Horizon Europe

Project: 101079789

D3.1 – Data policy and Data Management Plan of EIRENE

WP 3 – Data and Ethics

WP Leader:	CNR
Date:	September 2024
Nature:	DEC
Dissemination level:	Public

Grant Agreement Number	101079789	Acronym	EIRENE PPP
Full title	Environmental Exposure Assessment Research Infrastructure Preparatory Phase Project		
Project URL	https://www.eirene.eu/		
Project Officer	Andreas Holtel, Andreas.HOLTEL@ec.europa.eu		

Delivery date	Contractual	30.09.2024	Actual	04.09.2024
Status	Final			
Nature	DEC			
Dissemination level	Public			

Responsible Partner	CNR			
Responsible Author	Nicola Pirrone		E-mail	nicola.pirrone@cnr.it
	Partner	CNR	Phone	+39.0984.493239
Other partners	VITO, RECETOX			
Co-Authors	Dirk Devriendt, Sylvie Remy, Jos Bessems, Shachar D'vir, Sergio Cinnirella, Ian M. Hedgecock, Francesco De Simone			

Document History

Institution	Date	Version
CNR	2024/09/04	v. final
RECETOX	2024/08/13	v.3 (revision of v.2)
VITO	2024/07/25	v.2 (revisio of v.1)
CNR	2024/04/30	v.1 (draft)

Table of Contents

1. Foreword

EIRENE responds to the need to facilitate research into human health and well-being at all life stages, accounting for environmental, sociological, and economic drivers. This complex interdisciplinary research environment, involving widely varying institutions, will require an extremely well-curated approach to data acquisition, documentation, storage, and dissemination.

Within the implementation stage of EIRENE (from 2026 -2030) research fields will be covered which may or may not have overlapping data types, data processing, and data access needs, requirements, or regulations. EIRENE aims to enable collaborations across the de facto data boundaries that exist between the research communities involved in exposome research.

The Data Management Plan (DMP) is a 'living document' providing guidelines to help harmonise and maximise the accessibility and usability of data throughout the EIRENE life cycle. EIRENE data is derived from laboratory-based experiments, field measurement campaigns, routine monitoring, satellite observations, and biomedical research data. Clearly, a 'one fits all' management plan to handle such highly variable data types does not exist. For example, biomedical and health data brings with it many regulations regarding data access, data processing, and necessary data anonymisation or pseudonymisation[1] processes, while such requirements are unnecessary for measurements of atmospheric particulate composition. This Data Management Plan (DMP) lays down the guidelines for the future activities of EIRENE, to foster the 'as open as possible, as closed as necessary' in the FAIR principles of accessibility, to enhance reusability while ensuring conformation with controller and processing agreements, as well as national and international data protection regulations.

Since much of the data in EIRENE will be historical, it will be necessary to harmonise it as much as possible, leveraging the FAIR principles.

2. Introduction

The Environmental Exposure Assessment Research Infrastructure Preparatory Phase Project (EIRENE PPP) aims to prepare the implementation of the distributed Research Infrastructure for "EnvIRonmental Exposure assessmeNt in Europe" (EIRENE RI) by enabling the development of advanced technologies and complementary services on the characterisation of complex environmental exposures and their impact on the European population.

Due to the peculiarity of a real multidisciplinary approach and the ambition to provide unique web services for research studies and policy decisions, data management is a very critical

[1] According to *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data* (GDPR) 'pseudonymisation' means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person; while 'anonymization' is defined in an ISO standard (ISO 29100:2011) as the "process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party".

aspect to be addressed for the forthcoming infrastructure. Both data generated in the broader research community and collected for exposome studies will undergo the typical data archiving, sharing, preservation and use processes that require decisions on interoperability and metadata standards.

It is imperative that EIRENE RI data activities comply with the EU's open science policy, which states that "FAIR (Findable, Accessible, Interoperable, and Re-usable data) and (as) open (as possible) data sharing should become the default for the results of EU-funded scientific research." In light of the objectives and methodology of the EIRENE RI, the logic behind this is inescapable.

