Making Data FAIR

Anca Vlad
Research Data Services Team
UK Data Service

Managing and sharing research data for transparency and FAIRness

24 January, 2020
Overview Part 1: Documentation & Metadata

- FAIR Principles
- Is data Findable?
- Is data Accessible?
- Is data Interoperable?
- Is data Reusable?
- Exercise – 20 minutes
FAIR principles for repositories

Findable

Accessible

Interoperable

Re-usable

https://www.force11.org/group/fairgroup/fairprinciples
Is data Findable?

The dataset should have a unique, permanent “address” that would allow it to be found using discovery portals. This is referred to as a persistent identifier and is normally part of that dataset’s citation.


Another aspect that can make a dataset more findable is rich, machine readable metadata that describes the dataset and is used to index datasets in data catalogues and allow them to be found efficiently through discovery portals. Example: UKDS Data Catalogue, CESSDA Data Catalogue.

Key: Does the dataset have a persistent identifier?
Documenting your data

• Enables you to understand the data if/when you return to it.
• Sufficient information for future researchers to understand and use the data

• If using your data for the first time, what would a new user need to know to make sense of it?

• The UK Data Archive uses data documentation to:
  • Supplement a data collection with documents and research instruments
  • Ensure accurate processing and archiving
  • Create a catalogue record (metadata) for a published data collection
Include as documentation

- Data collection methodology and processes: sampling, sample size, fieldwork protocol, experiment protocol, interviewer instructions
- Codebook, user guide (for quantitative data)
- Information sheet, consent form (blank versions)
- Questionnaires, show cards, topic guides
- Transcripts: header with context information: data and place of interview, interviewer, interviewee details (in line with consent form) etc.
- Data list: overview of key information about each interview, a map of the data collection (for qualitative data)
- Links to reports and publications (preferably DOIs where possible)
Data-level documentation

- All structured, tabular data should have adequate variable names, variable and value labels
- Variable names might include:
  - Question number system matching questions in the questionnaire used e.g. Q1a, Q1b, Q2, Q3b
  - Numerical order system e.g. V1, V2, V3
  - Meaningful abbreviations or combinations of abbreviations referring to meaning of the variable e.g. ‘oz%=percentage ozone’, ‘GOR=Government Office Region’, ‘moocc=mother occupation’
  - For interoperability across platforms, variable names should not be longer than 8 characters and without spaces
Data-level documentation

Similar principles for variable labels:
• Be brief, maximum 80 characters
• Include unit of measurement where appropriate
• Reference the question number of a survey or questionnaire
e.g. variable ‘q11hexw’ with label ‘Q11b: hours spent taking physical exercise in a typical week’ – the label gives the unit of measurement and a reference to the questions number (Q11b)
• Coding or classification schemes used, with a bibliographic reference
e.g. Standard Occupational Classification 2000; ISO 3166 alpha-2 country codes

For value labels:
• Codes of, and reasons for, missing data
• Avoid blanks, system missing or ‘0’ values e.g. ’99= not recorded’, ‘98= not provided (no answer)’, ‘97=not applicable(skipped)’, ’96= not known’, ’95=error’
Organising data

• Plan in advance how to best organise data (project specific)
• Use a logical structure and ensure collaborators understand

Examples
• Hierarchical structure of files, grouped in folders e.g. audio, transcripts and annotated transcripts
• Survey data: spreadsheet, SPSS, relational database
• Interview transcripts: individual well-named files
Is the data Accessible?

Can be open but not necessarily.

Some data needs to be placed under access restrictions due to privacy concerns, consent agreements, disclosure risk or commercial interests.

Key: access should be implemented using a standardised protocol -
   - terms and conditions governing access and reuse should be clear, standardised and transparent.
Managing access to data

Open
- available for download/online access under open licence without any registration

Safeguarded
- available for download / online access to logged-in users who have registered and agreed to an End User Licence (e.g. *not identify any potentially identifiable individuals*)
- special agreements (depositor permission; approved researcher)
- embargo for fixed time period

Controlled
- available for remote or safe room access to authorised and authenticated users whose research proposal has been and who have received training
Is data Interoperable?

Datasets are Interoperable if they are machine readable (metadata) and they are in specific formats, language and vocabularies and/or ontologies.

