



HAL
open science

Umbrella Data Management Plans to Integrate FAIR Data: Lessons From the ISIDORE and BY-COVID Consortia for Pandemic Preparedness

Romain David, Audrey Richard, Claire Connellan, Katharina Lauer, Maria Luisa Chiusano, Carole Goble, Martin Houde, Isabel Kemmer, Antje Keppler, Philippe Lieutaud, et al.

► To cite this version:

Romain David, Audrey Richard, Claire Connellan, Katharina Lauer, Maria Luisa Chiusano, et al.. Umbrella Data Management Plans to Integrate FAIR Data: Lessons From the ISIDORE and BY-COVID Consortia for Pandemic Preparedness. CODATA Data Science Journal, 2023, DATA MANAGEMENT PLANNING ACROSS DISCIPLINES AND INFRASTRUCTURES, 22, pp.35. 10.5334/dsj-2023-035 . hal-04210490

HAL Id: hal-04210490

<https://hal.science/hal-04210490v1>

Submitted on 18 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Umbrella Data Management Plans to Integrate FAIR Data : Lessons From the ISIDORe and BY-COVID Consortia for Pandemic Preparedness

COLLECTION:
DATA MANAGEMENT
PLANNING ACROSS
DISCIPLINES AND
INFRASTRUCTURES

PRACTICE PAPER

Ju[ubiquity press

ROMAIN DAVID

AUDREY S. RICHARD

CLAIRE CONNELLAN

KATHARINA B. LAUER

MARIA LUISA
CHIUSANO

CAROLE GOBLE

MARTIN HOUDE

ISABEL KEMMER

ANTJE KEPPLER

PHILIPPE LIEUTAUD

CHRISTIAN OHMANN

MARIA
PANAGIOTOPOULOU

SARA RAZA KHAN

ARINA RYBINA

STIAN SOIL
AND-REYES

CHARLOTTE WIT

RUDOLF WITTNER

RAFAEL ANDRADE
BUONO

SARAH ARNAUD
MARSH

PAULINE AUDERGON

DYLAN BONFILS

JOSE-MARIA CARAZO

REMI CHARREL

FREDERIK COPPENS

WOLFGANG FECKE

CLAUDIA FILIPPONE

EVA GARCIA
ALVAREZ

SHERAZ GUL

HENNING HERMJAKOB

KATJA HERZOG

PETR HOLUB

LUKASZ KOZERA

ALLYSON L. LISTER

JOSÉ LÓPEZ-
CORONADO

BÉNÉDICTE MADON

KURT MAJCCEN

WILLIAM MARTIN

WOLFGANG MÜLLER

ELLI PAPAPOPOULOU

CHRISTINE M. A. PRAT

PAOLO ROMANO

SUSANNA-ASSUNTA
SANSONE

GARY SAUNDERS

NIKLAS BLOMBERG

JONATHAN EWBANK

BY-COVID PROJECT
COMMUNITY

ISIDORe PROJECT
COMMUNITY

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

ABSTRACT

The Horizon Europe project ISIDORe is dedicated to pandemic preparedness and responsiveness research. It brings together 17 research infrastructures (RIs) and networks to provide a broad range of services to infectious disease researchers. An efficient and structured treatment of data is central to ISIDORe's aim to furnish seamless access to its multidisciplinary catalogue of services, and to ensure that users' results are treated FAIRly. ISIDORe therefore requires a data management plan (DMP) covering both access management and research outputs, applicable over a broad range of disciplines, and compatible with the constraints and existing practices of its diverse partners.

Here, we describe how, to achieve that aim, we undertook an iterative, step-by-step, process to build a community-approved living document, identifying good

CORRESPONDING AUTHOR:

Dr Romain David

European Research
Infrastructure on Highly
Pathogenic Agents (ERINHA
AISBL), 98 rue du Trône
B-1050 Bruxelles, BE

romain.david@erinha.eu

KEYWORDS:

data management plan; DMP; research data management; RDM; data management quality; research data; pandemic; outbreak; data services harmonisation; data reuse; FAIR principles; data stewards

TO CITE THIS ARTICLE:

David, R, Richard, AS, Connellan, C, Lauer, KB, Chiusano, ML, Goble, C, Houde, M, Kemmer, I, Keppler, A, Lieutaud, P, Ohmann, C, Panagiotopoulou, M, Khan, SR, Rybina, A, Soiland-Reyes, S, Wit, C, Wittner, R, Buono, RA, Marsh, SA, Audergon, P, Bonfils, D, Carazo, J-M, Charrel, R, Coppens, F, Fecke, W, Filippone, C, Alvarez, EG, Gul, S, Hermjakob, H, Herzog, K, Holub, P, Kozera, L, Lister, AL, López-Coronado, J, Madon, B, Majcen, K, Martin, W, Müller, W, Papadopoulos, E, Prat, CMA, Romano, P, Sansone, S-A, Saunders, G, Blomberg, N and Ewbank, J. 2023. Umbrella Data Management Plans to

practices and processes, on the basis of use cases, presented as proof of concepts. International fora such as the RDA and EOSC, and primarily the BY-COVID project, furnished registries, tools and online data platforms, as well as standards, and the support of data scientists. Together, these elements provide a path for building an umbrella, FAIR-compliant DMP, aligned as fully as possible with FAIR principles, which could also be applied as a framework for data management harmonisation in other large-scale, challenge-driven projects. Finally, we discuss how data management and reuse can be further improved through the use of knowledge models when writing DMPs and, how, in the future, an inter-RI network of data stewards could contribute to the establishment of a community of practice, to be integrated subsequently into planned trans-RI competence centres.

Integrate FAIR Data : Lessons From the ISIDORE and BY-COVID Consortia for Pandemic Preparedness. *Data Science Journal*, 22: 35, pp. 1–15. DOI: <https://doi.org/10.5334/dsj-2023-035>

INTRODUCTION

Research data can have enduring value, as long as scientists can use, reuse and combine data sets (Jacobs & Humphrey 2004). Effective ‘data sharing’ requires that the conditions for access and reuse are clear. In the case of ‘open data’, there are no limitations. Data that is not open can still be made accessible, as long as the procedures for obtaining it are defined (Hutson 2022). The need to formalise the conditions that govern data collection, storage and reuse is one of the motivations for the establishment of data management plans (DMPs). These serve as framework documents in research projects (Smale et al. 2020) and are necessary tools for data reuse across various scientific and engineering fields. DMPs are now required for many grant applications, including those of various international and national funding bodies, such as the European Commission (see Horizon Europe Data Management Plan Template [HE DMP]) or the French National Research Agency [ANR].

There are ten essential items in any research DMP (Michener 2015): (i) funders’ requirements, (ii) data to be collected, (iii) data organisation, (iv) data documentation, (v) data quality assurance, (vi) data storage and preservation strategy, (vii) project’s data policies (licensing, ethical considerations, etc.), (viii) data dissemination, (ix) team members’ roles and responsibilities, and (x) budgetary aspects. While usually relatively straightforward for single investigator-driven projects, DMPs are more difficult to define and maintain in the case of international, multidisciplinary and/or multi-partner research projects (e.g. Stall et al. 2020). Moreover, an effective DMP must not become obsolete after a project is completed (Vines et al. 2014), and needs to address the implementation of Findable, Accessible, Interoperable, Reusable (FAIR) data principles (Wilkinson et al. 2016). Here, we describe the steps taken to establish a DMP for ‘Integrated Services for Infectious Disease Outbreak Research’ (ISIDORE), a large consortium constituted in response to the SARS-CoV-2 pandemic (Richard, Stefanyan & Ewbank 2022), with the support of its sister European Open Science Cloud (EOSC) project BeYond-COVID (BY-COVID).

CHALLENGES FOR THE ISIDORE PROJECT AND ITS DATA

INTRODUCTION TO ISIDORE

ISIDORE is a three-year project coordinated by the European Research Infrastructure on Highly Pathogenic Agents (ERINHA), funded in 2022 through the Horizon Europe Research and Innovation Programme. It aims to support researchers through the provision of free of charge access to services, to facilitate rapid research response to infectious disease threats, and improve the EU’s readiness to epidemic-prone pathogens.

