allea
All European
Academies

Sustainable and FAIR Data Sharing in the Humanities

ALLEA Report | **February 2020**

Published in Berlin by ALLEA - All European Academies

February 2020

This ALLEA report has been prepared by ALLEA's Working Group E-Humanities, whose members are listed at the end of the report.

Through its Working Groups, ALLEA provides input on behalf of European academies into pressing societal, scientific and science-policy debates and their underlying legislations. With its work, ALLEA seeks to ensure that science and research in Europe can excel and serve the interests of society.

ALLEA is especially thankful to the main authors and editors of the report: Natalie Harrower, Beat Immenhauser, Maciej Maryl and Timea Biro. Invaluable contributions were also received from Bernard Rentier and Tito Orlandi, who were essential in the earlier phases of drafting this report, and Eveline Wandl-Vogt, who helped define the structure of this document and facilitate the hands-on workshop at the DARIAH Event.

For citation purposes, please use the following:

ALLEA (2020). Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities. Edited by Natalie Harrower, Maciej Maryl, Beat Immenhauser , andTimea Biro. Berlin. DOI: 10.7486/DRI.tq582c863

Pictures for cover and back cover, and inner pages, unless indicated otherwise, are by Shutterstock

Table of Contents

3 PREFACE

5 INTRODUCTION

8 IDENTIFY
8 Research Data in the Humanities

11 PLAN
11 Data Management Plans (DMPs)

14 COLLECT, PRODUCE, STRUCTURE and STORE
14 Types and Formats
17 Metadata
19 Data Models

21 DEPOSIT, PRESERVE and SHARE
21 Legal Aspects
24 Licences
28 Trustworthy Digital Repositories and Persistent Identifiers

31 DISSEMINATE

34 LEGACY DATA

36 CONCLUSIONS

39 ACKNOWLEDGEMENTS

Preface

Why these recommendations, and why now?

The FAIR principles (Findable, Accessible, Interoperable, Reproducible) were formally published in 2016 (Wilkinson et al., 2016), following a 2014 Lorentz centre workshop in Leiden, Netherlands, on 'Jointly designing a data FAIRPORT'. Since that time, the adoption of the principles as best practice guidance in data preparation, stewardship and sharing has been rapid and widespread globally, with numerous researchers, working groups, government bodies, professional bodies, and data organisations working to both refine and expand how the principles can be adopted by scholarly communication practices. Research funders are incorporating FAIR data practices into their guidelines, requiring applicants to submit data management plans and ensure that they create FAIR research outputs. Building on the open data pilot in Horizon 2020, the European Commission plans to incorporate Open Science principles and practices across the programme in their next research funding framework, Horizon Europe, explicitly noting the requirement for open access to research data in line with the maxim 'as open as possible, as closed as necessary,' and responsible research data management in line with FAIR principles (data does not have to be open to be FAIR, but openness is made meaningful through FAIR). Similarly, FAIR is being incorporated into the formation and ongoing definition of the European Open Science Cloud (EOSC). It is clear that at the current moment – to say the least – the FAIR principles will persist in shaping the management and sharing of research data for some time.

There are also indications that FAIR is becoming influential beyond the research sector, as principles for managing and providing access to cultural heritage data held by heritage/memory institutions or the 'GLAM' sector (Galleries, Libraries, Archives, Museums). For examples, we can look to supportive statements on FAIR by Europeana, as well as scholarly arguments emerging from the sector itself. Heritage institutions have been moving towards making their collections more accessible through digital means, either by creating digital surrogates for online access, or archiving and making available born-digital artefacts. For many humanities researchers, these digital collections are crucial inputs to research, so the application of FAIR across the research and cultural sectors has the potential to significantly improve data sharing between researchers and heritage institutions as well.

It is clear that at the current moment – to say the least – the FAIR principles will persist in shaping the management and sharing of research data for some time.

In 2015, the ALLEA Working Group E-Humanities published *Going Digital: Creating Change in the Humanities.* The report addressed how the humanities could harness digital approaches and data management processes in ways that would enhance scholarship and ensure that research outputs were sustainable over the long term. The recommendations before you now build on the broader discussion in the *Going Digital* report to make targeted recommendations to humanities scholars on 'FAIRifying' their data. The elements that define the principles are mostly agreed, but the pathway to implementing them is still being developed[1]; there is a clear understanding that the creation and use of data differs significantly by discipline, so implementation will require approaches that are shaped by disciplinary requirements and practices. The report of the European Commission's expert group on FAIR data, published in late 2018, argues that the successful implementation of FAIR principles generally requires significant resources at the disciplinary

[1] For example, the FAIR data maturity model working group of the global Research Data Alliance has been building, through significant community input, core criteria to assess the implementation level of the principles, and will release a stable version of guidelines in March 2020. Also, the FAIR working group of the EOSC Executive Board, which is assessing FAIR initiatives across Europe, will release recommendations on the implementation of Open and FAIR practices within the EOSC at the end of 2020.

level to develop data-sharing frameworks. Interoperability across disciplines to facilitate interdisciplinary research is imperative to the goals of Open Science[2], but the development of data sharing cultures and methods generally starts with disciplines. Our intention in this report is to provide recommendations to humanities scholars, with the understanding that the humanities themselves are diverse and data practices and demands vary significantly. We expect that some parts of this report will be more relevant to some researchers than to others, and also that many parts will be relevant to researchers outside the humanities.

The Process

To build these recommendations, the Working Group E-Humanities incorporated the most up-to-date developments in the FAIR landscape, surveying reports published by the European Commission, seeking out future directions articulated by the projects and groups building the EOSC (European Open Science Cloud) and noting statements and activities on research data from relevant networks, such as DARIAH, CLARIN, OPERAS, the SHAPE-ID project, and the Research Data Alliance (RDA). We drafted a series of recommendations mapped to the key phases of the data management lifecycle, and then ran an open consultation process over the period of two months to gather broad feedback from humanities researchers. The open consultation was launched at the ALLEA General Assembly in Bern in May 2019, and a workshop on the recommendations was held in partnership with the DARIAH Digital Methods and Practices Observatory working group (DiMPO) later that month as part of the DARIAH annual event in Warsaw. Contributors were invited to comment and suggest edits by way of an open Google Doc. Contributions were welcomed from all, with a particular effort to attract feedback from “researchers and practitioners working in disciplines within the humanities, policy makers and representatives of all public and private organisations working in the field.” The open consultation received over 200 comments and editing suggestions, which were each carefully considered, and used to develop the final version of the recommendations. We are grateful to all contributors, as the feedback significantly helped to expand and clarify certain aspects of our draft recommendations, and the result is a much richer set of recommendations. A list of contributors from the workshop and open consultation follows at the end of this document.

We hope that you find these recommendations useful.

Dr Natalie Harrower
Chair of the ALLEA Working Group E-Humanities

The open consultation received over 200 comments and editing suggestions, which were each carefully considered, and used to develop the final version of the recommendations.

[2] Several terms are in use, and we do not make significant distinctions between them in this report. ‘Open Science’ is the preferred term adopted by the European Commission, ‘Open Research’ substitutes ‘research’ for ‘science’ in an effort to emphasise that all disciplines are included, and not just those under the English language understanding of ‘science,’ and ‘Open Scholarship’ adds emphasis to the sharing of knowledge as early as possible in the research process. Many issues addressed in Open Science are also being addressed by experts in ‘Scholarly Communication.’

Introduction

We live in a data-driven world. The increasing volume and ubiquity of data is driving the digital revolution, deeply impacting social, economic and political developments across the globe in a range of areas, such as health, transport, energy, environment, agriculture, journalism, innovation, marketing and public policy. No field of investigation is immune to the 'data phenomenon' as the production, collection, and availability of data directly affects research practice across disciplines and sectors. Procedures and systems for classifying and organising knowledge are crucial to manage the 'data deluge' and will inevitably evolve as data continue to grow.

New analytical methods and tools are being developed to exploit this abundance of data in all fields of research, leading to many questions and challenges. For instance, analysing large corpora requires the use of automated tools and methods that may not be commonly employed in humanities research methods. While tools can be used to simplify data processing, they do not always allow the fine-grained analyses required by the methodologies and theoretical frameworks employed in the humanities. Moreover, visualisation of massive datasets may highlight important, large-scale trends, but it tends to transform vast corpora of complex data into a synthetic and necessarily reduced representation of information, which can lead to the criticism that complex realities have been oversimplified in the process. It does not mean, however, that we need to be reluctant to try new approaches or choose between the two. For instance, in textual studies many would argue that distant reading or macroanalysis (i.e. computational approaches) needs to be supplemented by close reading (i.e. informed philological interpretation of particular texts or excerpts) in order to fully appreciate the results of a massive data mining. Hence the synthesis of both approaches yields innovative results.