The first version of the Data Management Plan was developed during the Design Phase of EIRENE (EIRENE RI Design Study - Final Report, Annex V). The current version describes possible data sources and data management aspects, including collection, organization, archiving, preservation, sharing, interoperability, and metadata standards. This DMP outlines the lifecycle of research data and knowledge generated within the EIRENE framework, in line with the European Open Science ambitions and in compliance with the European Open Science Cloud (EOSC) aims.

It is important to keep in mind that EIRENE is currently in the Preparatory phase and funded by the EIRENE PPP Coordination and Support project. It does not offer at present any services nor collect or process data. While several pilots are planned during the project, they will only be used to simulate procedures and the user portal.

3. DMP in EIRENE RI

In the first stage, data providers/collators will be expected to complete the DMP template. In a first stage, the Horizon Europe template has been adopted, although during the EIRENE RI lifetime this will most probably be updated to satisfy user requirements and reflect changes in legislation, technical protocols and emerging technologies. In the case where project participants make use of publicly available datasets, such as air quality monitoring data, satellite retrievals of NO_2, SO_2 or Aerosol Optical Depth for example, the appropriate metadata should be included, and the data template below made available.

The DMP template is available online (https://ec.europa.eu › docs › temp-form › report), and there are tools such as DMP online (https://dmponline.dcc.ac.uk/) provided by the Digital Curation Centre (DCC), which can be used collaboratively. The Horizon Europe (HEU) template is integrated in DMP online.

The HEU DMP template is straightforward, requiring some information on the derivation (project, campaign, funding, etc) of the data and then the data details: the template questions to be answered are shown in green in the next paragraphs.

In its final form, EIRENE will cater to different exposure research communities, which will have both established data conventions and varied requirements with regard to the type of metadata that is useful to them (e.g. domain specific search vocabularies), the information that should be part of a DMP (e.g. licensing info, data usage restrictions, governance conditions or provenance concerns) and how this should be integrated with or can be instrumentalized by the infrastructure. At the same time, making data available in a FAIR way assumes the availability of FAIR metadata, which in its turn includes the information that composes the DMP, implying that also the DMP has to be a structured, FAIR resource, presented in an interoperable machine-readable format.

Taking these aspects together, EIRENE will accommodate data resources (datasets, metadata records, algorithms, etc) from multiple research domains, with varying interoperability

characteristics and at differing levels of FAIR maturity, while also enabling and even promoting FAIR reuse by combining both high level common standards with domain specific metadata schemas and DMP templates.

Both the desired content of the DMPs and their integration with the enabling infrastructure will evolve with the developing EIRENE infrastructure architecture, which is another reason this DMP document is considered a "living document".

4. Data Summary

In the following sections text reported in italic, green is derived from the HEU DMP template and tracks the preparation of the EIRENE DMP.

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

This is likely to be the most common case in the pilot operations and pilot projects. Re-use of biomedical data will require special attention as described in the ELSI guidelines (D3.2 – ELSI guidelines) and further addressed by the Ethics Advisory Board that will be established by the Consortium. The EIRENE RI Mission is to establish a sustainable research infrastructure enabling the advancement of exposome research in Europe by bringing together complementary capacities available in the member states, harmonising them and upgrading to address current scientific and societal challenges in the areas of chemical exposures and population health (*from the EIRENE RI Design Study - Final Report*).

The EIRENE RI will be founded on newly developed computational, data interpretation and modelling tools provided through the Knowledge Platform and specific Virtual Labs, operated at multiple research clouds.

To make an effective and holistic analysis possible of data on exposure, susceptibility and effect markers together with data on environmental and social stressors, food basket and consumer product contamination, high quality data managed following interoperable standards are required. Such data can then be processed with bioinformatics tools, omics-based approaches, remote sensing technologies, GIS-based computational platforms and exposure models using advanced artificial intelligence and machine learning methods.