ISIDORE brings together 17 major European life-science research infrastructures (RIs) and infectious disease networks, in a multidisciplinary consortium of 154 partners¹ across 32 countries. Collectively, the consortium offers access to an integrated portfolio of cutting-edge resources and research services covering fundamental research, such as structural biology and bio-imaging, as well as diagnostic, therapeutic and vaccine preclinical development, clinical research, epidemiology, social sciences and regulatory matters.

¹ The Consortium brings together 154 partners under the umbrella of ERINHA, BBMRI, EATRIS, ECRIN, Elixir, EMBC, EMERGEN, Euro-BioImaging, EU-OPENSREEN, EVA, Infrafrontier, Infravec, Instruct, MIRRI, Sonar-Global, Transvac, and VetBioNet.

DIVERSITY OF PARTNERS AND SERVICES

As an international project, ISIDORE's DMP needs to align with different federal and European requirements. While some of the RIs participating in ISIDORE had already collaborated in the frame of RI cluster projects, such as [RI-VIS](#), [CORBEL](#) and [EOSC-Life](#), the ISIDORE project represents a unique challenge, due to the diversity of participating RIs and networks, in terms of their respective sizes, expertise, legal structures and/or working practices. This required a new approach to coordination and collaboration among the different partners. It also made the elaboration of plans for data management, to favour the fast generation of FAIR policies for data sharing, reuse and interoperability, and open access to existing methods and software, such as online tools, workflows and registries, particularly challenging. Data standardisation and accessibility can, however, be maximised by adherence to established principles in data production and management processes, principally FAIR, CARE ([Carroll et al. 2020](#)) and TRUST ([Lin et al. 2020](#)), as well as by the implementation of requirements and recommendations from international fora (e.g. [CoData](#), Research Data Alliance [[RDA](#)], World Data System [[WDS](#)], [GoFAIR](#) or the [EOSC](#)). Furthermore, to avoid the duplication of effort and a lack of standardisation, it is critical to identify and adapt pre-existing online tools, community-adopted standards and processes, and implement any relevant complementary ones, especially from other multidimensional data science networks. The diversity of partners and services included in ISIDORE requires its DMP to be broadly interpretable, while preserving the necessary individual specificities, so as not to become meaningless, by preserving two aspects: the general principles that will be used to approach data management, and the specifics for services and use cases.

DIVERSITY OF PARTNER MATURITY

In addition to complying with the various laws, ISIDORE's DMP must also be compatible with the consortium partners' diverse data handling procedures. While some, especially in Life Science (LS)-RI communities with extensive experience in data handling (e.g. structural biology, [INSTRUCT](#); high-content screening, [EU-OPENSREEN](#); biological and biomedical imaging, [EuroBioImaging](#)) have a solid set of internal procedures, implemented in close exchange with the relevant repositories (e.g. [PDB](#), [ECBD](#), [BioImage Archive](#), [EMPIAR](#)), others still rely primarily on the diligence of individual researchers. The challenge is therefore to accommodate this high heterogeneity. It requires first determining precisely the current state-of-the-art in data management across the RIs and networks, to reconcile differences and avoid duplication of the efforts already in place. The aim is to create an overarching core DMP that encompasses all existing processes and defines minimal procedures for those partners that do not currently have a fully developed DMP.

DEALING WITH SENSITIVE DATA

A further challenge is linked to the type of data that is generated within the ISIDORE project. Certain partners produce data with intrinsic biosecurity and/or biosafety issues. For others, there are concerns regarding the potential of identifying individuals from clinical or social science data. There is therefore a need to ensure that any approach to sensitive data FAIRification is legally and ethically compliant and ensures confidentiality and privacy.

The ability to conduct meta-analyses of data from cross-disciplinary studies is especially important in emergency situations, such as in a pandemic. A well-designed DMP that also ensures linking and accessing information on already existing data and relevant research findings (such as on pathogens of concern) will provide a roadmap not only on how to handle the data, but also on the establishment of processes to deal with a mix of sensitive and non-sensitive metadata and assess potential risks for data misuse.

ISIDORE PROJECT DMP BUILDING METHOD

A SURVEY TO DEFINE THE LANDSCAPE

To overcome the challenges described above, and conceive a robust DMP, ISIDORE's strategy was to build on existing components. The participating RIs and networks were therefore surveyed about their data management tools, and those of the researchers who are their end users. One important finding of this survey was that even for users of the more data-mature RIs, DMP tools

are often not integrated into their project workflows. Therefore, the ISIDORe DMP must indicate i) where researchers can find guidance for the stewardship and preservation of both data and software, ii) where and how DMP tools and content can be easily and efficiently integrated. For the latter, repository selection, licensing, formats and metadata recommendations, including use of identifiers for building linkages to external systems that provide data interoperability, as defined by the FAIR core principles, must be listed and updated during the project. Lastly, roles, responsibilities and the periodicity for updates must be clearly established.

ALIGNMENT WITH EXISTING RECOMMENDATIONS

ISIDORe implements the data policy description workflow that was created by the Digital Curation Centre (on behalf of the FAIRsFAIR project) and FAIRsharing (Sansone et al. 2019). It is designed to help with the creation of FAIR-aligned data policies DCC_FAIRsharing, adhering to FAIRsharing policy record metadata, the FAIRsFAIR FAIR Data Policy Checklist (Davidson et al. 2022) and RDA (Hrynaszkiewicz et al. 2020) specifications. In addition, all checklist fields and RDA-endorsed policy features within its scope are available within FAIRsharing policy records, making them accessible to both humans and machines. This creates a FAIR data policy ‘workflow’ from i) FAIR Data Policy Checklist to ii) deposition of the policy and assignment of a digital object identifier (DOI), to iii) submission of that policy into FAIRsharing and/or other appropriate repositories. Following this workflow will ensure the FAIRification of ISIDORe data management policies.

MAPPING REQUIREMENTS FOR THE DMP

The ISIDORe DMP will make use of pre-existing data repository concepts and integrate available data management tools and guidelines as much as possible (Figure 1), including RDMKit, the FAIR Cookbook (Rocca-Serra et al. 2023), the Data Stewardship Wizard (DSW) (Pergl et al. 2019), the ARGOS (Papadopoulou et al. 2020) DMP service, and the Infectious Diseases Toolkit (IDTk). Despite this relatively esoteric landscape, the DMP should be easy to understand by the different ISIDORe partners, so that they can apply its guidelines properly. The toolbox developed for discovery of digital objects related to sensitive data, harmonised across six life-science infrastructures and developed using an iterative process, will serve as an example of how to deal with issues arising across research infrastructures (David et al. 2022).

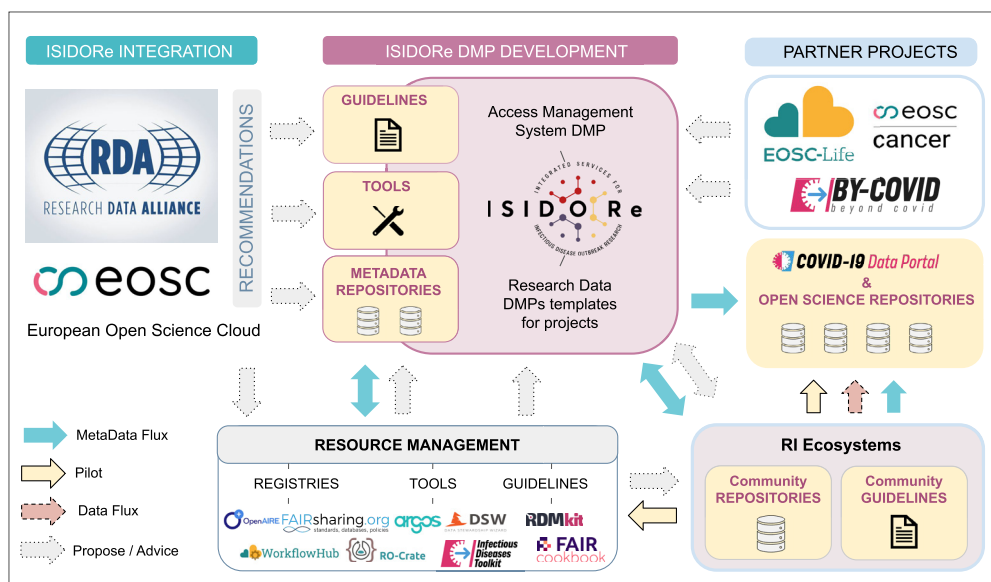


Figure 1 DMP building scheme for the ISIDORe project.

