

■ FAIR DATA AND DATA MANAGEMENT REQUIREMENTS IN A COMPARATIVE PERSPECTIVE: HORIZON 2020 AND FWF POLICIES

von Daniel Spichtinger und Susanne Blumesberger

Abstract: *In this paper we provide a comparative perspective on the open data and data management requirements in the European Union's Horizon 2020 programme and those of a national funder, the Austrian FWF. We consider that such a comparative analysis of the requirements pertaining to research data management can help avoiding duplication and assist researchers when drawing up data management plans for their respective funders. We conclude that, although there are some differences in terminology and specific requirements, both the FWF and Horizon 2020 DMPs essentially cover the same ground.*

Keywords: *Data Management Plan; Data Management; Citizen Science; Open Science; Open Data; FWF; Horizon 2020; Research Data Management*

FAIRE DATEN UND ANFORDERUNGEN AN DAS DATENMANAGEMENT IN VERGLEICHENDER PERSPEKTIVE: HORIZON 2020 UND FWF POLICIES

Zusammenfassung: *In diesem Beitrag stellen wir eine vergleichende Perspektive auf die Anforderungen an Open Data und an das Datenmanagement im Horizont-2020-Programm der Europäischen Union und die eines nationalen Förderers, des österreichischen FWF, vor. Eine vergleichende Analyse der Anforderungen an Forschungsdatenmanagement kann aus unserer Sicht dazu beitragen, Doppelarbeit zu vermeiden, und die Forscher bei der Erstellung von Datenmanagementplänen für ihre jeweiligen Fördergeber zu unterstützen. Wir kommen zu dem Schluss, dass trotz einiger Unterschiede in der Terminologie und den spezifischen Anforderungen, der DMP des FWF und des EU-Programms Horizont 2020 im Wesentlichen sehr ähnlich sind.*

Schlagwörter: *Data Management Plan; Data Management; Citizen Science; Open Science; Open Data; FWF; Horizon 2020; Forschungsdatenmanagement*

DOI: <https://doi.org/10.31263/voebm.v73i2.3504>

© Daniel Spichtinger, Susanne Blumesberger



This work – excluding individual logos and images – is licensed under a [Creative Commons License Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Data is sometimes described as the 21st century's most valuable resource.¹ A recently updated study found that the open data market size for EU27+ will reach EUR 199.51 billion in 2025 in a conservative scenario and EUR 334.20 billion in an optimistic scenario. This results in a growth potential for the open data market of EUR 134.69 billion as compared to the 2015 baseline.² This is why data is sometimes described as “the new oil”, although this metaphor does not take into account that data can be reused, while oil cannot. A more apt metaphor would therefore be to compare data with a form of indispensable *renewable* energy.

Focusing on *research* data, policy makers & funders around the globe promote open research data due to benefits for science, economy and society. Research data are thus increasingly conceptualized as inherently valuable products of scientific research, rather than components of the research process that have no value in themselves.³ In this paper, we provide a comparative perspective on the data management requirements implemented by a national funder, the Austrian FWF, and the European Union's Horizon 2020 programme for research and innovation.

2. Open Data and FAIR data management in Horizon 2020⁴

From 2014 to 2016 the European Commission ran an initial research data pilot scheme (ORD Pilot) in some thematic areas of Horizon 2020, with the possibility for grantees to opt-out in case of commercialisation and Intellectual Property Rights (IPR), privacy concerns, national security issues or other significant concerns. As of the work programme 2017 this pilot was extended to all thematic areas of Horizon 2020 (open data as the default), whilst retaining the robust opt-outs described above. These opt-outs can be invoked at any time: during the application stage but also during the implementation phase (in the latter case through an amendment). For the uptake of the ORD pilot from 2014 to 2016 (when its scope was more restricted), figures show an opt-out rate of 35% in the core areas of the pilot. The most important reasons for opt-outs were IPR concerns followed by privacy concerns and projects which do not expect to generate data.⁵

The open data requirement applies primarily to the data needed to validate the results presented in scientific publications. Other data can also be provided by the beneficiaries on a voluntary basis. Costs associated

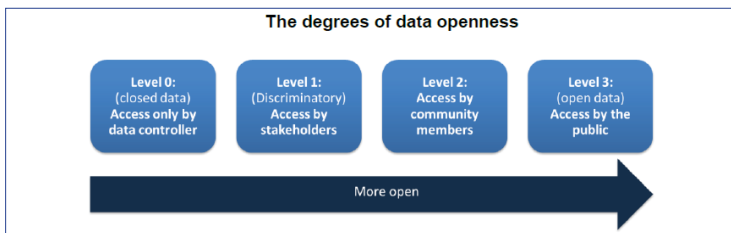
with open access to research data can be claimed as eligible costs of any Horizon 2020 grant,⁶ but only during the duration of the project.