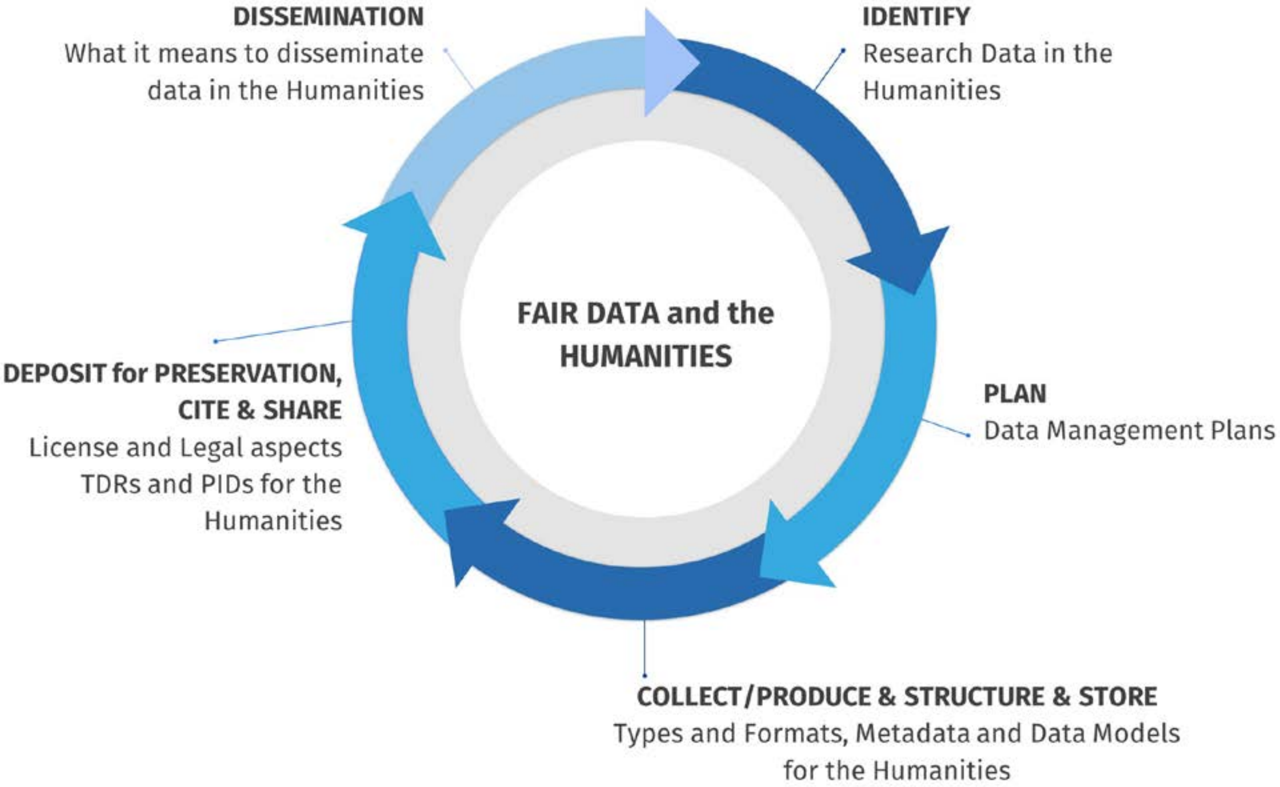
The use or reuse of data varies considerably and is subject to ongoing debate and critical reflection. Public authorities harness data to steer policy or identify social trends, and the commercial sector strongly benefits from the exploitation of large datasets. In a landscape where the opening of datasets created through publicly funded research is increasingly mandated, researchers may question the use of these data, and analyse the role of public institutions and private companies in the production, sharing or appropriation of data. The recent headlines about misuses of data, such as the Cambridge Analytica case or the efforts by Russian parties to sow discord in the lead-up to the 2016 US presidential election, points to the need for ongoing attention to data governance, and the difficult balance between the openness of public data and the protection of privacy. However, data-driven approaches provide complex models to represent, analyse and discuss multifaceted issues. The scholarly imagination has a unique role to play here, as it can identify questions and programmes to harness data in ways that are not guided by narrow commercial or political interests.

The recommendations provide an introduction and further reading on data management, and how data can be constructed, stored, presented, and published in such a way that they can be retrieved, accessed, reused, and interoperable.

The purpose of these recommendations is not to focus on wider societal questions of data governance, but to assist humanities researchers who aim to make their data FAIR: **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable. The recommendations provide an introduction and further reading on data management, and how data can be constructed, stored, presented, and published in such a way that they can be retrieved, accessed, reused, and interoperable. While the implementation of these principles was first proposed in the life, natural and technological sciences, it is now clear that the principles concern the opening, communication, appropriation and reuse of research data, whatever these may be, and are applicable to research outputs across all disciplines. Managing data in line with Open Science and with FAIR principles fosters a transition from human-readable data to machine-readable data, which in turn will require considerable reflection and adaptation by "interpretive" disciplines like the humanities. For researchers, the value of making data FAIR, and accessing FAIR data, is unprecedented access to research assets and analytical tools to interrogate those assets. The goal of this document is to serve as a starting point for scholars, archivists and institutional leadership to creatively engage with this change.

The structure of this report is mapped to the key phases of the data management lifecycle, which is visualised below. Every section contains an introductory part, followed by the recommendations of the Working Group and some indications for further information. In defining research data, we consider all inputs and outputs to the research process and distinguish data from research publications. The lifecycle approach is useful because it presents data management as a series of chronological steps, while also acknowledging that, in a well-functioning ecosystem of research exchange and knowledge building, the outputs of one research undertaking become the inputs to another venture.

Unless otherwise stated, recommendations are focused on digital data, with the understanding that many researchers also work with physical data (see section on Identify) that has its own management and preservation/conservation requirements that are beyond the scope of this document.


DISSEMINATION
What it means to disseminate data in the Humanities
IDENTIFY
Research Data in the Humanities
FAIR DATA and the HUMANITIES
PLAN
Data Management Plans
COLLECT/PRODUCE & STRUCTURE & STORE
Types and Formats, Metadata and Data Models for the Humanities
DEPOSIT for PRESERVATION, CITE & SHARE
License and Legal aspects
TDRs and PIDs for the Humanities

FURTHER READING

ALLEA. Working Group E-Humanities. https://allea.org/e-humanities/

Collins, S., Harrower, N., Haug, D., Immenhauser, B., Lauer, G., Orlandi, T., Romary, L., Wandl-Vogt, E. (2015). Going Digital: Creating Change in the Humanities - ALLEA Working Group E-Humanities report. https://www.allea.org/wp-content/uploads/2015/07/Going-Digital_digital-version.pdf

Common Language Resources and Technology Infrastructure (CLARIN). FAIR. https://www.clarin.eu/fair

European Commission. FAIR Working Group, EOSC Executive Board. https://www.eoscsecretariat.eu/working-groups/fair-working-group

European Commission. H2020 Online Manual - Open access & Data management https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

Hodson, S., Jones, S., Collins, S., Genova, F., Harrower, N., Laaksonen, L., ... Wittenburg, P. (2018). Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data. Luxembourg: Publications Office of the European Union. http://doi.org/10.2777/54599

Isaac, A., Freire, N., (2019). Europeana and the FAIR principles for research data. https://pro.europeana.eu/post/europeana-and-the-fair-principles-for-research-data

Research Data Alliance. FAIR Data Maturity Model Working Group. https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Mons, B. (2019). The FAIR guiding principles for scientific data management and stewardship (vol 15, 160018, 2016). *Scientific Data*, 6. http://doi.org/10.1038/sdata.2016.18

IDENTIFY

Research Data in the Humanities

Introduction

In the humanities, we all use research data, although we may not be aware of it. It is like in the case of Monsieur Jourdain, the title character of Molière's *Le Bourgeois gentilhomme*, who learnt, to his great satisfaction, that unwittingly he had been speaking prose all his life. With research data in the humanities it is exactly the same: you are using it, even if you don't know it, and once you realise it, it will affect your research workflow forever.

Although the term 'data' intuitively seems to be more at place in natural or social sciences (e.g. survey data, experimental data), currently, due to a widespread application of digital means to our academic workflows, scholars in the humanities seem to be recently more eager to consider their sources and results as research data. Some areas in the humanities have long traditions of data-driven research, such as computational linguistics or social and economic history, and some have published clear frameworks for disciplinary data governance, such as the European Association of Social Anthropologists. One reason for the uneven adoption of the term could be that in the humanities, "[w]e resist the blanket term 'data' for the very good reason that we have more and precise terminology (e.g. primary sources, secondary sources, theoretical documents, bibliographies, critical editions, annotations, notes, etc.) available to us to describe and make transparent our research processes" (Edmond & Tóth-Czifra, 2018:1). The resistance to 'data' in the humanities, as an oversimplifying abstraction of complex phenomena, was voiced by many critics, most notably by Johanna Drucker (2011), who opposed the objectifying term 'data' (something given) and proposed to use 'capta' (something captured, taken) instead. This constructivist criticism draws our attention to the fact that data in the humanities are also an effect of operationalisation and interpretive processes.

We could then define data in the humanities broadly as all materials and assets scholars collect, generate and use during all stages of the research cycle. In this report we focus on digital assets, but they are obviously also stored in non-digital formats, like a printed book, handwritten notes, catalogues, etc. There are many typologies and categorisations of research data, depending on disciplines, formats, or media (See section on Types and Formats). Given the introductory nature of this section, let us go through some examples of what research data in the humanities could be.

We could then define data in the humanities broadly as all materials and assets scholars collect, generate and use during all stages of the research cycle. In this report we focus on digital assets.

Probably the most obvious thing that comes to mind, when we think about data in the humanities, are all kinds of lists, tables or matrices containing organised, numerical, categorical, and ordinal information in different disciplines in the humanities, which are usually the outputs (or by-products) of research activities. For instance, the population of French medieval cities, a list of Nobel Prize for literature winners by country, the number of titles and print runs for Victorian novels, a list of participants in the Society of Independent Artists exhibition in 1917, the GDP of European countries before and after Brexit, etc.

However, we may also think of objects, the primary focus of scholarly inquiry, as our research data. Possible study objects used in the humanities, in both physical and digital form, may include historical artefacts, digital (incl. digitised) documents, images (2D or 3D), sound and video recordings. These may entail among others archaeological finds, medieval manuscripts, poetry texts, social media posts, paintings, 3D scans of architecture, recording of a theatre performance, and so on. All digital objects could be stored as facsimile or surrogate, i.e. image of the physical object, or in some transformed form, e.g. transcripts, texts generated through Optical Character Recognition of a scan, or even a scholarly edition of the source text (i.e. text combined from different versions by the editor).

These objects could be further enriched by scholars who may add information about them or interpretations of their features in the form of annotations. For instance, a digital critical edition of correspondence may contain an XML mark-up of people, objects and locations that appear in the text. These can be further defined in footnotes or critical commentary. These annotations could serve as research data also independently from the artwork itself such as in geographical visualisation of places mentioned in a text.

Object description could also entail its formal properties as in the case of a bibliographical record which provides data about the object, i.e. descriptive metadata (e.g. contributors, title, publisher, place, date, number of pages), or a similar catalogue description of the collection of an art object, GPS coordinates for the archaeological site, and so on. Metadata also perform an important task of pointing researchers to objects which may not be available openly. To facilitate this direction, metadata should follow the FAIR principles.

Finally, let us briefly consider the issue of why it is so important that we recognise our assets as research data and act according to the guidelines presented in the next sections. Let us conclude with an example. In *Graphs, Maps and Trees*, Franco Moretti (2005) discusses a complex evolution of British novelistic genres (1740-1900), using an elaborate graph to show that the lifecycle of most of those genres spanned only one generation. Yet, in order to come up with this conclusion, he had to amass and collate the research material previously collected and presented - not always explicitly - in numerous analytical studies on individual genres. So, the availability of data generated in multiple previous studies allowed entirely new insights to appear. This demonstrates the huge potential of the dispersed, lower-scale data that are sitting in our publications or hard-drive folders, which – when made accessible and aggregated – can open the path to new, original research and could be reused by others.