The ISIDORe DMP is being developed by integrating available DMP guidelines, resources and advice provided by participating RI communities, partner EU projects (such as those shown) and the overall recommendations by the RDA and EOSC. The established DMP procedures, FAIR guidelines, data repositories and metadata will be collected through platforms and tools (such as those illustrated). A live representation of the standards and databases used within the DMP will be available as a FAIRsharing collection. While many LS-RI partners already have DMP workflows in place, the goal and challenge of the ISIDORe project will be to set up cross-disciplinary pipelines, onboarding additional communities and adapting DMP procedures to meet the needs of infectious diseases research (data and tools) for the different disciplines in the consortium.

An investigation of data sharing statements (DSS), as included within the trial registry entries of COVID-19 related studies, revealed strong differences in how the request for details of data sharing has been interpreted, and, when data sharing is described as possible, a huge heterogeneity in the specification of the access procedures (Canham et al. 2022). This indicates that the instructions for the inclusion of DSSs within trial registry data need to be clearer, and where possible, statements should only use standardised categories to indicate the type of access that will be available.

To organise this homogenisation, ISIDORE MetaData mobilisation is based on an access management system (AMS). The main purpose of the AMS is to administer responses to calls, and requests for services, and to collect feedback from users in a standardised and centralised manner. This AMS, currently in development, will allow the management of the catalogue of services, project submission information with its associated metadata, and create a community-validated registry of projects (both those submitted and accepted). For certain RIs, any data generated by the provision of a service to a user is channelled into existing registries and repositories. This will be indexed in the ISIDORE AMS and syndicated through metadata tools and registries (described in the Supplementary Material). Metadata mobilisation will be achieved through existing RI AMSs when they exist (Figure 1); associated syndication processes are described in the ISIDORE DMP.

Since the project’s contributors are numerous and diverse, the DMP should be sufficiently flexible to fit these broad needs, whilst retaining minimal common requirements in standards, protocols and rules for metadata, traceability and provenance of data. This work on the ISIDORE DMP may introduce elements that could significantly impact the DMPs of individual project’s members, making them more general and potentially more ambitious.

FAIR-compliant DMPs covering service management and delivery have, up until now, generally been designed and implemented at the level of the RIs and networks. This type of DMP is distinct from the DMP covering the production of scientific data (Figure 2). For the latter, the onus to support users in DMP compliance has previously generally fallen on the individual service providers (e.g. national or regional core facilities). ISIDORE provides an opportunity to explore ways to harmonise user support. In this, as for all data challenges, ISIDORE is supported by BY-COVID. The BY-COVID project aims to make data and analysis tools related to infectious disease outbreaks more interoperable and accessible, not only to scientists but also to medical staff in hospitals, and policy makers, such as government officials, or indeed any other potential user.

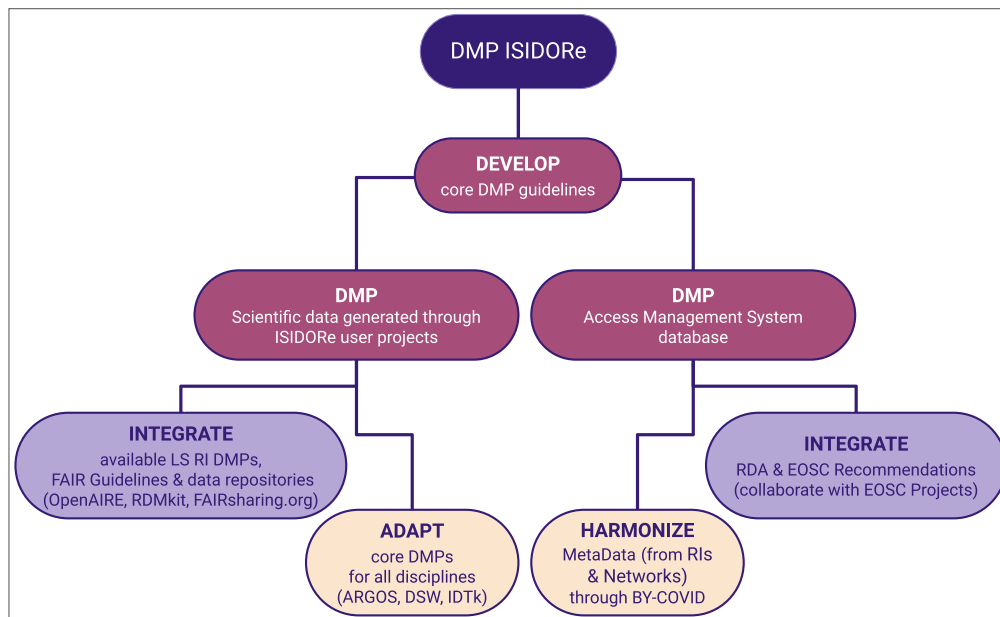


Figure 2 The two sides of ISIDORE’s DMP approach. ISIDORE requires a data management plan (DMP) covering both access management systems (AMSs) and scientific data generated through ISIDORE user projects. The scientific data part integrates available LS RI DMP components and adapts them for all disciplines. The AMS part harmonises management metadata from RIs and networks with the support of BY-COVID and in line with EOSC/RDA recommendations. Both approaches are applicable over a broad range of disciplines and must be compatible with the constraints and existing practices of ISIDORE’s diverse partners.

The two projects are working together: BY-COVID provides support, tools, data discovery and dissemination strategies, to enable data from user projects implemented by ISIDORE to be as FAIR as possible, while ISIDORE promotes the adoption of FAIR (meta)data standards by partners, via its network of data stewards, and tries to ensure that data is made available in the public domain, as widely as possible, preferably via dedicated sites such as the [COVID-19 Data Portal](#). As digital data needs to be actively managed over time to ensure continued availability and useability, depositing data resources with a trusted digital archive helps ensure that they are curated and handled according to good practices in digital preservation ([ICPSR_DMP](#)). Through the COVID-19 data portal, BY-COVID provides a platform for data discovery across a broad range of disciplines, with the flexibility to add any relevant data resource proposed by ISIDORE, through a lightweight metadata format jointly developed by both projects ([Hermjakob et al. 2022](#)). The collaboration between the two projects will lay the groundwork for Europe’s

future pandemic preparedness with FAIR data pipelines, enabling rapid response to outbreak situations, and a broad framework for answering research questions across research disciplines and country borders.

BY-COVID also provides standardised data management and analysis methods (partially based on the Galaxy Project), protocols, and training, to ensure FAIR is an integral part of the data pipeline. Building on existing infrastructure, data repositories are explorable through a single platform (COVID-19 data platform, and in the future, the EBI pathogen portal) allowing users to deposit, access and analyse pathogen data more easily. BY-COVID's data source catalogue of choice is FAIRsharing. It describes relationships among the data sources (databases, knowledge bases, repositories) and assists with the selection, documentation and visualisation of standards. The use of RO-crate (Eguinoa et al. 2022), based on FAIR Digital Object (FDO) principles, guarantees efficient data sharing, including data and metadata compliance using community-adopted standards such as bioschemas and common workflow language. Workflow repositories such as [WorkflowHub](#) that include the use of persistent and unique identifiers (e.g. DOIs [[cool URIs](#)]) ensure findability and discoverability of computational workflows and pipelines. Data discovery is further enabled by the BY-COVID data discovery network, based on the principles of the Beacon Network and employing an Application Programming Interface (API) that allows, for instance, discovery of genomic, phenotypic and clinical data.

BUILDING A BRIDGE BETWEEN PARTNERS' DATA

The implementation within ISIDORE of most user projects will generate experimental data. Ensuring that researchers adopt common DMP principles to deal with their data will require provision of advice at an early stage of their projects, and guidance on FAIR principles and the available tools and resources that fit their projects' data needs, as well as open science requirements. The current DMP strategy in Euro-BioImaging, for example, offers incoming users a consultation with the RI's FAIR Data steward before data is acquired. On the one hand, this raises awareness of the available guidelines, repositories and tools, and on the other hand, it brings data generation onto the DMP track. It also fosters timely adoption of the DMP with regards data provenance and life cycle management requirements, as well as readiness for data integration into open science frameworks after data generation and analysis are completed. It will be important for ISIDORE to establish cross-consortium exchanges between individual FAIR data stewards from the different RIs, to support the harmonisation of DMP principles and procedures.