As a main obligation beneficiaries must create a Data Management Plan (DMP) by month 6 of each project.⁷ Such a DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project.⁸ This emphasises the more general need for good data management, rather than just openness. The fact that openness of research data is embedded in a wider context is well expressed in the FAIR principles,⁹ an acronym for making data findable, accessible, interoperable and reusable, developed by the Force 11 community of scholars, librarians, archivists, publishers and research funders. While there is certainly an overlap between degrees of openness and FAIR (as part of accessibility), these two concepts are not synonymous. As part of making research data FAIR a DMP should thus include information on:

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

Since it is impossible to foresee at an early stage of a project how research data will develop in detail, the DMP should be considered a living document and updated over the course of the project, at least in time with the periodic evaluation/assessment of the project or, if such an evaluation does not take place, at the end of the project concurrent with final reporting or review.

When developing a DMP it is important to cross-reference and consider the consortium agreement and the relevant Intellectual Property provisions. The principle of “as open as possible, as closed as necessary” is a good yardstick in this regard. As a practical guideline the OECD project on enhanced access to data includes different levels of data openness, which highlights that open data does not need to be a binary concept:¹⁰



After creating a DMP, relevant digital research data should then be **deposited**, preferably in a research data repository, either an institutional repository (of a university, research institute etc) or a subject specific repository. The Registry of Research Data Repositories (re3data)¹¹ is a useful starting point for finding an appropriate repository. Furthermore, the Zenodo repository¹², run jointly by OpenAIRE and CERN, can be used, in particular if there is no other relevant repository for the project's research data.

As a next step, **open access** to the deposited data should then be provided to enable users to access, mine, exploit, reproduce and disseminate the data free of charge. In the case of databases¹³ this entails assigning an appropriate licence (recommended: Creative Commons CC-BY or CC0). Data needed to validate the results presented in scientific publications should be made open as soon as possible (**but not necessarily immediately**); in particular for data not related to publications the legal obligations do in most cases¹⁴ not prohibit setting a data embargo period in the data management plan. Finally, information should be provided on **tools and instruments** needed for validating the results. Where possible, the data creators should also provide such tools and instruments (e.g. specialised software or software code, algorithms, analysis protocols, etc.).

3. Requirements from national funders: the example of the FWF

A mandatory data management plan must be submitted for projects approved by the FWF since 1 January 2019.¹⁵ The FWF DMP is based on Science Europe's "Core Requirements for Data Management Plans".¹⁶ The DMP covers the following areas:

- a. Data characteristics, including information on source code (if applicable)
- b. Documentation and Metadata
- c. Data Availability and Storage
- d. Legal and ethical aspects

It may also be stated that no data will be generated or analysed. Concerning the data characteristics, the project staff must answer questions such as:

- e. What kinds of data/source code will be generated or reused (type, format, and volume)?

- f. How will the research data be generated and which methods will be used?
- g. How will you structure the data and handle versioning?
- h. Who is the target audience?

The section **Documentation and Metadata** contains questions about metadata standards, documentation of data and data quality control, including which metadata standards are used, whether the data are machine-readable, whether they are compliant to the FAIR Principles and which quality assurance processes will be adopted. **Data Availability and Storage** includes questions about the data sharing strategy and the data storage strategy. In this section, details must be given about the repository selected, which persistent identifiers will be assigned, which data should be archived long-term, how long the data will be accessible after the end of the project, what storage costs will be incurred and whether there are technical obstacles to making the data inaccessible. The section on **legal aspects** includes the questions: “Are there any legal barriers to making the research data fully or partially accessible? Who owns the data? What licence for reuse are you planning to attach to the data? Are there any restrictions on the re-use of the data? If so, why?” Whether there are ethical reasons not to make the data freely available and how sensitive data will be handled during and after the project must be declared in the **ethical aspects** section.

