

TUB-DMP

Guidance for Creating a Data Management Plan in Horizon 2020 Projects

– Handout –

Monika Kuberek

Version 1.0, 25 July 2018

Online available

DOI: <http://dx.doi.org/10.14279/depositonce-7199>

Licence

This work is licenced under the Creative Commons Licence
Attribution 4.0 International (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/>

Contact

SZF-Team, Technische Universität Berlin, University Library

E-Mail: team@szf.tu-berlin.de



The Technische Universität Berlin (TU Berlin) is operating a research data infrastructure with different tools and services to support its scientists.¹ The “heart” of this infrastructure is DepositOnce², the repository for research data and publications of TU Berlin, which guarantees the long-term availability, verification, citability and re-usability of research results, i.e. publications as well as research data. Another component of the infrastructure is TUB-DMP,³ a Web tool for creating Data Management Plans. The research data management and the research data infrastructure at TU Berlin are aligned with the FAIR data principles (Findable, Accessible, Interoperable, Re-usable).⁴ TUB-DMP has been developed according to these requirements in order to provide the scientists with a documentation tool that enables the FAIR data management of their research data.

What is a Data Management Plan?

According to Horizon 2020 “a DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project.”⁵ Generally, a DMP documents the context in which research data have been generated and how they will be managed, maintained and preserved – to the purpose to provide information that is necessary to re-use the research data in the future. A DMP contains e.g. the names of the project investigators, the project title, the volume of the research data, information about their preservation, etc.

TUB-DMP

To facilitate the task of creating a DMP, the Web tool TUB-DMP has been developed as a value-added service: (1) TUB-DMP provides a step by step workflow following a checklist. (2) The general project information is auto-imported from the central project database of TU Berlin. (3) There are pre-filled fields with information about the data management strategy of DepositOnce (*DepositOnce Standard Value*). TUB-DMP provides a Horizon 2020 template (see below). Among other functions TUB-DMP has a versioning function and complies with the requirements of Horizon 2020.

Horizon 2020

The EU Framework Programme for Research and Innovation *Horizon 2020* aims to improve and maximise access to and re-use of research data. In the Extended Open Research Data (ORD) Pilot⁶ not all research data needs to be open and there is an opt-out possibility, but it is recommended to create a Data Management Plan (DMP) on a voluntary basis even if the opt-out approach is used.

A Data Management Plan does not have to be submitted at the proposal stage.⁷ However, in your proposal you should provide information about the repository you will use and its data management strategy. **For DepositOnce, you will find the appropriate information about its data management strategy (*DepositOnce Standard Value*) in the Horizon 2020 template below.** You may also fill the Horizon 2020 template in TUB-DMP as far as needed and benefit from the advantages of TUB-DMP.

If you participate at the ORD Pilot a first version of the DMP must be submitted within the first 6 month of the project and updated during the research project life cycle. **We recommend the use of TUB-DMP.**

Instructions to fill in the template in TUB-DMP

1. Login into TUB-DMP with your TUB-account: <https://dmp.tu-berlin.de>.
2. Initiate your DMP by entering your **project number** and an **institutional** (NOT personal) **email-address**. If you have no project number, use a random number.
3. After a plausibility check, the SZF team will get back to you and you can start your entry. TUB-DMP provides a Horizon 2020 template with a step by step workflow.
4. When you open your DMP for the first time, several fields are already filled in. **Note, that you can change the content of all fields according to your project.**
You must not fill in all fields of TUB-DMP, only the fields which are relevant for your project. **Note, that only the fields which you filled in are exported into the PDF.**
You can interrupt your data entry at any time and continue later.
5. When you have finished your DMP you can send it by email to your partners or export in PDF format.

General Project Information

In TUB-DMP, almost all General Project Information is **auto-imported** from the project database of the Research Department of TU Berlin.

