Title
Users and uses of a digital data archive: A case study of DANS

Authors
Borgman, Christine L.
Sands, Ashley E.
Golshan, Milena S.

Publication Date
2017-02-22

Users and uses of a digital data archive: A case study of DANS

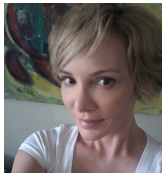
Christine L. Borgman, Ashley E. Sands, Milena S. Golshan


Christine Borgman


Peter Darch


Ashley Sands


Irene Pasquetto


Bernie Randles


Milena Golshan

Project collaborators:
DANS: Andrea Scharnhorst, Henk van den Berg
DANS visiting scholars
ANDS: Andrew Treloar
LANL: Herbert van de Sompel

UCLA Center for Knowledge Infrastructures

https://knowledgeinfrastructures.gseis.ucla.edu

Data Archiving and Networked Services

DANS

https://dans.knaw.nl/en

KNOWeSCAPE 2017, Sofia, Bulgaria
February 22, 2017

Overview of talk

- Motivation for study of DANS users and uses
- Research questions and methods
- Findings from interviews
- Discussion of results
- Implications for knowledge infrastructures

UCLA Center for Knowledge Infrastructures

Data Archiving and Networked Services
DANS

Knowledge Landscapes and Knowledge Infrastructures

- **Knowledge landscapes:** "large knowledge spaces and systems that organize and order them"

  KNOWeSCAPE. http://knowescape.org/about/

- **Knowledge infrastructures:** "robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds"

  Edwards, P. N. (2010). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: The MIT Press.

Sustaining Access to Scholarship

depends upon building better

Knowledge Infrastructures

http://knowledgeinfrastructures.org

Open access policies

- Research Councils of the UK
- European Union
- Australian Research Council
- U.S. Federal research policy
- Taiwan, China, India…
- Individual countries, funding agencies

Supported by wellcome trust

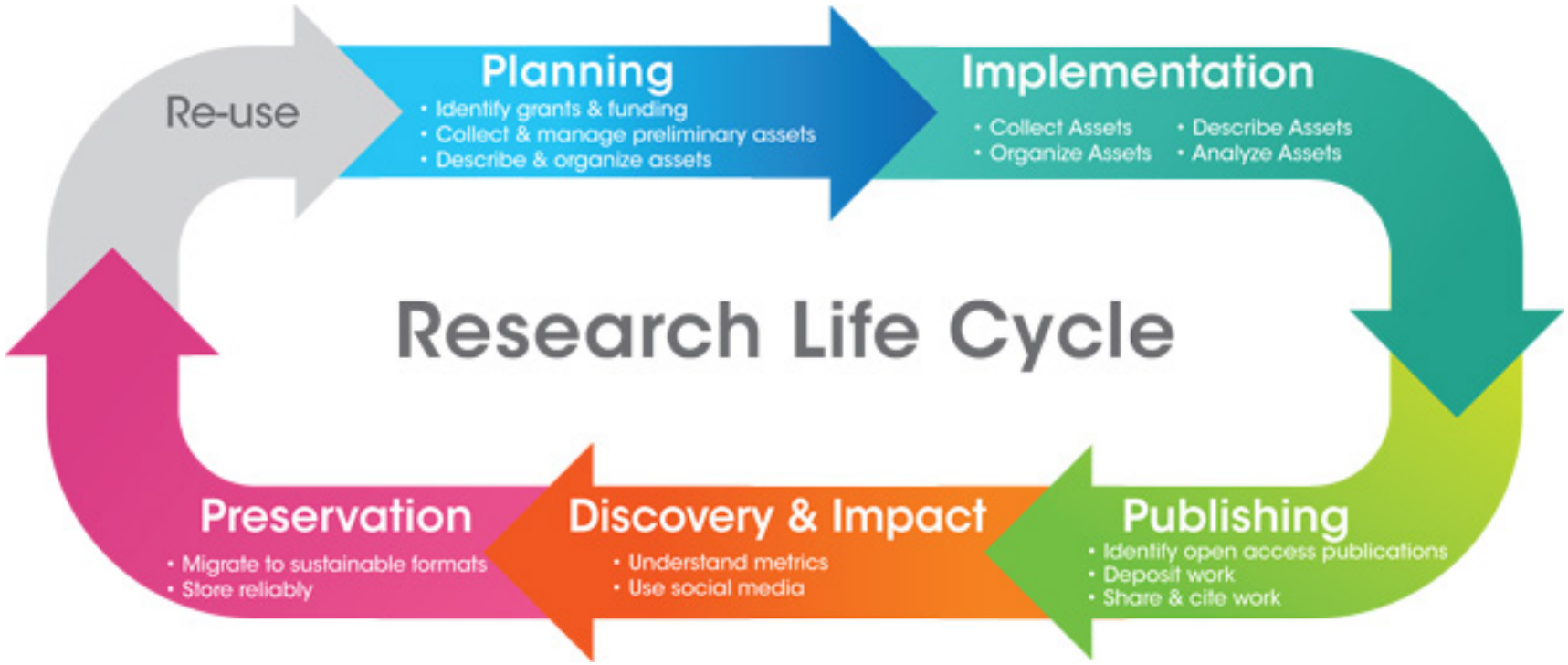
Australian Government
National Health and Medical Research Council

