

UCLA

Presentations

Title

Motivations for Sharing and Reusing Data: Complexities and Contradictions in the Use of a Digital Data Archive

Permalink

<https://escholarship.org/uc/item/5sb7p80v>

Authors

Borgman, Christine L.
Scharnhorst, Andrea
Berg, Henk van den
et al.

Publication Date

2016-10-05

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Motivations for Sharing and Reusing Data: Complexities and Contradictions in the Use of a Digital Data Archive

Christine L. Borgman

Distinguished Professor and Presidential Chair in
Information Studies

University of California, Los Angeles

@scitechprof

Project collaborators:

DANS: Andrea Scharnhorst, Henk van den Berg

UCLA: Ashley Sands, Milena Golshan

ANDS: Andrew Treloar

Seminar Presentation, Quello Center on Media,
Communication, and Information Policy

Michigan State University

October 5, 2016



Christine Borgman



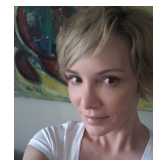
Peter Darch



Ashley Sands



Irene Pasquetto



Bernie Randles



Milena Golshan



UCLA Center for
Knowledge Infrastructures

<http://christineborgman.info>

<https://knowledgeinfrastructures.gseis.ucla.edu>

Data Archiving and Networked Services

DANS

<https://dans.knaw.nl/en>



Data sharing policies



- OECD
- European Union
- U.S. Federal research policy
- Research Councils of the UK
- Australian Research Council
- Individual countries, funding agencies, journals, universities



Supported by
wellcometrust



Australian Government
National Health and Medical Research Council



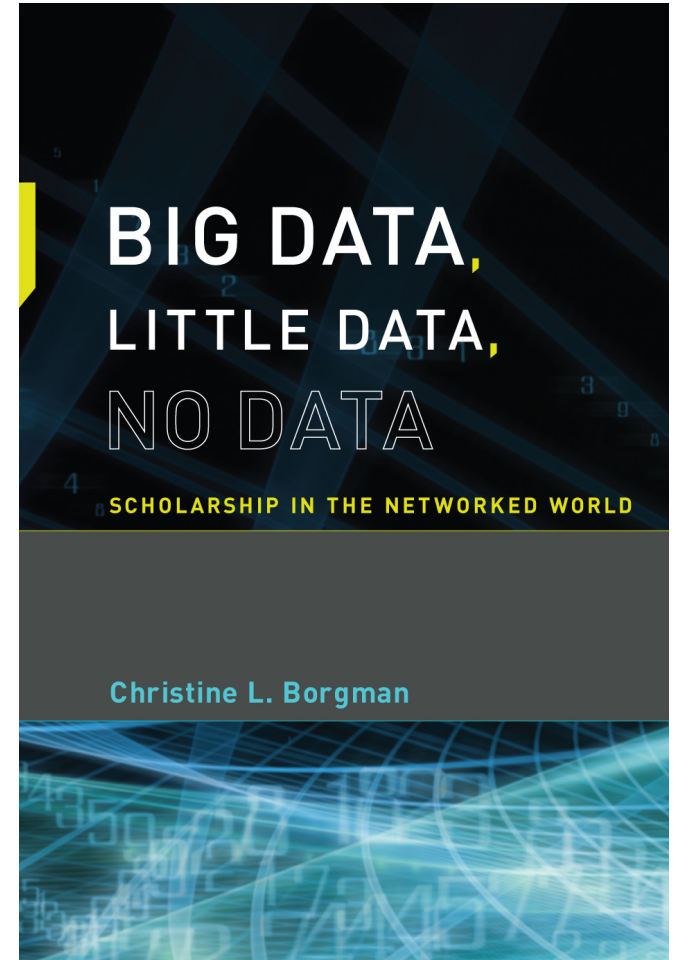
National Science Foundation
WHERE DISCOVERIES BEGIN

Policy RECommendations for Open Access to Research Data in Europe



Why Share Research Data?

- To reproduce research
- To make public assets available to the public
- To leverage investments in research
- To advance research and innovation



MIT Press, 2015

Stakeholders in Data

- Researchers
- Data creators
- Data users / reusers
- Funding agencies
- Publishers
- Universities
- Libraries, archives
- Educators, students
- Public...

