Data Archiving and Networked Services

Trusted Digital Archives and the Data Seal of Approval

Matthias Hemmje (FTK)
Peter Doorn (DANS)

APARSEN-EGI-Community-Forum Training on Data Preservation
Helsinki, 22nd of May 2014

What is DANS?

Institute of Dutch Academy and Research Funding Organisation (KNAW & NWO) since 2005

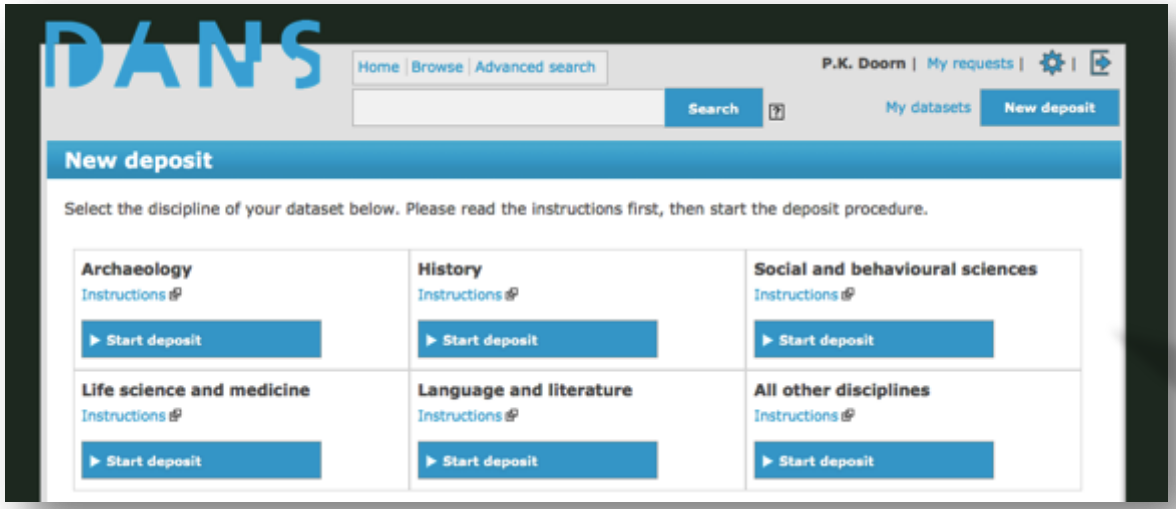
First predecessor dates back to 1964 (Steinmetz Foundation), Historical Data Archive 1989

Mission: promote and provide permanent access to digital research information

DANS

Driven by data

DANS' services



EASY: Electronic Archiving System for self-deposit

Data Seal of Approval

NARCIS: Gateway to scholarly information In the Netherlands

persid

Persistent Identifier
URN:NBN resolver

DANS

Driven by data

Trust in research data

- Trust is at the very heart of storing and sharing data
- Trust involves:
  - Data creators
  - Data users
  - Data repositories
  - Funders

