Generic DMP Evaluation Rubric

This is an evaluation rubric for DMPs based on the generic DMP template of Ghent University. It is inspired by and based on the example rubric (CC0) from the DART project (https://osf.io/26b9r/), which was presented during the IDCC 2016 DMP Workshop in Amsterdam on 25 February 2016.

	PERFORMANCE LEVELS		
PERFORMANCE CRITERIA	**Complete/detailed**	**Addressed issue, but incomplete/too vague**	**Did not address issue**
1. *Describes what **types of data** will be collected, captured or created/generated in the project.*	Clearly describes data type(s), distinguishing between digital/non-digital data, quantitative/qualitative data, data formats (e.g. texts, spreadsheets, images, videotapes, audio files, surveys, databases, field notes etc.) and data collection methods (e.g. experimental, observational, simulation, derived/compiled… data). E.g.: *"This project will produce qualitative observational data from interviews and fieldwork conducted at various locations across Finland between January and June 2017. Raw data will comprise digital audio recordings of interviews, digital images and hand-written field notes. Audio files will be transcribed into digital text documents and notes will be digitized (via manual transcription) to prepare them for analysis."*	Some details about data types are provided, but not enough for someone outside the project to fully understand what sort of data are involved.	No details about data types are provided.
2. *Indicates whether the project will **generate its own data** or **reuse existing data**.*	Clearly states whether new data are generated or existing data are re-used in this project, and explains why. Also clearly specifies the source(s) of any existing data being reused.	States whether new data are generated or existing data are re-used, but does not explain why and/or provides no information about the source(s)	There is no indication of whether new data are generated or existing data are reused.

	E.g.: *"The project will reuse existing datasets on local mobility that are publicly available on the city of Ghent's Open Data Portal. These data will provide sufficient information to perform a quantitative analysis of the use of public transport services. However, to allow a qualitative analysis of local perceptions of public transport, the project will also generate new data, as there are no existing qualitative data about that aspect of our study."*	of the existing data being reused.	
3. Describes *what **file formats** will be used for the data.*	Clearly describes file formats in which data will be stored during data collection, processing and analysis, and explains why. E.g.: *"Spreadsheet data will be stored in the .sav format as they will be created and analysed with the statistical software program SPSS, which is widely used in the social sciences and for which our institution has a license."*	Only partially describes file formats that will be used for storing data and/or the rationale.	Does not describe file formats used for storing data and does not provide a rationale.
4. *If **unusual and/or proprietary file formats** are used during data collection, processing and analysis, the plan proposes solutions for converting data into file formats that are more suitable for long-term accessibility and use.*	Clearly explains how data files will be converted into more accessible/sustainable formats and any complicating factors (e.g. loss of information). If files will not be converted, describes how they will otherwise be made accessible to interested parties. E.g.: *"ATLAS.ti files will be converted to the open .xml format to facilitate sharing with collaborators across software, and to ensure that the files remain readable in the future."*	Vaguely or only partially explains how data files will be converted into more accessible/sustainable formats and any complicating factors, or vaguely describes how they will otherwise be made accessible to interested parties.	Does not explain how data files will be converted into to more accessible/sustainable formats or otherwise made accessible to interested parties.
5. *Gives an estimate of the **data volume** that is expected to be collected/generated.*	Expected total amount of data (MB, GB, TB…) is clearly stated. Information can be provided on how the researcher arrives at the estimate, such as a reference to the data volume of similar projects or other calculations.	Expected total amount of data is vaguely specified. E.g.: *"The project will produce a large amount of data."*	Expected total amount of data is not specified.