*Digital data is software dependent, so endangered by obsolescence of software/hardware.*

Formats used should be:

- community agreed (vary across disciplines)
- open (as opposed to proprietary)
- unencrypted
- suitable for long-term preservation

The metadata will also need to use a community agreed standards and vocabularies (such as the [DDI Schema](#)), and contain links to related information using persistent identifiers.
Machine-readable metadata (xml) for qualitative data

<text xml:lang="en">

Where were you born?

What was your parent's house like?

Well I think they had two furnished rooms in those days or one then they moved to another place. They always had about two rooms. It was only small, myself and two brothers. I was the eldest. They were business people, they had a business, they acquired a business in Drury Lane.

What sort of business was it?

It was a hairdressing business and then they had one also in Charing Cross Road and then came back bought another business in what is now Bulstrode Street, in High Street, Marylebone. So we were just a bit above the ordinary class you see.

How old were you ... when you moved?

In Drury Lane I still think I was, a little child of about 4 or 5 I should imagine. I may have been a bit older because my brother was born then.

How old were your brothers?

One is six years and one has just passed away so I think he is just about eight years younger than me. They went to school. When we moved they went to a very nice school in Castle Street connected with St. Martins in the Fields a sort of little subsidiary school something like that where they went a little while

</text>
Machine-readable data elements: Qualibank

Qualibank demo: https://discover.ukdataservice.ac.uk/QualiBank

How data collections are included in Qualibank
File formats

File format examples:

• structured, open standard, machine-readable format e.g. (text) PDF/A, HTML, Plain text, (images) TIFF, JPEG 2000, GIF, (audio) MP3, AIFF, WAVE, (video) MOV, MPEG, AVI, (Tabular data) CSV

• structured, open standard, non-machine-readable format, e.g. PDF, HTML, JPG

• proprietary format, e.g. doc (Word), .xls (Excel), .ppt (PowerPoint), .sav

Best formats for long-term preservation:

standard, interchangeable and open

UK Data Service optimal file formats for various data types

Digital Preservation Coalition guidance on preservation formats
Is the data Reusable?

- Does it have accurate and relevant attributes and provenance information? (machine readable metadata or text format)
- Does it meet domain-relevant community standards to allow it to be reused?

- Licencing* - Stating clear re-use rights is like having a warm 'Welcome' on the doormat of your dataset.

  To make re-use as likely as possible, choose a licence which:
  - makes data available to the widest audience possible
  - makes the widest range of uses possible

*More on Licencing in Part 2
Exercise: FAIR data

• As part of this exercise, we will ask you to assess the FAIRness of a dataset.

• Use the four FAIR criteria to decide if the dataset is Findable, Accessible, Interoperable and Reusable and what are possible solutions.

• You can use our question list or an online, self-assessment tool:
For exercise: Interoperable section – What format is the data available in?

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Packed Size</th>
<th>Modified</th>
<th>Created</th>
</tr>
</thead>
<tbody>
<tr>
<td>BG.UK.PR.12.M.23 transcript.rtf</td>
<td>156912</td>
<td>23680</td>
<td>2018-07-14 18:30</td>
<td></td>
</tr>
<tr>
<td>BG.UK.PR.1.F.54 transcript.rtf</td>
<td>191307</td>
<td>25927</td>
<td>2018-07-14 18:27</td>
<td></td>
</tr>
<tr>
<td>BG.UK.AC.13.F.47 transcript.rtf</td>
<td>439027</td>
<td>57060</td>
<td>2018-08-22 19:12</td>
<td></td>
</tr>
<tr>
<td>BG.UK.AC.11.F.35 transcript.rtf</td>
<td>239232</td>
<td>32800</td>
<td>2018-07-14 19:26</td>
<td></td>
</tr>
<tr>
<td>BG.UK.AC.8.F.38 transcript.rtf</td>
<td>572397</td>
<td>60729</td>
<td>2018-07-15 14:52</td>
<td></td>
</tr>
<tr>
<td>BG.UK.AC.5.F.60 transcript.rtf</td>
<td>493826</td>
<td>53897</td>
<td>2018-07-15 16:25</td>
<td></td>
</tr>
<tr>
<td>BG.UK.AC.3.F.28 transcript.rtf</td>
<td>583370</td>
<td>28492</td>
<td>2018-07-15 14:20</td>
<td></td>
</tr>
</tbody>
</table>
Exercise: file formats
Overview Part 2: Preparing and publishing data

Prepare to deposit data
Access
Licencing
Repository options
File formats exercise
ReShare Demo (10 min)
How to prepare for sharing your data?