Project Title	
Project Description / Abstract (German)	
Project Description / Abstract (English)	
Begin and End of the Project	
Project Stage	<i>[To fill in manually: Individual answer according to your project]</i>
Principal Investigator (FirstName LastName)	
Other Project Investigators (FirstName LastName)	
External Project Partner	
Funding Source	
Funding Program	
Grant Reference Number	<i>[To fill in manually: Individual answer according to your project]</i>
Project Management Organization	
Project Data Contact	<i>[To fill in manually: Individual answer according to your project]</i>

Horizon 2020 Template

The sections of the template and the questions are taken from the *Horizon 2020 FAIR Data Management Plan (DMP) template*.⁸ The use of the template is recommended by the EU commission but voluntary.

1. Data summary	
a) What is the purpose of the data collection / generation and its relation to the objectives of the project?	<i>[Individual answer according to your project]</i>
b) What types and formats of data will the project generate / collect?	<i>[Individual answer according to your project]</i>
c) Will you re-use any existing data and how?	<i>[Individual answer according to your project]</i>
d) What is the origin of the data?	<i>[Individual answer according to your project]</i>
e) What is the expected size of the data?	<i>[Individual answer according to your project]</i>
f) To whom might the data be useful ('data utility')?	<i>[Individual answer according to your project]</i>

2. FAIR data	
2.1 Making data findable, including provisions for metadata	
a) Are the data produced and / or used in the project discoverable and identifiable? <i>Guidance: Discoverability and identifiability of the data means: The data are provided with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers).</i>	
	<i>Text Suggestions (DepositOnce Standard Value):</i> The data are stored in DepositOnce, the repository for Research Data and Publication of the

	<p>Technische Universität Berlin (TU Berlin) (see also 2.2 f). All data in DepositOnce are provided with metadata. DepositOnce automatically assigns a DOI (Digital Object Identifier) to every submitted record and each of its versions, for research data as well as for publications. Research data may be linked to the corresponding publications and vice versa via their DOIs.</p>
b) What naming conventions do you follow?	<p><i>[Individual answer according to your project]</i></p>
c) Will search keywords be provided that optimize possibilities for re-use?	<p>Text Suggestions (DepositOnce Standard Value): As additional metadata, search keywords and Dewey Decimal Classification categories are assigned to the data.</p>
d) Do you provide clear version numbers?	<p>Text Suggestions (DepositOnce Standard Value): One of the core functions of the repository DepositOnce is versioning (see also section 2.2 f), which allows new versions of published records while previous versions are kept available. To every new version, a new persistent identifier (DOI) is assigned. Previous and new versions are linked to each other automatically.</p>
e) What metadata will be created?	<p>Text Suggestions (DepositOnce Standard Value): The stored data are described by using the standard metadata schema Qualified Dublin Core. To meet the requirements of the DepositOnce service partners, e.g. the DOI registration agency DataCite or Open Access initiatives like OpenAIRE, there are some additional qualifiers.</p>
<p>2.2 Making data openly accessible</p>	
a) Which data produced and / or used in the project will be made openly available as the default?	<p>Guidance: <i>If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.</i></p>
	<p>Text Suggestions (DepositOnce Standard Value): By default, all metadata in DepositOnce are openly available as soon as the record is published, even if there are restrictions like an embargo on the publications or research data themselves.</p>
b) How will the data be made accessible (e.g. by deposition in a repository)?	<p>Text Suggestions (DepositOnce Standard Value): All data are stored in DepositOnce, the repository for research data and publications of TU Berlin (see also section 2.2 f). All metadata in DepositOnce are made publicly available in the sense of Open Access. To enable search engines and service providers to index content which is stored in DepositOnce, SZF has taken all reasonable steps: like generating sitemaps, offering an OAI-PMH interface etc. DepositOnce is registered in OpenDOAR and publishes its records in OpenAIRE. DepositOnce is included in common search engines, e.g. Google Scholar, BASE – Bielefeld Academic Search Engine and others. Furthermore, the DOI registration agency DataCite itself acts as a data provider: While registering a DOI, all important metadata are sent to DataCite.</p>
c) What methods or software tools are needed to access the data?	<p>Guidance: e.g., if you are using proprietary programs or file formats to generate and store your research data</p>
	<p><i>[Individual answer according to your project]</i></p>
d) Is documentation about the software needed to access the data included?	<p><i>[Individual answer according to your project]</i></p>

