



# Data Seal of Approval

[www.datasealofapproval.org](http://www.datasealofapproval.org)

Quality guidelines for digital research data



*Data Archiving and Networked Services*



DANS is an institute of KNAW and NWO

ICPSR | INTER-UNIVERSITY CONSORTIUM FOR  
POLITICAL AND SOCIAL RESEARCH



KONINKLIJKE NEDERLANDSE  
AKADEMIE VAN WETENSCHAPPEN



Max Planck Institute  
for Psycholinguistics



ISBN 978 9490 531 027

## Introduction

This document contains a total of sixteen guidelines for the application and verification of quality aspects with regard to creation, storage and (re-)use of digital research data in the social sciences and humanities. These guidelines serve as a basis for granting a 'Data Seal of Approval' by the Data Seal of Approval Board.

## Background

When DANS was established by the two Dutch science organizations KNAW and NWO, they assigned it the task of developing a Seal of Approval for data, to ensure that archived data can still be found, understood and used in the future.<sup>1</sup> In 2008 the first edition of the Data Seal of Approval, written by Laurents Sesink, René van Horik and Henk Harmsen, was presented in an international conference. In spring 2009 the Data Seal of Approval was handed over to an international Board.

## Objectives

The quality guidelines formulated in this document are of interest to researchers and institutions that create digital research files, to organizations that archive research files, and to users of research data.

The objectives of the Data Seal of Approval are to safeguard data, to ensure high quality and to guide reliable management of research data for the future without requiring the implementation of new standards, regulations or high costs.

The Seal of Approval

- Gives researchers the assurance that their research results will be stored in a reliable manner and can be reused
- Provides research sponsors with the guarantee that research results will remain available for reuse
- Enables researchers, in a reliable manner, to assess the repository where research data are held.
- Allows data repositories to archive and distribute research data efficiently

---

<sup>1</sup> DANS – Data Archiving and Networked Services – is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), and is also supported by the Netherlands Organization for Scientific Research (NWO). Since its establishment in 2005, DANS has been providing storage of and continuous access to research data in the social sciences and humanities.

## About the guidelines

The criteria for assigning the Data Seal of Approval to data repositories are in accordance with, and fit in with, national and international guidelines for digital data archiving such as *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive* as developed by NESTOR;<sup>2</sup> *Digital Repository Audit Method Based on Risk Assessment* (DRAMBORA) published by the Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE);<sup>3</sup> and *Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist* of the Research Library Group (RLG).<sup>4</sup> Furthermore the following has been taken into account: *Foundations of Modern Language Resource Archives* of the Max Planck Institute<sup>5</sup> and *Stewardship of Digital Research Data: A Framework of Principles and Guidelines* published by the Research Information Network.<sup>6</sup> The guidelines in this document can be seen as a minimum set distilled from the above proposals.

Fundamental to the following guidelines are five criteria, that together determine whether or not the digital research data may be qualified as sustainably archived:

- The research data can be found on the Internet.
- The research data are accessible, while taking into account relevant legislation with regard to personal information and intellectual property of the data.
- The research data are available in a usable format.
- The research data are reliable.
- The research data can be referred to.

These criteria will be addressed below in the guidelines.

The associated guidelines relate to the implementation of these criteria and focus on three stakeholders: the *data producer*, the *data repository* and the *data consumer*.

1. The *data producer* is responsible for the quality of the digital research data.
2. The *data repository* is responsible for the quality of storage and availability of the data: data management.
3. The *data consumer* is responsible for the quality of use of the digital research data.

---

2 See: <<http://edoc.hu-berlin.de/docviews/abstract.php?id=27249>> [accessed May 14, 2009].

3 See: <<http://www.digitalpreservationeurope.eu/announcements/drambora/>> [accessed May 14, 2009].

4 See: <<http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>> [accessed May 14, 2009].

5 Peter Wittenburg, Daan Broeder, Wolfgang Klein, Stephen Levinson and Laurent Romary. See: <<http://www.lat-mpi.eu/papers/papers-2006/general-archive-paper-v4.pdf>> [accessed May 14, 2009].

6 See: <<http://www.rin.ac.uk/data-principles>> [accessed May 14, 2009].

# 1. DATA PRODUCERS

The starting point for assigning the Seal of Approval to data is the deposit of research data into a data repository qualified to this end (see Section 2 *Data repositories*).

## Guidelines relating to the quality of digital research data

The quality of the digital research data is determined by:

1. Their intrinsic scientific and scholarly quality
2. The format in which the research data and supporting information are stored
3. The documentation (metadata or contextual information) regarding the research data

### 1.1 The (scientific and scholarly) content

**1. The *data producer* deposits the research data in a data repository with sufficient information for others to assess the scientific and scholarly quality of the research data and compliance with disciplinary and ethical norms.**

Scientific and scholarly quality criteria indicate to what degree the research data are of interest to the business of research. The assessment by experts and colleagues in the field is the main deciding factor for the quality of research data. Transparency in terms of adherence to ethical norms in relevant disciplines facilitates the assessment of data content.

