

Data management aspects in the social sciences

Marjan Grootveld, DANS (Twitter @MarjanGrootveld)

Presenting also slides by Marion Wittenberg and Peter Doorn, DANS

Workshop on Active DMPs – Geneva, 28-30 June 2016

On the agenda

- DANS services
- Social science traits
- Example datasets
- Data management training
- My personal concerns



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



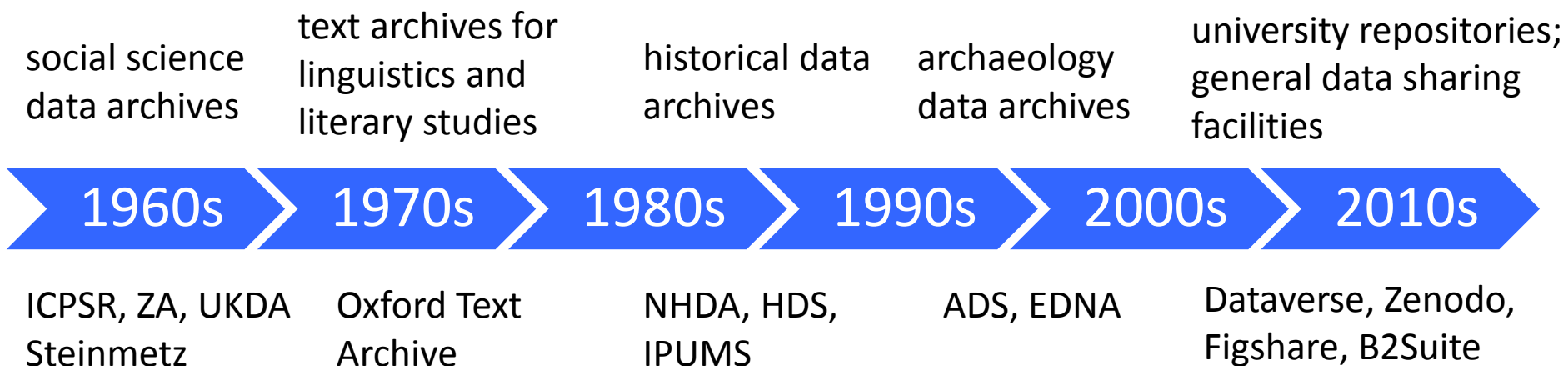
Data Archiving in Humanities and Social Sciences

Data collection and data processing

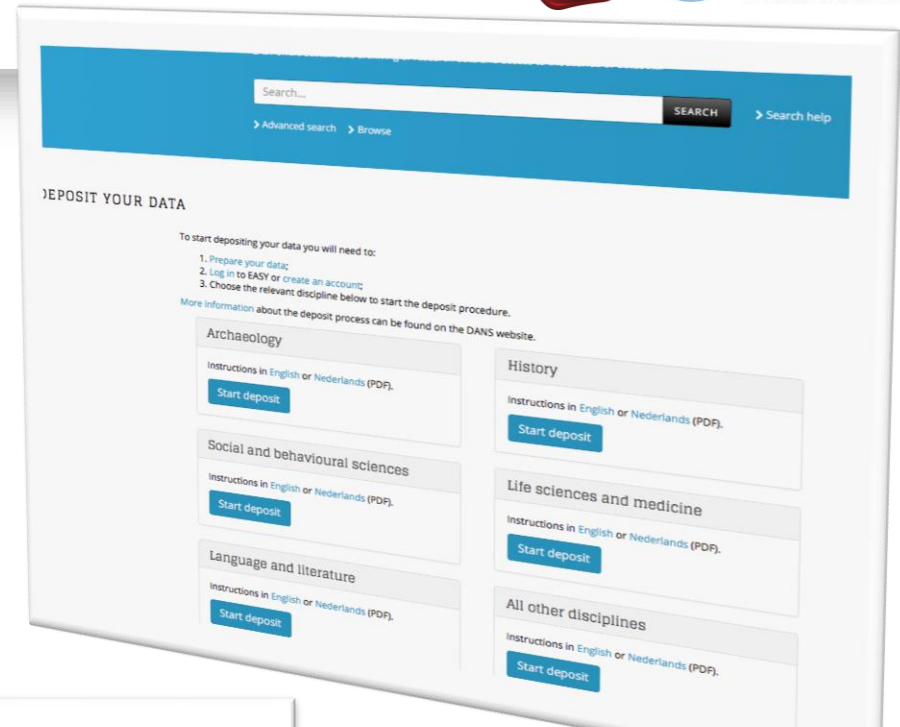
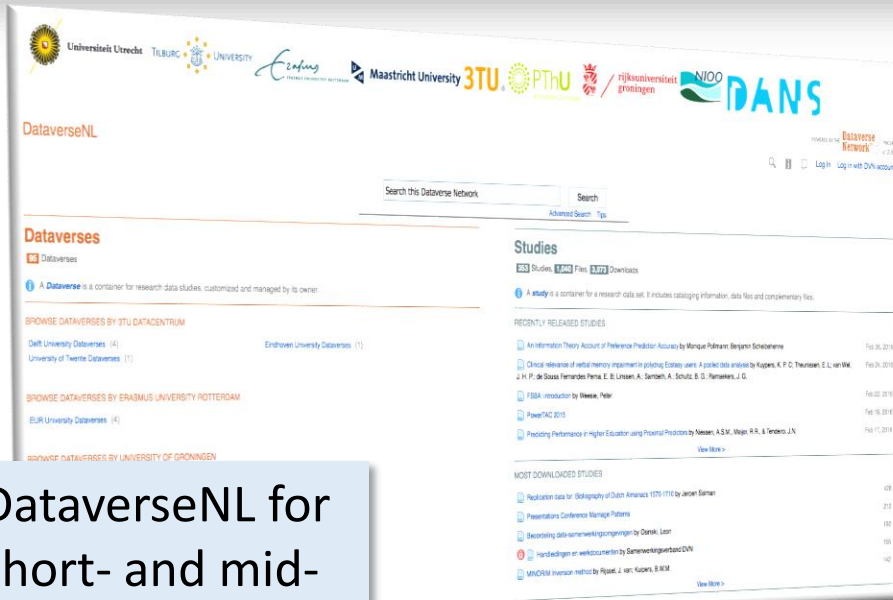
→ awareness of the value of preserving data for re-use:

- for validating the results of earlier research
- for comparative analysis
- for secondary analysis: answering new research questions with existing data

Emergence of data archives:



Core online services



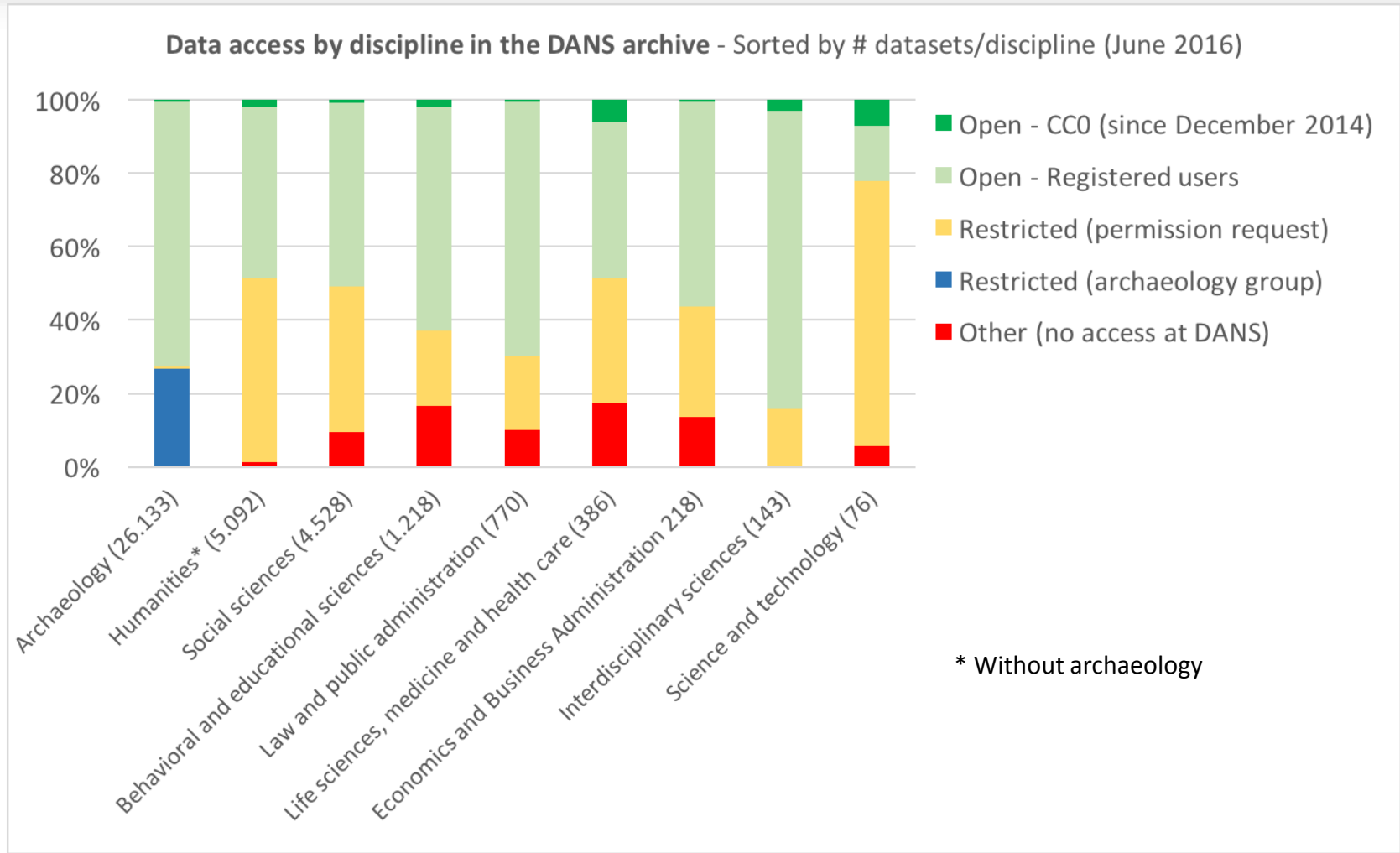
DataverseNL for short- and mid-term storage