Lack of incentives to share data



- Rewards for publication
- Effort to document data
- Competition, priority
- Control, ownership

<http://www.buildingsrus.co.uk/.../target1.htm>



Data



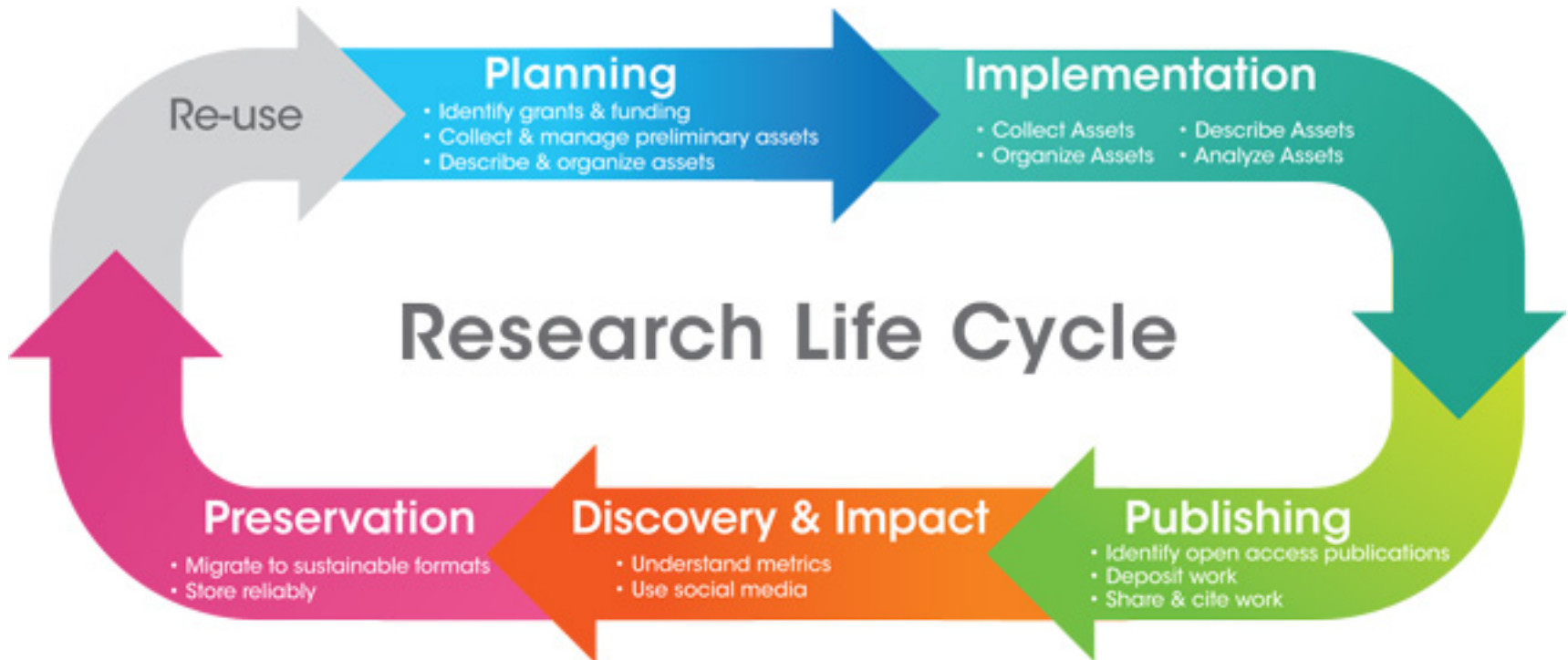
Data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.