As of its conceptual definition the exposome addresses its assessments by means of multidisciplinary data, which include for example factors indoor and outdoor environment, socioeconomics, lifestyle, and the individual’s ability to cope with various stressors such as infection or stress.

D3.2 – ELSI guidelines have already provided a preliminary analysis on new and existing data that will be required to further understand the links between environmental exposure and health diseases.

What types and formats of data will the project generate or re-use?

The data used in EIRENE RI can be classified into three major categories:

- Environmental: (monitoring data from measurement sites/campaigns). Data acquisition should be carefully and extensively documented, according to the FAIR principles. It is not expected that this data will be sensitive, but the data owner/provider must be clearly identifiable.
- Human studies: (statistics from population-based studies on morphometry, health and epidemiology). Data access will be via national nodes to ensure that all appropriate

data protection processes comply with international and national legislation and that ethics requirements are met.

- Laboratory Services/Capabilities: (data generated in EIRENE physical and virtual laboratories) this is effectively a shop window for the infrastructure's technical expertise and capacity. Visiting scientists will need to be made aware of the local/national legislation which covers the work they do in other countries.

Table 1 – Possible data categories generated or collected in EIRENE.

Data Category	Notes	Examples
Environmental (Earth Observation)	Include monitoring data on chemicals and ancillary parameters, which will have the geographical component.	pollution levels, land-use change, water quality, soil quality, vegetation
Human studies (Census)	Includes demographic and social statistics.	population, fertility, mortality
Human studies: Laboratory analyses (Nucleotide sequence-based)	Including genomics, epigenomics, metagenomics and transcriptomics.	genome, gene and transcript sequences
Human studies: Laboratory analyses (Biological and biochemical markers)	Includes nucleic acid–based biomarkers.	gene mutations or polymorphisms and quantitative gene expression analysis, peptides, proteins, lipids metabolites
Human studies: Health & lifestyle	Including health, lifestyle and nutrition, social environment, and psychology and stress.	health status, lifestyle
Human studies: External and internal exposure	Include modelled environmental data based on location profile of data subject, human biomonitoring data assessing internal chemical (and metabolite) exposure	Exposure to a particular chemical or chemical group

In detail, they can include environmental, social, economic and health disciplines (Table 1).

The EIRENE RI will use a range of data types and formats, from simple .txt, .csv and .xlsx files to gridded NetCDF and GRIB files, as well as GIS and MYSQL formats. No restrictions on data formats are set and where possible use of "inherently FAIR" (or more easily FAIRifiable) formats such as RO-Crate or FAIR Digital Objects will be encouraged.

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

Exposome studies in EIRENE cannot exist without data generation and re-use because the systematic evaluation of the exposome represents a challenge of simultaneous assessment of

thousands of synthetic and natural chemicals with a wide range of physicochemical properties, concentrations, and biological effects. Data availability can open new directions in population-based research enabling joint assessment of the chemical exposure and their downstream biological effects.

The re-use, particularly of biomedical data, should be well explained and justifiable while respecting the ethic-legal framework.

What is the expected size of the data that you intend to generate or re-use?

This is going to be highly variable and cannot be estimated in this phase of the EIRENE implementation.

What is the origin/provenance of the data, either generated or re-used?

The origin of data can be internal and external. Internal data will be originated through research activities in EIRENE that can be either laboratory analyses on human specimens or data/model analyses carried-out on collected datasets. External data originate from research, monitoring and surveys carried out independently of EIRENE. It is extremely important that all data providers are suitably acknowledged, where appropriate citing the literature that they suggest.

Extreme attention will be placed on protecting the Intellectual Property of data created and collected in and outside EIRENE.

To whom might your data be useful ('data utility'), outside your project?

Environmental data produced within the EIRENE RI will be of interest to local, regional, and national institutions. In our case this may also apply to datasets generated from analyses performed on combinations of existing datasets.