Already, one solution to the problem of a comprehensive, broadly applicable yet specific DMP has been identified. The DMP can be divided into sections, including a general, overarching one, describing overall data management principles (so called 'core DMP'), and RI/field-specific sections. This may be achieved through several curated sets of questions (knowledge models) created in the Data Stewardship Wizard (DSW) and pre-filtered by the different ISIDORE partners, depending on their domain and how the data covered by a specific knowledge model is to be disseminated. The generation of RI-specific resources and repositories will greatly benefit from existing data-related or domain-specific collections (such as those collected through FAIRsharing, RDMKit and IDTK), and through the existing collaboration of RIs with data and workflow repositories. Indeed, in certain cases, these may have already paved the way for easy and fast submission of data.

Projects that require the provision of services by two or more RIs, and whose datasets must therefore comply with the requirements and recommendations of those RIs, present a special challenge. Again, the introduction of a two-tier DMP will help to overcome this, although the participating RIs will still need to reach a consensus on the recommendations to be given to users regarding their specific data constraints. One can imagine that this will also be facilitated by the variability of knowledge models within the DSW. Ideally, the RI or network (preferably via its data steward) should be automatically alerted of possible conflicting recommendations or overlapping issues, and thus ISIDORE will be able to provide users with access to coherent and curated catalogues of data, resources and repositories.

Building alignment between consortium partners on the core DMP principles and tools, as well as on the implementation of available resources to develop harmonised DMP workflows, can

be fostered by consolidating a consortium-wide data steward team/committee. Data steward representatives from consortium RIs and networks are to be invited to participate in shaping and curating ISIDORE's DMP, in collaboration with partner projects and initiatives such as those developed through [BY-COVID](#), [EOSC-Life](#) and [EOSC4Cancer](#), and taking into account the procedural details of their respective DMPs ([García Álvarez, Mayrhofer & Holub 2021](#); [Blomberg, Sarntivijai & Mayrhofer 2020](#)). This will allow the creation of sustainable exchange and cross-community validated solutions on the RI landscape.

SENSITIVE DATA WITHIN ISIDORE DMP

Based on the experience of the CORBEL project, with its multi-stakeholder task force, sensitive-data experts previously examined major issues associated with the sharing of sensitive data from individuals. They developed a consensus document on providing access to such individual participant data, using a broad interdisciplinary approach. This included 10 principles and 50 recommendations ([Ohmann et al. 2017](#)), representing the fundamental requirements of any framework used for the sharing of clinical research data. The document, to be included in ISIDORE's DMP, covers the following main areas: making data sharing a reality (e.g. cultural change, academic incentives, funding), consent for data sharing, protection of trial participants (e.g. de-identification), data standards, rights, types, and management of access (e.g. data request and access models), data management and repositories, discoverability, and metadata. Its adoption helps promote and support data sharing and reuse among researchers, adequately inform trial participants, protect their rights, and provides effective and efficient systems for preparing, storing and accessing data.

ACCESS MANAGEMENT SYSTEMS AS A GLUE

The main objective of the ISIDORE access management system (AMS) is to administer responses to calls for proposals and requests for free access to the project's services, as well as to collect feedback from users in a standardised and centralised manner. Some of the ISIDORE partners already have a pre-existing AMS, key to the facilitation of good data management in access provision within ISIDORE. Therefore, ISIDORE has conceived a central ISIDORE AMS system as a decoupled 'headless' component that will feed and be fed by the partners' pre-existing AMSs. For that purpose, minimal interoperability elements with the central ISIDORE AMS have been defined which the partners' AMSs will have to implement in order to provide the central system with up-to-date, formatted content regarding the details of the services proposed and the status of the requests for services that they process.

For partners that do not have a pre-existing AMS, or cannot ensure the transmission of the set of minimally-defined data, a new basic AMS system that matches the minimal interoperability criteria with the central AMS will be implemented. This could either be developed from scratch or be based on elements from one of the two pre-existing systems in use by partner RIs, [ARIA](#), and the [EVAg portal](#). [EU-OPENSREEN](#), [Euro-BioImaging](#), [MIRRI](#), [EMBRC](#) and [Instruct](#), use the [ARIA](#) AMS. They also use it for managing user requests submitted in response to ISIDORE's calls for proposals. [ARIA](#) can handle many aspects of RI operations, from proposal submission, and the evaluation process, to keeping a record of users' on-site visits and their feedback. Such a system enables all application and access provision data associated with a proposal to be collected and findable, and is compliant with the General Data Protection Regulation ([GDPR](#)). Additional developments are planned for [ARIA](#) to allow for the output of an executed project, such as publications and datasets, to also be findable via the AMS. This would be a powerful tool to allow data traceability. [EVA-GLOBAL](#)'s AMS is currently based on the Drupal open-source content management system (CMS) and also implements GDPR compliance. Similar to [ARIA](#), the [EVA](#) portal manages many aspects of RI operations and service provision. Currently used by the [EVA](#) and [Infravec](#) networks, it has the advantage of being user-friendly and has demonstrated great adaptability in the context of rapid responses to epidemic and pandemic situations (including the Zika epidemic, COVID-19 pandemic and most recently, the 2022 Mpox outbreak). This flexibility would be a great asset, especially considering ISIDORE's expected role in rapid research response.

EXPLOITING THE ISIDORE FRAMEWORK

The ISIDORE project is a unique opportunity for the 17 participating RIs and networks to work together to establish a new approach for service provision, to improve pandemic preparedness and accelerate research in response to an unexpected outbreak. The project needs to reconcile the partners' disparate philosophies, processes and practices, as well as ensuring that they all embrace common contemporary standards, while implementing best practices. The task is made more challenging, but technically more feasible, since ISIDORE has been designed to collaborate closely with its sibling data-focused project, BY-COVID. A necessary part is the constitution of a shared DMP. Researchers report that the benefits from a well-prepared DMP outweigh the initial investment needed to implement rigorous data management practices (Burnette, Williams & Imker 2016). Our overarching aim is thus to establish a working consensual umbrella DMP for this broad, multifaceted and complex community that will be useful for both the ISIDORE members, and the researchers who profit from the consortium's services.

THE ITERATIVE PROCESSING OF THE DMP

Our goal here was to describe the joint efforts of ISIDORE and BY-COVID towards the establishment of a complete and harmonised DMP, through a living document, allowing contributions from all stakeholders, prior to finalisation of a definitive and functional version.

As DMPs require upgrading in an iterative fashion by all partners involved in their creation, the use of a web-based DMP tool that allows users to contribute in parallel, manages versioning, and provides guidance in building the document, is paramount during the complete maturation cycle of the project's DMP. In principle ISIDORE will use ARGOS, rather than any of the other available tools (e.g. DSW, EOSC, DMPOnline).

HOW TO EXPLOIT PRE-EXISTING EFFORTS

The ISIDORE DMP has to deal with two facets of the project, the RI/network-centric access provision, and the service-provider level data generation. Thus, it cannot simply be a merging of institutional-level DMPs, but it needs to be designed through a dynamic agreement among all partners, based on the best current protocols to manage and exchange data in the respective fields. The partners with more advanced best practices act as drivers, laying the groundwork to define the content of the DMP, while the active contribution of all the other partners ensures that all needs and perspectives are appropriately considered and integrated in the set-up of a general proof of concept, with successive refinements and adaptations to the respective specificities and requirements.

Stakeholders involved need to develop equitable and sustainable financial models for data sharing, to ensure the long-term resourcing of data preparation and storage, as well as the request and sharing process. These tasks are not easily predictable, and thus not easily linked to initial funding. The discussion regarding sustainable business models for data infrastructures is ongoing, and it is difficult to identify a preferred model yet. A particular problem is that, while many established national and international data repositories have core streams of income from research funders, these sources of income are usually short-lived and may be vulnerable to change in priorities or in responsibilities (Ohmann et al. 2017).

DIVERSITY AND SPECIFICITIES INTEGRATION

To integrate FAIR, CARE and TRUST principles, ISIDORE necessarily adapted basic recommendations to this interdisciplinary, international project. It additionally needed to take into consideration the fact that the individual partners in the project could have conflicting priorities or may have had their own prior implementation of FAIR data sharing. This requires an inclusive harmonisation for data management practices, as a basis for future cross-disciplinary meta-analyses.

