





DATABASE

Open Access



Data management plans as linked open data: exploiting ARGOS FAIR and machine actionable outputs in the OpenAIRE research graph

Elli Papadopoulou^{1*} , Alessia Bardi² , George Kakaletis³ , Diamadis Tziotzios³ , Paolo Manghi²  and Natalia Manola¹ 

Abstract

Background Open Science Graphs (OSGs) are scientific knowledge graphs representing different entities of the research lifecycle (e.g. projects, people, research outcomes, institutions) and the relationships among them. They present a contextualized view of current research that supports discovery, re-use, reproducibility, monitoring, transparency and omni-comprehensive assessment. A Data Management Plan (DMP) contains information concerning both the research processes and the data collected, generated and/or re-used during a project's lifetime. Automated solutions and workflows that connect DMPs with the actual data and other contextual information (e.g., publications, fundings) are missing from the landscape. DMPs being submitted as deliverables also limit their findability. In an open and FAIR-enabling research ecosystem information linking between research processes and research outputs is essential. ARGOS tool for FAIR data management contributes to the OpenAIRE Research Graph (RG) and utilises its underlying services and trusted sources to progressively automate validation and automations of Research Data Management (RDM) practices.

Results A comparative analysis was conducted between the data models of ARGOS and OpenAIRE Research Graph against the DMP Common Standard. Following this, we extended ARGOS with export format converters and semantic tagging, and the OpenAIRE RG with a DMP entity and semantics between existing entities and relationships. This enabled the integration of ARGOS machine actionable DMPs (ma-DMPs) to the OpenAIRE OSG, enriching and exposing DMPs as FAIR outputs.

Conclusions This paper, to our knowledge, is the first to introduce exposing ma-DMPs in OSGs and making the link between OSGs and DMPs, introducing the latter as entities in the research lifecycle. Further, it provides insight to ARGOS DMP service interoperability practices and integrations to populate the OpenAIRE Research Graph with DMP entities and relationships and strengthen both FAIRness of outputs as well as information exchange in a standard way.

*Correspondence:
Elli Papadopoulou
elli.p@athenarc.gr

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords Research Data Management, Data Management Plans, FAIR DMPs, Machine Actionable, Research Graphs, Knowledge graphs, Open Science

Background

Funders' and institutions' mandates on open and FAIR [7]¹ data management resulted in new demands for responsible research conduct, affecting traditional practices followed by stakeholders and posing a dynamic cultural shift in science and scholarly communication, globally. The UNESCO Recommendation on Open Science is a clear indication towards that direction, uniting nations and their leaders into acknowledging the benefits and acting upon Open Science adoption for sustainable and prosperous societies [17]. In Europe, responsible data management is considered a vital part of Open Science incorporated in Framework Programmes (Horizon2020, HorizonEurope), policy documents (Recommendation on access to and preservation of scientific information) [13] and implementation actions (European Open Science Cloud) [2]. Researchers and students are expected to embrace and thrive in the new paradigm of Open Science since their grants or graduations heavily rely on complying with Research Data Management (RDM) policies. In this context, one way RDM can be supported is through the use of tools that facilitate the writing of Data Management Plans (DMPs).

DMPs have traditionally been *free-form text documents* that describe the processes, tools, and effort undertaken in data handling and management activities, including FAIR implementation, throughout a project lifecycle. The structure and expected content of the DMPs is provided in *model templates* prepared by the respective authorities who will receive researchers' DMPs upon its submission. The most common requirements for DMPs are: (1) Data description and collection or re-use of existing data, (2) Documentation and data quality, (3) Storage and backup during the research process, (4) Legal and ethical requirements, codes of conduct, (5) Data sharing and long-term preservation, (6) Data management responsibilities and resources [20]. Today, DMPs are expanding their purpose to include management of software and other research outputs [21].

There have been early efforts to facilitate the DMP authoring process involving the development of software that guides researchers into the DMP contents and instructs them on how to answer the required questions to then download and share their work as .pdf files. DMPRoadmap² is the first software for DMPs, developed back in 2010-11 by the Digital Curation Centre and the University of California Curation Center. As the first tool

to provide DMP workflows and guidance, DMPRoadmap had been widely adopted by the scientific community. In the later years, however, community needs have raised to accommodate intrinsic elements of the evolving FAIR ecosystems (e.g. Persistent Identifiers - PIDs, standards, machine readable licenses) calling for new research data services to be built and others to be reformed according to those standards [18]. Following such demand, new DMP software solutions are developed, such as ARGOS³, Data Stewardship Wizard⁴ and the Research Data Management Organizer⁵.

Recently, Research and Innovation (R&I) ecosystems have started to realize Knowledge Graphs for their ability to present a contextualized view of current research that supports discovery, re-use, reproducibility, monitoring, transparency and omni-comprehensive assessment [1]. In particular, Open Science Graphs (OSGs) are scientific knowledge graphs representing different entities of the research lifecycle (e.g. projects, people, research outcomes, institutions) and the relationships among them. They provide insight on the wealth of information in Open Science environments which, according to RDA⁶, can be classified to at least the following subsets: thematic graphs, citation graphs and monitoring graphs⁷. Examples of OSGs are Open Citations⁸, which focuses on bibliographic and citation data; Research Graph⁹, which connects scientific literature, research data, project grants and researchers; the Open Research Knowledge Graph, which focuses on research papers and the knowledge available in their full-texts; the FREYA PID Graph¹⁰, which includes interlinked entities provided they are assigned a persistent identifier; and the OpenAIRE Research Graph (OpenAIRE RG)¹¹[5], which includes descriptive metadata, provenance metadata and links between research results of any kind (literature, software, datasets, and other), organisations, grants, research communities and infrastructures.