6. FAIR data

The 'FAIR Guiding Principles for scientific data management and stewardship' were published in Scientific Data (Wilkinson et al., 2016). The guidelines were suggested to improve the Findability, Accessibility, Interoperability, and Reuse of data, emphasising on machine-actionability. Since early 2018, the GO FAIR community has been working toward the implementation of the FAIR Guiding Principles (https://www.go-fair.org), whose framework provides a Three-point 'how to' guidance:

- Typically, the FAIRification process begins when a community of practice considers its domain-relevant metadata requirements and other policy considerations, and formulates these considerations as machine-actionable metadata components. These considerations can be guided in Metadata for Machines (M4M) Workshops.
- The re-usable metadata schemata produced in the M4M compose part of the larger FAIR Implementation Profile (FIP).
- The FAIR Implementation Profile in turn guides the choice and configuration of FAIR infrastructure, for example the use of FAIR Data Points (FDP) or FAIR Digital Objects (FDO) which contribute to a global Internet of FAIR Data and Services. (https://www.go-fair.org/how-to-go-fair/)

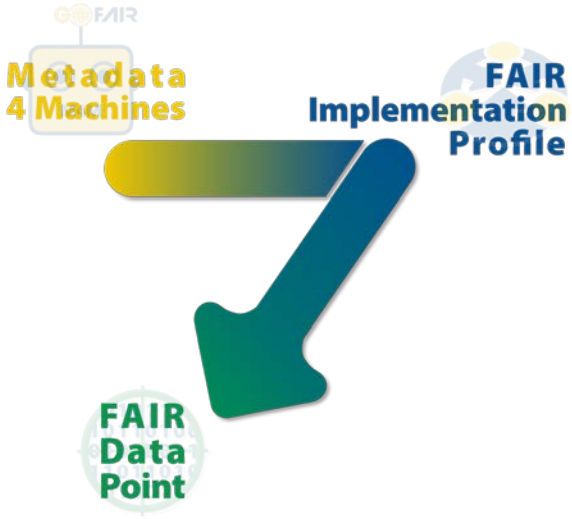


Figure 1 - The official icons for the Three-Point FAIRification Framework (Metadata for Machines Workshops, FAIR Implementation Profile and FAIR Data Point) have been registered at Zenodo under the Creative Commons Attribution Share Alike 4.0 International licence.

It should be noted that FAIR does not equate to 'open' ,see the following link for details, (https://www.gofair.foundation/openess). However FAIR principles apply to restricted as well as open data. Open data is fully available, and may be used also for commercial purposes. There are a number of reasons for which data within the EIRENE RI will be restricted, for perfectly justifiable reasons. These would include; confidentiality, privacy, legislative reasons, IPR motives and consent. In this case the motivation for the restrictions should be clearly stated, and the conditions for re-use of the data or the procedure(s) for gaining access to the data explained.

The following section reports a possible implementation of FAIR principles (FAIRification) in EIRENE having in mind that only general information is available on new and existing data that will be part of the infrastructure.

6.1. Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

A persistent identifier (PID) is a long-lasting reference to a resource (https://www.openaire.eu/what-is-a-persistent-identifier), with the primary purpose of providing the information required to reliably identify, verify, and locate it. Different PID types can be applied for different kinds of resources. In the current research environment, we most commonly see those for objects (publications, data, software, such as Uniform Resource Identifier (URI), Digital Object Identifier (DOI), Archival Resource Key (ARK), Handle System, Persistent Uniform Resource Locator (PURL)) and those for people (researchers, authors, contributors, such as Open Researcher & Contributor Identifier (ORCID), International Standard Name Identifiers (ISNIs)). Details on PIDs for objects, their difference and application can be discovered at https://arks.org/about/comparing-arks-and-other-identifiers/.

Generally speaking, a PID is assigned when an object is deposited in a repository, for example in figshare (https://figshare.com/), Harvard Dataverse Network (https://dataverse.harvard.edu/), Kaggle (https://www.kaggle.com), Network Data Exchange (https://www.ndexbio.org), Open Science Framework (https://osf.io/), Permanent Identifiers for the Web (https://w3id.org/), dPID (https://www.dpid.org/) and Zenodo (https://zenodo.org/).

