Pla de gestió de dades per a Horitzó Europa

(Versió 1, 20 d'octubre de 2022)

(Doc.CO23/07) (4 RDM\DMP\Eina DMP\Plantilles\HorizonEurope\DataManagementPlan_HorizonEurope-CA.docx, 17.10.22)

Aquest document té com a objectiu donar suport als investigadors en l'elaboració del seu Pla de Gestió de Dades (*Data Management Plan, DMP*). S'adreça específicament a projectes finançats en el marc del programa Horitzó Europa de la Comissió Europea per crear un pla de gestió de dades FAIR.

A continuació es mostra:

- Amb un número, els camps que es requereixen a Horitzó Europa.
- Amb una lletra majúscula, els elements que s'han de tenir en compte a l'hora d'emplenar cada camp.
- Amb una lletra minúscula, les descripcions de cada element i una mostra d'exemples reals.

Aquest document ha estat elaborat pel Grup de Treball de Suport a la Recerca del CSUC, que està format per representants de les següents universitats: Universitat de Barcelona, Universitat Autònoma de Barcelona, Universitat Politècnica de Catalunya, Universitat Pompeu Fabra, Universitat de Girona, Universitat de Lleida, Universitat Rovira i Virgili, Universitat Oberta de Catalunya, Universitat de Vic-Universitat Central de Catalunya, Universitat Ramon Llull i Universitat de les Illes Balears.

Exemples de plans de gestió de dades estan disponibles en línia.[1]

Versió digital: http://hdl.handle.net/2072/531778

[1] Actris (Grant 654109), Citilab (Grant 635898), ConnectingGEO (Grant 641538), DIMENSION (Grant 688003), DR-BOB (Grant 696114), EGI-Engage (Grant 654142), FREME (Grant 644771), iCirrus (Grant 644526), MAGIC (Grant 689669), MAMI (Grant 688421), MMT (Grant 645487), RAMCIP (Grant 643433), SatisFactory (Grant 636302), Solidus (Grant 649489), Step (Grant 649493), Tandem (Grant 654206), UMobile (Grant 645124), U-Turn (Grant 635773), WaterInnEU (Grant 641821)

Informació preliminar

El lliurable de DMP definitiu ha d'incloure altra informació preliminar: el logotip del projecte, el nivell de disseminació, les revisions històriques, una taula del contingut i una llista dels acrònims utilitzats.

Consulta el "Periodic report template" (o els formularis al Portal del Participant) o contacta amb la teva institució.

1. Resum de les dades

1.A Quina és la finalitat de la generació o reutilització de dades i la seva relació amb els objectius del projecte?

1.A a) Descripció

Explicar de manera breu la finalitat de la recollida/generació o reutilització de dades i la relació de les dades amb els objectius del projecte.

1.A b) Exemple real

Ex. 1 The data will originate from measurements, calibrations, comparisons and validations. It will be used in meeting the project's objectives and in conference and peer-reviewed publications.

Experimental data will be collected by the consortium in order to meet objectives 1 - 4. Measurement and calibration data will result from objectives 1 and 3 and comparison and validation data from objectives 2 and 4. Data from questionnaires and market surveys will be used to support end-user uptake (objective 5).

Ex. 2 Collecting and making available the data of the analysis of superconducting materials to support the credibility and raise the quality of the scientific publications based on those data. Ease the exchange of data within the Consortium and promote the distributed characterization of samples with different methods. Permit follow-up projects and further generations of students continuing the work to build upon existing data sets, to validate the results and to document the improvement of materials and production techniques in a verifiable manner. This approach will ensure a durable impact of this EC funded project beyond the project period.

The objective of the project is to advance the performance of superconducting wires and at a later stage thin films by gaining a better understanding of the material behavior, the influence on the production techniques on the performance and to elucidate performance limitations (e.g. quality factor for superconducting thin films on substrate, current limits in wires under high-magnetic field conditions). Managed collection and publication of the data shall help establishing a durable library of results that can help documenting the performance evolution across several years and to permit other researchers validating the results independently.

1.B Quins tipus i formats de dades generarà o reutilitzarà el projecte?

1.B a) Descripció

Descriure el contingut i abast de les dades. Les dades de recerca es generen per diversos motius i a través de diversos processos, i poden ser dels següents tipus:

- Observacionals: dades capturades en temps real (neuroimatges, dades de mostres, dades de sensors, dades d'enquestes, etc.).

- Experimentals: dades captades per equips de laboratori (seqüències gèniques, cromatogrames, dades de camp magnètic, etc.).
- Simulació: dades generades a partir de models d'assaig (climàtics, matemàtics, econòmics, etc.).
- Derivades o compilades: dades reproduïbles però difícils de reproduir (mineria de textos i dades, models 3D, bases de dades compilades, etc.).
- Referència: conjunts de dades conglomerats (bases de dades de seqüències gèniques, estructures químiques, portals de dades espacials, etc.).
- Altres

També s'ha d'indicar el format de les dades (text, numèric, imatge, etc.).