Trust comes on foot, but leaves on horseback

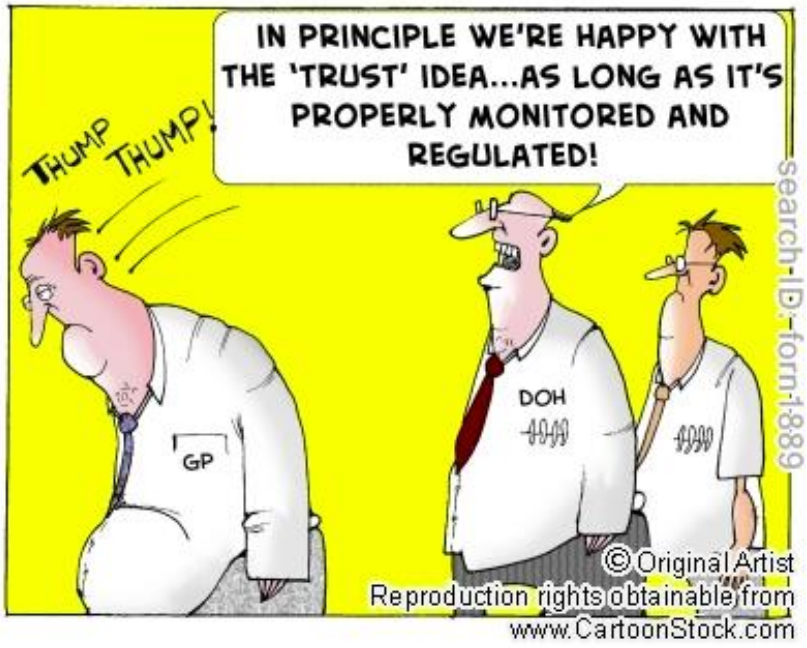
DANS

Driven by data

What is trust built on?

- Dedicate yourself (mission statement)
- Do what you promise (stable, sincere and competent reputation)
- Be transparent (peer review, get certified)

"They don't trust each other to share research."