The adoption of the FAIR principles has encouraged researchers to comply with data and metadata standards. This need for 'FAIRification' has spurred several international efforts (cf. description in Supplementary Material), and members of both the ISIDORE and the BY-COVID

consortia are either directly involved in these efforts or engaged in the exploitation of these approaches in the context of their respective projects. One aim of these efforts is to facilitate interoperability between disparate data types. The interoperability principle in FAIRification can only come into play when the semantics of the content are well-defined across heterogeneous data sources (Holub et al. 2018). Ontologies are one of the semantic tools that are frequently used to support interoperability (Haque et al. 2022). The ISIDORE project, therefore, might benefit from the development of an ontology project as a guide to the elaboration of the DMP, especially for the section concerning the implementation of FAIR principles and open data (distinct from the OBO academy's domain ontologies). Such an ontology project would aim at clarifying the relationships between the entities (in the ontological sense) and metadata, as well as the selection of data standards and data types. Reusing project partners' existing metadata and data standards might simply be done by merging the information from each partner infrastructure in ISIDORE and agreeing on the most consistent one in case of equivalent candidates. A better approach would, however, be to evaluate for each data/metadata standard, if a 'linked data', 'linked open data' or even better, 'FAIR linked open (meta)data' standard exists that could be used (Frey & Hellmann 2021). Using such linked (meta)data standards would contribute to the semantic web and thus highly increase the potential findability of ISIDORE data, as well as its capacity to be interconnected. It would also give ISIDORE data additional meaning and boost the potential of its syntactic power in data mining processes. Wikidata entities and their associated metadata could be good FAIR-linked open metadata candidates for ISIDORE data, as they are an essential part of the current semantic web, validated and maintained by a large community. They could be expanded by the ISIDORE community with consistent novel inputs (Waagmeester et al. 2020). Thus, in an ideal world, by using linked data as metadata, we might enable ISIDORE data to be linked to the semantic web.

FURTHER STEPS AND PERSPECTIVES

The ISIDORE project has a mission to define and implement best practices in collaborative data management, with a fundamental objective to promote their adoption by the community. ISIDORE has the potential to lead the way with innovative aspects related to data controllers, data processors, and future proofing data sets (both data type and format) regarding storage, processing and transfer. Additional aspects to be considered will include data security, GDPR compliance, quality control, data anonymization, analysis and traceability of data, as well as communication and knowledge management. Indeed, the heart of the work will consist of data mobilisation planning, to convince all partners of the importance of replicating methods for which there is already an established implementation, or at least a proof of concept, all the while taking into account the human dimension (FAIR literacy [David et al. 2020] and the availability of qualified personnel for support, training and the assessment of results [Mons, B. 2020]). The quality criteria that are identified at each step will have to be constantly revisited. In the end, dissemination, support for DMP application, training and (self-)assessment must be FAIR and re-used, and could potentially feed into RDA recommendations. To conclude, while many aspects of the ideal DMP laid out here are achievable with current resources, an ultimate challenge for the ISIDORE community could be the implementation of machine actionable DMPs (Miksa, Walk & Neish 2020) that adhere to all DMP Common Standards (RDAMSC). But for that to happen, as for most research projects, FAIR DMP Literacy provided by a well organised network of data stewards must first be improved. This should become more feasible once the Competence Centres already active in some fields become more generalised, including in the planned EOSC cluster project roadmap.

DATA ACCESSIBILITY STATEMENTS

CC-BY licence – Data contact: RD. Survey & survey data will be made available after anonymisation with the ISIDORE DMP deliverable.

Preprint for Zenodo: <https://doi.org/10.5281/zenodo.7516084> https://docs.google.com/document/d/1AfLuhEFZl7suqWXbRvgyF_dSpmfhzrQ676mbTx-pLjI/edit?usp=sharing.

PREPRINT submitted: 02/06/2023 in Data Science Journal (DSJ) Special issue title: [Data Management Planning across Disciplines and Infrastructures](#).

The additional file for this article can be found as follows:

- **Supplementary material.** Data Management Tools and definitions. DOI: <https://doi.org/10.5334/dsj-2023-035.s1>

FUNDING INFORMATION

This work is mainly a product of ISIDORe (Integrated Services for Infectious Disease Outbreak Research) funding from the European Union's Horizon Europe research and innovation programme under grant agreement N°101046133. It is also a product of the BY-COVID Project funding from the European Union's Horizon Europe research and innovation programme under grant agreement N°101046203. Complementary support was provided through the EOSC-Life European program under grant agreement N°824087. Special acknowledgement to AM, DD, KE, MT, PG, SA for filling the survey without being an author of this paper.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHORSHIP CONTRIBUTIONS

Conceptualization: ARi, CC, JE, NB, RD

Original Draft Preparation: CC, JE, RD

Writing Paper: ARi, AK, ARy, CC, CG, CO, CW, IK, KL, JE, MLC, MH, MP, NB, PL, RD, RW, SRK, SS

Reviewing and editing paper: ARi, AK, ARy, CC, CG, CO, CW, IK, KL, JE, MLC, MH, MP, NB, PL, RD, RW, SRK, SS, RA, SAM, PA, DB, JMC, RC, FC, WF, CF, EGA, SG, HH, KH, PH, LK, AL, JLC, BM, KM, WMa, WMü, EP, CP, PR, SAS, GS

Figure conception: ARy, RD

Figure contributions: ARy, RB, SA, RD

Survey conception: ARi, CC, JE, RD

Survey completion (within the authors): GS, KH, KL, KM, PA, PL, RD, SR, WMa

AUTHOR AFFILIATIONS

Dr Romain David  orcid.org/0000-0003-4073-7456

European Research Infrastructure on Highly Pathogenic Agents (ERINHA), 98 rue du Trône B-1050 Bruxelles, BE

Dr Audrey S. Richard  orcid.org/0000-0002-0207-0139

European Research Infrastructure on Highly Pathogenic Agents (ERINHA), 98 rue du Trône B-1050 Bruxelles, BE

Ms Claire Connellan  orcid.org/0000-0001-6449-8512

European Research Infrastructure on Highly Pathogenic Agents (ERINHA), 98 rue du Trône B-1050 Bruxelles, BE

Dr Katharina B. Lauer  orcid.org/0000-0002-4347-7525

ELIXIR Hub, Wellcome Genome Campus, Hinxton, CB10 1SD, UK

Pr Maria Luisa Chiusano  orcid.org/0000-0002-6296-7132

Dept. of Agricultural Sciences, University of Naples Federico II, and Stazione Zoologica Anton Dorhn, Naples, IT

Pr Carole Goble  orcid.org/0000-0003-1219-2137

The University of Manchester, Oxford Road, Manchester, M13 9PL, UK

Dr Martin Houde  orcid.org/0000-0001-8630-204X

European Research Infrastructure on Highly Pathogenic Agents (ERINHA), 98 rue du Trône B-1050 Bruxelles, BE

Dr Isabel Kemmer  orcid.org/0000-0002-8799-4671

Euro-BioImaging ERIC Bio-Hub, European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, DE

Dr Antje Keppler  orcid.org/0000-0003-4358-2269

Euro-BioImaging ERIC Bio-Hub, European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, DE

Mr Philippe Lieutaud  orcid.org/0000-0002-5080-3456

European Virus Archive – GLOBAL (EVA-GLOBAL), unité des virus émergents – [UVE – UMR190] – Aix Marseille Université, IRD 190, INSERM U1207, FR

Pr Christian Ohmann  orcid.org/0000-0002-5919-1003

European Clinical Research Infrastructure Network (ECRIN), Kaiserswerther Str. 70, 40477 Düsseldorf, DE

Dr Maria Panagiotopoulou  orcid.org/0000-0002-4221-7254

European Clinical Research Infrastructure Network (ECRIN), 5-7 rue Watt 75013 Paris, FR

Dr Sara Raza Khan  orcid.org/0000-0003-1587-6171

European Clinical Research Infrastructure Network (ECRIN), 5-7 rue Watt 75013 Paris, FR

Dr Arina Rybina  orcid.org/0000-0002-5609-5710

Euro-BioImaging ERIC Bio-Hub, European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, DE