EASY: certified long-term Electronic Archiving System for self-deposit

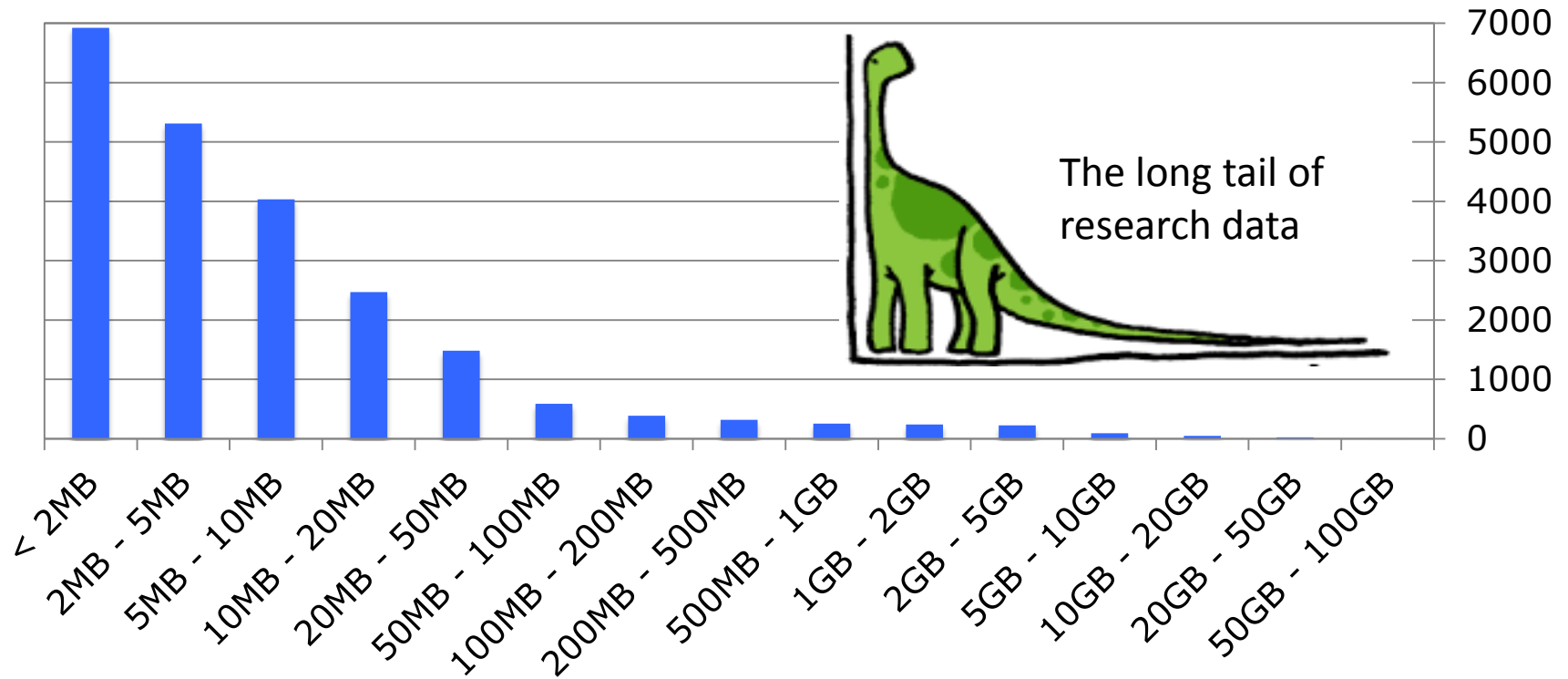


NARCIS: Gateway to scholarly information in the Netherlands

Data access by discipline in DANS archive



Datasets in DANS archive according to size







RDM support: DANS DMP brochure

Describing the data
Are existing data reused or are new data generated?
What type(s) of data is concerned; file size; growth rate?
Standards and metadata, or everything required to find and use the data
Which metadata standards are used (findability)?
Which coding etc. is used for enabling future linking to other data (interpretation, interoperability)?
Which software and hardware are used (interpretation, usability)?
What is documented and stored to enable replication?
Ethical and legal
In collecting or generating data, how is required permission obtained from the data supplier/ test subjects / ...? Which restrictions, if any, apply during research?
How are sensitive data protected during and after the project?
What are the agreements in the case of stakeholders leaving (early)?
Are data available Open Access after the project, perhaps after an embargo period? If not, which conditions apply?
Storage and archiving
How is sufficient storage and back-up capacity organised during the project, including version management? Are the expenses for this covered, and if not, ...?
Where, and how long after completing the project, are data available for follow-up research and verification? Is this a Trusted Digital Repository with an international certification? If not, how will the data be findable, sustainably accessible and usable? Consider metadata and persistent identifiers, such as DOI, Handle and URN.
Are the expenses for (preparing the data for) archiving covered?

Data Archiving and Networked Services
DANS

Research Data Management in 4 Steps

An effective research data management plan provides a timely insight into facilities and expertise necessary both during and after your research, which is useful for yourself, your senior staff, data supporters and research funders. An effective plan furthermore enables the reuse of research data, open if possible, protected if necessary.

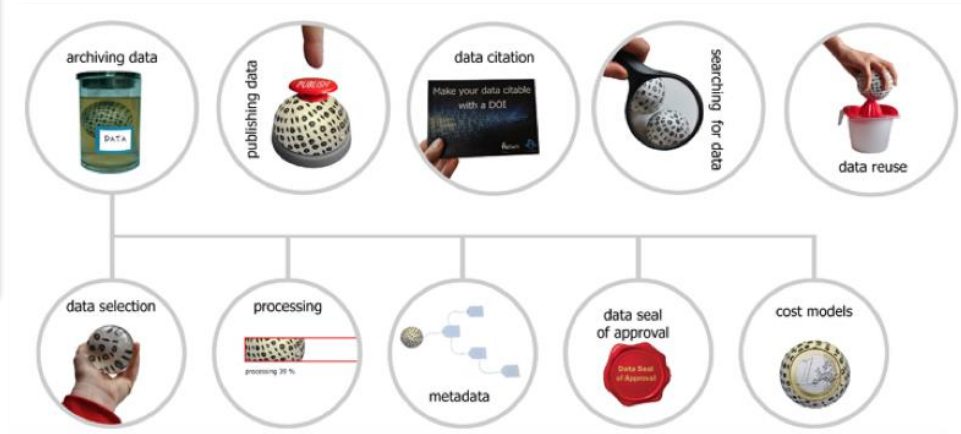
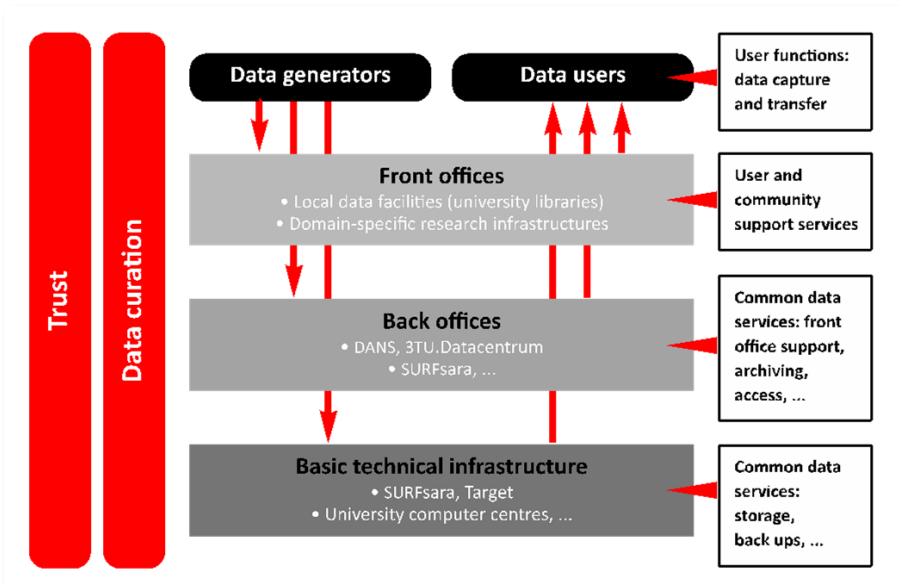
- 1. Establish permissions regarding the data**

