Horizon 2020
European Union Funding
for Research & Innovation

European Research Council
Executive Agency

ATTOSTRUCTURA

Structured attosecond pulses for ultrafast nanoscience

Grant Agreement: 851201

Data Management Plan V.1.0

Deliverable 1.1

The project “Structured attosecond pulses for ultrafast nanoscience” (ATTOSTRUCTURA) has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme Grant Agreement No 851201.

VNiVERSiDAD D SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL

Abstract

About this document

- Work package in charge: WP1 Participation in the Open Research Data Pilot
- Actual delivery date for this deliverable: 31 August 2021
- Dissemination level: PU (for public use)
- Nature: Report
- Lead beneficiary: University of Salamanca (USAL)
- Autor: Carlos Hernández-García (USAL), Irene Huerta Illera (USAL)
- Contact details: Irene Huerta (i_huerta@usal.es)
- Version: 1.0
- Draft/Final: Final

Historical of changes

Version	Publication Date	Change description
1.0	August 2021	Original version

Document summary

This document describes the data management life cycle for the data that will be generated, collected, and processed during or as result of the project ATTOSTRUCTURA in all relevant work packages. This plan will be review throughout the project and updated, if necessary, in the time with the project evaluation.

DMP includes information regarding the guides, lines and normative used as reference, the information necessary for the correct identification and definition of the different datasets that will be obtained and generated during the project, the selection of the repositories and supplementary tools that will be used for the data management and the strategy for the publication of the datasets related to scientific publications or potentially commercial exploitable or patentable results.

Finally, the first dataset generated in the project is described and specific management plan strategy is defined.

This document is based on the *Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020* document version 1.1 of 21 April 2017.

Acknowledgements

The work described in this publication has received funding under the European Union's Horizon 2020 research and innovation programme (Grant Agreement number 851201).

Disclaimer

This document contains information on the ATTOSTRUCTURA core activities, findings and outcomes. Any reference to content in this document should clearly indicate the authors, source, organization and publication date.

The present report was prepared by ATTOSTRUCTURA Project coordinator USAL. The report was originally submitted to the European Commission as Project Deliverable D1.1 in August 2021.

The content of this publication is the sole responsibility of the ATTOSTRUCTURA consortium and cannot be considered to reflect the views of the European Commission.

List of abbreviatures

CC	Creative Commons
DMP	Data Management Plan
DMT	Data Management Tools
DO	Digital object
DoA	Description of Action
ERC	European Research Council
EU	European Union
FAIR	findable, accessible, interoperable and re-usable
HHG	High Harmonics Generation
OCD	Open Data Commons
PI	Principal investigator
PID	Persistent identifier
PM	Project Manager
USAL	University of Salamanca
WP	Working package

References and tools

- Open Access Guidelines for researchers funded by the ERC[1]
- Guidelines on Implementation of Open Access to Scientific Publications and Research Data (Version 1.1., April 2017)[2]

[1] https://erc.europa.eu/managing-project/open-access
[2] https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf

- 'Guidelines on FAIR Data Management in H2020 (Version 3.0, July 2016)[3]
- Open Research Data and Data Management Plans (Version 3.1, July 2019)[4]
- ERC Model Grant Agreement [5]
- Horizon 2020 Annotated Grant Agreement - ERC specific annotations to Articles 29.2 and 29.3 (pages 363 – 365)[6]
- ZENODO. Available at: http://www.zenodo.org/
- DataCite
- Sherpa Romeo. Available at: https://v2.sherpa.ac.uk/romeo/
- Sherpa Juliet. Available at: https://v2.sherpa.ac.uk/juliet/
- OpenAIRE H2020 Project. Available at: https://www.openaire.eu/
- Gredos Repository. Available at: https://gredos.usal.es/?locale-attribute=en

[3] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
[4] https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf
[5] https://ec.europa.eu/research/participants/data/ref/h2020/mga/erc/h2020-mga-erc-multi_en.pdf
[6] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf

1. Introduction

1.1. Project description

ATTOSTRUCTURA[7] (Structured attosecond pulses for ultrafast nanoscience) is a research project funded by the European Research Council (ERC) within the framework of its Excellent Science program, in the topic ERC-2019-STG - ERC Starting Grant. The project is led by PhD. Carlos Hernández-García and the University of Salamanca is the host institution.

ATTOSTRUCTURA began in March 2020 and has an expected duration of 66 months (60 + 6 months of extension), with which its completion date is by the end of August 2025.

Project studies the way in which twisted light (light exhibiting orbital angular momentum) interacts with the matter and how using high-frequency harmonics it's possible to vary the twist in time. Project focuses on three of the main challenges regarding the use of structure light on the field of the ultrafast science:

- To pioneer the generation of x-ray, sub-attosecond 3D structured light pulses.
- To disentangle the rules of light-matter interaction at the sub-fs level with structured x-ray light.
- To observe and control the magnetic properties of dielectrics at ultrafast, attosecond, timescales by using structured x-ray pulses.

For this, project defines a general objective that aims *to design of new theoretical tools to understand, develop, and propose experiments at the present and future frontiers of structured attosecond science*. This general objective is articulated in four secondary objectives through which the project and its methodology is organized.

- To push today's limits of structured light driven by HHG in atoms and molecules towards hard x-rays photon energies and sub-attosecond pulse durations.
- To pioneer the scenario of nonlinear laser-matter interaction in 2D solids driven by structured pulses.
- To explore the angular momentum dynamics in the interaction of ultrafast structured light pulses and nanoscale matter.
- To pioneer the application of structured ultrashort pulses in ultrafast nano-magnetization: towards the first proposal for the next generation experiments in attomagnetism.

[7] https://cordis.europa.eu/project/id/851201

The dissemination and availability of other researchers to access to the results obtained is something that is present in the project already in the definition of the methodology. Working package 1 (*WP1 – Towards the generation of structured x-ray sub-attosecond pulsed trough HHG atoms and molecules*) includes one fulled dedicated task, the task *WP1.T5 - To implement a multiplatform user-friendly HHG code to be freely used by the EU community*.

Task WP1.T5 aims the implementation of a free user-friendly multiplatform for the dissemination of the code and the numerical tools obtained as result of the project. The implementation and start-up will be carried out progressively, starting with the creation of a single source user-friendly HHG code, which after several phases of testing and improvement will become a tool that other laboratories and researchers can use.

1.2. Data Management Plan (DMP)

Purpose and motivation

The purpose of the Data Management Plan (DMP) is to summarize the strategy to be used for the identification, definition, protection and dissemination of the data generated within the framework of the project.

The DMP covers the complete research data life cycle. It describes the types of research data that will be generated or collected during the project, its organization in datasets, the standards that will be used, how the research data will be preserved and what parts of the datasets will be shared for verification or reuse. It also reflects the current state of the agreements on data management and must be consistent with exploitation and IPR requirements.

As part of making research data findable, accessible, interoperable and re-usable (FAIR principles), a DMP should include:

- Sufficient detailed information to describe the dataset including the scientific and the technical approach.
- the handling of research data during and after the end of the project
- what data will be collected, processed and/or generated
- which methodology and standards will be applied
- whether data will be shared/made open access and
- how data will be curated and preserved (including after the end of the project).

The DMP is not a fixed document; on the contrary it will evolve during the lifespan of the project. As project progress data generated will be changed in type and volume. On consequence, DMP will be revised, maintained, or modified as the research advanced.

This first version of the DMP includes an overview of the datasets to be produced by the project, and the specific conditions that are attached to them.

Structure and scope

Templates provided by the ERC and the guide document 'Guidelines on FAIR Data Management in H2020' will be used as reference for definition of the management conditions of each of the datasets. Different questions are defined through which the DMP and its scope are structured.