To prepare your data for future deposit and sharing:

• capture information and documentation (metadata) during the data collection process that will allow understanding of your data, such as variable descriptions, survey questions, data collection methodology and information

• check, validate and clean your data

• ensure you are organising, naming and versioning data files meaningfully

if data contain personal or confidential information, prepare to gain participant consent to share data with future researchers and create where possible an anonymised version
Documentation

Planning ahead will save you time and help keep things organised. A useful exercise is to think about the information that a stranger to the project would need in order to understand, replicate or reuse the data.

Planning to archive the data at the end of the project and where in particular can be useful to know in advance, as guidance/formats/metadata standards can vary across repositories.

- Project-level documentation includes information about the study, what were the main research questions, type of data was collected to answer these questions.
- Data-level documentation includes information at the level of individual data files, such as an interview transcript or a particular variable in a dataset.
- Metadata: ‘Data about data’ – machine readable information needed to catalogue and discover the data (DDI compliant).
## Anonymisation

<table>
<thead>
<tr>
<th>Quantitative</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remove direct identifiers</td>
<td>Remove direct identifiers, or replace with pseudonyms</td>
</tr>
<tr>
<td><em>e.g.</em> names, address, institution, photo</td>
<td>• Avoid blanking out; use pseudonyms or replacements</td>
</tr>
<tr>
<td>• Reduce the precision/detail of a variable through aggregation <em>e.g.</em> birth year instead of date of birth, occupational categories rather than jobs; and, area rather than village</td>
<td>• Identify replacements with [brackets]</td>
</tr>
<tr>
<td>• Generalise meaning of detailed text variable <em>e.g.</em> occupational expertise</td>
<td>• Plan or apply editing at time of transcription</td>
</tr>
<tr>
<td>• Restrict upper lower ranges of a variable to hide outliers <em>e.g.</em> income, age</td>
<td>• Consistency throughout project</td>
</tr>
<tr>
<td>✅ Keep an anonymization log</td>
<td>• Avoid over-anonymising – removing information in text can distort data, make them unusable, unreliable or misleading; so balance anonymisation with the need to preserve context</td>
</tr>
<tr>
<td></td>
<td>✅ Keep an anonymization log</td>
</tr>
</tbody>
</table>
Data list example