Dr Stian Soiland-Reyes  orcid.org/0000-0001-9842-9718

Department of Computer Science, The University of Manchester, Manchester, M13 9PL, UK & Informatics Institute, University of Amsterdam, LAB42, Science Park 900, 1098 XH Amsterdam, NL

Dr Charlotte Wit  orcid.org/0000-0002-5461-8354

EU-OPENSREEN, Robert-Rössle-Straße 10, Building 87, 13125 Berlin, DE

Dr Rudolf Wittner  orcid.org/0000-0002-0003-2024

Biobanking and BioMolecular Resources Research Infrastructure (BBMRI-ERIC), Neue Stiftingtalstrasse 2, 8010 Graz, AT

Dr Rafael Andrade Bueno  orcid.org/0000-0002-6675-3836

VIB-UGent Center for Plant Systems Biology, Ghent, BE

Dr. Sarah Arnaud Marsh  orcid.org/0000-0003-4194-043X

Institut Pasteur (IP), Genetics and Genomics of Insect Vectors Unit (Infravec) & Department of Anthropologie and Ecology of Disease Unit (Sonar-Global), 25-28 rue du Docteur Roux, 75724 Paris Cedex 15, FR

Dr Pauline Audergon  orcid.org/0000-0002-1624-7388

Instruct-ERIC, Oxford House, Parkway Court, John Smith Drive, Oxford, OX4 2JY, UK

Mr Dylan Bonfils  orcid.org/0000-0002-2734-3420

European Research Infrastructure on Highly Pathogenic Agents (ERINHA), 98 rue du Trône B-1050 Bruxelles, BE

Pr Jose-Maria Carazo  orcid.org/0000-0003-0788-8447

CNB-CSIC, Instruct-ES, Campus Universidad Autónoma, 28049, Madrid, ES

Pr Remi Charrel  orcid.org/0000-0002-7675-8251

European Virus Archive – GLOBAL (EVA-GLOBAL), unité des virus émergents – [UVE – UMR 190] – Aix Marseille Université, IRD 190, INSERM U1207, Marseille, FR

Dr Frederik Coppens  orcid.org/0000-0001-6565-5145

VIB-UGent Center for Plant Systems Biology, Ghent, BE

Dr Wolfgang Fecke  orcid.org/0000-0003-1917-0037

EU-OPENSREEN, Robert-Rössle-Straße 10, Building 87, 13125 Berlin, DE

Dr Claudia Filippone  orcid.org/0000-0001-5483-7656

European Research Infrastructure on Highly Pathogenic Agents (ERINHA), 98 rue du Trône B-1050 Bruxelles, BE

Dr Eva Garcia Alvarez  orcid.org/0000-0002-3522-5088

BBMRI-ERIC, Biobanking and BioMolecular Resources Research Infrastructure (BBMRI-ERIC), Neue Stiftingtalstrasse 2, 8010 Graz, AT

Dr Sheraz Gul  orcid.org/0000-0003-2543-1643

Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Schnackenburgallee 114, 22525 Hamburg, Germany Fraunhofer; Cluster of Excellence for Immune-Mediated Diseases CIMD, Schnackenburgallee 114, 22525 Hamburg, DE

Mr Henning Hermjakob  orcid.org/0000-0001-8479-0262

European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, UK

Dr Katja Herzog  orcid.org/0000-0002-1389-8118

EU-OPENSREEN, Robert-Rössle-Straße 10, Building 87, 13125 Berlin, DE

Dr Petr Holub  orcid.org/0000-0002-5358-616X

BBMRI-ERIC, Biobanking and BioMolecular Resources Research Infrastructure (BBMRI-ERIC), Neue Stiftingtalstrasse 2, 8010 Graz, AT

Dr Lukasz Kozera  orcid.org/0000-0002-9015-4518

BBMRI-ERIC, Biobanking and BioMolecular Resources Research Infrastructure (BBMRI-ERIC), Neue Stiftingtalstrasse 2, 8010 Graz, AT

- Dr. Allyson L. Lister**  orcid.org/0000-0002-7702-4495
Data Readiness Group, Oxford e-Research Centre, University of Oxford, UK
- Dr. José López-Coronado**  orcid.org/0000-0001-7442-8091
CECT-University of Valencia, Agustín Escardino 9,46980 Paterna (Valencia), ES
- Dr Bénédicte Madon**  orcid.org/0000-0001-8608-3895
La Rochelle Université, Littoral, Environnement et Sociétés (LIENSs), UMR 7266 CNRS – La Rochelle University, 2 rue Olympe de Gouges, 17000 La Rochelle, FR
- Mr Kurt Majcen**  orcid.org/0000-0002-4041-4887
BBMRI-ERIC, Biobanking and BioMolecular Resources Research Infrastructure (BBMRI-ERIC), Neue Stiftingtalstrasse 2, 8010 Graz, AT
- Dr William Martin**  orcid.org/0000-0003-4839-1752
European Vaccine Initiative, Vossstrasse 2, Geb. 4040, 69115 Heidelberg, DE
- Dr Wolfgang Müller**  orcid.org/0000-0002-4980-3512
Heidelberg Institute for Theoretical Studies (HITS), Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, DE
- Dr Elli Papadopoulou**  orcid.org/0000-0002-0893-8509
ATHENA Research Center/OpenAIRE, Artemidos 6 & Epidavrou, 15125, Athens, GR
- Dr Christine M.A. Prat**  orcid.org/0000-0002-6963-8673
European Virus Archive – GLOBAL (EVA-GLOBAL), unité des virus émergents – [UVE – UMR190] – Aix Marseille Université, IRD 190, INSERM U1207, Marseille, FR
- Dr Paolo Romano**  orcid.org/0000-0003-4694-3883
IRCCS Ospedale Policlinico San Martino, Largo Rosanna Benzi 10, 16132 Genova, IT
- Prof. Susanna-Assunta Sansone**  orcid.org/0000-0001-5306-5690
Data Readiness Group, Oxford e-Research Centre, University of Oxford, UK
- Dr Gary Saunders**  orcid.org/0000-0002-7468-0008
EATRIS-ERIC, De Boelelaan 1118, 1081 HZ Amsterdam, NL
- Dr Niklas Blomberg**  orcid.org/0000-0003-4155-5910
ELIXIR Hub, Wellcome Genome Campus, Hinxton, CB10 1SD, UK
- Dr Jonathan Ewbank**  orcid.org/0000-0002-1257-6862
European Research Infrastructure on Highly Pathogenic Agents (ERINHA), 98 rue du Trône B-1050 Bruxelles, BE

REFERENCES

- Bishop, B, Gunderman, H, Davis, R,** et al. 2020. Data curation profiling to assess data management training needs and practices to inform a toolkit. *Data Science Journal*, 19(1): 4. DOI: <https://doi.org/10.5334/dsj-2020-004>
- Blomberg, N, Sarntivijai, S** and **Mayrhofer, MTh.** 2020. EOSC-Life data management plan. *Zenodo*. DOI: <https://doi.org/10.5281/zenodo.4010442>
- Burnette, MH, Williams, SC** and **Imker, HJ.** 2016. From plan to action: successful data management plan implementation in a multidisciplinary project. *Journal of eScience Librarianship*, 5(1): 6. DOI: <https://doi.org/10.7191/jeslib.2016.1101>
- Canham, S, Felder, G, Ohmann, C** and **Panagiotopoulou, M.** 2022. Identification of COVID-19 clinical studies intending to share individual participant data for secondary use: Protocol for a pilot study. *Zenodo*. DOI: <https://doi.org/10.5281/zenodo.7064624>
- Carroll, SR, Garba, I, Figueroa-Rodríguez, OL,** et al. 2020. The CARE principles for indigenous data governance. *Data Science Journal*, 19: 43. DOI: <https://doi.org/10.5334/dsj-2020-043>
- David, R, Mabile, L, Specht, A,** et al., **RDA – SHaring Reward and Credit (SHARC) Interest Group R.D.A.** 2020. FAIRness literacy: The Achilles' heel of applying FAIR principles. *Data Science Journal*, 19(1): 32. DOI: <http://doi.org/10.5334/dsj-2020-032>
- David, R, Ohmann, C, Boiten, JW,** et al. 2022. An iterative and interdisciplinary categorisation process towards FAIRer digital resources for sensitive life-sciences' data. *Sci Rep*, 12: 20989. DOI: <https://doi.org/10.1038/s41598-022-25278-z>
- Davidson, J, Grootveld, M, Verburg, M,** et al. 2022. FAIR-enabling Data Policy Checklist. *Zenodo*. DOI: <https://doi.org/10.5281/zenodo.6225775>
- Eguinoa, I, Suchánek, M, Knaisl, V,** et al. 2022. BioHackEU22 Report: Enhancing research data management in Galaxy and Data Stewardship Wizard by utilising RO-Crates. *BiorXiv (in prep)*. DOI: <https://doi.org/10.37044/osf.io/24jst>
- Frey, J** and **Hellmann, S.** 2021. FAIR linked data – towards a linked data backbone for users and machines. In *Companion Proceedings of the Web Conference 431–435 (Association for Computing Machinery, 2021)*. DOI: <https://doi.org/10.1145/3442442.3451364>
- García Álvarez, E, Mayrhofer, M** and **Holub, P.** 2021. BY-COVID – D8.2 – Data Management Plan. *Zenodo*. DOI: <https://doi.org/10.5281/zenodo.6884816>