National Science Foundation
WHERE DISCOVERIES BEGIN

National Institutes of Health
Turning Discovery Into Health

Policy RECommendations for Open Access to Research Data in Europe

When to invest in data?



http://www.lib.uci.edu/dss/images/lifecycle.jpg

Research Questions

1. What are the roles of data archives in knowledge infrastructures?
2. How do stakeholder roles in digital archives vary?
   a. Contributors to a digital data archive
   b. Consumers of data from a digital archive
   c. Archivists who manage the digital data archive

UCLA Center for Knowledge Infrastructures

Data Archiving and Networked Services
DANS

DANS as a Digital Data Archive

- Netherlands government funding; about 50 staff
- Data Seal of Approval
- 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

KNAW NWO

Driven by data

Research Methods

- Document analysis
- Ethnography
- Mining transaction logs
- Interviews
  - DANS data contributors
  - DANS data consumers
  - DANS archivists and staff

A. Scharnhorst, C. Borgman, DANS 2014

DANS Interviews 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) 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

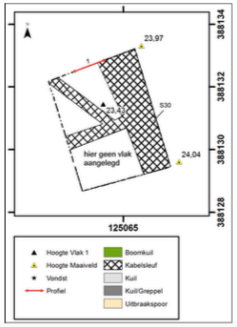
Findings

- Data types
- Data contributors
- Data consumers
- DANS archivists and staff
- Infrastructure roles of data archives

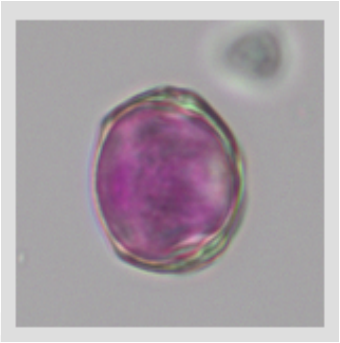
DANS data

Afbeelding 8 (links). Proefsleuf nr. 1, Profiel nr. 2. Foto (Fotonummer 3) genomen naar het zuiden.

Afbeelding 46. Proefsleuf nr. 7, Vlak 1. Duidelijk zichtbaar is de verstoring door kabels en leidir 13) genomen naar het noordoosten.



Afbeelding 47. Proefsleuf nr. 7, Vlak 1. Schaal 1: 100.



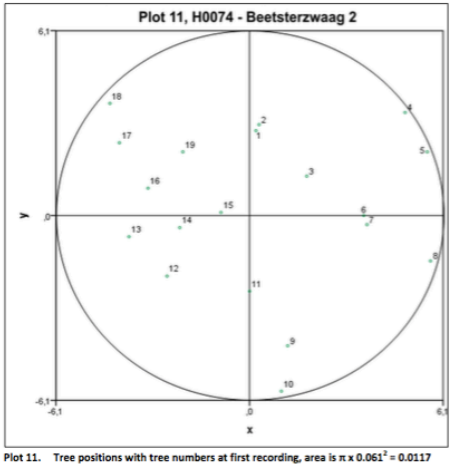
Wilgen, L. R. van (SOB Research) (2104).DANS. https://doi.org/10.17026/dans-z5y-tdb6

