

ADR UK Flagship Dataset Documentation Guidance



MINIMUM STANDARDS FOR DOCUMENTATION

Date	October 2024
Author	Karen Powell, with significant contributions from across the ADR UK partnership
Sign-off	Operational Management Group
Version	V1.0

Document circulation / Readership

This guidance is for anyone producing documentation for ADR UK flagship datasets. This document is findable on the ADR UK website on the ADR UK flagship dataset page.

The intended internal circulation or readership for this document is as follows:

- ADR UK Internal mailing list
- ADR UK Operational Management Group
- ADR UK Communications Steering Group
- ADR England Data Linkage Steering Group.

Future updates

This document will be reviewed in March 2026 if not before. If you have any feedback that you'd like to share to improve future iterations of this document, please use this [online form](#).

Definition of terms

The terms listed in the table below have the associated definition within this document. See also the [ADR UK Glossary](#).

Term	Definition
User Guide	A guide written for potential users of a dataset (researchers, analysts) which (together with the data dictionary) provides enough information about the dataset that they may write a robust and feasible research application to use the data.
Data Dictionary	Information that describes key characteristics of a dataset, particularly the variables. The information can be used by researchers to write an application to use the dataset for a research project.
Metadata	Information that describes a dataset or a data item. This includes information that provides context to the data, such as how it was collected, coverage of the data, publication date, as well as data quality dimensions, and data characteristics such as summary statistics.
ADR UK flagship dataset	Linked or linkable administrative datasets (either administrative linked to administrative data or administrative linked to other data) of significant research value and that we anticipate will have wide appeal to researchers. The flagship datasets are findable in the ADR UK Data Catalogue and available for all accredited researchers to apply to access via one of ADR UK's trusted research environments. They have approved legal gateways for research and are publicly documented.

Table of Contents

1	Minimum standards of documentation.....	3
2	Document guidance	4
2.1	User Guide	5
2.2	Data Dictionary	8
3	Appendix 1: ADR UK publication types	8

1 Minimum standards of documentation

These minimum standards of documentation must be met for a dataset to be promoted as an ADR UK flagship dataset. By promoted we mean to be featured in the ADR UK website, funding opportunities, newsletters, printed materials, etc.

- i. **All ADR UK flagship datasets must have a user guide and data dictionary** (or equivalent documents with different names).
- ii. The documents should give potential research-users of the data **sufficient information about the data to write robust and feasible research applications** to use the data, as detailed in the following guidance.
- iii. **These documents will be publicly available** and clearly signposted to on the ADR UK website, and in the data catalogue.
- iv. The equivalent documentation **will also available inside the secure research environment**.
- v. The documents should be **updated with every new version or refresh of the data**.

The following guidance is for minimum standards of documentation going forward. It is anticipated it will take time for all ADR UK flagship dataset documentation to reach the minimum standard. There isn't a requirement to retrospectively overhaul existing documentation.

2 Document guidance

The following guidance can be employed with some flexibility and modification where appropriate, for example:

- flexibility in the scope, (e.g., focused on a single source dataset that is 'linkable')
- flexibility in the format or type of the document (e.g., print or digital).

The expectation is that this guidance should be followed wherever possible.

Users in a trusted research environment will not have internet access and will not be able to access hyperlinks. When using hyperlinks, please consider requesting essential documentation that is linked to, is also uploaded to the secure environment.

2.1 User Guide

	Sub-headings in User Guide	Suggested content
1	Version control and contacts	<p>Version control of the document including date of creation or update, amendments made and corresponding version of the dataset.</p> <p>Document ownership e.g. organisation name and contact information for any researcher queries.</p>
2	Introduction	<p>Short introduction including:</p> <ul style="list-style-type: none"> • overview of the dataset • purpose and content of the document • intended audience • motivation for the linkage: summary of research potential of the data, policy needs, links to areas of research interest, comparable datasets.
3	About the component data (individual source data)	<p>Information about the component datasets and what data is included from them such as:</p> <ul style="list-style-type: none"> • data owners and data collection details • summary of the scope of data in terms of years, geographical coverage and missingness • summary of known exclusions, caveats and limitations • link to the list of variables and their description (e.g. data dictionary). <p>If available include:</p> <ul style="list-style-type: none"> • links to further information about caveats and limitations (e.g. methodology report) • information about how researchers or data users can provide feedback on their experience of data • information about relevant community groups and how to join them. • A sample of example (artificial) records to help researchers understand the data.