- Giraldo, O, Alves, R, Bampalikis, D**, et al. 2022. A FAIRification roadmap for ELIXIR software management plans. 1st International Conference on FAIR Digital Objects (FDO 2022). *Research Ideas and Outcomes*, 8: e94608. DOI: <https://doi.org/10.3897/rio.8.e94608>
- Goble, C, Soiland-Reyes, S, Bacall, F**, et al. 2021. Implementing FAIR digital objects in the EOSC-Life workflow collaboratory. *Zenodo*. DOI: <https://doi.org/10.5281/zenodo.4605654>
- Goble, C, Bacall, F, Soiland-Reyes, S**, et al. 2022. WorkflowHub – a FAIR registry for workflows. *F1000Research*, 11. DOI: <https://doi.org/10.7490/f1000research.1118984.1>
- Haque, AKMB, Arifuzzaman, BM, Siddik, SAN**, et al. 2022. Semantic web in healthcare: A systematic literature review of application, research gap, and future research avenues. *Int J Clin Pract*. 2022: 6807484. DOI: <https://doi.org/10.1155/2022/6807484>
- Hermjakob, H, Kleemola, M, Moilanen, K**, et al. 2022. BY-COVID – D3.1 – Metadata standards. Documentation on metadata standards for inclusion of resources in data portal. *Zenodo*. DOI: <https://www.doi.org/10.5281/zenodo.6885016>
- Holub, P, Kohlmayer, F, Prasser, F**, et al. 2018. Enhancing reuse of data and biological material in medical research: From FAIR to FAIR-Health. *Biopreserv Biobank*, 16: 97–105. DOI: <https://doi.org/10.1089/bio.2017.0110>
- Hrynaszkiewicz, I, Simons, N, Hussain, A**, et al. 2020. Developing a research data policy framework for all journals and publishers. *Data Science Journal*, 19: 5. DOI: <https://doi.org/10.5334/dsj-2020-005>
- Hutson, M**. 2022. Taking the pain out of data sharing. *Nature*, 610: 220–221. DOI: <https://doi.org/10.1038/d41586-022-03133-5>
- Jacobs, JA and Humphrey, C**. 2004. Preserving research data. *Commun. ACM*, 47(9): 27–29. DOI: <https://doi.org/10.1145/1015864.1015881>
- Kesisoglou, I, Cosgrove, S, Derycke, P**, et al. 2022. Glossary of commonly used terms in the field of health data research – developed by the EU project HealthyCloud. *Zenodo*. DOI: <https://doi.org/10.5281/zenodo.6787119>
- Lin, D, Crabtree, J, Dillo, I**, et al. 2020. The TRUST Principles for digital repositories. *Sci Data*, 7: 144. DOI: <https://doi.org/10.1038/s41597-020-0486-7>
- Michener, WK**. 2015. Ten simple rules for creating a good data management plan. *PLoS Comput Biol*, 11: e1004525. DOI: <https://doi.org/10.1371/journal.pcbi.1004525>
- Miksa, T, Jaoua, M and Arfaoui, G**. 2020. Research object crates and machine-actionable data management plans. *First Workshop on Data and Research Objects Management for Linked Open Science (DaMaLOS) at The 19th International Semantic Web Conference (ISWC 2020)*. DOI: <https://doi.org/10.4126/frl01-006423291>
- Miksa, T, Walk, P and Neish, P**. 2020. RDA DMP common standard for machine-actionable data management plans. RDA recommendation. DOI: <https://doi.org/10.15497/rda00039>
- Mons, B**. 2020. Invest 5% of research funds in ensuring data are reusable. *Nature*, 578: 491–491. DOI: <https://doi.org/10.1038/d41586-020-00505-7>
- Ohmann, C, Banzi, R, Canham, S**, et al. 2017. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open*, 7: e018647. DOI: <https://doi.org/10.1136/bmjopen-2017-018647>
- Pergl, R, Hooft, R, Suchánek, M**, et al. 2019. Data Stewardship Wizard: A tool bringing together researchers, data stewards, and data experts around data management planning. *Data Science Journal*, 18(1): 59. DOI: <http://doi.org/10.5334/dsj-2019-059>
- Papadopoulou, E, Bardi, A and Kakalettris, G**. 2020. Data management plans and linked open data: Exploiting machine actionable data management plans through open science graphs. *First Workshop on Data and Research Objects Management for Linked Open Science (DaMaLOS) at The 19th International Semantic Web Conference (ISWC 2020)*. DOI: <https://repository.publisso.de/resource/frl:6423283>
- Richard, A, Stepanyan, D and Ewbank, J**. 2022. Pandemic preparedness – Europe launches research consortium. *Nature*, 603(7900): 228. PMID: 35260850. DOI: <https://doi.org/10.1038/d41586-022-00669-4>
- Rocca-Serra, P, Gu, W, Ioannidis, V**, et al. 2023. The FAIR Cookbook – The essential resource for and by FAIR doers. *Sci Data*, 10: 292. DOI: <https://doi.org/10.1038/s41597-023-02166-3>
- Sansone, S-A, McQuilton, P, Rocca-Serra, P**, et al. 2019. FAIRsharing as a community approach to standards, repositories and policies. *Nat Biotechnol*, 37: 358–367. DOI: <https://doi.org/10.1038/s41587-019-0080-8>
- Smale, N, Unsworth, K, Denyer, G**, et al. 2020. A review of the history, advocacy and efficacy of data management plans. *International Journal of Digital Curation*, 15(1): 1–29. DOI: <https://doi.org/10.2218/ijdc.v15i1.525>
- Stall, S, Specht, A, Corrêa, PLP**, et al. 2020. PARSEC data and digital output management plan and workbook. *Zenodo*. DOI: <https://doi.org/10.5281/zenodo.3891426>
- Soiland-Reyes, S, Sefton, P, Crosas, M**, et al. 2022a. Packaging research artefacts with RO-Crate. *Data Science*, 5(2). DOI: <https://doi.org/10.3233/ds-210053>

- Soiland-Reyes, S, Sefton, P, Jael Castro, L**, et al. 2022b. Creating lightweight FAIR digital objects with RO-Crate. 1st International Conference on FAIR Digital Objects (FDO 2022). *Research Ideas and Outcomes* 8: e93937. DOI: <https://doi.org/10.3897/rio.8.e93937>
- Vines, TH, Albert, AYK, Andrew, RL**, et al. 2014. The availability of research data declines rapidly with article age. *Current Biology*, 24: 94–97. DOI: <https://doi.org/10.1016/j.cub.2013.11.014>
- Waagmeester, A, Stupp, G, Burgstaller-Muehlbacher, S**, et al. 2020. Wikidata as a knowledge graph for the life sciences. *eLife*, Mar 17; 9: e52614. DOI: <https://doi.org/10.7554/eLife.52614>
- Wilkinson, MD, Dumontier, M, Aalbersberg, IJJ**, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3: 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>
- Wittner, R, Holub, P, Müller, H**, et al. 2021. ISO 23494: Biotechnology – Provenance Information Model for Biological Specimen And Data' In: Glavic, B., Braganholo, V., Koop, D. (eds) Provenance and Annotation of Data and Processes. IPAW IPAW 2020 2021. Lecture Notes in *Computer Science*, vol 12839. Springer International Publishing, Cham, 222–225. DOI: https://doi.org/10.1007/978-3-030-80960-7_16
- Wittner, R, Mascia, C, Gallo, M**, et al. 2022. Lightweight distributed provenance model for complex real-world environments. *Sci Data*, 9: 503. DOI: <https://doi.org/10.1038/s41597-022-01537-6>