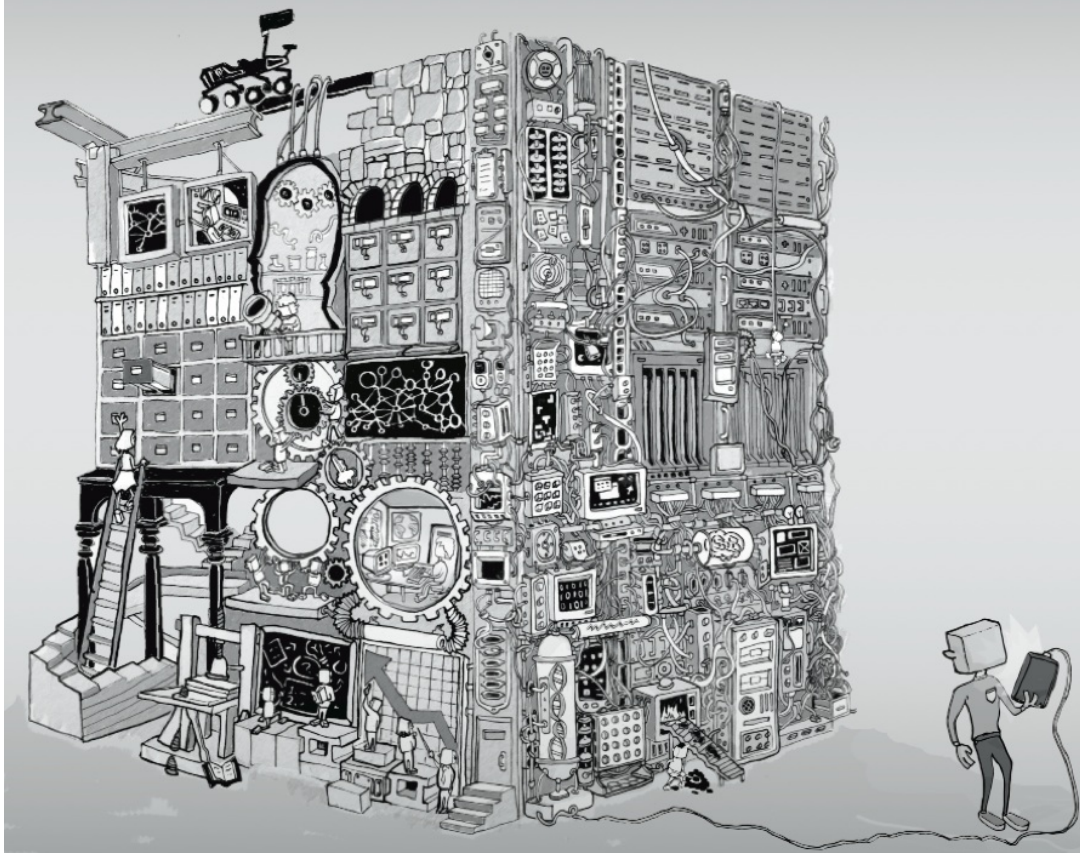
C.L. Borgman (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press

When to invest in data?



When to invest in data?