1. Administrative data
State the purpose of the data collection/generation
Explain the relation to the objectives of the project
Specify the types and formats of data generated/collected
Specify if existing data is being re-used (if any)
Specify the origin of the data
State the expected size of the data (if known)
Outline the data utility: to whom will it be useful

Table 1. Issues to be addressed regarding data summary for DMP definition

2. FAIR data	
Findable	Outline the discoverability of data (metadata provision) Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? Outline naming conventions used Outline the approach towards search keyword Outline the approach for clear versioning Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
Accessible	Specify which data will be made openly available? If some data is kept closed provide rationale for doing so Specify how the data will be made available Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Specify where the data and associated metadata, documentation and code are deposited Specify how access will be provided in case there are any restrictions

Interoperable	Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
Re-usable	Specify how the data will be licensed to permit the widest reuse possible Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why Describe data quality assurance processes Specify the length of time for which the data will remain re-usable

Table 2. Issues to be addressed regarding FAIR principles for DMP definition

3. Allocation of resources
Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
Clearly identify responsibilities for data management in your project
Describe costs and potential value of long term preservation

Table 3. Issues to be addressed regarding the allocation of resources for DMP definition

4. Data security	Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Table 4. Issues to be addressed regarding data security, ethical aspects and others for DMP definition

It is possible that first version of the DMP may not be able to answer all the questions posed. In the different revisions that the document will be carried out, an attempt will be made to complete and expand the scope of the DMP to optimize the management of the data, its dissemination and reuse.

Revision and update

DMP plan will be review and updated as minimum in time with the periodic reports of the project. Based on the scientific and financial reporting periods defined in project Grant Agreement the preliminary DMP review calendar is set.

Number of revision	Project Month	Date
V.2	M30	August 2022
V.3	M36	February 2023
V.4	M54	August 2024
V.5	M66	August 2025

Table 5. DMP review calendar

In addition, the DMP will be reviewed in the event of significant modifications, such as the appearance of a new type of data or dataset not initially contemplated, the possibility of exploiting any of the results obtained or changes in the conditions of the Grant Agreement. PI and PM will be responsible of the review and update of DMP when necessary.

1.3. Legal framework and principles for DMP definition

European Research Council

As legal basis for the definition of the data management plan, articles 29.2 and 29.3 of the Grant Agreement will be taken into consideration.

> ***29.2 Open access to scientific publications***
>
> *Each beneficiary must ensure open access (free of charge, online access for any user) to all peer-reviewed scientific publications relating to its results.*
>
> *In particular, it must:*
>
> *(a) as soon as possible and at the latest on publication, deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications;*
>
> *Moreover, the beneficiary must aim to deposit at the same time the research data needed to validate the results presented in the deposited scientific publications.*
>
> *(b) ensure open access to the deposited publication — via the repository — at the latest:*
>
> *(i) on publication, if an electronic version is available for free via the publisher, or*
>
> *(ii) within six months of publication (twelve months for publications in the social sciences and humanities) in any other case.*
>
> *(c) ensure open access — via the repository — to the bibliographic metadata that identify the deposited publication.*

> *The bibliographic metadata must be in a standard format and must include all of the following:*
>
> *- the terms ["European Union (EU)" and "Horizon 2020"]["Euratom" and Euratom research and training programme 2014-2018'];*
>
> *- the name of the action, acronym and grant number;*
>
> *- the publication date, and length of embargo period if applicable, and*
>
> *- a persistent identifier.*

Figure 1. Extract from article 29.2 of the Grant Agreement document

> ***29.3 Open access to research data***
>
> *Regarding the digital research data generated in the action ('data'), the beneficiary must:*
>
> *(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:*
>
> *(i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;*
>
> *(ii) other data, including associated metadata, as specified in the data management plan;*
>
> *(b) provide information — via the repository — about tools and instruments at the disposal of the beneficiary and necessary for validating the results (and — where possible — provide the tools and instruments it self).*
>
> *This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.*
>
> *As an exception, the beneficiary does not have to ensure open access to specific parts of its research data if the achievement of one of the action objectives, as described in Annex 1, would be jeopardized by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access.*

Figure 2. Extract from article 29.3 of the Grant Agreement document

These articles are developed in the guides generated by the European Commission and by the ERC itself. The guides used in this report are included in the References and tools section at the beginning of the document.

FAIR Principles

The concept and the problems from which the FAIR ecosystem developed emerged in 2007. It is at that moment when the first voices are heard that indicate that it is not enough to share data and other materials related to research, since without a adequate contextualization and without complementing them with adequate metadata, their discovery, interpretation and use by other researchers is limited. It appears again in 2012 in the seminal Royal Society report and in 2013 it is mentioned

in the G8 Science Ministers 'Statement and in the European Commission's first set of data guidelines for the Horizon 2020 framework program.

FAIR principles began to be developed in 2014[8] as a result of the forum that took place at the Lorentz Center in Leiden promoted by FORCE11. In this forum representatives of all stakeholders related to eScience, the dissemination and sharing of data and research results gathered to discuss the basic principles and practices necessary for the correct dissemination and exchange of data.

From these discussions, it was concluded that by defining a set of principles and practices that would serve as a guide and were agreed upon by the scientific community, it could be guaranteed and facilitated that both humans and machines could discover, access, reuse, quote and operate. data more easily. In 2016 FAIR principles were published[9].

The speed, complexity and amount of data generated today is increasing. To deal with this fact, the use of computer tools is increasingly common and necessary. That is why the FAIR principles place special emphasis on the ability of computer systems to find, access, operate and use data with them.

The FAIR principles have been embraced as a basis and reference by many organizations and communities that promote research and development. Among them is also the European Commission and the ERC as seen by the fact that they include said principles in their guides and reference documents. FAIR principles are used also as reference and base for some of the data management tools (DMT) developed as Zotero, OpenAire or such as repositories.

The FAIR principles and the FORCE11 forum in which they were generated remain open for discussion and improvement.

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the process.

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

Accessible

Once the user finds the required data, she/he/they need to know how can they be accessed, possibly including authentication and authorisation.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

[8] https://www.force11.org/fairprinciples

[9] Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles

I3. (Meta)data include qualified references to other (meta)data

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

Figure 3. FAIR principles detailed description

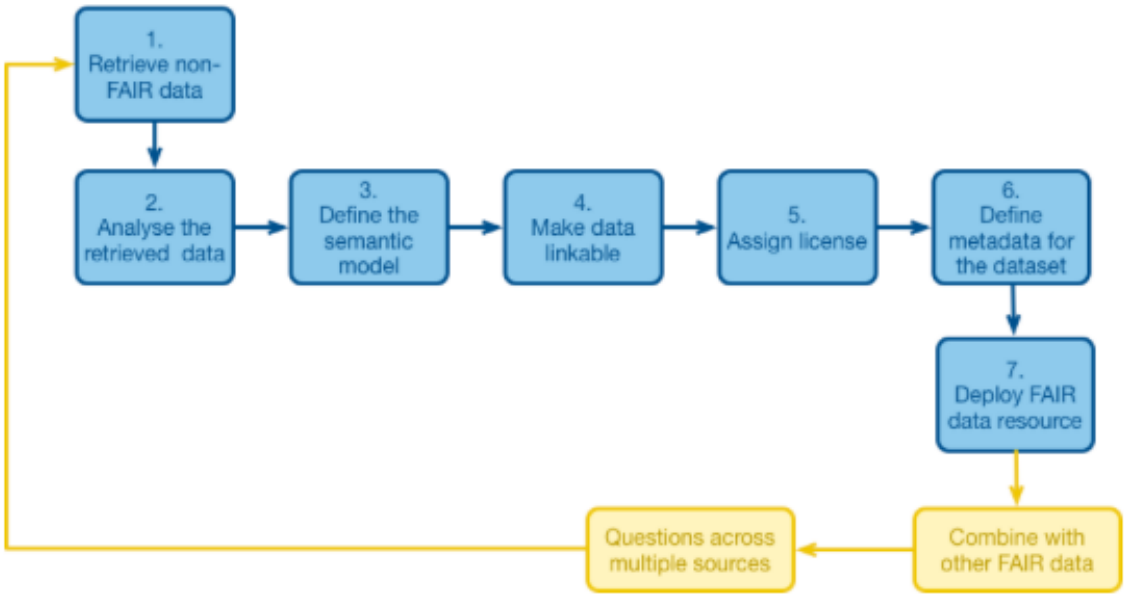


Figure 4. Conversion of a raw data set into a FAIR data group[10]

[10] https://www.go-fair.org/fair-principles/fairification-process/

FAIR and OpenAccess are different concepts. FAIR does not mean open, as data can be FAIR but shared under restrictions. However, there are greater benefits when data are open and FAIR. The European Commission formulations "as open as possible, as closed as necessary" perfectly define the differences between the concepts FAIR and OPEN, and how they can be applied.