DANS

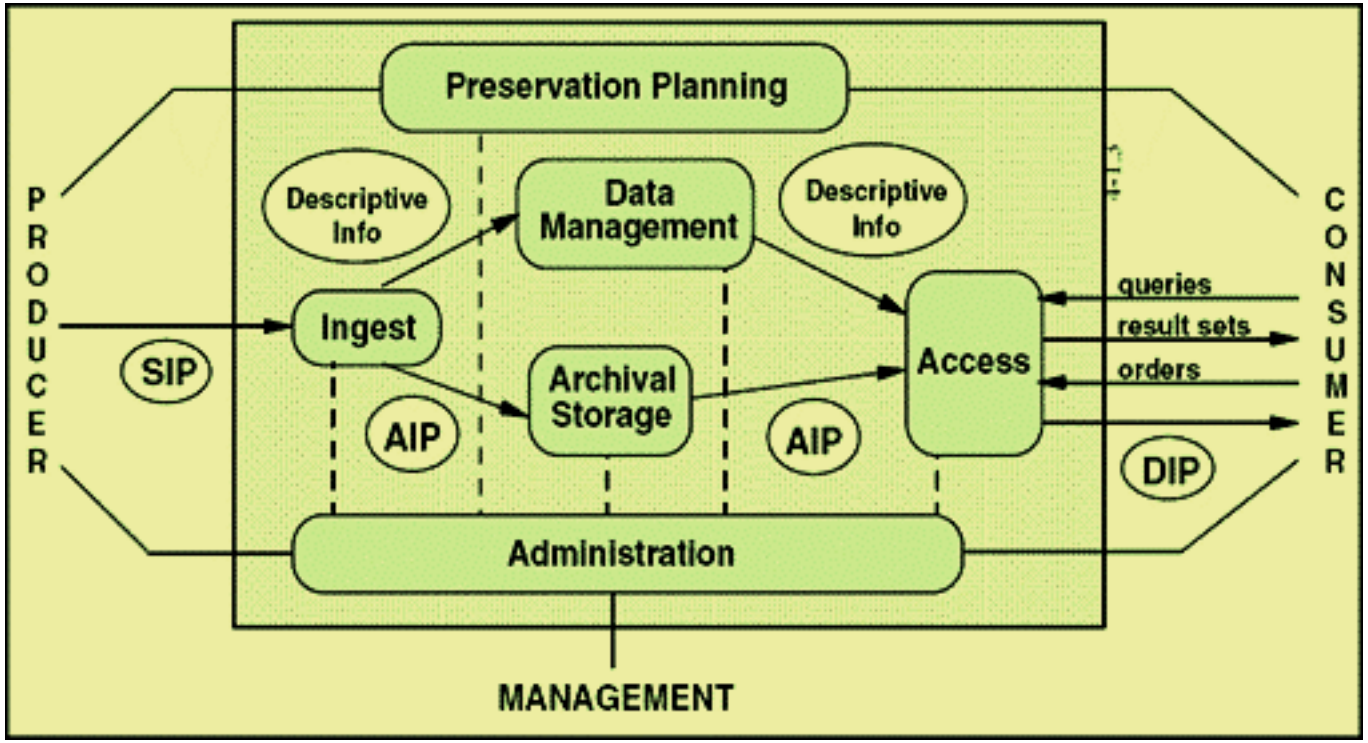
Driven by data

The need for trusted digital repositories

- In 1996 the Consultative Committee for Space Data Systems established a task force that developed OAIS (Open Archival Information System, accepted as ISO 14721 in 2002), a high-level model for the operation of digital archives.
- Independent auditing was deemed necessary to certify OAIS-compliance and thus engender trust.
- Development of OAIS auditing metrics began in 2003 and resulted in Trustworthy Repositories Audit & Certification (TRAC 2007).
- TRAC is the basis of the Trusted Digital Repository (TDR) document that was accepted as ISO 16363 in 2012.

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56510

OAIS – Open Archival Information System



SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package

Note: OAIS <> OAI
OAI = Open Archives Initiative, which develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content.

DANS

Driven by data

Trust in data archives: an example

UK • DATA ARCHIVE

THE UK'S LARGEST COLLECTION OF DIGITAL RESEARCH DATA IN THE SC
HUMANITIES

HOME | ABOUT US | CREATE & MANAGE DATA | DEPOSIT DATA | **HOW WE CURATE DATA**

HOW TO CURATE DATA | STANDARDS OF TRUST

- HOW WE CURATE DATA
- THE PROCESS
- OUR QUALITY CONTROL
- OUR PRESERVATION POLICY
- TRUSTED DIGITAL REPOSITORIES

STANDARDS OF TRUST / OVERVIEW

OVERVIEW | **DATA SEAL OF APPROVAL** | **ISO16363** | **DIN 31644**

Any organisation which provides access to data over a long period of time should be fully trusted only with a public statement describing the practices they follow and the provenance of data they provide. Standards of trust are critical.

DANS

Driven by data

ESFRI Research Infrastructures and Trust

DARIAH
Digital Research Infrastructure for the Arts and Humanities

Compliance as a DARIAH Trusted Digital Repository

The reliability and trustworthiness of digital preservation programmes are of utmost importance to all users, providers and

Organisations which join DARIAH or begin engaging in repository type activities will be expected to meet certain minimum standards. This will enable DARIAH to vet organisations and recommend practices to be followed. All organisations should meet policy requirements in order to be counted as Trusted.

CLARIN ERIC
Common Language Resources and Technology Infrastructure

Requirements for CLARIN Centres

"Centres need to have a proper and clearly specified repository system and participate in a quality assessment procedure as proposed by the Data Seal of Approval or MOIMS-RAC approaches"

metadata
accessing data
data harmonisation
ethics
cessda
social science data
confidentiality
data catalogue
data preservation

Building Trust: CESSDA Self-Assessment Project

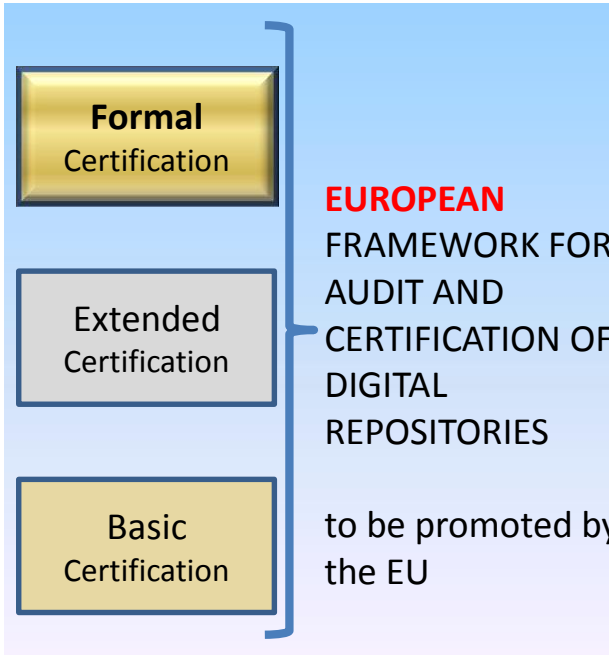
Participants from fifteen CESSDA member organisations discussed the CESSDA-ERIC requirements and agreed upon using the Data Seal of Approval (DSA) guidelines as a tool to gain information on the level of their conformance with the DSA and the CESSDA-ERIC requirements.

