



DARIAH Winter School in Prague

Open Data Citation for Social Sciences and Humanities

24th to 28 of October 2016

Session 3: Data Management Plan

3-Data Management Plan	3
What is research data?	3
A new model of openness for research data	3
Supported by European and national initiatives	4
Direct benefits for researchers	5
Why manage data?	6
Research data management	7
Definition	7
Research data lifecycle	7
Research data management services	7
Creating Data Management Plans (or DMPs)	8
FAIR Data	9
H2020 framework requirements	9
Data Management Plans : How-to guide	11
Components of a DMP	11
Crucial points to address	11
Responsibility	11
Results management	12
Data Collection	12
Datasets management	12
Description and metadata	12
Documentation and Metadata	12
Example: DataCite metadata standard	13
Formats	13
Backup plan	13
Selection and Preservation	14
Data repositories	14
Data access and sharing	15
Share research data with a data paper	15
Publish your sharing policy with the Data Reuse Charter	15
Conclusion	16
A research good practice	16
The DMP Aide-mémoire	17
Appendix : DMP tools	17
DMPonline Exercise	17
Contact	18

3-Data Management Plan

Marie Puren & Charles Riondet, INRIA

Data management can offer many advantages, like higher quality data, increased visibility and better citation rate. In this approach, research data is an asset and a resource that can be shared with mutual benefits for the person who share the data and the person who collect the data. Yet, the Open Science movement implies radical changes for many researchers.

What is research data?

We can find a simple definition on the [website of the Boston University Libraries](#): “Data are distinct pieces of information, usually formatted in a special way”. But it is difficult to clearly define “research data”, because research data is challenging:

- there is no consensus on the definition;
- it varies according the discipline;
- it differs according to the research funder.

The [University of Bristol](#) define “research data” as follows: “Research data is created as a direct result of ‘doing research’. It excludes teaching materials and administrative documents (such as job descriptions, emails or financial reports). Research data comes in an endless variety of formats”. For the Boston University Libraries, research data is “data that is collected, observed, or created, for purposes of analysis to produce original research results”. These data can be: observational, experimental, generated from test models (simulation), derived or compiled (like text and data mining), reference or canonical (for instance, gene sequence data banks). Therefore, research data can adopt multiple forms like: text or Word documents, spreadsheets, laboratory notebooks, questionnaires, videotapes, photographs, slides, samples, databases, methodologies, output for analysis software, standards, etc.

According to these definitions, “research data” could be defined as:

- data that help to do research;
- data that could be collected, created and analysed;
- data that come in multiple formats.

The term "dataset" is used to describe a collection of research data: “A digital dataset might comprise a single element [...] [or] a collection of related elements” ([Oxford Research Data Website](#)). Thus, a dataset is a compilation of research data. It could gather together data in a single document - like in a CSV (Comma-Separated-Values) for instance - or a series of data.

A new model of openness for research data

The movement for Open Science promotes a new model of openness, with an important impact on research data - in particular on data sharing. The main aspects of openness are: availability and access, reuse and redistribution, and universal participation. This new model

tries to gradually replace traditional ways of thinking in the international research community.

Further information:

- [Andreas E. Neuhold](#), work based on “[The taxonomy tree](#)”, FOSTER (Facilitate Open Science Training for European Research)
- Michael Nielsen, [Reinventing Discovery: The New Era of Networked Science](#), Princeton University Press, 2011.

Generally, we consider that Open Science rests on six main pillars:

- Open Data
- Open Source
- Open Methodology
- Open Peer Review
- Open Access
- Open Educational Resources

The Open Data movement fosters initiatives to open data, which means that “anyone can freely access, use, modify, and share for any purpose” (Open Knowledge International, “[The Open definition](#)”). In this new framework, “It has become increasingly apparent that scientific data should be considered a product in much the same way journal articles or conference proceedings are [...].” Felicia LeClere, “[Too Many Researchers Are Reluctant to Share Their Data](#)”, *The Chronicle of Higher Education*, 2010.

Supported by European and national initiatives

In 2013, the Pilot on Open Research Data ([ORD Pilot](#)) announced the European engagement to open research data in the *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020* as part of the Horizon 2020 Research and Innovation Programme. The Pilot “aims to improve and maximise access to and reuse of research data generated by projects for the benefit of society and the economy”. Two types of data are concerned:

- data needed to validate results in scientific publications;
- any data considered valuable by the project.