	E.g.: *"We expect that each station will collect around 300 KB of data per month, amounting to approximately 3.6 MB per station for the 12-month observation period. As there are 10 stations, the total amount of data produced is expected be about 36 MB."*		
6. *Identifies **methods** used for collecting, capturing or generating data.*	Clearly describes how data will be collected, captured or generated, including information about e.g. methodology, disciplinary standards or conventions, instruments/software/infrastructure, selection criteria for data sources, quality control measures used etc. where relevant. E.g.: *"Our archeological data will be generated from fieldwork in the Ave valley and further laboratory study of artefacts. An intensive survey of the entire valley will be followed by a more detailed survey of visible standing structures across the valley. Basic survey units will be mapped in the field with handheld GPS and satellite image print-outs; they will then be digitised and checked for overall consistency. Standing structures in each survey unit will be mapped in more detail using GPS and historical aerial photos. Artefacts collected during the surveys will be further studied in the laboratory and attributes will be entered in a database. Artefact descriptions will be based on the controlled vocabulary of the* Getty Art & Architecture Thesaurus*."*	Provides some details on how data will be collected, captured or generated, but not enough for someone outside the project to understand the methods used to obtain the data and get a sense of their quality. E.g.: *"Data will be generated by field sensors."*	Does not describe how data will be collected, captured or generated.
7. *Describes how data (files) will be **organised.***	Clearly describes the structuring system (e.g. folder and file hierarchy, database etc.), naming conventions, and version control strategies that will be used for organising data (files). E.g.: *"Each experiment will have its own folder containing the sub-folders "Data", "Documentation" and "Software". File names will include the project acronym, experiment name, any relevant specification of the file content, and a version number (e.g. ECG_Gamble10_Readme_v1). For raw data, there will be a single master file that will be left untouched; working*	Vaguely describes the structuring system, naming conventions, and version control strategies that will be used for organising data (files), or does not address all three aspects of data organisation.	Does not describe the structuring system, naming conventions, and version control strategies that will be used for organising data (files).

	copies will be created for processing and analysis. Git will be used to track changes in processing and analysis codes."		
8. *Identifies what **metadata standards** and/or other **data documentation** approaches will be used.*	Clearly describes what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, now and in the future (e.g. in terms of documentation levels and types required, procedures used, where this information is recorded…). Also indicates whether or not a (discipline-specific) metadata schema will be used, and explains rationale. *E.g.: "Descriptive metadata of data items will be captured in XML files in accordance with the EML schema, which is an international metadata standard for the field of ecology. In addition, datasets will be accompanied by a separate readme.txt file providing study-level documentation including the field methods used for data collection."*	Vaguely describes the approach that will be followed to capture the necessary accompanying information, and/or only includes a vague statement regarding the use of a metadata schema and/or its rationale.	There is no description of the approach followed to capture the necessary accompanying information, and there is no statement regarding the use of a metadata schema.
9. *Identifies whether **personal** and/or **otherwise confidential data** will be used.*	Clearly explains to what extent personal data and/or data that should otherwise be considered confidential will be used in the project. *E.g.: "The project will conduct a survey among students. Completed questionnaires and tabular data will only contain participant numbers and will not record any identifying information. I will hold a separately stored, non-public document linking participant numbers to students' real names for administrative purposes. Until this document is destroyed after the required 5-year retention period, the totality of the information I hold (research data and separate identifying information) will constitute personal data."*	Vaguely explains to what extent personal data will be used, and/or does not explain why data should otherwise be considered confidential.	Does not discuss whether personal and/or otherwise confidential data will be used.
10. *If personal/confidential data are used, the plan describes what*	Clearly identifies what permissions will be sought via informed consent forms (or otherwise, e.g. via ethical committee procedures) for handling data (including data collection, processing, archiving, and especially also sharing). Also clearly describes the measures that will be taken to safeguard	The plan is vague about what permissions will be sought for handling data, and/or about the measures taken to	Does not discuss what permissions will be sought for handling data, and does not

***permissions** will be obtained and what strategies will be used to **protect privacy and confidentiality**.*	privacy and confidentiality while storing and transferring data (e.g. anonymisation, access control, encryption…). *E.g.: "As the project involves human research participants, prior permission to collect, process, archive and share data will be sought from my faculty's ethical committee and, via informed consent forms, from participants. Files containing personal data will be stored in access-controlled folders on the university central disk space (with access restricted to project team members) to protect privacy. As stated in the consent forms, only de-identified data files will be shared more broadly and archived for the longer term with a trusted data repository."*	safeguard privacy and confidentiality.	describe measures taken to safeguard privacy and confidentiality.
11. *Identifies what **intellectual property rights** may apply to the data and accompanying documentation.*	For every dataset, clearly describes who owns any intellectual property rights in the data and accompanying documentation. Also indicates whether IPR issues might affect the ability to (re)use, archive and/or share data, and, if applicable, how these issues may be resolved. *E.g.: "My research will generate a database of photographs I made during fieldwork. I will own any copyright in the images, whereas the university will own any copyright and database rights in the database itself (as stipulated in the university's Valorization regulation). Therefore, I will seek permission from the university to make the database publicly available online."*	Vaguely or only partially describes who owns any intellectual property rights in the data and accompanying documentation, and/or whether IPR issues might affect the ability to re(use), archive and/or share data and how any issues may be resolved.	Does not describe who owns any intellectual property rights in the data and accompanying documentation, and whether IPR issues might affect the ability to re(use), archive and/or share data.
12. *If data will be externally shared (i.e. outside the own research team), the plan describes under what **license** they will be made available.*	Clearly explains how permissions (and conditions) for future reuse of data will be communicated (e.g. by using standardised licenses such as CC or ODC licenses). *E.g. "Data from the project that can be shared will be made available under a Creative Commons Attribution license (CC-BY 4.0), so that users have to give credit to the original data creators."*	Vaguely explains how permissions (and conditions) for future reuse of data will be communicated.	Does not explain how permissions (and conditions) for future reuse of data will be communicated.

