Open Science governance: the role of persistent identifiers and metadata standards

ISABEL ABEDRAPO ROSEN [1]

PABLO SÁNCHEZ-NÚÑEZ [2]

RICARDO HARTLEY BELMAR [3]

*Author affiliations can be found in the back matter of this article

ABSTRACT

While Open Science emphasises openness and reproducibility, governance documentation does not necessarily incorporate these features. It raises concerns, especially compared to government policy mandates emphasising transparency and accountability. Persistent identifiers (PIDs) play a crucial role in enabling the discoverability, accessibility, and traceability of scholarly outputs. However, PIDs see widespread adoption among individual practitioners but slower adoption within institutional and regulatory bodies. This discrepancy leads to uneven metadata usage and highlights the need for a more unified approach to PIDs across the scholarly ecosystem. This essay analyses 46 Open Science governance documents to pinpoint essential areas for improvement. The inconsistencies across documents, the absence of digital object identifiers (DOIs), and varied recognition ability by bibliographic managers underscore the urgent need for standardisation. Embracing Open Science offers a promising avenue to unify stakeholders in a collective push towards bolstering the integrity and efficiency of research, thereby ensuring more robust governance.

KEYWORDS:

Persistent Identifiers (PIDs), Meta-research, Open Science, Governance Documentation, FAIR Data, Discoverability, Accessibility, Traceability.

CORRESPONDING AUTHOR:

ISABEL ABEDRAPO ROSEN

isa.abedrapo@gmail.com

Introduction

As noted by Wilkinson, researchers underscore the critical importance of openness and reproducibility, mainly focused on FAIR Data Principles (Wilkinson et al., 2016); despite this emphasis, a stark contrast arises when examining the governance policies at the heart of this movement; often, the structures of governance documentation inadequately reflect these core values (standardisation, version control, persistent identifiers, and comprehensive metadata). This discrepancy underscores the urgent need for governance frameworks to embody the principles of Open Science better, ensuring a coherent and robust application of its ideals throughout the scientific community (Nosek et al., 2015).

This difference becomes even more profound when juxtaposed against government policy mandates (Davis, 2015), driving us to the question: Why doesn't Open Science apply the same rigour of its mandates to its governance? Open Science's policy mandates have recognised the significance of transparency in their decision-making, formulation, and implementation processes (Gustafsson, 2019). Moreover, policies are constantly revised, necessitating transparent record-keeping of their evolutions (Cruz et al., 2019). Transparent record-keeping ensures accountability, enables public scrutiny, and guarantees that policies are based on reliable evidence. Furthermore, documenting the research, data, and analysis that inform policy decisions and maintaining explicit version control of policy documents is crucial for tracking changes and ensuring consistency over time. Recording consultation and stakeholder engagement processes remain essential (Silveira et al., 2023). It is urgent to re-envision the governance structures of Open Science, ensuring they uphold and exemplify the principles they advocate.

Though aspiring to ideals like transparency, traceability, and version control, the Open Science Framework reveals gaps in its practical application. Transparency is not uniformly implemented across disciplines, with some fields withholding data and methodologies (Collier, 2015; Gulland, 2018; Miyakawa, 2020). Similarly, traceability suffers from inconsistent documentation of data provenance, obstructing reproducibility (Hernández & Colom, 2023; Shiffrin et al., 2018). While standard in software development, version control remains underutilised for datasets and research protocols (Klump et al., 2021; Koehler Leman et al., 2020; *Research in Progress Blog Version Control for Scientific Research*, 2013). Persistent identifiers (PIDs) and metadata schemes are essential tools for findability, accessibility, and interoperability, and their consistent application is crucial for adhering to FAIR principles (Hernández & Colom, 2023). Addressing these shortcomings is necessary for improving the reliability and overall impact of scientific research.

Roles and Responsibilities in Implementing Open Science

The stakeholders involved in Open Science can vary depending on the specific context and objectives of each initiative. Different authors have explored and categorized these stakeholders (Anger et al., 2022; Boon et al., 2021; Fecher & Friesike, 2013; Pontika et al., 2015). Additionally, educational resources such as the “OpenSciency Introductory Resources on Open Science - Stakeholders of Open Science: Who practices responsible Open Science and for whom?” (Almarzouq et al., 2023) provide further insights into these roles. Across these sources, a consensus emerges around at least five key stakeholder groups: researchers or scientists, institutions and funding agencies, publishers and platform providers, policymakers, and society. Each group contributes uniquely to implementing and sustaining Open Science practices, emphasizing collaboration and shared responsibility in achieving openness, transparency, and reproducibility in research.

