

*Data Archiving and Networked Services*

## **Trusted Digital Archives**

**Peter Doorn**

Data Archiving and Networked Services



**ALLEA | ALL European Academies**

European Federation of Academies of  
Sciences and Humanities

*14th General Assembly  
29/30 April 2013 | Berlin  
Berlin-Brandenburg Academy of Sciences and Humanities*



# Contents

- Two slides about DANS
- Data fraud and data quality
- Trust and responsible data management
  - About trust and certification
  - how to build trust
  - European framework for certification
  - DSA, DIN, ISO
  - pros and cons of certification
- The role of journals
  - Data Availability Policies
  - Data reviews
- Conclusions



# What is DANS?

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

Mission: promote  
permanent access  
to digital research  
information

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



# DANS main activities and services

- Encourage researchers to self-archive and reuse data by means of our Electronic Archiving System EASY
- Our largest digital collections are in archaeology, social sciences and history (moving into other domains)
- Provide access, through [Narcis.nl](https://narcis.nl), to thousands of scientific datasets, e-publications and other research information in the Netherlands
- Data projects in collaboration with research communities and partner organisations
- Advice, training and support (Data Seal of Approval, Persistent Identifier Infrastructure)
- R&D into archiving of and access to digital information

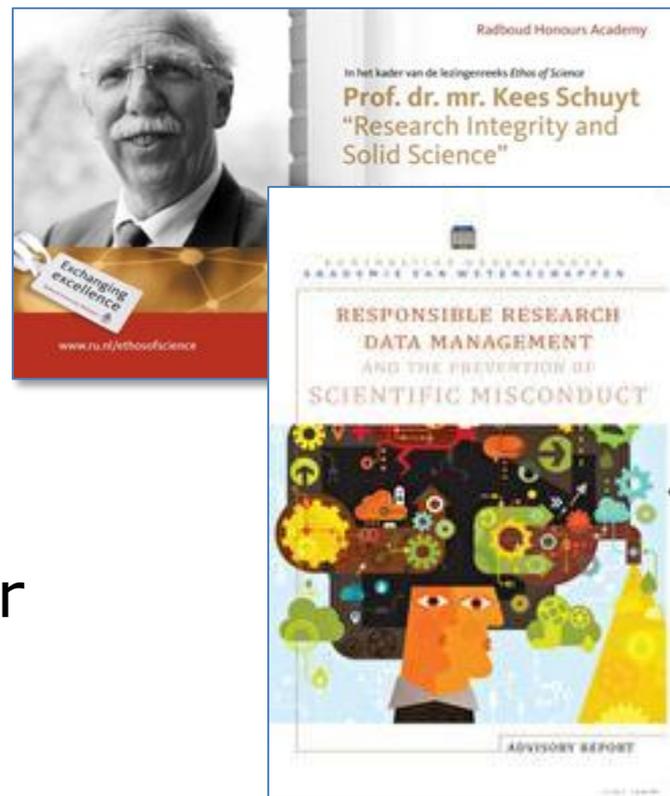






## The KNAW "Schuyt report" on data practices

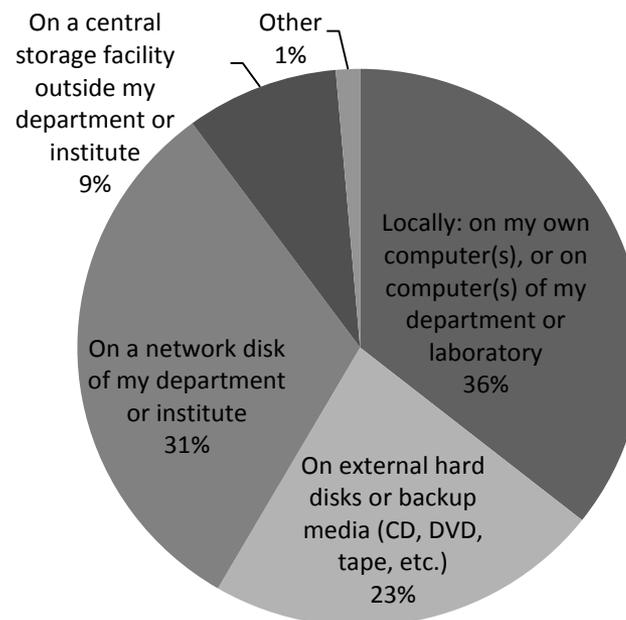
- 94 interviews with researchers
- Variation across and within disciplines
- Pattern: data management in small-scale research more risky than in big science
- Risk: missing checks and balances, especially in phase after granting research proposal and before publication
- Peer pressure is an important control mechanism