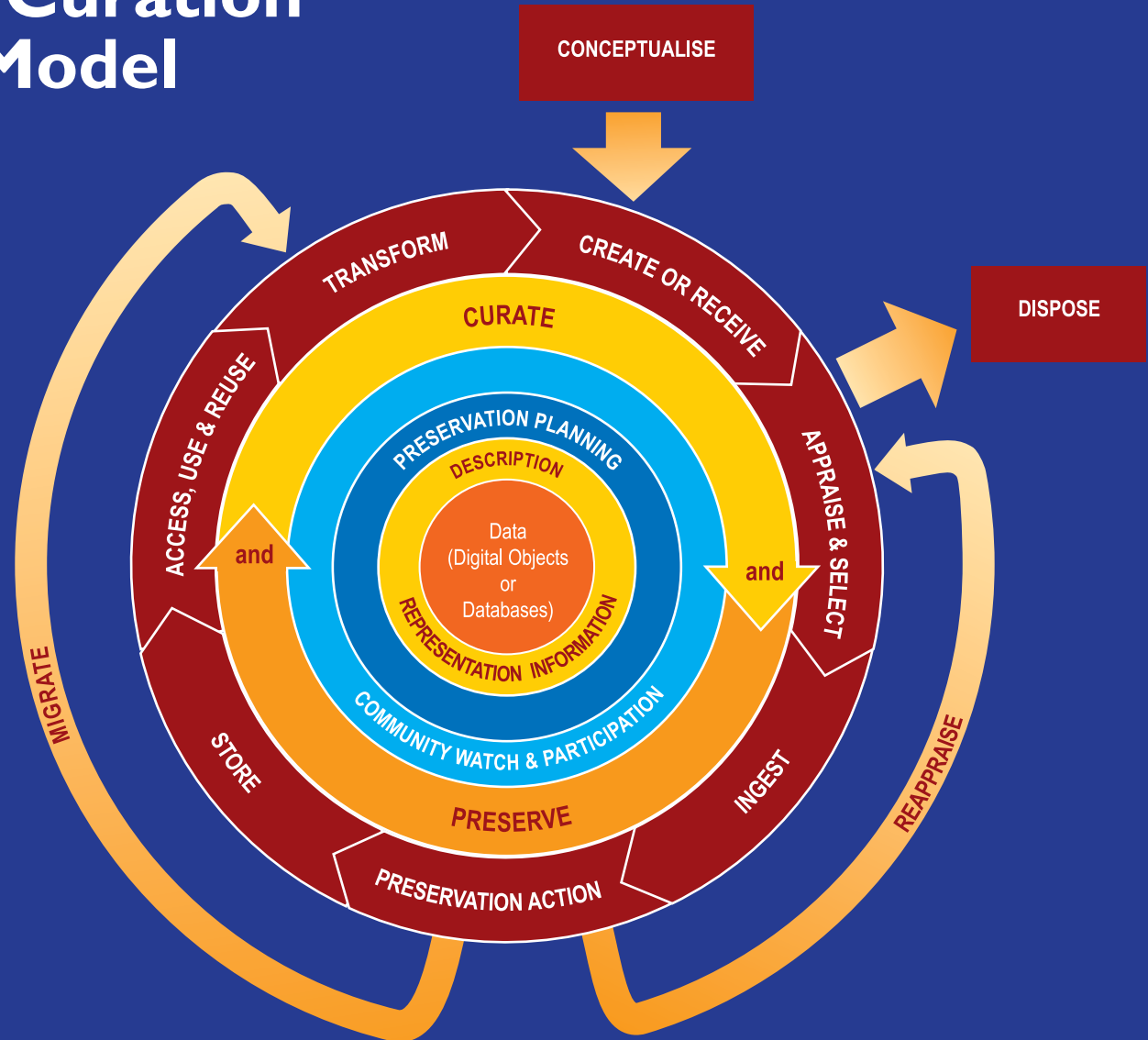
Knowledge Infrastructures:
Intellectual Frameworks and Research Challenges

*Report of a workshop sponsored by the National Science Foundation and the Sloan Foundation
University of Michigan School of Information, 25-28 May 2012*

