European digital repository certification: the way forward

Ingrid Dillo (DANS)

Workshop The Concept of Trust

Stockholm, 4 June 2014

DANS is an institute of KNAW en NWO

Driven by data

Content

- DANS
- Data and trust
- European framework for certification
  - DSA, DIN/NESTOR, ISO
- Horizon 2020 and EUDAT2

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

Data Seal of Approval

The Dataverse Network® Project

Dutch Dataverse Network

NARCIS: Gateway to scholarly information In the Netherlands

Driven by data

'Who' is DANS?

About 50 employees: IT-specialists, archivists, information experts, policy makers and communication advisors

DANS

Driven by data

DANS associated partner in EUDAT

Definition of research data

Factual records, which may take the form of numbers, symbols, text, images or sounds,

used as primary sources for research,

commonly accepted in the research community as necessary to validate research findings.

DANS

Driven by data

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

The Dutch data landscape in 32 interviews and a survey

Insight into digital preservation of research output in Europe

PARSE insight

survey report

- On a central storage facility outside my department or institute 9%
- Other 1%
- Locally: on my own computer(s), or on computer(s) of my department or laboratory 36%
- On a network disk of my department or institute 31%
- On external hard disks or backup media (CD, DVD, tape, etc.) 23%
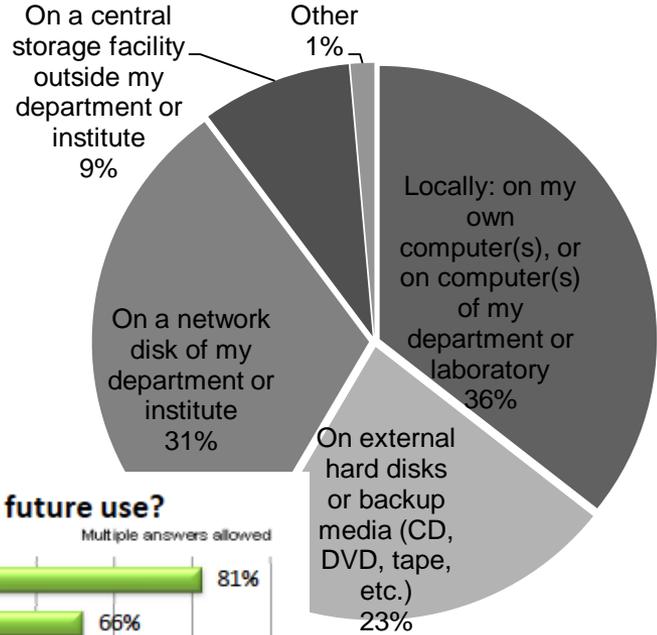
Where do you as a researcher store your data for future use?

Multiple answers allowed

Storage location	%
Computer at work	81%
Portable storage carrier	66%
Organisational server	59%
Computer at home	51%
Submitted with journal (at publisher)	15%
Digital archive of organisation	14%
Digital archive of discipline	6%
Other	3%
Don't store digital research data	3%
External web service	2%

DANS

Driven by data
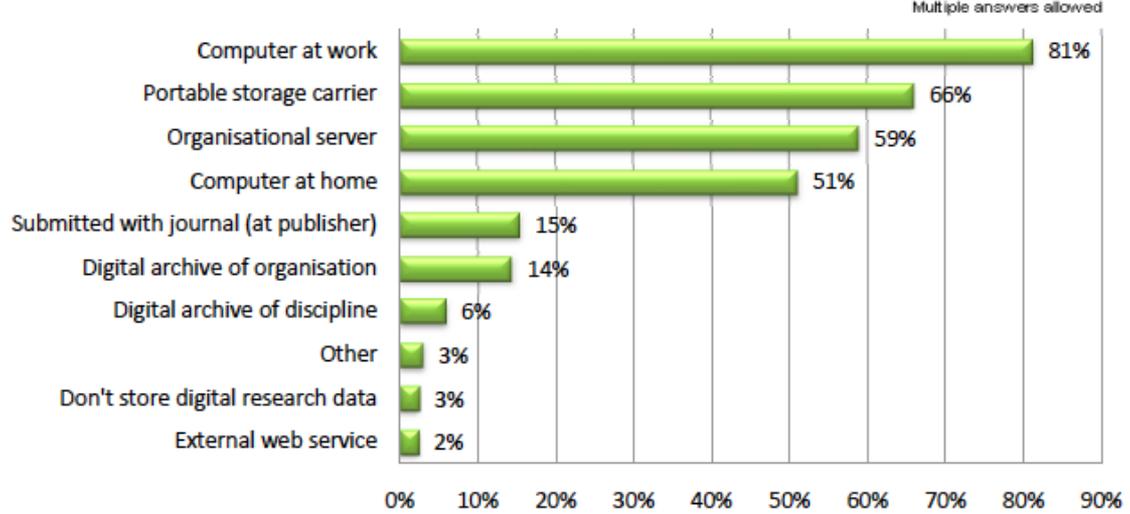
Report Norwegian Research Council

- 80% of the researchers agree that open access to research data enhances research, and that it is an ethical obligation of research to make research data available for validation.