The distinction between Individuals or Practitioners (e.g., researchers, laboratories, and research groups) and Institutional or Regulatory Bodies (e.g., governments, policymakers, and funding agencies) highlights their distinct but interconnected roles. While practitioners focus on transparency, collaboration, and accessibility in research outputs, institutional bodies establish frameworks, standards, and indicators to guide and evaluate the adoption of Open Science principles. This differentiation recognises the operational autonomy of individuals while acknowledging the regulatory and evaluative roles of institutions. Both groups share core values such as ethics, transparency, accountability, and inclusivity, contributing to the traceability, accessibility, and interoperability of research outputs.

It becomes necessary to consider how these roles interact through technological infrastructures (Borgman, 2015; Kratz & Strasser, 2015; Piwowar & Vision, 2013). Despite the tensions within both domains, they exhibit shared characteristics that promote dialogue between these contexts. These commonalities include technical aspects, such as PIDs, connectors between entities enabling the establishment of relationships and the exchange of information contributing to FAIR principles (Barcelona Declaration on Open Research Information et al., 2024), and values-based aspects, such as ethics principles, integrity, transparency, accountability, and inclusivity.

By leveraging PIDs, both groups can enhance the reliability of the evaluation of the scientific process, thereby collaboratively advancing the goals of Open Science. These shared values and tools strengthen collaboration, promote interoperability, and foster trust in Open Science governance.

The Pivotal Role of Persistent Identifiers in Open Science Governance

Persistent identifiers (PIDs) are unique, long-lasting identifiers assigned to digital entities or objects like research articles, datasets, or researchers. Unlike other identifiers, PIDs are designed to remain stable over time, ensuring that they persist even when information systems change. They play a crucial role in enabling the discoverability, accessibility, traceability, and evaluation of scholarly outputs in Open Science (Cousijn et al., 2019; Fenner et al., 2019). High-quality metadata associated with PIDs ensures clarity and consistency in data representation, facilitating better integration and analysis across platforms. For example, metadata standards like those implemented in the DOIs allow research outputs to be seamlessly linked to funding information, institutional affiliations, and related datasets, creating a unified ecosystem that improves discoverability and reuse.

We suggest a hierarchy design that presents the relationship between groups, values, objects, and PIDs (Figure 1). This design showcases how practitioners and institutions are linked through open science practices, supported by a variety of identifiers (e.g., DOI, ORCID iD, and ROR), fostering transparency and interoperability in research outputs.