1.4. Obtaining and defining research data and datasets

Although related to each other, the data generated in the project will be organized in different datasets. The organization of the data in different groups will facilitate its definition, management, dissemination and use. The definition of the different datasets will depend not only on the type of data that comprise it, but on the use to which that data is going to be given and the conditions under which it will be shared and preserved.

After analysing the data obtained to date in the framework of the project and estimating the data that will be generated in the future, data groups have been defined. In section Dataset definition and management, detailed information will be given on the groups of data generated and the strategies developed for their diffusion, reutilization and preservation among others.

For the management and dissemination of data to be correct and to take advantage of its full potential, it is not enough to upload and share the raw data.

Data should be complemented with additional information and represented in common – and ideally open – file formats and be richly documented using metadata standards and vocabularies adopted by the given research communities to enable interoperability and reuse. Same consideration must be taking in consideration when code is shared.

- Digital object (DO): The data, code or other research result. At the most basic level, they are made up of a binary sequence or a data stream. For it to be reused and comply with the FAIR criteria, it must be represented in a suitable format and accompanied by a persistent identifier (PID) and the corresponding metadata.
- Identifiers: Digital objects must be assigned a unique and persistent identifier (PID) such as a DOI code. This allows to establish links to the object and facilitates the citation, in addition to allowing the traceability of its use. In addition to the data, there are other elements to which identifiers must be given, such as the author of the data (ORCID), the project, the financing entities or the participating organizations.
- Standards and codes: Digital objects must be published in open and documents formats. This enables This enables others to reuse them as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code use to process and analyse the data.

- Metadata: Generation of a suitable context. In order DO to be usable and assessable it must be accompanied with metadata and documentation. Metadata will be used not just to facilitate identification and search of data but also its use and understand when, why and how were created.

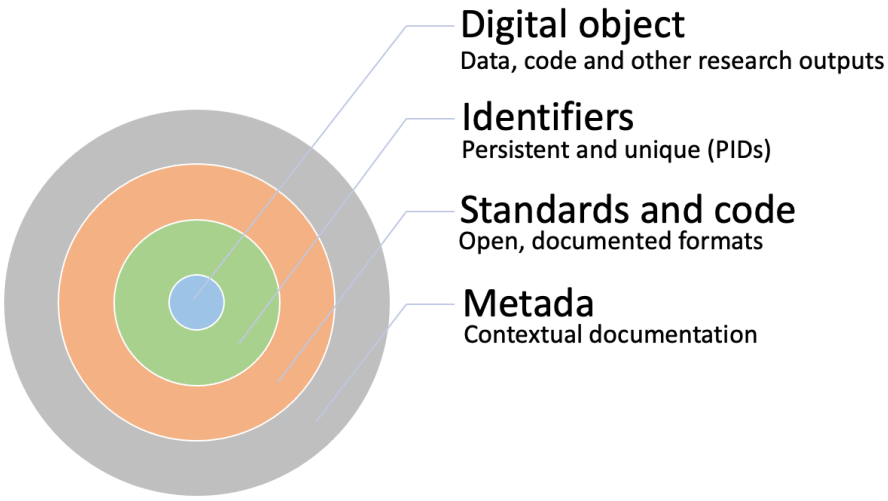


Figure 5. FAIR object structure[11]

Definition of the digital object

A correct selection and definition of the data is critical for the management and use of the data.

For each of the identified datasets, a form like the one indicated below will be completed. The right column briefly discusses, as a guide, the considerations that must be considered when defining each of the indicated aspects

Some of the aspects and / or characteristics of the data sets will have to be defined by the creator of the data or by the person responsible for its management. Others will depend on the repository that has been chosen for their dissemination and sharing or on the specific conditions of the project or organization where they have been generated.

11 TURNING FAIR INTO REALITY - Final Report and Action Plan from the European Commission Expert Group on FAIR Data. European Commission Expert Group on FAIR Data, November 2018

General information about dataset and data obtaining	
Data summary	*Summary of the data including a brief description and name. Name must be created using descriptive labels, so datasets can be easily identified even for those users not directly related to the project. Names must follow a pattern, includes date and version. Names must not include empty spaces, uncommon characters (#, &., ?, etc…) and to be limited to 32 characters.*
Purpose	*Briefly explain the objective of the data generated and / or collected and its relationship with the objectives of the project. Also indicate for whom they can be useful*
Source of data	*Indicate the origin of the data, if it is generated within the project (project objective or working package) or indicate the source from which they have been obtained if they are collected or re-used.*
Methodologies	*Indicate the methodologies used for the generation, collection, and organization of the data*
Potential Reuse	*Specify how this dataset could be reused in other context and to whom/which fields of research might be useful for.*
Type, format, volume	*Indicate, approximately, the volume of the dataset considering the volume of data in terms of storage, backups, costs, and access. Estimate data volume in MB / GB / TB and gradual growth to ensure additional storage and technical support can be provided.* *Give information regarding the type of data (numeric, textual, image, audio, video) and the format. Selection of the format must be to ensure maxima usability. Open standard formats are preferable and accepted by repositories. Also long term re-use of data is easily when standard formats are used. Examples of preferred file format choices include ODF or LaTeX or TXT, not Word; ASCII, not Excel; MPEG-4, not Quicktime; TIFF or JPEG2000, not GIF or JPG; XML or RDF, not RDBMS*
Structure	*Indicate how the data will be organized during the project, mentioning for example conventions, version control, and folder structures. Consistent, well-ordered research data will be easier to find, understand, and re-use.*

Version Handling	*Describe how the version control of files will be organized. Generally, is defined by repository policies. Repository has to be chosen to ensure that:* – *It allows easy control of the versions either through a numerical or status system and that also allows the creation of relationships between the different versions.* – *The PID must be related to the latest version* – *The latest version must be the one that appears visible, although access to previous versions can be allowed* – *The repository must keep a copy of all the uploaded versions of the same dataset*
Quality Control Processes	*Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeated samples or measurements, standardised data capture, data entry validation, peer review of data, or representation with controlled vocabularies.*
Documentation and metadata	
Metadata	*Indicate which metadata will be use to identify and discover the data.* *Indicate which metadata standards (for example DDI, TEI, EML, MARC, CMDI) will be used.* – *What contextual details (metadata) are needed to make your data meaningful?* – *What form or format will the metadata describing your data take? Which metadata standards will you use? If there is no applicable standard, how will you describe your data in a way that will make them accessible to others?* – *How will metadata files be generated for each of the data sets that you produce? Who will do the work of data description and how will the costs be borne?* – *Who on your team will be responsible for ensuring that metadata standards are followed and are correctly applied to the corresponding data sets?*
Terminology, Identifiers	*Outline the identifiability of data and refer to standard identification mechanism. Use of persistent and unique identifiers such as Digital Object Identifiers. Outline the approach towards search keywords, including:* – *The list of keywords for the dataset*