Although important, DMPs are not yet integral parts of the ecosystem, and consequently, OSGs. Moreover, the

¹ Findable, Accessible, Interoperable and Reusable.

² DMPRoadmap source code: <https://github.com/DMPRoadmap/roadmap>.

³ ARGOS/OpenDMP source code: <https://code-repo.d4science.org/MaDgiK-CITE/ARGOS>.

⁴ Data Stewardship source code: <https://github.com/ds-wizard/>.

⁵ RDMO source code: <https://github.com/rdmorganiser>.

⁶ Research Data Alliance.

⁷ Open Science Graphs for FAIR data Interest Group: <https://www.rd-alliance.org/groups/open-science-graphs-fair-data-ig>.

⁸ <https://opencitations.net/>.

⁹ <https://researchgraph.org/>.

¹⁰ <https://www.project-freya.eu/en>.

¹¹ <https://graph.openaire.eu>.

adoption of ma-DMPs, i.e. machine actionable exports of DMPs that provide a structured way of organizing their content, is limited. Until today, DMPs are typically produced and submitted as deliverables outside of scientific workflows that support the publication of FAIR research outputs. Those practices pose several limitations on the DMP lifecycle, from the discovery and re-use of DMPs, to their enhancement with other information and further exploitation [15, 16].

In this respect, ARGOS (argos.openaire.eu) [6] was built and integrated into OpenAIRE to provide to the scientific community a solution that produces and publishes FAIR and ‘*machine actionable*’ DMPs (ma-DMPs) [12]. OpenAIRE is the largest source for Open Access Scholarly Communication in Europe and a pillar infrastructure to the architecture and operations of the EOSC¹². Since 2007, OpenAIRE has enabled interoperability of literature and data repositories [9, 10] and has aggregated, deduplicated, contextualized and made available metadata from millions of scientific information building the OpenAIRE RG. In order to enhance DMP practices and FAIRness, ARGOS explored the possibilities offered by other underlying services of the OpenAIRE service portfolio.

First, ARGOS applied the DMP Commons Standard [18] to achieve a minimum layer of interoperability to its DMP outputs. Then, ARGOS moved with publishing its DMPs to Zenodo to complete the DMP *lifecycle* and *enhance* FAIRification of its outputs (DMPs are assigned DOIs and licenses). Finally, with DMPs being exposed in OpenAIRE (via Zenodo), the OpenAIRE RG became ARGOS main trusted source to: populate DMPs with standardized information, and create links with other outputs and activities (datasets, projects, funders).

This paper, to our knowledge, is the first to explore the modelling, representation and linking of DMPs in OSGs. It explains the mapping activity that led to ARGOS DMP service integration with the OpenAIRE RG. The paper presents ARGOS as a case study for FAIR and ma-DMPs exposed in OSGs and highlights the new potential unlocked in the DMP landscape, from discovery to enrichment and exploitation.

Construction and content

ARGOS DMP service is based on the OpenDMP open-source software¹³ that was developed in collaboration between OpenAIRE and EUDAT CDI¹⁴. OpenDMP offers the possibility of semantic tagging in a multi-granular manner (dataset level) and of connecting with other

services and workflows to validate and exchange information, automate the DMP publication and writing processes and apply the FAIR principles to its outputs.

ARGOS is an instance of the OpenDMP open source software¹⁵, which is configured in the OpenAIRE ecosystem, and is available through the OpenAIRE Service catalogue¹⁶ and the European Open Science Cloud (EOSC)¹⁷. In this section, we present how the OpenDMP ARGOS instance builds on and produces FAIR and machine actionable processes and outputs. Those are seen as prerequisites to the mapping activity between ARGOS and the OpenAIRE RG because they enrich the DMP content and help produce better mapping results.

A. OpenDMP model and interoperability

The OpenDMP software features, among other things, a Template editor which is used by Administrators who define personalised structures and rules (e.g. mandatory fields, types of input supported, such as APIs) for DMPs according to RDM policies. An example is the H2020 DMPs template [11] that can be used in ARGOS to create DMPs for projects funded under the European Commission H2020 framework programmes [4]¹⁸.

The Template editor is a key enabler of compatibility with the RDA standard as it incorporates a mechanism that maps ARGOS OpenDMP fields contained in templates to template entities, including RDA dataset entities.

OpenDMP has two more editors: the DMP Editor and the Dataset editor (embedded in the DMP editor) which are used by researchers who describe their datasets/outputs (e.g. how FAIR principles were followed throughout the RDM lifecycle) and project information (e.g. funding and beneficiaries) in DMPs. DMPs produced in OpenDMP enable values in DMPs to be fetched from external APIs, like OpenAIRE APIs¹⁹, ORCID API²⁰, and EOSC API²¹.