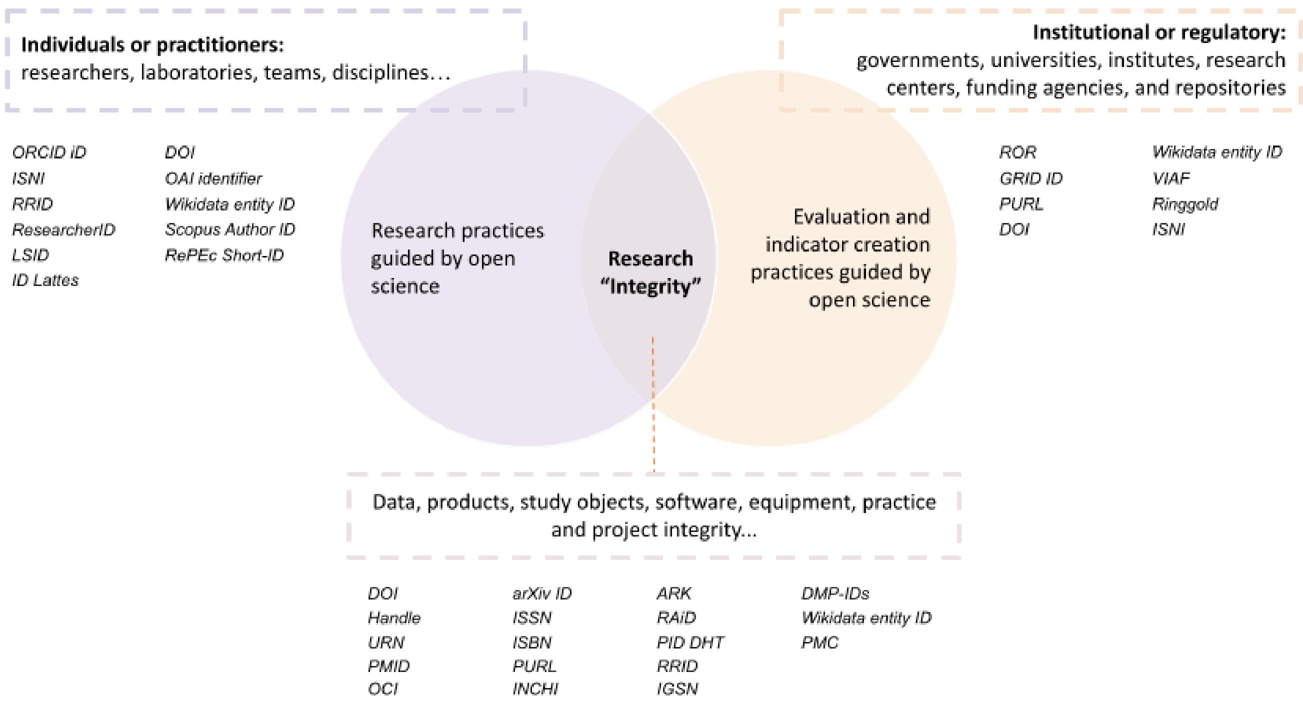

Figure 1. Persistent identifiers hierarchy design based on the following authors: (ARK Alliance, 2023; CERN, 2020; Cousijn et al., 2021; De Castro et al., 2023; Denk, 2014; Golodoniuc et al., 2017; Jones et al., 2011; Knoth et al., 2022; Marín-Arraiza, 2022; Peroni & Shotton, 2019; Seidlmayer et al., 2022; Stein et al., 2003). Source: Own elaboration.

Persistent identifiers can have several positive impacts on the governance of Open Science, highlighting their critical role in ensuring reliability, traceability, and the long-term preservation of research outputs. Here are some key points:

- **Unambiguous identification:** Accurate referencing and attribution (Piwowar & Vision, 2013).
- **Linking and discoverability:** Enhanced connectivity between research outputs (Mayernik, 2018).
- **Enhanced citation practices:** Reliable citation tracking and impact assessment (Fenner et al., 2019).
- **Reproducibility and transparency:** Easier tracing and accessing resources such as datasets, code, or other research (Starr et al., 2015).
- **Long-term preservation:** Support the long-term preservation and accessibility despite URL changes (Meadows & Haak, 2018; Parsons et al., 2019).

Additionally, the Joint Declaration of Data Citation Principles (JDDCP) (Data Citation Synthesis Group, 2014) offers principle-based recommendations for scholarly data citation, including persistent identifier schemes, resolution behaviour, metadata requirements, and best practices for ensuring machine-actionability citations. These principles are applicable to a wide range of Open Science actors, such as publishers, scholarly organisations, and persistent data repositories, and their comprehensive nature allows them to be extended to other stakeholders (Starr et al., 2015).

PIDs and their kernel, which store metadata, attributes, and values, serve as crucial links that enable interaction between PIDs and entities (Weigel et al., 2018). These connections, accessible to both machines and humans, transcend simple data exposure by creating a dynamic tapestry of interactivity, wherein each PID, through its unique identifier, plays a specific and interconnected role within the larger narrative (Golodoniuc et al., 2017; Lannom et al., 2015). These roles and relationships yield a more profound understanding of how entities within the system interact, facilitating deeper insights into the structure of the scholarly ecosystem and leading to more meaningful conclusions (Hardisty et al., 2021).