https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:62234/tab/2

Study 4. HOSP plots

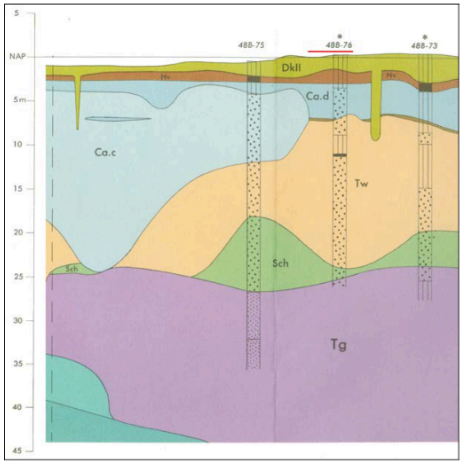
The HOSP plots of study 4 are all circular plots with a varying radius. The y axes is orientated to the North. Each plot is divided in one or more subplots of 0.01 ha, the total number depending on its rounded area. A plot with a radius of 9.1 m has as area 0.0260 ha. Rounded in units of 0.01 ha the area consists of 3 units. So the plot is divided in 3 subplots and the circle is divided in three equal parts of 120°, moving clockwise from the North. Those dividing lines for the subplots are not drawn in the maps.

Plot 11



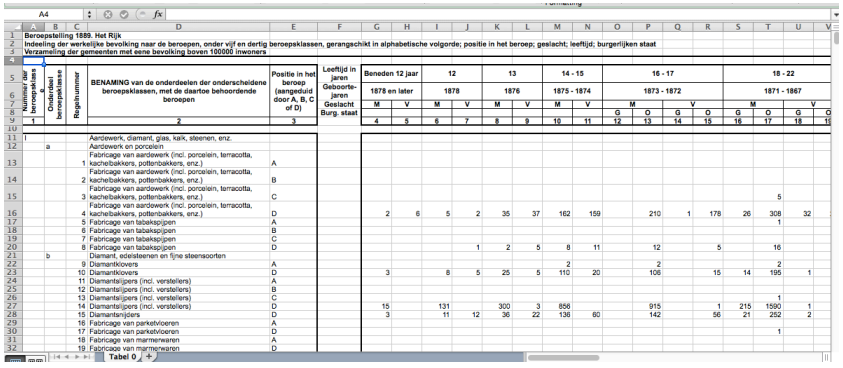
Plot 11. Tree positions with tree numbers at first recording, area is $\pi \times 0.061^2 = 0.0117$ ha. There is only one subplot.

https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:63812/tab/2



Afbeelding 6. De globale ligging van het plangebied (rood gemarkeerd), geprojecteerd (zie Afbeelding 5) van de Geologische Kaart van Nederland.

Bosch, J. E. van den (SOB Research) (2102). DANS. https://doi.org/10.17026/dans-2cg-tg88



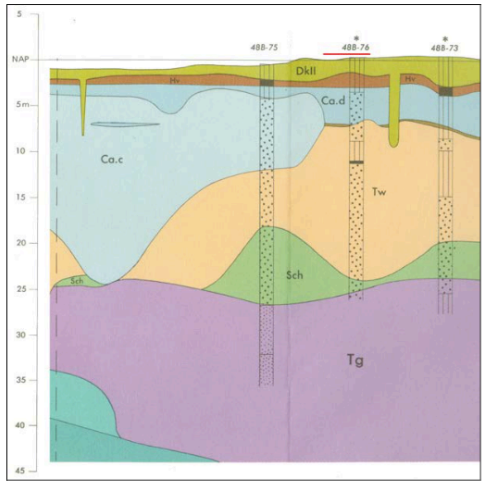
Beroepstelling 1889. Het Rijk
Indeeling der werkelijke bevolking naar de beroepen, onder vijf en dertig beroepsklassen, gerangschikt in alphabetische volgorde; positie in het beroep; geslacht; leeftijd; burgerlijken staat
Verzameling der gemeenten met eene bevolking boven 100000 inwoners

