

Closing the Grey Gap – An Open Data Intensive Challenge

Peter Doorn, director DAM
Dominic Farace, Greynet



Open Science 2020: Harmonizing Current OA practices with H2020 Guidelines

Pisa, 8 April 2014





Contents

- Data fraud and the awareness of data policies
- To share or not to share...
- Different data (sharing) cultures
- Three slides about DANS and Research Data NL
- Where do we go from here?

nature

International weekly journal

News

Report finds massive fraud at Dutch universities

Investigation claims dozens of social-psychology studies contain faked data.



Science Insider

Breaking analysis from the world's leading science policy

Report: Lord of the Flies? Charged Dozens of Scientists (UPDATE)

on 31 October 2011, 7:05 PM

the guardian

Public sector can combat fraud with data sharing

Outsourcing is not the only game for procurement, says **Graham Kemp**, and the public sector needs to be less of a security risk but knowledge to be shared

The New York Times

Fraud Case in Netherlands a Red Flag for Psychology Research

By BENEDICT W. L. ... psychologist in the Netherlands who published widely in professional journals and made up entire experiments, a committee has found. Experts say the case reveals deep flaws in the way science is done in a field that has only recently earned a fragile reputation.



SPIEGEL ONLINE

Niederlande
Frühjahr Psychologie steht
Forschungen

PLOS

accelerating the publication of science

Home Browse For Readers For Authors and Reviewers

RESP

OPEN ACCESS

Do Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data

Driven by data

September 2011: Diederik Stapel, Social Psychology

November 2011: Don Poldermans, Cardiovascular Medicine

June 2012: Dirk Smeesters, Experimental Social Psychology

October 2012: Mart Bax, Cultural Anthropology



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

Driven by data



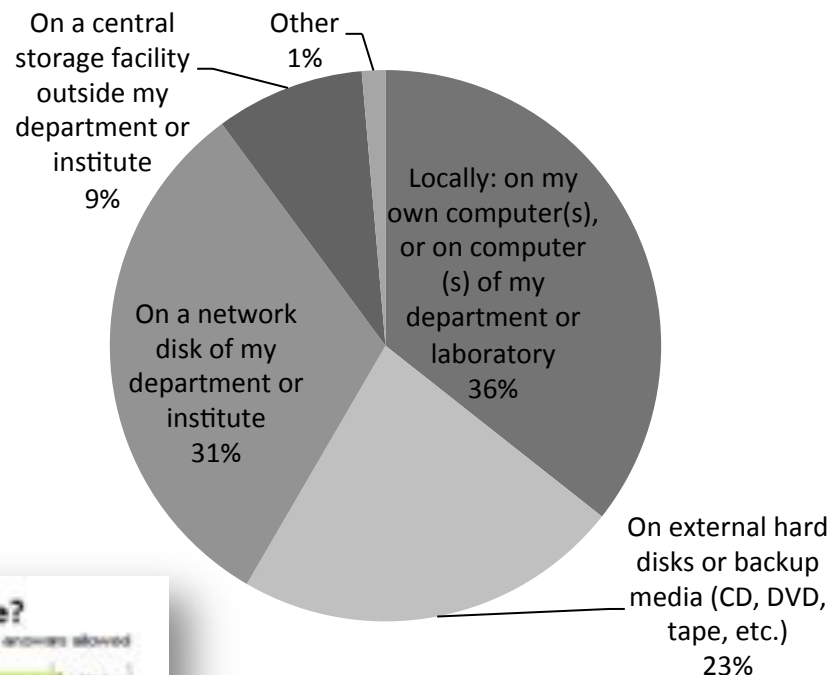
Increased awareness of need for Data Policies

- **Dutch Academy (KNAW)** “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”.
- **Dutch Universities** are developing data policies
- **Dutch research funding organisations NWO and ZonMw**: DMP and data sharing are requirements for funding projects
- **Science Europe**: Access to research data first priority area on S.E. Roadmap (December 2013)
- **EU research funding programme Horizon 2020**: Guidelines on data management
- **EU Vice-President Neelie Kroes**: “Data is the new gold” (Riding the Wave report, 2010)