Do you have permission from stakeholders (test subjects, data supplier or ethical commission), to share data with colleagues? Are the data available for follow-up research?
- 2. Document your research and data**

Store both raw and processed data. Data that do not lead to publication may also be significant. Via which metadata and coding can the data be found, understood and linked to other data sources? Which software and equipment are required for replicating the research? Which file formats are concerned?
- 3. Organise storage and access**

How big are the data files? Do you have sufficient storage capacity for intermediary versions? How do you protect sensitive data? Archiving the data in a Trustworthy Digital Repository upon completing your research ensures sustainable access, including a persistent identifier for citations.
- 4. Keep your research data management plan up-to-date**

Does your research data management plan contain up-to-date and relevant information for data supporters in the library/ IT department/ legal desk, for your research director, funder and project partners? Make sure that changes to your project are reflected in your plan.

http://www.dans.knaw.nl/en/about/organisation-and-policy/information-material?set_language=en

Research Data Netherlands



Collaboration of DANS, 4TU.ResearchData and SURFsara to promote sustained access to and responsible re-use of digital research data



Essentials 4 Data Support <http://datasupport.researchdata.nl/en>

Large players in Social Science data

CESSDA

Consortium of European Social Science Data Archives

NATIONAL DATA
SERVICES

Find your National Data Service



<http://cessda.net/>

ICPSR

A partner in social science research

Find & Analyze Data

quick search for data

GO

Membership in ICPSR

Deposit Data

ICPSR Summer Program

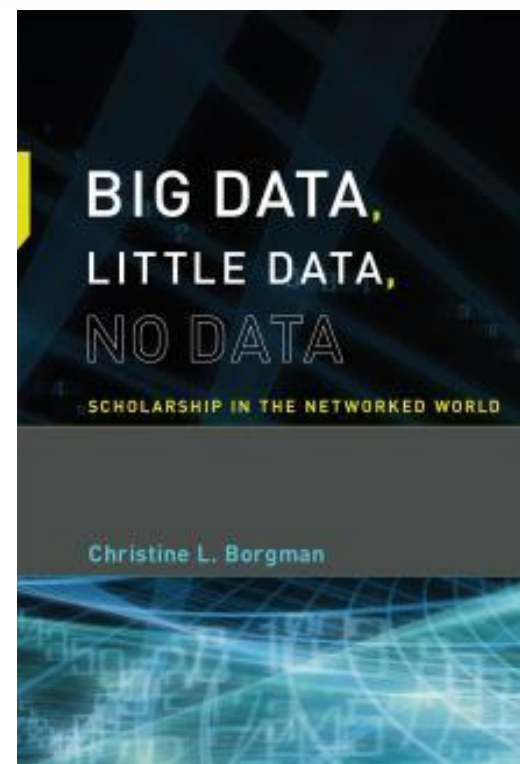
Teaching & Learning with ICPSR

Data Management & Curation

<http://www.icpsr.umich.edu/>

Borgman: Data Scholarship in the Social Sciences

- 'The social studies encompass research on human behavior in the past, present, and future' (p.125)
- 'The social sciences articulate their research methods more explicitly than do most fields' (p. 126)
- '...characterized more by shared knowledge than by shared technical infrastructures' (p.157)
- 'diffuse data sources, fuzzy boundaries between fields, political sensitivity of topics, and the array of stakeholders' (p.160)



Christine L. Borgman: *Big data, little data, no data – Scholarship in a networked world.*
MIT Press, 2015.

Social science traits (over-generalised!)