The relationships enabled by PIDs and their kernel allow for a comprehensive recognition and understanding of the system's structure, facilitating the extraction of critical insights and meaningful deductions. The relationships enabled by PIDs and their kernel allow for a comprehensive recognition and understanding of the system's structure, facilitating the extraction of critical insights and meaningful deductions (Lannom et al., 2015). By unlocking the potential for a deeper understanding of these relationships, PIDs and their kernel (DataCite Metadata Working Group, 2021) play a pivotal role in enabling the scholarly ecosystem to generate significant insights and foster innovation. Acting as a bridge, PIDs link the research outputs of individuals to the activities of institutional and regulatory bodies, such as evaluation, funding, and regulation, ensuring a cohesive and interoperable framework for Open Science governance.

Showing uncharted territories: the intermittent characteristics of a sample of Open Science governance documents

In a review exercise of a sample of 46 documents about Open Science governance, conducted in August 2023, a diverse set of institutions and regulatory actors was included, such as UNESCO, LA Referencia, and the European Commission, considering institutional diversity, public availability and geographical representation. These documents were selected based on their relevance to Open Science governance and their representation of global and regional policy frameworks. The selection aimed to include a range of geographical contexts and organizational types, ensuring a comprehensive perspective on existing practices. The review considered the following variables: Document name, file type, institution – country, PDF/A[1], presence of DOI, other PIDs, and whether the metadata was recognized by Zotero, Mendeley, and in PDF properties. The complete table is accessible at the Zenodo Repository (Hartley Belmar & Abedrapo Rosen, 2023). Table 1 shows a summary.

PDF document	Is PDF/A	Has PID	Metadata recognised by Zotero	Metadata recognised by Mendeley	Metadata in PDF properties
No	41	31	14	9	20
Yes	5	15	32	37	26

Table 1.“Review of documents about Open Science”: Synthesis source: Own elaboration.

Analysis of findings

The review revealed significant inconsistencies in the adoption of technical standards and metadata practices across the documents analyzed. Table 1 highlights the key characteristics of these governance documents, including their use of persistent identifiers (PIDs), metadata compatibility, and long-term preservation formats.

The data shows that only 31% of the documents included DOIs, providing a basic level of traceability and discoverability. In addition, the PDF/A format, essential for long-term preservation, was used in just five documents, suggesting a limited focus on ensuring accessibility over time. Metadata embedded in PDF properties was inconsistently applied, which impacted the recognition of these documents by bibliographic tools like Zotero and Mendeley.

Also seems to exist distinct patterns based on the type of issuing organization and their regional origins:

International Organizations (e.g., UNESCO, European Commission, OECD): These institutions generally provide governance documents with a high degree of standardization, often including persistent identifiers (PIDs) such as DOIs for reports and recommendations. However, metadata implementation within documents, such as using PDF/A for long-term preservation, remains inconsistent.

[1] PDF/A is an ISO-standardized version of the Portable Document Format (PDF) specialized for archiving and long-term preservation of electronic documents (Oettler, 2013).

Regional Networks (e.g., LA Referencia, COAR, African Open Science Platform): While these networks promote the adoption of PIDs, their reports display variations in DOI usage. Instead, many rely on internal or institutional identifiers, which limits interoperability across platforms.

National Governments and Funding Agencies (e.g., UKRI, ANID, NIH, NSF): These entities enforce the use of PIDs for research outputs funded under their mandates, yet their own policy documents frequently lack persistent identifiers. This disconnect between regulatory expectations and internal documentation practices weakens overall transparency.

Universities and Research Institutions: These show the most heterogeneity in standards adoption. Some institutions integrate RORs and DOIs into their governance strategies, while others still rely on closed systems or non-interoperable document references.

It is essential to consider that the use of metadata analysis techniques can be detrimental when metadata is missing from policy documents because there are structural obstacles, such as significant, non-OCR (optical character recognition) prepared and unidentified documents, as well as inconsistencies in manually tagged metadata (Laparra et al., 2023). Such gaps in standardization highlight broader challenges in the implementation of PIDs and metadata in institutional and regulatory bodies. While individual researchers have embraced these practices more widely (Porter, 2024), institutions often lag behind, resulting in documents that are less interoperable, harder to discover, and more difficult to preserve for future use.

The inconsistency in metadata also affects how these documents are accessed and cited. Zotero recognized metadata in only 14 documents, while Mendeley successfully identified metadata in 20. This limitation poses challenges for those relying on these tools to manage and cite governance documents, reducing their accessibility and usability.

