Arts and Humanities Research Council Data Management Plan Rubric

Authors:

Rosie Higman, rosie.higman@manchester.ac.uk, University of Manchester

Mary Donaldson, mary.donaldson@glasgow.ac.uk, University of Glasgow

Purpose:

This rubric is designed as a checklist or marking aid for those reviewing data management plans for submission to the Arts and Humanities Research Council (AHRC). The Data Management Plan should outline the project's approach to managing data. It is mandatory to include for all Leadership Fellows, Research Grants and Follow on Funding applications but is not required for Research Networking.

Documents used:

AHRC data management plan text for funding guide

Digital Curation Centre guide to the AHRC policy

AHRC Research Funding Guide

AHRC Peer Review College Handbook

AHRC FAQs on new data management plan template

AHRC Peer Review news and updates

The draft rubric is available at:
https://docs.google.com/spreadsheets/d/1QaNrXZYWQG4fi1MQVxsxbDQsx5B4i79hE7RCC9dZwM4/edit?usp=sharing

Performance Criteria		Performance Levels		
		Detailed	Incompletely addressed	Not addressed
Section 1	What types of data will be created?	Clear descriptions of the types of data being created with details of formats and approximate dataset sizes where appropriate.	Some description of the data being created but it is unclear or incomplete (based on subsequent answers)	Minimal information about what data types are being created.
Section 1	Why are these data types being proposed?	The plan explains why the particular data type and format is being used and how it will contribute to the project/answering the research question.	Some mention of why these data types are being used but it is not clear how they will contribute to the project.	No explanation given for why the proposed data types are required for the research or how they will aid in answering the research questions.
Section 2	What methodology is being proposed?	Methodologies are described clearly for each data type and information given about any intermediate data type produced before the final data.	Methodologies are described but either not clearly or not for all data types outlined in section 1.	Minimal description of the methodologies being proposed
Section 2	How is this methodology suited to the data types or digital outputs?	As well as linking methodologies to the data types the plan explains how the methodology will produce the data type.	It is clear which methodology is related to each data type but not necessarily why that methodology suits a data type	No mention of how the methodologies proposed connect to the data being collected or created.
Section 2	How is the project team suitable for the digial/data aspects of the work?	The plan details the skills necessary to deliver the data and digital aspects of the project, and explains how the project team either deliver these skills, or will be trained to do so (eg transcribers will be trained in TEI/XML).	The plan mentions skills necessary to deliver the project, but does not indicate how the project team deliver these skills. OR the plan mentions skills necessary but does not cover the full range of the project.	There is no mention of the digital or data skills needed to deliver the project, or how the project team will provide these skills.
Section 2	How will the institution's support teams be supporting these methodologies?	Where methodologies use specialist techniques, equipment or processes it is clearly described how the research will be supported in implementing their proposal.	There is some mention of institutional support for the methodologies involved in the project, but it is not clear if the support covers all the specialist methodologies involved in the project.	No mention of how the researcher will be supported in using these methodologies by their institution

Performance Criteria		Performance Levels		
		Detailed	Incompletely addressed	Not addressed
Section 3	Has an appropriate storage solution for the duration of the project been described?	The data storage for the project is clearly described, covers all the data to be stored and is suitable, so far as it is possible to judge. The plan may also reference institutional storage policies or pages.	There is some description of the data storage solution the project will use but it is not clearly described or does not cover all of the data being produced.	It is not clear where the data will be stored during the project or the storage solution proposed is inappropriate.
Section 3a	Is the backup process described appropriate?	The back-up process for active data storage is clearly described or referenced, and is appropriate for the data to be collected.	Data backup is mentioned, but no detail, or link to institutional policy is provided. Back-up process described might be inadequate for the data being collected and stored.	No backup process is described or the one described is inappropriate or inadequate for the data that is being collected.
Section 4a	Has an appropriate long-term storage solution been described?	The long-term storage plan for the data is described. This might be a repository or other appropriate solution. The solution(s) identified cover all the data to be retained.	A long-term storage plan is mentioned, but detail may be lacking or the solution(s) identified may not cover all the data to be retained.	No long-term storage plan is mentioned, or the solution proposed is inappropriate, either for the data to be retained, or does not comply with funder requirements.
Section 4b	How long will the data be stored for and is this appropriate to the project?	The long-term retention schedule is described for all data. The retention period is appropriate to the data and in keeping with any consent from participants.	The long-term retention schedule is mentioned, but may not cover all data or may be inadequate or inappropriate.	No long-term retention schedule is mentioned.
Section 4c	Has long-term storage costs been described in the plan?	The costs for the long-term storage are clearly described. Alternatively, it is stated that the data will be stored for the long-term in a repository with no ingest costs.	Cost for long-term storage are mentioned, but no detail. Costs may not appear to cover all the data. Costs may appear to be inappropriate for the storage option indicated.	No costs for long-term storage are mentioned.
Section 5a	Has the value of the data to the disciplinary area been outlined?	The value to all relevant disciplinary areas have been clearly outlined for each data type. Consideration has been given to the different types of value data can provide and these are described appropriately. It is	The value of the data to the disciplinary area is mentioned but it is not clear or may be poorly explained. There may be missing details about which disciplines may benefit from this	The value of the data to the disciplinary community is not mentioned.