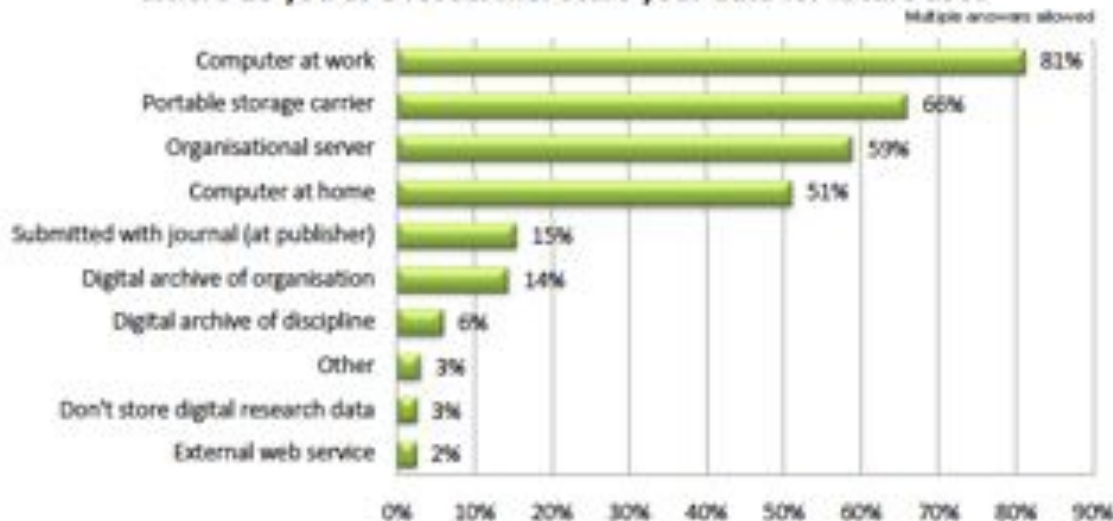




Data sharing cultures across scientific domains



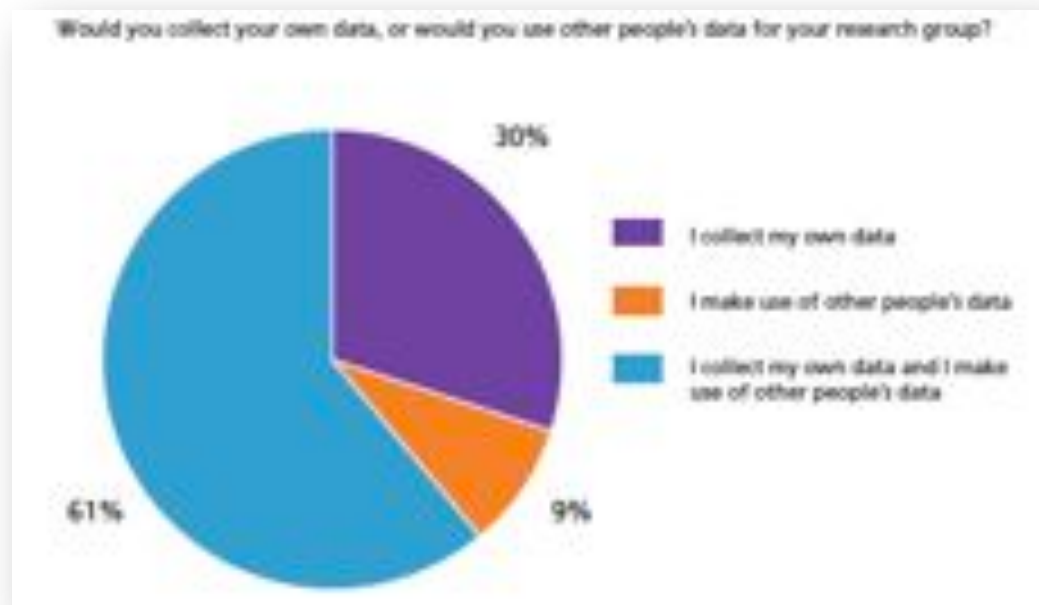
Where do you as a researcher store your data for future use?