We could relate those inconsistencies to several key functions in Open Science governance:

Identification and Citation: The lack of DOIs makes it difficult to properly cite key governance documents in meta-research and policy evaluation studies.

Discovery: Open Science governance documents remain difficult to index in bibliographic databases and search engines without interoperable identifiers and metadata.

Reproducibility: The absence of structured metadata within these documents hinders effective tracking of policy evolution.

Preservation: Only five analyzed documents used PDF/A, indicating that most texts lack long-term preservation assurances.

Opportunities for improvement

To address these gaps, governance bodies must prioritize the adoption of technical standards that align with Open Science principles. Recommended actions include:

- Ensuring the systematic use of PIDs, such as DOIs and RORs, in all governance documents to improve traceability.
- Standardizing metadata practices, including embedding structured metadata in PDF properties and ensuring compatibility with bibliographic tools.
- Adopting PDF/A format to enhance document preservation and long-term usability.
- Providing resources and training to institutional staff on the importance of persistent identifiers and metadata in Open Science governance.

By implementing these measures, institutional and regulatory bodies can ensure their documentation supports interoperability, transparency, and the long-term goals of Open Science.

CONCLUSION

Open Science represents a transformative vision for research, emphasizing openness, transparency, and reproducibility. However, as this brief study reveals, governance documentation often falls short of these ideals, particularly in the adoption of standardisation and PIDs. The findings highlight significant inconsistencies in metadata practices, DOI usage, and long-term preservation efforts, which hinder traceability, discoverability, and interoperability—key pillars of a cohesive Open Science ecosystem.

Persistent identifiers are not merely a technical solution; their success depends on trust, collaboration, and shared responsibility across the research ecosystem. Persistency requires more than technological infrastructure—it is a social and institutional commitment to ensuring the long-term reliability and accessibility of scholarly outputs. Without this commitment, the promise of PIDs risks failure, potentially resulting in what he describes as a "citational armageddon," undermining the foundations of Open Science (Askitas, 2010).

Bridging the gap between individual practitioners and institutional and regulatory bodies is critical to addressing these shortcomings. Governance frameworks must not only encourage the consistent adoption of PIDs but also foster the cultural and operational shifts needed to sustain them. Embedding shared values of transparency, accountability, and inclusivity into governance policies will strengthen trust and collaboration across all levels of the research ecosystem (Ananthakrishnan et al., 2020).

The recommendations outlined—mandating the use of PIDs, standardising metadata practices, adopting preservation-friendly formats like PDF/A, and providing training for institutional staff—align with the broader vision of Open Science. By integrating these measures, governance frameworks can promote interoperability and inclusivity while ensuring that Open Science principles are applied effectively in practice.

As the scientific landscape continues to evolve, aligning governance policies with these core principles will enhance the credibility and impact of Open Science. A commitment to both technical standards and shared community values will facilitate seamless collaboration, improved interoperability, and efficient knowledge dissemination.

DATA AVAILABILITY

The data supporting the findings of this study are openly available in Zenodo at https://doi.org/10.5281/ZENODO.10015205, under the title “Review of documents about Open Science”.

FUNDING INFORMATION

This work has been supported by the project TWIN4MERIT, which received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No. 101079196.

AUTHOR AFFILIATIONS

Isabel Abedrapo Rosen orcid.org/0000-0001-9990-0436
1. Universidad Central de Chile, Santiago, Chile, https://ror.org/0577avk88 Email: isa.abedrapo@gmail.com

Pablo Sánchez-Núñez orcid.org/0000-0001-7845-9506
2. OpenScienceLab, Department of Library and Information Science, Universidad Carlos III de Madrid, Getafe, Spain. https://ror.org/03ths8210 Email: psnunez@bib.uc3m.es

Ricardo Hartley Belmar orcid.org/0000-0001-5709-3533
3. Remolino Consultores, Chile rtleyb@icloud.com

AUTHOR CONTRIBUTIONS

Isabel Abedrapo Rosen: Conceptualization, Data curation, Methodology, Validation, Formal analysis, Investigation, Writing – review & editing, Supervision
Pablo Sánchez-Núñez: Conceptualization, Writing – original draft
Ricardo Hartley Belmar: Conceptualization, Methodology, Investigation, Writing – original draft, Supervision

