aménagement du territoire. Plus le temps passe et plus des initiatives seront prises – par des acteurs publics ou privés, ce qui est en soi souhaitable mais rendra plus difficile la coopération et la mutualisation si la Région n'est pas elle aussi clairement engagée. Pour cela, une fois l'engagement politique déterminé, il faut choisir des domaines et proposer des données certes utilisables mais pas encore de façon très élaborée, respectant toutes les normes connues etc.

- Avoir une réflexion générale et un engagement précis, des élus et des services sur la relation avec le service public, la politique des données, la politique de réutilisation dans l'étape présente - entre ce qui est obligatoire légalement et ce qui est interdit, les usages actuels des données et les méthodes correspondantes. Savoir qui est responsable. Préciser les priorités.

- Adopter une heuristique de court terme pour réagir aux demandes et aux offres – dont une politique de licence.

- Etre en veille et pour cela échanger, s'inscrire dans des réflexions collectives, associer les autres acteurs régionaux.

- Expliquer aux personnels concernés, valoriser cette nouvelle organisation des données et activités, former.

- Se saisir d'opportunités fortes : évènements, coopérations régionales sur un secteur, un projet

- Ouvrir un répertoire / portail et le partager avec un maximum de partenaires, y compris hors données publiques – et coopérer avec les autres répertoires, même dans la région (villes, associations, acteurs économiques, plates-formes spécialisées) en veillant aux normes.

- Intégrer la préoccupation dans les contrats, dont ceux de délégation de service public, les règlements de syndicats mixtes, de GIP etc.

- Organiser la concertation de l'ensemble des collectivités et proposer des services à ceux qui le souhaitent, mutualiser.

- Assurer un suivi : mettre en place des indicateurs souples, recueillir les informations auprès du personnel, auprès des acteurs du territoire, maintenir des relations de partenariat avec les réutilisateurs, observer les évolutions dans d'autres régions. Assurer une veille particulière sur des thèmes considérés comme stratégiques ou sensibles.

## 8.4 L'OTeN

### 8.4.1 Des attentes et un potentiel

Les interlocuteurs rencontrés au cours de l'étude, et pas seulement dans les Conseils régionaux, considèrent qu'il manque un espace de réflexion et un outil de coordination pour les Régions sur la questions de la réutilisation des données publiques et plus largement sur les stratégies et les méthodes en matière de données. L'OTeN est attendu aussi par ses partenaires potentiels que sont la FING et le GFII.

Comme cela a été indiqué plusieurs fois, il faut un guide des bonnes pratiques, et il y a aujourd'hui peu de gens qui savent traiter de la question des données publiques du point de vues organisationnel et stratégique. Compte tenu de l'évolution rapide de la situation, ces bonnes pratiques doivent au demeurant être suivies en permanence et de façon interactive.

La spécificité de l'action régionale suppose de pouvoir suivre des projets, de se confronter pour des actions groupées, par exemple sur les questions de relation avec les grands acteurs nationaux et l'Etat, en liaison avec l'ARF. Ainsi ne s'agit-il pas seulement d'observer mais de construire un outil pour l'action, même si c'est au sein d'un Observatoire...

Il ne s'agit pas d'atteindre sur tous les sujets à une position commune, même si ce sera souvent plus souhaitable, en tout cas plus efficace, mais de disposer d'un lieu d'échange sur les problématiques régionales et d'un outil de coopération pour ceux qui souhaiteront mutualiser des méthodes et des outils pour leurs projets, ou qui voudront mener une action commune.

#### 8.4.2 Suivi et accompagnement de projets

La communauté que constitue l'OTeN peut appuyer des projets, qui seront alors suivis par les membres de l'association ou un groupe d'intérêt particulier en son sein. En effet, et surtout dans un domaine en construction, c'est en participant concrètement à des projets et en jouant de la synergie entre les membres qu'il est possible de dégager les bonnes pratiques et de mutualiser rapidement les résultats.

Une telle démarche est envisagée en Région Midi-Pyrénées pour accompagner les collectivités.

#### 8.4.3 Vers un outil d'information, d'échange et de coordination

Dans une situation très mouvante, un outil efficace doit :

- fournir des informations à jour – il pourra d'ailleurs pour cela se relier aux autres outils du domaine en France (FING, GFII, ETALAB etc.) et renforcer les observations sur les gestions régionales des données
- permettre l'organisation de groupes d'intérêt et de discussions
- permettre l'organisation d'un travail de groupe pour des projets
- constituer progressivement une base de cas

Techniquement, cet outil est à rechercher dans ceux du Web 2.0 car il s'agit de créer un réseau social, où des groupes pourront s'organiser.<sup>37</sup>

<sup>37</sup> Des outils comme Jamespot (utilisé pour le réseau social du GFII) sont bien adaptés. Un CMS, notamment Wordpress, est aussi envisageable mais n'a pas la même souplesse.



