



Data Archiving and Networked Services

European digital repository certification: the way forward

Ingrid Dillo (DANS)

EUDAT 3rd User Forum
Prague, 24 April 2014

DANS is an institute of KNAW en NWO



Driven by data



Content



- DANS
- Data and trust
- European framework for certification
DSA, DIN, ISO
- Other approaches
- Horizon 2020



What is DANS?

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

Mission: promote
and provide
permanent access
to digital research
information

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





DANS main services

- EASY: Electronic archiving system
- NARCIS: gateway to scholarly information in NL
- Expertise: RDM and certification of TDRs



[Home](#) [Services & Support](#) [EUDAT Communities](#) [EUDAT Events](#) [Working Groups](#) [News & Publications](#)

Search

Extending our network - DANS becomes an Associated Partner

Release Date: 20/03/2014



Proliferation of data



Trend of data sharing/open data policies;
recognition of the value of data

Advantages:

- Transparency and replication of research
- Re-use of data

Challenges:

- Data management
- Access and preservation





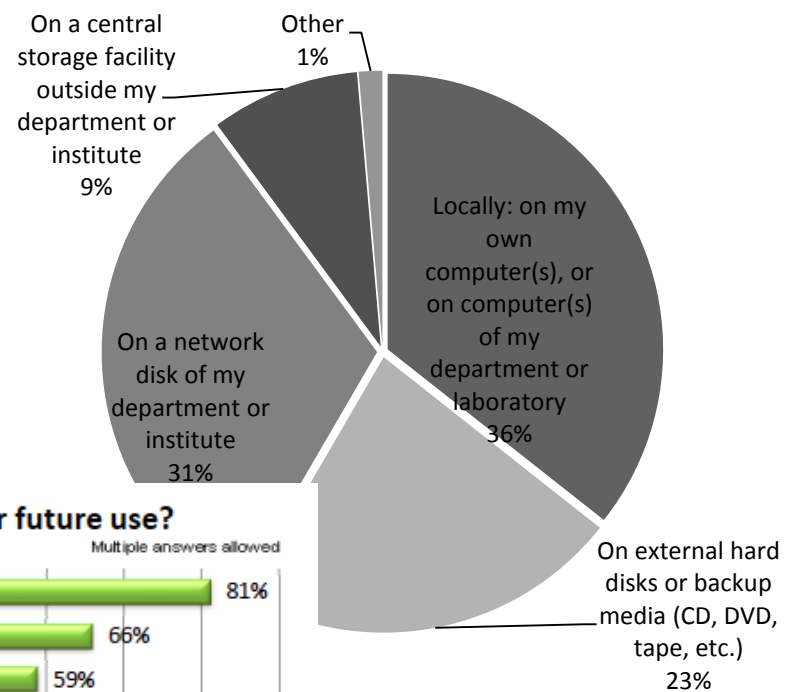
Reality of data sharing



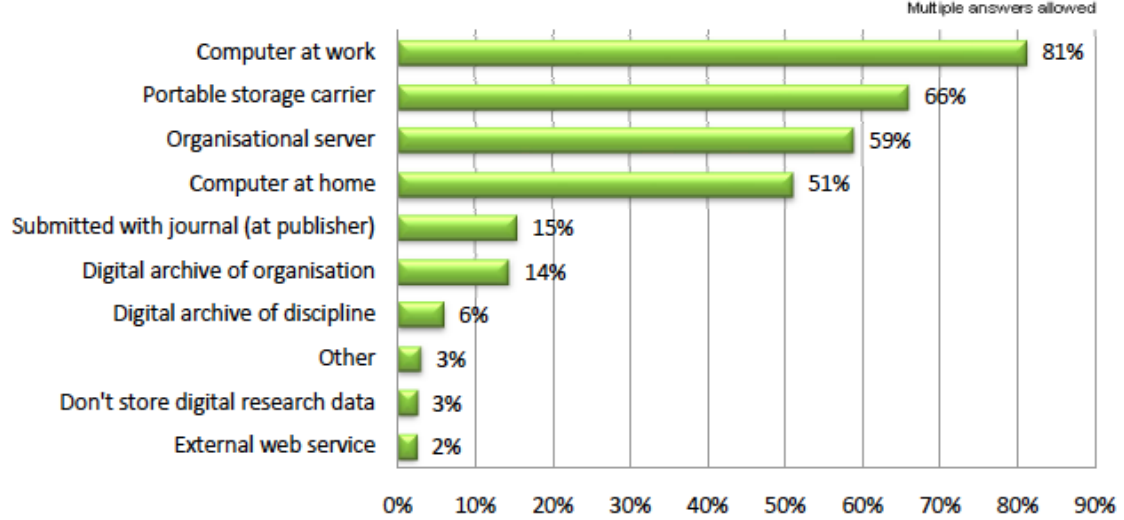
Insight into digital preservation of research output in Europe

PARSE insight

survey report



Where do you as a researcher store your data for future use?