Nummer van beroepsklasse	Onderdeel beroepsklasse	Rubrieknummer	BENAMING van de onderdeelen der onderscheidene beroepsklassen, met de daartoe behoorende beroepen	Positie in het beroep (aangeduid door A, B, C of D)	Leeftijd in jaren / Geboortejaren / Geslacht / Burg. staat	Beneden 12 jaar		12		13		14 - 15		16 - 17				18 - 22			
						1878 en later		1878		1876		1875 - 1874		1873 - 1872				1871 - 1867			
						M	V	M	V	M	V	M	V	M	V	G	O	G	O	G	O	G	O
1			2	3		4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19		
I			Aardewerk, diamant, glas, kalk, steenen, enz.																				
	a		Aardewerk en porcelein																				
		1	Fabricage van aardewerk (incl. porcelein, terracotta, kachelbakkers, pottenbakkers, enz.)	A																			
		2	Fabricage van aardewerk (incl. porcelein, terracotta, kachelbakkers, pottenbakkers, enz.)	B																			
		3	Fabricage van aardewerk (incl. porcelein, terracotta, kachelbakkers, pottenbakkers, enz.)	C															5				
		4	Fabricage van aardewerk (incl. porcelein, terracotta, kachelbakkers, pottenbakkers, enz.)	D		2	6	5	2	35	37	162	159	210	1	178	26	308	32				
		5	Fabricage van tabakspijpen	A														1					
		6	Fabricage van tabakspijpen	B																			
		7	Fabricage van tabakspijpen	C																			
		8	Fabricage van tabakspijpen	D					1	2	5	8	11	12		5		16					
	b		Diamant, edelstenen en fijne steensoorten																				
		9	Diamantkijkers	A								2		2				2					
		10	Diamantkijkers	D		3		8	5	25	5	110	20	108		15	14	195	1				
		11	Diamantslijpers (incl. verstellers)	A																			
		12	Diamantslijpers (incl. verstellers)	B																			
		13	Diamantslijpers (incl. verstellers)	C														1					
		14	Diamantslijpers (incl. verstellers)	D		15		131		300	3	856		915		1	215	1590	1				
		15	Diamantwrijvers	D		3		11	12	36	22	136	60	142		56	21	252	2				
		16	Fabricage van parketvloeren	A																			
		17	Fabricage van parketvloeren	D														1					
		18	Fabricage van marmerwaren	A																			
		19	Fabricage van marmerwaren	D																			

https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:65225/tab/2

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
  - Institutional librarian
  - Head of the department



Afbeelding 6. De globale ligging van het plangebied (rood gemarkeerd), geprojecteerd (zie Afbeelding 5) van de Geologische Kaart van Nederland.

Bosch, J. E. van den (SOB Research) (2102). DANS. https://doi.org/10.17026/dans-2cg-tg88

Why do they contribute data to DANS?

- Meet legal requirements
- Get credit for data
- Share data with others
- Preserve data for long term
- Control access to their data

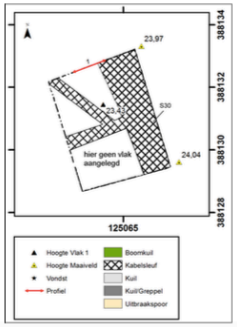
Afbeelding 8 (links). Proefsleuf nr. 1, Profiel nr. 2. Foto (Fotonummer 3) genomen naar het zuide

Wilgen, L. R. van (SOB Research) (2104).DANS. https://doi.org/10.17026/dans-z5y-tdb6

Controlling access to deposited data

- Contributor goals
  - To identify consumers of their data
  - To control how their data are used
- Consumer requirements
  - Register with DANS
  - Request permission from data contributor