In EIRENE, we suggest using ARK for data PIDs. This choice can be justified by looking at the differences between the PIDs (Table x).

Capabilities	ARK	DOI	Handle	PURL	URN
Decentralised admin and inferenceable syntax	Yes	No	No	No	No
Flexible metadata and persistence statements	Yes	No	No	No	No
Identifiers extensible during resolution	Yes	No	No	Yes	Yes
Free, non-paywalled, unlimited PID numbers	Yes	No	No	Yes	Yes

Details can be found on the ARK webpage so we highlight only a few of the advantages of ARK (https://arks.org/about/comparing-arks-and-other-identifiers/).

ARK can be implemented locally with open-source tools and work on a range of metadata and object types (dataset, database, file, etc.). This means that PIDs. can be created without metadata.

By design (being built for generic applications) ARK fits naturally with physical objects like samples or monitoring stations/campaigns and is able to create shorter identifiers.

ARK can be used in open infrastructures consistent with the values of EIRENE and can link directly to the objects instead of requiring landing pages. Moreover, one identifier enables millions of children by means of the suffix passthrough.

In summary, ARKs are the only mainstream, non-siloed, non-paywalled identifiers that can be provided by EIRENE itself, freely providing links directly to objects, which are machine- and human-friendly since they do not require an extra human navigation step for common tasks.

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

To the extent possible all collected and created datasets will be provided with metadata, according to the standards adopted in different disciplines. Sharing will occur through open-source catalogue tools, such as Geonetwork, a catalogue application to manage spatially referenced resources (https://geonetwork-opensource.org/).

Several resources that will be part of the EIRENE Data Infrastructure can already be discovered through standard metadata (i.e.: ISO 19115-x, https://www.iso.org/standard/80874.html), e.g. mercury concentration in the atmosphere (https://sdi.iia.cnr.it/gos4mcat). The Unified Modeling Language (UML) representation of the ISO 19115-2 Metadata is reported in Annex 1. Metadata are XML-based and defined using XML Schemas.

In the case of health data different standards are applied according to specific requirements. The Data Catalogue Application profile (DCAT-AP) for data portals in Europe (https://op.europa.eu/en/web/eu-vocabularies/dcat-ap) is a specification based on the Data

Catalogue vocabulary for describing datasets in Europe. It can enable cross-data portal searches for data sets and make them more easily searchable across borders and sectors. The HealthData@EU project (https://ehds2pilot.eu/) is developing an extension for descriptive metadata standards for health data. The application profile specification for metadata records meets the specific application needs. The UML representation of the DCAT-AP Metadata is reported in Annex 1.

Dataset discovery will be made possible by including the resource PID in the metadata.

Will search keywords be provided in the metadata to optimise the possibility for discovery and then potential re-use?

With the help of controlled vocabularies metadata will be enriched and discoverability facilitated. Adoption of already available controlled vocabularies (i.e. AgroVoc and/or EuroVoc) is preferred but they can be integrated with discipline vocabularies. An example is the thesaurus developed for mercury pollution and developed for improved discovery of mercury measurements (http://www.gos4m.org/download/gos4m-thesaurus-developed-under-i-gosp-e-shape-projects).

Will metadata be offered in such a way that it can be harvested and indexed?

As described above, metadata prepared according to standards will be provided for browsing and data access through open-source applications. These tools are suitable for both machine-to-machine and human-to-machine interactions. Once prepared in a standard format metadata can be harvested and/or indexed by common catalogue tools.

6.2. Making data accessible

Data accessibility refers to the easy retrieval of datasets for valuable analyses. A comprehensible and usable environment can facilitate data use to researchers with varying degrees of technical expertise. The central aspect of data accessibility is the tailoring of accessing tools

Repository:

Will the data be deposited in a trusted repository?