## Documents de référence

De très nombreux documents sont accessibles sur le Web, en particulier aux adresses indiquées dans l'annexe «Sites».

Ceux qui sont référencés ici sont des rapports et des études qui ont représenté ou représentent une contribution importante ou une étape dans le processus de développement de la réutilisation des données publiques comme dans la réflexion sur le sujet.

Le classement en rubriques est indicatif – beaucoup de documents traitent de plusieurs aspects.

Directive européenne

Textes juridiques français

Analyses et politiques d'ensemble, guides généraux

Données culturelles

Géographie et environnement

Information scientifique et technique - Open Access

Informations internationales – Etat des lieux - Economie

Licences

Redevances

Répertoires

Technique

Transports

### Directive européenne

Directive 2003/98/CE du Parlement et du Conseil du 17 novembre 2003 concernant la réutilisation des informations du secteur public

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32003L0098:HTML>

### Textes juridiques français

Ci-dessous une liste de textes juridiques importants, accessible sur

<http://www.legifrance.gouv.fr/>

2011

Décret n° 2011-577 du 26 mai 2011 relatif à la réutilisation des informations publiques détenues par l'Etat et ses établissements publics administratifs - JORF n°0123 du 27 mai 2011

Loi n°78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal - Version consolidée au 19 mai 2011

Circulaire du 26 mai 2011 relative à la création du portail unique des informations publiques de l'Etat « data.gouv.fr » par la mission « Etalab » et l'application des dispositions régissant le droit de réutilisation des informations publiques

Décret n° 2011-194 du 21 février 2011 portant création d'une mission « Etalab » chargée de la création d'un portail unique interministériel des données publiques

2010

Ordonnance n° 2010-1232 du 21 octobre 2010 portant diverses dispositions d'adaptation au droit de l'Union européenne en matière d'environnement

2006

■ Circulaire du Premier ministre 29 mai 2006 : réforme des dispositions régissant l'accès aux documents administratifs et institution d'un droit de réutilisation des informations publiques

■ 2005

■ Décret n°2005-1755 du 30 décembre 2005 relatif à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques, pris pour l'application de la loi n° 78-753 du 17 juillet 1978.

Ordonnance n°2005-650 du 6 juin 2005 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques.

1978

Loi n°78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal

## Analyses et politiques d'ensemble, guides généraux

2011

### FING

Guide pratique de l'ouverture des données publiques territoriales - Guide à l'usage des territoires et de leurs partenaires.

Version de travail.

<http://fing.org/?Guide-pratique-de-l-ouverture-des>

### Ecole des Ponts ParisTech

Les Données publiques au service de l'Innovation et de la Transparence. Pour une politique ambitieuse de réutilisation des données publiques. Rapport pour la Délégation aux Usages de l'Internet, dans le cadre du portail

□ Proxima Mobile de services aux citoyens sur terminaux mobiles.

<http://www.proximamobile.fr/>

2010

### Aquitaine Europe Communication

Les données publiques, guide juridique & pratique

<http://www.aecom.org/Vous-informer/Juridique-TIC/Guides-juridiques/Les-donnees-publiques-decembre-2010>

### Conseil d'orientation de l'édition publique et de l'information administrative

Rapport 2010, présenté à Monsieur le Premier ministre en application de l'article 1er du décret n° 2010-32 du 11 janvier 2010

<http://www.dila.premier-ministre.gouv.fr/missions/comite-orientation-edition-publique.html>

2000

### Commission européenne

Commercial exploitation of Europe's public sector information - final report, Préparé par Pira International Ltd., University of East Anglia and KnowledgeView Ltd.

[http://www.e-gouvernement.be/bibliographie/ce\\_0009.htm](http://www.e-gouvernement.be/bibliographie/ce_0009.htm)

1999

### Commission européenne

Livre vert « L'information émanant du secteur public dans la société de l'information, une ressource clef pour l'Europe ».

<http://www.cordis.lu/econtent/publicsector/greenpaper.html>

## Commissariat général du Plan - France

Diffusion des données publiques et révolution numérique – Rapport de Dieudonné MANDELKERN, Bertrand du MARAIS, La Documentation Française, Paris.