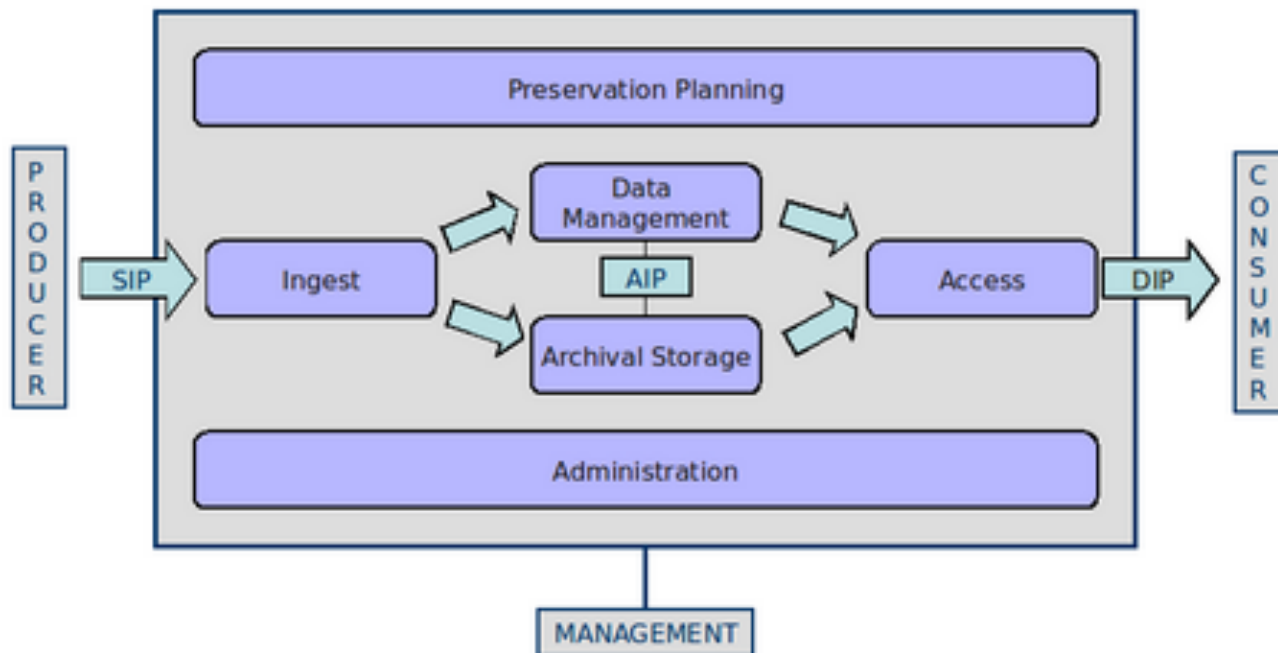
<http://knowledgeinfrastructures.org>



The DCC Curation Lifecycle Model



Open Archival Information System (OAIS) reference model (ISO-STD 14721)



<https://wiki.archivematica.org/Overview>

Search...



Trusted Data Services for Global Science



WORLD DATA SYSTEM

- Home
- About
- Community
- Data and Services
- Publications
- News
- Events

SciDataCon 2016

Advancing the Frontiers of Data in Research

Programme Available and Registration Still Open!
11–13 September 2016, Denver, Colorado, USA

Latest News

WDS Blog

Webinar #10: An Introduction to the DSA–WDS Common Requirements

Share

Upcoming Events & Deadlines

12 to 16 Dec 2016

AGU 2016 Fall Meeting

Publishing and Managing Data: The case for Trustworthy Digital Repositories

Data Seal of Approval

Latest News

WDS Blog

Data Seal
of Approval

Towa

There a
determi
qualifies

Webinar #10: An Introduction to the DSA–WDS Common Requirements

Share



29 Aug 2016

On Wednesday, 31 August at 12:00 UTC, Ingrid Dillo (WDS-SC Vice-chair) gave An Introduction to the DSA–WDS Common Requirements in the 10th WDS Webinar . Although this Webinar was open to all, it was particularly focussed towards current and prospective WDS Regular Members, who will transition to the new DSA–WDS Common Requirements from October. Since 2012, a Working Group consisting ...

[Read more](#)

DANS: Data Archiving and Networked Services

Data Archiving and Networked Services

Nederlands

Contact

Search this website



HOME

DEPOSIT

SEARCH

TRAINING AND CONSULTANCY

PROJECTS

ABOUT DANS

NEWS AND EVENTS

DANS promotes **sustained access** to digital research data files and encourages researchers to **archive** and **reuse** data.



ARCHIVING

Deposit your datasets in EASY or send research data and publications to NARCIS.

DEPOSIT



REUSE

Find datasets, publications, researchers, projects and institutions via NARCIS and EASY.

SEARCH



TRAINING & CONSULTANCY

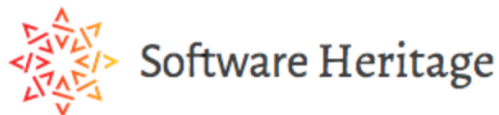
Let DANS advise you on data management and certification of digital archives.

ADVICE FROM DANS

SPOTLIGHT

DANS and Inria sign the MoU for Software Heritage

DANS will collaborate with Inria (France) on the development of the Software Heritage Initiative. On June 30 Inria officially announced the collaboration with DANS with the launch of the Software Heritage website.



NEWS

Your data paper in Research Data Journal for the Humanities and Social Sciences

RDJ, published by Brill publishers and DANS, is a peer reviewed e-only open access journal. Authors can submit their data papers online.

>>>

DANS
@DANSKNAW

U heeft nog ongeveer twee weken voor het indienen van KDP-voorstellen. Bezoek onze website voor meer informatie: bit.ly/2cSTIT4

3h

DANS



DANS Services and Systems

- Netherlands government funding; about 50 staff
- Member of World Data Systems, Data Seal of Approval, other partnerships
- Participates in international research infrastructure projects
- 50+ years of social science and humanities data
- Published datasets in EASY: 32,000+
- About 3.5 million files; total 5TB

Data Archiving and Networked Services

DANS

DANS is an institute of KNAW and
NWO



Driven by data

Services and Systems



Figure 2: Information system and services at DANS (no claim of completeness)

Research Questions

1. Who contributes data to DANS? How, when, why, and to what effects?
2. Who consumes data from DANS? How, when, why, and to what effects?
3. What role do archivists at DANS play in acquiring and disseminating data?