These data have to be made available for other researchers, industries and citizens. Nevertheless, if the research will be jeopardized or if intellectual property and personal data will be threatened by making data open, projects are allowed to opt out.

So, “the ORD pilot applies primarily to the data needed to validate the results presented in scientific publications. Other data can also be provided by the beneficiaries on a voluntary basis, as stated in their Data Management Plans.” ([H2020 Programme Guidelines on FAIR Data Management in Horizon 2020](#), Version 3.0, 26 July 2016, p.3.)

In July 2016, the ORD Pilot has been extended to cover all Horizon 2020 funded projects. In this updated version of the *Horizon 2020 Programme Guidelines*, open access becomes the default setting for research data generated in Horizon 2020 ([H2020 Programme Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020](#), Version 3.0, 26 July 2016, p.8). This European program is particularly inspired by Anglo-Saxon policies. For instance, American funding agencies like the *National Institutes*

of Health ([NIH](#)) since 2003 and the *National Science Foundation* ([NSF](#)) since 2010, fostered by the American Government, impose to their funded researchers to subscribe to their research data sharing policies on open access. In United Kingdom, most of the funding agencies require that funded researchers made their research data openly available, with the help of dedicated structures such as the *Joint Information Systems Committee* ([JISC](#)) and its *Digital Curation Center* ([DCC](#)), dedicated to data management in the UK.

Direct benefits for researchers

This new model of openness aims to offer new resources that can be exploited by economy and by research. Sharing research data provides direct benefits to researchers, but, some of them are still reluctant to share them. As Felicia LeClere stated, “Data sharing is a bit like going to the dentist. We can all agree that it is a good thing to do and intrinsic to good scientific practice. In reality, however, researchers tend to view data sharing with a mix of fear, contempt, and dread” ([The Chronicle of Higher Education](#), 2010). Fortunately, the situation is gradually evolving, but sharing (or not) rests most of the time on the shoulders of the researchers. Researchers need to be clearly aware of the benefits of sharing their research data:

- It fulfills requirements of:
 - Funders
 - Journals
- It increases research impact and visibility
 - By getting credit for research outputs
 - By boosting citation rate
- It saves time
 - By planning ahead the research
 - By being more efficient (data and methods already explained)
- It preserves data
 - By depositing in a repository
 - By making accessible unpublished data with a citable links. Videos, posters, full methods can be published and used with full citable links via permanent DOI
- It ensures higher quality data
 - Maintaining data integrity
 - Managing and documenting data throughout its life cycle will allow you and others to understand and use your data in the future.
- It promotes innovation and potential new data uses
 - By creating new collaborations between data users and data creators
 - By encouraging new research in a field.
- It maximises transparency and accountability
 - By allowing scrutiny of research findings
 - By improving and validating research methods
 - By reducing fraud
- It supports Open Access
- It helps less rich institutions and countries to do research
- It makes good science and contribute to scientific progress

Funders and research institutions can also take advantage of this model of openness as well:

- Maximising return on investment
- Reducing the cost of duplicating data collection
- Getting access to great resources for education and training

Why manage data?

For yourself:

- Keep yourself organized
- Control the various versions of your data
- Systematically control the quality of your data
- Make backups to avoid data loss
- Format data for reuse (by yourself or others)
- Be prepared: document your data for your own recollection and reuse (by yourself or others).

For funders, data are valuable assets and they are expensive and time consuming to collect. Data should be managed to:

- Maximize the effective use and value of data and information assets
- Be assured that the quality of data is continually improved (data accuracy, integrity, integration, timeliness of data capture and presentation, relevance and usefulness)
- Ensure appropriate use of data and information
- Facilitate data sharing
- Ensure sustainability and accessibility for reuse in science

Research data management

Definition

“Data management is integral to the process of conducting research.”

University of Leicester, [“When do you need to think about RDM”](#)

Data management has to be seen as the baseline of the research lifecycle. In this regard, it should be designed as early as possible and evolve all along the research project. This practice allows researchers to plan and decide how they will “collect, organise, manage, store, backup, preserve and share [...] data during a research project, and after the project is complete”. Good research data management is the “key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse” ([Guidelines on FAIR Data Management in Horizon 2020](#), Version 3.0, 26 July 2016, p.3).