<http://www.ladocumentationfrancaise.fr/rapports-publics/994001620/index.shtml>

1997

Association Communication publique

Diffusion des données publiques et marché de l'information - Rapport Pierre Soubie.

Voir auprès de l'association (rapport plus en ligne) <http://www.communication-publique.fr/>

1992

## Commission européenne

Report on Access to Public Information (prepared by Policy Studies Institute, London and Centre de Recherches Informatiques et Droit, Namur for the European Commission).

<http://europa.eu.int/ISPO/legal/en/publaw/publaw.html>

## Données culturelles

2011

## Commission européenne

PSI re-use in the cultural sector - Final report - Curtis+Cartwright Consulting Ltd - Genevieve Clapton, Dr Max Hammond, Nick Poole

[http://ec.europa.eu/information\\_society/policy/psi/docs/pdfs/report/cc462d011\\_1\\_1final\\_report.pdf](http://ec.europa.eu/information_society/policy/psi/docs/pdfs/report/cc462d011_1_1final_report.pdf)

□ 2009

## Ministère de la culture et de la communication

Partager notre patrimoine culturel - Propositions pour une charte de la diffusion et de la réutilisation des données publiques culturelles numériques – La documentation française.

Conclusions du groupe de travail sur le patrimoine culturel numérisé remises à Madame la ministre de la culture et de la communication.

<http://www.ladocumentationfrancaise.fr/rapports-publics/104000652/index.shtml>

## Géographie et environnement

2010

## Association française pour l'information géographique (AFIGEO)

Catalogue des infrastructures de données géographiques françaises. 5<sup>èmes</sup> rencontres des dynamiques régionales en information géographique - Orléans, 3 et 4 juin 2010 - AFIGÉO / EUROGI / eSDI-NET +

[http://www.dewynterconseil.fr/images/AFIGEO\\_Catalogue2010.pdf](http://www.dewynterconseil.fr/images/AFIGEO_Catalogue2010.pdf)

## France – Transposition Inspire

Ordonnance n° 2010-1232 du 21 octobre 2010 portant diverses dispositions d'adaptation au droit de l'Union européenne en matière d'environnement JORF n°0246 du 22 octobre 2010

<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000022934766&dateTexte=&categorieLien=id>

2007

## Union européenne - Directive INSPIRE

DIRECTIVE 2007/2/CE DU PARLEMENT EUROPÉEN ET DU CONSEIL du 14 mars 2007 établissant une infrastructure d'information géographique dans la Communauté européenne - Journal officiel de l'Union européenne 25/04/2007

[http://europa.eu/legislation\\_summaries/environment/general\\_provisions/l28195\\_fr.htm](http://europa.eu/legislation_summaries/environment/general_provisions/l28195_fr.htm)

### Conseil national de l'information géographique

La directive INSPIRE – Fiche CNIG du n° 110 2007

[http://www.cnig.gouv.fr/Front/docs/cms/direc-insp\\_123651176165657900.pdf](http://www.cnig.gouv.fr/Front/docs/cms/direc-insp_123651176165657900.pdf)

### Information scientifique et technique - Open Access

2010

Consortium universitaire de publications numériques (Couperin), CNRS – INIST, ministère de l'enseignement supérieur et de la recherche, avec six pays d'Europe du Sud, dans le projet SELL (Southern European Library Link).

Open Access in France A State of the Art Report

<http://www.couperin.org/fr/presentation/nos-cooperations-nationales-et-internationales/sell/item/764-etat-de-lopen-access-en-france>

### Informations internationales – Etat des lieux - Economie

2011

Espagne – Ministère de l'industrie, du tourisme et du commerce

Réutilisation de l'information du secteur public. Characterization Study of the infomediary Sector. Projet aporta.

[http://www.ontsi.red.es/articles/detail.action?id=5340&request\\_locale=en](http://www.ontsi.red.es/articles/detail.action?id=5340&request_locale=en)

<http://www.aporta.es>

### Ambassade de France au Royaume uni

Data.gov.uk : l'ouverture des données publiques au Royaume uni

<http://digit4cities.files.wordpress.com/2011/05/louverture-des-donnc3a9es-publiques-au-royaume-uni.pdf>

### CSC

Unchartered waters, the state of opendata in Europe 2011

<http://www.csc.com>

2010

### The national Archives

The United Kingdom Report on the Re-use of Public Sector Information, report 2010 : Unlocking PSI potential

<http://www.nationalarchives.gov.uk>

2006

### Union européenne

Measuring European Public Sector Information Resources (MEPSIR) - Makx Dekkers, Femke Polman, Robbin te Velde, Marc de Vries