1.B b) Exemple real

Ex. 1 The majority of the data will be in ASCII (American Standard Code for Information Interchange) data files, eg comma separated variable (CSV) format, which can be imported into rich-text files for word-processing or into spreadsheets. If specialised software is used, then information about free readers will be provided. Data will be generated in the following formats:

• Graphics: jpeg, odg, pdf, png, pttx
• Tables: odsu, opj, xlsx
• Text: docx, pdf, txt
• Other: nb, cpp

Ex. 2 The openly accessible data will be the comprehensive result data sets of characterized samples that are used to create the figures and plots in scientific publications, such that other researchers can compare their results easier and such that further results including historic data can be produced quicker. The data are value tables in Open Document Spreadsheet format (.ODS) for limited amounts of data with typed columns. For larger quantities of numeric data, UTF-8 encoded, comma separated value in textual format files (.CSV) with column value and data format description (FORMAT.TXT) will be used. In addition, images and raw measurement data files as provided by the measurement instruments will be stored on a project-internal data storage platform. Data files and images will be included in the open data sets. Proprietary raw data delivered by the measurement instruments will not be published. For all published files, a document record and change track will be included (author contact information, status, version, change reason and date, description of contents, title, origin of the data including a brief description of the measurement and/or experiment setup) in a separate metadata file for each characterization action called METADATA.ODS.

1.C Reutilitzaràs les dades existents i per a què les reutilitzaràs? Indiqueu els motius si s'ha considerat la reutilització d'alguna dada existent però s'ha descartat

1.C a) Descripció

Si es reutilitzen un conjunt de dades, cal especificar la font de la qual s'ha extret, per exemple, d'un repositori rellevant. Si es compren o reutilitzen fonts de dades existents, s'ha d'explicar com s'han abordat qüestions com ara els drets d'autor i els drets de propietat intel·lectual.

Quan es creen fonts de dades noves, cal explicar per què no es poden reutilitzar les fonts de dades existents.

1.C b) Exemple real

Ex. 1 Some of the project's tasks will use existing data in hdf, txt and xlsx formats. These data will be used in the validation of the project's results.

Ex. 2 Existing data from ongoing R&D projects in the scope of the <AcronymProject> study on superconducting wires and thin films will serve as a basis for the data files.

Ex. 3 Selected, existing images and data from the databases of the partner museums (<Partner1>, <Partner2>, <Partner3>...) will be used in specific tests, such as the storage tests in WP6. The final kind of data that will be created is that which is information in project deliverables, which must be preserved, made accessible and passed on to subsequent persons working in <AcronymProject>.

1.D Quin és l'origen/procedència de les dades, ja siguin generades o reutilitzades?

1.D a) Descripció

Si les dades es generen dins del projecte, cal indicar la font de les dades.
Si es recullen les dades, cal indicar la font de la qual s'han extret.
Si es reutilitzen les dades, cal indicar la font de la qual s'han extret.

1.D b) Exemple real

Ex. 1 The existing data will originate from several sources, which will include: partner's pre-existing data, data from the scientific literature, real-world measurement data and data from simulation experiments. The data collected from domestic properties will remain confidential and will not be included in the repository.

Ex. 2 The data stem from experiments and measurement campaigns performed by the ESRs and their colleagues at the beneficiary institutes: 1. Phase A: Superconducting wires and tapes: <Partner1>, <Partner2>, <Partner3>... 2. Phase B: Superconducting thin films: <Partner1>, <Partner2>, <Partner3>....

Ex. 3 These data have been digitised in diverse earlier projects.

1.E Quina és la mida esperada de les dades que es pretenen generar o reutilitzar?

1.E a) Descripció

Indicar el volum aproximat dels conjunts de dades. Cal considerar les implicacions dels volums de dades en termes d'emmagatzematge, còpia de seguretat, cost i accés. Calcular el volum de dades en MB / GB / TB i com aquest creixerà per assegurar-se que es pugui proporcionar emmagatzematge addicional i suport tècnic necessari.

1.E b) Exemple real

Ex. 1 The expected size of the data is not currently known, but it is likely to be <10 GB with individual files being ≤1 MB.

Ex. 2 The size of the data is today not known. Initial experience with storing results from different kind of measurements will permit revising this initial data management plan. The main relevant data sizes will stem from images such as microscopic sample characteristic that are stored in high-resolution bitmap format. However, the total data set size for a single sample characterization is expected to be in the order of tens of MB only.

Ex. 3 The size of the data handled by <AcronymProject> is quite small, such as less than 10 GB, except in the tests of the data infrastructure in WP6, where the project needs experience of managing large volumes of data, as explained above.

1.F A qui li poden ser útils les dades ("utilitat de dades"), fora del projecte?

1.F a) Descripció

Indicar el/s grup/s que poden estar interessats en les dades.

1.F b) Exemple real

Ex. 1 The data will be suitable for use by other research groups working on the following topics: biogas, biomethane, energy gases. It will also be useful for standards committees including ISO/TC193/SC1/WG25 Biomethane Working Group, ISO/TC 158 Analysis of Gases and regulators.

Ex. 2 Within the Consortium:

The data sets will be shared within the consortium as the working baseline to produce the scientific publications, to verify and validate the results through repeated experiments at different locations and as a baseline for a comprehensive documentation of the superconducting material performance evaluation in the scope of the world-wide Future Circular Collider technology R&D program.

Beyond the Consortium:

The data can be used by independent researchers to understand better the contents and conclusions of the scientific publications, which base their findings on the data. Furthermore, independent researchers can use the files to produce figures and publications, showing comparisons of their own results and the <AcronymProject> results. Scientists can also use

the data files to repeat the experiments and measurements to verify and validate the <AcronymProject> research. Finally, the data sets may also be used by scientific writers and the press to produce high-quality infographics, demonstrating the impact potentials of the technology.