Research data management usually involves:

- creating a Data Management Plan (DMP) submitted along with a research funding application to explain how data will be managed both during and after a project
- storing research data safely throughout a project and sharing with authorised colleagues
- at the end of a project, cataloguing data and making them available in a long-term repository

Research data lifecycle

Research data management takes into account that data has a longer lifespan than the research project that creates them. The research data lifecycle integrates distinctive features by separating the research process into stages and by taking into account the expanded lifetime of research data. In this approach, at each stage of the research process must be implemented specific research data management practices. Five essential steps compose the research data lifecycle and cover the best practices that you should follow during a research project:

- Plan and fund
- Collect and Analyse
- Preserve and store
- Publish and Share
- Discover and reuse

Research data management services

Today the increasingly collaborative nature of research invites to develop research data management services or RDM services. Researchers need to exchange data and want to use effective systems to store, access and share data securely. In United Kingdom for instance, the *Engineering and Physical Sciences Research Council (EPSRC) Framework on Research Data* has stimulated the development of RDM services within many UK higher

education institutions. Higher Education Institutions have a key role to play in research data stewardship. These data are an asset for institutions, bringing benefits and impact for the institution as much as for the researcher. In order to reach these benefits, effective systems and support services need to be in place.

[Developing an RDM service](#) in an higher education institution implies to:

- Identify areas of responsibility for university management, support, administrative services and researchers
- Have a strategy to develop coherently
- Develop an institutional policy
- Create a long-term business plan with objectives, predicted costs, resource deployment and anticipated benefits
- Provide guidance, training and support to researchers
- Create guidelines to help researchers when they apply for grants and have to submit an outline data management plan
- Make space for storage (safeguarded and accessible)
- Assess your research data
- Create “good” metadata by supporting mechanisms for registering metadata
- Be aware of compliance with institutional, national and international policies
- Guarantee the respect of legal requirements on privacy and confidentiality
- Track the impact of research data with [metrics](#)

Creating Data Management Plans (or DMPs)

The *Guidelines on FAIR Data Management in Horizon 2020* provide a clear definition of Data Management Plans: “Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated ” ([H2020 Programme Guidelines on FAIR Data Management in Horizon 2020](#), Version 3.0, 26 July 2016, p.4).

A DMP is a formal document that outlines what you will do with your data both during and after your research project. It describes the data you expect to acquire or generate during the course of a research project, how you will manage, describe, analyze, and store those data, and what mechanisms you will use at the end of your project to share and preserve your data.

Several funders now make data sharing mandatory and applicants must provide a data management plan.

DMP key features:

- It is a regularly updated roadmap
- It is a standardised document
- Its content varies depending on projects’ requirements and funding agencies’ requests
- It focuses on data and datasets collected, created, analyzed by the research project.

It is a deliverable of the project, but not a “technical” document:

- It materializes the data policy of a project
- It sums up the goals and actions that will be implemented

- It meets funder's requirements

Digital data requires an “active management”, it means:

- an ongoing maintenance (backup, migration, conversion, etc.) all along the data lifecycle
- an action plan in terms of data quality, technical feasibility, financial viability

In this context, data management is not data stewardship and means optimizing resources for a specific purpose. It allows ones to:

- Identify and make visible the actions to be conducted;
- Plan key stages, deadlines and critical time periods.

=> This active management goes hand in hand with “digital curation”. It is an ongoing process that requires time and resources and consists in selecting, preserving, maintaining, collecting and archiving digital assets:

“Data curation activities enable data discovery and retrieval, maintain data quality, add value, and provide for reuse over time. This new field includes representation, archiving, authentication, management, preservation, retrieval, and use” ([Digital Humanities Curation Guide](#)).

FAIR Data

In January 2014, researchers, professional data publishers and funding agencies met upon the request of the [Netherlands eScience Center](#) and the Dutch Techcentre for the Life Sciences ([DTL](#)) at the Lorentz Center in Leiden. They agreed to support a minimal set of principles and practices: “data providers and data consumers - both machine and human - could more easily discover, access, interoperate, and sensibly reuse, with proper citation, the vast quantities of information being generated by contemporary data-intensive science”. They are the [FAIR principles](#). According to these principles, data should be:

- Findable with descriptive metadata and persistent identifiers
- Accessible in that it can be always obtained by machines and humans upon appropriate authorization, through a well-defined protocol
- Interoperable by using open formats, common standards, documented data specification and consistent vocabularies/ ontologies
- Re-usable with clear rights and appropriate licence.

The European commission endorsed the FAIR principles and wish to see them applied in H2020 funded projects ([H2020 Programme Guidelines on FAIR Data Management in Horizon 2020](#), Version 3.0, 26 July 2016).