RECOMMENDATIONS

» Think of all your research assets as research data that could be potentially reused by other scholars. Consider how useful it would be for your own work if others shared their data.

» Familiarise yourself with the FAIR Data Principles before you start collecting data and building corpora e.g. FORCE11: the FAIR Data Principles, GO-FAIR: FAIR Data Principles and discuss with colleagues and experts to build a better understanding.

» Digitally document all your research and data collection work -- at the beginning of a project it is difficult to judge which information of the research process will be important and valuable later on.

» Use well-established tools to facilitate your research work, as many of them allow data sharing e.g. MIT Libraries Digital Humanities: Tools and Resource Recommendations.

» Browse humanities datasets and consider whether your own assets could be published in a similar fashion (e.g. Humanities Commons, UK Data Archive, ARCHE re3data.org filtered for humanities).

» When you start producing data, keep this maxim of Open Science in mind: data should be 'as open as possible and as closed as necessary'.

FURTHER READING

Drucker, J. (2011). Humanities Approaches to Graphical Display. *Digital Humanities Quarterly* 005, no. 1. http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html

Edmond, J., & Tóth-Czifra, E. (2018). Open Data for Humanists, A Pragmatic Guide. Zenodo. http://doi.org/10.5281/zenodo.2657248

Gitelman, L. (2013). ”Raw Data" Is an Oxymoron. Cambridge, Massachusetts: The MIT Press.

Moretti, F. (2005). Graphs, Maps, Trees: Abstract Models for a Literary History. London: Verso.

Moscati, P. (2016). Jean-Claude Gardin and the Evolution of Archaeological Computing. *Les nouvelles de l'archéologie*. 144. 10-13. http://doi.org/10.4000/nda.3457

Tóth-Czifra, E. (2019). The risk of losing thick description: Data management challenges Arts and Humanities face in the evolving FAIR data ecosystem. halshs-02115505. https://halshs.archives-ouvertes.fr/halshs-02115505/document

Wuttke, U. (2019). Here be dragons: Open Access to Research Data in the Humanities. https://ulrikewuttke.wordpress.com/2019/04/09/open-data-humanities/

PLAN

Data Management Plans (DMPs)

Introduction

Managing data for their eventual sharing and reuse is a process that requires attention and planning, so researchers should plan and allocate time for data management early in their research project, and track progress towards goals alongside other project deliverables and milestones. This includes accounting for any costs that may arise through data preparation, curation, and preservation, if applicable. Planning how to manage data is an essential research practice. It has multiple benefits for the researcher, projects and organisations and increasingly, it is part of grant conditions by research funders. Data management plans (DMP) describe how data will be created, collected, managed, documented, described, shared and preserved before, during and after a research activity is conducted. They also specify any restrictions on data use, including aspects related to data ownership and intellectual property, contractual obligations, sensitive data and privacy concerns. The goal is to ensure that data are handled appropriately throughout the research activity.

With proper planning, researchers are better equipped to maximise the use (and reuse) potential of data in current and future research. There are many arguments supporting data management as a key research practice. Enumerating and organising data from the beginning creates efficiencies that save time and effort, and supports planning for various stages of research, which can enable problem-solving early in the research process. Good data management facilitates the reuse of data, which helps avoid duplication of effort, and mitigates against data loss. It also supports collaboration, facilitates continuity across projects, and improves the visibility and impact of research outputs. By being aware of the sensitivity of data and how to securely manage this (personal) data, data breaches will more likely be prevented. Pragmatically, good data management is also increasingly becoming best practice, and DMPs are a new de facto standard requirement by funders. Note also that many publishers are increasingly requiring Data Availability Statements (DAS) aimed at providing information on where the datasets or other research outputs that support a given publication can be found and accessed. The DAS format, different from that of a DMP, depends on a journal's requirements and recommended templates but essentially provides links or indications to the location of the various datasets supporting the research.

Good data management facilitates the reuse of data, which helps avoid duplication of effort, and mitigates against data loss. It also supports collaboration, facilitates continuity across projects, and improves the visibility and impact of research outputs.

Common DMP components include:

1. Define the data – in terms of content, format, size, etc.

2. Methodology for data creation and/or collection and quality assurance of data collected, in particular ensuring the validity of the data is supported via rich context metadata.

3. Data models, how will the data be structured/organised and used.

4. Documentation - how will data be documented and described, and what metadata will you create, in which standard.

5. Legal and/or ethical issues.

6. How will the data be stored and accessed during the research activity.

7. How will the data be preserved for the long term, and any relevant policies governing retention or disposal.

8. How will the data be shared – organise data in open, standardised formats, with assigned persistent identifiers that facilitate their reuse and (machine-) readability in a long-term sustainable perspective.

9. Data ownership and responsibilities: who is responsible for data management? And who owns the data? Who manages the life cycle?

RECOMMENDATIONS

» If applicable, determine if the body funding your research has particular requirements for a DMP, or offers a template for framing your plan. If there is no required template, choose an existing appropriate one (e.g. via DMPOnline).

» Devise a DMP prior to collecting data. Define and plan for your data: all research projects deal with data. If your project includes the analysis of text corpora, for example, then the corpora themselves are data, and you should make sure they are clearly described, documented, and managed according to the FAIR principles so your research is reusable by others.

» Plan documentation of metadata: In order for your data to be comprehensible in the future and/or reusable by others, they will need descriptive metadata created according to a common schema to understand the context/purpose of the research. The richer the metadata, the more intelligible and useful the dataset (see section on Metadata).

» Use standardised terminology to increase interoperability. Consider employing vocabularies or ontologies that follow FAIR principles to increase interoperability and findability (e.g. see FAIRsharing.org).

» Consider the right questions to be answered in your DMP that can account for discipline-specific requirements. The DMP templates suggested by funders are quite high level and provide generic guidance for file naming or versioning conventions, database structuring, and can be a good start. Tools like the dmponline.dcc.ac.uk provide discipline specific examples that can be of further reference.

» DMP as living documents: Update your data management plan regularly in order to take into account any potential relevant changes such as using new data types and/or models, technology, new institutional data management policies, reassessing legal aspects or licences for legal compliance etc.

» Depending on the size of the organisation: think of providing institutional support for research data management (RDM); organise information sessions to raise awareness about good research data management, and the risks of not managing it early.

» If possible, consider involving library and/or repository support staff from the initial stages of research data management planning to discuss the best solutions, specifications, standards and protocols along which the repository operates. Repository staff can also assist scholars with understanding any specific data management requirements and associated costs.

» Factor the cost of research data management (time or human resources) into budgetary requirements at the point of application.

FURTHER READING

Consortium of European Social Science Data Archives (CESSDA) Training Working Group (2017 - 2018). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC.

Digital Curation Center (DCC) DMPonline (web based tool templates, examples and guidelines on how to create a DMP according to the requirements of major UK, European, Dutch and Swiss research funders) https://dmponline.dcc.ac.uk/

DCC. (2013). Checklist for a Data Management Plan. v.4.0. Edinburgh: Digital Curation Centre. http://www.dcc.ac.uk/resources/data-management-plans

DMPtool. Data management general guidance. https://dmptool.org/general_guidance

Jisc. (2019). Research data in arts, humanities and social sciences. https://rdmtoolkit.jisc.ac.uk/plan-and-design/research-data-in-arts-humanities-and-social-sciences/

Research Data Management Organiser RDMO https://rdmorganiser.github.io/ demo instance: https://rdmo.aip.de/projects/

Sansone, S., McQuilton, P., Rocca-Serra, P. et al. (2019). FAIRsharing as a community approach to standards, repositories and policies. *Nat Biotechnol* 37, 358–367 http://doi.org/10.1038/s41587-019-0080-8

Science Europe.(2018). Practical Guide to the International Alignment of Research Data Management. D/2018/13.324/4.https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf

Tóth-Czifra, E. (2019). DARIAH Pathfinder to Data Management Best Practices in the Humanities. https://campus.dariah.eu/resource/dariah-pathfinder-to-data-management-best-practices-in-the-humanities

University of Bristol Research Data Service. (2013). DMP Sample AHRC Technical Plan. Retrieved from DMPonline http://www.dcc.ac.uk/sites/default/files/documents/adocs/AHRC_Databris_Religion_DMP.pdf

COLLECT, PRODUCE, STRUCTURE and STORE

Types and Formats

Introduction

Once research assets are identified researchers need to decide about the final shape of their data, taking into consideration the data type (i.e. what kind of data will be collected), format (i.e. what computer file format will be used), and typologies. Research produces a variety of data types and formats, and not all of these are born-digital. Humanities researchers often digitise physical objects, but also create research data that are not digital, such as manually annotated text or hardcopy field notes. While the focus of these recommendations is on digital data (and its 'FAIRification') some references are made to physical data in different sections.

Different disciplines may employ different data types, but also different typologies. The collection typology, for instance, distinguishes between primary data, requested and collected by a researcher through first-hand research (e.g. experiment, survey, observation, text-mining), and secondary data, or resources which already exist and could become the subject of analysis (e.g. texts, objects, descriptions). Hence, one may expect that primary data will be more at home in the social sciences and secondary data in the humanities.

An interesting case of secondary data is the assets available in cultural heritage institutions, or in the GLAM sector. While the data types of these sources are not per se qualitatively different than data drawn from elsewhere, there has been significant effort over the last decades on the part of memory institutions to create digital surrogates of their artefacts or holdings, and often to make these available to researchers and/or the public. The data made available by GLAM institutions may or may not be in digital format, and it may or may not be shared in a way that conforms to FAIR standards. Recent efforts are being made to align data from these institutions with FAIR principles, and a report from DARIAH recommends that researchers can assist in this process by sharing metadata they have created with the originating institution, where suitable.

While the data types of cultural heritage and GLAM institutions are not per se qualitatively different than data drawn from elsewhere, there has been significant effort over the last decades on the part of memory institutions to create digital surrogates of their artefacts or holdings, and often to make these available to researchers and/ or the public.