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

The EIRENE RI will itself be a trusted repository, so these questions should not arise. We may, however, reflect on the required properties to make EIRENE a “trusted” repository and the information we can provide to potential users on its trustworthiness.

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

Environmental data will be openly available while human data, sharing and sharing prerequisites will depend on subject consent, national and international laws, and the ethics guidelines. EIRENE will allow for capturing and publishing any relevant licensing and usage information in a FAIR way.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Embargos will most likely apply to new data and will be down to the data providers to specify the timeline for free and open access according to their funder's requirements (when applicable), the sensitivity of the data, and the time required for initial publication. Results will be published in Open Access journals referencing the data (PID) and directing to the procedures for access.

Will the data be accessible through a free and standardised access protocol?

Initially, not all data will be accessible through standard protocols, as EIRENE will need to find, collate, and catalog it. Where possible, data will be made accessible through free and standard protocols such as HTTP(S), FTP, SFTP, REST, SOAP, MQTT and CoAP.

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

This is partially covered in deliverable D3.2 - Ethical Guidelines. Health data is subject to both national and international regulation.

How will the identity of the person accessing the data be ascertained?

EIRENE will establish a data access request (DAR) system compliant with the EU GDPR to track the use of data and ensure the proper use of health data. Online identity verification can be done through multi-factor authentication (MFA), for example, username, password, and One-Time-Password received by a registered phone.

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

Access to personal or sensitive data may involve approval by the ethics committee of the data providers, hospitals, clinical trial leaders, national health institutions, etc., and so will be out of the hands of the EIRENE RI. Providing data to parties outside the EIRENE RI will be subject to national and international laws and those same ethics committees of the data providers. EIRENE RI will record relevant metadata for personal and sensitive data to provide accessibility information transparently and could if this would be decided to be in the scope from an infrastructure architecture standpoint, leverage this information to facilitate certain data access request workflows as part of the infrastructure services provided to the governing stakeholders for the datasets hosted in EIRENE RI.

Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

This should be the default. Only in some cases will the metadata contain information to access the data.

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

The EIRENE RI national data centres will provide a decentralised repository for EIRENE data and therefore the both data and metadata will remain as long as the RI exists. When the RI is eventually wound up then there will be protocols to follow to ensure the safety of the metadata and possibly the data.

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

This is most unlikely in the case of data from EIRENE. Some know-how may be required to use modelling of meteorological datasets in NetCDF or GRIB format for example, but one imagines that the researchers using these datasets have the necessary skills.

6.3. Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

For example:
DCAT / DCAT-AP- generic metadata schemes. Describe datasets and increases compatibility between data catalogues
EEA metadata - used for air quality measurement/modelling
INSPIRE Directive metadata - spatial data
PARC Chemical Risk Assessment vocabularies
PARC Environmental Metadata standards (building upon a.o. NORMAN)
PARC Human BioMonitoring Metadata standards (building upon a.o. HBM4EU)

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

The answer should be yes for EIRENE RI projects

E.g. I-ADOPT Framework - variable description ontology facilitating interoperable observations

Will your data include qualified references[1] to other data (e.g. other data from your project, or datasets from previous research)?

Datasets originated in EIRENE will be provided with metadata reporting lineage and scientific reference and links when available. Datasets collected will be properly provided with citation in respect of Intellectual Property Right and licence, if available.

6.4. Increase data re-use

In many instances it will be EIRENE that is re-using data, and participants should absolutely ensure that data provenance is thoroughly documented.

As a general rule the ‘as open as possible, as closed as necessary’ approach should apply.

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licences, in line with the obligations set out in the Grant Agreement?

Will the data produced in the project be usable by third parties, in particular after the end of the project?

Will the provenance of the data be thoroughly documented using the appropriate standards?

Describe all relevant data quality assurance processes.