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

Almarzouq, B., Azevedo, F., Batalha, N., Bayer, J., Bell, T., Bhogal, S., Black, M., Brown, S., Campitelli, E., Chegini, T., Dunleavy, D., Ee, Y. K., El-Gebali, S., Erdmann, C., Ferdush, J., Fouilloux, A., Hall, S. M., Kherroubi Garcia, I., Klusza, S., … Yehudi, Y. (2023). *Opensciency—A core open science curriculum by and for the research* *community* (Versión 1.0.0). Zenodo. https://doi.org/10.5281/ZENODO.7392118

Ananthakrishnan, R., Chard, K., D'Arcy, M., Foster, I., Kesselman, C., McCollam, B., Pruyne, J., Rocca-Serra, P., Schuler, R., & Wagner, R. (2020). An Open Ecosystem for Pervasive Use of Persistent Identifiers. *Practice and Experience in Advanced Research Computing*, 99-105. https://doi.org/10.1145/3311790.3396660

Anger, M., Wendelborn, C., Winkler, E. C., & Schickhardt, C. (2022). Neither carrots nor sticks? Challenges surrounding data sharing from the perspective of research funding agencies—A qualitative expert interview study. *PLOS ONE*, *17*(9), e0273259. https://doi.org/10.1371/journal.pone.0273259

ARK Alliance. (2023, febrero 4). *ARK Alliance*. ARK Alliance. https://arks.org/

Askitas, N. (2010). What Makes Persistent Identifiers Persistent? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1639996

Barcelona Declaration on Open Research Information, Kramer, B., Neylon, C., & Waltman, L. (2024). *Barcelona Declaration on Open Research Information*. https://doi.org/10.5281/ZENODO.10958522

Boon, W., Duisterwinkel, C., Strick, M., & Thunnissen, M. (2021). *Open Science & Stakeholder Engagement: Why, how, and what could be improved?* [Report]. Utrecht University. https://dspace.library.uu.nl/handle/1874/416090

Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. The MIT Press. https://doi.org/10.7551/mitpress/9963.001.0001

CERN. (2020, junio 9). *Persistent identifiers for objects*. https://library.cern/submit-and-publish/persistent-identifiers/pids-for-objects

Collier, R. (2015). Is withholding clinical trial results "research misconduct"? *Canadian Medical Association Journal*, *187*(10), 724-724. https://doi.org/10.1503/cmaj.109-5053

Cousijn, H., Braukmann, R., Fenner, M., Ferguson, C., Van Horik, R., Lammey, R., Meadows, A., & Lambert, S. (2021). Connected Research: The Potential of the PID Graph. *Patterns*, *2*(1), 100180. https://doi.org/10.1016/j.patter.2020.100180

Cousijn, H., Feeney, P., Lowenberg, D., Presani, E., & Simons, N. (2019). Bringing Citations and Usage Metrics Together to Make Data Count. *Data Science Journal*, *18*, 9. https://doi.org/10.5334/dsj-2019-009

Cruz, M., Dintzner, N., Dunning, A., Van Der Kuil, A., Plomp, E., Teperek, M., Turkyilmaz-van Der Velden, Y., & Versteeg, A. (2019). Policy Needs to Go Hand in Hand with Practice: The Learning and Listening Approach to Data Management. *Data Science Journal*, *18*, 45. https://doi.org/10.5334/dsj-2019-045

Data Citation Synthesis Group. (2014). *Joint Declaration of Data Citation Principles*. Force11. https://doi.org/10.25490/A97F-EGYK

DataCite Metadata Working Group. (2021). *DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs v4.4* [Application/pdf]. 82 pages. https://doi.org/10.14454/3W3Z-SA82

Davis, R. C. (2015). Git and GitHub for Librarians. *Behavioral & Social Sciences Librarian*, *34*(3), 158-164. https://doi.org/10.1080/01639269.2015.1062586

De Castro, P., Herb, U., Rothfritz, L., & Schöpfel, J. (2023). *The gradual implementation of organisational identifiers (OrgIDs)*. Zenodo. https://doi.org/10.5281/ZENODO.7327535

Denk, C. (2014). *ORCID, ResearcherID, Scopus Author ID - ¿Qué son y para qué sirven?* https://investigacion.us.es/docs/apoyo/ORCID.pdf