	– *The sources or references that have been used to define the keywords.* – *The protocols for reviewing and updating keywords* – *The use of standard vocabulary to allow interdisciplinary operability*
Documentation	*Consider what other documentation is needed to enable re-use. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, and so on.* *Consider how this information will be captured and where it will be recorded (for example in a database with links to each item, a 'readme' text file, file headers, code books, or lab notebooks)*
Storage, preservation and safety	
Storage	*Describe where/how the data will be stored*
Data Back-Up	*Described how the data will be backed up during research activities and how often the backup will be performed. It is recommended to store data in least at two separate locations.* *Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution. Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended to be the unique option*
Data Recovery	*Explain how the data will be recovered in the event of an incident.*
Risk Management	*Consider data protection, particularly if your data is sensitive for example containing personal data, politically sensitive information, or trade secrets. Describe the main risks and how these will be managed.*
Data Access	*Explain who will have access to the data during the research and how access to data is controlled, especially in collaborative partnerships.*
Preservation Life	*Specify the length of time for which the data will remain re-usable. Consider who decides what data will be kept and for how long.*
Public data diffusion	

Findability	*Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?*
Availability	*Specify how the data will be made available. Indicates the repository or equivalent system where data will be deposited and the repository conditions.*
Interoperability	*Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.*
Procedures	*Specify what methods or software tools are needed to access the data. Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?*
Restrictions	*Consider strategies to minimise restrictions on sharing. These may include anonymising or aggregating data, gaining participant consent for data sharing, gaining copyright permissions, and agreeing a limited embargo period.*
Citation	*Describe the citation format to be used for the dataset and the minimum information it must contain (creator (s) or contributor (s); date of publication; title of dataset; publisher; identifier or URL; version; date accessed). Check if there is a preferable format.*
Ethic and legal requirements	
Personal data issues	*Consider whether there are any ethical or legal personal data issues than can have an impact on data sharing.*
Intellectual Property	*Specify if some considerations regarding the intellectual property (copyright, patents, trademarks or trade secrets) are applicable to the dataset*
Reuse License	*Specify how the data will be licensed to permit the widest reuse possible.*
Third-Party Data Restrictions	*Specify whether the data produced and/or used in the project is useable by third parties, after the end of the project. If the re-use of some data is restricted, explain why.*
Embargoes	*Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is required.*

	Reasons for embargoes may include time to publish or seek patents. If an embargo is sought, specify why and for how long, bearing in mind that research data should be made available as soon as possible. Research funders expect timely release. They typically allow embargoes but not prolonged exclusive use
Resources and responsibilities	
Human Resources	*Indicate, if they exist, the human resources that will be allocated to data management. Both short-term (for the duration of the project) and long-term.*
IT Resources	*Indicate, if they exist, the technological resources that will be allocated to data management. Both short-term (for the duration of the project) and long-term.*
Economic Resources	*Indicate, if they exist, the economic resources that will be allocated to data management. Both short-term (for the duration of the project) and long-term (after project termination). If possible estimate the cost of the dataset management.*
Responsibility	*Outline the roles and responsibilities for all activities e.g. data capture, metadata production, data quality, storage and backup, data archiving & data sharing. Consider who will be responsible for ensuring relevant policies will be respected. Individuals should be named where possible.*

1.5. Data management responsibilities

Different groups of data will be generated because of investigations, experiments, theoretical models, and simulations. Datasets must be created, managed and stored appropriately and in line with applicable legislation.

The ultimate responsibility for sharing and managing the datasets rests on the principal investigator (PI) and the project manager (PM). Some of those responsibilities include:

- To ensure dataset are shared though an adequate platform easily accessible by any user who wishes to consult or use them and respecting the applicable legislation.
- To ensure dataset integrity and compatibility for its use during and beyond the project lifetime.
- Manage the different versions of the databases that are generated. Ensuring that the latest version generated is available and that there is no confusion between the different versions published.
- Keep the databases updated by adding the new data generated or correcting errors and mistakes if necessary.
- To perform necessary backups to ensure data are secure and that it will not be lost in the event of a system failure.

Quality control, validation dataset before publication and definition of the metadata associated to each dataset will be responsibility of the investigator that generates the data. Definition of metadata is a critical step since metadata constitutes an underlying definition and description of the data and facilitate its finding and use.

In general, as a step prior to the publication of a group of data, the PI will consult with the researchers participating in the project to analyse the potential for exploitation of said data or the results associated with them and decide the suitability of its publication.

1.6. Selected repositories and tools used

The transformation of raw data into data that complies with the FAIR principles requires the use of different tools and applications. This section describes the tools and applications that will be used for the management of the data generated in the project and its publication in Open Access.

Repositories

The publication of data in OpenAccess repositories is one of the objectives of data management. For the correct storage and publication of the data and for its dissemination to have the greatest scope, it is necessary to choose a suitable

repository. The selection should be made not only in relation to its theme and target community but also in relation to its long-term sustainability and the way in which the data is stored and referenced.

A repository must guarantee:

- data are stored securely
- data are easily found, accessible and reusable
- data are described in a standard way, allowing the use of metadata
- to specify a license to manage access and reuse

There is many repositories, free or paid, private or financed by public institutions, where the data generated in the project can be shared in Open Access. As the number of repositories increases, so does the difficulty in choosing the most appropriate one for the data to be published and to ensure that the repository meets the conditions and requirements. To ensure that the appropriate repository is chosen, different organizations have developed tools where to certify the characteristics and conditions of the repositories and where to find specific repositories focused on specific research fields.

The following tools and websites have been used to choose the repositories. In this way, their validity, the conditions of use and their dedication as generalist or specific repositories within a certain scientific field have been verified.

- Core Trust Seal[12]
- Re3data[13]
- OpenDOAR[14]

Initially **Gredos** (https://gredos.usal.es), the institutional repository of the University of Salamanca, was chosen as repository for the diffusion of the research data. It allows the dissemination and sharing of scientific articles, academic documents, and research data. In the case of research data, Gredos adheres to the Berlin Declaration[15] and therefore undertakes to ensure the integrity and preservation of the data uploaded to the repository. To this end, a series of measures have been developed to ensure its accessibility, legibility and future use, including making backup copies, updating software and hardware, migrating to safer formats, etc ... Gredos accepts formats open or world-popular formats. Regarding the use of metadata, Gredos follows the DublinCore scheme.

[12] https://www.coretrustseal.org/certified-repositories/
[13] https://www.re3data.org
[14] https://v2.sherpa.ac.uk/opendoar/
[15] https://openaccess.mpg.de/Berlin-Declaration

Although Gredos is the official repository of the University of Salamanca and is supported by the entity, its scope is not large enough to ensure maximum dissemination of the project data. That is why other repositories have been selected.

Through Re3data, different repositories have been identified. For the selection, both the scientific field in which they are focused has been considered, as well as the fact that they use persistent identifiers to identify each group of data and that they are certified repositories.

A repository has not been found that meets all the desired requirements (PID, certificate, specific to the field of study and with global diffusion). From the available repositories these have been selected.

	FigShare	DIGITAL.CSIC	Zenodo
Subject	Multidisciplinary	Multidisciplinary	Multidisciplinary
Certificates and standards	None	Data Seal of Approval	None
PID	DOI / ORCID	Hdl DOI	DOI / ORCID
Metadata standards	DataCite Metadata Schema Dublin Core	Dublin Core	DataCite Metadata Schema Dublin Core
Institution	Commercial	Spanish Government, Ministry of Economy and Competitiveness Spanish National Research Council	European Commission CERN OpenAIRE
License	CC0	CC OCD	CC0
Country	International	Spain	European Union
Link	https://figshare.com	https://digital.csic.es	https://zenodo.org

Table 6. Summary of the characteristics of the selected repositories

As stated previously, it has not been possible to find a specific repository in the project field that also meets the desired characteristics in relation to certification and the use of PID and metadata. Of the three multidisciplinary repositories selected, the one with the greatest participation in fields related to the scope of the project will be chosen. To do this, the number of records available in each repository associated with a group of keywords will be evaluated. The chosen keywords were: attosecond pulse, photonic, high harmonic generation, ultrafast laser, structured light and magnetic field.