An essential element for achieving interoperability with standards and 3rd party systems, given the flexibility of the ARGOS underlying data model, which assumes almost unlimited user-driven structuring of dataset descriptions, is the concept of “semantic tags”. Semantic tags are tags that are attached to elements of a data set description template by their authors, so that, in due

¹²<https://www.openaire.eu/openaire-and-eosc>.

¹³<https://gitlab.eudat.eu/dmp/OpenAIRE-EUDAT-DMP-service-pilot/-/tree/master>.

¹⁴<https://eudat.eu/eudat-cdi>.

¹⁵Wiki page: <https://gitlab.eudat.eu/dmp/OpenAIRE-EUDAT-DMP-service-pilot/-/wikis/home>.

¹⁶<http://catalogue.openaire.eu/service/openaire.ARGOS>.

¹⁷<https://marketplace.eosc-portal.eu/services/argos?fromc=data-management>.

¹⁸The Horizon Europe template is also available in ARGOS.

¹⁹Application Programming Interfaces.

²⁰<https://orcid.org/>.

²¹<https://providers.eosc-portal.eu/openapi>.

time, they utilized properly by an informed component that handles an intended action, be it an export or an interaction with an external system.

B. ARGOS lifecycle and FAIR DMPs

A DMP's lifecycle in ARGOS consists of a predefined number of internal states followed by versioning and publication workflows. A DMP is in draft state while being composed, and while draft it might be either validated (i.e. passing all the requirements of the dataset templates that each of its contained datasets conform to) or non-validated, which is computed substate. Validated DMPs can advance to a finalized state, where editing is restricted, unless the DMP is reverted to draft again. The DMP may be published both internally, i.e. listing of published DMPs offered by the service, or externally.

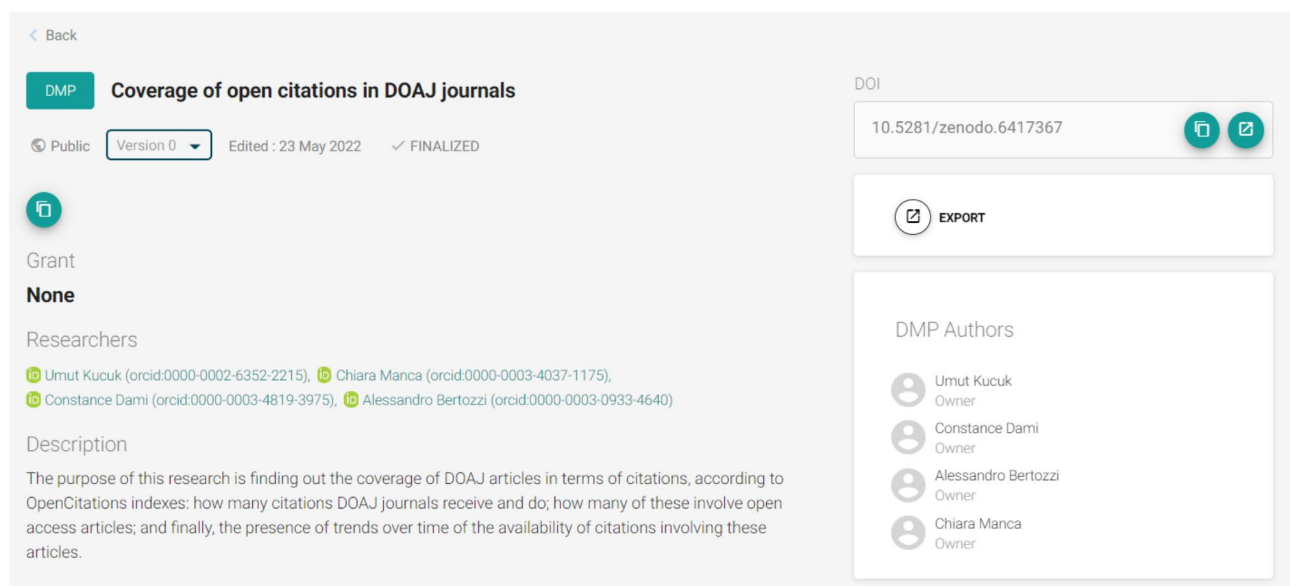
As shown in Figs. 1 and 2, ARGOS is fully integrated with Zenodo²², the catch-all repository of OpenAIRE hosted by CERN, and thus offers the option to publish DMPs as outputs in an open and FAIR manner (.pdf and .json DMP representations), by assigning DOIs²³, descriptive metadata, and licenses and by supporting DMPs as living documents through versioning.

Once a DOI is issued from Zenodo, reverting a DMP to draft state is not possible and a new version must be created in order to add modifications. Following one cycle of publication, new versions of a DMP may be created and linked to their predecessor ones, while the latter continue to survive in the system for reference.

The following table highlights how the ARGOS OpenDMP instance has enabled FAIRification of its outputs Fig. 3.