DANS

Driven by data

Certification of digital repositories

- International framework
- 3 standards
- 3 levels (basic, extended, formal)



DANS

http://www.trusteddigitalrepository.eu

Driven by data

Framework levels

- **Basic Certification** is granted to repositories which obtain Data Seal of Approval
- **Extended Certification** is granted to Basic Certification repositories which *in addition* perform a structured, externally reviewed and publicly available self-audit based on ISO 16363 or DIN 31644
- **Formal Certification** is granted to repositories which *in addition t*o Basic Certification obtain full external audit and certification based on ISO 16363 or DIN 31644

Certification Standards: Data Seal of Approval (DSA)

- DANS initiative (2005/6)
- International Board
- 16 guidelines
- Self assessment
- Transparency
- 24 seals awarded since 2010

Data producers are responsible for the quality of research data, repositories for storage and long-term access, and users for correct use of data

The research data:

- can be found on the Internet
- are accessible (clear rights and licenses)
- are in a usable format
- are reliable
- can be referred to (persistent identifier)

DANS

http://datasealofapproval.org/

Driven by data

The Guidelines 2014-2015

Guidelines Relating to Data Producers:

1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data and compliance with disciplinary and ethical norms.
2. The data producer provides the data in formats recommended by the data repository.
3. The data producer provides the data together with the metadata requested by the data repository.

Guidelines Related to Repositories (4-8):

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 including, when applicable, regulations governing the protection of human subjects.
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 work flows across the data life cycle.

Guidelines Related to Repositories (9-13):

9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.
10. The data repository enables the users to discover and use the data and refer to them in a persistent way.
11. The data repository ensures the integrity of the digital objects and the metadata.
12. The data repository ensures the authenticity of the digital objects and the metadata.
13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.

Guidelines Related to Data Consumers (14-16):

14. The data consumer complies 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 the relevant sector 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 data.

DSA self-assessment & peer review

- Complete a self-assessment in the DSA online tool. The online tool takes you through the 16 guidelines and provides you with support
- Submit self-assessment for peer review. The peer reviewers will go over your answers and documentation
- Your self-assessment and review will not become public until the DSA is awarded.
- After the DSA is awarded by the Board, the DSA logo may be displayed on the repository's Web site with a link to the organization's assessment.

Certification Standards: DIN 31644

- *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive – NESTOR*, Deutsche National Bibliothek
- 34 criteria
- Extended self-assessment process offers digital archives a harmonised and practical method of checking whether they are trustworthy. If the reviewed assessment yields a positive result they are entitled to publicise this by using the nestor Seal for Trustworthy Digital Archives.