Another basic typology distinguishes between quantitative and qualitative data. The former concerns the data that could be expressed in a numerical form (both continuous or discrete), e.g. number of coins found on burial sites as well as their measurements. The latter can usually be expressed in language to denote a certain quality of the object or its description, e.g. the testimony of the burial-site keeper or the colour of the coins.

Data types may be distinguished by the basic media they use, i.e. word, image, sound, or the combination of these. For instance, a field notebook, interview, performance recording, photographs, footnotes. Data types also may be considered according to their structure. In a paper on data in the humanities, Schöch (2013) discusses the different levels of structure: structured (database), semi-structured (XML), and unstructured (plain text). Furthermore, such structure could be linear (e.g. table), hierarchical (e.g. tree-like structure) or multi-relational (e.g. network). For a fuller discussion on this, see the section on Data Models.

Once the data type is selected one needs to decide upon the format, i.e. the kind of file in which the data will be encoded. File format choices are an extremely important component of any research data management planning and digital preservation and sharing strategy. There are different data formats for similar data types. For instance, the same text could be stored in a TXT (plain-text), ODT (formatted) or XML (structured) format. The choice of the format should reflect both its type and the desired research use. For instance, storing a manuscript and its OCR transcription in one PDF/A file makes sense for human-interpretation purposes. However, computer-assisted research on these texts would require a machine-readable format, such as a transcript in XML (with markup reflecting various features of manuscript such as line-breaks, highlights, headings, etc.), or at least a plain text file (TXT). Images are best stored in a high-resolution, lossless image format (e.g. TIFF) for quality and accessibility over time. These examples show how tightly the data types and formats are interconnected with the actual research questions and activities.

When making these choices all researchers and data curators should be aware that as technology develops, digital formats – proprietary or not –also change quickly and could become obsolete or corrupted, which renders them unreadable. Hence, it is recommended to use open formats or widely adopted standard formats from the beginning, or export data and results achieved with proprietary software to standard outputs. Standard formats and open formats, because of their wide usage and community support, are expected to be more sustainable and accessible over the long term. For more on digital preservation, see the section on Trusted Digital Repositories.

RECOMMENDATIONS

» Search for advice and recommendation from your community, look for widely used formats possibly with documented standards, consider in advance any software dependencies that are created by your format choice.

» Make sure your data formats are preferred by your preservation repository to ensure long term access and facilitate re-use. (see the part on Plan above)

» The same information could be expressed through different data types and formats. For instance, a list of bibliographical records could be expressed as a table in CSV format or a mark-up text in XML. Before you make a choice you can look up what types and formats other researchers use for similar data or check the preferred formats for digital humanities data, following one of the links listed below.

FURTHER READING

Angelaki, G., Badzmierowska, K., Brown, D., Chiquet, V., Colla, J., Finlay-McAlester, J., … Werla, M. (2019). How to facilitate cooperation between humanities researchers and cultural heritage institutions. Guidelines. Warsaw, Poland: Digital Humanities Centre at the Institute of Literary Research of the Polish Academy of Sciences. http://doi.org/10.5281/zenodo.2587481

Data Archiving and Networked Services (DANS). File formats. https://dans.knaw.nl/en/deposit/information-about-depositing-data/before-depositing/file-formats

Digital Repository of Ireland. (2018). Factsheet No.3: File Formats. https://repository.dri.ie/catalog/jw82mv08x

Digital Research Infrastructure for the Arts and Humanities (DARIAH). Heritage Data Reuse Charter. https://datacharter.hypotheses.org/ See also: https://www.dariah.eu/activities/open-science/data-re-use/

DMPtool. General Guidance on Formats https://dmptool.org/general_guidance#file-formats

FAIRsharing.org. Standards.https://www.fairsharing.org/standards

Koster, L., Woutersen-Windhouwer, S. (2018). FAIR Principles for Library, Archive and Museum Collections: A proposal for standards for reusable collections. *The code4lib journal*. 40. http://journal.code4lib.org/articles/13427

Library of Congress. Recommended Formats Statement http://www.loc.gov/preservation/resources/rfs/TOC.html

Research Data Alliance. Data Type Registries (DTR) Working Group. https://www.rd-alliance.org/groups/data-type-registries-wg.html

Schöch, C. (2013). Big? Smart? Clean? Messy? Data in the Humanities'. *Journal of Digital Humanities* 2, no. 3. http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/

UK National Archives. Guidance Notes http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/guidance/

Metadata

Introduction

While it is central to the FAIR principles to make data accessible, accessibility alone does not render data usable. For data to be reusable, they must be accompanied by sufficient information to communicate the contents of the dataset, the purposes or circumstances under which it was created, and the ways in which it could be reused. Simply described, metadata are 'data about data', or information used to identify and describe data, and they are one of the key building blocks of FAIR data. There are different categories of metadata, including descriptive, preservation, technical and administrative metadata; these recommendations will focus on the descriptive category. The European Commission's expert group on FAIR data argues that central to the realisation of FAIR is the concept of a FAIR digital object – an elemental 'bundle' that includes the research data, the persistent identifier (PID), and "metadata rich enough to enable them to be reliably found, used and cited."

Metadata can be understood as a series of fields that describe data and other research objects in consistent and standardised ways, much like the bibliographic record of a library book. Metadata standards have been created, often by different research communities and disciplines, to provide optimal, tailored ways of describing data. They provide a formal, shared, schematic way of representing knowledge through common language. Early in the research process, aim to identify a metadata standard that is suitable for your discipline or domain, and one that is compatible with the repository in which you will deposit your dataset(s) (See the section on Trusted Digital Repositories and Persistent Identifiers). Some of the standards commonly adopted for humanities data are Dublin Core, which is flexible and widely used in the Digital Humanities, MARC, a library cataloguing standard, and EAD, which is used by archives. While standards vary by discipline, the important thing is to choose an existing standard, instead of creating a new schema for organising your metadata. Standards schemas will facilitate deposit in repositories, and foster interoperability with other datasets.

Metadata standards have been created, often by different research communities and disciplines, to provide optimal, tailored ways of describing data. They provide a formal, shared, schematic way of representing knowledge through common language.

When collecting data during the research process, or storing datasets that have been created through the research process, one should aim to create metadata that describes this data at as early a stage as possible, as this will facilitate working organisation of research data, and make it easier to create the metadata that will accompany the eventual deposit of data for sharing and reuse (see section on Data Management Plans).

Furthermore, metadata itself should be persistently accessible, even if the data it describes is restricted or no longer available.

RECOMMENDATIONS

» A good starting point is to consult the Metadata Standards Directory, a community-maintained directory hosted by the Research Data Alliance: https://rd-alliance.github.io/metadata-directory/

» Metadata works best when terminology is consistent, e.g. naming conventions are followed, spelling is normalised, and so on. Depending on the complexity and size of your metadata, consider using a tool such as Open Refine to 'clean' your metadata.

» For greater searchability and interoperability, researchers should also consider using controlled vocabularies to identify common terminology when populating metadata fields. For example, the Library of Congress maintains a controlled vocabulary for subject headings: http://id.loc.gov/authorities/subjects.html

» For findability, the metadata should include a clear and explicit reference to the dataset it describes, through the inclusion of a PID in the metadata (see section on Trustworthy Data Repositories and Persistent Identifiers).

» Make your metadata as rich as possible in order to better contextualise your data and facilitate reuse. Consider more detailed descriptions, and fuller provenance information, as well as using the spectrum of available metadata fields.

» Metadata should be machine-readable.

FURTHER READING

Digital Repository of Ireland. (2016). DRI Guidelines, Digital Repository of Ireland [Distributor], Digital Repository of Ireland [Depositing Institution], https://doi.org/10.7486/DRI.r4958092r

Digital Repository of Ireland. (2019). Vocabularies. https://dri.ie/vocabularies

Fordham University Libraries. Digital Humanities: Understanding Metadata. https://fordham.libguides.com/DigitalHumanities/Metadata

Higgins, S. (2007). What Are Metadata Standards. http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards

Open Refine (open source tool). https://openrefine.org/

Research Data Alliance Metadata Standards Directory Working Group. Metadata Standards Directory. http://rd-alliance.github.io/metadata-directory/

Data Models

Introduction

The modelling of information about artefacts and abstract concepts has always been an important issue in the humanities research process. A critical apparatus of an edition or the cataloguing of archaeological findings are recognised ways of representing knowledge within certain disciplines. Although there is a long tradition of debate among schools of scholars on how this best should be done, these debates always refer to defined representational systems. With the digital turn, these systems of information modelling have to be transferred into data models, and the artefacts – usually digital objects – have to be described in an adequate manner and set in relation to one another and to abstract concepts.

Within the research process, the process of data modelling is therefore of great importance and should be approached in an informed way. It should be done at the initial phase of the research process, when the research topics, the available data and the research interest have been defined and clarified to some extent. It is useful to allow sufficient time for this period within the research process, as later changes of the data structures or even of the data model can be very time-consuming and inefficient. Furthermore, common data models can make good use of many standardised tools for processing, annotating, accessing, analysing or illustrating data and foster their sustainability and reusability.

The first step of data modelling usually consists of defining the entities (e.g. artefacts and concepts), their attributes (e.g. name, age, measurements) and their relations (e.g. author of, mother of, published by), followed by the mapping of the conceptual model in the technology best suited for the research project. Three types of data models are most relevant for the humanities. For well structured, uniform data (e.g. registries of custom revenues, matriculation registers of universities, catalogues of natural history collections) the model of choice is a database (often a relational database), because it allows quick and easy access to the informational content and it enables quantitative analysis. It must be possible to enter the data as records in one or several tables. For low or variably structured data such as texts (letters, editions etc.) the document-oriented XML-model has become a widespread standard. Data represented in XML must have a hierarchical order expressed in a tree structure (e.g. book -> chapter -> paragraph -> phrase -> word). Graph databases usually structured in the Resource Description Framework (RDF) describe qualitative and/or quantitative relations between entities and, therefore, can be used to express their semantic meaning. The data must be representable as a semantic triple in the form of subject – predicate – object (e.g. Shakespeare is the author of Hamlet) and be structured through ontologies. Within the humanities the RDF data model was first used for cultural heritage data, more recently it is used by a growing community for linked data (e.g. network analysis) and especially for Linked Open Data (LOD) in the context of Open Data and the semantic web.