13. *Describes what **storage and backup procedures** will be in place to prevent data loss.*	Clearly describes the locations, storage media and procedures that will be used for storing and backing up digital and non-digital data during research. *E.g.: "A primary copy of digital files will be stored on a shared university network drive. DICT is in charge of backing up this network drive. In addition, I will make my own daily backups of important files on an external hard drive. Paper-based data and documentation files will be stored in clearly labeled folders in my office cabinet. In addition, they will be digitally scanned as a backup."*	Only partially or vaguely describes locations, storage media and procedures used for storing and backing up digital and non-digital data during research.	Does not describe locations, storage media and procedures used for storing and backing up digital and non-digital data during research.
14. *Describes what **security measures** will be in place to protect data from unauthorised access and changes.*	Clearly describes the measures (in terms of physical security, network security, and security of computer systems and files) that will be taken to ensure that stored and transferred data are safe. *E.g. "Data will be handled in accordance with university information security guidelines, involving measures proportionate to their nature and the risks involved. Lab computers and external drives will be password-protected, and the rooms in which they are kept will be locked when no lab members are present. Security upgrades to operating systems will be regularly performed. Read-only permissions will be implemented for raw data files so that they cannot be edited."*	Only partially or vaguely describes the measures (in terms of physical security, network security, and security of computer systems and files) that will be taken to ensure that stored and transferred data are safe.	Does not describe the measures that will be taken to ensure that stored and transferred data are safe.
15. *Describes which data and accompanying documentation will be **retained** for preservation and/or sharing after publication or completion of the project.*	Clearly indicates which (versions of) digital and non-digital data and accompanying documentation will be retained or destroyed, and explains rationale. *E.g.: "To comply with journal data availability policies and enable verification of published claims and data reuse, de-identified data underpinning articles will be preserved after publication, together with associated documentation files, so that they can be made widely available for the long term. As stated in the consent forms, the raw data files containing personally identifying information will be retained for a limited period only to allow for internal scientific integrity review by authorized*	Vaguely or only partially describes what data and accompanying documentation will be retained or destroyed, and/or the rationale.	Does not describe what data and accompanying documentation will be retained or destroyed, and does not provide a rationale.

	persons in compliance with applicable ethical and university regulations; after this retention period, however, they will be securely destroyed. Pilot data will not be retained either, as they have little reuse value."		
16. *Indicates **how long** data and accompanying documentation will be preserved.*	Clearly specifies the intended preservation period for all of the retained data and accompanying documentation (e.g. 5 years, 10 years, long-term…), and provides a motivation. E.g.: *"Archived data and documentation files will be preserved for 10 years after the end of the project, as required by the project funder."*	Vaguely or only partially specifies the intended preservation period for retained data, and/or the motivation. Or the intended preservation period can be inferred because the plan indicates that data will be deposited in a data repository or archive.	Does not specify the intended preservation period for retained data, and does not provide a motivation.
17. *Indicates **how data and accompanying documentation will be archived** during the specified preservation period.*	Clearly describes what means, facilities etc. will be used for effectively preserving digital and non-digital data and accompanying documentation, including details about any suitable data repository that has been identified and its requirements for data deposit (e.g. accepted file formats, required documentation/metadata, accepted file sizes). Where no trusted data repository will be used, explains how effective preservation will be ensured. *E.g.: "Raw data and documentation files will be offered for deposit to 4TU.ResearchData, which is a DSA-certified data repository accepting research data in the field of engineering and preserves them for a minimum of 15 years. Files will be offered in the repository's preferred formats (.txt, .xml and JCAMP), and as the volume of data does not exceed 10GB the repository will not charge for the deposit."*	Vaguely or only partially describes what means, facilities etc. will be used for effectively preserving digital and non-digital data and accompanying documentation, and/or does not explain how effective preservation will be ensured where no trusted data repository will be used.	Does not describe what means, facilities etc. will be used for effectively preserving digital and non-digital data and accompanying documentation.