<table>
<thead>
<tr>
<th>Interview ID</th>
<th>Farmer code</th>
<th>Age</th>
<th>Farm Scheme</th>
<th>Farm type</th>
<th>Size of Farm (hectare)</th>
<th>Number of Holdings</th>
<th>Date of Interview</th>
<th>Interviewer Name</th>
<th>No of Pages</th>
<th>Text File Name</th>
<th>Audio File Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Be1</td>
<td>35-45</td>
<td>Beekingham</td>
<td>Beef</td>
<td>362</td>
<td>1</td>
<td>04.12.2006</td>
<td>Helena</td>
<td>28</td>
<td>6371int001</td>
<td>6371int001</td>
</tr>
<tr>
<td>2</td>
<td>Be2</td>
<td>45-65</td>
<td>Beekingham</td>
<td>Arable</td>
<td>364</td>
<td>1</td>
<td>05.12.2006</td>
<td>Helena</td>
<td>21</td>
<td>6371int002</td>
<td>6371int002</td>
</tr>
<tr>
<td>3</td>
<td>Be3</td>
<td>45-65</td>
<td>Beekingham</td>
<td>Arable</td>
<td>372</td>
<td>2</td>
<td>06.12.2006</td>
<td>Helena</td>
<td>22</td>
<td>6371int003</td>
<td>6371int003</td>
</tr>
<tr>
<td>4</td>
<td>Be4</td>
<td>45-65</td>
<td>Beekingham</td>
<td>Arable</td>
<td>194</td>
<td>3</td>
<td>07.12.2006</td>
<td>Helena</td>
<td>18</td>
<td>6371int004</td>
<td>6371int004</td>
</tr>
<tr>
<td>5</td>
<td>Be5</td>
<td>55-65</td>
<td>Beekingham</td>
<td>Arable</td>
<td>108</td>
<td>1</td>
<td>01.02.2008</td>
<td>Helena</td>
<td>19</td>
<td>6371int006</td>
<td>6371int006</td>
</tr>
<tr>
<td>6</td>
<td>Be6</td>
<td>45-65</td>
<td>Beekingham</td>
<td>Arable</td>
<td>1254</td>
<td>2</td>
<td>15.02.2007</td>
<td>Quentin</td>
<td>15</td>
<td>6371int008</td>
<td>6371int008</td>
</tr>
<tr>
<td>7</td>
<td>Bu1</td>
<td>55-65</td>
<td>Bushley</td>
<td>Mixed</td>
<td>101</td>
<td>1</td>
<td>13.02.2007</td>
<td>Quentin</td>
<td>29</td>
<td>6371int007</td>
<td>6371int007</td>
</tr>
<tr>
<td>8</td>
<td>Bu2</td>
<td>&gt;65</td>
<td>Bushley</td>
<td>Mixed</td>
<td>97</td>
<td>1</td>
<td>15.02.2007</td>
<td>Quentin</td>
<td>15</td>
<td>6371int008</td>
<td>6371int008</td>
</tr>
<tr>
<td>9</td>
<td>Bu3</td>
<td>&gt;65</td>
<td>Bushley</td>
<td>Arable</td>
<td>194</td>
<td>4</td>
<td>15.03.2007</td>
<td>Quentin</td>
<td>21</td>
<td>6371int009</td>
<td>6371int009</td>
</tr>
<tr>
<td>10</td>
<td>Bu4</td>
<td>55-65</td>
<td>Bushley</td>
<td>Mixed</td>
<td>202</td>
<td>1</td>
<td>15.03.2007</td>
<td>Helena</td>
<td>19</td>
<td>6371int010</td>
<td>6371int010</td>
</tr>
<tr>
<td>11</td>
<td>Cu1</td>
<td>35-45</td>
<td>Cuddyarch</td>
<td>Dairy</td>
<td>64</td>
<td>1</td>
<td>08.05.2007</td>
<td>Helena</td>
<td>19</td>
<td>6371int011</td>
<td>6371int011</td>
</tr>
<tr>
<td>12</td>
<td>Cu2</td>
<td>55-65</td>
<td>Cuddyarch</td>
<td>Dairy</td>
<td>189</td>
<td>2</td>
<td>08.05.2007</td>
<td>Helena</td>
<td>18</td>
<td>6371int012</td>
<td>6371int012</td>
</tr>
<tr>
<td>13</td>
<td>Cu3</td>
<td>55-65</td>
<td>Cuddyarch</td>
<td>Mixed livestock</td>
<td>76</td>
<td>1</td>
<td>08.05.2007</td>
<td>Helena</td>
<td>18</td>
<td>6371int013</td>
<td>6371int013</td>
</tr>
<tr>
<td>14</td>
<td>Cu4</td>
<td>45-65</td>
<td>Cuddyarch</td>
<td>Mixed livestock</td>
<td>193</td>
<td>1</td>
<td>09.05.2007</td>
<td>Helena</td>
<td>24</td>
<td>6371int014</td>
<td>6371int014</td>
</tr>
<tr>
<td>15</td>
<td>Cu5</td>
<td>55-65</td>
<td>Cuddyarch</td>
<td>Dairy</td>
<td>89</td>
<td>1</td>
<td>09.05.2007</td>
<td>Helena</td>
<td>14</td>
<td>6371int015</td>
<td>6371int015</td>
</tr>
<tr>
<td>16</td>
<td>Cu7</td>
<td>&gt;65</td>
<td>Cuddyarch</td>
<td>Mixed livestock</td>
<td>190</td>
<td>4</td>
<td>11.05.2007</td>
<td>Helena</td>
<td>20</td>
<td>6371int016</td>
<td>6371int016</td>
</tr>
<tr>
<td>17</td>
<td>Cu8</td>
<td>55-65</td>
<td>Cuddyarch</td>
<td>Mixed livestock</td>
<td>109</td>
<td>2</td>
<td>11.05.2007</td>
<td>Helena</td>
<td>22</td>
<td>6371int017</td>
<td>6371int017</td>
</tr>
<tr>
<td>18</td>
<td>Id1</td>
<td>35-65</td>