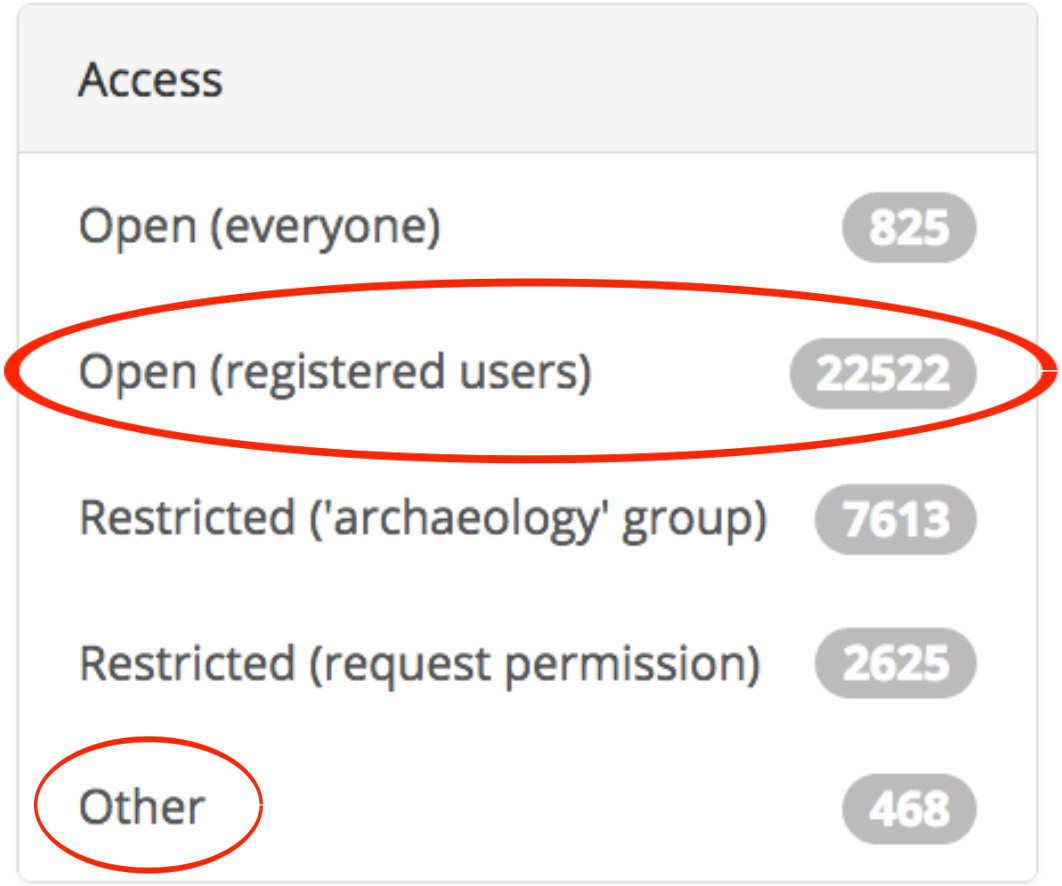
Afbeelding 46. Proefsleuf nr. 7, Vlak 1. Duidelijk zichtbaar is de verstoring door kabels en leidir
13) gemomen naar het noordoosten.



Afbeelding 47. Proefsleuf nr. 7, Vlak 1. Schaal 1: 100.

Wilgen, L. R. van (SOB Research) (2104).DANS. https://doi.org/10.17026/dans-z5y-tdb6

Types of Access



https://easy.dans.knaw.nl/ui/browse

Granularity of access

Overview | Description | Data files (6)

You need to log in to be able to view/access (some of) the files. **Log In**

You need to have special permission to be able to access (some of) the files. You can request permission after logging in.

Download | View details

Dataset Contents / original

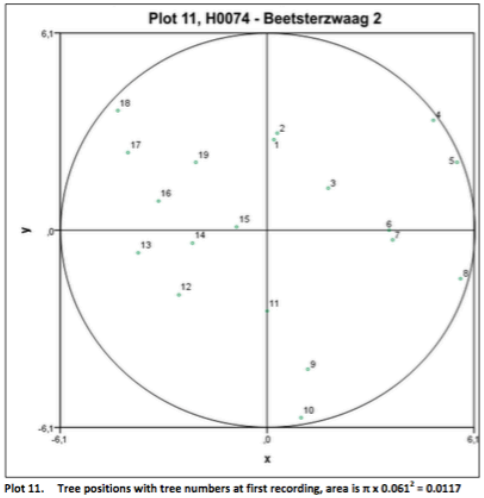
Dataset Contents
original

	Name	Size	Accessible
	FEM growth and yield data Monocultures - Silver Birch - plot data book.pdf	353295	Yes
	FEM growth and yield data Monocultures - Silver Birch - plot information.csv	6282	Requires granted permission request
	FEM growth and yield data Monocultures - Silver Birch - plot level.csv	21235	Requires granted permission request
	FEM growth and yield data Monocultures - Silver Birch - tree level.csv	190875	Requires granted permission request
	FEM growth and yield data Monocultures - Silver Birch - tree maps atlas.pdf	1017162	Yes
	Read me - FEM growth and yield data Monocultures - Silver Birch.pdf	205772	Yes

Study 4. HOSP plots

The HOSP plots of study 4 are all circular plots with a varying radius. The y axes is orientated to the North. Each plot is divided in one or more subplots of 0.01 ha, the total number depending on its rounded area. A plot with a radius of 9.1 m has as area 0.0260 ha. Rounded in units of 0.01 ha the area consists of 3 units. So the plot is divided in 3 subplots and the circle is divided in three equal parts of 120°, moving clockwise from the North. Those dividing lines for the subplots are not drawn in the maps.