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

All data produced in EIRENE will be licensed via the Creative Commons Licensing scheme to ensure attribution (CC BY). There are tools to aid researchers in choosing a licence for their data (e.g. https://choosealicense.com/, https://creativecommons.org/choose/). It is possible to also make sure that re-use of the data does not lead to research products or data sets with a more restrictive licence by using the CC-BY-SA licence. SA here stands for ‘Share Alike’.

Open datasets, tools, models, workflows will be indexed via OpenAIRE (https://catalogue.openaire.eu/home).

Data which is re-used within the EIRENE RI will be used in full respect of the licence with which it is made available.

7. Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Within the EIRENE RI it is expected that all research outputs will be FAIR, as this is fundamental to the RI’s philosophy. The interpretation of “FAIR” is very often in the eye of the beholder, though, and EIRENE will be able to make a difference by enabling EIRENE data and metadata records to be linked to external resources, thus being able to serve as a FAIR placeholder for non-FAIR or FAIR-but-not-interoperable resources to the exposure research community.

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

See above.

8. Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?

Making data and research outputs FAIR is part of the remit of the EIRENE RI and therefore intrinsic to the activity of the Research Infrastructure. EIRENE will transparently provide cost and governance information to its infrastructure users, so they can reliably estimate the costs involved and independently evaluate the governance characteristics of the service.

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

See above.

Who will be responsible for data management in your project?

Each of the national data hubs will have a data management team. Internationally, representatives from each of the national data hubs will form an oversight group to ensure as far as possible harmonised data management practices, metadata protocols. This group will also be the point of reference for the sharing of sensitive data according to national and international regulations regarding the treatment of such data.

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

This is the responsibility of the EIRENE RI, and the pilots and projects within the RI will be able to make use of the RI infrastructure. EIRENE will transparently provide preservation information to its infrastructure users, allowing them to include this in the metadata records accompanying their datasets.

9. Data security

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

Will the data be safely stored in trusted repositories for long term preservation and curation?

This will be part of the EIRENE RI remit, and the RI will provide both infrastructure and security appropriate to the data type.

10. Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?

This is covered by the Ethics Guidelines and will apply to all projects within the EIRENE RI

11. Other issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

Not foreseen

12. Oversight

The EIRENE 'steering committee' should set up a DMP review and advisory board to oversee the development of the DMPs for individual pilots and projects to ensure that they comply with all the Ethics Guidelines and also all the FAIR principles expected for the EU, ESFRI and the EIRENE RI 'constitution'

13. References

Wyborn, L.A., Friedrich, C., Rawling, T., Wu, M., Squire, G., Klump, J.F., Fraser, R., 2018. Building a multipurpose Geoscience Virtual Research Environment to cater for multiple use cases, a range of scales and diverse skill sets. In: American Geophysical Union, Fall Meeting 2018 abstract #IN33E-0887. https://ui.adsabs.harvard.edu/abs/2018AGUFMIN33E0887W/abstract.

Squire, G., et al., 2018. Scientific software solution centre for discovering, sharing and reusing research software. In: American Geophysical Union Fall Meeting 2018. American Geophysical Union, Washington, D.C. https://agu.confex.com/agu/fm18/meetingapp.cgi/Paper/459873

Schimel, D., Keller, M., Berukoff, S., Kao, R., Loescher, H.W., Powell, H., Kampe, T., Moore, D., Gram, W., 2011. NEON Science Strategy; Enabling Continental-Scale Ecological Forecasting. Pub. NEON Inc., Boulder CO, p. 55.