Psychology: a closed data community

- Psychologists hardly have a tradition of sharing data.
- Differences between sub-disciplines within psychology in terms of their view on data sharing and data archiving.





To share or not to share...

What is in it for the researcher? Benefits:

- Visibility
- Citation: researchers who share data are cited more often than others



Piwowar, H., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ PrePrints*, 1:e1. doi:10.7287/peerj.preprints.1

"[...] we find a robust citation benefit from open data [...]

*there is a direct effect of third-party data reuse that persists for years [...]
a substantial fraction of archived datasets are reused, and the intensity of
dataset reuse has been steadily increasing since 2003."*



Six common objections to data sharing... and how to overcome them (1-4)

Why not share?	How to overcome?
1. No one else can understand the complexity of my data	Document the data, describing the conditions of the research
2. If someone else analyzes my data, they may come up with a different answer disproving my perspective	By considering different perspectives on the same data set, we will come closer to the “right” answer
3. Someone else may find something new in my data that I did not see	Finding something new in an existing data set will increase the return on investment in the data collection
4. I have not finished analyzing my data, and I will make it available once my analysis is complete	Research is a never-ending story... A published paper suggests that the data have been substantially analyzed; thus sharing at this point seems appropriate



Six common objections to data sharing... and how to overcome them (5-7)

Why not share?	How to overcome?
5. It is my data that I worked very hard to collect, and no one else has the right to it.	Publicly funded data should be publicly available. A publication implies that research results are to be shared. Reviewers and readers should have access to the primary data on which publications are based.
6. I cannot trust or understand the data produced in another laboratory	If this is not possible, who can we trust the the scientific literature? This is the mirror image of 1.
7. Documenting my data for others costs me time for which I get no credit	You will be rewarded by more citations; your work will be more visible and your data will be cited as well.

Adapted from Stephen H. Koslow (2000), 'Should the neuroscience community make a paradigm shift to sharing primary data?', *Nature America Inc.*, 3:9 (September), p. 863-865.