Trust

Trust is at the very heart of storing and sharing data

- Users
- Depositors
- Funders





What do we rely on?

RANG IS
ALLEEN
RANG ALS
ER RANG
OP STAAT

RANG
TONNEMA N.V. SNEEK

CRISTAL
FRUITS

'n produkt van de KINGFABRIEKEN TONNEMA N.V. SNEEK

You can rely on us



Can you?





What is a trusted digital repository?



Things are not always what they say they are.

Things do not always state what they are.





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."



© Original Artist
Reproduction rights obtainable from
www.CartoonStock.com



Trust in data archives: an example



THE UK'S LARGEST COLLECTION OF DIGITAL RESEARCH DATA IN THE SOCIAL AND 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

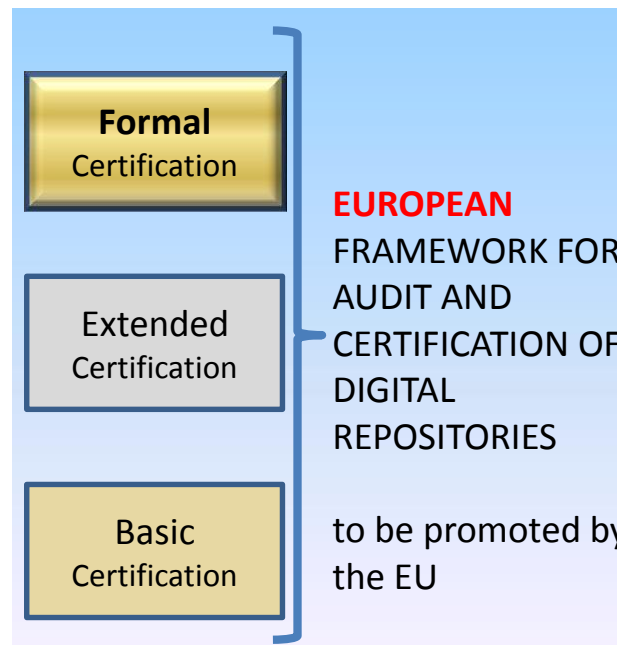
Any organisation which provides access to data over a long period of time should be fully trusted only with a public statement describing the practices they follow and the provenance of data they provide. Standards of trust are critical.



Certification of digital repositories



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





DSA: basic certification

- DANS initiative (2005/6)
- International Board (2009)
- 16 guidelines
- Self assessment&review
- Transparency
- Almost 30 seals awarded since 2010; 25 coming

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)

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 data.

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

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

4. The data producer provides the research data in a format that is suitable for digital archiving and promulgates it.

5. The data producer provides the research data in a format that is suitable for digital archiving and promulgates it.

6. The data producer provides the research data in a format that is suitable for digital archiving and promulgates it.

7. The data producer provides the research data in a format that is suitable for digital archiving and promulgates it.

8. The data producer provides the research data in a format that is suitable for digital archiving and promulgates it.

9. The data producer provides the research data in a format that is suitable for digital archiving and promulgates it.

10. The data producer provides the research data in a format that is suitable for digital archiving and promulgates it.

11. The data producer provides the research data in a format that is suitable for digital archiving and promulgates it.

12. The data producer provides the research data in a format that is suitable for digital archiving and promulgates it.

13. The data producer provides the research data in a format that is suitable for digital archiving and promulgates it.

14. The data producer provides the research data in a format that is suitable for digital archiving and promulgates it.

15. The data producer provides the research data in a format that is suitable for digital archiving and promulgates it.

16. The data producer provides the research data in a format that is suitable for digital archiving and promulgates it.



DIN 31644: extended certification

- 34 criteria written by German NESTOR-group and adopted in Germany as DIN31644
- Self-assessment procedure by NESTOR leads to NESTOR seal
- Review of the assessment by 2 reviewers, appointed by NESTOR
- Self assessment and evidence on website
- No seals acquired yet..

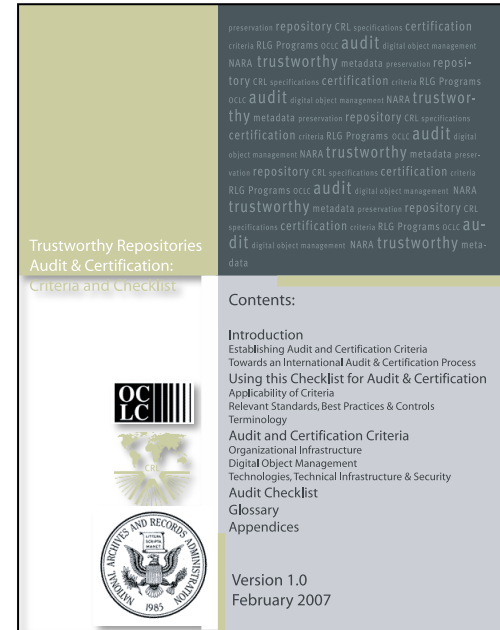


http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.htm |



ISO 16363: formal certification

- 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 16919: Requirements for bodies providing audit and certification of candidate trustworthy digital repositories
- No ISO certifications yet..