Even if no data is used or produced, a short explanation must be given. The DMP must not exceed a length of 10,000 characters (including spaces). The website states that “the DMP is to be viewed as a living document that can be modified throughout the project. Any changes made to the DMP should be documented, and its final version must be included in the final grant report.”¹⁷

The FWF mandates open access for research data on which the research publications of the project are based. If, for legal, ethical or other reasons, open access to these data is not or only partially possible, this must be specified in the DMP. Open access to all other research data from a project is at the discretion of the principal investigator. The selected repositories must be listed in re3data. Data should be deposited in such a way that it can be re-used without restrictions (e.g., CC BY or a similar open licence). Deposited datasets must be citable by means of a persistent identifier (e.g. DOI).

4. Comparison, Findings & Conclusions

The table below provides a comparative perspective on the two funder requirements, as outlined in the previous sections:

Category	Horizon 2020	FWF
<i>DMP requirements</i>		
Mandate	A DMP must be submitted if the project does not opt-out of open data requirement	DMP must be submitted for projects approved by FWF
Timing	DMP to be submitted within 6 months after project start	DMP must be submitted once project is approved
Length	No requirements	Max 10.000 characters
Updates	“DMP as a living document”	
	Update in time for project review(s) and/or for final review	Include final version in final grant report
<i>Data characteristics requirements</i>		
FAIR principles	Integrated into DMP structure	As a subpart of DMP
Licence	“appropriate licence”, CC-BY or CCO recommended	CC-BY or similar licence
<i>Openness requirements</i>		
Opt-outs	may be claimed for all or certain datasets. Reasons must be provided.	Need to be explained by PI
Type of data	Data underlying publication, optional for other data	Mandatory for data underlying publications, other data within discretion of PI
Timing	“As soon as possible”, data embargoes possible in particular for data not related to a publication.	Not specified
Include tools and software	include “where possible”	Not specified

<i>Other</i>		
Costs of DM	Eligible as part of the grant	“costs for the preparation, archiving, open access and later use of research data in repositories can be requested.”

Tab. 1: Horizon 2020 and FWF requirements (Source: authors' own creation)

Our **main finding and conclusion** is that although there are some differences in terminology and specific requirements, both the FWF and Horizon 2020 DMPs essentially cover the same ground. However, a further harmonisation of the FWF template to Horizon 2020 specifications would be desirable in order to avoid duplication of effort for researchers. This could be done in the context of the preparations of Horizon Europe, the successor programme to Horizon 2020, which is likely to come into effect in 2021 (given there is agreement among Member States on the budget).

Judging from what is already known about the content and structure of **Horizon Europe**, data management will be essential to the new programme; opt-outs are likely to remain in place for providing open access to data but even so, provisions for the appropriate curation and preservation of data (even if not open), and thus an obligatory Data Management Plan, are likely to be a requirement.¹⁸

Mag. Daniel Spichtinger, M. A.
 ORCID: <https://orcid.org/0000-0002-9601-8592>
 Ludwig Boltzmann Gesellschaft, Grant Service & Policy
 E-Mail: daniel.spichtinger@lbg.ac.at

Mag.^a Dr.ⁱⁿ Susanne Blumesberger
 ORCID: <https://orcid.org/0000-0001-9018-623X>
 University of Vienna, Library and Archives
 E-Mail: susanne.blumesberger@univie.ac.at

Acknowledgements

This article is based on the input to a Data Management Workshop held during the Austrian Citizen Science Conference 2019.^{19 20}

References

- Austrian Citizen Science Conference 2019, program. https://www.citizen-science.at/images/Konferenz/Konferenz2019/2019_Konferenzprogramm.pdf
- Blumesberger, Susanne (2019). Wie mache ich meine Daten FAIR? Erfahrungen aus der Universität Wien. <https://doi.org/10.25365/phaidra.139>
- Capgemini Invent Study (2020). The Economic Impact of Open Data Executive Summary. <https://www.capgemini.com/news/capgemini-invent-study-the-economic-impact-of-open-data/>
- The Economist (2017). “The world’s most valuable resource is no longer oil, but data.” <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>
- European Commission, Annotated Model Grant Agreement (2019). https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf
- European Commission, Template Horizon 2020 Data Management Plan (2016). https://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx
- European Parliament and European Council, Directive 96/9/EC (11 March 1996). <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML>
- European Research Council (ERC), ERC Data Management Plan Template (2017). https://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-erc-tpl-oa-data-mgt-plan_en.docx
- FAIR principles. <https://www.force11.org/group/fairgroup/fairprinciples>
- FWF, Data Management Plan Template (DMP) Guide (2019). https://www.fwf.ac.at/fileadmin/files/Dokumente/Open_Access/FWF_DMPTemplate_e.pdf
- FWF, Research Data Management. <https://www.fwf.ac.at/en/research-funding/open-access-policy/research-data-management/>
- Leonelli, Sabina (2013). Why the Current Insistence on Open Access to Scientific Data? Big Data, Knowledge Production, and the Political Economy of Contemporary Biology. *Bulletin of Science, Technology & Society* 33(1-2), 6–11. <https://doi.org/10.1177/0270467613496768>
- OECD, Danish Business Authority, Enhanced Access to Data: Reconciling risks and benefits of data re-use OECD Expert Workshop, Copenhagen (2017). <https://www.oecd.org/internet/ieconomy/oecd-expert-workshop-enhanced-access-to-data-copenhagen-programme.pdf>