4	Data linkage methodology	<p>Information about how the data was linked including:</p> <ul style="list-style-type: none"> • legal gateway for the linkage • who did the linkage • linkage methodology (e.g. deterministic, probabilistic) and details of algorithms and their settings (or a link to these details). See example of probabilistic indexing match report. • linkage key; linkage spine information • linkage quality and linkages rates • donor imputation (if relevant) • strengths and limitations of the linked/linkable data asset • summary of the gaps in coverage, linkage quality and variable specific limitations • link to the data dictionary. <p>If 'linkable' data: provide detail on how the data can be linked e.g., indexes, or common identifiers.</p>
5	Research potential (if available)	<p>If there is additional information about the research potential of the data beyond what is included in the introduction include a dedicated section to describe what research is possible with the data.</p>
6	Accessing the data	<p>Information about how to access the data and associated governance including:</p> <ul style="list-style-type: none"> • application process and approvals required as part of the legal gateway • any specific information about accessing the data that is helpful to the user: details of the trusted research environments, conditions of use and access options including SafePods and remote access • output clearance policy and details of any acknowledgements required • ethical considerations: restricted variables; ethical approvals required and responsibilities • contact details of the data owner or producer (or other nominated expert contact) for the research to access additional support and ask queries about the dataset. • information on who is the resource creator and data controller. <p>If known, whether it is possible for researchers to request the data to be linked to other datasets (optional).</p>

7	<p>Appendix with links to further information (publicly available user guide only) plus list of hyperlinked documents to be uploaded to the secure environment.</p>	<p>Any additional materials that provide more detail about the data such as:</p> <ul style="list-style-type: none"> • links to the data dictionary and metadata catalogue entry • (artificial) example records that allow researchers to see what data looks like • academic journals or other published research outputs • further reports about the data quality or emerging limitations e.g., from researchers • repositories of code used by other researchers • links to published statistics based on the data (if applicable) • information available inside the trusted research environment e.g., methodology reports. <p>As researchers in the trusted research environment won't have internet access, please highlight which of the above information, should be made available alongside the user guide within the secure environment.</p> <p>Provide contact details so the researcher can request that additional documentation is uploaded into the secure environment if they notice that the publicly available hyperlinks above do not match the documentation within the secure environment.</p>
---	---	---

2.2 Data Dictionary

The content in the below table may be covered by a complete entry on a metadata catalogue, and this would be considered sufficient. There isn't an expectation of an additional document.

	Sub heading in data dictionary	Content
1	Version control	
2	Dataset information	<ul style="list-style-type: none"> • Name (using agreed naming conventions where applicable) • Dataset version number and time period • Description • Collection methods • Frequency of collection • Coverage: spatial coverage; geographical levels; start and end dates - associated to dataset versions/waves • Distribution release date • Mode • Permissions • DOI name and link/Citation
3	Source file information	<ul style="list-style-type: none"> • Name • Size • Frequency • Coverage <p>This section is distinct from the dataset information section and should refer to distinct files/tables within a dataset.</p>
4	Variable catalogue	<ul style="list-style-type: none"> • Variable names • Descriptions • Types • Values • Length • Summary statistics e.g. minimum and maximum values for numeric variables or number of categories for categorical variables • Any data quality aspects such as completeness or validity • Notes (plus, lookups) • Restricted variables • Limitations, caveats specific to each of the variables • Links to further information • Link to relevant section within the guidance document

3 Appendix 1: ADR UK publication types

The following table illustrates the publications ADR UK currently supports. Please note the differences between User Guides and Data Explained.

Document name	Author	Intended audience	Purpose
User Guide	Those responsible for the dataset curation and/or linkage.	Dataset users or potential users such as researchers or analysts	Provides enough information about the dataset that (with the data dictionary), researchers may write a robust and feasible research application to use the data.
Data Dictionary	Those responsible for the data curation or linkage.	Dataset users or potential users such as researchers or analysts	Describes the key characteristics of a dataset, particularly the variables, to help researchers write an application to use the dataset.
Data Explained	A researcher who has used a dataset for a research project.	Dataset users such as researchers or analysts Data owners	Summarises experiences and learning from working with a dataset to: - help guide future researchers using the data - provide feedback into future dataset development and documentation.
Data Insight	A researcher who has used a dataset for a research project.	Policymakers or practitioners	Summarises the research done, what the researcher found, and why it matters to influence policy and practice.