

## Access and re-use

A significant portion of efforts to curate and preserve research data serves the purpose of re-use. By making data accessible, archives facilitate good science: discussion, innovation, contribute to making research cost-effective, and improve the quality of education and training. To reap the benefits an archive must offer the ability to retrieve and allow users to obtain data, in addition to responding to user queries about data and documentation.

Access is a matter of trust. Trust from data creators, who expect the archive to preserve their data and disseminate as stipulated in the deposit agreement and in accordance with the archive's access policy; and trust from the users, who depend on the archive to provide them with uncorrupted, authentic, comprehensible data. At the same time, archives can benefit from interaction with users that providing access to archived data entails. This interaction gives archives an opportunity to stay informed about, for example, changes in the way their designated communities handle and work with data, and their preferred modes of access to data.

### Licensing data for re-use

Archives do not usually seek or assert ownership of data. Instead, ownership remains with the data creator or depositor, who agrees to allow the archive to manage dissemination via a license agreement. For this reason, archives can only hold data where the rights holders have given permission for archiving and re-use.

The license is a legal agreement negotiated between the archive and depositor often based on license options. The standard options include open (accessible to all), non-commercial use only, or case-by-case depositor permission required for re-use. Licenses are usually non-exclusive, allowing data creators freedom to host and disseminate the data elsewhere on separate terms. Alternative options for licensing data include various permutations of Creative Commons<sup>1</sup> licenses and suitably, Open Data Commons Licenses<sup>2</sup>. Archives are often expected to ensure descriptive metadata and documentation are open access under these licenses.

### Recommended introductory resources

Ball, A. (2014). *How to License Research Data*. Retrieved from <http://www.dcc.ac.uk/resources/how-guides/license-research-data>

Kreutzer, T. (2014). *Open Content – A practical guide to using Creative Commons licences*. Retrieved from [https://meta.wikimedia.org/wiki/Open\\_Content\\_-\\_A\\_Practical\\_Guide\\_to\\_Using\\_Creative\\_Commons\\_Licences](https://meta.wikimedia.org/wiki/Open_Content_-_A_Practical_Guide_to_Using_Creative_Commons_Licences)

<sup>1</sup> <http://www.creativecommons.org>

<sup>2</sup> <http://opendatacommons.org/>

## Legal frameworks for data re-use

While the movement towards “open data” is significant, the bulk of social science data curated by archives remains outside the public domain. Archives therefore take on a responsibility of due diligence to protect the intellectual property rights of data creators and an ethical protection of research participants. The result is that to access data, users often need to sign a legal agreement accepting terms and conditions by which they can obtain and use data. The license can contain civil or criminal sanctions to deter abuse of data collections by malevolent or misguided users.

Examples for license agreements from CESSDA archives (English language versions):

Czech Republic - CSDA	<a href="http://archiv.soc.cas.cz/sites/default/files/smlouva_csd_a_en.pdf">http://archiv.soc.cas.cz/sites/default/files/smlouva_csd_a_en.pdf</a>
Denmark - DDA	<a href="http://samfund.dda.dk/dda/dokumenter/Rek_uk.pdf">http://samfund.dda.dk/dda/dokumenter/Rek_uk.pdf</a>
Finland - FSD	<a href="https://services.fsd.uta.fi/docs/terms-of-use?lang=en">https://services.fsd.uta.fi/docs/terms-of-use?lang=en</a>
France - Réseau Quetelet	<a href="http://www.reseau-quetelet.cnrs.fr/spip/rubrique.php?id_rubrique=67&amp;lang=en">http://www.reseau-quetelet.cnrs.fr/spip/rubrique.php?id_rubrique=67&amp;lang=en</a>
Germany - GESIS	<a href="http://www.gesis.org/en/services/data-analysis/data-archive-service/usage-regulations/">http://www.gesis.org/en/services/data-analysis/data-archive-service/usage-regulations/</a>
Lithuania - LiDA	<a href="http://www.lidata.eu/en/index.php?file=files/eng/data/terms_for_use_data.html">http://www.lidata.eu/en/index.php?file=files/eng/data/terms_for_use_data.html</a>
Slovenia - ADP	<a href="http://www.adp.fdv.uni-lj.si/eng/za_uporabnike/o_podatkih/#omejitve">http://www.adp.fdv.uni-lj.si/eng/za_uporabnike/o_podatkih/#omejitve</a>
Sweden - SND	<a href="http://snd.gu.se/en/search-and-order-data/conditions">http://snd.gu.se/en/search-and-order-data/conditions</a>
United Kingdom - UK Data Archive	<a href="http://ukdataservice.ac.uk/get-data/how-to-access/conditions.aspx#/tab-end-user-licence">http://ukdataservice.ac.uk/get-data/how-to-access/conditions.aspx#/tab-end-user-licence</a>


See Horton (2013) for a comparison of CESSDA license agreement.