18. *Identifies whether there are **factors limiting the ability to externally share** (i.e. outside the own research team) data and accompanying documentation.*	Clearly states if the external sharing of (some) data and accompanying documentation should be restricted or delayed, and if so, why. *E.g.: "Because my research data are part of a potentially patentable invention, sharing will need to be delayed to investigate patent protection first. Before any disclosure can take place, my research results will need to be reported to the university's TechTransfer office, which will determine whether releasing data will need to be embargoed until after a patent application has been filed."*	Vaguely indicates whether the external sharing of (some) data and accompanying documentation should be restricted or delayed, and/or does not explain why if that is the case.	Does not state if external sharing of (some) data and accompanying documentation should be restricted or delayed.
19. *Describes **how** data and accompanying documentation will be externally shared (i.e. outside the own research team).*	Clearly describes when (some of) the data and accompanying documentation will be made available to people outside the project, under what conditions, and by what means (e.g. via a data repository, project website, upon request…). If data will not be shared via a trusted data repository, explains how long-term access will be ensured. E.g.: *"My dataset will be made available upon publication of the associated journal article to approved users via the Zenodo repository. Users wishing to download the data from Zenodo will have to request permission. Metadata for the dataset will also be made publicly available via my institutional repository (together with a link to the research data record in Zenodo) as required by my institution's research data policy, and to further aid discovery."*	Only partially or vaguely describes when (some of) the data and accompanying documentation will be made available to people outside the project, under what conditions, and by what means, and/or does not explain how long-term access will be ensured where data will not be shared via a trusted data repository.	Does not discuss how (some) data will be externally shared.
20. *Identifies the individual(s) **responsible** for maintaining the data management plan, and for implementing data management for the*	Clearly lists the person(s) responsible for the DMP and for overseeing its day-to-day implementation, and – if multiple individuals are involved – also their role. E.g.: *"The project's principal investigator is responsible for developing and updating the DMP, and for overseeing its implementation by the project team. Each team member is responsible for managing the active data from*	The plan is vague about the person(s) responsible for the DMP and for overseeing its day-to-day implementation, and/or – if multiple individuals are involved – about their role.	Does not discuss who is responsible for the DMP and for overseeing its day-to-day implementation.

project on a day-to-day basis.	*the experiments he/she conducts in the manner outlined in the DMP (i.e. data collection, documentation, storage & backup). Team members are also responsible for depositing data underpinning publications on which they are the lead author in the agreed data repository, so as to make the data publicly available. Preserving and providing long-term access to the project's data will be outsourced to the repository."*		
21. *Describes what happens to the data **if a researcher involved in the project leaves** his/her department or the university.*	Clearly describes what arrangements are in place regarding data access, preservation and use in the event of a researcher leaving. *E.g.: "Team members joining the project will have to sign a declaration outlining what happens with data in the event of their departure. The principal investigator will retain access to any of the departing team member's active project data and associated documentation files via the group's shared network drive, and will retain the right to use them for further research. The PI will also be responsible for selecting project data for preservation and for archiving them after project completion. Since the project does not involve personal or confidential data, departing team members will be allowed to take a copy of the project data they collected and use them for further research. However, they will need to notify the PI before publishing their findings from the data or the data themselves, and properly acknowledge him. In turn, the PI will need to acknowledge the departing team member as the original data creator in any publications."*	Vaguely describes what arrangements are in place regarding data access, preservation and use in the event of a researcher leaving.	Does not describe what arrangements are in place regarding data access, preservation and use in the event of a researcher leaving.
22. *Indicates whether **resources** are needed – beyond what the university already provides – to implement the DMP.*	Clearly states whether any additional resources will be required to execute the DMP (e.g. for external storage, hosting, specialist training, etc.), and if so, provides details. *E.g.: "I will need discipline-specific training on using the DDI metadata schema to document my research data. Therefore, I plan to attend a DDI primer workshop organised by an established social science data centre, such as ICPSR or GESIS-DAS."*	The plan is vague about whether any additional resources will be required to execute the DMP, and/or provides no details of the additional resources required.	Does not state whether any additional resources will be required to execute the DMP.