RECOMMENDATIONS

» Data models go FAIR: the FAIR Guiding Principles, correctly applied, ensure data are findable, accessible, interoperable and reusable. Data modelling should take this into account by using formal, easily accessible languages for knowledge representation, providing persistent identifiers, open standards, well documented Application Programming Interfaces (API), generic user interfaces and rich metadata. The FAIRification process developed by the GO FAIR initiative offers a system on how to shape the data modelling.

» Use open standards, and whenever possible, standardised technologies and procedures should be used. The World Wide Web Consortium W3C maintains several standards relevant for data models like XML and RDF. Within XML the Text or Music Encoding Initiative TEI/MEI or specific expressions of them have become standards for text or music editions. The query language SPARQL and the representation tool for linked data JSON-LD are common standards for RDF (refers to FAIR principle I.1).

» Prefer human and machine-readable systems: coding of data models and of the actual data that is both human and machine-readable in a unified way provides better sustainability and long-term accessibility than machine-readable only code (binary codes), that may use different formats for data model description and the actual data. For both, hierarchical data models and graph-based data, various serialisations (file formats) are available that fulfil this condition (XML, TEI/XML, Turtle, N3, RDF/XML), whereas SQL based technologies need bigger efforts.

» Normalise as much as possible: to avoid redundant information, the content of databases should be normalised as far as possible, using for example authority files like VIAF and identifiers like DOI, ARK, ISNI, GND and the like. To foster the exchange of data, standardised vocabularies and ontologies are needed as well, but an overall ontology for the humanities has not yet been established. The ontology CIDOC-CRM and especially some extensions are well on their way to become a reference model for cultural heritage information.

» Data models follow the data management plan (DMP): when establishing a data model, researchers should keep the whole lifecycle of their data in mind, as it should be outlined in a DMP. Therefore, an extensive documentation of the data model, its software and tools are highly relevant and facilitates the transfer of data in a secure and trusted repository in order to keep them accessible. The same is true here: the more you use open standards for your data model, the easier this task becomes.

FURTHER READING

Flanders, J. (Ed.), Jannidis, F. (Ed.). (2019). The Shape of Data in Digital Humanities. London: Routledge, https://doi.org/10.4324/9781315552941

GO-FAIR. FAIRification process. https://www.go-fair.org/fair-principles/fairification-process/

GO-FAIR. GO StRePo. https://www.go-fair.org/implementation-networks/overview/fair-strepo/

DEPOSIT, PRESERVE and SHARE

Legal Aspects

Introduction

Sharing data inevitably raises questions about intellectual property rights and privacy. With the digital turn, digital content is becoming more accessible, but it can still be subject to protections and researchers need to be cognisant of these when practicing good data management and planning. In some disciplines, the requirements of securing patent protection may restrict the ability of scientists to share data, as recently discussed by in The Need for Intellectual Property Rights Strategies at Academic Institutions (2019), a recent ALLEA statement prepared by its Permanent Working Group on Intellectual Property Rights. More relevant to humanities data is the question of copyright. As the name implies, copyright basically protects the form in which creative content is expressed against unauthorised copying. It does not protect ideas themselves, merely their expression in concrete form, and there must be an element of human creativity if copyright protection is to apply. Thus, data generated or collected in the e-humanities may potentially be subject to copyright in whole or in part. In addition, particular challenges may arise when some of the items within a data set themselves are subject to third party rights.

These legal issues can affect artistic and literary works like editions, pictures, films, sounds and other recordings, but also software or databases, and it is important to know that they are regulated according to the territoriality principle. Subtle variations exist between different legal systems in how copyright is interpreted. This is a complex subject where it is impossible to give more than general guidance and in case of doubt or where necessary appropriate local experts should be consulted. For certain topics the EU has created an overarching legal framework like the GDPR, the Directive on the Legal Protection of Databases, or the Directive on Copyright in the Digital Single Market, relevant for text and data mining (TDM) within the research context.

Data generated or collected in the e-humanities may potentially be subject to copyright in whole or in part. In addition, particular challenges may arise when some of the items within a data set themselves are subject to third party rights.

In practice, digital humanists can make good use of some checklists to determine whether and how data relevant for their research are subject to legal regulation. Some important questions that have to be solved concern topics like:

- Which national legislation applies to other researchers' work I use in my project?

- Do I have the right to collect, preserve and provide access to the data of my project?

- Is there sensitive information that could connect to some privacy issues?

- Are there risks of exposing the identity of human participants in my study?

- Am I allowed to digitally reproduce material and (re-)publish it in a digital reproduction?

RECOMMENDATIONS

» Clarify all legal issues at the beginning of your research project and include the findings of this process in the data management plan.

» Use checklists adequate to your research topic/discipline.

» Check the resources indicated by DARIAH, CLARIN. (see further reading).

» In the case of personal data ensure that only relevant people can access the data and that these are clearly identified (see GDPR).

» Ask for consent to share anonymised data and establish transparent and well-documented anonymisation routines that consider not just direct identifiers, but also how a combination of indirect identifiers could reveal identities. (See for example the guide on informed consent in the CESSDA data management expert guide).

» Avoid collection of (sensitive and non-sensitive) personal data when possible.

» Get legal support (IPR, copyright, patents, trademarks etc.) from your home institution. If there is no dedicated office for this purpose, try to get information from your university library, as its staff are often confronted with such issues.

» If you need permission from the copyright holder in order to use sources like images for your publication, try to get one that covers both printed and digital copies.

» Finally, check the recommendations in the section on Licences, that are closely related to this section.

FURTHER READING

ALLEA. Permanent Working Group Intellectual Property Right. https://www.allea.org/working-groups/overview/permanent-working-group-intellectual-property-rights/

Common Language Resources and Technology Infrastructure (CLARIN). Bibliography/Further reading on Legal and Ethical Issues https://www.clarin.eu/content/bibliographyfurther-reading-legal-and-ethical-issues

Common Language Resources and Technology Infrastructure (CLARIN). Legal Information Platform. https://www.clarin.eu/content/legal-information-platform

Consortium of European Social Science Data Archives (CESSDA) Training Working Group (2017 - 2018). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. Anonymisation: https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/5.-Protect/Anonymisation

Digital Research Infrastructure for the Arts and Humanities (DARIAH). Working-group Ethics and Legality in the Digital Arts and Humanities (ELDAH). https://www.dariah.eu/activities/working-groups/ethics-and-legality-in-the-digital-arts-and-humanities-eldah/

European Commission. Digital Single Market. Copyright in the Digital Single Market. https://ec.europa.eu/digital-single-market/en/copyright

European Commission. Digital Single Market. Protection of databases. https://ec.europa.eu/digital-single-market/en/protection-databases

European Commission. General Data Protection Regulation (GDPR). https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04

Galina, I., Gil, A., Padmini, R. M., Zafri, V. (2017). Copyright and Creator Rights in DH Projects: A Checklist. http://dx.doi.org/10.17613/M6148V

Hannesschläger, V. (2018). The Legal Checklist for Digital Cultural Heritage Projects. Legal Issues for Open DH. https://legaldh.hypotheses.org/292

Hannesschläger, V., Scholger, W. (2018). ACDH Tool Gallery 4.2: Intellectual property rights, data privacy and licensing tools: theory and practice. Zenodo. https://doi.org/10.5281/zenodo.1470537

UK Data Archive. (2011). Managing and Sharing Data. https://ukdataservice.ac.uk/media/622417/managingsharing.pdf

World Health Organisation. Templates for informed consent forms. https://www.who.int/ethics/review-committee/informed_consent/en/

Licences

Introduction

Researchers are “prosumers” who produce and consume information and knowledge of other researchers. This section focuses on their role of producing knowledge and on ways to foster its diffusion by clear legal boundaries. In the humanities, texts are quite often closely intertwined with underlying data, which form an indispensable part of digital publications. Traditional conceptions of copyright like “All Rights Reserved” raise obvious problems for data sharing in the context of publications. In general: if machine readable data is to be shared, the recipient, in order to use the data effectively, will most likely need to make a local copy for analysis, or for merging with other data sets, or to extract some subset of the data. For this reason, our recommendation is to avoid applying any legal restrictions that do not embrace the principle of openness. The Reusability FAIR principle recommends that data and metadata are released with a clear, human and machine readable data usage licence, in order to avoid legal ambiguity that could limit their reuse (FAIR principle R1.1).

Our recommendation is to avoid applying any legal restrictions that do not embrace the principle of openness. The Reusability FAIR principle recommends that data and metadata are released with a clear, human and machine readable data usage licence, in order to avoid legal ambiguity that could limit their reuse.

The reuse of data in general should be guided by a licence that is as open as possible. Nevertheless, the question may arise as to whether open licences pose a particular problem for humanities data. Special attention must be paid to elements of different origins in texts and data collections, so that it might be necessary to consider different levels of copyright, which impacts on licensing. A common example is the use of images, for which permission was granted, in an article. Open access and free use can only be granted to content of which one owns the copyright or that is already part of the public domain. But in general, humanities scholars can be faced with similar challenges as in other disciplines, such as considerations around sensitive data or concerns about plagiarism. An important distinction must be made between the incorrect use of licensed content and unethical scientific behaviour: Plagiarism or the alteration of foreign content without proof of the source with deceptive intent primarily violates ethical scientific behaviour. For further guidance on research integrity, see the ALLEA European Code of Conduct for Research Integrity, which has been used as a reference document for all Horizon 2020 funded research grants. In 2002, the non-profit corporation Creative Commons (CC) developed a system of easily applicable licences that gives the licensor a choice of licences for the reuse of works ranging from as little restricted as possible to limited in various ways. With the version 4.0 released in 2013 the CC-licences are designed to serve as global licences and thereafter gained a standard-like standing not only in the creative sector but also within the scientific community. They are mainly used for creative works and for publications (including data). The CC system offers four types of licences regulating the use of works (derivative and no derivative) in terms of copying, distributing, displaying, performing and remixing by the licensees: Note that not all combinations are considered open (see Recommendations below).

- CC BY (Attribution): Work may be used by giving credit to the author(s).

- CC SA (Share-alike): Work may be distributed under an identical licence.

- CC NC (Non-commercial): Licensees may use work only for non-commercial purposes.

- CC ND (No Derivative): Licensees may only disseminate the verbatim work, derivative or remixed copies of the licensed version are excluded.

These four types can be used in six combinations:

Open access and free use can only be granted to content of which one owns the copyright or that is already part of the public domain. But in general, humanities scholars can be faced with similar challenges as in other disciplines, such as considerations around sensitive data or concerns about plagiarism.