e)	Is it possible to include the relevant software (e.g. in open source code)?
	<i>[Individual answer according to your project]</i>
f)	Where will the data and associated metadata, documentation and code be deposited? Guidance: Preference should be given to certified repositories which support open access where possible
	<p><i>Text Suggestions (DepositOnce Standard Value):</i></p> <p>The project uses the research data infrastructure and the research data management procedures of the Technische Universität Berlin (TU Berlin). They are operated by the Service Center Research Data and Publications (SZF), a virtual organization, where the University Library, the IT-Service-Center tubIT and the Research Department of TU Berlin cooperate to support the scientists of TU Berlin in all questions concerning their research data (https://www.szf.tu-berlin.de/). The research data infrastructure complies with the requirements of the funding organizations, e.g. the European Commission. The technical core of the research data infrastructure is the institutional repository DepositOnce. Other IT applications developed by SZF are TUB-DMP, a web-tool to create data management plans, and PDF/A-Conversion, a web-tool to validate and convert PDF/A documents. The IT applications of SZF are integrated into the IT infrastructure of TU Berlin which is run by tubIT.</p> <p>SZF developed a sustainable strategic concept to store the research data: In DepositOnce, researchers shall store their research results, i.e. consolidated research data and all information which is necessary to verify or re-use the data (e.g. scripts, calculations, software etc.).</p> <p>DepositOnce is based on the open source repository software DSpace. In DepositOnce, research data as well as publications of all kind of disciplines and all kind of formats can be stored. DepositOnce provides a workflow for the metadata description and the upload of the data files. There is a record for each data set (research data as well as publications). Each record has a persistent identifier (DOI). Research data may be linked to the corresponding publications and vice versa via their DOIs. According to the principle of "Good Scientific Practice" the data files cannot be changed after they have been published in DepositOnce. Therefore, one of the core functions of DepositOnce is versioning, which allows new versions of published records while previous versions are kept available (e.g. in order to guarantee correct citation). To every new version, a new persistent identifier (DOI) is assigned. Previous and new versions are linked to each other automatically. Another core function consists in the possibility to set an 'embargo' (i.e. a blocking period) to block the publication of full texts and research data until a certain date.</p> <p>DepositOnce supports open access where possible. In its "Open Access Policy", TU Berlin encourages the scientists to make their research data open access available.</p> <p>Every new IT application at TU Berlin has to go through an approval procedure which is conducted by the Data Protection Officer of TU Berlin and the Staff Council. All IT applications of SZF have successfully passed this procedure. TU Berlin doesn't have a Research Data Management (RDM) policy yet (in preparation in 2018). Since 2002 TU Berlin has a directive for safeguarding good scientific practice: "Statute on the Safeguarding of Good Academic Practice at TU Berlin".</p>
g)	Have you explored appropriate arrangements with the identified repository?
	<i>[Individual answer according to your project]</i>
h)	If there are restrictions on use, how will access be provided?
	<p><i>Text Suggestions (DepositOnce Standard Value):</i></p> <p>All metadata in DepositOnce are openly accessible as soon as the record is published, even if there are restrictions like an embargo on the publications or research data themselves. In this way, it is always possible to contact the author of the data to ask for individual agreements on accessing the data, even if there are general restrictions.</p>
i)	Is there a need for a data access committee?
	<i>[Individual answer according to your project]</i>

