



New Developments in Data Management Support

for Universities, Research Institutes and Higher Education in The Netherlands

Peter Doorn, Vyacheslav Tykhonov & Marion Wittenberg (DANS)

IASSIST 2017, Lawrence, Kansas, May 23-26, 2017

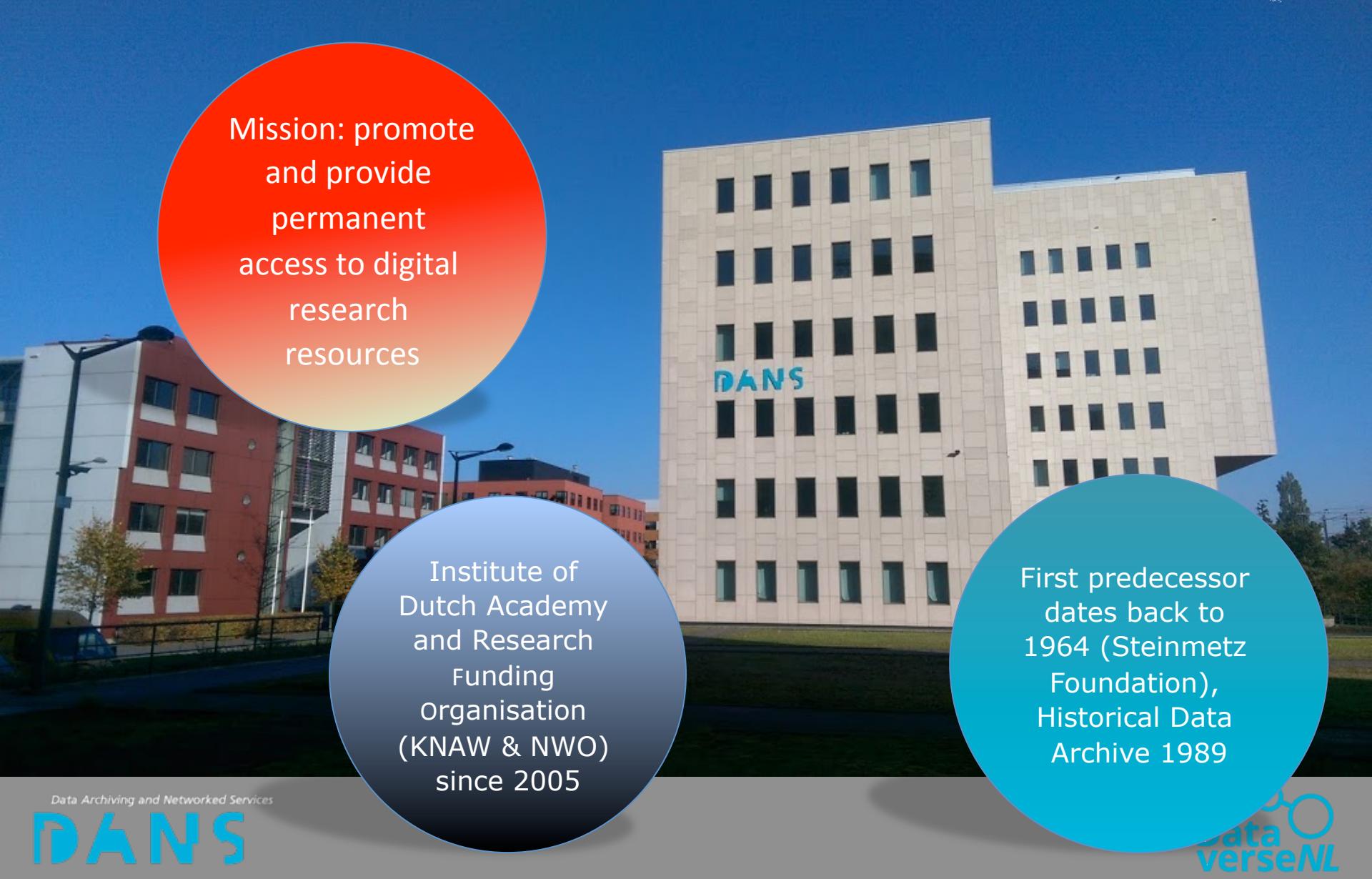
Session B2: Repository Strategies across Communities

Wednesday, 24 May 2017, 10:30 – 12:00, Room Big 12



@pkdoorn @dansknaw

Data Archiving & Networked Services

A photograph of several modern buildings under a clear blue sky. In the foreground, there's a red brick building and a white building. In the background, there are more buildings, including a prominent light-colored multi-story building with many windows. A street lamp is visible on the left.