Keywords/Repository	Figshare	CSIC	Zenodo
Attosecond pulse	195	49	13
Photonic(s)	31,987	5,956	845
High harmonic generation	1,565	34	10
Ultrafast laser	1,197	77	21
Structured light	44,723	57	50
Magnetic fields	20,914	3,300	589

Table 7. Number of records found for the different keywords used.

After analyzing the results and characteristics of the selected repositories, it has been decided to use the Figshare repository. It is assumed that this repository is where a greater scope will be achieved in the dissemination of project data.

Metadata and vocabulary

Metadata specifications and standards are essential to data interoperability and reuse. A metadata record is a file that captures all details about a data set that another researcher would need to make use of the data set in a separate or related line of inquiry. Metadata captures the who, what, when, where, why, and how of the data you produce. There are three different types of metadata.

- Descriptive metadata describes and identifying information resources at the local (system) level to enable searching and retrieving (e.g., searching an image collection to find paintings of animals) and at the Web-level to enable users to discover resources (e.g., search the Web to find digitized collections of poetry). Sample elements: unique identifiers, physical attributes (media, dimensions), bibliographic attributes (title, author/creator, language, keyworks)
- Structural metadata facilitates navigation and presentation of electronic resources. Structural metadata provides information about the internal structure of the resources including page, section, chapter, index, and table of contents, describes the relationship among different datasets or blinds related files.
- Administrative metadata facilitates short- and long-term management and procession of the digital datasets. Administrative metadata includes technical information regarding the creation and the quality control, the preservation actions, the management rights and the requirements for use and access.

Due to its nature and purpose, it is important to use standardized metadata. In this way, the interoperability of the data is facilitated and that other users locate them.

There are different metadata schemes, some of a general type and others developed specifically for a scientific field or a specific type of data.

A metadata schema related to the project fields was searched, but none was found[16]. On the other hand, it is logical that if a repository is to be used for the storage and dissemination of data groups, the metadata schema imposed by the repository should be used.

Selected repositories use metadata schemas:

- DataCite Metadata Scheme[17]. The DataCite Metadata Schema is a list of core metadata properties chosen for an accurate and consistent identification of a resource for citation and retrieval purposes, along with recommended use instructions. Last version (Metadata Schema 4.4) was released on March 30th 2021.

  This metadata schema includes 20 properties, some required and some optional or recommended. The mandatory properties are identifier (with mandatory type sub-property); creator (with optional given name, family name, name identifier and affiliation sub-properties); title (with optional type sub-properties); publisher; publication year; resource type (with mandatory general type description sub- property)[18].

  In relation to the citation style, DataCite is used in different fields so it follows a generalist citation style valid for different disciplines

  Creator (PublicationYear): Title. Publisher. (resourceTypeGeneral). Identifier

- Dublin Core Metadata Scheme[19] is a basic, domain-agnostic standard which can be easily understood and implemented. It is one of the best known and most widely used metadata standards. The original metadata schema was formalized in 1998 in the Internet Engineering Task Force standard RFC 5791. It is currently published in other standards: RFC 5791 (2010), Z39-85-2012, and ISO 15836-1: 2017

  This metadata includes fifteen generic, widely used elements -- Creator, Contributor, Publisher, Title, Date, Language, Format, Subject, Description, Identifier, Relation, Source, Type, Coverage, and Rights[20]. Last version was updated in 2020 January.

License selection and definition

[16] http://rd-alliance.github.io/metadata-directory/standards/
[17] http://schema.datacite.org
[18] DataCite Metadata Working Group. (2021). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. https://doi.org/10.14454/3w3z-sa82
[19] https://www.dcc.ac.uk/resources/metadata-standards/dublin-core
[20] https://www.dublincore.org/specifications/dublin-core/

In order for other researchers to use the data, it is necessary to define the legal context in which they can use it. The license is the legal language that describes the way in which the data can be used or modified by others than the people who own it

Although it is possible to write your own license to regulate the conditions of use and exploitation of the data, the most common is to use one of the standard licenses that exist. There are two main organizations that write the language for the licenses: Creative Commons (CC) and the Open Data Commons (ODC).

In general, three types of licenses can be distinguished:

- *Attribution.* This license lets others distribute, change, and build upon your work, even commercially, as long as they credit you for the original creation. CC-BY and DC-By licenses are this type.
- *Attribution - Share Alike*. You let others copy, distribute, display, and modify your work, as long as they distribute any modified work on the same terms. If they want to distribute modified works under other terms, they must get your permission first. The CC-BY-SA and ODC-ODbL licenses are this type.
- *Public Domain*. All rights worldwide to this work are waived by the author. Others may copy, modify, and distribute the work, even for commercial purposes, all without asking permission. The CC0 and ODC-PPDL licenses are this type.

To facilitate the selection of the most appropriate license for each type of data, it is possible to use wizards or guided tutorials. Two of the best known are Creative Commons[21] and EUDAT[22]. For the selection of the license, the tool developed by EUDAT will be used as it includes the Public Domain licenses and allows differentiating between data and software.

[21] https://creativecommons.org/choose/
[22] https://ufal.github.io/public-license-selector/

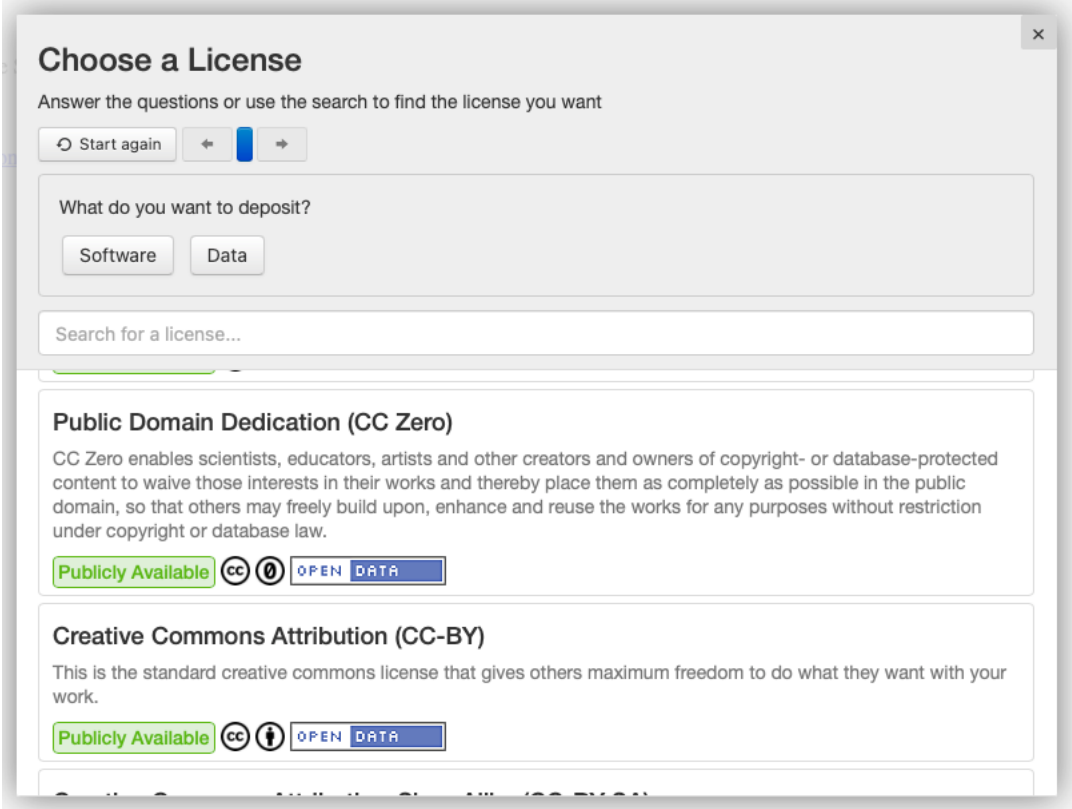


Figure 6. Home screen of the license selector developed by EUDAT

Whenever possible, the license selector will be used to identify which of the available licenses is the most appropriate for each group of data. However, it must be borne in mind that repositories have their own policies regarding which licenses to use and that they may force the use of a specific one.