What is 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



Our services



EASY: Electronic Archiving System for self-deposit



Data Seal of Approval



Persistent Identifier
URN:NBN resolver



NARCIS: Gateway to scholarly
information In the Netherlands

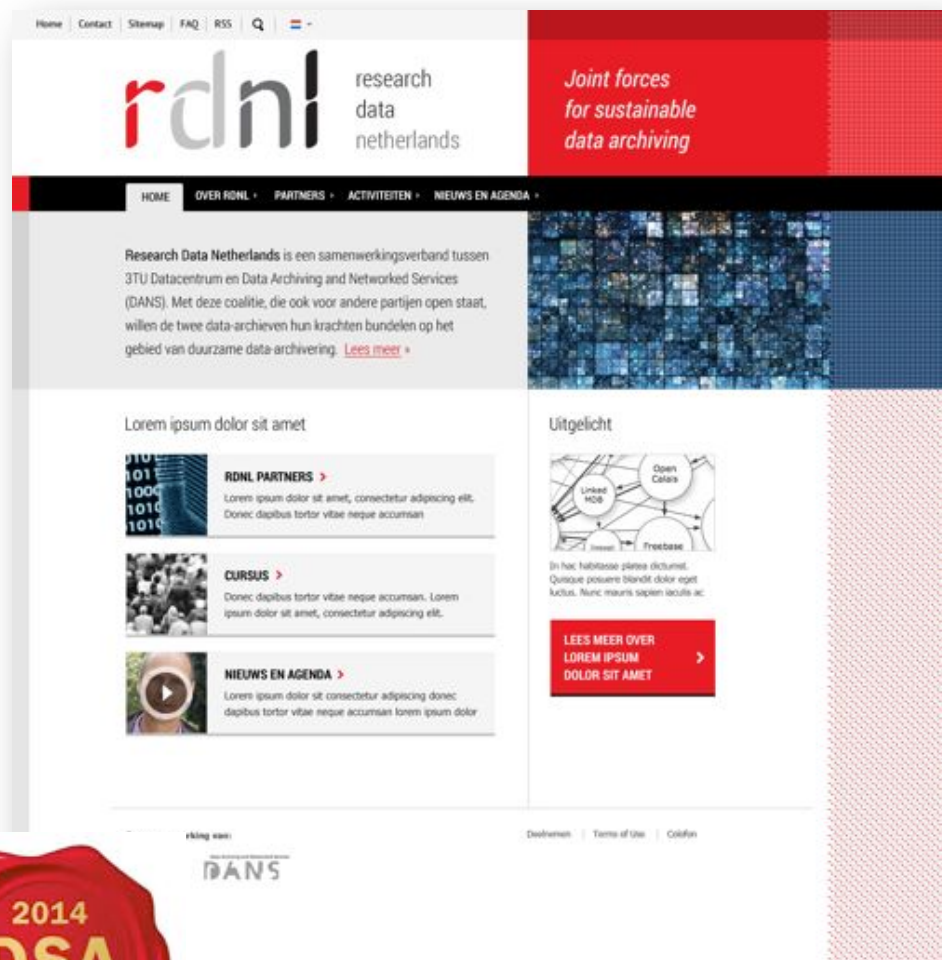
Driven by data

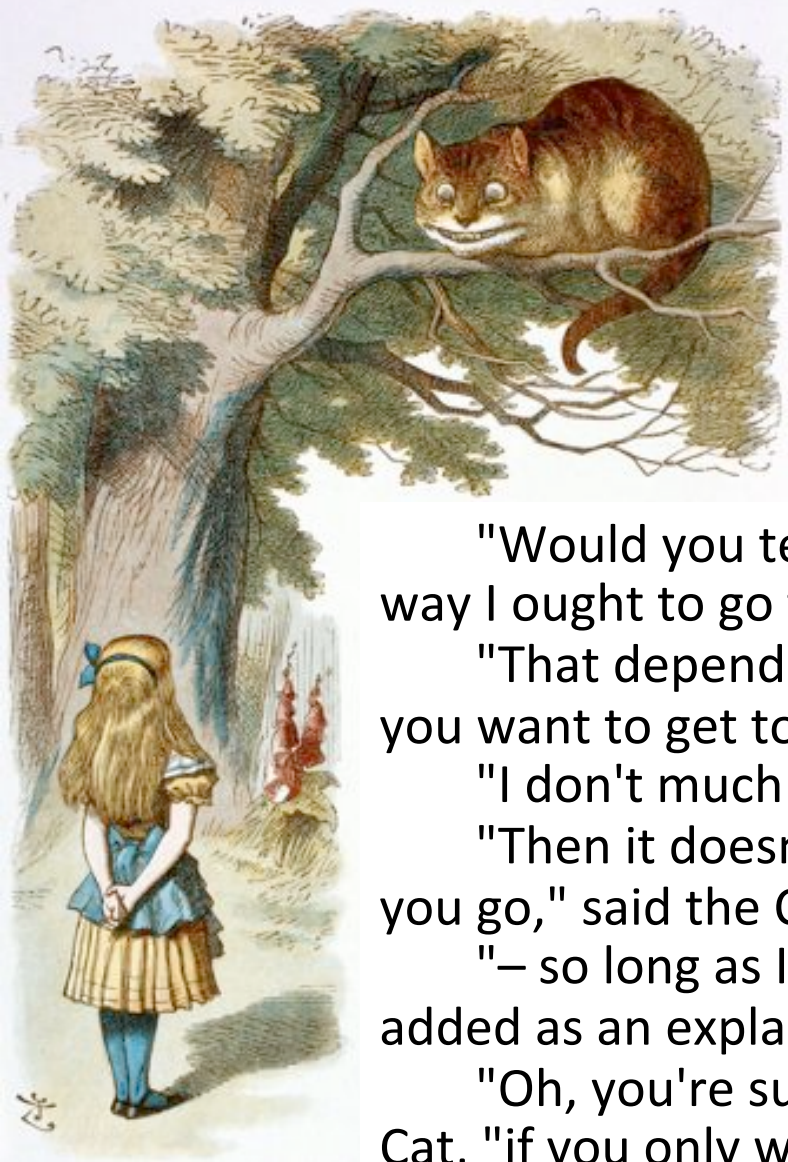


Research Data Netherlands

www.researchdatanetherlands.nl

- Mission: the promotion of sustained access and responsible re-use of digital research data
- Builds upon existing cooperation of DANS & 3TU.Datacentrum (D4L training, Dutch Dataprize)