BY, BY-SA, BY-NC, BY-NC-SA, BY-ND, BY-NC-ND. Creative Commons also introduced the quite often used CC0 (Zero) licence that corresponds to a large extent with the public domain, meaning that the author waives as many rights as legally possible.

Different licence systems are more or less appropriate for data versus software. For example, a widely used open licence specifically for software is the MIT licence. The GNU General Public Licence (GNU GPL) became, especially since its third version of 2007, a widely accepted way to licence the free re-use of software. Licensees are allowed to run, study, share and modify the software and its code. For databases, depending on the national legal situation, it must be noted that the level of creation determines the copyright. The Open Knowledge Foundation created in 2007 the Open Data Commons designed for open licences especially for data and databases. The foundation released three free licences:

- Public Domain Dedication and Licence (PDDL): Corresponds with CC0.

- Attribution Licence (ODC-By): Corresponds with CC BY.

- Open Database Licence (ODC-ODbL): Corresponds with CC BY-SA.

Of course, not all data can be openly shared. There may be privacy concerns, issues of commercial confidentiality, questions of security, etc. which preclude full open data sharing. The maxim as always is "as open as possible, as closed as necessary". But if it is possible to be open, and the decision has been taken whether as a matter of institutional policy, regulatory requirement or personal conviction to go for Open Data sharing, then for the reasons outlined above, the CC-BY route is to be recommended as avoiding ambiguity and making clear that the data can be freely used subject to proper attribution. Regardless of whether or not you open your data, it is still good practice to make it FAIR.

RECOMMENDATIONS

» Proper entitlement: first of all, identify who owns the data, i.e. whether you are entitled to license your work. You may only attribute a licence to a work of which you are the copyright holder. If there are co-authors, you have to agree with them on the licence. Furthermore, you are not allowed to license the works of the public domain. You should also be aware of whether there are any licensing requirements from the funding organisation or the data repository.

» Determine the necessary and sufficient level of access restrictions. Some data cannot be shared openly but can still be shared under certain restrictions while at the same time protecting the data. See for instance the CLARIN licensing framework for language data or the CESSDA access categories for qualitative and quantitative data (interviews, survey data etc).

» Use free and standardised licences: In order to benefit from the possibility of sharing data since the digital turn and to foster Open Science, use a licence as free as possible. The Open Knowledge Foundation and the Open Access Scholarly Publishers Association only acknowledge CC BY, CC BY-SA and CC0 as compatible with Open Access. Remember that the CC BY licence reflects a long-established element of good scientific conduct: You may quote a work, or parts of it, in your publication as long as you indicate the source of your quotation correctly, otherwise it's plagiarism.

» For editors of journals and repositories managers: Avoid applying more restrictive licences like NC (non-commercial) or ND (no derivatives) just to be 'on the safe side'. NC can produce unintendedly limiting side-effects to potential re-users, as it is not quite clear whether the setting of a re-used work has commercial aspects or not. ND originates from the creative sector and is thought of as an instrument to protect the integrity of a work of art, such as a music composition. Many humanities scholars also want to protect their works from misuse and therefore are in favour of a ND licence. However, the risk of misuse through derivatives in the humanities is often quite low, so one has to balance this potential risk against the potentially unintended constraints imposed by ND, such as restrictions against reuse of publications in text and data mining procedures. Keep in mind that anybody deliberately deriving original content and thoughts by other scholars with misleading intention violates ethical scientific behaviour, whether a work is put under and ND licence or not.

» Use a licence selector: If you are uncertain about the licence you want to choose, a licence selector will help you to make an informed decision. For works in the scope of Creative Commons there is the Licence Chooser, for data and software the Public Licence Selector, which also includes CC licences and other licences mentioned above, as well as the Choose a Licence tool. These tools try to suggest the most open and suitable licence based on users' requirements.

» Make your licence machine-readable. Once you have selected the appropriate CC-licence, you can insert it as a defined text module in your publication and make it easily comprehensible by the corresponding icon. When you have a website, use the HTML code offered by Creative Commons for the icon. This will ensure that search engines will be able to find publications selected by certain licences. Choosing an open licence and making it machine-readable can give your work an extra boost of diffusion (refers to FAIR Principle R1.2).

FURTHER READING

ALLEA (2017). European Code of Conduct for Research Integrity, Revised Edition. https://allea.org/code-of-conduct/

Common Language Resources and Technology Infrastructure (CLARIN). Licenses and CLARIN categories. https://www.clarin.eu/content/licenses-and-clarin-categories

Consortium of European Social Science Data Archives (CESSDA) Training Working Group (2017 - 2018). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. Access Categories https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/6.-Archive-Publish/Publishing-with-CESSDA-archives/Access-categories

Creative Commons. Frequently asked questions https://creativecommons.org/faq/

Creative Commons. Wiki. https://wiki.creativecommons.org/ and https://wiki.creativecommons.org/wiki/License_RDF

GO-FAIR. R.1.1 regulation of the FAIR Principles on licenses: https://www.go-fair.org/fair-principles/r1-1-metadata-released-clear-accessible-data-usage-license/

Grabus, S., & Greenberg, J. (2019). The Landscape of Rights and Licensing Initiatives for Data Sharing. Data Science Journal, 18(1), 29. DOI: http://doi.org/10.5334/dsj-2019-029

Open Access Scholarly Publishers Association. Licensing frequently asked questions https://oaspa.org/information-resources/frequently-asked-questions/

Open Data Commons. Making Your Data Open: A Guide. https://opendatacommons.org/guide/

Open Knowledge Foundation. Conformant licenses. http://opendefinition.org/licenses/

Kamocki, P., Ketzan, E. (2014). Creative commons and language resources: general issues and what's new in CC 4.0. Mannheim, Germany: Institut für Deutsche Sprache. urn:nbn:de:bsz:mh39-32241

Kingsley, D. (2016). Is CC-BY really a problem or are we boxing shadows? Unlocking Research. https://unlockingresearch-blog.lib.cam.ac.uk/?p=555

Trustworthy Digital Repositories and Persistent Identifiers

Introduction

For data to be managed over the long term, and made accessible in a continuous and sustained way, it should be deposited in a location that ensures trusted, ongoing stewardship of the data. Researchers depositing their data and those accessing it for reuse should be assured that data sets are authentic, retrievable, annotated sufficiently to understand the context of their creation, and assigned licence information that clarifies the conditions of reuse.

There are many different ways to store data during the research process, with researchers commonly saving data on personal computers, external hard drives, USB flash drives, institutional servers, or in the cloud through various storage services. Most of us are conscious that we need to back up this data in order to avoid loss from accidental deletion, loss, overwriting, or the failure of storage media. However, storage is not the same as preservation, because digital data are fragile and subject to corruption and degradation over time. File formats or the software and hardware required to access them may also become obsolete. Data published on websites can become inaccessible when links break, pages are moved, or the website disappears. Over time, technology, human actions (or inaction) and environmental factors challenge the integrity of data, so simply 'backing up' that data is not sufficient: it must be preserved. Digital preservation is not a single action, but a process that is designed to ensure digital data are continuously accessible into the future, through all of the changes that time and technology can inflict.

Trustworthy Digital Repositories (TDRs) are designed to meet the challenges of storing data over the long term – to preserve it so that it is findable, accessible, reusable, and the integrity of what was deposited is maintained. Protocols can also be put into place to ensure that sensitive data have restricted access, or certain data sets are embargoed for periods of time where necessary. Certification standards exist to assess TDRs against a range of criteria to ensure trustworthiness and international best practice in digital preservation. For example, the CoreTrustSeal (CTS) was created via the Research Data Alliance, a grassroots, researcher-led organisation, and included the harmonisation of previous standards – one of which was developed by the humanities and social sciences communities (Data Seal of Approval). The report of the European Commission's expert group on FAIR data recommends deposit in certified trusted digital repositories, as these repositories have demonstrated that they meet the requirements of long-term preservation and access. Similarly, the rules of participation in the European Open Science Cloud (EOSC), which are being developed through 2020 by the EOSC working groups, will likely underline the requirement to use certified repositories.

Preservation is generally considered valuable as a goal only when access to the preserved material is provided. For access to be trusted over time, digital data, or 'digital objects' should be provided with a persistent identifier (PID) so that data can be located even if their location on the internet changes. A PID is a globally unique, persistent and resolvable identifier that is based on an openly identified schema. PIDs create stable links for objects, and increasingly are the preferred method for citation and reuse, enabling consistent attribution and tracking. PIDs can identify many different entities, from born-digital objects (documents, data, software) to physical objects (people, samples), to conceptual entities (organisations, projects). Examples commonly used for data include DOIs, ARKs, and Handle, but identifiers should also be applied to other entities, such as authors/researchers (ORCIDs), projects (RAIDs), and permanent locations on the web (PURL).

PIDs also facilitate citation, and for increased findability, links should be created between publications and their associated datasets (bidirectional linking). These links are often created through metadata. Initiatives are well underway to support this linking, and their maturity is being developed. For example, Scholix provides a high level interoperability framework for exchanging information about the links between scholarly literature and data, and is widely supported by journal publishers, data centres, and global service providers. The FREYA project, funded under the European Commission's Horizon 2020 programme, aims to extend the infrastructure for persistent identifiers (PIDs) as a core component of open research, and to connect PIDs to each other in standardised ways.

RECOMMENDATIONS

» To ensure the best possible stewardship of your data, choose to deposit it in a digital repository that is certified by a recognised standard such as the CoreTrustSeal. The Registry of Research Data Repositories (re3data) provides a good starting point, noting disciplines, standards, content types, certification status and more. FAIRsharing (manually curated information on standards, databases, policies and collections) allows you to search databases by subject, and includes entries tagged 'Humanities and Social Sciences'.

» Use disciplinary repositories where they exist, as they are more likely to be developed around domain expertise, disciplinary practices and community-based standards, which will promote the findability, accessibility, interoperability and ultimately the reuse and value of your data. The level of curation available in a repository is key to data quality and reusability.

» Datasets should be assigned persistent identifiers (PID). Most repositories that are designed for long-term preservation will automatically assign or 'mint' persistent identifiers for your datasets, so choosing a quality repository will automate this step. Consider as well signing up for ORCID, a free service that assigns persistent identifiers to individuals/authors.