It is not enough to select a license, but it is also necessary to ensure that whoever uses this data is aware of the conditions under which they can use it[23]. For this it is advisable to include information regarding the license in the file. In a simple way, it is enough to add a small paragraph to the file indicating what type of license is applied, however it is recommended that this information be easily readable by machines and automatic systems. For this, it is recommended that the license information be specified in the characteristics of the file (for example in the "Properties" section of Excel files) or in an introductory file in the case of packed data sets. It is also recommended that the license information is clearly visible on the page from which it will be downloaded.

Other tools

SHERPA

[23] Ball, A. (2014). 'How to License Research Data'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/how- guides

The tools grouped in the SHERPA application (https://www.jisc.ac.uk/sherpa), promoted and supported by Jsic, collect and associate information related to the open access publication of data and articles. It also allows to compare them to analyze and define the most appropriate publication strategy based on the requirements of the financing entity.

- SherpaRomeo allows to know in a simple way the conditions that must be met to be able to share in open access a publication from a scientific journal. It includes information regarding the embargo time that is applied, the type of document that can be shared (accepted, manuscript, etc ...), the owner of the rights of the publication, the license and the type of repository, web or base broadcast where the document can be shared.
- SherpaJuliet provides information regarding the funder's conditions for open access publications.
- OpenDOAR enables the identification of repositories and the collection of information about the policies about data, metadata, content and preservation.

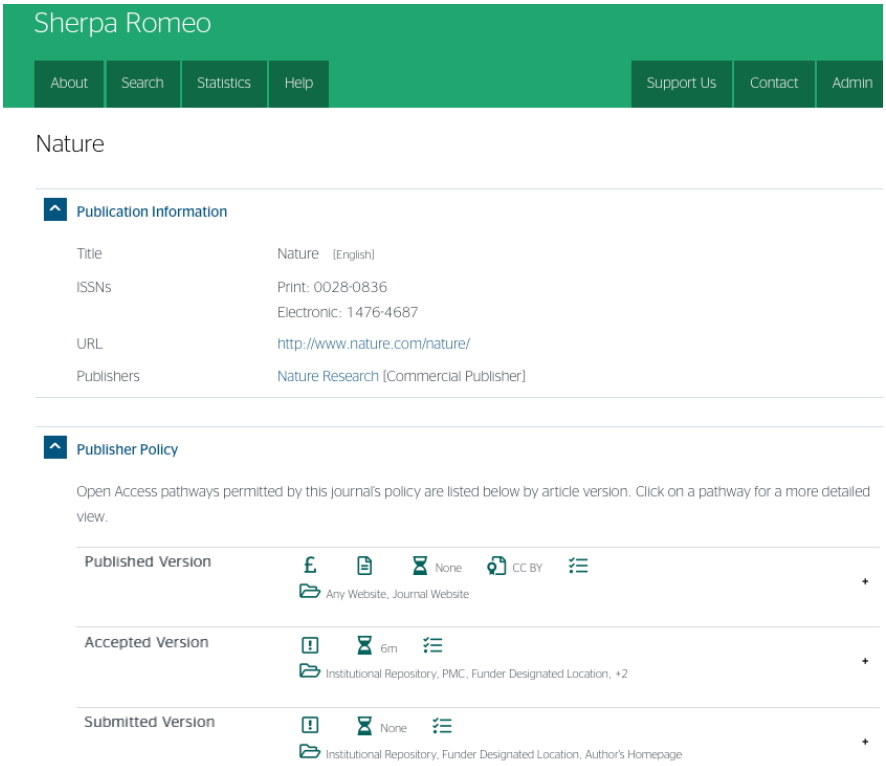


Figure 7. Example of information provided by SherpaRomeo regarding OpenAccess publishing policies

Sherpa tools are databases that collect information published by journals, publishers and funding institutions respectively. The information provided by these tools should be used only as an informative means and must be confirmed by consulting the information provided directly by the journal, publisher or funding institution.

2. Data associated to scientific publications or exploitable results

For those data associated to scientific publications or potentially exploitable results, specific procedures have to been defined in order to ensure correct dissemination and protection.

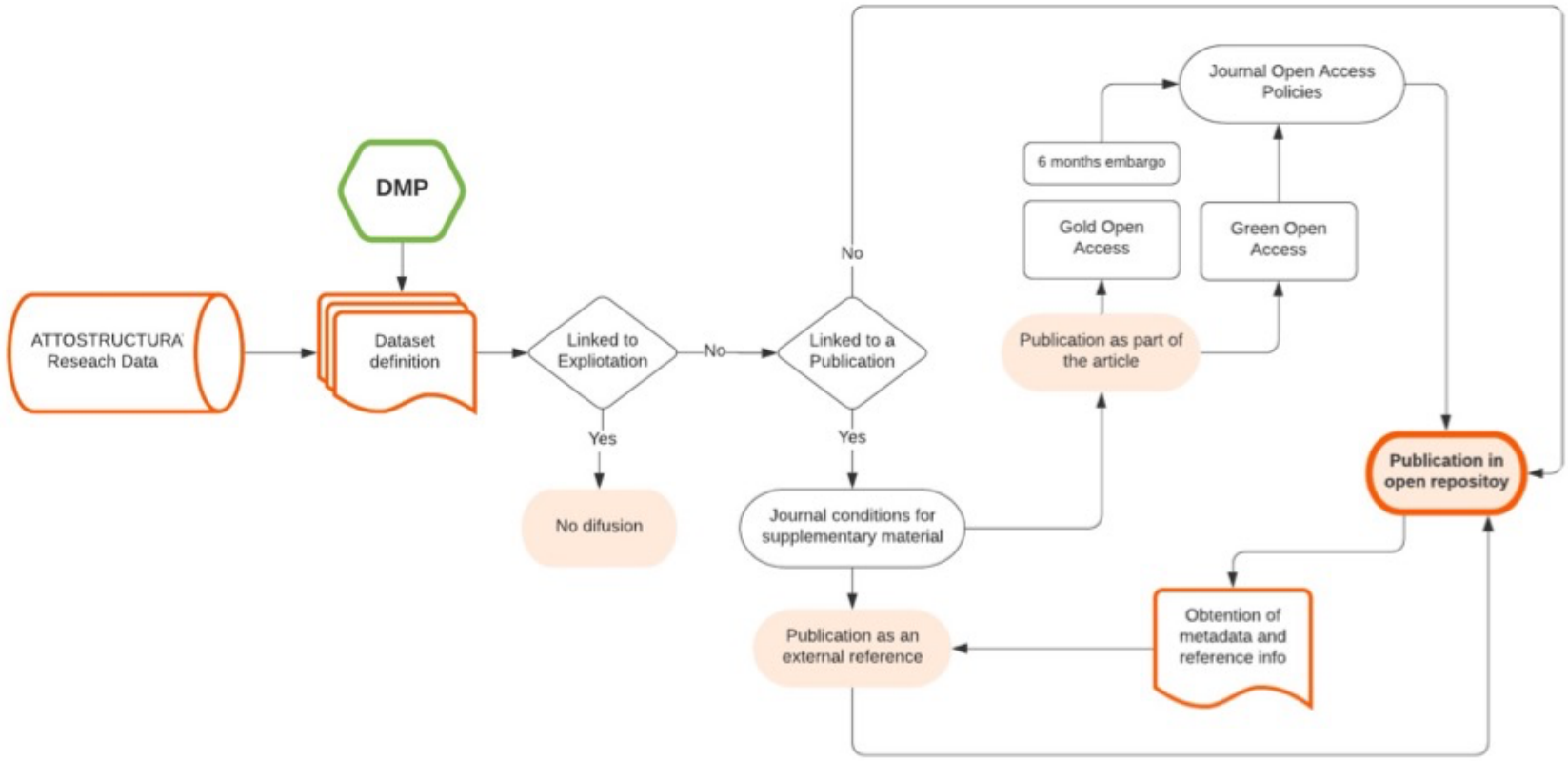


Figure 8. Decision scheme for the publication in Open Access of the data.