- Expanding cooperation on back-office functions
- Open to other trusted digital repositories





Where should we go from here?

"Would you tell me, please, which way I ought to go from here?"

"That depends a good deal on where you want to get to," said the Cat.

"I don't much care where —" said Alice.

"Then it doesn't matter which way you go," said the Cat.

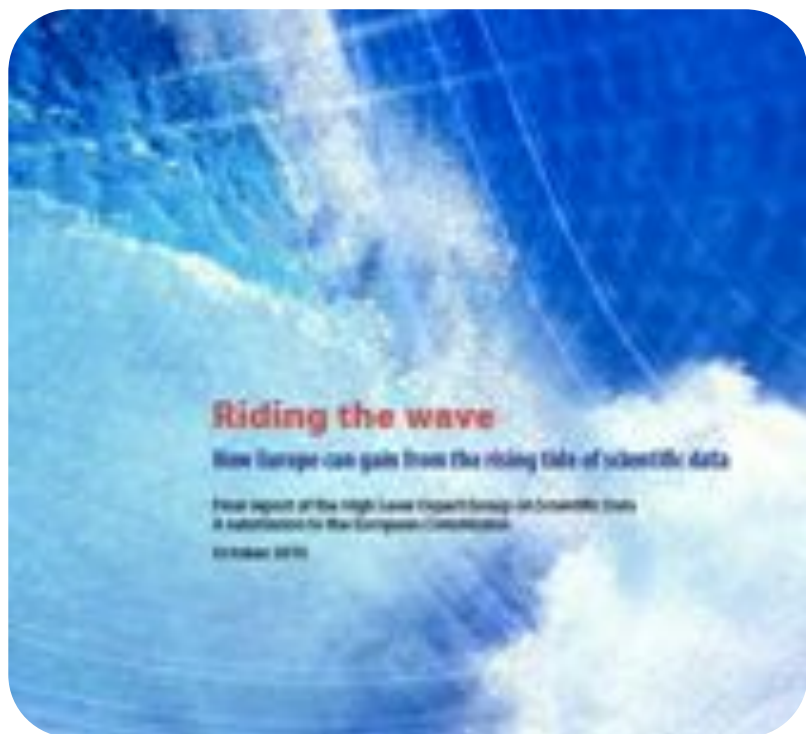
"— so long as I get *somewhere*," Alice added as an explanation.

"Oh, you're sure to do that," said the Cat, "if you only walk long enough."