<td>Idle</td>
<td>Arable</td>
<td>158</td>
<td>3</td>
<td>07.02.2007</td>
<td>Quentin</td>
<td>17</td>
<td>6371int018</td>
<td>6371int018</td>
</tr>
<tr>
<td>19</td>
<td>Id1b</td>
<td>35-65</td>
<td>Idle</td>
<td>Arable</td>
<td>158</td>
<td>3</td>
<td>07.02.2007</td>
<td>Quentin</td>
<td>17</td>
<td>6371int019</td>
<td>6371int019</td>
</tr>
<tr>
<td>20</td>
<td>Id2</td>
<td>45-65</td>
<td>Idle</td>
<td>Dairy</td>
<td>150</td>
<td>1</td>
<td>08.02.2007</td>
<td>Quentin</td>
<td>17</td>
<td>6371int020</td>
<td>6371int020</td>
</tr>
<tr>
<td>21</td>
<td>Id2b</td>
<td>45-55</td>
<td>Idle</td>
<td>Dairy</td>
<td>150</td>
<td>1</td>
<td>08.02.2007</td>
<td>Quentin</td>
<td>19</td>
<td>6371int021</td>
<td>6371int021</td>
</tr>
<tr>
<td>22</td>
<td>Id3</td>
<td>35-45</td>
<td>Idle</td>
<td>Arable</td>
<td>380</td>
<td>3</td>
<td>01.02.2006</td>
<td>Helena</td>
<td>27</td>
<td>6371int022</td>
<td>6371int022</td>
</tr>
<tr>
<td>23</td>
<td>Mo1</td>
<td>&gt;65</td>
<td>Morda</td>
<td>Mixed</td>
<td>138</td>
<td>1</td>
<td>12.03.2007</td>
<td>Helena</td>
<td>31</td>
<td>6371int023</td>
<td>6371int023</td>
</tr>
<tr>
<td>24</td>
<td>Mo3</td>
<td>35-45</td>
<td>Morda</td>
<td>Arable</td>
<td>157</td>
<td>2</td>
<td>13.03.2007</td>
<td>Helena</td>
<td>19</td>
<td>6371int024</td>
<td>6371int024</td>
</tr>
<tr>
<td>25</td>
<td>Mo4</td>
<td>55-65</td>
<td>Morda</td>
<td>Mixed livestock</td>
<td>122</td>
<td>1</td>
<td>13.03.2007</td>
<td>Helena</td>
<td>19</td>
<td>6371int025</td>
<td>6371int025</td>
</tr>
<tr>
<td>26</td>
<td>Mo5</td>
<td>&gt;65</td>
<td>Morda</td>
<td>Mixed</td>
<td>142</td>
<td>2</td>
<td>14.03.2007</td>
<td>Helena</td>
<td>14</td>
<td>6371int026</td>
<td>6371int026</td>
</tr>
<tr>
<td>27</td>
<td>Mo6</td>
<td>&gt;65</td>
<td>Morda</td>
<td>Mixed livestock</td>
<td>19</td>
<td>1</td>
<td>14.03.2007</td>
<td>Helena</td>
<td>15</td>
<td>6371int027</td>
<td>6371int027</td>
</tr>
<tr>
<td>28</td>
<td>Mo7</td>
<td>&gt;30</td>
<td>Morda</td>
<td>Dairy</td>
<td>74</td>
<td>1</td>
<td>30.05.2007</td>
<td>Helena</td>
<td>22</td>
<td>6371int028</td>
<td>6371int028</td>
</tr>
<tr>
<td>29</td>
<td>Mo8</td>
<td>55-65</td>
<td>Morda</td>
<td>Mixed livestock</td>
<td>48</td>
<td>4</td>
<td>30.05.2007</td>
<td>Helena</td>
<td>19</td>
<td>6371int029</td>
<td>6371int029</td>
</tr>
<tr>
<td>30</td>
<td>Mo9</td>
<td>&gt;65</td>
<td>Morda</td>
<td>Mixed</td>
<td>278</td>
<td>1</td>
<td>31.05.2007</td>
<td>Helena</td>
<td>23</td>
<td>6371int030</td>
<td>6371int030</td>
</tr>
<tr>
<td>31</td>
<td>Mo10</td>
<td>35-45</td>
<td>Morda</td>
<td>Beef</td>
<td>81</td>
<td>1</td>
<td>01.06.2007</td>
<td>Helena</td>
<td>21</td>
<td>6371int031</td>
<td>6371int031</td>
</tr>
<tr>
<td>32</td>
<td>Mo11</td>
<td>35-45</td>
<td>Morda</td>
<td>Mixed</td>
<td>109</td>
<td>2</td>
<td>01.06.2007</td>
<td>Helena</td>
<td>22</td>
<td>6371int032</td>
<td>6371int032</td>
</tr>
<tr>
<td>33</td>
<td>Mo12</td>
<td>35-45</td>
<td>Morda</td>
<td>Mixed livestock</td>
<td>51</td>
<td>1</td>
<td>11.10.2007</td>
<td>Helena</td>
<td>34</td>
<td>6371int033</td>
<td>6371int033</td>
</tr>
<tr>
<td>34</td>
<td>Ro1</td>
<td>&gt;65</td>
<td>Rother</td>
<td>Rent and let out</td>
<td>57</td>
<td>1</td>
<td>20.03.2007</td>
<td>Helena</td>
<td>19</td>
<td>6371int034</td>
<td>6371int034</td>
</tr>
<tr>
<td>35</td>
<td>Ro3</td>
<td>&gt;65</td>
<td>Rother</td>
<td>Sheep</td>
<td>49</td>
<td>1</td>
<td>21.03.2007</td>
<td>Helena</td>
<td>33</td>
<td>6371int035</td>
<td>6371int035</td>
</tr>
<tr>
<td>36</td>
<td>Ro4</td>
<td>35-45</td>
<td>Rother</td>
<td>Mixed</td>
<td>182</td>
<td>1</td>
<td>21.03.2007</td>
<td>Helena</td>
<td>24</td>
<td>6371int036</td>
<td>6371int036</td>
</tr>
</tbody>
</table>
File naming conventions and best practice