In summary, ARGOS DMPs are findable because they are minted DOIs directly from Zenodo, which in turn is registered in the re3data data repositories registry²⁴ and is indexed by OpenAIRE. Metadata of ARGOS DMPs include PIDs, such as ORCIDs for researchers, and are exposed to Zenodo, shared under CC0 license. To access work performed on DMPs, OpenDMP supports a number of configurable providers for login to the system (SAML 2.0/OpenID Connect/ OAuth2.0), including OpenAIRE for the ARGOS instance. Additionally, ARGOS DMPs are provided with access rights (open and restricted) to allow specific conditions to be set along with their license information. To make DMPs interoperable and enable their exchange, apart from the DMP Common Standard supported, DMPs are exported in formats that are machine readable and machine actionable, namely .xml and .json. The xml format is a full native representation of a DMP, exposing ARGOS model, containing elements accompanied with tags that may be utilized for interoperability with other systems. The .json format conforms to the specs introduced by RDA ma-DMP working group. From the Research Graph, links are created between ARGOS DMPs and other outputs. Furthermore, DMPs can be versioned at any time of the DMP lifecycle and are assigned licenses to enhance reusability practices.

²²Zenodo <https://zenodo.org>.



Back

DMP Coverage of open citations in DOAJ journals

Public Version 0 Edited: 23 May 2022 FINALIZED

DOI: 10.5281/zenodo.6417367

EXPORT

DMP Authors

- Umüt Kucuk (orcid:0000-0002-6352-2215) Owner
- Constance Dami (orcid:0000-0003-4037-1175) Owner
- Alessandro Bertozzi (orcid:0000-0003-0933-4640) Owner
- Chiara Manca (orcid:0000-0003-4819-3975) Owner

Description

The purpose of this research is finding out the coverage of DOAJ articles in terms of citations, according to OpenCitations indexes: how many citations DOAJ journals receive and do; how many of these involve open access articles; and finally, the presence of trends over time of the availability of citations involving these articles.

Fig. 1 Published DMP record (with DOI) in ARGOS platform

²³Digital Object Identifiers.

²⁴<https://www.re3data.org/repository/r3d100010468>.

There is a **newer version** of this record available.

April 6, 2022 Data management plan Open Access

Coverage of open citations in DOAJ journals

Constance Dami
 Project member(s)
 Constance Dami
 Researcher(s)
 Constance Dami (orcid:0000-0003-4819-3975); Alessandro Bertozzi (orcid:0000-0003-0933-4640); Umut Kucuk (orcid:0000-0002-6352-2215); Chiara Manca (orcid:0000-0003-4037-1175)

The purpose of this research is finding out the coverage of DOAJ articles in terms of citations, according to OpenCitations indexes: how many citations DOAJ journals receive and do; how many of these involve open access articles; and finally, the presence of trends over time of the availability of citations involving these articles.

Preview

Cannot preview file
 Sorry, we are unfortunately not able to preview this file.

Name	Size	
DMP_None_0.pdf <small>md5:66f5de25c2050d36424d38f3deb53ff7</small>	118.6 kB	Preview Download
DMPNone0.json <small>md5:b6ab90da64da7ccbe2049ff44bc2278d</small>	4.1 kB	Preview Download

182 views
56 downloads
[See more details...](#)

Indexed in

Publication date:
April 6, 2022

DOI:
DOI: [10.5281/zenodo.6417368](https://doi.org/10.5281/zenodo.6417368)

Related identifiers:
 Identical to
<https://argos.openaire.eu/external/zenodo/12d0cab3-21ca-4023-ab21-98b9b63a909a>

Communities:
Argos DMP outputs

License (for files):
[Creative Commons Attribution 4.0 International](#)

Versions

Version 4 10.5281/zenodo.7418355	Dec 9, 2022
Version 3 10.5281/zenodo.6573767	May 23, 2022
Version 2 10.5281/zenodo.6560550	May 18, 2022
Version 1 10.5281/zenodo.6491184	Apr 26, 2022

Fig. 2 The published DMP of Fig. 1 in Zenodo exposed in both .pdf and .json

Findability	Accessibility	Interoperability	Reusability
<ul style="list-style-type: none"> ● PIDs of DMPs ● Indexed by repositories and aggregators ● DMP Metadata 	<ul style="list-style-type: none"> ● AAI ● Licensing Info, Access rights ● REST API 	<ul style="list-style-type: none"> ● DMP Common Standard ● Linked metadata (Qualified references) ● XML & OpenXML exports 	<ul style="list-style-type: none"> ● Licensing DMPs ● Versioning

Fig. 3 FAIR DMPs in ARGOS

Methods

The process of integrating ARGOS FAIR ma-DMPs with the OpenAIRE Research Graph concentrated on mapping the data models of OpenAIRE [5], OpenDMP [19] and the DMP Common Standard. This was a vital step to build the framework for information exchange and semantics adoption between a ma-DMP service and an Open Science Graph.

The aim was to identify what information captured in DMPs was already available in the graph, and vice versa, as well as how entities and properties are expressed in all settings, thus contributing to the mapping results. Since the DMP Common Standard is used to exchange DMPs in a standard way, it had a prominent place in this activity for providing the minimum set of DMP related properties that comprise a ma-DMP. Then, identified gaps and weaknesses were reported and prioritized for implementation as new entities and properties in both ARGOS and the OpenAIRE RG.

The mappings were further analysed to report possible new relationships and properties that were not included in metadata records of DMPs found in OpenAIRE, but that are instead available in ma-DMPs like those produced and exposed by ARGOS.