[http://ec.europa.eu/information\\_society/policy/psi/actions\\_eu/policy\\_actions/mepsir/index\\_en.htm](http://ec.europa.eu/information_society/policy/psi/actions_eu/policy_actions/mepsir/index_en.htm)

### Licences

2011

## Etalab

Licence ouverte

<http://ddata.over-blog.com/xxxyyy/4/37/99/26/licence/Licence-Ouverte-Open-Licence.pdf>

## APIE Agence du patrimoine immatériel de l'état

Documents de cadrage ; les différents types de licences

<http://www.apiefrance.fr>

**Creative Commons France, Regards citoyens, Veni vidi libri, Open knowledge foundation**

Déclaration sur l'opendata en France

<http://donneeslibres.info/opendataFR.pdf>

2010

## The national Archives (Royaume Uni)

Open government licence for public sector information

<http://www.nationalarchives.gov.uk/doc/open-government-licence>

## Redevances

2011

### GFII

Principes de tarification de la réutilisation des informations publiques  
Position du GFII en vue de la rédaction des recommandations du COEPIA  
février 2011

<http://www.gfii.fr>

□ 2010

## CNRS: Université de Strasbourg

La valorisation des informations du secteur public (ISP) : un modèle économique de tarification optimale. Julien Pénin (coordinateur), Caroline Hussler, Valentine Millot, Patrick Rondé, Haniyeh

<https://www.apiefrance.fr/>

## Répertoires

2010

### APIE

Guide méthodologique d'aide à la mise en place d'un répertoire des informations publiques au sein d'un ministère

[www.apiefrance.fr](http://www.apiefrance.fr)

## Technique

2011

### GFII

Réutilisation des informations publiques et formats Position du GFII en vue de la rédaction des recommandations du COEPIA

<http://www.gfii.fr/uploads/docs/reutilisation-des-informations-publiques-et-formats.pdf>

2009

### DGME

■ Référentiel Général d'Interopérabilité RGI

■ [www.modernisation.gouv.fr](http://www.modernisation.gouv.fr)

■ Transports

■ 2010

**CETE Méditerranée** (ministère de l'équipement)

Accès aux données décrivant l'offre de transport collectif - état des lieux

<http://www.cete-mediterranee.fr/tt13/www/IMG/pdf/accesdonneesTC.pdf>

2007

**Plate-Forme de Recherche et d'Expérimentation pour le Développement de l'Information Multimodale (PREDIM)**

Note relative à l'accès aux informations concernant la mobilité et le transport des personnes et leur réutilisation

<http://www.predim.org/spip.php?article2705>

## Sites

Les sites indiqués ici sont ceux d'institutions et associations nationales ou de collectivités ayant ouvert un répertoire. Compte tenu de l'extension du domaine et des débats sur le sujet, cette liste ne prétend pas à l'exhaustivité. Elle a vocation à évoluer en permanence jusqu'à la constitution éventuelle d'un outil base de données / portail.

Organismes publics

- *Etalab, Mission du Premier ministre en charge de l'ouverture et du portail de l'Etat*

<http://blog.etalab.gouv.fr/>

- *APIE, Agence du patrimoine immatériel de l'Etat*

<https://www.apiefrance.fr/>

- *CADA, Commission d'accès aux documents administratifs*

<http://www.cada.fr/>

- *DILA, Direction de l'information légale et administrative*

<http://www.dila.premier-ministre.gouv.fr/>

- *CNIL, Commission nationale de l'informatique et des libertés*

<http://www.cnil.fr/>

Associations nationales ayant ouvert un chantier opendata

- *FING, Fondation Internet nouvelle génération*

<http://fing.org/>

- *GFII, Groupement français de l'industrie de l'information*

<http://www.gfii.fr>

- *OTeN, Observatoire des territoires numériques*

<http://oten.fr>

Associations citoyennes sur les opendata

- *LiberTIC*

<http://libertic.wordpress.com>

- *Regards citoyens*

<http://www.regardscitoyens.org>

Autres associations impliquées (sectorielles, spécialisées)

(voir aussi "Géographie")

- *ACSEL*

Association de l'économie numérique (entreprises)

<http://www.associationeconomieenumerique.fr/>

- *AFCDP, Association française des correspondants à la protection des données personnelles*

<http://www.afcdp.net/>

- *FIGEC, Fédération nationale de l'information d'entreprise et de la gestion de créances*