Ex. 3 The data from these limited pilots will be useful for users and institutions who may be considering similar technologies in their digitisation and data management work. This applies in particular to the experiments carried out by WP6, but also the others. In particular, the digitised data from the experiments in WP3 will make apparent the quality of the digitisation results achieved with the new technologies. The data in the experiments of WP5 will be useful for the museums.

2. Dades FAIR

2.1 Fer que les dades es puguin trobar, inclòs el subministrament de les metadades

2.1.A Les dades s'identificaran mitjançant un identificador persistent?

2.1.A) Descripció

Explicar com s'assignen les dades i les metadades a un identificador globalment únic i eternament persistent (DOI, Handle...).

2.1.A b) Exemple real

Ex. 1 The institutional repository provides a unique URL to access the document with the format https://repository/record/1234.

Ex. 2 The repository assigns Handle/DOIs for persistent identification and citability of the dataset.

2.1.B Es proporcionaran metadades riques per permetre el descobriment? Quines metadades es crearan? Quins estàndards per disciplina o generals se seguiran? En cas que els estàndards de metadades no existeixin en la disciplina, fer un resum de quin tipus de metadades es crearan i com.

2.1.B a) Descripció

Metadades riques proporcionades: Les metadades han de documentar com es van generar les dades, sota quina llicència i com es poden reutilitzar. A més, les metadades ajuden a descobrir les dades i proporcionen el context per a la correcta interpretació per part d'altres investigadors.

Metadades creades i estàndards: Indicar els estàndards de metadades que s'utilitzaran. Es recomana utilitzar estàndards de metadades específics de la disciplina. Consultar estàndards de metadades.

Si no s'utilitzen estàndards de metadades, indicar quines metadades es generaran (manualment o automàticament) i com.

2.1.B b) Exemple real

Ex. 1 The metadata standard used to describe the dataset will be the Dublin Core Schema, as it is a flexible and common used standard and is also the one adopted by the repository.

Ex. 2 Metadata are created manually by depositors in the deposit form at the repository.

Ex. 3 (1) The data are expected to be provided in ANSI SQL, XML or text (ASCII) format. For this dataset, data citation and metadata practices derived from the community will be considered.

(2) There are no standards for these logs. A possible solution is project servers such as AAA servers. In this case, the logs would include the attributes defined by "project".

Ex. 4 Each file associated with data will be accompanied with unique specified metadata to allow ease of access and re-usability. Below, the form to be followed is presented.

Ex. 5 Standards such as the Dublin Core and ISO/IEC 11179 Metadata Registry (MDR), which addresses issues in the metadata and data modelling space, will be considered.

Ex. 6 There are many different metadata standards for many different types of data and it may not be possible to find one that fits all purposes. Therefore, a pragmatic and feasible approach is to agree on a common and minimal catalogue metadata schema for those datasets that are published in public catalogues and data repositories and to use data-type specific schema extensions, if necessary.

In general, the Zenodo deposition metadata domain model which is based on DataCite's metadata schema minimum and recommended terms will be used for open data generated by the project and deposited in an appropriate repository.

2.1.C Es proporcionaran paraules clau de cerca a les metadades per optimitzar la possibilitat de descobriment i després, de reutilització potencial?

2.1.C a) Descripció

Indicar com es crearan les paraules clau de cerca de contingut per optimitzar la recuperació i la reutilització.

2.1.C b) Exemple real

Ex. 1 Data must be findable easily, rapidly and identically. Therefore, exact and standard measures have to be used to identify the data sets. This can include the definition and use of naming conventions, search keywords, version numbers, metadata standards and standard data identifiers.

Ex. 2 All open project results deposited in a repository will provide search keywords together with their metadata. Keywords for open data will be selected from controlled vocabularies that are suitable for the specific type of data.

2.1.D S'oferiran metadades de manera que es puguin recol·lectar i indexar?

2.1.D a) Descripció

Les metadades s'han de proporcionar estructurades mitjançant l'Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) de manera que es permeti l'intercanvi amb altres repositoris. Al mateix temps, les metadades proporcionades han de ser el més detallades possible per permetre que s'indexin i que les dades siguin cercables i recuperables.

2.1.D b) Exemple real

Ex. 1. Datasets published in CORA.RDR will be harvested and indexed using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). They will be also indexed in OpenAIRE, RECOLECTA, Google Dataset Search and Mendeley Data.

2.2 Fer accessibles les dades

2.2.A Quines dades produïdes i/o utilitzades en el projecte es posaran obertament a disposició per defecte?

2.2.A) Descripció

Descriure com es compartiran les dades, inclosos els procediments d'accés, els períodes d'embargament (si n'hi ha) i la definició de si l'accés serà obert o restringit a col·lectius específics. Si algunes no poden estar obertament disponibles, s'ha de justificar per què.

2.2.A b) Exemple real

Ex. 1 All data produced by the experiments of WP3, WP4, WP5, and WP6, which has been described above, will be made openly available. This is any imagery and results of automatic or computer-assisted human interpretation of the data, which can be seen in the imagery. This does not mean that also the details of the equipment used and algorithms used in the interpretation will be made openly available, as these may contain proprietary information. In Zenodo, the option exists to provide open access, embargoed access, closed access.

Ex. 2 All of the data associated with scientific publications will be made openly available as the default unless there is a specific reason not to publish the data. Datasets which cannot be shared – voluntary restrictions Other data may be made available on a case-by-case basis if it is relevant for third parties.