» To facilitate findability of all research outputs, bidirectional links should be created between publications related outputs, such as data (using PIDs).

» Include the richest metadata possible with your deposited data so that others can find it, understand the parameters under which it was created, and understand the conditions under which they can access and/or reuse it. See recommendations in this report in the sections on Licences and Metadata for more information.

FURTHER READING

Australian National Data Service (ANDS). Persistent identifiers: awareness level. https://www.ands.org.au/guides/persistent-identifiers-awareness

CoreTrustSeal. Core Certified Repositories. https://www.coretrustseal.org/why-certification/certified-repositories/

Digital Curation Centre. How to Cite Datasets and Link to Publications. DCC, 2015. http://www.dcc.ac.uk/resources/how-guides/cite-datasets

Digital Preservation Coalition (DPC). (2015). Digital Preservation Handbook. 2nd Edition https://www.dpconline.org/handbook and Digital Preservation Topical Notes Series: https://www.dpconline.org/knowledge-base/dp-topical-notes

Digital Research Infrastructure for the Arts and Humanities (DARIAH). Data Deposit Recommendation Service for humanities researchers. https://ddrs-dev.dariah.eu/ddrs

European Open Science Cloud (EOSC). About EOSC. https://www.eosc-portal.eu/about/eosc

F1000. Repositories (FAIR1000 be FAIR be Open) (Decision tree for choosing a suitable repository) https://f1000.com/resources/FAIR_Open_Repositories.pdf

FREYA project. About. https://www.project-freya.eu/en/about/mission

Harrower, N., Cassidy K. (2017). Why Storage is not Preservation: A conversation by the Digital Repository of Ireland: https://www.dri.ie/why-storage-not-preservation-conversation-surrounded-conservation

Hellström, M., Heughebaert, A., Kotarski, R., Manghi, P., Matthews, B., Ritz, R., ... Wittenburg, P. (2019). Initial Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC) (Version 1.0). Zenodo. http://doi.org/10.5281/zenodo.3574203

International DOI Foundation. Key facts on the Digital Object Identifier System. https://www.doi.org/factsheets/DOIKeyFacts.html

Martone M. (ed.). (2014). Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. San Diego CA: FORCE11. https://doi.org/10.25490/a97f-egyk

N2T. Archival Resource Key (ARK) Identifiers. https://arks.org/e/ark_ids.html

ORCID. https://orcid.org/

Registry of Research Data Repositories (Re3data). https://www.re3data.org/

Research Activity Identifier (RAID). https://www.raid.org.au/

Research Data Alliance. Repository Audit and Certification DSA–WDS Partnership WG Recommendations. http://doi.org/10.15497/RDA00019

Research Data Alliance (RDA) FAIRsharing WG recommendations https://rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases-wg/outcomes/fairsharing

Scholix - A Framework for Scholarly Link exchange http://www.scholix.org/home

DISSEMINATE

Introduction

The rapid uptake of the FAIR data principles as part of a wider movement towards Open Science is changing how scientists and scholars collect, curate, preserve and share their research data. In particular the principle of "as open as possible, as closed as necessary" is aimed at guiding researchers in their efforts to strike a balance between sharing data and the need to account for issues around sensitive data/legal aspects. Overall this shift has also brought a focus on maximising data use and potential not only for future research but also in other areas (e.g. private sector) and for other categories of potential users (e.g. citizen scientists).

For researchers this comes with the benefit of increasing the likelihood of citations, acknowledgements and collaborations while from the funder's perspective this translates to better value for research investments and increased potential for innovation. It should also be viewed as the crowning point of the data management lifecycle as it relies on good practices in data curation, deposit in a relevant repository and ensuring rich metadata and having persistent identification as a starting point. Active dissemination around data, once the data have been made FAIR, needs to become a key research data management best practice.

Active dissemination around data, once the data have been made FAIR, needs to become a key research data management best practice.

RECOMMENDATIONS

» Humanities scholars are encouraged to take advantage of the frameworks, networks and resources that facilitate the discoverability and wider reuse of research:

- Domain registries, portals, harvesters, e.g. Re3data and FAIRsharing.org
- Platforms e.g. Europeana, AGATE
- Researcher profiles e.g. ORCID

» Share online your data and all supporting materials such as presentations, posters, blogs, data papers etc. and consider using social media for wider outreach, cite using persistent identifiers.

» Consider publishing a data paper either as a preprint or via a dedicated journal for data papers. An emerging practice supporting the FAIR principles, publishing data papers about data sets increases findability as well as reuse, as these provide the key information about specific datasets. e.g. Journal of Open Humanities Data, Research Data Journal for the Humanities and Social Sciences.

» Talk about your research outside academia, consider diverse audiences, such as journalists, policy makers, private companies or citizen scientists as Open Science is ultimately promoting the involvement of a wider audience in scientific research.

» Consider non-traditional channels and formats to present your data: infographics or interactive data visualisations, online exhibition or digital tours, websites or apps, executive summary/lay summary, also consider a wider use of national languages.

» Promote/prepare your datasets for use in class (schools or HEI) or for Hackathons (e.g. Coding Da Vinci).

» As an institution, actively also showcase and provide institutional channels that researchers can leverage, and reward data dissemination.

» Encourage and support pedagogic approaches which include student production and curation of open research data, and use of existing open datasets as open educational resources (OER).

» While considering how to open up your research as much as possible, be aware that you have to take the proper approach to self-archiving and using a trusted repository to make sure you are enabling the discovery, access and citation of your work. While many researchers find Academia.edu and ResearchGate useful as dissemination aids, these should not be used as solutions for self-archiving. (See section on Trusted Digital Repositories and Persistent Identifiers.)

FURTHER READING

AGATE: A European Science Academies Gateway for the Humanities and Social Sciences https://agate.academy/

Alliance of Digital Humanities Organisations (ADHO) http://adho.org/

Bezjak, S., Clyburne-Sherin, A., Conzett, P., Fernandes, P., Görögh, E., Helbig, K... Heller, L. (2018). Open Science Training Handbook (Version 1.0). Zenodo. http://doi.org/10.5281/zenodo.1212496 and online https://open-science-training-handbook.gitbook.io/book/

Europeana: the European digital platform for cultural heritage https://www.europeana.eu/portal/en and the Impact Playbook https://pro.europeana.eu/what-we-do/impact

European research infrastructure for the development of open scholarly communication in the social sciences and humanities (OPERAS) https://operas.hypotheses.org/

Humanities Commons. https://hcommons.org/

National Institute for Health Research (NIHR UK). (2019). How to disseminate your research https://www.nihr.ac.uk/documents/how-to-disseminate-your-research/19951

Padilla, Thomas, Allen, Laurie, Frost, Hannah, Potvin, Sarah, Russey Roke, Elizabeth, & Varner, Stewart. (2019, May 20). Santa Barbara Statement on Collections as Data --- Always Already Computational: Collections as Data (Version 2). Zenodo. http://doi.org/10.5281/zenodo.3066209

Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimisation and Synergies (PARTHENOS) Training Module. Manage, Improve and Open Up Your Research Data https://training.parthenos-project.eu/sample-page/manage-improve-and-open-up-your-research-and-data/

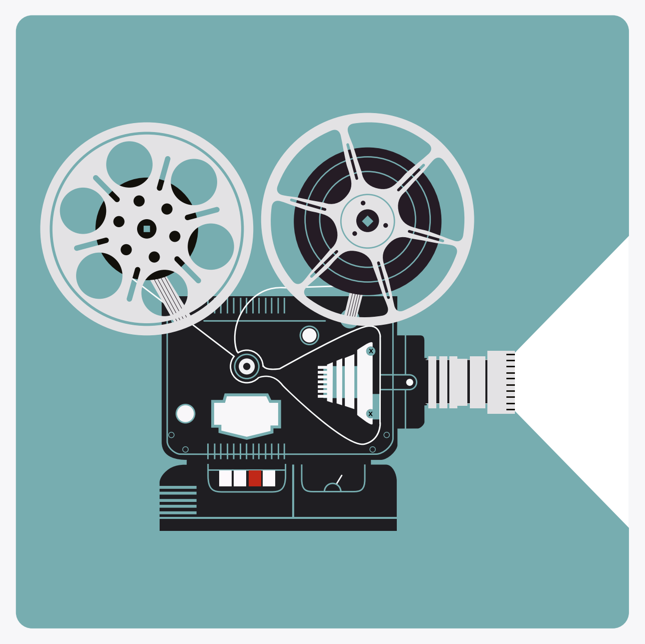
LEGACY DATA

Introduction

Legacy data refers to data that were created and/or stored using previous tools and systems, which may be now outmoded or obsolete. It is a multifaceted notion to describe analogue and digital resources that risk being lost, or that are not in a suitable form to be reused, for various reasons (for example data stored on old media or old digital data). The former links more to digital preservation and (retro) digitisation as steps towards FAIR Data, whereas the latter requires a conversion or migration to up-to-date formats and accessibility. This section acknowledges that legacy data may come with particular challenges or steps in the FAIRification process.

In the humanities context, we usually think of legacy data as a certain array of analogue or digital data, like collections, text editions, lexicographical works, dictionaries and underlying materials like questionnaires and card indexes, all usually stored in libraries, archives or museums and research institutions. Resources in analogue form that are not managed by a professional institution can be at immediate risk of getting lost, when the carrier medium is damaged or is in danger of becoming unusable, like documents with water damage, flammable nitrate-based negatives or newspapers on acid paper. Digital data can also be endangered, for example when files are corrupted, storage media have degraded, or the software and/or hardware required for access has disappeared or is no longer compatible with contemporary tools. In such cases, digitisation as a substitution in order to rescue the resources is highly indicated.
When legacy data are not in immediate danger of degradation or loss, researchers, data owners or institutions – whoever is responsible or interested in these data – may consider converting them to a form that follows the principles of FAIR data, provided they have sufficient funding to do so. As no standardised quality assessment of projects including the repurposing of legacy data has been developed so far, we can only mention some general aspects to be taken into account:

- Before activities are undertaken, it must be clarified whether an eventual secondary publication of the data is legally possible.



- Peer groups/stakeholders should identify and assess important data that is highly relevant for future research and promote it for conservation and digitisation. Ensure that the technical approach to data conversion is up to date and adequate to the standards relevant for the concerned discipline(s).

- Assessment of legacy data is generally enriched through a multi-disciplinary approach, which fosters the wider reusability of legacy data.

