

CESSDA Data Access Policy

June 2016

This policy was agreed by the CESSDA General Assembly on 14 June 2016.

CESSDA Data Access Policy

Principle 1: The responsibilities of a Service Provider (SP) regarding access to any data collection shall be managed by a licence/agreement between the data owner and the SP.

Principle 2: The responsibilities of a user regarding the use of any data collection shall be managed by a licence/agreement between the Service Provider and the user.

Principle 3: All data holdings shall be available to anyone regardless of status, nation or type of use (except redistribution) unless there are other known requirements which prevent it.

Principle 4: All data holdings, documentation and metadata of each service provider shall be available free at the point of access for public research and education.

Principle 5: All data holdings of each Service Provider shall be identifiable on the public web.

Principle 6: Each data collection shall have its access conditions clearly displayed.

Principle 7: Each data collection shall have documentation which allows its use.

Principle 8: Documentation shall be accessible without restriction and at no cost unless there are other known requirements which prevent it.

Principle 9: Resource–discovery metadata in a structured format shall be openly available for any third–party to harvest.

Principle 10: Resource–discovery metadata shall, by 2020, include a persistent identifier.

Principle 11: Access conditions to data shall, by 2022, be fully interoperable.

Principle 12: Access conditions shall be applied only taking into account: disclosure risk, intellectual property, and reasonable first use period.

Principle 13: Service Providers shall advise data owners on the most appropriate access conditions.

Principle 14: This policy shall be reviewed by the Director, every five years or more frequently if required.

Implementation notes

These notes are for guidance, and do not need to be approved. They form a living document which deal with interpretation. With the principles as written we cannot avoid a difference between expectation and actuality, and we cannot avoid local differences of implementation because of cost, local interpretation of legislation, and differences in terms of funders' opinions.

Principle 1

In order for Service Providers to act within the law it is necessary to have a licence (or agreement) for redistribution in place for each data collection. The licence should state the rights owner and the data controller and it should give the conditions under which the Service Provider can provide access. These are traditionally known as Depositor's Licences and are legally binding contracts.

There will (inevitably) be some cases where the right's holder/data controller is unknown. Examples include creators who have died and there is no clarity over the inheritor; creators as organisations which have no clear successor; and creators as organisations which have a clear successor, but that successor is not eligible to manage those rights (e.g., in Estonia, data created during the Soviet regime). In such cases treating these data collections as 'orphan works' will be sensible.

As noted, these are traditionally called (in English) Depositor's Licences. However, with separation of roles, we should distinguish between Data Owners (those who are the rights holders in the data), Data Controllers (who for data where there is a risk of disclosure of personal information are the 'risk managers' of those data and Data Creators (those who create the data). These three actors are not identical. The licence/agreement in this principle should be between the Data Owner (and Data Controller if different) and the Service Provider (or its legal entity).

Principle 2

In order to manage the licence with between the SP and the Data Owner/Controller, an agreement (or licence) between the SP and the data user is necessary. This protects the SP against any use which the Data Owner might object to. Licences and agreements do not have to be written.

This agreement may be limited to a set of terms and conditions, but it would be important to ensure that users are actually agreeing to those terms rather than just having their existence drawn attention to. For example, one SP (at the time of writing) has on their website that users are asked to "please observe" the conditions of use. It is likely that this does not make an agreement. If the wording were to say: "By downloading data from this site I agree to follow the following terms and conditions..." this would constitute an agreement.

We anticipate that the statement as it stands will also cope with data being accessed through an API.

During the implementation phase of this policy, CESSDA will publish recommended Terms and Conditions (see Recommendations below) and the CESSDA ERIC will use these as part their negotiation with aspiring members. The SPF do not believe that a single set of licences will be possible in the near future. The benefits accrued from a single set of licences would more than offset by the time spent renegotiating licences with existing data owners.

Principles 1 and 2

Where SPs make data available under an Open licence (e.g., CC4) then principles may be implemented differently. In the case of a CC4 licence the agreement is explicitly between the 'data owner' and the 'user'.

Principle 3

This principle does not provide an exhaustive list of exclusions. However, the key requirements which may prevent data from being made available for re-use include: privacy and consent issues, contractual issues, intellectual property requirements (especially with linked data), national security (political science), and trade secrets. There may be others which are not anticipated here.

Privacy requirements explicitly include the appropriate management of consent for data sharing by data subjects.

Legal requirements include both contractual (subject to data owner permission) and statute law (national legislation). Therefore if consent for commercial use is not given there are privacy requirements which would limit access, etc.

Principle 12 also explicitly deals with different access conditions. We must accept that this principle is idealistic since data owners/data controllers are the final arbiters of access conditions, since it is they not Service Providers who are the data owners. Thus they decide; our obligation to these principles is to help data owners come to the right decision.