The following data will not be made publicly available:
• Data obtained with the permission of third parties, but the third parties have not agreed to make the data publicly available.
• Data that discloses the identity of a manufacturer.
• Data that compromises the protection of a partner(s) intellectual property. The level of data made available will also be considered, for example, pre-processed data will not be provided unless there is a clear reason for doing so.
Datasets which cannot be shared - legal and contractual reasons All of the data from the project will be made available, except for market or customer survey data, which are commercially sensitive and cannot be shared.

2.2.B Com es faran accessibles les dades (per exemple, depositant en un repositori)?

2.2.B a) Descripció

Descriure com es compartiran les dades, és a dir, qui tindrà accés al conjunt de dades. Es pot crear un procediment per fer que les dades siguin accessibles temporalment a altres membres del grup, socis del projecte i al públic en general. S'ha d'indicar si les dades seran d'accés obert i en quin període raonable. Una possibilitat és oferir-les juntament amb les publicacions. Si es requereixen períodes d'embargament, aquí és on cal especificar-los.

2.2.B b) Exemple real

Ex. 1 The data will be deposited in the storage systems which will be tested by WP6, as appropriate (<Repository1>, <Repository2>, <Repository3>). Links from <AcronymProject> website will be provided to these storage systems. By their service definition, the data stored at <Repository1> remains permanently available. Permanent access to the data on national <Repository2> and <Repository3> tests is not foreseen. Data from the digitisation pilots may remain permanently available if published on <Repository4>. These arrangements will be revisited after the data from the pilots has been created.

Ex. 2 Once processing, quality control, organisation, analysis and publication are complete, the data will be made accessible by deposition in open access repositories (eg Zenodo).

2.2.C Quins mètodes o eines de programari es necessiten per accedir a les dades?

2.2.C a) Descripció

Incloure qualsevol requisit tècnic per a l'accés i la reutilització de les dades.

2.2.C b) Exemple real

Ex. 1 Web browser and/or application programming interfaces (API) offered by these storage systems, complemented by customized tools developed by users in specific domains. Zenodo provides basic robust, fast services. Anything on top of it is envisioned to be layered, and not necessarily part of the Zenodo infrastructure. For example, viewing and searching multiple images has to be handled outside Zenodo, e.g., by using <ExampleURL> that is currently being developed by <Partner1> for the domain-specific Biodiversity Literature Repository.

Ex. 2 The data will be accessible using the following software: MS Office, Matlab, Mathematica, Origin, Open Office, Adobe Reader, Image Viewer.

2.2.D És necessària la documentació sobre el programari per accedir a les dades incloses?

2.2.D a) Descripció

També s'ha d'incloure la documentació del programari que es necessita per accedir a les dades.

2.2.D b) Exemple real

Ex. 1 If accessed through the API, documentation will be needed.

Ex. 2 Standard publicly available software will be used where possible, but if specialist software tools are developed, i.e. created within Matlab, a short text file (e.g. ASCII) will be provided with the data file to explain the software required.

2.2.E És possible incloure el programari pertinent (per exemple, en codi obert)?

2.2.D a) Descripció

En cas que es tracti de programari específic, incloure'lsi és possible. Per exemple, en codi obert.

2.2.D b) Exemple real

Ex. 1 Any such software has already been released by the providers of these storage systems.

Ex. 2 The majority of the software programmes are available as commercial products or as freeware. For the software developed in the project, the source code will be deposited in the repository (eg Zenodo).

2.2.F On es dipositaran les dades i metadades, documentació i codi associats?

2.2.F a) Descripció

Indicar el repositori en el qual s'emmagatzemaran les dades i les metadades, documents i codi associats. Pot ser el mateix repositori o diferents repositoris per als diferents tipus de contingut, per exemple, el codi es podria dipositar en un repositori específic per al codi. Hi ha disponible un document de recomanacions per seleccionar els repositoris de dades de recerca del CSUC.

És important utilitzar un repositori que proporcioni enllaços permanents (DOI, handle) a les dades per tal de facilitar la cerca i la citació.

2.2.F b) Exemple real

Ex. 1 The data will be deposited in the storage systems which will be tested by WP6, as appropriate (national OSC, EUDAT, Zenodo). Links from <AcronymProject> website will be provided to these storage systems.

Ex. 2 The data and associated metadata, documentation and code will either be deposited in the open access repository called Zenodo or in Open Access Repository (<ExampleURL>).

2.2.G Heu explorat els criteris apropiats amb el repositori identificat?

2.2.G a) Descripció

Indicar si s'ha explorat adequadament quins són els requisits del repositori identificat.

2.2.G b) Exemple real

Ex. 1 We have already explored the appropriate arrangements with the national cloud services in Finland (CSC), EUDAT through the work of <AcronymProject> pilot, and Zenodo through the work of the <Disciplinary> Literature Repository community.

Ex. 2 Yes, Open Access Repository is functional and it correctly labels datasets with a metadata scheme that is compatible with DataCite).

2.2.H Si hi ha restriccions d'ús, com es facilitarà l'accés?

2.2.H a) Descripció

En cas que l'accés públic a les dades estigui restringit per qualsevol motiu justificat, especificar si les dades serien accessibles per a un soci individual, per a tots els socis o sota sol·licitud. Especificar els procediments de com sol·licitar l'accés a les dades restringides i en quines condicions es concedirien. A més, especificar si les restriccions s'aixecaran després d'un període de temps.