- When public funding is involved, free access to the data is a prerequisite, as well as the implementation of the FAIR principles in order to make them interoperable and reusable for research and other purposes.

- The sustainable institutional setting of the converted legacy data can be provided by an assessable data management plan.

- These basic objectives are relevant both for analogue and digital legacy data. They only differ in procedural ways: Analogue data first have to be digitised, digital data have to be converted, both have to be annotated, enriched, interlinked etc.

RECOMMENDATIONS

» Data have to be curated: Try to avoid any data to become legacy data at risk of getting lost. As a researcher develop a DMP but also ensure you realise it at the end of a project.

» Attention must be paid to copyright and licensing conditions associated with legacy data.

» Funding for legacy data: Funding for (retro)digitisation of relevant analogue resources or for repurposing digital data is very difficult to obtain, if it is not part of a research project (and even then). Funding organisations are invited to provide specified funding schemes for this kind of project.

» Safe places for research data: Former legacy data are -- as all digital research data -- in the need of a sustainable infrastructure that secures the efforts done for converting and/or repurposing them. Building up and maintaining suitable long-term infrastructure for research data are crucial success factors to reach this objective.

» Make legacy data open and FAIR: When you get the chance to deal with legacy data use it to create open accessibility and involve a broader public in the sense of the Open Science initiative, and set a high value on interoperability and reusability of the (meta-)data according to the FAIR principles.

FURTHER READING

Abgaz, Y., Dorn, A., Piringer, B., Wandl-Vogt, E., Way, A. (2018). Semantic Modelling and Publishing of Traditional Data Collection Questionnaires and Answers. ***Information,*** 9(12), 297; https://doi.org/10.3390/info9120297

Digital Preservation Coalition. "Legacy Data." Digital Preservation Handbook. https://www.dpconline.org/handbook/organisational-activities/legacy-media

Research Data Alliance. Data Rescue Interest Group https://www.rd-alliance.org/groups/data-rescue.html

Research Data Alliance Data Rescue Interest Group (2017). Guidelines. https://www.rd-alliance.org/guidelines-data-rescue-0

Conclusions

The FAIR data principles have in a short time become highly influential, and while the research community should continue to reflect on, critique and refine the values and goals that underpin these principles, it is clear at the present moment that a paradigm shift in research practice and methodology is occurring globally. FAIR data has been noted as one of the European Commission's 'eight ambitions of Open Science' and it is actively shaping policy by governments, research funders and institutions. Research supports and training are being developed by libraries, departments in higher education institutions, and national or transnational infrastructures, and data are being stewarded and shared under FAIR principles with varying degrees of adoption. The influence of FAIR is moving beyond the research sector to other areas where data sharing and reuse is central to mandates, such as the GLAM (Galleries, Libraries, Archives, Museums) sector, which curates important collections that are also inputs to research, particularly in the humanities.

The report is aligned with the research data lifecycle to facilitate the integration of 'FAIRifying' practices into research workflows. The series of recommendations are meant to provide concrete steps and considerations, and serve as a starting point for implementing best practice in FAIR data management. At the same time, it is important to note that precise implementation pathways for FAIR, and metrics to assess the 'FAIRness' of research outputs, are still being developed. These developments are being shaped by disciplinary requirements and efforts, as well as by discipline-agnostic initiatives and interoperability efforts that aim to facilitate interdisciplinary research as an imperative for the goals of solving major 'societal challenges' through Open Science. Researchers should work within their disciplinary communities to find suitable solutions for different aspects of the principles, but also consider broader networks that bring disciplines together. ALLEA provides an excellent forum for this, in its contributions to borderless and universal science, and its dual mandate to protect excellence in scientific research, while simultaneously safeguarding high ethical standards and academic freedom. Many other networks exist for reflecting on and advancing best practices in research, navigating the quickly evolving landscape of scholarly communication, and participating in its collective development. For example, the Research Data Alliance (RDA) is a global community-driven organisation dedicated to better data sharing and re-use that tackles both social/cultural challenges as well as technical ones, and sets its itinerary based on researcher interests. RDA is unique in that it is open to all scientific and research domains and disciplines, providing an international platform to define best practices and standards and make them openly available for communities to adopt.

Awareness of the FAIR principles and willingness to adopt them is not sufficient to transform data practices in any discipline. The paradigm shift requires effort, and this effort, which impacts on many roles in the research and higher education sectors, requires incentives, support, and recognition for adoption to be successful.

It is important to acknowledge that the success of the FAIR principles is dependent not just on researchers but is highly connected to systemic changes required in wider research culture. Awareness of the principles and willingness to adopt them is not sufficient to transform data practices in any discipline. The paradigm shift requires effort, and this effort, which impacts on many roles in the research and higher education sectors, requires incentives, support, and recognition for adoption to be successful. This larger cultural shift is beyond the scope of this report, but it is essential. The European Commission's expert group on FAIR underlines the importance of developing a culture of FAIR, which requires recognition for practising data stewardship, as well as training and capacity building across the research system. Universities, research centres, academies, policy makers and funding bodies must review their evaluation methods in order to promote adhesion and commitment to the principles and practices that underpin FAIR data management, because, particularly at these early stages, researchers, data stewards, IT professionals, librarians and archivists, and many others in the research ecosystem need certainty that their involvement will be perceived and recognised in ways that are beneficial to assessment and career progression. Similarly, the development of data management skills must be widely supported and nurtured. The present recommendations therefore join other voices in encouraging research institutions, policymakers and funders to fundamentally review their research support services, as well as their definitions of the roles and activities that feed into research under this new paradigm. In the end, the driving force behind data sharing is to advance research, and in advancing research, tackle the challenges facing our increasingly fragile world.

> ***We encourage research institutions, policymakers and funders to fundamentally review their research support services, as well as their definitions of the roles and activities that feed into research under this new paradigm.***

About the ALLEA Working Group E-Humanities

The ALLEA Working Group E-Humanities is charged with identifying and raising awareness for priorities and concerns of the Digital Humanities, contributing to the Open Science and Open Access agenda from a humanities and social sciences perspective, and building consensus for common standards and best practices in E-Humanities scholarship and digitisation. The Group's first publication, Going Digital: Creating Change in the Humanities, made recommendations around archival sustainability and data training required for achieving Open Access and Open Data goals across the humanities.

Currently, the Working Group E-Humanities is focusing on the European Open Science and Open Research agendas, identifying growth opportunities for humanities scholarship, as well as the contributions humanities methodologies can make to truly opening research.

Members of the ALLEA Working Group E-Humanities

- Dr Natalie Harrower (Chair) – Royal Irish Academy
- Dr Beat Immenhauser – Swiss Academies of Arts and Sciences
- Professor Gerhard Lauer – Chair of Digital Humanities, University of Basel (Special Member)
- Professor Maciej Maryl – Institute of Literary Research of the Polish Academy of Sciences
- Professor Tito Orlandi – The National Academy of the Lincei
- Professor Bernard Rentier – The Royal Academies for Science and the Arts of Belgium
- Mag. Eveline Wandl-Vogt – Austrian Academy of Sciences
- Timea Biro (Secretariat) – Royal Irish Academy

Read more: https://allea.org/e-humanities

Acknowledgements

Through the public consultation process and focused efforts such as the workshop Let's Be FAIR: Forging Organisational Recommendations on Research Data in the Humanities collocated with the DARIAH Annual Event 2019, we have collected over 200 comments and edits to the draft document opened for community feedback.

We list below the names of the contributors, acknowledging their support in shaping the Recommendations in a way that is both driven by and for the community. Their volunteer contributions and valuable feedback was crucial and a key indicator of the interest the topic has for the Humanities.

Individual contributions:

- David Bloomfield
- Matthew Cannon
- Viola Capkova
- Birte Christensen-Dalsgaard
- Marie-Louise Coolahan
- Stefan Decker
- William Farrell
- Raman Ganguly
- Julie Gent
- Andrea Goethals
- Rebecca Grant
- Marjan Grootveld
- Leo Havemann
- Gunn Inger Lyse
- Kathleen James-Chakraborty
- Neil Jefferies
- Adeline Joffres
- Sacha Jones
- Patrick Juola
- Radoslaw Kowalski
- John Laudun
- Eric Laureys
- Jacky Leung
- J Love
- Peter McQuilton
- Hollydawn Meunier
- Thomas Padilla
- Hugh Paterson III
- Esther Plomp
- Lai Rina
- Michelle Ryder
- Patrick Sahle
- Susana Sansone
- Amelia Sanz
- Mari Sarv
- Thomas Schmidt
- Barbara Sierman
- Ana Slavec
- Daniel Spichtinger
- Chris Stary
- Pavel Straňák
- Nick Thieberger
- Erzsébet Tóth-Czifra
- Tomasz Umerle
- Suzan van Dijk
- Rene van Horik
- Remco van Weenendaal
- Beatrijs Vanacker
- Ulrike Wuttke

Contributions from organisations

- The International Association of Scientific, Technical and Medical Publishers (STM)
- The European Council of Doctoral Candidates and Junior Researchers (Eurodoc) Open Science Working Group

About ALLEA

ALLEA is the European Federation of Academies of Sciences and Humanities, representing more than 50 academies from over 40 EU and non-EU countries. Since its foundation in 1994, ALLEA speaks out on behalf of its members on the European and international stages, promotes science as a global public good, and facilitates scientific collaboration across borders and disciplines.

Academies are self-governing bodies of distinguished scientists drawn from all fields of scholarly inquiry. They contain a unique human resource of intellectual excellence, experience and multidisciplinary knowledge dedicated to the advancement of science and scholarship in Europe and the world.

Jointly with its members, ALLEA seeks to improve the conditions for research, to provide the best independent and interdisciplinary science advice available, and to strengthen the role of science in society. In doing so, ALLEA channels the expertise of European academies for the benefit of the research community, decision-makers and the public. Outputs include science-based advice in response to societally relevant topics, as well as activities to encourage scientific cooperation, scientific reasoning and values through public engagement.

ALLEA is constituted as a non-for-profit association and remains fully independent from political, religious, commercial or ideological interests.

allea | All European Academies

CONTACT US

ALLEA | All European Academies
Jägerstraße 22/23
10117 Berlin
Germany

☎ +49 (0)30-3259873-72
✉ secretariat@allea.org
🌐 www.allea.org
🐦 @ALLEA_academies