Performance Criteria		Performance Levels		
		Detailed	Incompletely addressed	Not addressed
		clear how this value will be facilitated by data release and/or collaborations.	data being created/shared. There may be data types from section 1 which are not mentioned in this section.	
Section 5a	Have possible future uses of this data been described?	Future uses of this data have been clearly described with consideration given to the different possible audiences for this data. The descriptions of the data uses are clear, specific (as much as possible given the research has yet to happen), and realistic. Any limitations on future uses of the data are described and justified.	It is noted that the data can be re-used but there is little or no detail about these possible uses. Where future re-uses are mentioned they are only vaguely described. Where there are limitations on future uses of the data these are not well described or justified.	No possible future use of this data have been described nor reasons given why the data cannot be reused.
Section 5b	When will the data be released?	The point at which data will be released is clearly identified. This might be a date (in relation to the lifetime of the project) or triggered by an event such as the publication of major findings.	It is noted that the data will be release, but no clear timeframe / trigger event is identified.	No indication is given as to when the data will be released.
Section 5b	If the data will not be released within 3 years has this been justified?	If necessary, reasons for delaying the release of the data are described and are justified. This question may not be addressed if there are no plans to delay release of the data.	Reasons for delaying release of the dataset are given, but have not been justified or may not be reasonable (with respect to the funder requirements).	The plan indicates that the release of the dataset will be delayed, but no reasons or justification are given.
Section 5b	How will different audiences be made aware of this data?	Discovery routes for the dataset are described. Eg. data availability statements in associated papers; statement of availability on the project website; deposit in a repository which is searchable / aggregated on a discovery portal etc.	It is noted that the data will be discoverable, but details are not given.	No mention of data discoverability is made.

Performance Criteria		Performance Levels		
		Detailed	**Incompletely addressed**	**Not addressed**
Section 5c	Will the data require future updating?	It is clearly stated whether or not the data will require updating in the future and why this is the case.	There is a mention of updating the data in the future but it is vague and not clear why this is needed.	It is not explained whether or not the data will need to be updated in there future.
Section 5c	Has an appropriate plan for updating the data been described?	There is a clear plan for updating the data in the future described including the resources required, frequency of updates, and point at which the updates will no longer be required. The resources, frequency of updates and eventual end date for updates are appropriate, costed and realistic.	A plan for updating the data is described but this is vague (lacking mention of resourcing and frequency of updates) or inappropriate for the project with unrealistic resourcing suggested	There is no plan for updating the data.
Section 5d	Is there a clear explanation of whether the data will be open or charged for? Where charges have been proposed have these been justified?	The license and charging model (if applicable) for the dataset are clearly described and justified. Charges for different groups have been outlined where appropriate and the justification recognises the likely audiences for the dataset (section 5b) and their ability to pay. The charges proposed are sufficient to maintain the dataset and there is a clear mechanism described for collecting these charges.	There is some explanation of a charging model or open licence for the dataset but this is not clear and the reasons behind this decision have not been explained. Where there is some justification of a charging model this is incomplete, unclear or fail to consider the audience this dataset will be used by.	There is no mention of charging models or open licences for this dataset. Where charges have been mentioned there is no justification of these.
Section 5e	Have the costs of sharing been described?	The researcher has considered what the costs of sharing will be, including storage, preparing the data to be shared and gaining permission from any other parties which have IP in their dataset. These costs are clearly described and included in the grant.	The costs of sharing have been outlined but there is little detail about what these costs will be or evidence behind the costs suggested.	There is no reference to the costs of sharing the data.

Performance Criteria		Performance Levels		
		Detailed	Incompletely addressed	Not addressed
Section 6a	Have any legal or ethical issues in collecting the data been described and addressed?	Legal and/or ethical issues around data collection are described. Solutions / mitigations for these issues are presented. Issues and proposed solutions are appropriate and in line with funder and legal requirements. Alternatively, these issues are not discussed, and the reviewer feels the project is not likely to incur these issues, or the author states that the project does not incur these issues. Information about dataset licensing may also be included here.	Legal and/or ethical issues around data collection are described, but may not cover all of the data to be collected. Solutions are presented, but may not be appropriate, or solutions may not be proposed.	Legal and ethical issues around data collection are not mentioned and the reviewer feels that the project is likely to incur these issues.
Section 6b	Have any legal or ethical issues in storing and sharing the data been described and addressed appropriately?	Either, as indicated in response to the previous question, this project will not incur these issues, so this is not discussed, or there is a discussion of how the legal / ethical issues around storing data will be addressed. These might include participant consent for long-term storage of data; participant consent for sharing of anonymised data; permission to store or share third party data; withholding sharing of a subset of data for which appropriate consent could not be secured.	Legal and ethical issues around storing and sharing data are mentioned, but either don't consider all of the data to be stored / shared, or don't propose adequate solutions / mitigations.	Legal and ethical issues around data storage and sharing are not mentioned and the reviewer feels that the project is likely to incur these issues.