- Quantitative research, e.g. surveys (lots of variables > codebook needed) and qualitative research, e.g. interviews and observations
- May involve individual people > ethical issues, informed consent forms, sensitive or anonymised data
- Often longitudinal research (e.g. the start of the International Social Survey Programm (ISSP) was in 1972)
- Mixed attitude towards sharing and reusing data, e.g.
 - Political scientists are used to sharing data
 - Economists often explore private third-party data (cannot be released or archived afterwards)
 - Sociotechnical researchers cannot release or reproduce all materials (lab journals remain property of the lab) (Borgman, p. 149)
 - For psychologists research methodology may have more value than the data
- Recent NL tendency (Oldenburg): publication packages along with publication: data + statistical syntax queries

Beau Oldenburg: *Integriteit en duurzaamheid in het digitale tijdperk.*

White paper DANS, 2015. <http://www.dans.knaw.nl/> (in Dutch)



Example dataset 1

Persistent identifier	DOI: 10.17026/dans-z6t-wmuv URN: urn:nbn:nl:ui:13-ijuk-2s
Creator	Corten, Dr. R. (Utrecht University) DAI: info:eu-repo/dai/nl/304835145
Title	Two experiments on cooperation and reputation in networks
Description	<p>The purpose of this study was to test the hypothesis by Raub and Weesie (1990) that embeddedness in social networks fosters cooperation in repeated dyadic Prisoner's Dilemmas via reputation effects. Data were collected in two computerized laboratory experiments. In the experiments, subjects played 40 rounds of a game in which they played dyadic Prisoner's Dilemmas with the other members of their group. The experimental manipulation involved whether subjects could observe the actions of all other group members or not.</p> <p>This dataset is provided for replication purposes for the following publication: Corten, R. Buskens, V., Rosenkranz, S. and Cook, K.S. (2016). "Reputation effects in social networks do not promote cooperation: an experimental test of the Raub & Weesie model". Forthcoming in PLOS ONE. Further details of the experimental design are provided there. An example of the instructions used is provided in the file repexp - Instructions - emb_V3.pdf.</p>
Date created (ISO 8601)	2016-06-15
Access rights	Open access for registered users - Unrestricted access for all registered EASY users

<input type="checkbox"/>	Name ▲	Size ▾ 5 MB
<input type="checkbox"/>	 repexp_cortenetal_2016_V02.por	4916062
<input type="checkbox"/>	 repexp_cortenetal_2016_V02.sav	5022075

DDI - Data Documentation Initiative

<http://www.ddialliance.org/>

International standard for describing data from the social, behavioral, and economic sciences

Documenting data with DDI facilitates interpretation and understanding - both by humans and computers

Codebook and Lifecycle



See also <http://rd-alliance.github.io/metadata-directory/standards/>

DDI-Codebook

DDI-Codebook is a light-weight version of the standard, intended primarily to document simple survey data

To make DDI codebooks you can make use of [the NESSTAR publisher](#)

Example DANS [NESSTAR server](#)



Example 2: inspect survey outcomes online

nesstar.dans.knaw.nl/webview/ Zoeken

DANS EASY online analysis tool

DESCRIPTION TABULATION ANALYSIS

Dataset: Dutch parliamentary election study, 1994 - DPES'94

Variable var014: Q7 Frequency watching NOS tv newscast

LITERAL QUESTION
Kunt u met behulp van dit kaartje aangeven hoe vaak in de week u het Nos-journaal volgt?

1. (vrijwel) dagelijks
2. 3 a 4 keer per week
3. 1 a 2 keer per week
4. minder dan 1 keer per week
5. n.v.t. (geen t.v.)

DESCRIPTIVE TEXT
 English translation:
Could you indicate on this card how often you generally watch the NOS television news?

Values	Categories	N	
1	almost daily	1279	70.7%
2	3-4 times per week	268	14.8%
3	1-2 times per week	157	8.7%
4	less than once a wk	92	5.1%
5	does not own tv set	14	0.8%
7	DK	2	

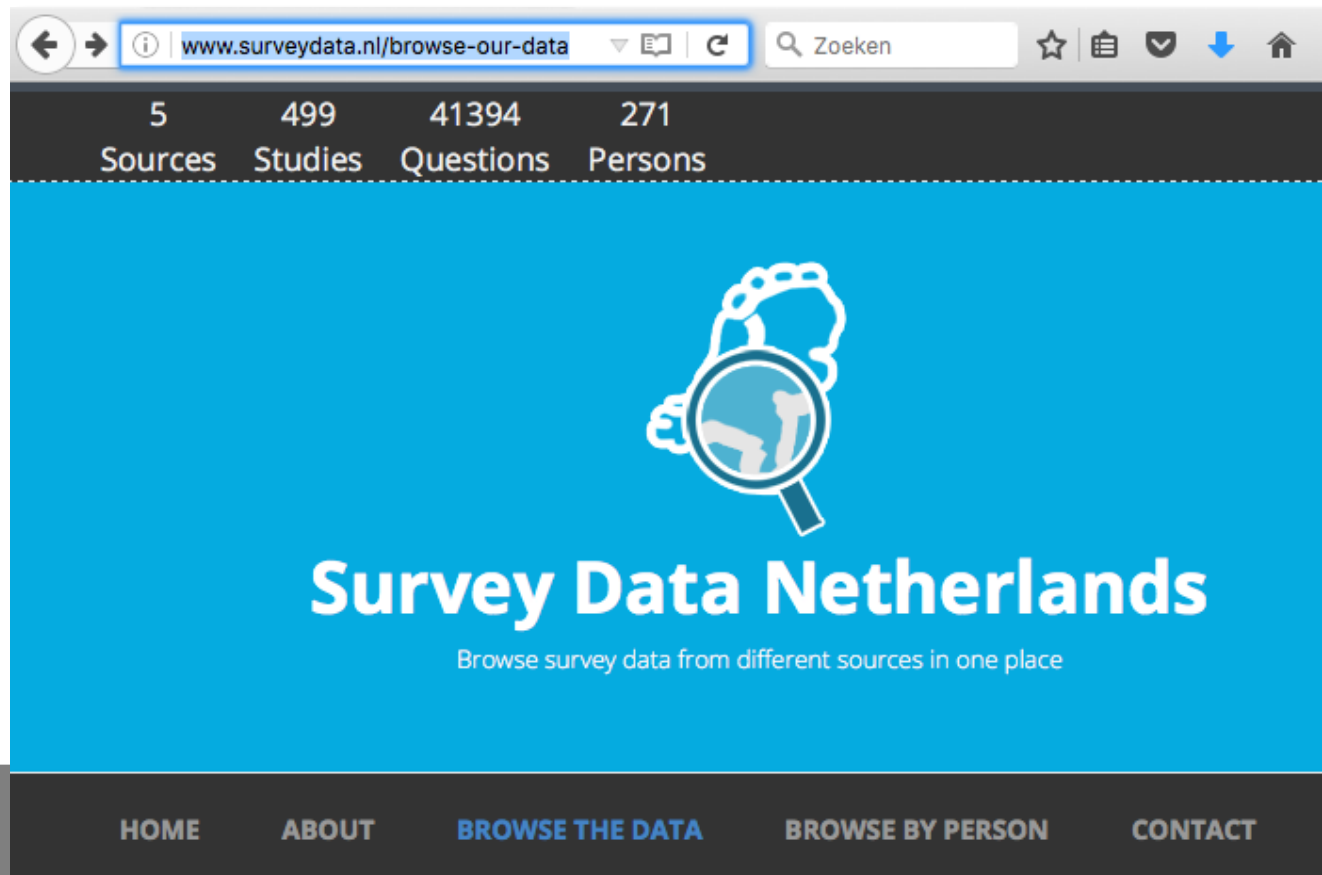
SUMMARY STATISTICS
Valid cases 1810
Missing cases 2
This variable is numeric