Plot 11

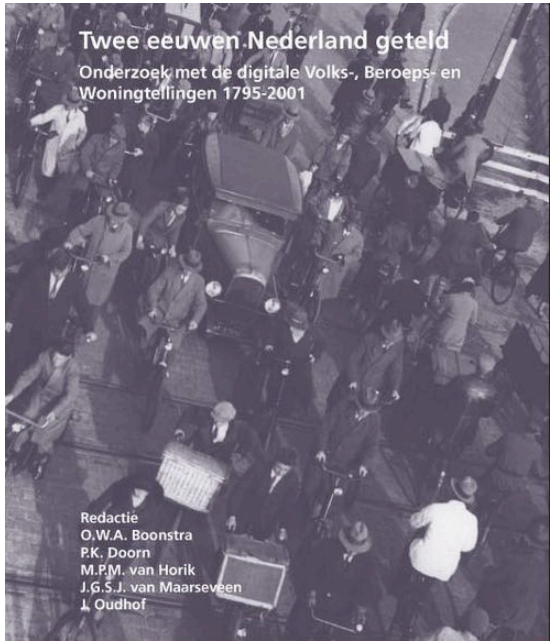


Plot 11. Tree positions with tree numbers at first recording, area is π x 0.061² = 0.0117 ha. There is only one subplot.

https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:63812/tab/2

Who consumes data from DANS?

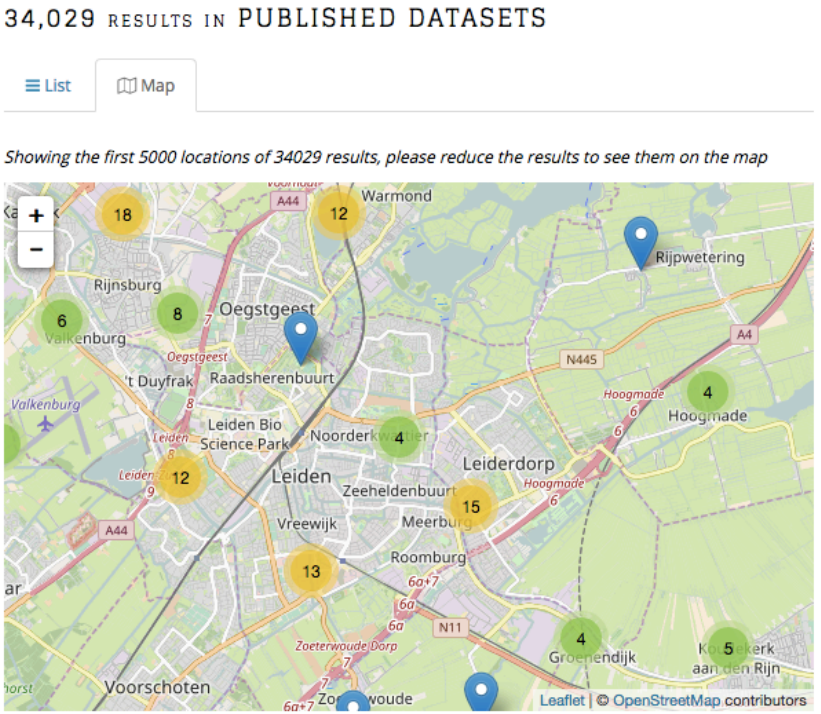
- Archaeologists
  - Companies that conduct site surveys
  - University researchers
- Researchers in other fields
  - Public policy, economics, geography
  - Sociology, linguistics, oral history…
- Students, teachers, visiting scholars
- Local history guides, genealogists
- Museum curators, amateur scientists...



http://www.volkstellingen.nl/nl/vt_bundel/index.jpg

How do users search in DANS?

- Browse topics
- Browse spatial regions
- Search for place names
  - Archaeology sites
  - Building sites
- Few known item searches
- Open access datasets are most used



https://easy.dans.knaw.nl/ui/?wicket:interface=:1:1:::

How are DANS data used?