## Disseminating sensitive data

A challenge for archives is how to facilitate the use of good quality data that is highly sensitive or disclosive. For example, macro data, complex linked data, or sensitive personally disclosive material. If archives are to make data available to the “fullest extent possible”, they need to address methods of secure data dissemination that allow data to be still reusable without compromising the identity of research participants. While aggregating data or anonymisation can enable re-use of some sensitive data, there comes a point where these strategies compromise data quality to render the data unsuitable for re-use.

These cases require implementation of additional access checks and controlled access. This often takes the form of restrictive licenses and physical forms of dissemination, for example making the data only available on encrypted CD-R rather than downloadable. A further step is to permit onsite-supervised use of data. For example, only approved researchers may use certain data collections at Inter-university Consortium for Political and Social Research’s (ICPSR) secure data enclave<sup>3</sup>. Any analysis of these data is monitored to protect confidentiality and is conducted on dumb-terminals with no external storage medium access. The researcher then receives approved analysis outputs on storage devices supplied by ICPSR.

<sup>3</sup> <http://www.icpsr.umich.edu/icpsrweb/content/ICPSR/access/restricted/index.html>



Meanwhile, the UK Data Archive's Secure Data Service<sup>4</sup> allows approved researchers to work on secure data remotely, but over a controlled network, that prevents transfer or downloads to a local computer or printer. In both cases, sensitive data never leaves the archive.

### User authentication

Data are often subject to restrictions on the right to use. This necessitates controls as to who can access data. The basis for control is Authentication and Authorization Infrastructure (AAI). There are two types, direct and federated AAI. The direct approach is to issue a unique user name and password to each individual user. However, this can be resource intensive for the service and identity providers. Therefore, the basis of a Federated AAI is to have one user name for users within that institution. Potential users of data are managed and authenticated by their home organization (the identity provider) and authorized by the service provider who holds the data. This federated system works on a basis of trust between identity and service provider, known as Where Are You From (WAYF). This is based on the exchanging of public keys (SSL encryption). However, while federations are widely adopted, they are problematic for users who do not know to which federation they belong. Consequently, a system called DiscoJuice has the option of attaching weight to important identity providers by usage or location.

#### Recommended introductory resource

DASISH online tutorial "AAI - Guarding the key to your data", available at <http://training.dasish.eu/training/2/>.

#### Examples of federated AAI

- UK Access Management Federation: <http://www.ukfederation.org.uk/>
- Shibboleth: <http://shibboleth.net/about/index.html>
- Athens: <http://www.openathens.net/>
- DiscoJuice: <http://discojuice.org/>

### Persistent identifiers (PIDs)

It is critical data creators receive credit for subsequent re-use. For archives, ensuring proper citation of data is a responsibility of care and professional recognition of the depositor, funder, and integral to the wider data lifecycle. This includes a responsibility to ensure links and citations allow future readers and users to discover the data collection. It is standard to use http Universal Resource Identifier links, but this does not protect against "link rot" which occurs when domain names change or file systems alter thus rendering the hyperlink outmoded.

A Digital Object Identifier (DOI®) applies a case-sensitive Unicode character string to permanently and uniquely identify an object with metadata associated to that DOI®. Therefore, referring to an online document by its DOI® is a stable persistent approach. The DOI® string divides into a prefix identifying the registrant of the name, and a suffix chosen by the registrant to identify the specific object associated with that DOI®. However, the DOI® system is not open. Organizations have to pay to become members of the federation that develops and controls the system. Nonetheless, in October 2014 more than 100 million DOI® names had been assigned by over 5,000 organizations (see [www.doi.org/factsheets/DOIKeyFacts.html](http://www.doi.org/factsheets/DOIKeyFacts.html)).

As members of the DataCite consortium<sup>5</sup>, GESIS and the German National Library of Economics, Leibniz Information Centre for Economics (ZBW) offer da|ra, the DOI registration service for social

---

<sup>4</sup> <http://ukdataservice.ac.uk/get-data/how-to-access/accesssecurelab.aspx>

<sup>5</sup> <https://www.datacite.org/>



science and economic data in Germany<sup>6</sup>.

## References

Horton, L. (2013). A look at CESSDA and data re-use licenses. In *IASSIST Conference, May 28 - 31, 2013, Cologne*. Retrieved from <http://dasish.eu/dasishevents/iassistws/dasish-cessda-iassist.pptx>



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

---

<sup>6</sup> <http://www.da-ra.de/en/home/>