INTERVIEWER INSTRUCTIONS
Showcard 2 presented, listing response alternatives.

- Dutch parliamentary election study, 1994 - DPES'94
 - Metadata
 - Variable Description
 - First pre-election wave
 - Political interest and communication
 - Q2 Reads about national news?
 - Q3 Talks about national problems?
 - Q4 Reads about foreign news?
 - Q5 Interested in politics?
 - Q2-5 Political interest score
 - Q6 Daily newspaper - first answer
 - Q6 Daily newspaper - second answer
 - Q6 Daily newspaper - third answer
 - Q6 Daily newspaper-fourth to eighth answer
 - Q7 Frequency watching NOS tv newscast
 - Q8 Frequency watching RTL4 tv newscast
 - Most important national problem
 - Party adherence
 - Party membership
 - Government policy satisfaction
 - Vote intention Second Chamber 1994
 - Previous voting behavior Second Chamber 1989
 - Political position issues - euthanasia
 - Political position issues - crime
 - Political position issues - differences in income
 - Political position issues - nuclear plants
 - Political position issues - ethnic minorities
 - Voting behavior in 1994 municipal elections
 - Political Idealism
 - Sympathy scores for political parties
 - Sympathy scores for politicians
 - Coalition preference
 - Political environment during adolescence
 - Family values
 - Left and right in politics
 - Political position issues - asylum seekers
 - Meaning of life
 - Religion of respondent
 - Religion of partner

DDI-Lifecycle

DDI-Lifecycle is designed to document and manage data across the entire life cycle, from conceptualisation to data publication, analysis and beyond. E.g. [Survey Data Netherlands](http://www.surveydata.nl)



The screenshot shows a web browser window displaying the Survey Data Netherlands website. The browser's address bar shows the URL www.surveydata.nl/browse-our-data. The website header features a dark grey bar with the following statistics: 5 Sources, 499 Studies, 41394 Questions, and 271 Persons. Below this is a large blue banner with a white icon of a hand holding a magnifying glass over a globe. The text on the banner reads "Survey Data Netherlands" in large white font, followed by the tagline "Browse survey data from different sources in one place" in a smaller white font. At the bottom of the page is a dark grey navigation bar with the following links: HOME, ABOUT, BROWSE THE DATA, BROWSE BY PERSON, and CONTACT.

Category	Count
Sources	5
Studies	499
Questions	41394
Persons	271

Question 1989W1Q006

<< prev | next >>

Selection of respondents: all respondents

 ENGLISH

Question Text:


Could you indicate on this card how often you generally watch the NOS television news?

Answer Type: Categories

- 1 almost daily
- 2 3 - 4 times per week
- 3 1 - 2 times per week
- 4 less than once a week
- 5 does not own tv set

Instruction:

Showcard 2 presented

 Dutch

Question Text:

Kunt u met behulp van dit kaartje aangeven hoe vaak u gewoonlijk het NOS-journaal (het nieuws) volgt op de t.v.?

Answer Type: Categories

- 1 (vrijwel) dagelijks
- 2 3 a 4 keer per week
- 3 1 a 2 keer per week
- 4 minder dan 1 keer per week
- 5 n.v.t. (geen t.v.)

Instruction:

Overhandig Kaart 2

Related Variables

Variable	Dataset	Study Unit
var009	DPES1989	Dutch Parliamentary Election Studies > Dutch Parliamentary Election Study 1989

Related Questions

Question	Questionnaire	Study Unit
1994W1Q007	Dutch Parliamentary Election Studies 1994	Dutch Parliamentary Election Studies > Dutch Parliamentary Election Study 1994
1998W1Q007	Dutch Parliamentary Election Studies 1998	Dutch Parliamentary Election Studies > Dutch Parliamentary Election Study 1998
2002W1Q0071	Dutch Parliamentary Election Studies 2002-2003	Dutch Parliamentary Election Studies > Dutch Parliamentary Election Study 2002-2003
2003W1Q0071	Dutch Parliamentary Election Studies 2002-2003	Dutch Parliamentary Election Studies > Dutch Parliamentary Election Study 2002-2003
2006W1Q007	Dutch Parliamentary Election Studies 2006	Dutch Parliamentary Election Studies > Dutch Parliamentary Election Study 2006
2010W1Q012	Dutch Parliamentary Election Studies 2010	Dutch Parliamentary Election Studies > Dutch Parliamentary Election Study 2010
2012SQ03	Dutch Parliamentary Election Studies 2012	Dutch Parliamentary Election Studies > Dutch Parliamentary Election Study 2012