j) Are there well described conditions for access (i.e. a machine readable license)?	<i>Text Suggestions (DepositOnce Standard Value):</i> Each license in DepositOnce has a URI (Uniform Resource Identifier): CC licenses and Right Statements as well as software licenses (e.g. MIT License, GNU General Public License, etc.).
k) How will the identity of the person accessing the data be ascertained?	<i>Text Suggestions (DepositOnce Standard Value):</i> See Section 2.2 h.
2.3 Making data interoperable	
a) Are the data produced in the project interoperable?	Guidance: <i>Interoperability of data means: allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins).</i>
	<i>Text Suggestions (DepositOnce Standard Value):</i> DepositOnce is based on the open source repository software DSpace. It is one of the principles in the concept of DepositOnce to use standard interfaces, protocols, metadata, etc. For data exchange, standard interfaces and standard protocols like OAI-PMH, SWORD (Simple Web-service Offering Repository Deposit), etc. are implemented. Using standard metadata schemas in DepositOnce, metadata can easily be converted into other metadata schemas.
b) What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?	<i>Text Suggestions (DepositOnce Standard Value):</i> See Section 2.1 e.
c) Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?	<i>Text Suggestions (DepositOnce Standard Value):</i> As far as possible, standard vocabularies and machine readable file formats are used when storing research data in DepositOnce. In the course of the quality check after the submission of the data, SZF-team verifies whether the data corresponds to the metadata standard.
d) In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?	<i>[Individual answer according to your project]</i>
2.4 Increase data re-use (through clarifying licences)	
a) How will the data be licensed to permit the widest re-use possible?	Guidance: <i>e.g. Creative Commons, GPL, MIT, or GNU licenses</i>
	<i>[Individual answer according to your project]</i>
b) When will the data be made available for re-use?	Guidance: <i>If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>
	<i>[Individual answer according to your project]</i>
c) Are the data produced and / or used in the project useable by third parties, in particular after the end of the project?	Guidance: <i>If the re-use of some data is restricted, explain why.</i>
	<i>[Individual answer according to your project]</i>
d) How long is it intended that the data remains re-usable?	<i>[Individual answer according to your project]</i>

e) Are data quality assurance processes described?
<i>[Individual answer according to your project]</i>

3 Allocation of resources

a) What are the costs for making data FAIR in your project? Guidance: Please calculate the expected costs (software, human resources) for the research data management. <i>Note that the data management plan tool TUB-DMP is a free service of SZF and is maintained by the University Library of TU Berlin. For the web tool TUB-DMP no additional hardware or software is required. Furthermore, you have the possibility to publish your research data on the repository DepositOnce of TU Berlin without any costs for the service and without costs for the data storage (up to a certain maximum amount of data).</i>

[Individual answer according to your project]

b) How will these be covered? Guidance: Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).
--

[Individual answer according to your project]

c) Who will be responsible for data management in your project?

Text Suggestions (DepositOnce Standard Value):

According to the concept of the research data infrastructure of TU Berlin, the task sharing between the stakeholders is as following:

- * The researchers are responsible for the metadata management during the project as well as for the quality check of the research results: After having collected a large amount of data during the project they select those data that shall be preserved. In the submission process to DepositOnce they describe these data and upload the data files. Researchers are also responsible to create new versions of published submissions if necessary. The submission process of a new version is the same as the primary process.
- * The SZF-Team at the University Library is responsible for the quality check of the metadata and the formal check of the submitted research data: The uploaded data files are not published immediately but stored in an intermediate store. The SZF-Team checks e.g. whether the metadata fields are filled properly, whether the PDFs are in a valid PDF/A format, etc. If there are any questions, the SZF-Team contacts the submitter of the data. Accepted submissions are published by the SZF-Team and stored in DepositOnce. The SZF-Team is also responsible for the dissemination of the data.
- * tubIT is responsible for the IT infrastructure of TU Berlin which includes the safe storage of the data.

d) Are the resources for long term preservation discussed? Guidance: Costs and potential value, who decides and what data will be kept and for how long.
--

[Individual answer according to your project]

4 Data security

a) Is the data safely stored in certified repositories for long term preservation and curation?