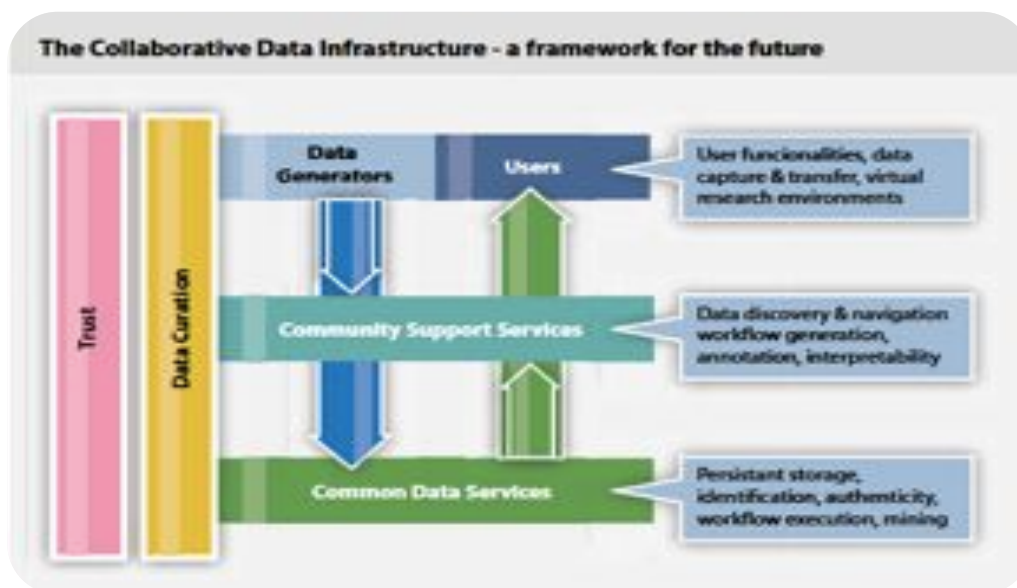




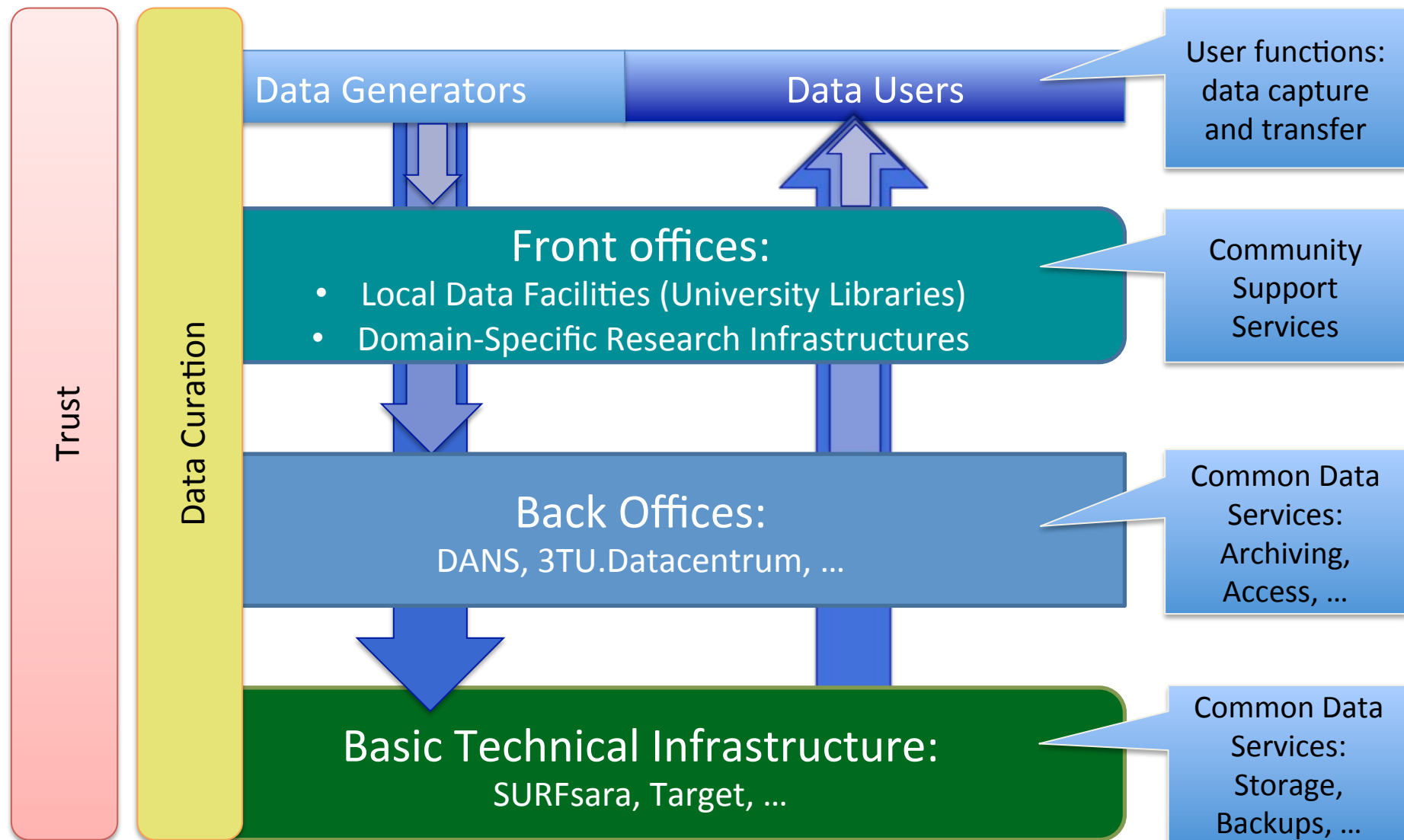
Riding the wave (Oct. 2010)