2.1. Exploitable results

As general method, potentially exploitation of data and the results will be evaluated by the researchers participating in the project and the host institution before its publication and diffusion. If exploitation is possible, dataset associated to these results will be no published until exploitation is solved.

2.2. Scientific publications

Peer review publications are expected as result of the project. According to DoA two publications per year are planned to be published in Open Access Journals directly related with Optic, Multidisciplinary Sciences or Physis fields.

Publications obtained within ATTOSTRUCTURA project will be Open Access published following the *29.2 Article of Grant Agreement guidelines*. Wherever possible, publication fees will be paid to facilitate open access to publications and reduce embargo time.

In any case, as soon as possible and always within 6 months after its publication, machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication will be upload into an open repository for scientific publications.

Scientific publications from ATTOSTRUCTURA project will be published in University of Salamanca's repository Gredos and OpenAire repository. Open access publication may be delayed for a few months due to embargo conditions imposed by journals and publishers. Conditions defined for each journal or publisher for the publication of open access repositories will be check in Sherpa Romeo web page to ensure that legal conditions and requirements are fulfilled.

Changes in the policies of scientific journals and in the paradigm of the dissemination of scientific articles makes it increasingly frequent that the data on which these publications are based are published together with the articles. In this way, the authors allow other researchers to analyse and use the original data to validate the results shown in the article or to develop their own hypotheses.

Magazines and editorials are promoting in an increasingly active way the publication and dissemination of the basic data from which the published results are obtained. Many of them include in their publication conditions the obligation to include them as supplementary material or to refer the data from a public access repository.

Datasets used to obtain the results presented in articles and necessary to validate them will be upload to repositories so they could be totally available. Datasets used for each publication will be uploaded independently, in individual datasets, well differentiated from other similar data so that it is uneasily known which data has been used in the preparation of the article.

Submission to repositories will be done according to the statements of *article 29.3 of ATTOSTRUCTURA Grant Agreement* and recommendations of the guidelines referenced in section *References and tools* of this document. In addition to the basic bibliographic information of the publication all available metadata that facilitate identification will be included in the open access repository. As much as the chosen repository allows it, the information that will be added will be the following:

- EU funding acknowledgement:

  Contributor: "European Union (EU)", "Horizon 2020", “European Research Council”

- Peer Reviewed type (e.g. accepted manuscript; published version).
- Embargo Period (if applicable):

  End date.

  Access mode.

- Project Information:

  Grant number: “851201”

  Name of the action: “ERC Starting Grant”

  Project Acronym: “ATTOSTRUCTURA”

  Project Name: “Structured attosecond pulses for ultrafast nanoscience”

- Publication Date.
- Persistent Identifier:

  Authors and Contributors. Wherever possible identifiers should be unique, non-proprietary, open and interoperable (e.g. ORCID number, D.O.I code, etc…)

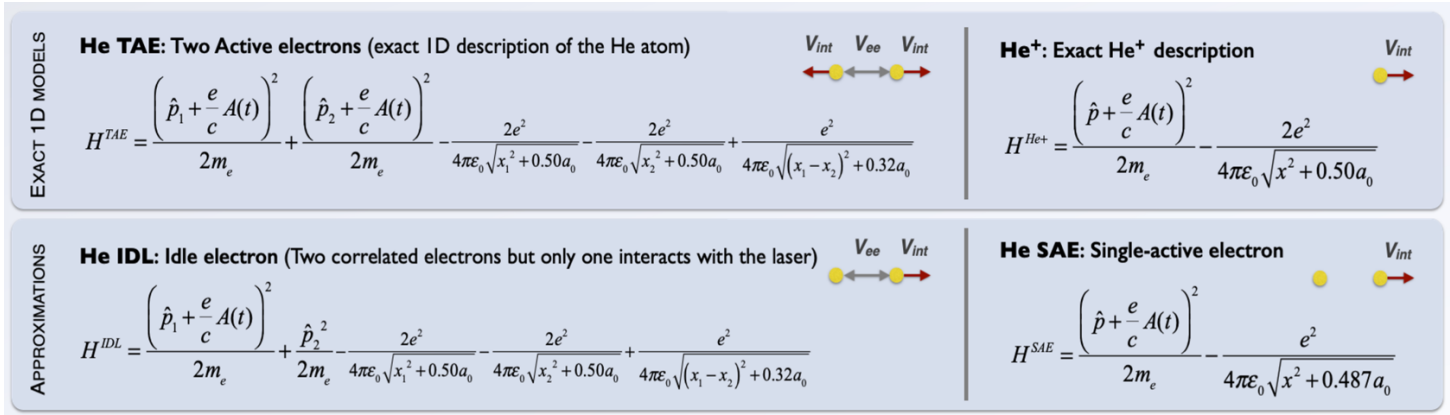
3. Dataset definition and management

3.1. Project general information

Project basic information for DMP contextualization	
Project Name	Structured attosecond pulses for ultrafast nanoscience
Project Acronym	ATTOSTRUCTURA
Funder	European Reseach Council (ERC)
Grant Reference Number	851201
Principal Investigator (ID)	Carlos Hernández – García
Host Institution	University of Salamanca
Project data contact	carloshergar@usal.es
DMP first version	August 2021 (V1.0)
DMP last update	'--

3.2. Datadocument_PRR2020

General information about dataset and data obtaining	
Data summary	We provide results of the numerical integration of the one-dimensional Time-Dependent Schrödinger Equation (TDSE) of the helium atom in the interaction with a strong electromagnetic field. We compare different theoretical models: (i) Two Active Electrons (TAE), which is the exact 1D description of the He atom, (ii) Idle Electron model (IDL), considering two-electron correlation, and only one electron interacting with the external field, (iii) Single-Active Electron approximation (SAE), and (iv) the exact 1D description of the helium cation (He^+). The methodology and data interpretation are publicly available in [1] A. de las Heras, C. Hernández-García, and L. Plaja, "Spectral signature of back reaction in correlated electron dynamics in intense electromagnetic fields", Phys. Rev. Res. **2**, 033047 (2020). Concretely this dataset contains: - High-order harmonic spectra for different wavelengths (400 nm-1030 nm) and peak intensities ($(1.6 - 10) \times 10^{14}$ W/cm^2) of the driving laser field. - Excited atomic transitions after the laser pulse (515 nm wavelength and 1.6×10^{14} W/cm^2 peak intensity) in the TAE and IDL descriptions. - Energy levels of the He atom (TAE) and the cation He^+. - Wavefunction of the second excited state of He. One-electron energy dependence on the position of a secondary electron.
Purpose	The objective of the generated data is to gain further insights about multi-electron correlations in laser-matter interactions by analysing the helium atom as one of the simplest multi-electron systems, where we can isolate elemental correlation mechanisms. This numerical data can be used to compare with future experimental results or other numerical calculations.
Source of data	This data was obtained in the numerical integration of the TDSE, and the diagonalization of the time-independent Hamiltonian of He and He^+ in a one-dimension description.