H2020 framework requirements

Projects funded under the Pilot on Open Research Data were required to produce a first version of a DMP as a deliverable during the first six months of the project. At the research proposal stage, all projects submitting to “Research and Innovation Actions” and “Innovation Actions” had to provide a short outline of their data management policy, evaluated under the “Impact” criterion. Since July 2016, a revised version of the 2017 work programme extends the Open Research Data pilot “to cover all the thematic areas of

Horizon 2020” requiring all the Horizon 2020 funded projects to provide a Data management Plan.

A template is provided in the Annex 1 of the *Guidelines on FAIR Data Management in Horizon 2020* (version of 26 July 2016). More detailed versions can then be submitted as additional deliverables at later stages of the project but also when any significant changes occur such as the generation of new data sets or changes in consortium agreements. The first DMP must fulfill minimal requirements:

- A description of data to be generated or collected
- The standards and metadata that will be used
- The data sharing or how datasets will be shared
- The archiving and preservation: procedures which will ensure the preservation of data, including backup and storage.

Data Management Plans : How-to guide

Components of a DMP

There are five main categories of information that should be included in a DMP:

- Information about the data and its format
- information about the metadata content and formats
- policies for access, sharing, and reuse of data
- long-term storage
- budget considerations for data management (salary time for data preparation and documentation, hardware and software requirements, etc.)

Crucial points to address

- Responsibility
- Results management
- Back up plan
- Intellectual property rights
- Becoming of the data after the project (hosting in a long-term perspective, access policies, etc.)

Responsibility

It has to be addressed for each step of the DMP => Outline the roles and responsibilities for all activities: data capture, metadata production, data quality control, storage and backup, data archiving & data sharing. Individuals should be named where possible. For collaborative projects the coordination of data management responsibilities across partners should be expressed clearly. Data management is not just the responsibility of the researcher who has created or collected the data, various parties are involved in the research process and may play a role. It is crucial that roles and responsibilities are assigned and not just presumed.

Researcher is the DMP coordinator, responsible for the data and its description, but there are other actors:

- computer engineer (hosting, security, infrastructural aspects)
- Archivist (broad sense): interlocutor for data selection, standards choices, mappings, durations and technical solutions
- research staff designing research, collecting, processing and analysing data
- laboratory or technical staff generating metadata and documentation
- database designer
- external contractors involved in data collection, data entry, transcribing, processing or analysis
- support staff managing and administering research and research funding, providing ethical review and assessing Intellectual Property rights
- institutional IT services providing data storage, security and backup services

- external data centres or web archives that facilitate data sharing

Results management

Data Collection

Two steps

- Document the data creation process or the data collection process for existing data, and the methods of data acquisition.
- Characterization of the data:
 - Raw or derived,
 - Purpose,
 - volume estimation,
 - type (quantitative, qualitative, survey data, experimental measurements, models, images, audiovisual data, samples, etc.)

Datasets management

For each dataset, the DMP should give minimal information:

- Reference & name (Identifier for the dataset)
- Description (metadata of your data):
 - description of the data that will be generated or collected
 - origin (if collected)
 - nature & scale
 - whether it underpins a scientific publication.
 - to whom it could be useful
 - existence of similar data and the possibilities for integration & reuse.
- Which formats/standards are used for this data?

The DMP should contain rules, like a file naming system or a filing plan and people involved in the research should comply to the normative information.

Description and metadata

Is the data understandable by an outside researcher?

The actual description of the data differs from the dataset management on the targeted audience. The audience of the latter was fellow researchers in a project, funders. The former audience is other researchers that will reuse your data.

Documentation and Metadata

Metadata is data documentation. It includes contextual details about data collection and any information that is important for using and understanding the data. A DMP should express if the metadata is:

- Machine/human readable
- Standardized: [DublinCore](#), DataCite Metadata Schema, homemade format
- Automatically or manually captured
- Stored in databases, text files, or as headers in your files (Cf `teiHeader`)

- Created with controlled vocabularies or any internal conventions

Example: DataCite metadata standard

[DataCite](#) is a consortium of several libraries and research institutes that provide Persistent identifiers (DOIs) for research data and a metadata format to describe them:

- Identifier
- Creators
- Titles
- Publisher
- Publication Year
- Resource type
- Format
- Subjects
- Languages
- Version
- description

Formats

Open + Interoperable + Well spread in the research community => Standard

This is applicable both for your data and your metadata: using standardised and interchangeable or open lossless data formats ensures the long-term usability of data. For example, .csv and .txt are non-proprietary and are likely to be readable in the future, regardless of software availability.