1st recommendation: Develop an international framework for a Collaborative Data Infrastructure

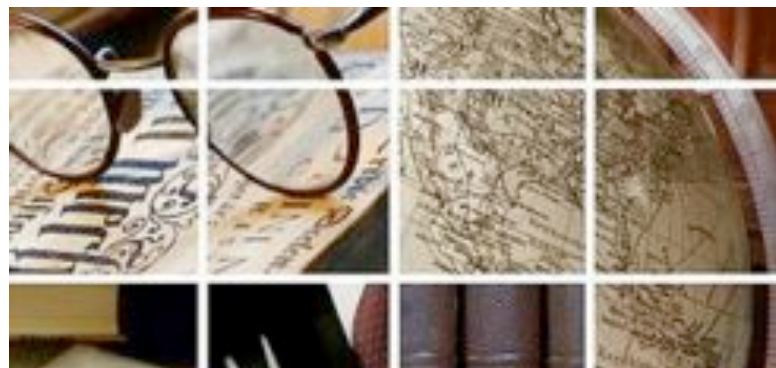


The federated data infrastructure: a collaborative framework





Changes in process of knowledge creation



Explicit knowledge can be captured, written down and presented in documents and databases.

Tacit knowledge

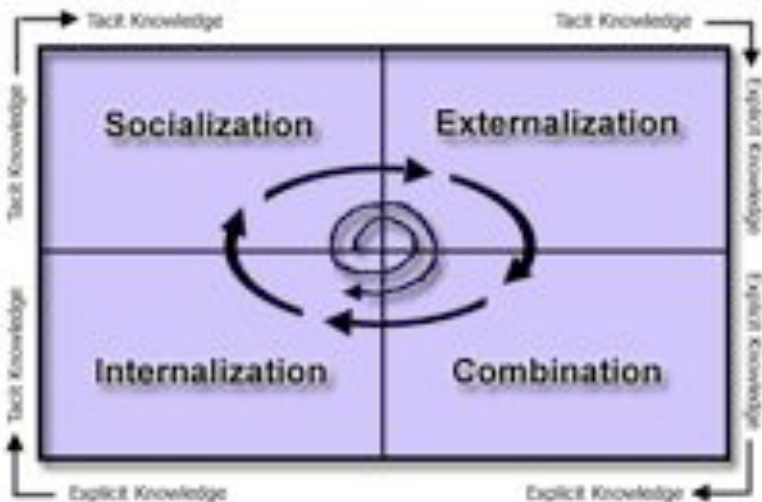
Is the knowledge in our heads. It is much less concrete and more difficult to document and measure, however, it is more valuable because it provides context.