re3data.org – Registry of Research Data Repositories. <https://www.re3data.org/>

Science Europe Practical Guide to the International Alignment of Research Data Management (2018). <https://www.scienceurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management/>

Spichtinger, Daniel and Siren, Jarkko (2017). The Development of Research Data Management Policies in Horizon 2020. Filip Kruse und Jesper Boserup Thestrup (eds.). Research Data Management – A European Perspective, De Gruyter SAUR: Berlin/Boston, 11-23. <https://doi.org/10.1515/9783110365634-002> (Open Access available at: <https://zenodo.org/record/1188886>)

Zenodo. <https://zenodo.org/>

- 1 The Economist (2017) “The world’s most valuable resource is no longer oil, but data.” <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>
- 2 Capgemini Invent Study (2020) The Economic Impact of Open Data Executive Summary. <https://www.capgemini.com/news/capgemini-invent-study-the-economic-impact-of-open-data/>
- 3 Leonelli, Sabina (2013) Why the Current Insistence on Open Access to Scientific Data? Big Data, Knowledge Production, and the Political Economy of Contemporary Biology. *Bulletin of Science, Technology & Society* 33(1-2) 6–11. <https://doi.org/10.1177/0270467613496768>
- 4 For a good overview from the perspective of the legal provisions, see European Commission, Annotated Model Grant Agreement 2019. Article 29.3, page 251 ff. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf
- 5 Spichtinger, Daniel and Siren, Jarkko (2017). The Development of Research Data Management Policies in Horizon 2020. Filip Kruse und Jesper Boserup Thestrup (eds.). Research Data Management – A European Perspective, De Gruyter SAUR: Berlin/Boston, 11-23. <https://doi.org/10.1515/9783110365634-002>
- 6 under the conditions defined in the H2020 Grant Agreement, in particular Article 6.2.D.3,
- 7 Note that while a Horizon 2020 proposal is not expected to contain a fully developed DMP, data management as such should be addressed under the impact criterion, as relevant to the project.
- 8 Voluntary templates for a Horizon 2020 DMP are available from the European Commission and the ERC. European Commission, Template

- Horizon 2020 Data Management Plan 2016. https://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx; European Research Council (ERC), ERC Data Management Plan Template 2017. https://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-erc-tpl-oa-data-mgt-plan_en.docx
- 9 <https://www.force11.org/group/fairgroup/fairprinciples>
 - 10 OECD, Danish Business Authority, Enhanced Access to Data: Reconciling risks and benefits of data re-use OECD Expert Workshop, Copenhagen 2017. <https://www.oecd.org/internet/ieconomy/oecd-expert-workshop-enhanced-access-to-data-copenhagen-programme.pdf>
 - 11 <https://www.re3data.org/>
 - 12 <https://zenodo.org/>
 - 13 Defined by the EU as “a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means”, see Article 1(2) Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML>
 - 14 Note that more immediate open data requirements are in place for data related to public health emergencies.
 - 15 FWF, Data Management Plan Template (DMP) Guide 2019. https://www.fwf.ac.at/fileadmin/files/Dokumente/Open_Access/FWF_DMPTemplate_e.pdf
 - 16 Science Europe Practical Guide to the International Alignment of Research Data Management 2018. <https://www.scienceurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management/>
 - 17 FWF Research Data Management <https://www.fwf.ac.at/en/research-funding/open-access-policy/research-data-management/>
 - 18 However, as an important *caveat*, a key document, the future Model Grant Agreement of Horizon Europe, has not been released at the time of writing.
 - 19 See the program: https://www.citizen-science.at/images/Konferenz/Konferenz2019/2019_Konferenzprogramm.pdf
 - 20 See the slides of the presentation of Susanne Blumesberger: Wie mache ich meine Daten FAIR? Erfahrungen aus der Universität Wien. <https://doi.org/10.25365/phaidra.139>