<http://www.figec.com/>

- ■ *GESTE, Groupement des éditeurs de services en ligne*

<http://www.geste.fr/>

Répertoires nationaux privés

- *Data Publica*

<http://www.data-publica.com/>

- *Nos données*

"Le datagov citoyen" par Regards Citoyens

<http://www.nosdonnees.fr/>

Géographie - Géolocalisation

- *IGN – Géoportail, Géocatalogue*

<http://www.ign.fr>

<http://www.geoportail.fr/>

<http://www.geocatalogue.fr>

- *Inspire*

<http://inspire.jrc.ec.europa.eu/>

<http://inspire.ign.fr/>

- *OpenStreetMap*

<http://www.openstreetmap.org/>

[http://wiki.openstreetmap.org/wiki/FR:Main\\_Page](http://wiki.openstreetmap.org/wiki/FR:Main_Page)

■ *AFIGEO, Association française pour l'information géographique* - <http://www.afigeo.asso.fr/>

■ *SPDG, Syndicat professionnel de la géomatique*

<http://www.spdg-geomatique.org/site/exindex.htm>

Syndicat professionnel de la géomatique

■ *Normes*

<http://www.geonames.org/>

■ *Des sites régionaux*

CRAIG - Centre régional auvergnat d'information géographique

<http://www.craig.fr/>

CRIGE PACA - Centre régional de l'information géographique

<http://www.crige-paca.org/>

APEM - Association pyrénéenne d'économie montagnarde

<http://www.apem.asso.fr/>

## Collectivités

□ ■ *CR Aquitaine et CG Gironde*

<http://datalocale.fr/>

■ *CG Saône et Loire*

<http://www.opendata71.fr/>

■ *Brest*

[http://espcol002.brest-metropole-oceane.fr/sites/wss00192/pages\\_web/tout\\_public.aspx](http://espcol002.brest-metropole-oceane.fr/sites/wss00192/pages_web/tout_public.aspx)

■ *Montpellier*

<http://opendata.montpelliernumerique.fr>

■ *Paris*

<http://opendata.paris.fr/opendata/jsp/site/Portal.jsp>

■ *Rennes*

<http://www.data.rennes-metropole.fr/>

## Datajournalisme



- OWNI

<http://owni.fr>

- EchosOpendata

<http://blogs.lesechos.fr/les-echos-opendata-r114.html>

Union européenne

- Union

[http://ec.europa.eu/information\\_society/policy/psi/index\\_en.htm](http://ec.europa.eu/information_society/policy/psi/index_en.htm)

- Association PSI Alliance, Public Sector Information Alliance  
(communauté de la réutilisation)

<http://www.psialliance.eu/>

Plate-forme d'information soutenue par le programme européen eContent

<http://epsiplatform.eu/>

- Projet LAPSI

<http://www.lapsi-project.eu>

Réseau thématique européen sur les aspects juridiques des données publiques

## International : les répertoires pionniers

- ■ Etats-Unis

<http://www.data.gov/>

- Royaume Uni

<http://data.gov.uk/>

## Standards sur les outils de base

(sur les formats de base, mais il faut rappeler que l'essentiel est dans les normes par domaines, en particulier dans la géolocalisation).

- Web Consortium

Government Linked Data (GLD) Working Group

[http://www.w3.org/2011/gld/wiki/Main\\_Page](http://www.w3.org/2011/gld/wiki/Main_Page)

RDF (Resource Description Framework)

[http://www.w3.org/standards/techs/rdf#w3c\\_all](http://www.w3.org/standards/techs/rdf#w3c_all)

- Consortium Schema.org Google, Microsoft, Yahoo!

<http://schema.org/>

## Le monde du libre

(axe de réflexion aussi présent dans LiberTIC et autres cf. aussi Open Street

■ Map)



■ Veni Vidi Libri (aspects juridiques, licences)



blog.vvlibri.org



■ Creative Commons

<http://creativecommons.org/about>

<http://fr.creativecommons.org>

■ Open Knowledge Foundation

[http://fr.wikipedia.org/wiki/Open\\_Knowledge\\_Foundation](http://fr.wikipedia.org/wiki/Open_Knowledge_Foundation)

## Offres internationales

■ Google, Google public data explorer

<http://www.google.com/publicdata>

■ Microsoft, Open Government Data Initiative

<http://www.microsoft.com/government/ww/public-services/initiatives/Pages/open-government-data-initiative.aspx>

<http://ogdisdk.cloudapp.net/>

□ ■ Socrata

<http://www.socrata.com/>

### Mentions légales :

Directeur de publication : Philippe OURLIAC

Maître d'œuvre : Norbert Paquel pour CANOPE

Emmanuel Pavageau pour 2iM SA

Janvier 2012

Réalisation  **PublicPartner**  
Communication publique