2.2.H b) Exemple real

Ex. 1 There are no restrictions on use, except when CC BY-NC license has been chosen. <AcronymProject> should address question of sensitive data (e.g. location of protected plants), but <AcronymProject> will avoid working with any sensitive data. If personal data is received in questionnaires, which <AcronymProject> will receive, such data shall be anonymised before making available outside the project.

Ex. 2 There are no restrictions on the use of the published data, but users will be required to acknowledge the consortium and the source of the data in any resulting publications.

2.2.I És necessari un comitè d'accés a les dades?

2.2.I a) Descripció

Especificar per què o per què no és necessari un comitè d'accés a les dades.

2.2.I b) Exemple real

Ex. 1 Because of the small scale of these experiments, there is no need for a data access committee.

Ex. 2 This consortium will have a data access committee. Their remit will be to select the data that will be openly accessible on a case by case basis. Ethical aspects and data security, including intellectual property requirements, will be considered. If necessary, some or all of a potential publication's data will be withheld. This will be decided in consultation with the relevant partner(s).

2.2.J Hi ha condicions d'accés ben descrites (és a dir, una llicència llegible per màquina)?

2.2.J a) Descripció

Descriure quines són les condicions d'accés definides pel repositori que heu triat (també podeu indicar la URL d'on prové la informació). Per exemple: una llicència llegible per màquina.

2.2.J b) Exemple real

Ex. 1 The Creative Commons licenses supported by the GBIF will be used. These include CC0, CC-BY, and CC BY-NC (see <ExampleURL>). Zenodo supports a large array of widely used as well as domain specific, machine-readable licences. The owner of the data will determine which of these licenses will be used when data is posted on <AcronymProject> repositories. However, it is the project's recommendation to choose CC0 for data and CC-BY for media and avoid CC-BY-NC which has issues in some national jurisdictions.

Ex. 2 Yes, Zenodo provides well-described conditions for access (see http://about.zenodo.org/policies/).

2.2.K Com es comprovarà la identitat de la persona que accedeix a les dades?

2.2.K a) Descripció

Descriure el procediment establert pel repositori per determinar la identitat de la persona que accedeix a les dades.

2.2.K b) Exemple real

Ex. 1 Identity of the person accessing the data will not be directly ascertained. However, we expect users to follow the standard norms of scientific citation and use of the data in this context will be tracked through scientific citation.

Ex. 2 Users are required to register to use the repository.

2.3 Fer que les dades siguin interoperables

2.3.A Especificar quins vocabularis, estàndards o metodologies de dades i metadades es seguiran per fer que les dades siguin interoperables per permetre l'intercanvi i la reutilització de dades dins i entre disciplines? Se seguiran les millors pràctiques d'interoperabilitat aprovades per la comunitat? Quines?

2.3.A) Descripció

Explicar quins vocabularis, estàndards o metodologies de dades i metadades es seguiran per facilitar la interoperabilitat. La interoperabilitat de dades del projecte permet l'intercanvi i la reutilització de dades entre investigadors, institucions, organitzacions, països, etc. Adherir-se als estàndards de formats que siguin, en la mesura del possible, compatibles amb programes i aplicacions obertes.

2.3.A b) Exemple real

Ex. 1 The data produced in the project will be interoperable as the datasets will adhere to standardised formats: ASCII, txt, csv, xml, tiff. If MS Office, pdf viewer or image viewer cannot be used, a text (ASCII) file will be provided with the dataset that explains where a free reader can be obtained.

Ex. 2 Final data will be published in CORA.RDR, the repository of Consorci de Serveis Universitaris de Catalunya (CSUC). This repository follows the Open Archives Initiative model, which allows interoperability with the OAI-PMH metadata transmission protocol (Open Archive Initiative - Protocol for Metadata Harvesting). This protocol allows visibility of the documents from different platforms and collectors: Google Scholar, BASE, CORE, etc. This data repository is OpenAIRE compliant and meets all the requirements of metadata required by the European Commission.

Persistent IDs are provided for each document (DOI) and author identifiers (ORCID) are included in the metadata. The metadata standard used to describe the dataset is the DDI's metadata schema compatible with the Dublin Core, a flexible and commonly used standard that is also adopted by the european OpenAIRE repository.

We use the vocabulary [to be completed] being a standard vocabulary in this field.

2.4 Augmentar la reutilització de dades (mitjançant llicències aclaridores)

2.4.A Com s'aportarà la documentació necessària per validar l'anàlisi de dades i facilitar la reutilització de les dades?

2.4.A) Descripció

També s'ha d'incloure la documentació necessària per validar l'anàlisi de dades i facilitar la reutilització de les dades (per exemple, fitxers readme amb informació sobre metodologia, llibres de codis, neteja de dades, anàlisis, definicions de variables, unitats de mesura, etc.).

2.4.A b) Exemple real

Ex. 1 Metadata records will accompany the data files in order to describe, contextualise and facilitate external users to understand and reuse the data.

2.4.B Quina llicència es donarà a les dades per permetre la reutilització més àmplia possible, d'acord amb les obligacions establertes en el Contracte de subvenció?