- To assess geographic regions
- To use map data in other research
- To compare to other data
- To create new products, e.g., local history guidebooks

http://www.intlconnect.illinois.edu/jobsearch

DANS as Invisible Infrastructure

- DANS data delivered via external websites
- DANS as “back office-front office” services
- DANS as metadata harvesting source

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

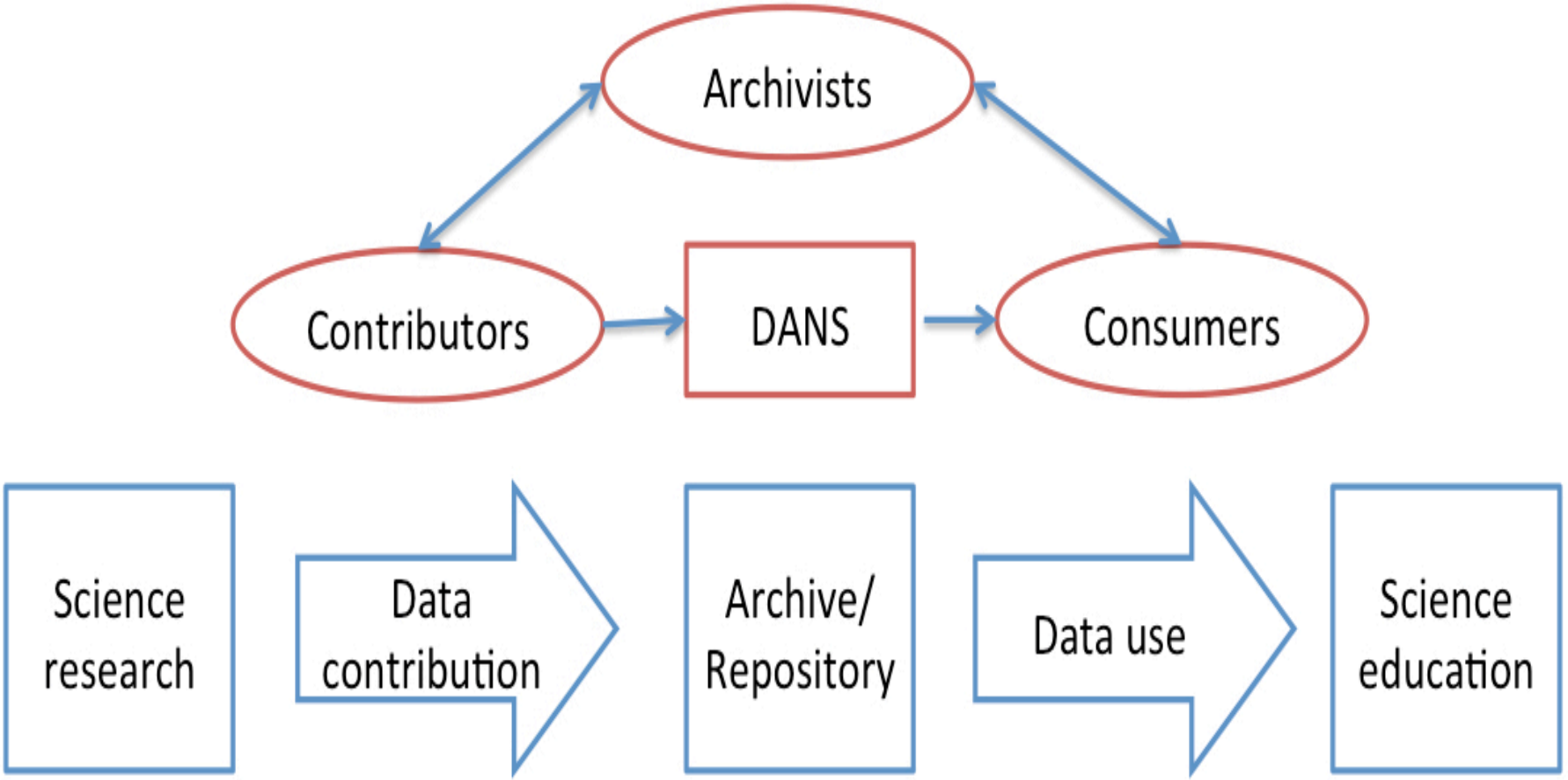
Visible roles of archivists

- Acquire data
  - Work with contributors
  - Seek useful data
- Disseminate data
  - Assist users in searching
  - Outreach to communities
- Staff help desk