The *data producer* therefore provides sufficient information to enable fellow researchers to assess the research data (see Section 1.3 *Documentation* and guideline 3).

## 1.2 Data formats

**2. The *data producer* provides the research data in formats recommended by the data repository.**

The bits that form any digital object are organized according to the rules for a particular data format. Various data formats exist for digital objects. For all formats, there is a risk that they may become obsolete. This creates a chance that the data object may become unusable. For storage of data objects *preferred formats* are used. *Preferred formats* are formats that a data repository can reasonably assure will remain readable and usable. Usually, these are the *de facto* standards employed by a particular research discipline.

## 1.3 Documentation

**3. The *data producer* provides the research data together with the metadata requested by the data repository.**

The *data producer* provides the research data with information about the context of the data (metadata). There is a distinction between descriptive, structural and administrative metadata. These must be provided in accordance with the guidelines of the data repository.

- a. *Descriptive metadata* consist of information required to find research data and add transparency to their meaning (definition and value) and importance. Examples of descriptive metadata are the data elements of the Dublin Core Element Set,<sup>7</sup> with fields such as creator, type, and date.
- b. *Structural metadata* indicate how different components of a set of associated data relate to one another. These metadata are needed to be able to process the research data. When data are coded, the codebook will be a component of the structural metadata.
- c. *Administrative metadata* are required to enable permanent access to the research data. This concerns the description of intellectual property, conditions for use and access and the preservation metadata needed for durable archiving of the research data.

The data repository specifies the level of producer-created metadata required and provides the tools for its effective capture.

---

<sup>7</sup> See: <<http://www.dublincore.org>> [accessed May 14, 2009].

# 2 DATA REPOSITORIES

## Guidelines relating to the quality of the data repository

The *data repository* is responsible for access and preservation of digital research data in the long term. Two factors, in particular, determine the quality of the data repository:

1. The quality of the organizational framework in which the data repository is incorporated (*organization and processes*)
2. The quality of the *technical infrastructure* of the data repository

A data repository is designated a *Trusted Digital Repository* (TDR) according to the requirements of the Data Seal of Approval if it meets guidelines 4-13 below and if it enables data-producers and data-consumers to adhere to guidelines 1-3 and 14-16.

Organizations that play a role in digital archiving and are establishing a *Trusted Digital Repository* minimally possess a sound financial, organizational and legal basis in the long term.

### 2.1 Organization and processes

4. The *data repository* has an explicit mission in the area of digital archiving and promulgates it.
5. The *data repository* uses due diligence to ensure compliance with legal regulations and contracts.
6. The *data repository* applies documented processes and procedures for managing data storage.
7. The *data repository* has a plan for long-term preservation of its digital assets.
8. Archiving takes place according to explicit workflows across the data life cycle.
9. The *data repository* assumes responsibility from the data producers for access to and availability of the digital objects.
10. The *data repository* enables the users to utilize the research data and refer to them.
11. The *data repository* ensures the integrity of the digital objects and the meta-data.

The information contained in the digital objects and metadata is complete and does not change relative to the originals.

**12. The *data repository* ensures the authenticity of the digital objects and the meta-data.**

This pertains to the degree of reliability of the original and to the provenance of the data. Existing relationships between data sets and explicit links are maintained.

## 2.2 Technical Infrastructure

**13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.**

The technical infrastructure constitutes the foundation of a *Trusted Digital Repository*. The OAIS reference model,<sup>8</sup> an ISO standard, is the *de facto* standard for using digital archiving terminology and defining the functions that a data repository fulfils, amongst other things.

---

<sup>8</sup> Details regarding the OAIS reference model can be found at:  
< <http://public.ccsds.org/publications/archive/650x0b1.pdf> > [accessed May 14, 2009].

# 3 DATA CONSUMERS

## Guidelines relating to the quality of the use of digital research data

The *data consumer* uses the digital research data in compliance with the guidelines below:

- 14. The *data consumer* must comply with access regulations set by the data repository.**
- 15. The *data consumer* conforms to and agrees with any codes of conduct that are generally accepted in higher education and research for the exchange and proper use of knowledge and information.**
- 16. The *data consumer* respects the applicable licences of the data repository regarding the use of the research data.**

The quality of the use of research data is determined by the degree to which the data can be used without limitation for scientific and scholarly research by the various target groups, while complying with certain codes of conduct.

The open and free use of research data takes place within the relevant legal frameworks and the policy guidelines as determined by the national authorities.

With regard to accessing information, the *data consumer* is bound by national legislation. The data repository may have separate access regulations, which include restrictions imposed by the laws of the country in which the data repository is located. Access regulations should be based on relevant international access standards (e.g., Creative Commons) as much as possible.

Most nations have legal frameworks relating to the ethical use and re-use of data. These frameworks range from the statutory — which protect the privacy of individuals — to formal codes of conduct which inform ethical research. Repositories must be aware of these local legal frameworks and ensure that they are taken into account when providing data for re-use.