But, researchers are strongly encouraged to use community standards to describe and structure data. To help researchers' finding their way in the data formats jungle, we are happy to announce the next release of the Parthenos [Standardization survival kit](#) developed by INRIA, as "a comprehensive online environment aiming at providing basic information, documentation and resources concerning standards applicable in a wide scope of digitally based humanities and cultural heritage research activities." The idea is to gather in one place useful information created by researchers for research project about sharing good practices, guidelines, pieces of code, as a single environment, as a part of the big Parthenos infrastructure which goal is to foster communication and collaborative work between digital humanities researchers.

Backup plan

The DMP must contain information about the storage conditions and the backup procedures of the data during the research. In particular, DMP readers should know how eventual incidents are anticipated.

Technical information are strongly recommended, like the frequency of backups, number of copies, crypting solutions, restoring plans, etc.

Of course, these questions are related to financial and human resources questions: server space costs must be evaluated. The responsibility of the tasks must be identified.

=> Storage = Budget + Anticipation

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Some selection criteria:

- Anticipate the futures uses and reuses
- Legal or policy aspects
- Potential value
- Ratio cost/benefit

Datasets and the associated metadata, software and algorithms used might have to be preserved, for example, the European Code of Conduct for Research Integrity demands to archive primary and secondary data for a « substantial period » ([European Science foundation, 2011](#)). In general, any raw data should be kept with any data products that were particularly expensive or time consuming to obtain should be preserved. You should then find out archives or data centers that are commonly used in your discipline. Data centers usually last longer than lab or personal websites. Besides, your data management plan should describe what data transformations and formats need to be preserved to ensure future usability of your data. Finally, you should identify the person who will be responsible for maintaining contact information with the data center, it is especially important if there are restrictions on data use, for instance a requirement that potential users contact the data collector before reusing data.

Data repositories

The appropriate solution to store your research data is, in many cases, a data repository, which provides (in theory) sustainability. Many exists like [EUDAT](#), [Nakala](#), [Re3data](#), [Zenodo](#).

Important criteria are:

- Data available for reuse (Harvesting, API, etc.)
- Citability
- Visibility
- Transparency
- Links to papers
- Preservation

It is possible to choose a repository according to quality criteria, certified by the [Data seal of approval](#). First created by the Data archiving and Networked services ([DANS](#)), in the Netherlands, the certification process is now an international board (mostly European) that gives a seal to repositories based on quality criteria:

- The data can be found on the Internet
- The data are accessible (clear rights and licenses)
- The data are in a usable format
- The data are reliable
- The data are identified in a unique and persistent way so that they can be referred to

Data access and sharing

A major requirement for any DMP is the description of how data will be shared. As explained before, most of the funding institutions asking for a DMP have in the meantime a specific policy regarding data access and sharing. Therefore, the DMP should gather all information about:

- access procedures and policies
- embargo periods (if any)
- outlines of technical mechanisms for dissemination & necessary software and tools for reuse
- definition of access (widely open or restricted to specific groups)
- data sharing mechanisms (underlying data of a scientific paper, data paper, research data repository, project website, ...)
- if the dataset cannot be shared, it should be explained (ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related)
- Unique identification of the data and its producers : “Where possible, contributors should also be uniquely identifiable, and data uniquely attributable, through identifiers which are persistent, non-proprietary, open and interoperable (e.g. through leveraging existing sustainable initiatives such as ORCID for contributor identifiers and DataCite for data identifiers).” ([Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020.](#))

Underlying issues are ethics and intellectual property rights, and the potential constraints for the data reuse must be identified as well: cite the right holders or the way to contact them, explain the ethical issues possibly encountered (consent, privacy, sensitive data), ... Note that it can have consequences on long time preservation: For example, patents data should be stored indefinitely.

Share research data with a data paper

Amongst the classical ways to share the data, an interesting one for researchers is the data paper, which means that the research data is editorialized and can be published in itself. A data paper is a scientific publication whose main goal is to describe a dataset or a group of datasets, more than analysis or research results and to give access to the described data. So it can be based on your DMP and makes your data more accessible for potential reusers. It makes it human readable and it can provide citation and peer review. Examples of the *Journal of Open Archeology Data* ([JOAD](#)) and *Research Data Journal for the Humanities and Social Sciences* ([RDJ](#)).