Fecher, B., & Friesike, S. (2013). Open Science: One Term, Five Schools of Thought. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2272036

Fenner, M., Crosas, M., Grethe, J. S., Kennedy, D., Hermjakob, H., Rocca-Serra, P., Durand, G., Berjon, R., Karcher, S., Martone, M., & Clark, T. (2019). A data citation roadmap for scholarly data repositories. *Scientific Data*, *6*(1), 28. https://doi.org/10.1038/s41597-019-0031-8

Golodoniuc, P., Car, N. N. J., & Klump, J. (2017). Distributed Persistent Identifiers System Design. *Data Science Journal*, *16*, 34. https://doi.org/10.5334/dsj-2017-034

Gulland, A. (2018). Drug companies are unwilling to share information on trial protocols, study finds. *BMJ*, k416. https://doi.org/10.1136/bmj.k416

Gustafsson. (2019). Learning from the Experiences of the Intergovernmental Panel on Climate Change: Balancing Science and Policy to Enable Trustworthy Knowledge. *Sustainability*, *11*(23), 6533. https://doi.org/10.3390/su11236533

Hardisty, A., Addink, W., Glöckler, F., Güntsch, A., Islam, S., & Weiland, C. (2021). A choice of persistent identifier schemes for the Distributed System of Scientific Collections (DiSSCo). *Research Ideas and Outcomes*, *7*, e67379. https://doi.org/10.3897/rio.7.e67379

Hartley Belmar, R., & Abedrapo Rosen, I. (2023). *Review of documents about Open Science* [Dataset]. Zenodo. https://doi.org/10.5281/ZENODO.10015205

Hernández, J. A., & Colom, M. (2023). *Repeatability, Reproducibility, Replicability, Reusability (4R) in Journals' Policies and Software/Data Management in Scientific Publications: A Survey, Discussion, and Perspectives*. https://doi.org/10.48550/ARXIV.2312.11028

Jones, A. C., White, R. J., & Orme, E. R. (2011). Identifying and relating biological concepts in the Catalogue of Life. *Journal of Biomedical Semantics*, *2*, 7. https://doi.org/10.1186/2041-1480-2-7

Klump, J., Wyborn, L., Wu, M., Martin, J., Downs, R. R., & Asmi, A. (2021). Versioning Data Is About More than Revisions: A Conceptual Framework and Proposed Principles. *Data Science Journal*, *20*, 12. https://doi.org/10.5334/dsj-2021-012

Knoth, P., Budko, V., Pavlenko, V., & Cancellieri, M. (2022). *OAI Identifiers: Decentralised PIDs for Research Outputs in Repositories*.

Koehler Leman, J., Weitzner, B. D., Renfrew, P. D., Lewis, S. M., Moretti, R., Watkins, A. M., Mulligan, V. K., Lyskov, S., Adolf-Bryfogle, J., Labonte, J. W., Krys, J., RosettaCommons Consortium, Bystroff, C., Schief, W., Gront, D., Schueler-Furman, O., Baker, D., Bradley, P., Dunbrack, R., … Bonneau, R. (2020). Better together: Elements of successful scientific software development in a distributed collaborative community. *PLOS Computational Biology*, *16*(5), e1007507. https://doi.org/10.1371/journal.pcbi.1007507

Kratz, J. E., & Strasser, C. (2015). Making data count. *Scientific Data*, *2*(1), 150039. https://doi.org/10.1038/sdata.2015.39

Lannom, L., Broeder, D., & Giridhar Manepalli. (2015). *Data Type Registries working group output* [Pdf]. https://doi.org/10.15497/A5BCD108-ECC4-41BE-91A7-20112FF77458

Laparra, E., Binford‐Walsh, A., Emerson, K., Miller, M. L., López‐Hoffman, L., Currim, F., & Bethard, S. (2023). Addressing structural hurdles for metadata extraction from environmental impact statements. *Journal of the Association for Information Science and Technology*, *74*(9), 1124-1139. https://doi.org/10.1002/asi.24809

Marín-Arraiza, P. (2022). Madurez de sistemas de identificadores persistentes: Oportunidades en el contexto español. *Anuario ThinkEPI*, e16a06. https://doi.org/10.3145/thinkepi.2022.e16a06