WEB REFERENCES

- ANR** 2019. Available at: <https://anr.fr/en/latest-news/read/news/the-anr-introduces-a-data-management-plan-for-projects-funded-in-2019-onwards/> [Last accessed 31 May 2023]
- ARGOS** n.d. Available at: <https://argos.openaire.eu/> [Last accessed 31 May 2023]
- ARIA** n.d. Available at: <https://instruct-eric.org/help/about-aria> [Last accessed 31 May 2023]
- BioImage Archive** n.d. Available at: <https://www.ebi.ac.uk/bioimage-archive/> [Last accessed 31 May 2023]
- BY-COVID** n.d. Available at: <https://by-covid.org/> [Last accessed 31 May 2023]
- Canserv** n.d. Available at: <https://www.canserv.eu/> [Last accessed 31 May 2023]
- CoData** n.d. Available at: <https://codata.org/> [Last accessed 31 May 2023]
- Cool URIs** n.d. Available at: <https://www.w3.org/Provider/Style/URI.html.en> [Last accessed 31 May 2023]
- CORBEL** n.d. Available at: <https://www.corbel-project.eu/> [Last accessed 31 May 2023]
- COVID-19 Data Portal** n.d. Available at: <https://www.covid19dataportal.org/> [Last accessed 31 May 2023]
- COVID-19 Workflows** n.d. Available at: <https://covid19.workflowhub.eu/> [Last accessed 31 May 2023]
- Data Policy** n.d. Available at: <https://stats.oecd.org/glossary/detail.asp?ID=4454> [Last accessed 31 May 2023]
- DCC_FAIRsharing** n.d. Available at: <https://dcc.ac.uk/blog/fairsharing-and-dcc-collaborate-align-policy-metadata> [Last accessed 31 May 2023]
- DMPOnline** n.d. Available at: <https://marketplace.eosc-portal.eu/services/dmponline> [Last accessed 31 May 2023]
- DSW** n.d. Available at: <https://ds-wizard.org/> [Last accessed 31 May 2023]
- ECBD** n.d. Available at: <https://ecbd.eu/> [Last accessed 31 May 2023]
- ELIXIR** n.d. Available at: <https://elixir-europe.org/> [Last accessed 31 May 2023]
- EMPIAR** n.d. Available at: <https://www.ebi.ac.uk/empiar/> [Last accessed 31 May 2023]
- EOSC** n.d. Available at: <https://eosc.eu/> [Last accessed 31 May 2023]
- EOSC4Cancer** n.d. Available at: <https://eosc4cancer.eu/> [Last accessed 31 May 2023]
- EOSC-Life** n.d. Available at: <https://www.eosc-life.eu/> [Last accessed 31 May 2023]
- ERINHA** n.d. Available at: <https://www.erinha.eu/> [Last accessed 31 May 2023]
- EU-OPENSREEN** n.d. Available at: <https://www.eu-openscreen.eu/> [Last accessed 31 May 2023]
- Euro-BioImaging** n.d. Available at: <https://www.eurobioimaging.eu/> [Last accessed 31 May 2023]
- EVAg** n.d. Available at: <https://www.european-virus-archive.com/access-evag-ressources> [Last accessed 31 May 2023]
- FAIR Cookbook** n.d. Available at: <https://faircookbook.elixir-europe.org> [Last accessed 31 May 2023]
- FAIRsharing** n.d. Available at: <https://fairsharing.org/> [Last accessed 31 May 2023]
- FAIRsharing Community Curation** n.d. Available at: https://fairsharing.org/community_curation [Last accessed 31 May 2023]
- FAIRsharing EOSC-Life** n.d. Available at: <https://fairsharing.org/EOSCLife> [Last accessed 31 May 2023]
- FAIRsharing Policy Record Metadata** n.d. Available at: <https://fairsharing.gitbook.io/fairsharing/additional-information/policy-content-and-scope> [Last accessed 31 May 2023]
- GDPR** n.d. Available at: <https://gdpr-info.eu/> [Last accessed 31 May 2023]
- Go-FAIR** n.d. Available at: <https://www.go-fair.org/> [Last accessed 31 May 2023]
- HE DMP** n.d. Available at: <https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf> [Last accessed 31 May 2023]
- ICPSR_DMP** n.d. Available at: <https://www.icpsr.umich.edu/web/pages/datamanagement/dmp/framework.html> [Last accessed 31 May 2023]

IDTk n.d. Available at: <https://www.infectious-diseases-toolkit.org/> [Last accessed 31 May 2023]
INSTRUCT n.d. Available at: <https://instruct-eric.org/> [Last accessed 31 May 2023]
ISIDORE n.d. Available at: <https://isidore-project.eu/> [Last accessed 31 May 2023]
ISO 20691 n.d. Available at: <https://fairsharing.org/ISO20691> [Last accessed 31 May 2023]
IVOA n.d. Available at: <https://fairsharing.org/IVOA> [Last accessed 31 May 2023]
Interoperability Resource within ELIXIR n.d. Available at: <https://elixir-europe.org/platforms/interoperability/rirs> [Last accessed 31 May 2023]
NIH n.d. Available at: <https://www.nia.nih.gov/research/blog/2022/09/data-management-and-sharing-nih-policy-details-and-guidance> [Last accessed 31 May 2023]
OBO Academy's n.d. Available at: <https://oboacademy.github.io/obook/tutorial/project-ontology-development/> [Last accessed 31 May 2023]
OSTP n.d. Available at: <https://www.whitehouse.gov/ostp/news-updates/2022/08/25/ostp-issues-guidance-to-make-federally-funded-research-freely-available-without-delay/> [Last accessed 31 May 2023]
PDB n.d. Available at: <https://www.rcsb.org/> [Last accessed 31 May 2023]
RDA n.d. Available at: <https://www.rd-alliance.org/> [Last accessed 31 May 2023]
RDA FAIRsharing WG n.d. Available at: <https://www.rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases.html> [Last accessed 31 May 2023]
RDAMSC n.d. Available at: <https://rdamsc.bath.ac.uk/> [Last accessed 31 May 2023]
RDMKit n.d. Available at: <https://rdmkit.elixir-europe.org/> [Last accessed 31 May 2023]
RI-VIS n.d. Available at: <https://ri-vis.eu/> [Last accessed 31 May 2023]
RO-Crate n.d. Available at: <https://w3id.org/ro/crate> [Last accessed 31 May 2023]
WDS n.d. Available at: <https://worlddatasystem.org/> [Last accessed 31 May 2023]
WorkflowHub n.d. Available at: <https://workflowhub.eu/> [Last accessed 31 May 2023]
Zenodo n.d. Available at: <https://zenodo.org/> [Last accessed 31 May 2023]

David et al.
Data Science Journal
DOI: 10.5334/dsj-2023-035

15

TO CITE THIS ARTICLE:

David, R, Richard, AS, Connellan, C, Lauer, KB, Chiusano, ML, Goble, C, Houde, M, Kemmer, I, Keppler, A, Lieutaud, P, Ohmann, C, Panagiotopoulou, M, Khan, SR, Rybina, A, Soiland-Reyes, S, Wit, C, Wittner, R, Buono, RA, Marsh, SA, Audergon, P, Bonfils, D, Carazo, J-M, Charrel, R, Coppens, F, Fecke, W, Filippone, C, Alvarez, EG, Gul, S, Hermjakob, H, Herzog, K, Holub, P, Kozera, L, Lister, AL, López-Coronado, J, Madon, B, Majcen, K, Martin, W, Müller, W, Papadopoulou, E, Prat, CMA, Romano, P, Sansone, S-A, Saunders, G, Blomberg, N and Ewbank, J. 2023. Umbrella Data Management Plans to Integrate FAIR Data : Lessons From the ISIDORE and BY-COVID Consortia for Pandemic Preparedness. *Data Science Journal*, 22: 35, pp. 1–15. DOI: <https://doi.org/10.5334/dsj-2023-035>

Submitted: 15 December 2022

Accepted: 22 June 2023

Published: 18 September 2023

COPYRIGHT:

© 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.