DIN

nestor Seal 2013

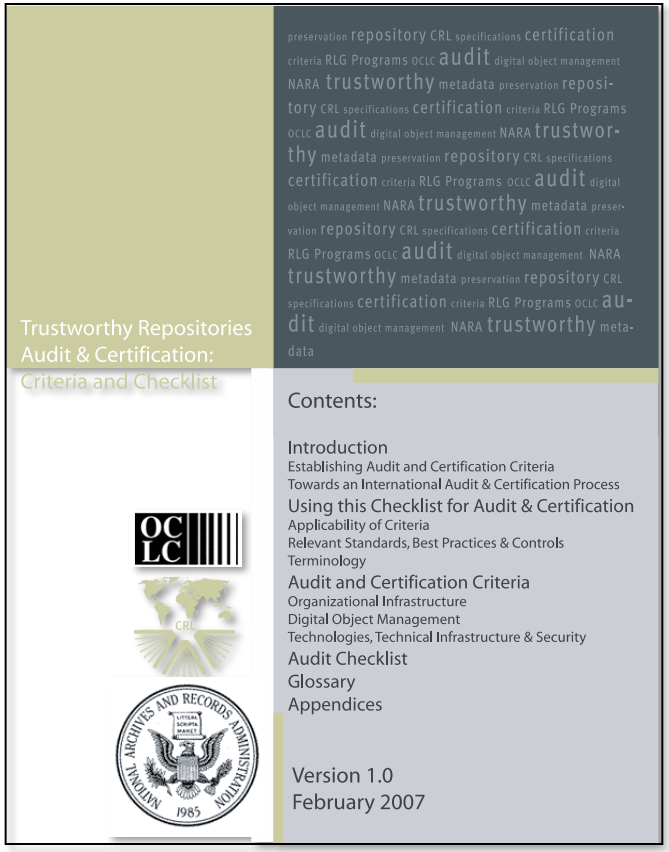
DANS

http://www.langzeitarchivierung.de

Driven by data

Certification Standards: ISO 16363

- Based on Open Archival Information System (OAIS) and Trusted Repository Audit and Certification (TRAC)
- Over 100 metrics
- Test audits 2011 by PTAB (Primary Trustworthy Digital Repository Authorisation Body
- Full external auditing process



ISO

DANS

http://www.iso16363.org/

Driven by data

Other certification/assessment procedures

- Certification and Assessment by Center for Research Libraries (CRL) - http://www.crl.edu/archiving-preservation
  - Metrics based on Trustworthy Repositories Audit and Certification checklist (TRAC)
  - CRL Certification Advisory Panel represents the various sectors of its membership
- The Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) - http://www.repositoryaudit.eu/
  - toolkit for use by repository administrators to (self) assess the risks to their digital archiving systems
- Certification (accreditation) of ICSU World Data System - http://www.icsu-wds.org/community/membership/certification
  - aims at transition from existing stand-alone WDCs and Services to a common globally interoperable distributed data system
  - criteria even less specific than DSA

On-going work on Trust

- Work Package on "Trust" within APARSEN project: http://www.alliancepermanentaccess.org/
- European Framework for Audit and Certification of Digital Repositories – now also in collaboration with accreditation of ICSU World Data System: http://www.trusteddigitalrepository.eu
- Research Data Alliance Interest Group on Certification: https://www2.rd-alliance.org/internal-groups/rdawds-certification-digital-repositories-ig.html

RESEARCH DATA ALLIANCE

Driven by data

Do we need certification?

No (devil's advocate)	Yes
Trustworthiness of digital repositories is an illusion • Too complicated to measure • Impossible to maintain over time • Too informal, too expensive and too time consuming • "Reputation and reuse are more important than certification" (Ron Dekker, NWO)	**Trustworthiness of digital repositories is necessary** • How else underpin claims to be a "trusted digital repository" • Different levels for different needs • To be repeated at certain intervals
Objective and consistent auditing is an illusion • "If auditing becomes a career, what will happen to objectivity?" (Helen Tibbo) • Impossible to guarantee consistency around the globe	**Objective and consistent auditing can be done** • Auditing is a career in many other areas • Requirements for bodies providing audit and certification of candidate Trustworthy Digital Repositories exist since 2011 • Some variation according to local requirements is not a problem

Data Archiving and Networked Services

Thank you for your attention

www.dans.knaw.nl
www.narcis.nl
peter.doorn@dans.knaw.nl

www.ftk.de
mhemmje@ftk.de