Of particular interest is the RDA Hackathon²⁵ that was organized in May 2020, which provided the opportunity for active conversation among DMP providers and the RDA community²⁶. ARGOS winning the Hackathon accelerated advancements in terms of interoperability and machine-actionability, also with respect to integrations with the OpenAIRE RG.

Results

The comparative study [14] between OpenDMP, OpenAIRE RG and the DMP Common standard models revealed common entities, properties, and relationships, clearly portraying the state-of-the-art for DMPs as well as the opportunities arising with the adoption and publication of ma-DMPs in OSGs.

Mapping between ARGOS and the DMP Common Standard

The mapping highlighted the need for updates in ARGOS to be fully compliant to the RDA standard and strengthen its ma-DMP outputs. In Fig. 4, we show the entities and their specific properties that were found to be absent from ARGOS.

Mapping between the OpenAIRE RG and DMP Common Standard

Given that some researchers publish their DMPs in repositories, the OpenAIRE RG already included metadata about few DMPs made available by content providers as records compliant to the OpenAIRE guidelines [9, 10]. Metadata includes properties and relationships that are useful for findability (persistent identifiers), accessibility (landing pages and download URLs of different versions of DMPs, access right information), citation and discovery (bibliographic metadata properties), and for tracking (links to funding grants and involved organisations). Observing the DMPs that were available in the OpenAIRE RG, indicated that they were “disguised as project deliverables”. Hence, OpenAIRE exploited the COAR vocabulary which introduced resource type “DMP”²⁷ in the latest release to enable the proper typing of DMPs that would clearly identify them on the OpenAIRE RG. Following that, Zenodo introduced a new type of resource so that users can select the type “Data Management Plan” upon deposition. Because ARGOS uses Zenodo as its publishing mechanism and Zenodo is exploited by OpenAIRE as one of its core data sources, ARGOS published ma-DMPs become immediately an integral part of the OpenAIRE RG. That way, ARGOS ma-DMPs are exposed with proper resource_type and are treated as independent entities in the OpenAIRE RG contributing to making searching for DMPs and their links with other outputs an established process.

As it is shown in Fig. 5, when analysing the mapping of the OpenAIRE and DMP Common Standard models, we observed that most of the ma-DMP entities can be mapped directly to the OpenAIRE RG Entities.

Those entities are:

- Identifier (*DMP_id*), indicating the persistent identifier of the DMP record.
- Title of the DMP.
- Date of creation and modification of the DMP record.
- Description regarding the context that the DMP is created.

Additionally, Fig. 6 shows that the mapping activity identified areas where the OpenAIRE RG can be strengthened.

ARGOS ma-DMPs enrich the OpenAIRE RG with the identified entities and properties using researchers’ input. This input might be inferred from an API selection, such as. OpenAIRE, EOSC, Zenodo, (authoritative source) or be the compilation of free text statements provided in the DMP (non authoritative source). In Fig. 7, we show which entities and properties and in what way ARGOS can provide them to enrich the graph with missing information.

²⁵RDA Hackathon on ma-DMPs: <https://github.com/RDA-DMP-Common/hackathon-2020>. A full report from ARGOS participation in the Hackathon will be published by the Data Science Journal (<https://datascience.codata.org/>).

²⁶<https://www.rd-alliance.org/groups/active-data-management-plans.html>.

²⁷https://vocabularies.coar-repositories.org/resource_types/c_ab20/.

Entity in the Common Standard	Properties in the Common Standard	Action(s) Taken in ARGOS
Contact	Contact_id (identifier; type) Mbox	Entity exists and new properties are added <ul style="list-style-type: none"> - to expose email information of the DMP creator. - to expose the role of contributors in the DMP process.
Contributor	Role	
Cost	Value	Introduced as new entity to provide information about expenditure per dataset and DMP (total).
Dataset	Language	Extending language reference at the level of datasets.
License_Dataset	Start_date	New property of license accommodates access rights imposed by start_date property.
Metadata	Description Language	Extending language reference at the level of metadata.
Ethical_issues_description	-	To be introduced as new entity.
Ethical_issues_exist	-	To be introduced as new entity.
Ethical_issues_report	-	To be introduced as new entity.
Project	End Grant_id (Funding_status) Start	Extending project entity , to include start-end date and funding status.

Fig. 4 DMP Common Standard elements not in ARGOS

Particularly for dataset metadata about ethics, security, quality, and preservation available in ma-DMPs, OpenAIRE could consider revisiting its guidelines [9, 10] so that this information always reaches the OpenAIRE data model from data deposits in OpenAIRE compliant repositories.

Besides expanding information captured in the OpenAIRE RG, additional efforts are needed to create value out of this captured information. Below we describe what relation types are available in the OpenAIRE RG and what changed to accommodate links between ma-DMPs and other entities.

The OpenAIRE Research Graph features a dedicated relation type between research products and project grants: “isProducedBy/produces” (a product is produced by a project; a project produces research product). This

type of relationship can exist between projects and research products of any type (publications, datasets, software and other types of research products), and is used by OpenAIRE to model the association between DMPs and project grants. Considering the importance assigned by funders to DMPs (e.g. for H2020 grants, the DMPs is a “living deliverable” that must be updated frequently during the lifetime of the project), and to support the development of added-value services on top of machine actionable DMPs, the OpenAIRE Research Graph will include a relationship with a dedicated semantics “hasDMP”/“hasProject”.