- 85% of the researchers reported storing their data on their own storage devices or on a server available at their institution.

- The researchers need a **better infrastructure for archiving and sharing data, systems for citing and crediting data as well as new guidelines, training and standards for sharing data.**
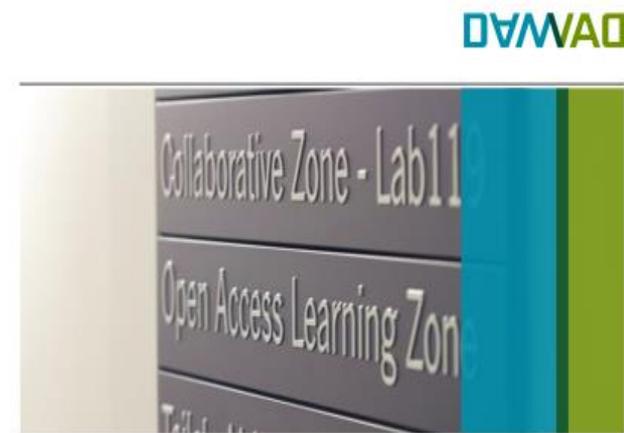
DANS

Driven by data

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

You can rely on us

Can you?

DANS

Driven by data

What is a trusted digital repository?

Ceci n'est pas une pipe.

IF YOU'D JUST *TRUST ME*...THEN I WOULDN'T HAVE TO LIE TO YOU!

GOSH! YOU'RE *RIGHT!* CAN YOU *EVER* FORGIVE ME?

©Last Kiss Inc. 2009

Things are not always what they say they are.
Things do not always state what they are.

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."

IN PRINCIPLE WE'RE HAPPY WITH THE 'TRUST' IDEA...AS LONG AS IT'S PROPERLY MONITORED AND REGULATED!

THUMP THUMP!

GP

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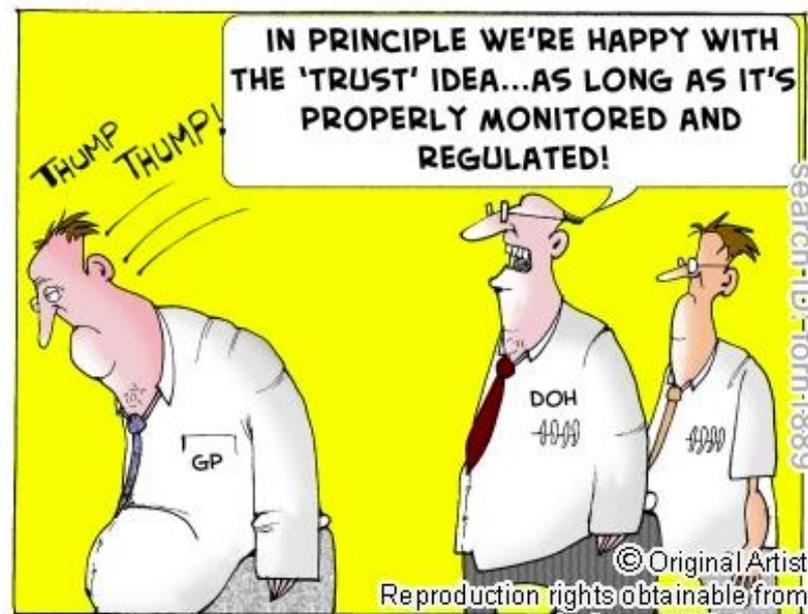
© Original Artist
Reproduction rights obtainable from www.CartoonStock.com

DANS

Driven by data

Trust in data archives: an example

UK · DATA ARCHIVE

THE UK'S LARGEST COLLECTION OF DIGITAL RESEARCH DATA IN THE SO[...] 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