Ex. 4: Interview project inspired DMP training

vi **veteraneninstituut** HOME VETERANEN & HUN MISSIES ZORG **DIENSTEN** ONTMOETEN NIEUWS OVER VI

DRAAGINSIGNES VETERANENPAS CHECKPOINT 'VETERAAN IN DE KLAS' **INTERVIEWCOLLECTIE** DIRECT REGELEN

Interviewcollectie

Ruim 600 interviews online te beluisteren in de Interviewcollectie Nederlandse Veteranen

Home » Diensten » Interviewcollectie

Ruim 600 interviews online te beluisteren

Hoe was het voor Kees om jaren van zijn familie verwijderd te zijn toen hij op Sumatra zat? Hoe ging Annemarie om met de lokale bevolking in Rwanda? Hoe beleefde Jan het om na het verlaten van de militaire dienst te integreren in de burgermaatschappij? Duik in de interviewcollectie en beluister de persoonlijke verhalen van ruim 600 veteranen. Van soldaat tot generaal, van hospik tot para, van gevechten op Nederlandse bodem in de Meidagen van 1940 tot de slag bij Chora in Afghanistan, alles komt aan bod in deze gevarieerde collectie.

Wat is de Interviewcollectie Nederlandse Veteranen?

600 interviews in DANS archive

Use case in *Essentials 4 Data Support* training



The What, Why and How of Data Management Planning

<http://datasupport.researchdata.nl>

DMP and data organisation assignments



Prepare a data management plan

Read the case well and write a data management plan.

Case

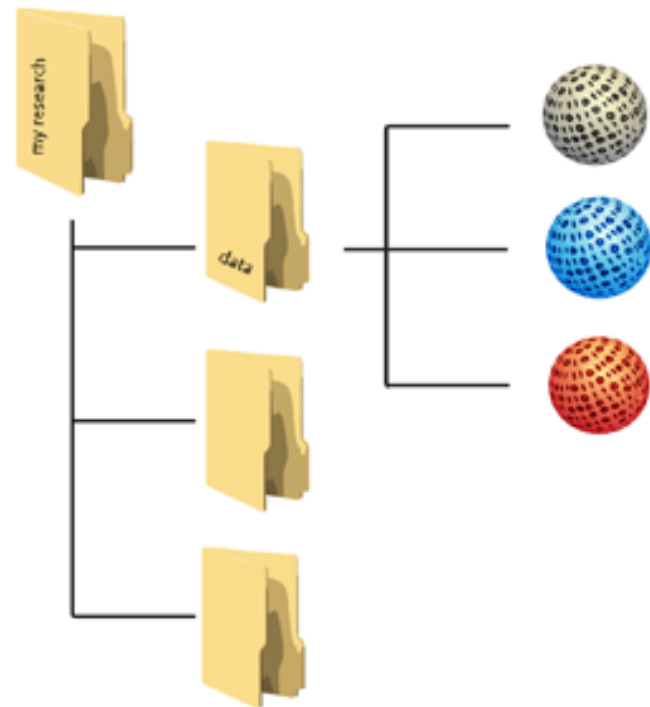
The project 'Veteran tapes' is allowed to start. The funding has been awarded and the funder expects a data management plan to be prepared.

Project proposal summary:

Since the end of the Second World War, the Netherlands has participated in several missions abroad. (...)

Design and definition of the project:

- Interviewing 600 people; video-taped.
- 20 people who are doing the interviewing; alignment is important.
- Structured interviews to enable comparison.
- The working language is Dutch.
- The people being interviewed will have a chance to see the video and can decide whether to show the interview freely, or to use it for research purposes only.
- The project will run for two years.
- At the end there will be a portal where the videos can be viewed when allowed to be accessed publicly. It is the intention to make all interviews available for research purposes.
- The audio stream (speech) will be transcribed. In doing so, it will be possible to search for interviews and to refer to illustrative passages.

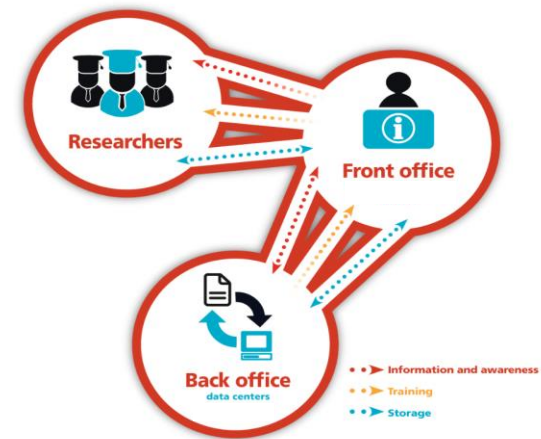


Design a data organisation for the Veterans project (folder structure, file naming convention, ...)