Publish your sharing policy with the Data Reuse Charter

An initiative by DARIAH, Parthenos and other partners and it has been presented by [Anne Baillot](#). It is another online environment as a work service where you create a profile (researcher, institution, cultural heritage institution or laboratory, another body that use primary or secondary cultural heritage data, data hosting body, etc.) and state your policy regarding the reuse of your own data.

The benefit of this charter is obvious: you have no more case by case agreement, no more blurry conditions; all is clear and set in one single place. Some basic features are:

- Register according to your personal or institutional profile
- Get in touch with cooperation partners and collections relevant to your activities
- Gather information on relevant topics such as licensing
- Gain visibility and recognition in the international research ecosystem
- Provide an opportunity for cooperation, retrieval of new collections because any institution, etc.
- Emphasize the important notion of citation (reference to the origin or the owner of data and provide sustainable tools for the further citation of your own data).

Conclusion

Making a DMP is defining how the data, within a project, will be:

- Described
- Shared
- Protected
- Preserved

A DMP contains:

- A data lifecycle description (including long term preservation)
- A data description
- A description of the data policy
- The associated costs

A DMP helps at secure and perpetuate data and is above all a way to see clearly a project's organization, on the data side. => Very strategic, but not technical

When?

Before the first data are created and Regularly updated

Why?

Funders wants it and it is a research good practice

Who?

A team work

A research good practice

A DMP formalizes inside a unique document a set of elements and information useful for the project monitoring and for a good management of the results. Its practical benefits are:

- Better understanding of the data
- Long-term research is easier
- Underlying data is more accessible
- Research more visible: better citability
- Save you time
- Allows to focus on research, increasing efficiency
- Prevents problems in understanding data and metadata in the future.

- Data are easier to preserve and archive
- Benefit for both yourself and others in your field. It might prevent duplication of scientific efforts to re-collect your data and it can lead to new and unanticipated discoveries you might not predict.
- Useful for PhDs: good practice, key data available, thesis and underlying data

=> optimization, "profitability" & perpetuation

The DMP Aide-mémoire

- Is there a model required by the institution/funder?
- Who will contribute to the DMP (team members, partner's projects)?
- Who can help (documentation professionals, IT, etc.)?
- Who will use the DMP?
- Use of an online tool?
- Come quickly with a first version
- Updates: required and/or desirable milestones
- Final version
- Identify datasets

Appendix : DMP tools

There are several tools available for helping the creation of data management plans. Two of the most commonly used are [DMPTool](#) and [DMP Online](#). Both operate as “wizards” and provide prompts for the user to fill out in order to create their data management plan. You can save your plan, print it, or export it to your computer. It includes also templates for H2020 projects.

DMPonline Exercise

- 1) create an account
- 2) choose a model
- 3) create and share a plan
- 4) identify a dataset
 - definition criteria of a dataset
 - reasoning of the decision (reproducibility, cost, etc.)
- 5) others datasets? (granularity, strategy and concrete practice, impact)
- 6) commenting fonction
- 7) export

Contact

Marie Puren and **Charles Riondet**, Ph.D., are junior researchers in Digital Humanities at the French Institute for Research in Computer Science and Automation ([INRIA](#)) in Paris, members of the [Alpage laboratory](#) (INRIA – Paris Diderot University). As collaborators to the [PARTHENOS](#) H2020 project, they focus their research on the development of standards for data management and research tools in Arts and Humanities, and they currently work on the creation of a Data Management Plan for this project.

Marie Puren also contributes to the [IPERION](#) H2020 project, especially by upgrading its Data Management Plan. After being a lecturer and a responsible for continuing education projects at the Ecole nationale des chartes, Marie Puren has been a visiting lecturer in Digital Humanities at the Paris Sciences et Lettres (PSL) Research University. Her main publications belong to fields including intellectual history of the XXth century, French studies and digital humanities. Marie Puren has been awarded a Ph.D. in History at the Ecole nationale des chartes – Sorbonne University. She holds Master's degrees in History and Political Science from the Institut d'Etudes Politiques de Paris, and in Digital Humanities from the Ecole nationale des chartes.

Charles Riondet, History PhD and archivist, is also involved in H2020 [EHRI](#) project as a metadata and standards specialist, with a focus on archival metadata (EAD, EAC-CPF).

Twitter: [@puren1406](#) & [@charlesriondet](#)

Email: marie.puren@inria.fr & charles.riondet@inria.fr