Naming provides organization, context & consistency.

- how to best organize files depends on the plan and organization of the study
- file name/base name (not including file format extension) = principal identifier of file
- use logical naming i.e. easy to identify and retrieve the file
- name elements: version number, date, content description, creator name

Best practice:
- meaningful & brief
- relevant to content
- no special characters, dots or spaces
- avoid using space, for separation use underscores
- dates used should be in format YYYY-MM-DD;
- include versioning (when appropriate) via filename: ascending, decimal version numbers
- avoid very long file names
Access

Open
- available for download/online access under open licence without any registration

Safeguardsed
- available for download / online access to logged-in users who have registered and agreed to an End User Licence (e.g. not identify any potentially identifiable individuals)
- special agreements (depositor permission; approved researcher)
- embargo for fixed time period

Controlled
- available for remote or safe room access to authorised and authenticated users whose research proposal has been and who have received training
In practice: data with access conditions


- 40 interview and diary transcripts are archived and available for reuse by registered users (**safeguarded access**)
- 3 interviews and 5 diaries were embargoed until 2015 (**Safeguarded – embargoed**)
- Audio files archived and only available with permission from depositor(s) (**Safeguarded – Special agreement**)

---

[UK Data Service logo]
Access Standard text:

Open data
*The Data Collection is available to any user without the requirement for registration for download/access.*

Safeguarded data
*The Data Collection is available for download to users registered with the UK Data Service.*

Safeguarded data with permission from the depositor
*The Data Collection is available for download to users registered with the UK Data Service. All requests are subject to the permission of the data owner or his/her nominee. Please email the contact person for this data collection to request permission to access the data, explaining your reason for wanting access to the data, then contact our Access Helpdesk.*

Embargo
*The UK Data Archive has granted a dissemination embargo. The embargo will end on [date] and the data will then be available in accordance with the access level selected.*
Licencing

- A licence agreement is a legal arrangement between the creator/depositor of the data set and the data repository, signifying what a user is allowed to do with the data.

- Dependent on access level, the less restrictive the access level, the less restrictive the licence will be:
  - For open data - Creative Commons with variations
  - For safeguarded (restricted access) data – End User Licence with variations
<table>
<thead>
<tr>
<th>Licence</th>
<th>Can I copy &amp; redistribute the work?</th>
<th>Is it required to attribute the author?</th>
<th>Can I use the work commercially?</th>
<th>Am I allowed to adapt the work?</th>
<th>Can I change the licence when redistributing?</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC0</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>CC BY</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>CC BY-SA</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>CC BY-ND</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>CC BY-NC</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>CC BY-NC-SA</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>CC BY-NC-ND</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>
How to choose a licence?

Be sure you know who owns the data - you can only archive and publish data you own (or if you have permission).

• Is open access appropriate for your data? Then use Creative Commons Licences.