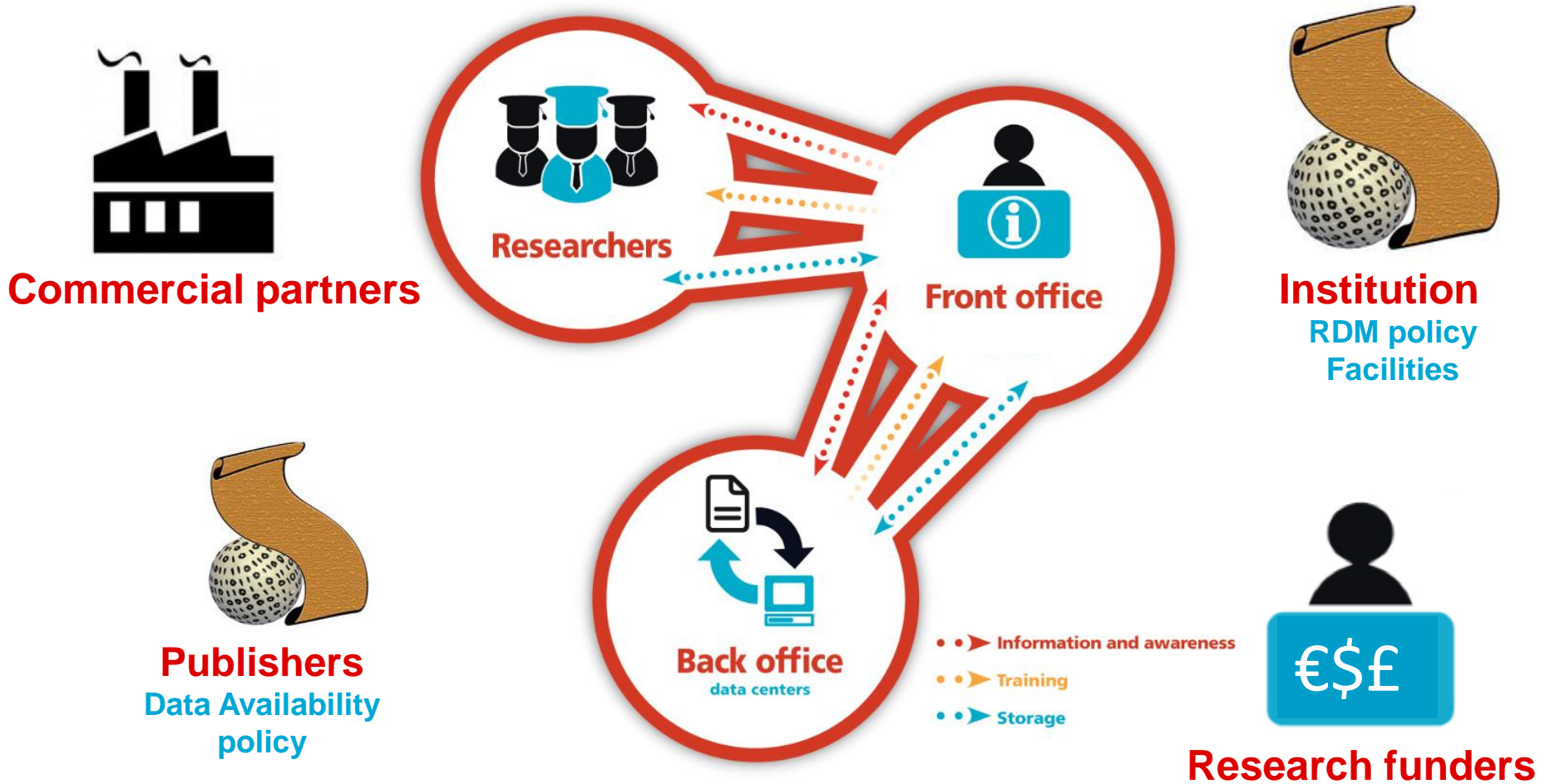
<http://datasupport.researchdata.nl/en/>

Outcome of the assignments

- Writing the DMP is always a real confidence booster.
- Discussing the data organisation for 10 minutes gives already a lot of insight.
- A dataset contains more than the data...
- Common assumption that ALL files are either Open or Restricted. (Relevant for H2020 practice to address different subsets in the DMP.)
- Realisation that planning RDM is teamwork.



Stakeholders in RDM





NON
PECUNIAE INVESTIGATIONIS CURATORE
SED VITAE FACIMUS
PROGRAMMAS DATORUM PROCURATIONIS

(Not
for the research funder
but for life we make
data management plans)

Image by Chrause via [wikimedia.org/wiki/File%3ANon_scolae.jpg](https://commons.wikimedia.org/wiki/File:Non_scolae.jpg)

On a personal note

1. In social sciences, with many long-tail data sets and small teams, using a simple and generic DMP template is a huge step forward.
2. But to align with e-humanities, text and data mining etc.:
3. Funders should require that (medium to) large projects comply with standards.
4. Data management is all in a day's work.
5. Planning is more important than the plan, and it is a team activity.



<http://bit.ly/28OfLIK>

Research Data Services, Expertise & Technology Solutions



Upcoming webinar:
How to write a DMP?
July 7, 11.00 CEST

Special Edition
June 2016

As several research funders are carrying out (Open) Data Pilots, **projects and researchers are increasingly required to deliver a Data Management Plan or DMP**. The DMP is a **useful instrument for researchers to reflect on and communicate about the way they will deal with their data**. It prompts them to think about how they will generate, analyse and share data during their research project and afterwards.

In this webinar **Sarah Jones (DCC) and Marjan Grootveld (DANS) will talk you through the aspects that Horizon 2020 requires from a DMP**. We'll discuss **examples from real DMPs** and also touch upon the **Software Management Plan**, which for some projects can be a sensible addition.

For introductory information about Research Data Management you can check the recordings and slides of our previous webinar: ["Research Data Management: An introductory Webinar from OpenAIRE and EUDAT"](#)

[REGISTER NOW!](#)

On a personal note

1. In social sciences, with many long-tail data sets and small teams, using a simple and generic DMP template is a huge step forward.
2. But to align with e-humanities, text and data mining etc.:
3. Funders should require that (mid to) large projects comply with standards.
4. Data management is all in a day's work.
5. Planning is more important than the plan, and it is a team activity.



marjan.grootveld@dans.knaw.nl

<http://www.dans.knaw.nl/>

<https://easy.dans.knaw.nl/> - DANS archive