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



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 managers of digital data. The preservation of

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.



CESSDA
social science data

metadata accessing data
data harmonisation ethics
data catalogue confidentiality
data preservation

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”

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.



H2020 Open data pilot

- Current Guidelines Open Data Pilot do not mention “trusted digital repositories”; Annex 2 mentions “certified repositories” once.
- Understandable: there are yet few certified TDRs.
- E-infrastructure call H2020:
*Services to ensure the quality and reliability of the e-infrastructure, **including certification mechanisms for repositories and certification services** to test and benchmark capabilities in terms of resilience and service continuity of e-infrastructures.*





Proposal RDM and Certification

- Initiative of DCC and DANS
- RDM and certification tools and services supporting the implementation of the open data policies of the European Commission, national funders and publishers.
- Focused project with tangible outcomes
- Limited number of dedicated and specialized partners



Proposal certification

- website with basic information on certification and tools for the different certification standards (awareness raising);
- online certification tools for the three standards;
- consultancy repositories;
- training of reviewers/certifiers of repositories;
- Dissemination (global outreach through RDA);
- Exploitation plan - development of a sustainable business model for the certification services



Proposal certification

- Coordination / cooperation with other proposals: EUDAT, EGI, OpenAire, APA?
- **Tile construction of complementary parts**, together providing the necessary conditions for the construction of a European network of TDRs
- EUDAT as a user community: feedback and testbed





EUDAT and trust



Collaborative Data Infrastructure: *"long-term, safe, trusted by its user communities"*

- Could certification have an added value in building trust into EUDAT services?
- Which building blocks of the CDI could be certified?
- How could this be incorporated into an EUDAT2 proposal?



Session 1

Policy, sustainability and certification

RESEARCH DATA ALLIANCE FOURTH PLENARY MEETING

22 – 24 September 2014
Amsterdam, the Netherlands | Meervaart conference centre

www.rd-alliance.org/rda-fourth-plenary-meeting.html

The Research Data Alliance (RDA) builds the social and technical bridges that enable open sharing of data to address cross-border & cross-disciplinary challenges.

The current global research data

landscape is highly fragmented, by disciplines or by domains, from oceanography, life sciences and health, to agriculture, space and climate, social sciences and humanities.



Reaping the fruits:

The Research Data Alliance Fourth Plenary Meeting in Amsterdam showcases the first concrete outputs from the RDA Working Groups

Johnny saw some fine
plums hanging,
Oh, like eggs, so very
large...
(Reaping the fruits from
the first Dutch children's
book "Kleine gedigten
voor kinderen" by Hieronymus van Alphen, 1778).

RDA PLENARY 4 MEETING IS CO-ORGANISED BY:

3TU.Datacentrum Data Archiving and Networked Services **DANS** netherlands Science center



UNIVERSITY OF AMSTERDAM

For more information on the Research Data Alliance – www.rd-alliance.org

RDA FOURTH PLENARY MEETING CO-LOCATED EVENTS

<https://www.rd-alliance.org/plenary4-colocated-events.html>



50 years Data Archiving and The Dutch Data Award



APARSEN: Alliance Permanent Access to the Records of Science in Europe Network



Creative-B: Workshop "Global Roadmap for biodiversity research infrastructures"



Data Science Research Center Amsterdam: new opportunities for Research Data



Data Seal of Approval Conference 2014



EUDAT 3rd Conference - Bringing data infrastructures to Horizon 2020



ORCID and DataCite: Towards Holistic Open Research

WHAT DATA SHARING PROBLEM ARE YOU TRYING TO SOLVE?

Get involved by becoming an RDA Member and signing up to the Working & Interest Group. Participation in the RDA is free and open to anyone who agrees to the RDA principles. Scientists come together through focussed Working Groups, exploratory Interest Groups to exchange knowledge, share discoveries, discuss barriers and potential solutions, explore and define policies and test as well as harmonise standards to enhance and facilitate global data sharing. Coupled with this RDA boasts a broad, committed membership of individuals and organisations dedicated to improving data exchange. <https://www.rd-alliance.org/user/register>



Thank you for your attention

We look forward to September...

<https://rd-alliance.org/rda-fourth-plenary-meeting.html>

RDA 4th Plenary 22-24 September
EUDAT 3rd Conference 24-26 September

..and hope to welcome you all in the Amsterdam!

ingrid.dillo@dans.knaw.nl