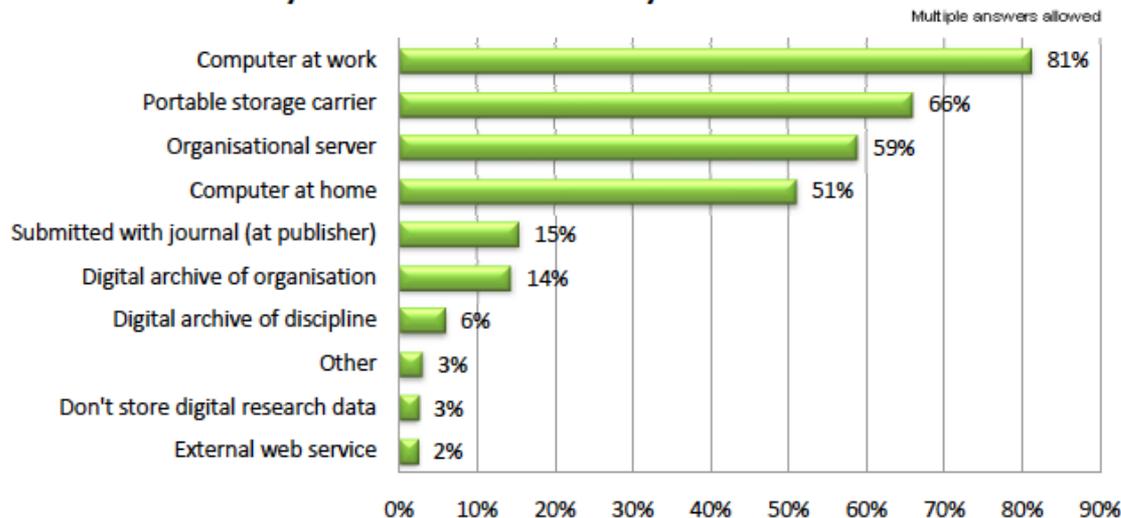
[http://www.knaw.nl/Content/Internet\\_KNAW/publicaties/pdf/20131009.pdf](http://www.knaw.nl/Content/Internet_KNAW/publicaties/pdf/20131009.pdf)



# Data sharing cultures across scientific domains



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



Insight into digital preservation of research output in Europe



survey report



## Dutch Academy (KNAW) endorses Schuyt recommendations

- Hans Clevers, President KNAW: “Maximum access to data promotes the pre-eminently scientific approach whereby researchers check one another’s findings and build critically on one another’s work”
- The Academy supports the free movement of data and results. Taking into account variations across and within scientific disciplines, free availability of data should be the default.
- We do not need additional rules or codes of conduct, but should focus on revitalising existing rules and making these better known.
- Examination of data management practices should become an integral part of official research evaluations.



# Trust: What do we mean? What do we rely on?





# What is a trusted digital repository?



Things are not always what they say they are.

Things do not always state what they are.





# Trust in research data

- Trust is at the very heart of creating, storing and sharing data:
  - Data creators
  - Data users
  - Data repositories
  - Funders
- Data quality
  - valid
  - accurate
  - consistent
  - integrity
  - timely
  - complete



Trust by definition  
implies uncertainty



## What is trust built on (in trusted archives)?

- 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."





# Standards of trust



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.**



# 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.



# Do we need certification of digital archives?

## No (devil's advocate)

Trustworthiness of digital repositories is an illusion



- Too complicated to measure
- Impossible to maintain over time
- Too informal, too expensive and too time consuming

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

## Yes

Trustworthiness of digital repositories is a necessity



- Three levels of certification, clear and balanced
- Different levels for different needs

Objective and consistent auditing can be done



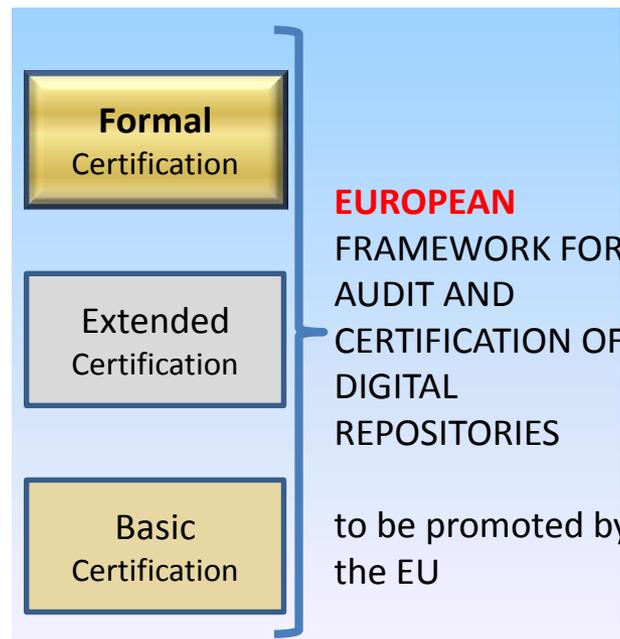
- Auditing is a career in many other areas
- Some variation according to local requirements is permitted