Methodologies	The computational method to solve the TDSE is based on the Crank-Nicolson method and a split-operator approach. We find the initial state via imaginary time propagation, and then we propagate it in real time with a time step $\Delta t = 0.02$ a. u. in a spatial grid with spacing $\Delta x_1 = \Delta x_2 = 0.16$ a. u. and box size $\lvert x_{1max} \rvert = \lvert x_{2max} \rvert = 200$ a. u. We implement a cos1/8 mask function at the limits of the spatial box (15% of the grid size) as an absorbing boundary. HHG spectra are computed as the squared modulus of the Fourier transform of the dipole acceleration. We apply a cos mask function at final times of the computed dipole acceleration (after the end of the laser pulse) to improve the signal-noise ratio in the HHG spectra, which is influenced by a non-smooth ending in the dipole acceleration. The driving field is modelled with a trapezoidal envelope (1 cycle of ramp-up, constant amplitude, 1 cycle ramp-down) or a sin2 envelope. The following figure shows the Hamiltonian in the four theoretical models:  For further details see ref. [1] A. de las Heras, C. Hernández-García, and L. Plaja, "Spectral signature of back reaction in correlated electron dynamics in intense electromagnetic fields", Phys. Rev. Res. 2, 033047 (2020).
Potential Reuse	This data can be useful for experimental or theoretical researchers studying multielectron effects or electron-electron correlation in the helium atom.
Type, format, volume	All the dataset is compressed in a .zip file of 29.2 MB containing the data in .txt format

Structure	The data is organized in .txt files with the following structure: "Dataset_515nm_E0067_8cy_trap.txt" and "Dataset_515nm_E0067_8cy_sin2.txt" contain the high-order harmonic spectra in all the models calculated for a 8 cycle trapezoidal or sin2 envelope, 515 nm wavelength and $1.6 \times 10^{14}\ \mathrm{W/cm^2}$ peak intensity. This data corresponds to figures 1-3,7 in ref. [1]. The file includes a first line of headers indicating the data in each column: (c1) Frequency in atomic units for all the models except SAE, (c2) TAE spectra, (c3) Contribution of neutral He to the TAE spectra, (c4) Contribution of cation He+ to the TAE spectra, (c5) He+ exact spectra, (c6) IDL spectra, (c7) Frequency in atomic units for SAE, (c8) SAE spectra. "Dataset_515nm_E012_8cy_trap.txt", "Dataset_515nm_E017_8cy_trap", "Dataset_400nm_E009_8cy_trap.txt", "Dataset_515nm_E009_8cy_trap.txt", "Dataset_800nm_E009_8cy_trap.txt", "Dataset_1030nm_E009_8cy_trap.txt" contain the high harmonic spectra in all models except IDL. This data corresponds to Figs. 5-6 in ref. [1]. The filenames indicate the wavelength in nm and the amplitude in atomic units. The file is structured in columns: (c1) Frequency in atomic units for all the models except SAE, (c2) TAE spectra, (c3) Contribution of neutral He to the TAE spectra, (c4) Contribution of cation He+ to the TAE spectra, (c5) He+ exact spectra, (c6) Frequency in atomic units for SAE, (7) SAE spectra. "Dataset_EnergyLevels_HeTAE.txt" and "DatasetEnergyLevels_HeION.txt" contain the energy levels obtained in the diagonalization of He TAE or He+ Hamiltonian (without laser field) in the 1D description. This data is displayed in Fig. 4(b) in ref. [1]. "Dataset_ExcitedState_HeTAE.txt" contains the 2D array of the 2nd excited state of He, displayed in the inset of Fig. 4(c) in ref. [1]. "Dataset_EnergyDependenceOnElectronPosition.txt" contains the one-electron energy as a function of the position of a secondary electron. It is calculated according to eq. (3) in ref. [1] and shown as a blue line in Fig. 4(c) in ref. [1].

Version Handling	Both repositories have different policies regarding the version handling. Gredos allows you to modify or update the data and metadata once they have been uploaded to the repository. To do so, it is necessary to contact the site administrator since it is an action that the user cannot perform. Figshare does allow the user to manage the different versions of the datasets and update them. Depending on the update that is carried out, according to Figshare's version management policy, a new version of the dataset will be created and / or the changes will be applied to the existing versions. Each of the versions have their own DOI and can be cited independently. The version DOI is created from the base DOI (last version dataset) by adding the v <x> suffix, where x is the version number.
Quality Control Processes	The researchers who have generated the data will be responsible for ensuring their quality and consistency before they are uploaded to the repository. To do this, they will use the tools they consider most appropriate based on the nature and quantity of the data generated. As in most cases the data will be associated with a publication in a scientific journal, the quality and revision of the data will be linked to the revision processes of the corresponding publication.
Documentation and metadata	
Metadata	Figshare metadata: data name, author with unique identifier (ORCID), dataset description, funding, categories, key words, item type, reference and licence. Gredos metadata: to be defined once submission at the repository was complete
Terminology, Identifiers	Keywords: Atomic Physics, Nonlinear Optics, Computational Physics, Strong-Field phenomena, High-order harmonic generation, electron-electron correlation, Time-Dependent Schrödinger Equation, Helium atom
Documentation	The file “readme.txt” summarizes the important information to re-use this data. The article [1] A. de las Heras, C. Hernández-García, and L. Plaja, "Spectral signature of back reaction in correlated electron dynamics in intense electromagnetic fields", Phys. Rev. Res. **2**, 033047 (2020) can also be used to further contextualize this data.

Storage, preservation and safety	
Storage	The data will be stored in two public open access repositories. In addition, copies of the data will also be kept on the servers of the researchers involved in the project.
Data Back-Up	Before uploading the data to the repository, and once they are consolidated and reviewed, a reference copy will be saved that will serve as a backup on the project's own servers and storage disks. Both Figshare and Gredos have strategies and tools to periodically generate backups of data and guarantee its security. In the case of Figshare, automatic backups of all data files and metadata are performed nightly, directly to Figshare-managed Amazon Web Service (AWS) S3 buckets. AWS S3 versioning allows for the immediate restore of content in case of deletion of a file.
Data Recovery	The selected repositories will use the generated backups to recover the data in the event of an accident. In the event that it is not possible to recover the data through the backups of the repositories themselves, the project's own backups may be used.
Risk Management	The data that make up this dataset does not include sensitive information (Personal data or industrial secrets) therefore it is not necessary to define any protection policy.
Data Access	Both Gredos and Figshare allow free access to any interested person without the need for registration or subscription.
Preservation Life	In both cases, the two selected repositories (Figshare and Gredos) include among their policy of use the commitment to preserve the data in the long term. Figshare stores the data on Amazon Web Services S3 storage, but also be deposited into Chronopolis for further preservation. Chronopolis is a digital preservation service based out of The University of California at San Diego. Gredos, on the other hand, includes tools and strategies to guarantee long-term data preservation.
Public data diffusion	
Findability	Digital Object Identifiers will be used. Figshare: https://doi.org/10.6084/m9.figshare.15188049.v1

	Gredos: dataset must be review and accepted in the repository. DOI will be included in next document review.
Availability	Data will be available at two public, full open repositories Figshare and Gredos.
Interoperability	For Figshare repository metadata schemes selected are: DataCite Metadata Schema and Dublin Core, Gredos repository uses DublinCore scheme.
Procedures	No specific methods or software are necessary to access the data.
Restrictions	The data does not have any restrictions on use or in relation to privacy or participation.
Citation	Figshare: de las Heras, Alba; Plaja, Luis; Hernández-García, Carlos (2021): Datadocument_PRR2020. figshare. Dataset. https://doi.org/10.6084/m9.figshare.15188049.v1 Gredos: dataset has to be review and accepted in the repository. Citation will be included in next document review.
Ethic and legal requirements	
Personal data issues	There is not ethical or legal personal data issues than can have an impact on data sharing.
Intellectual Property	No considerations regarding the intellectual property (copyright, patents, trademarks or trade secrets) are applicable to the dataset
Reuse License	Dataset will be licence following a CC by 4.0 licence attribution.
Third-Party Data Restrictions	No restrictions are applied
Embargoes	No embargoes are applicable
Resources and responsibilities	

Human Resources	During the duration of the project, the project's own human resources will be used. In the long term, if it is necessary host institution resources will be used.
IT Resources	During the duration of the project, the project's own technological resources will be used. In the long term, if it is necessary host institution resources will be used.
Economic Resources	During the duration of the project, the project's own economic resources will be used. In the long term, if it is necessary host institution resources will be used.
Responsibility	Roles and responsibilities regarding the data management have been already discussed in section 1.5 of this document.

List of Tables

List of figures