Interestingly, the DMPs that are currently available in the OpenAIRE Research Graph do not include explicit links to the datasets they refer to. Clearly, such links may not exist when the first version of a DMP is

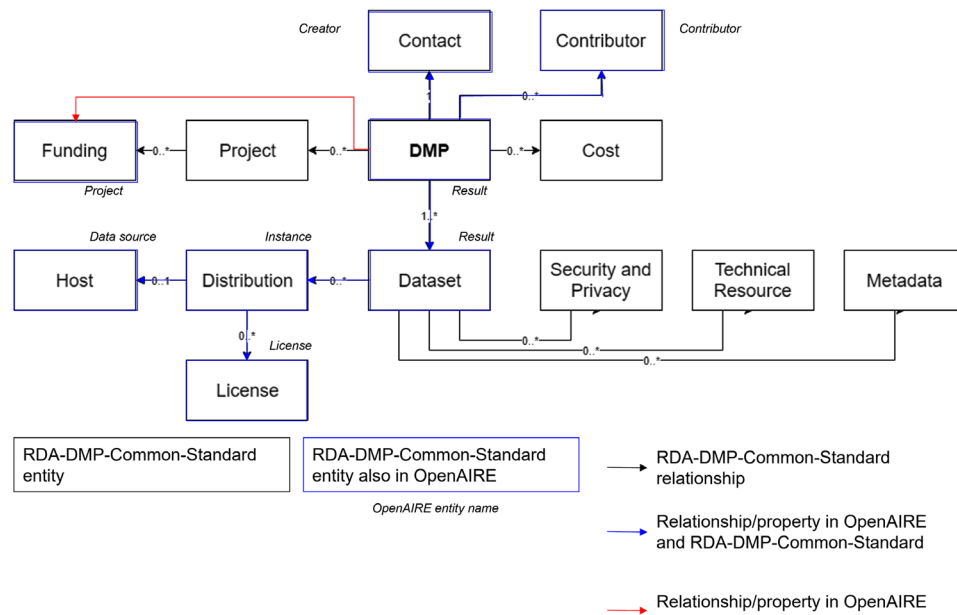


Fig. 5 RDA standard and the OpenAIRE Research Graph Model

published - because the datasets may not yet exist -, but the expectation is that the datasets are made available during the life-time of the project. Thanks to the different versions of the DMPs, therefore, it would be possible for OpenAIRE to add relationships with specific semantics between a DMP and the referred datasets. The relationships available in the OpenAIRE Research Graph, inspired from CERIF [8] and Datacite²⁸, do not include a specific semantics that could depict the association between a DMP and its datasets. OpenAIRE is therefore planning to add a new relationship with semantics “has-Dataset”, drawn from the core ontology of the RDA standard [3], and a corresponding inverse relationship, which is instead not defined in the standard, to link datasets to the DMPs (“hasDMP”).

Utility and discussion

Observations made during the mapping activity of entities and properties between the three data models of OpenDMP, OpenAIRE RG and DMP Common Standard showed that when direct mapping couldn't be fulfilled, information might still be able to be found in more abstract/ general fields of OpenAIRE or ARGOS though diverged in cardinality and / or data type; some information might still be covered by ARGOS DMP outputs as they enter the OpenAIRE RG or tweaked to accommodate the needs of ma-DMPs documentation or, rarely, they may be omitted in information exchange. OpenAIRE, also, highlighted the value added in contextualising ma-DMPs as they contain structured and specialised

information, especially about datasets, which cannot be found in original / traditional DMP documents.

Searching OpenAIRE for DMP outputs outside ARGOS, showed that DMPs are still typed as generic publications or reports, and not as data management plans. The introduction of a specific term for DMPs in global vocabularies, such as COAR's, is expected to improve the current status, although the adoption of the new term may take some time to be widespread. The recent update of the Datacite metadata schema²⁹ which includes DMP IDs, will significantly contribute towards wider adoption by service providers.

Furthermore, in order to conform to the needs of ma-DMP fixed schema, without losing the versatility of its templating mechanism, ARGOS software follows an approach that engages an extensible mechanism for attaching export format converters (ma-DMP being one of them) and semantic tagging of template elements that can be used “at-will” by those converters. The ma-DMP converter makes use of its knowledge of the fixed part of ARGOS data model as well as attributes attached to various dataset description fields in order to pick the data required for a ma-DMP file.

The flexibility of ARGOS machine actionable templating system combined with the integration of the publication mechanism of Zenodo and the interlinking with the OpenAIRE RG is crucial as it fosters change in the scene for DMPs content, distribution, exploitation, and reusability. According to the recent study of DMPs in Horizon

²⁸ https://support.datacite.org/docs/relaiontype_for_citation.

²⁹ <https://support.datacite.org/docs/a-crosswalk-of-datacite-and-rda-mapping>.