# Certification of digital repositories



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





## Three framework levels

- **Basic Certification** is granted to repositories which obtain the 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 to* Basic Certification obtain full external audit and certification based on ISO 16363 or DIN 31644





# Certification Standards: Data Seal of Approval (DSA)

- DANS initiative
- International Board
- 16 guidelines
- Self assessment
- Transparency
- 14 seals rewarded

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





# Certification Standards: DIN 31644

- *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive – NESTOR, Deutsche National Bibliothek*
- 34 criteria
- Test audits 2013

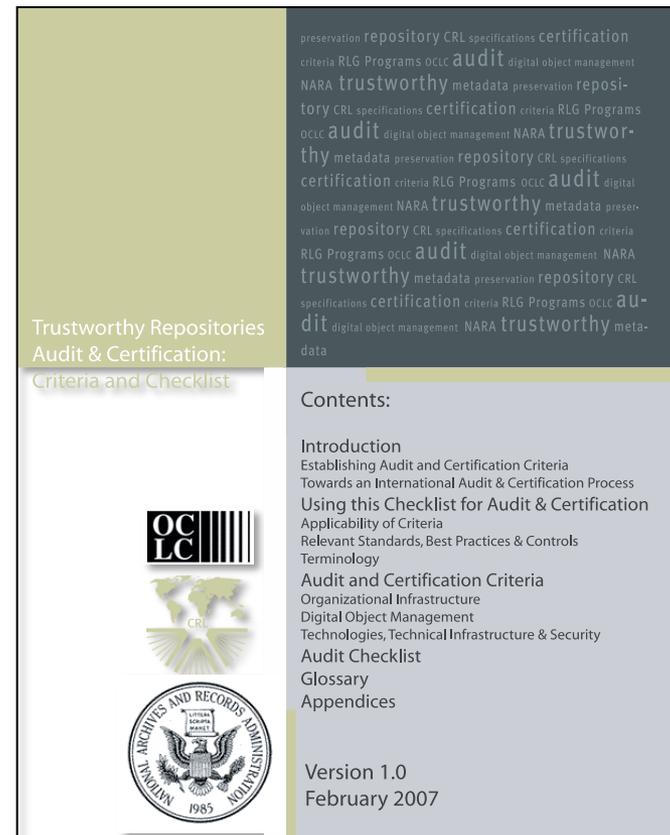


**DIN**



# Certification Standards: ISO 16363

- Based on Open Archival Information System (OAIS) and Trusted Repository Audit and Certification (TRAC)
- Over 100 metrics
- Test audits 2011
- Full external auditing process





## On-going work on Trust

- Work Package on “Trust” within APARSEN project
- European Framework for Audit and Certification of Digital Repositories – seek collaboration with ICSU World Data System (accreditation)
- Research Data Alliance Group on Certification





# Data quality and the role of journals

- Quality control of scientific output concentrates on publications: peer review
- Few journals have a data availability policy (DAP):
  - When a paper is submitted for peer-reviewed publication, data access for reviewers is required
  - When the paper is published, readers should be able to access the data
- Data repositories: marginal checking (metadata, format, privacy)
- Community reviews by data users

Aspect	Rating						Average rating
	(5)	(4)	(3)	(2)	(1)	(n/a)	
data quality	4	2	0	0	0	0	★★★★★ (4.67/5)
quality of the documentation	4	1	0	0	1	0	★★★★☆ (4.17/5)
completeness of the data	4	2	0	0	0	0	★★★★★ (4.67/5)
consistency of the dataset (if applicable)	4	1	0	0	0	1	★★★★★ (4.8/5)
structure of the dataset (if applicable)	4	2	0	0	0	0	★★★★★ (4.67/5)
usefulness of the file formats	3	2	0	1	0	0	★★★★☆ (4.17/5)

5 out of 6 reviewers of this dataset recommend the use of it.





## Conclusions

- Sharing data cannot prevent data fraud, but it increases the transparency of research, therefore reducing the risk of fraud
- Archiving and publishing data, via trustworthy data repositories, increases the reliance on data
- Funders should require that project proposals contain a data management plan, and that such plans contain a section on accessibility of data after publication of the results
- Not sharing data:
  - increases the risk of “sloppy science”
  - increases the risk of data fraud

S  
W  
A  
E

## *Data Archiving and Networked Services*

**Thank you for your attention**

[www.dans.knaw.nl](http://www.dans.knaw.nl)

[www.narcis.nl](http://www.narcis.nl)

[peter.doorn@dans.knaw.nl](mailto:peter.doorn@dans.knaw.nl)



DANS is an institute of KNAW en NWO

*Driven by data*