Mayernik, M. S. (2018). Scholarly resource linking: Building out a "relationship life cycle". *Proceedings of the Association for Information Science and Technology*, *55*(1), 337-346. https://doi.org/10.1002/pra2.2018.14505501037

Meadows, A., & Haak, L. (2018). How persistent identifiers can save scientists time. *FEMS Microbiology Letters*, *365*(15). https://doi.org/10.1093/femsle/fny143

Miyakawa, T. (2020). No raw data, no science: Another possible source of the reproducibility crisis. *Molecular Brain*, *13*(1), 24, s13041-020-0552-2. https://doi.org/10.1186/s13041-020-0552-2

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422-1425. https://doi.org/10.1126/science.aab2374

Oettler, A. (2013). *PDF/A in a Nutshell 2.0: PDF for long-term archiving*. Association for Digital Document Standards. https://pdfa.org/wp-content/uploads/2013/05/PDFA_in_a_Nutshell_211.pdf

Parsons, M. A., Duerr, R. E., & Jones, M. B. (2019). The History and Future of Data Citation in Practice. *Data Science Journal*, *18*, 52. https://doi.org/10.5334/dsj-2019-052

Peroni, S., & Shotton, D. (2019). *Open Citation Identifier: Definition*. 147574 Bytes. https://doi.org/10.6084/M9.FIGSHARE.7127816

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, *1*, e175. https://doi.org/10.7717/peerj.175

Pontika, N., Knoth, P., Cancellieri, M., & Pearce, S. (2015). Fostering open science to research using a taxonomy and an eLearning portal. *Proceedings of the 15th International Conference on Knowledge Technologies and Data-Driven Business*, 1-8. https://doi.org/10.1145/2809563.2809571

Porter, S. (2024, julio 26). Exploring Research Transformation through the lens of Persistent Identifiers. *Digital Science: TL;DR*. https://www.digital-science.com/tldr/article/exploring-research-transformation-through-the-lens-of-persistent-identifiers/

Research in progress blog Version control for scientific research. (2013, febrero 28). https://blogs.biomedcentral.com/bmcblog/2013/02/28/version-control-for-scientific-research/

Seidlmayer, E., Voß, J., Galke, L., & Schultz, C. (2022). *ORCID for Wikidata – Data enrichment for scientometric applications*.

Shiffrin, R. M., Börner, K., & Stigler, S. M. (2018). Scientific progress despite irreproducibility: A seeming paradox. *Proceedings of the National Academy of Sciences*, *115*(11), 2632-2639. https://doi.org/10.1073/pnas.1711786114

Silveira, L. D., Calixto Ribeiro, N., Melero, R., Mora-Campos, A., Piraquive-Piraquive, D. F., Uribe Tirado, A., Machado Borges Sena, P., Polanco Cortés, J., Santillán-Aldana, J., Couto Corrêa Da Silva, F., Ferreira Araújo, R., Enciso Betancourt, A. M., & Fachin, J. (2023). Taxonomia da Ciência Aberta: Revisada e ampliada. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência Da Informação*, *28*. https://doi.org/10.5007/1518-2924.2023.e91712

Starr, J., Castro, E., Crosas, M., Dumontier, M., Downs, R. R., Duerr, R., Haak, L. L., Haendel, M., Herman, I., Hodson, S., Hourclé, J., Kratz, J. E., Lin, J., Nielsen, L. H., Nurnberger, A., Proell, S., Rauber, A., Sacchi, S., Smith, A., … Clark, T. (2015). Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ Computer Science*, *1*, e1. https://doi.org/10.7717/peerj-cs.1

Stein, S. E., Heller, S. R., & Tchekhovski, D. (2003). An Open Standard for Chemical Structure Representation—The IUPAC Chemical Identifier. *Nimes International Chemical Information Conference Proceedings*. https://old.iupac.org/inchi/Stein-2003-ref1.html

Weigel, T., Plale, B., Parsons, M., Zhou, G., Luo, Y., Schwardmann, U., Quick, R., Hellström, M., & Kurakawa, K. (2018). *RDA Recommendation on PID Kernel Information*. https://doi.org/10.15497/RDA00031

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18