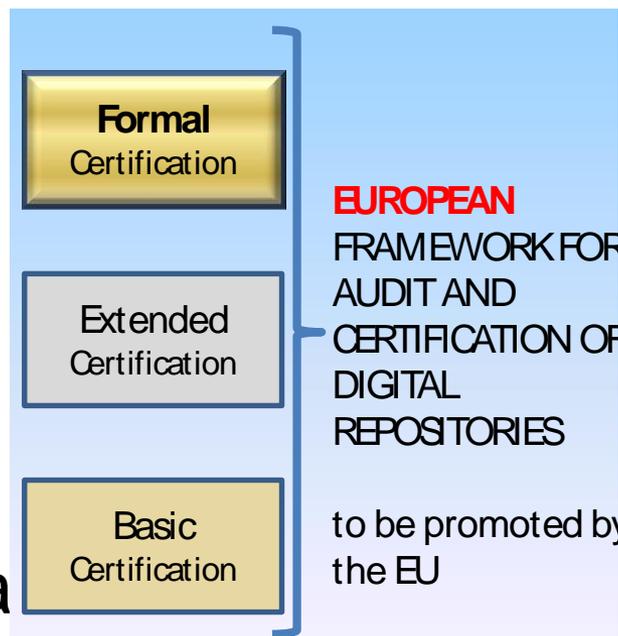OVERVIEW   **DATA SEAL OF APPROVAL**   **ISO16363**   **DIN 31644**

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

Certification of digital repositories

make
the digital world
more
reliable

- International framework
- 3 standards
- 3 levels (basic, extended, forma

Formal Certification

Extended Certification

Basic Certification

EUROPEAN FRAMEWORK FOR AUDIT AND CERTIFICATION OF DIGITAL REPOSITORIES

to be promoted by the EU

Data Seal of Approval

nestor Seal 2013

ISO

http://www.trusteddigitalrepository.eu

Driven by data

DSA: basic certification

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

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)

http://datasealofapproval.org/

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 l

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..

http://www.iso16363.org/

DANS

Driven by data
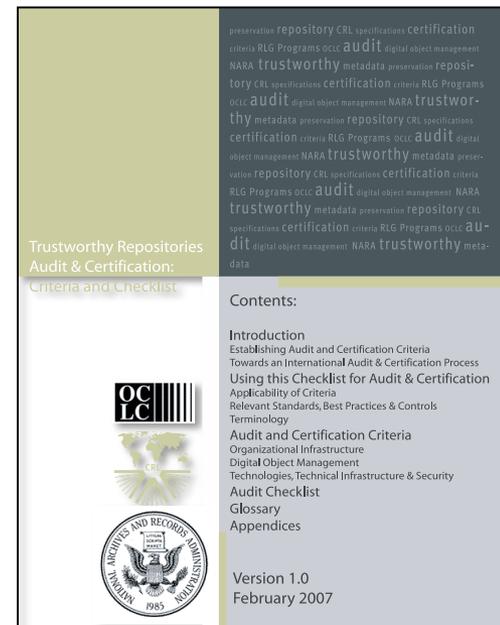
ESFRI Research Infrastructures and Trust
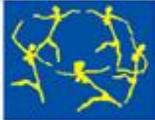
Compliance as a DARIAH Trusted Digital Repository

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.

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.

DANS

Driven by data

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.*

EUDAT2 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?

DANS

Driven by data

EUDAT2 Proposal

- Certification programme based on the DSA Guidelines:

  - Certification of a limited number of community repositories that make use of EUDAT services

  - Making "DSA-proof" part of the generic data centers that offer EUDAT services

  - Making the policy framework around the EUDAT services "DSA-proof" (SLAs, licensing, etc.)

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.

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).

Reaping the fruits:

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

RDA PLENARY 4 MEETING IS CO-ORGANISED BY:

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

RDA RESEARCH DATA ALLIANCE   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

rdnl	50 years Data Archiving and The Dutch Data Award
APARSEN Alliance Permanent Access to the Records of Science in Europe Network	APARSEN: Alliance Permanent Access to the Records of Science in Europe Network
CREATIVE-B	Creative-B: Workshop "Global Roadmap for biodiversity research infrastructures"
Data Science Research Center Amsterdam	Data Science Research Center Amsterdam: new opportunities for Research Data
Data Seal of Approval	Data Seal of Approval Conference 2014
EUDAT	EUDAT 3rd Conference - Bringing data infrastructures to Horizon 2020
ODIN	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*

We look forward to September...

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

22-24 September
Co-located events 21 & 24-26 September

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

Thank you for your attention

www.dans.knaw.nl

http://datasealofapproval.org/en/

ingrid.dillo@dans.knaw.nl

DANS is an institute of KNAW en NWO

Driven by data