2.4.B a) Descripció

Si les dades es posen a disposició d'altres investigadors i del públic en general, cal especificar quin grau de reutilització es permet. Aquest nivell de reutilització vindrà marcat per l'establiment de llicències. La CE proposa l'ús de llicències Creative Commons CC-BY o CC0, però n'hi ha d'altres.

2.4.B b) Exemple real

Ex. 1 The deliverables associated to the dataset are licensed through an All rights reserved license as they are working papers not intended to be re-used. Nevertheless, the database should be shared as a possible reusable dataset. For this reason, when deposited to the repository, an Attribution-NonCommercial license (by-nc) will be requested. The data is currently available for re-use from the project website and will also be findable and reusable through the final depositing repository (the institutional one or Zenodo) and from OpenAire, the latest by the end of the project.

Ex. 2 Wherever possible the data will be shared right after production following the Creative Commons 4.0 International License with Attribution (CC4BY). Experimental data test data will in some cases only become available after the end of the project or publication of the results, whatever comes first, and will be shared used the same CC4BY license.

2.4.C Les dades produïdes i/o utilitzades en el projecte són utilitzables per tercers, després de la finalització del projecte? Si la reutilització d'algunes dades està restringida, explicar per què.

2.4.C a) Descripció

En principi, les dades s'haurien de posar a disposició d'altres investigadors i del públic en general amb les menors restriccions possibles. No obstant això, hi pot haver diversos motius per no compartir-los: raons ètiques, protecció de dades personals, implicació de drets de propietat intel·lectual i/o industrial, interessos comercials, etc. S'ha d'especificar els motius pels quals no es compartirà un conjunt de dades.

2.4.C b) Exemple real

Ex. 1 IPRs and Privacy Issues. Data access and sharing activities will be rigorously implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and in the EU, as well as with the H2020 rules. Raw data collected through the interviews from externals the consortium sources may be available to the whole consortium or specific partners upon authorization of the owners. This kind

of data will not be available to the public. The results of the project will become publicly available based on the IPRs, as described in the Consortium Agreement.

Ex. 2 The full dataset will be confidential and only the members of the consortium will have access to it. Furthermore, if it is decided to make specific portions of it (e.g. metadata, statistics, etc.) widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section. Of course, these data will be anonymized so as not to have any potential correlation and identification of the ethical issues with their publication and dissemination.

Ex. 3 Each archived data set will have its own permanent repository ID and will be easily accessible. We expect most of the data generated to be made available without restrictions and only data sets subject to IPR and confidentiality issues will be restricted. Where this is going to be the case, agreements will be made based on the individual data sets. Requests for the use of the data by externals will be approved by the project consortium.

2.4.D Com es documentarà la procedència de les dades utilitzant els estàndards adequats?

2.4.D a) Descripció

Cal incloure informació sobre entitats, activitats i persones implicades en l'elaboració de dades.

2.4.D b) Exemple real

Ex. 1 The documentation and metadata of each dataset recognize the data provenance through proper citation of the source of information and entities using the formats usually accepted by the relevant scientific community.

2.4.E Es descriuen els processos d'assegurament de la qualitat de les dades?

2.4.E a) Descripció

Descriure quins són els processos d'assegurament de la qualitat de les dades. Com/quan s'implementaran les avaluacions internes de qualitat de les dades?

La qualitat de les dades es pot assegurar mitjançant diferents mesures. Aquestes inclouen validació de la mostra, replicació i comparació amb resultats d'estudis similars i control de la distorsió sistemàtica.

2.4.E b) Exemple real

Ex. 1 The quality of the dataset is guaranteed by the platform functioning.

Ex. 2 The data quality is ensured by different measures. These include validation of the sample, replication and comparison with results of similar studies and control of systematic distortion.

Ex. 3 Data quality assurance and control is central and the raison d'étre of this project. About 80% of the efforts spent in our Thematic Centres is directed at data quality assurance.

Ex. 4 For our research data collection, the quality control of the data can happen at various stages during the quality assurance process. Initial quality control is needed at the local level and early in the collection process. Additional controls will take place at a later stage of the data lifecycle. Final quality control of metadata takes place during its input into IMIS. The initial quality control of the data, during data collection, is the primary responsibility of the project data creator/owner, who must ensure that the recorded data reflect the actual facts, responses, observations and events. The quality of the data collection methods used strongly influences data quality, and documenting in detail how data are collected provides evidence of such quality. Errors can also occur during data entry. Data are digitised, transcribed, entered in a database or spreadsheet, or coded. Here, quality is ensured by standardised and consistent procedures for data entry with clear instructions.

3. Altres resultats de la recerca

3.A Hi haurà altres resultats de la recerca que es puguin generar o reutilitzar al llarg del projecte?

3.A a) Descripció

Explicar quins altres resultats de la recerca s'han generat en l'execució del projecte. Poden ser productes digitals com programari, fluxos de treball, protocols, models, etc. o productes físics com nous materials, anticossos, reactius, mostres, etc.

3.A b) Exemple real

Ex. 1 (Yes) It will be a series of new materials and samples derived from this research.

Ex. 2 (No) It won't be any other research output.