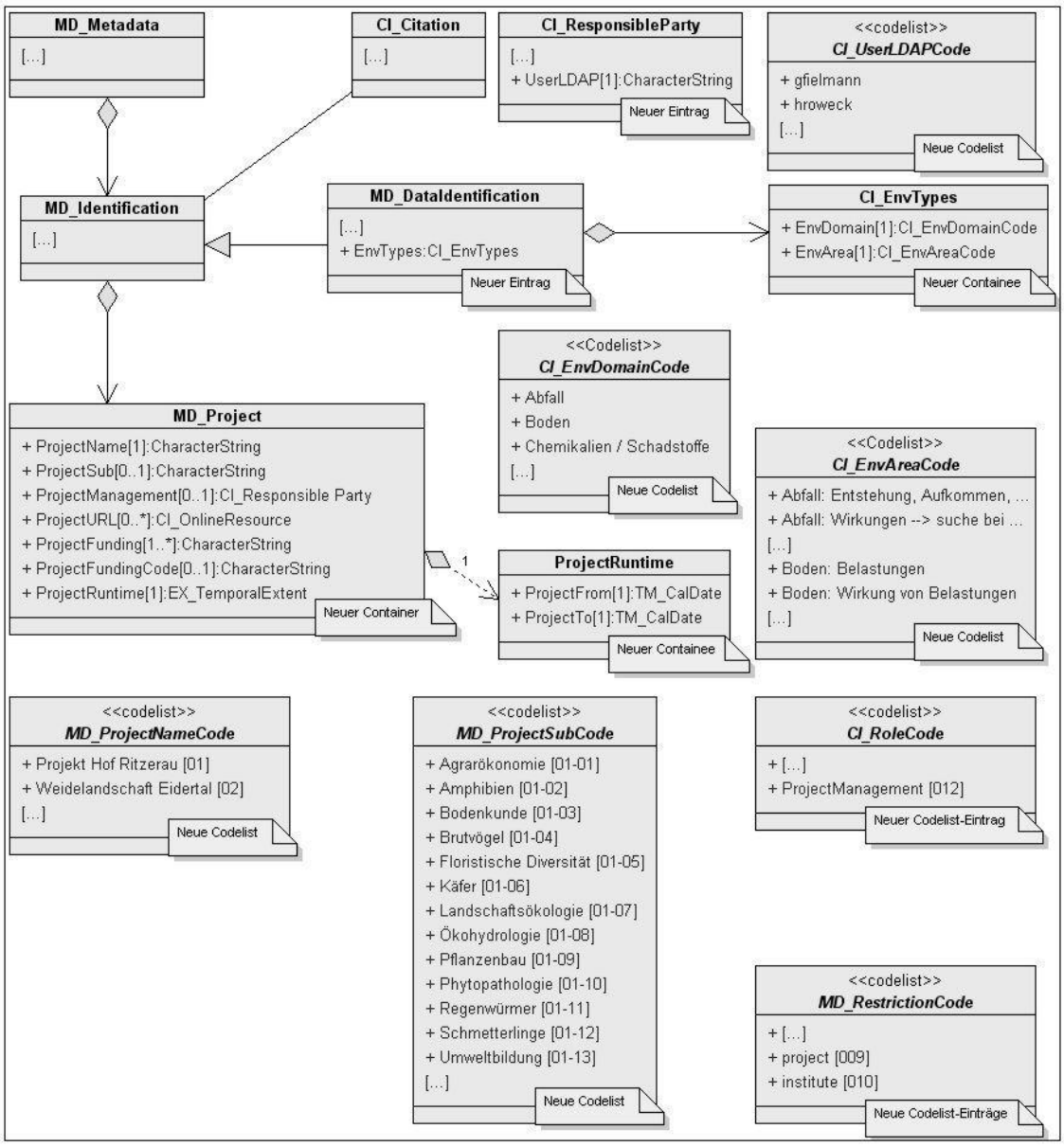
Fu, B.J., et al., 2010. Chinese ecosystem research network: progress and perspectives. Ecol. Complex. 7 (2), 225–233. https://doi.org/10.1016/j.ecocom.2010.02.007.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Annex 1

The Unified Modeling Language (UML) is a diagram that visualizes the structure of the system showing classes (i.e. structure, initial state and behaviour of objects) their attributes, operations (or methods), and the relationships among objects. The package-diagram is used to simplify complex class diagrams, by grouping classes into packages, which are collections of logically related UML elements.

UML package-diagram of ISO 19115 "Metadata" (ISO2, 2003)



UML package-diagram of DCAT-AP