The issue of registration came up in the discussion. CESSDA recommends that SPs mandate registration for access to all data holdings for the purposes of understanding the use and impact of data; however, it is recognised that in some countries local policy for open access will reject registration for this type of access. The majority of the CESSDA SPs would like the word 'open' included in this statement, but at present this is not possible. However, it would prefer to have the term 'open'" interpreted as it is by the OECD.¹

"Openness means access on equal terms for the international research community at the lowest possible cost, preferably at no more than the marginal cost of dissemination. Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based."

And not the Open Definition.²

"Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)."

The main reason to use the slightly outmoded OECD definition (as well as the underlying principles) is that it actually deals with research data whereas the Open Definition is designed for all digital objects. CESSDA would prefer to use the term 'intelligent openness' as it is used in the UK's Royal Society.³

¹ See here: <https://www.oecd.org/sti/sci-tech/38500813.pdf>

² See here: <http://opendefinition.org>

³ See: https://royalsociety.org/~media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

Principle 4

The design of this wording is to allow those service providers who already (or may in the future) make a 'handling' or 'special' charge for services which they provide, e.g., copying data to a CD, posting it, or maintaining a safe room. The principle for implementation is that access to the data is free, but that if the access method is a value-added service, it can be charged for.

The wording here follows that in the Statutes exactly, but there may be some implementations issues as different SPs interpret the term 'research' differently. On one hand public research, may be research by anyone in the public benefit, or it may be research by anyone. Scientific research may also be a distinct subset of research.

Principle 5

This means, in essence that the names (and other key information) of all of the data holdings of each SP should be publicly available. It doesn't mean they should be accessible. Thus, if SPs have some data which is 'preservation only' these should be identifiable. It is of course possible that this principle can not be fully implemented by all SPs since some of the holdings may need to be held confidentially. In these cases, we recommend that SPs put some explanatory text on their website which explains non-conformance to this principle.

The issue around older editions/versions or parts of current data collections needs further thought and explanation. It is not the intention of this policy to make SPs make available earlier editions or parts of data collections which are also available.

Principle 6

This seems to need no implementation guidance.

Principle 7

At this stage the implementation guidance will not provide a minimum level of documentation. However, the informally agreed baseline is that for a member of their 'designated community' the data collection should be independently understandable without recourse to the creator.

In those cases of legacy data where there is minimal documentation, and little hope of providing much without significant work, this rule can not be applied consistently. Therefore this applies for all new acquisitions to the Service Provider, and where the depositor is a known data owner. Exceptions should be allowable in other cases.

Principle 8

This wording is linked to the wording of Principle 3. There are cases where documentation (and or metadata) are deemed to be a disclosure risk in themselves. However, the key principle is that documentation and metadata should be freely available to all without restriction (except redistribution) regardless of the data access conditions of their related data collection.

Principle 9

From an implementation point of view this principle is to mandate the use of a protocol for specific (a minimum set not yet agreed) resource-discovery metadata to be available to anyone.

At present the minimum set of metadata for resource discovery is not agreed, but the four candidate pieces of information are currently supporting archive, name, access conditions and abstract. At present, it is right to assume that this principle only mandates resource-discovery metadata at a study level, however, SPs may wish make metadata at a more granular level available.

SPs will need to ensure that they have the rights to make any resource-discovery metadata which they use available in this way (cf Principles 1 and 4).

Principle 10

This principle is vague to allow for multiple solutions.

The purpose of this statement is not really to apply a technical solution, rather it is part of a strategy to increase the findability of our data holdings.

Principle 11

This principle is designed to allow for interoperability between Service Providers and to make things simpler for users of the CESSDA portal. It is expected to lead to greater harmonisation across SPs, but it is not designed to force complete harmonisation. There are differences of practice across data owners which the SPs cannot influence (see Principle 13) but we would like to. By interoperability, we mean, at the lowest level, that 'standard' access conditions are interpreted in the same way across service providers. So, registration means the user giving some details, but which details may differ between SPs.

Principle 12

This is another principle which is not necessarily to be implemented in the same way by each SP. It is a statement of intent which SPs should be taking into consideration when negotiating for data with data owners. A flow chart could easily show how, in principle, this can be implemented.

Principle 13

The only advocacy principle here. It relates especially to Principle 12. A number of SPs are already doing this successfully.

Principle 14

This principle mandates the review of this policy to the Director. There is no expectation that this is done without discussion with the SPs and the Service Providers' Forum (SPF). However, changes to this policy should be agreed by the General Assembly.