Tacit knowledge requires personal contact and trust to share effectively.



Driven by data

Spiral of Knowledge Creation
by Nonaka & Takeuchi (1995)





Changes in knowledge sharing

What is to remain stable, what should change?

- Research data:
 - RDM during research process
 - Archiving of stable versions
- Narrative text:
 - Introduction
 - Abstract
 - Methodology
 - Conclusions
 - Discussion
 - References/citations
- Software:
 - Maintain for sustainability

More Stable



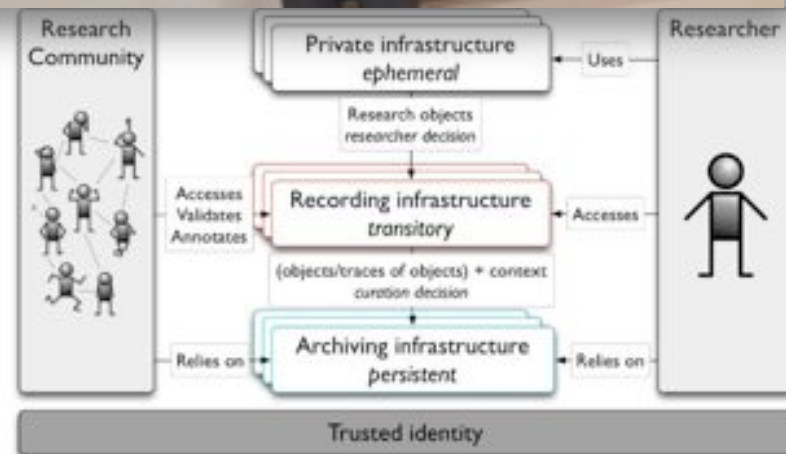
**More
Changeable**



Herbert van de Sompel ([Los Alamos](#)) and Andrew Treloar ([ANDS](#)) at DANS workshop on the 'Scholarly Archive of the Future' (20/1/2014)

Characterising the future

Hidden	← Research Process →	Visible
Fixed	← Nature of object →	Varying
Atomic	← Atomidity of object →	Compound
Discrete	← Process of making public →	Continuous
Delayed	← Speed of communication →	Instant
Publication + data proxies	← Communicated object →	Publication + linked data + linked models
Formal	← Nature of process →	Informal



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.



Reaping the fruits:

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

Already we admire the
stems hanging
on, the eggs so very
high...
(Reaping the fruits from
the first Dutch children's
book 'Nieuwe geestes-
keur kinderboek' by Willem
van Aelst, 1770).

RDA PLENARY 4 MEETING IS CO-ORGANISED BY:

3TU, Datacentrum  netherlands  center



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-co-located-events.html>



50 years Data Archiving and The Dutch Data Award



APARSEN: Alliance Permanent Access to the Records of Science in Europe Network



Creative-B: Workshop "Global Roadmap for biodiversity research infrastructures"



Data Science Research Center Amsterdam: new opportunities for Research Data



Data Seal of Approval Conference 2014



EUDAT 3rd Conference - Bringing data infrastructures to Horizon 2020



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>

Data Archiving and Networked Services

Thank you for your attention

www.dans.knaw.nl

www.narcis.nl

peter.doorn@dans.knaw.nl



The work DANS does
summarized in a 4 minute
video, introduced by
Neelie Kroes

<http://youtu.be/HJbo-OAaJ1I>

DANS is an institute of KNAW en NWO

Driven by data