Text Suggestions (DepositOnce Standard Value):

According to the directive for safeguarding good scientific practice of TU Berlin, the storage of the research data is guaranteed for 10 years. In cooperation with other partners and institutions (i.e. Zuse Institute Berlin and Kooperativer Bibliotheksverbund Berlin-Brandenburg), SZF is developing a concept for the long term preservation.

b) What provisions are in place for data security?

Guidance: Data security includes data recovery as well as secure storage and transfer of sensitive data.

Text Suggestions (DepositOnce Standard Value):

Operational concept of DepositOnce:

* DepositOnce uses the virtual server infrastructure of TU Berlin which is run by the IT Service Center tubIT. In this way, a high level of reliability and a high level of protection of the data and the application are guaranteed. All servers, networks and backup services used for DepositOnce are maintained by tubIT. The metadata as well as the research data are backed-up at least once a day in form of database dumps and files. A cronjob verifies file checksums once a week and ensures the data integrity.

Security concept of DepositOnce:

* The DepositOnce servers which store the research data and their metadata are part of the security concept of tubIT. They are located within the DMZ (Demilitarized Zone) of tubIT and are secured by several firewalls. The security concept strictly restricts any physical access to the data centre and any remote access to the servers.

* tubIT is responsible for the servers, network and backup services. The IT administrators at the University Library are responsible for the further enhancement and development of the software. They are also responsible for the recovery of the DepositOnce services out of the backup.

5 Ethical aspects

a) Are there any ethical or legal issues that can have an impact on data sharing?

Guidance: Ethical and legal issues can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the [Description of the Action \(DoA\)](#).

Text Suggestions (DepositOnce Standard Value):

All content which is stored in DepositOnce is anonymized according to the policy for safeguarding good scientific practice of TU Berlin.

Anonymization of personal data: When submitting a data file in DepositOnce the submitter confirms that the research data which he/she has submitted do not contain any personal data. If personal data are contained they are anonymized completely according to canonical standards and the human subjects have consented to the data collection as well as to the publication of the (anonymized) data.

Personal rights: When submitting a data file in DepositOnce the submitter also confirms that by submitting personal metadata (name and surname of the participating scientists) he/she acts in consent with all persons whose data he/she enters. He/she also confirms that he/she doesn't violate any personal rights by omitting the name of any participating scientist.

b) Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

[Individual answer according to your project]

6 Other issues

a) Do you make use of other national / funder / sectorial / departmental procedures for data management?

Guidance: If yes, please describe the procedures.

[Individual answer according to your project]

-
- ¹ The research data infrastructure and the research data management of TU Berlin are operated by the *Service Center for Research Data and Publications (SFZ)*. In SFZ, the University Library, the IT Service Center tubIT and the Research Department of TU Berlin cooperate to support the scientists of TU Berlin in all matters concerning their research results. All members of TU Berlin and, if applicable, their academic partners can use the tools and services of SFZ. More information about SFZ and its services: <https://www.szf.tu-berlin.de/>.
- ² Direct link to DepositOnce: <https://depositonce.tu-berlin.de>; test system: <http://testrepo.ub.tu-berlin.de>. For more information about DepositOnce see the FAQs: https://www.szf.tu-berlin.de/menue/faq/nutzung_von_depositonce.
- ³ Direct link to TUB-DMP: <https://dmp.tu-berlin.de>; demo system: <https://dmp-demo.ub.tu-berlin.de>. For more information about TUB-DMP see the FAQs: https://www.szf.tu-berlin.de/menue/faq/nutzung_von_tub_dmp.
- ⁴ Mark D. Wilkinson et al, *The FAIR Guiding Principles for scientific data management and stewardship*, Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016), <https://www.nature.com/articles/sdata201618>.
- ⁵ European Commission, *H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020*, Version 3.0, 26 July 2016, 4, last checked on 20.07.2018, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
- ⁶ European Commission, *H2020 Programme. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, Version 3.2, 21 March 2017*, 8ff., last checked on 20.07.2018, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.
- ⁷ See Footnote 5. The guidelines contain extensive information about DMPs and its handling in Horizon 2020 projects.
- ⁸ Ibid., 6–9.