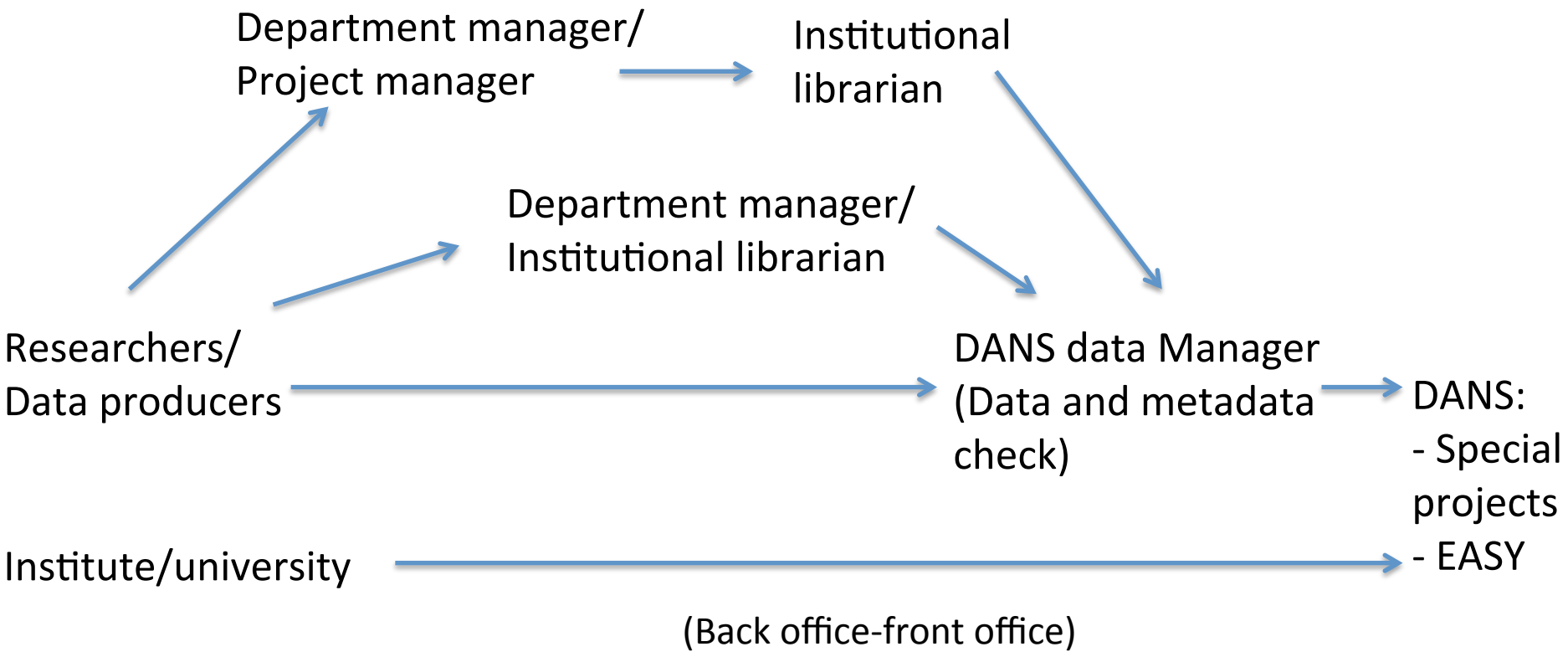
Invisible roles of archivists

- Curate data
  - Ingest, clean, verify anonymity
  - Migrate data to other formats
  - Describe, document, add metadata
  - Preserve the data in long term
- Bring expertise
  - Subject domain experts
  - Metadata and cataloging experts
  - Design and software engineering
  - Statistics and data analysis

Mediation and Stakeholders



Mediation and Stakeholders



Discussion of findings

- Data archive content is community driven
- Archives are a source of trust in data
- Archive uses and users are diverse
- Access to data is mediated by
  - Archivists
  - Contributors who retain control over data
  - Institutional policies for contribution
  - Services that deliver data

Implications for Infrastructures

- Invisible work shapes knowledge spaces
- Interdisciplinary data archives
  - Acquire a disparate array of datasets
  - Attract a diverse array of contributors and consumers
  - Are difficult to search due to their diversity
- Incentives to contribute data vary by stakeholder
  - Legal responsibility for archaeological reports
  - Assure preservation and access
  - Transfer responsibility while maintaining control
- Interoperability of data archives with publishers, libraries, project sites, could increase access

Acknowledgements

Christine Borgman

Peter Darch

Ashley Sands

Irene Pasquetto

Bernie Randles

Milena Golshan

Data Archiving and Networked Services

DANS

UCLA Center for Knowledge Infrastructures

ALFRED P. SLOAN FOUNDATION 1934