Entity in the Common Standard	Properties in the Common Standard	Action(s) Taken
Cost	Currency_code Description Title Value	Introducing new entity, Cost, to provide information about expenditure per dataset and DMP (total).
Dataset	Data_quality_assurance Preservation_statement Metadata Security_and_privacy (description; title)	Complementing Dataset, to include information about quality processes, security and preservation of data.
Dataset_Distribution	Byte_size	Complementing Distribution, to include the size of data per distribution.
Host	Backup_frequency Backup_type Storage_type Availability	Complementing Host, to capture additional elements regarding capacity, such as storage and back up.
License_Dataset	Start_date	Complementing License, to accommodate access rights imposed by start_date property.
Metadata	Description Language Metadata_standard_id (identifier; type)	Introducing new entity, Metadata, to provide information about standards used.
Ethical_issues_description	-	Entity to be introduced.
Ethical_issues_exist	-	Entity to be introduced.
Ethical_issues_report	-	Entity to be introduced.
Contact	Mbox	Complementing Contact to expose email addresses.
Contributor	Mbox	Complementing Contributor to expose email addresses.
Technical Resource	Description Name	Complementing Technical resource, to show links with software and instruments used in data management.
Project	Description	Complementing project, to include description that gives the context of the research performed.

Fig. 6 DMP Common Standard properties and entities not in OpenAIRE

Entities	Properties - Non authoritative source	Properties - Authoritative source
Contact		email
Contributor		email
Cost	Description Title Value	Currency_code
Dataset	Data_quality_assurance Preservation_statement Security_and_privacy (description; title)	Metadata
Metadata	Description Language Metadata_standard_id (identifier; type)	Description Language Metadata_standard_id (identifier; type)
Dataset_Distribution	Byte_size	
Host	Backup_frequency Backup_type Storage_type Availability	
License_Dataset	Start_date	
Technical_Resource	Description Name	Description Name
Ethical_issues_description		
Ethical_issues_exist		
Ethical_issues_report		
Project	Description	

Fig. 7 ARGOS DMP elements enriching the OpenAIRE RG

2020 [15], there are more than 1500 publicly available DMPs in CORDIS that most of them do not indicate the ways in which they can be re-distributed and used. That practice reverts the situation to the known problem of open access. The licenses assigned to ARGOS DMPs and the access rights indicated in its context, eliminate such confusion. Moreover, another study on Monitoring the Open Access Policy of Horizon 2020 [16], proves the importance of individual dataset descriptions in the exploitation of datasets and DMPs. That way, all datasets described in DMPs become individually identifiable in a granular way that links them with their associated metadata, repositories, back up policies etc. Similarly, re-used datasets become identifiable and can be further quantified, analysed and validated in reproducibility studies.

Moreover, it is foreseen that the OpenAIRE data model will be complemented with information that is currently not present, such as DMP cost or metadata. Cost information is not crucial to improve the level of FAIRness of DMPs or datasets, but it might enable analysis about data management costs useful to funders, organisations, and project administrators. Similarly, this kind of information is potentially useful to studies about responsible research and may be integrated to serve specific analysis or use cases related to RRI (Responsible Research & Innovation) monitoring.

Finally, this case study explored information carried out in published, i.e. DOIed DMPs. An opportunity that ARGOS has already started to exploit with Zenodo is pre-filling of ma-DMPs with information from deposited

metadata records. This automation highly contributes to the needed cultural shift that Open Science cultivates and incentivizes researchers to follow best practices while minimizing their compliance effort. Among the issues being examined in this implementation are capturing of the state of datasets that are pre-filled (e.g. reused dataset in a DMP) and adopting the pre-filling mechanism in other repository software, such as Dataverse. Further connections with data analysis PIDs are expected in the context of the Research Analysis Identifier System - RAISE project.

Conclusion

The paper aimed to provide a case study of how machine actionable DMPs can be contextualised and exploited in an Open Science Graph, and specifically the OpenAIRE Research Graph. ARGOS, the OpenAIRE DMP service, contributed to this work by paving the way to exposing ma-DMPs in OpenAIRE.

The paper explained how ARGOS is structured, how it applied the RDA standard to increase interoperability and machine-actionability of its outputs, how it integrated with other services to increase openness and FAIRness, and how it enriched the OpenAIRE RG. This is an ongoing effort between ARGOS and the OpenAIRE RG, supported thanks to the OpenAIRE API providing information about organizations, data sources and datasets, obtained by collecting metadata records from more than 12 K trusted scholarly communication sources (including Datacite, Crossref, re3data, OpenDOAR, Grid.ac).

The results of the comparative analysis that was performed between the data models of ARGOS and OpenAIRE Research Graph against the DMP Common Standard extended ARGOS with export format converters and semantic tagging, and the OpenAIRE RG with a DMP entity and semantics between existing entities and relationships. This enabled the integration of ARGOS machine actionable DMPs (ma-DMPs) to the OpenAIRE OSG, enriching and exposing DMPs as FAIR outputs.

Today, the work of this paper has shaped ARGOS in becoming the only DMP provider that completes the lifecycle by exposing its DMPs, including the ma-DMPs, directly to repositories and the only one to apply the whole set of the FAIR principles to its outputs so that DMPs are Findable, Accessible, Interoperable and Reusable.