Purpose

1. The primary objective of the policy was to put into a separate statement a series of principles which conformed to the intentions of the Statutes around data access conditions.
2. The secondary objective of the policy was to demonstrate CESSDA's commitment to both open science and the fullest protection of the privacy of data subjects, as well as, inter alia, protecting the intellectual property rights of the data owners.
3. The policy must also not contradict or be in opposition to any of the key standards already adopted by CESSDA.
4. Four specific benefits were kept in mind throughout the policy development process:
 1. data users should benefit from a uniform approach to data access;
 2. internal metadata users will benefit from a legally suitable statement for re-use;
 3. external metadata users will also benefit from free access to CESSDA's rich, resource discovery metadata;
 4. service providers will benefit from a unified approach to access, whereby workflows may be able to be shared/repurposed.
5. The primary aim of this piece of work is to construct a short, relatively generic policy (detailed enough to prevent confusion, but not so explicit as to make it impossible to implement by all of the members of the CESSDA consortium) and some implementation guidelines.
6. It was anticipated that it would be mandatory for all participating Service Providers in the consortium to implement this policy.
7. All principles (except the two which are dated) should come into immediate effect; however, with differential levels of maturity, there is an expectation that all SPs will have implemented all these principles within two years of the foundation of the ERIC. Not every eventuality relating to national legislation, local policies and cultural differences can be predicted, therefore some latitude has been written into these principles to allow for minor interpretive differences.

Background Notes

1. In May 2015 CESSDA MO published a call for a Work Plan Task for the construction of Data Access and Dissemination Policy. The contract was awarded to the UK Data Service. While the delivery agreement went to a single organisation, all CESSDA Service Providers, existing and prospective, have had multiple opportunities to feed into this policy development activity.
2. Following a period of desk research, including scrutiny of almost all the existing Service Providers' arrangements for data access, a discussion document was published on the CESSDA Basecamp in November 2015. At this stage the name of this policy was changed to a Data Access Policy, because of some differences of usage of the term 'dissemination'. The policy as constructed covers the exact conceptual area of the term in the current Statutes, and the new wording in the draft Statues for the ERIC. Both the Service Provider's Forum and the General Assembly were notified of the publication of this discussion document and the change of title.
3. The Discussion Document⁴ provided an overview of the rationale behind a Data Access Policy. In this document 13 principles were proposed, and 23 questions were asked. The questions were trialed internally within the UK Data Service for clarity.
4. Four detailed responses were received from CESSDA Service Providers. A fifth response was commissioned from an external expert with no formal knowledge of CESSDA.
5. A workshop was held in Bergen as part of the CESSDA SPF on 11 May 2016. Participants from all but two of the current members were present. The workshop presented each of the principles (some revised to take into account the responses to the questions) led by Matthew Woollard. Some further changes have been made in this document, and all SPs present agreed these principles were fit for purpose.
6. To ensure that all SPs had the possibility of commenting on these principles, they were posted to Basecamp before formal tabling at General Assembly.

⁴ If this file is not accessible through this link, just copy this URL into a browser: <https://asset1.basecamp.com/2505696/projects/10594445/attachments/192702644/1709e724f255a0fb82aefc1914e0ab0b0010/original/CESSDA-DAD-notes-distribute.pdf>

Additional Recommendations

At the time of approval of the CESSDA Data Access Policy, the General Assembly was asked to note the following points. These are maintained as an Annex to the Policy in order to maintain a record of decisions.

1. At present the application of uniform access conditions is likely to be burdensome for many SPs. (See Implementation notes under Principle 2). Therefore we recommend that:
 - CESSDA publish exemplar statements for terms and conditions for both data owner/data service and data service/data user agreements.
2. The discussion document also asked Service Providers questions about other areas relating to data access. The conclusions from the questions are noted here.
 - CESSDA should never act as a 'Service Provider'. Note: If there is a necessity for CESSDA to act as one partner in a data sharing agreement, then this agreement should not diverge from best SP practice, and sub-contractual responsibilities should be allowed for.
 - CESSDA does not need a formal Data Access Committee. Note: this is because at present it could have no devolved authority and thus no power.
 - CESSDA should not have a formal Ethics Review Panel. However anchoring scientific ethics within CESSDAs activities is important.
 - CESSDA SPs should manage information relating to informed consent alongside the relevant data.
 - CESSDA ERIC should have a working group (in the absence of a Board) which deals with these policy questions.

Acknowledgments

The construction of this document was coordinated by Matthew Woollard, UK Data Archive / UK Data Service. His time, and that of his colleagues, Hilary Beedham and Hervé L'Hours who both contributed to the policy statements and the construction of the discussion document was funded by CESSDA. Matthew would like to thank all of the CESSDA Service Providers for their invaluable help in providing both local knowledge and contributing to the discussion without which this Policy would not be so clear. CESSDA is grateful also for our General Assembly's discussion on the Policy and clarifying some points.