3.B Especificar quines de les preguntes relatives a les dades FAIR, poden aplicar-se a la gestió d'altres resultats de la recerca

3.B a) Descripció

Si s'obtenen altres resultats de la investigació, explicar com fer-los trobables, accessibles, interoperables i reutilitzables. Per a més informació, veure 2. Dades FAIR:

- Explicar si altres resultats de recerca s'identificaran mitjançant un identificador persistent, es proporcionaran metadades riques per permetre el descobriment, es recolliran i s'indexaran, quines seran les metadades creades, quins estàndards disciplinaris o generals se seguiran i si es proporcionaran paraules clau de cerca a les metadades.
- Explicar com seran accessibles aquests resultats de recerca des del repositori, com d'oberts o restringits seran i quina llicència s'utilitzarà.
- Explicar quins vocabularis, estàndards, formats o metodologies de dades i metadades se seguiran perquè els resultats de recerca siguin interoperables per permetre l'intercanvi i la reutilització d'aquests resultats dins i entre disciplines.
- Explicar com es farà perquè aquests resultats de recerca es reutilitzin. Quina documentació s'aportarà, quina llicència de reutilització s'aplicarà, etc.

3.B b) Exemple real

Ex. 1 See 2. FAIR Data real examples.

4. Assignació de recursos

4.A Quins seran els costos per fer FAIR les dades o altres resultats de recerca del projecte?

4.A) Descripció

Indicar el cost aproximat per fer les dades FAIR i com es pensa cobrir-les: costos directes i indirectes relacionats amb l'emmagatzematge, arxiu, reutilització, seguretat, etc.

4.A b) Exemple real

Ex. 1 There are no costs associated to the described mechanisms to make the database FAIR and long term preserved.

Ex. 2 The costs for depositing the dataset with the project, and subsequent resources required to make the dataset publicly available have been included within specific Work Packages within the project.

4.B Com es cobriran aquests costos?

4.B a) Descripció

Indicar com es preveu cobrir el cost de fer que les dades siguin FAIR, inclosos els costos addicionals d'arxiu i conservació.

Tenir en compte: que els costos relacionats amb la gestió de les dades i els resultats de recerca són elegibles com a part de l'ajut Horitzó Europa (si compleixen les condicions de l'acord de subvenció)

4.B b) Exemple real

Ex. 1 <AcronymProject> is managed and supported by a team of experts and is free of charge.

Ex. 2 The cost of preserving the database will be assumed by the <Partner1>.

Ex. 3 (1) A dedicated hard disk drive will probably be allocated for the dataset. No costs are currently foreseen regarding its preservation.

(2) The cost will be covered at the local hosting institute in the context of the project.

(3) The cost will be covered at the local hosting institute as a part of the standard network system maintenance.

4. C Qui serà el responsable de la gestió de les dades en el seu projecte?

4.C a) Descripció

Explicar les responsabilitats de gestió de dades en el projecte.

4.C b) Exemple real

Ex. 1 The project coordinator has the ultimate responsibility for the data management in the project and so, for the Marketplace platform management.

Ex. 2 Each partner has to respect the policies set out in this DMP. Datasets have to be created, managed and stored appropriately and in line with applicable legislation.

- The Project Coordinator has a particular responsibility to ensure that data shared through the website are easily available, but also that backups are performed and that proprietary data are secured.
- WP1 leader, will ensure dataset integrity and compatibility for its use during the project lifetime by different partners.
- Validation and registration of datasets and metadata is the responsibility of the partner that generates the data in the WP.
- Backing up data for sharing through open access repositories is the responsibility of the partner possessing the data.
- Quality control of these data is the responsibility of the relevant WP leader, supported by the Project Coordinator.

4.D Com s'assegurarà la preservació a llarg termini?

4.D a) Descripció

Indicar com es planifica la preservació a llarg termini i qui decideix quines dades es conservaran i durant quant de temps.

4. D b) Exemple real

Ex. 1 Regarding the question of long-term data preservation, no specific arrangements has been done in the consortium yet. However, with a great degree of confidence, it can be confirmed that it is the project coordinator with the help of local <AcronymProject> resources who will play the major role in this task.

5. Seguretat de les dades

5. A Quines disposicions hi ha o hi haurà per a la seguretat de les dades (inclosa la recuperació de dades, així com l'emmagatzematge / arxiu segur i la transferència de dades sensibles)?

5.A) Descripció

Descriure breument les mesures tècniques que s'implementaran a curt i mitjà termini per garantir la integritat de les dades (còpia de seguretat de les dades), la recuperabilitat (prevenció de la pèrdua de dades) i la seguretat (per evitar accessos no autoritzats).

5.A b) Exemple real

Ex. 1 Data collected from the research group for the Project will be digitised and stored on the University's <StorageService> which is subject to regular back-up that is controlled by the University's IT personnel. The IT department performs operations by type: mission-critical (user data, virtual machines, scientific results, etc.) and static (scientific data sets, intermediate files, etc.). Content will be checked regularly to preserve its integrity, security, and durability. These procedures are designed, set and applied in order to fully comply with personal data as ruled by Directive 95/46/EC (General Data Protection Regulation) and other current national legislation and institutional regulations. Research team members will have an appropriate access level according to their role in the project.

5.B Les dades s'emmagatzemaran de manera segura en repositoris de confiança per a la seva preservació i curació a llarg termini?

5.A) Descripció

Descriure on s'emmagatzemaran les dades de manera segura en un repositori de confiança per a la seva preservació i curació a llarg termini. Descriure també breument la configuració de seguretat del repositori escollit.

5.A b) Exemple real

Ex. 1 At the end of the project, all the final data will be deposited at the CORA.RDR data repository. The following list describes their security settings:

- Versions: Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and records are preserved.
- Replicas: All data files are stored in the CSUC Centre, primarily in Barcelona, with replicas in Consorcio Madroño in Madrid. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- Retention period: Items will be retained for the lifetime of the repository. The RDR has defined a lifetime for the repository of the next 10 years minimum.