Research Methods

- Document analysis
- Ethnography
- Mining transaction logs
 - to characterize communities
 - To draw interview samples
- Interviews
 - DANS archivists and staff
 - DANS data contributors
 - DANS data consumers



A. Scharnhorst, C. Borgman, DANS 2014

Data Archiving and Networked Services

DANS

UCLA Center for Knowledge Infrastructures

Interviews Conducted 2015-2016

Stakeholders/ Participants	Number of Interviews	Domain Expertise	Occupation
Data contributors	9	Archaeology, History, Paleogeography (3) Labor Economics (1) Linguistics (1) Oral Histories (1) Information Science (1) Theology (1) Plant Biology (1)	Academic staff (4) Cultural institution staff (3) Private company staff (1 interview with 2 staff) Unaffiliated (1)
Data consumers	8	Archaeology, History (6) Political Science, Sociology, Public Administration (3)	Academic staff (3) Cultural institution staff (2) Citizen scientists (1) Students (2)
DANS staff	10	Archaeology and humanities (6) IT Development (4)	Archivists Project Managers IT developers

Findings so far

- Characterizing DANS
- Data contributors
- Data consumers
- DANS archivists and staff
- Infrastructure roles

Data Archiving and Networked Services

DANS

UCLA Center for Knowledge Infrastructures

Borgman, C. L., Van de Sompel, H., Scharnhorst, A., van den Berg, H., & Treloar, A. (2015). Who Uses the Digital Data Archive? An Exploratory Study of DANS. In *Proceedings of Association for Information Science and Technology*.

<http://doi.org/10.1002/pr2.2015.145052010096>

Characterizing DANS

- Technical view: Transaction logs
 - Signons, searches,
 - uploads, downloads, etc
- Systems view
 - EASY
 - NARCIS
- Staff view
 - Organization
 - Community
- Users view
 - Content
 - Services



DATA AFTER RESEARCH

You can store your data in a sustainable manner with the online EASY archiving system upon completing the research.

[VISIT EASY](#)

[Read more about depositing data](#)



RESEARCH INFORMATION

You can add data on researchers, organisations and research projects to the NARCIS science portal.

[VISIT NARCIS](#)

Who contributes data to DANS?

- Archaeologists
 - Companies that conduct site surveys
 - University researchers
- Researchers in other fields
 - Public policy, economics, geography
 - Sociology, linguistics, oral history...
- Staff on behalf of data collectors



[http://
www.dayofarchaeology.com/
tag/the-netherlands/](http://www.dayofarchaeology.com/tag/the-netherlands/)

Why do people contribute data to DANS?

- Meet legal requirements
- Preserve data for long term
- Get credit for data
- Motivate citizen science participants
- Control access to data
- Use as background service



[http://
www.dayofarchaeology.com/tag/
the-netherlands/](http://www.dayofarchaeology.com/tag/the-netherlands/)

To Control Access to Their Data

- Permissions lock on individual files
- Prospective users of locked data register with DANS, ask permission of owner
- Permission conditions
 - Identity of user
 - Appropriate use
 - Appropriate proportion of dataset



<http://www.mapsofworld.com/netherlands/netherlands-river-map.html>

To Use DANS as a Background Service

- Data stored in DANS
- Data delivered from DANS via external websites
- Data served from other sites; DANS is backup
- DANS as “back office” to “front office” services of libraries
- DANS as harvesting source



<http://adst.org/oral-history/oral-history-interviews//>

Who consumes data from DANS?

- Archaeologists
 - Companies that conduct site surveys
 - University researchers
- Researchers in other fields
 - Public policy, economics, geography
 - Sociology, linguistics, oral history...
- Students, teachers, visiting scholars
- Local history guides, genealogists
- Museum curators, amateur scientists...



http://www.little-dixie.lib.mo.us/?page_id=42

How do users search in DANS?