Abbreviations

API	Application Programming Interface
DMP	Data Management Plan
DOI	Digital Object Identifier
EOSC	European Open Science Cloud
FAIR	Findability, Accessibility, Interoperability, Reusability
ma-DMPs	machine-actionable DMPs
OpenAIRE RG	OpenAIRE Research Graph

OSGs	Open Science Graphs
PID	Persistent Identifier
RDA	Research Data Alliance
RDM	Research Data Management
RG	Research Graph
RRI	Responsible Research & Innovation
SMEs	Small Medium Enterprises
URL	Uniform Resource Locator

Acknowledgements

The authors would like to thank the participants of the Damalos 2020 Conference and Spiros Athanasiou for their comments on the work presented in this paper.

Authors' contribution

EP and AB drafted the paper. EP, GK, DT and NM contributed information for the ARGOS service. AB and PM contributed information for the OpenAIRE Research Graph. All authors read and approved the submitted manuscript.

Funding

This work is partially supported by grant 101058479 of the European Commission.

Data Availability

The datasets generated and/or analysed during the current study are available in the ZENODO repository, <https://doi.org/10.5281/zenodo.5047766>.

Declarations

Ethics approval and consent

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests. Note: <https://zenodo.org/record/3944314#.YNzGbOgzY2w>.

Author details

¹OpenAIRE AMKE, 6 Artemidos, Athens 15125, Greece

²Consiglio Nazionale delle Ricerche, 1 Via Moruzzi 56124, Pisa, Italy

³Communication and Information Technologies Experts S.A, 22 Omiriou, Athens 16122, Greece

Received: 4 August 2021 / Accepted: 11 September 2023

Published online: 02 November 2023

References

- Aryani A, Fenner M, Manghi P, Mannocci A, Stocker M. Open Science Graphs must interoperate! 24th Int Conf Theory Pract Digit Libr (TPDL). 2020;Lyon, France. https://doi.org/10.1007/978-3-030-55814-7_16.
- European Commission. European Cloud Initiative - Building a competitive data and knowledge economy in Europe. Luxembourg: Office for Official Publications of the European Communities; 2016.
- Fajar J, Ekaputra, João Cardoso, Leyla Garcia, Marie Christine Jacquemot. The standard Ontology. Retrieved from: <http://semantics.id/ns/dco/index-en.html>.
- Data Management, European Commission Webpage., https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm, last accessed 2020/08/17.
- Manghi P, Bardi, Alessia, Atzori, Claudio, Baglioni, Miriam, Manola, Natalia, Schirrwagen, Jochen, & Principe, Pedro. (2019, April 17). The OpenAIRE Research Graph Data Model (Version 1.3). Zenodo. <https://doi.org/10.5281/zenodo.2643199>.

6. Papadopoulou E, Kakaletis, Georgios, Tziotzios, Diamadis, Moa, Hanne, & Hasan, Adil. (2020). ARGOS: a collaborative tool to plan and follow your data. Zenodo. <https://doi.org/10.5281/zenodo.3898249>.
7. Wilkinson M, Dumontier M, Aalbersberg I, et al. *Sci Data*. 2016;3:160018. <https://doi.org/10.1038/sdata.2016.18>. The FAIR Guiding Principles for scientific data management and stewardship.
8. CERIF data model. : <https://eurocris.org/services/main-features-cerif>.
9. Houssos N, et al. OpenAIRE guidelines for CRIS managers: supporting interoperability of open research information through established standards. *Procedia Comput Sci*. 2014;33:33–8.
10. Schirrwagen J, Baglioni M. (2018). OpenAIRE Guidelines for institutional and thematic repository managers 4.0 (Version 4.0.0). Zenodo.
11. European Commission. Template Horizon 2020 Data Management Plan (DMP). https://ec.europa.eu/research/participants/data/ref/h2020/other/gm/reporting/h2020-tpl-oa-data-mgt-plan-annotated_en.pdf.
12. Miksa T, Simms S, Mietchen D, Jones S. Ten principles for machine-actionable data management plans. *PLoS Comput Biol*. 2019;15(3):e1006750. <https://doi.org/10.1371/journal.pcbi.1006750>.
13. Official Journal of the European Union. COMMISSION RECOMMENDATION (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information. [online] Available at: <https://eur-lex.europa.eu/eli/reco/2018/790/oj>.
14. Papadopoulou E, Bardi A. (2021). Mapping Across OpenAIRE And ARGOS Data Models And The DMP Common Standard [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5047766>.
15. Spichtinger D. *Open Res Europe*. 2021;1:42. <https://doi.org/10.12688/openreseurope.13342.1>. Data Management Plans in Horizon 2020: what beneficiaries think and what we can learn from their experience [version 1; peer review: 2 approved, 1 approved with reservations].
16. European Commission. Report. Monitoring the Open Access Policy of Horizon 2020, <https://op.europa.eu/en/publication-detail/-/publication/56cc104f-0ebb-11ec-b771-01aa75ed71a1/language-en#>.
17. UNESCO, 2021. UNESCO Recommendation on Open Science. <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>.
18. DMP Common Standard. GitHub. (<https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>).
19. Kakaletis G, et al et al. 2022. TPDL 2022 Conference Proceedings. OpenDMP: orchestrating Data Management processes in research ecosystems. [Forthcoming].
20. Science, Europe. 2021. Practical Guide to the International Alignment of Research Data Management. https://www.scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf.
21. European Commission. 2021. Horizon Europe: Data Management Plan Template. <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/reference-documents;programCode=HORIZON>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.