• If access to your data should be regulated, then the licence would depend of the archive or repository you are choosing to deposit the data. In this case an End User Licence will be user to list the terms and conditions under which the data can be reused.
  Example: the UKDS End User Licence [https://ukdataservice.ac.uk/media/455131/cd137-enduserlicence.pdf](https://ukdataservice.ac.uk/media/455131/cd137-enduserlicence.pdf)
Example: UKDS End User Licence (for restricted access data)

2. **End User Licence (EUL) Summary text**

Eighteen points to help you understand the End User Licence (EUL). These pointers are for general guidance and you must read and understand the full EUL before agreeing to it. By accepting the EUL, you agree:

1. to use the data in accordance with the EUL and to notify the UK Data Service of any non-compliance you are aware of
2. not to use the data for commercial purposes without obtaining permission and, where relevant, an appropriate licence if commercial use of the data is required
3. that the EUL does not transfer any interest in intellectual property to you
4. that the EUL and data collections are provided without warranty or liability of any kind
5. to abide by any further conditions notified to you
6. to give access to the data collections only to registered users with a registered use (who have accepted the terms and conditions, including any relevant further conditions). There are some exceptions regarding the use of data collections for teaching and the use of data collections for Commercial purposes set out in an additional Commercial Licence.
7. to ensure that the means of access to the data (such as passwords) are kept secure and not disclosed to anyone else
8. to preserve the confidentiality of, and not attempt to identify, individuals, households or organisations in the data
9. to use the correct methods of citation and acknowledgement in publications
10. to send the UK Data Service bibliographic details of any published work based on our data collections
11. that personal data about you may be held for validation and statistical purposes and to manage the service, and that these will only be passed on to the following, in specific circumstances: the data collection depositor, your own institute or your research funder
12. to notify the UK Data Service of any errors discovered in the data collections
13. that personal data submitted by you are accurate to the best of your knowledge and kept up to date by you
14. to meet any charges that may apply
15. to offer for deposit any new data collections which have been derived from the materials supplied
16. will, destroy **all** copies of the data to the standards specified in point 1.16
17. will ensure that the data are destroyed to the standards specified in the Microdata Handling and Security: Guide to Good Practice;
18. that any non-compliance with the EUL will lead to immediate termination of your access to the services and could result in legal action against you.
Where to publish your data?

There are different ways to publish data. Your preference may depend on the existing practices in your discipline or on the expectations of your funder.

- Journal supplementary material service (Example: Plus One, Scientific Data)

- Institutional data repository (Example: Essex University Data Repository, ORDA - University of Sheffield’s data repository)

- General purpose repository (Example: Dryad, Figshare, Zenodo)

- Domain specific data repository(Example: UKDS -social science, ADS – Archaeology, GenBank – genetic sequence database)
Self-archiving or expert help?

- There is a difference between self-archiving without any help and archiving with the help of an expert.
- Self-archiving is a quick and easy way to publish data (ReShare, Datorium)
- Archiving with the help of an expert will enhance data quality (UKDS)
- Expert help is most likely to be available at a trusted domain repository or an institutional repository.

- Contact your local institution library or Research Office, they should be able to help.
ReShare Demo

ReShare is the UK Data Service's self-deposit repository, where researchers can archive, publish and share research data, as open or restricted access data.

Once you create a data collection in your account, UKDS staff will review your data collection before publishing data to the live system, for:

- disclosure risks
- copyright breaches
- validity of file formats
- level of documentation

Depositors are contacted if there are any problems the deposit; otherwise, they receive an email notification when the collection goes live.

Demo
Exercise: data list

Which interview elements from the example interview would you include in a data list for the collection of interviews?
Keep connected

• Subscribe to UK Data Service list: www.jiscmail.ac.uk/cgi-bin/webadmin?A0=UKDATASERVICE

• Follow UK Data Service on Twitter: @UKDataService

• Follow our RDM account on Twitter: @UKDSRDM

• Youtube: www.youtube.com/user/UKDATASERVICE
Contact

Enquiries/ Help Desk:

http://ukdataservice.ac.uk/help/get-in-touch.aspx

help@ukdataservice.ac.uk

Follow us on:

https://twitter.com/UKDataService

https://www.facebook.com/UKDataService

https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=UKDATASERVICE
Questions?

https://pbs.twimg.com/media/B7ZUntrCUAEQAgR.jpg