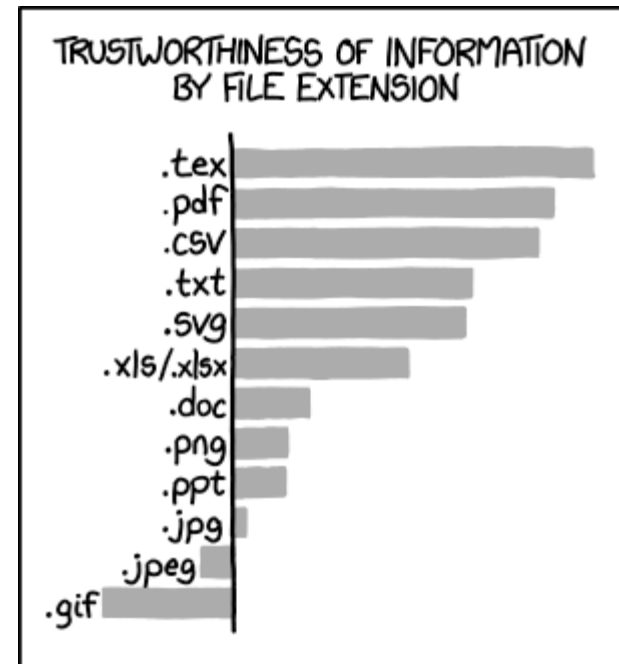
- Browse topics
- Search for place names
 - Archaeology sites
 - Building sites
 - Lat / long coordinates
- Browse spatial regions
- Rarely search for known items



[http://
www.intlconnect.illinois.edu/
jobsearch](http://www.intlconnect.illinois.edu/jobsearch)

How are DANS data used?

- Downloaded and saved
- Assess geographic regions
 - Archaeology sites
 - Building sites
- Read reports
- Use map data in other research
- Compare to other data
- Create new products – local history guidebooks



<https://xkcd.com/1301/>

Users knowledge of DANS

- Archaeology companies are heaviest users, only group that knows archivists by name
- Most other uses are infrequent, 1-2x/year or less, both contributing and consuming
- Assorted misconceptions
 - DANS is a “dark archive”
 - DANS is mostly quantitative data
 - Locked data are off limits to me
- Unfamiliar with some core features

What roles do archivists play?

- Acquire data
 - Work with contributors
 - Seek useful data
- Curate data
 - Ingest, clean, verify anonymity
 - Migrate / transfer to other formats
 - Describe, document, add metadata
 - Steward and sustain access for 10 years
- Disseminate data
 - Assist users in searching
 - Outreach to communities
- Staff help desks
- Bring expertise
 - Subject domain experts
 - Metadata and cataloging experts
 - Design and software engineering
 - Statistics and data analysis



<http://teleread.com/someone-is-right-on-the-internet-xkcd-turns-10-years-old/>

Discussion of findings so far-1

- Weblogs of limited value for user studies
- Uses and users far more diverse than expected
 - How to design for unknown users?
 - How to create and grow community for an archive?
- Archival activities favor acquisition of datasets
 - Continual acquisition necessary to grow the archive
 - How to promote and monitor uses of data?
 - How to balance preservation and access?

Discussion of findings so far-2

- Intermediaries play critical roles in data access
 - Archivists who acquire and curate data for use
 - Back offices who set policy for data use
 - Staff who contribute data on behalf of creators
- Public access is not equivalent to open access
 - Contributors may control access as condition of deposit
 - Potential consumers may or may not contact contributors for access

Infrastructure findings so far

- Data archives are essential components of knowledge infrastructures
 - Mediate access between contributors and consumers
 - Trusted institutions for long term stewardship
 - Invest in data through curation, migration, training, and assistance
- Data archives are visible components to some
 - Contributors to trusted archive
 - Consumers of trusted archive
- Data archives are invisible infrastructure to many
 - Data access via third party website, API, search engines, harvesters
 - Data access via alternative channels
 - Data contribution via home institutions
- Users come from unanticipated array of communities

Implications so far

- Research data archives are also used by non-researchers
 - Demonstrate value through diversity
 - Difficult to engage a “community” for support
- Archives serve data they do not own or control
 - Contributor control is an incentive for deposit
 - Controls place constraints on dissemination
 - Long term risk of orphaned data

Conclusions so far

- Data archives address goals broader and narrower than public policy suggests
- Data archives are essential for policy goals of data sharing and access
- More investment in description and search functions is needed to exploit archived data
- Data archives are often invisible infrastructure

Acknowledgements



Christine Borgman



Peter Darch



Ashley Sands



Irene Pasquetto



Bernie Randles



Milena Golshan

Data Archiving and Networked Services

DANS



UCLA Center for Knowledge Infrastructures