- Functional preservation: The RDR makes no promises of usability and understandability of deposited objects over time.
- File preservation: Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
- Fixity and authenticity: All data files are stored along with an MD5 checksum of the file content and the tabular file is stored with Universal Numerical Fingerprint (UNF).
- Files are regularly checked against their checksums to assure that file content remains constant.
- Succession plans: In case of closure of the repository, a guarantee has been made from RDR to migrate all content to suitable alternative institutional and/or subject-based repositories.

6. Aspectes ètics

6.A Hi ha, o podria haver-hi, alguna qüestió ètica o legal que pugui tenir un impacte en la compartició de dades?

6.A) Descripció

Descriure qualsevol problema ètic o legal que pugui tenir un impacte en la compartició de dades. Es poden tractar en el context de la revisió ètica. Si és rellevant, incloure referències a lliurables ètics i capítol d'ètica a la Descripció de l'Acció (DoA).

Especificar si el consentiment informat per compartir dades a llarg termini s'inclourà als qüestionaris sobre dades personals.

És important destacar en aquest punt qualsevol aspecte mencionat a l'article 34 dels acords del finançament "Article 34 — Ethics and research integrity".

Si les activitats de recerca impliquen infants, pacients, poblacions vulnerables, l'ús de cèl·lules mare embrionàries humanes (hESC) i embrions humans (hEs), humans, cèl·lules o teixits humans, dades personals, animals, països no comunitaris, medi ambient, salut i seguretat, intel·ligència artificial, altres qüestions ètiques (interacció home-màquina, es desenvolupa en nanotecnologia...), i qüestions transversals: possible mal ús dels resultats (Activitats que impliquen o generen materials, mètodes, tecnologies o coneixements que puguin ser mal utilitzats amb finalitats poc ètiques) s'ha de complir els principis ètics i la legislació nacional, comunitària i internacional pertinent.

6.A b) Exemple real

Ex. 1 All the activities carried out under the <AcronymProject> project comply with ethical principles and relevant national, EU and international legislation, for example the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights. The tasks for <AcronymProject> only concern basic research activities and the project does not involve humans, animals or cells. Due to the fact that the main domain of the <AcronymProject> project activity is related to materials science with the focus on refractory materials, the risk of having ethics issues during the project is extremely limited. Either way, within the <AcronymProject> DoA Part A, the workpackage 8 is devoted to the ethics issues which sets out the 'ethics requirements' that the <AcronymProject> project must comply with. One deliverable will be provided: D8.1 NEC -Requirement No. 1. In the framework of D8.1, all beneficiaries and partner organisations must confirm that the ethical standards and guidelines of Horizon2020 will be rigorously applied, regardless of the country in which the research is carried out.

Ex 2. The transfer of data on human subjects to the <AcronymProject> repository is only considered when: informed consents, ethics approval and – when applicable - approval by local data protection authorities cover the purpose that the data are envisaged to be used within <AcronymProject> and allow transfer of individual or aggregated data to the <AcronymProject> repository. All data that are transferred to the <AcronymProject> repository shall be either pseudonymised or completely anonymized. The Data Owner/Data Provider is responsible for the anonymization or prseudonymation process and for ensuring

that identifiable variables are not transferred to the <AcronymProject> repository. Directly identifiable variables include - but are not limited to - national ID number, name, phone number, ZIP-code, e-mail address, address, geographical coordinates (at a resolution that risks identification). One shall also be aware that a combination of just of few indirect identifying variables (such as birth data, gender, and zip-code) can be used to identify a large portion of individuals on any dataset. In this context, the Data Owner/Data Provider shall only provide such variables at the lowest possible resolution that is necessary to for analysis, e.g. district instead of zip-code; year of birth or age instead of birth date.

7. Altres temes

7.A Feu o fareu ús d'altres procediments nacionals/finançadors/sectorials/ departamentals per a la gestió de dades?

7.A) Descripció

Explicar els procediments nacionals / finançadors / sectorials / departamentals per a la gestió de dades que s'utilitzin.

7.A b) Exemple real

Ex. 1 As part of <University>'s commitment to ensuring FAIR and Open data, all research active staff (Postdoctoral fellows, PhD students) are expected to prepare DMPs for their own data, as per the University's Research Data Management Policy. The <University> data management policy defines research data as "the evidence that underpins the answer to the research question and can be used to validate findings regardless of its form." Thus, data covers quantitative and qualitative statements, raw data from measurements and derived data–either cleaned or extracted from a researcher's primary dataset or derived from an existing source.

Ex. 2 As well as European Commission policies on open data management, Project Partners must also adhere to their own institutional policies and procedures for data management:

Imperial College London:
- Recommended file storage options
- Encrypt sensitive information

UCAL
- Regolamento per la gestione dell'innovazione e della proprietà intellettuale e industriale. Rectoral Decree n.1597, 19/10/2015
- Codice di comportamento dell'Università della Calabria. Rectoral Decree n. 2653, 23/12/2014

University of Strathclyde Glasgow
- Information Security
- Research Code of Practice
- Research Data Policy

CPI
- IT policies for the company are set out in written policies which are subject to periodic review

FDB
- FDB has its own set of internal policies and procedures on data management.