Mission: promote
and provide
permanent
access to digital
research
resources

Institute of
Dutch Academy
and Research
Funding
Organisation
(KNAW & NWO)
since 2005

First predecessor
dates back to
1964 (Steinmetz
Foundation),
Historical Data
Archive 1989

DANS core data services

Archaeology
Instructions in English or Nederlands (PDF).
[Start deposit](#)

History
Instructions in English or Nederlands (PDF).
[Start deposit](#)

Social and behavioural sciences
Instructions in English or Nederlands (PDF).
[Start deposit](#)

Life sciences and medicine
Instructions in English or Nederlands (PDF).
[Start deposit](#)

Language and literature
Instructions in English or Nederlands (PDF).
[Start deposit](#)

All other disciplines
Instructions in English or Nederlands (PDF).
[Start deposit](#)

Data Archiving and Networked Services



NARCIS

The gateway to scholarly information in the Netherlands

> Submit Content to NARCIS

Search...

SEARCH

1,311,325
PUBLICATIONS

161,149
DATA SETS

64,662
RESEARCH

53,583
PEOPLE

2,948
ORGANISATIONS

500.000 Open Access
(40% → 25%)



NARCIS: Gateway to scholarly information In the Netherlands

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

Q About Guides Support Sign Up Log In

Dataverse

DataverseNL Dataverse

Metrics 6,588 Downloads

DataverseNL Dataverse Network

4TU CENTRE FOR RESEARCH DATA Vrije Universiteit Amsterdam Dataverse NIOO-KNAW Dataverse Tilburg University Dataverse

Search this dataverse... Find Advanced Search

HO

1 to 10 of 672 Results

Mindwandering during Attention Performance: effects of ADHD-Inattention Symptomatology, Negative Mood, Ruminative Response Style and Working Memory Capacity Feb 3, 2017 - Clinical Psychological Science Dataverse Jorinkman, Lies; Markus, Rob; Franklin, Michael; Dafsen, Jens van, 2017, "Mindwandering during Attention Performance: effects of ADHD-Inattention Symptomatology, Negative Mood, Ruminative Response Style and Working Memory Capacity", hdl:10411/20895, DataverseNL Dataverse, V1

Dataset_Bamelis_et_al_2015 Feb 1, 2017 - Clinical Psychological Science Dataverse Wetzelar, Pim; Arntz, Arnoud, 2017, "Dataset_Bamelis_et_al_2015", hdl:10411/20892, DataverseNL Dataverse, V1 Dataset of the study by Bamelis et al., 2015 These data were also used in the study by Wetzelar et al., 2017 Please refer to the file SupplInfo_Dataset_Bamelis_et_al_2015 for more information.

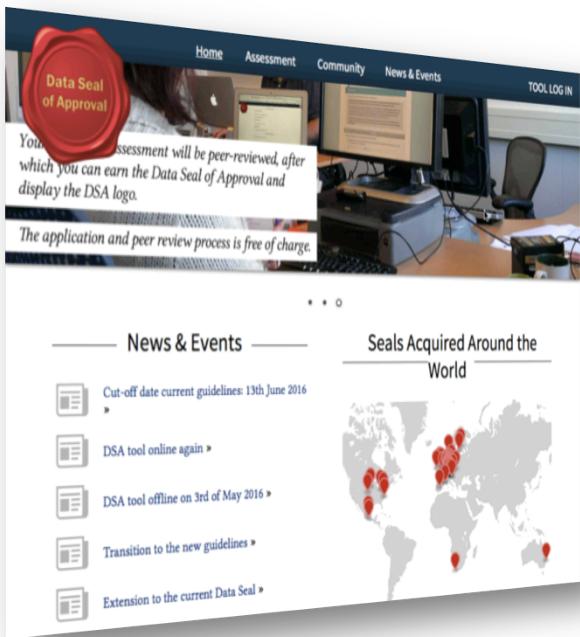
Sort ▾

DataverseNL for short/
intermediate term
storage (including RDM
during research)



Three additional services

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



Cradle of the
Data Seal of Approval

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



Persistent Identifier
URN:NBN resolver

Training & Consultancy



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

Three recent collaborative services



Background Archive



ELSEVIER



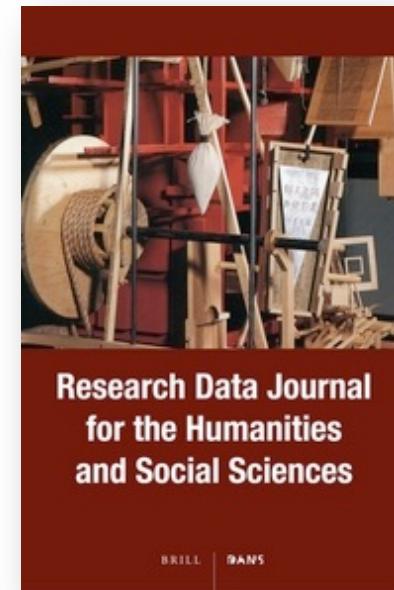
<https://data.mendeley.com/>

Software Heritage Archive



<https://www.softwareheritage.org/>

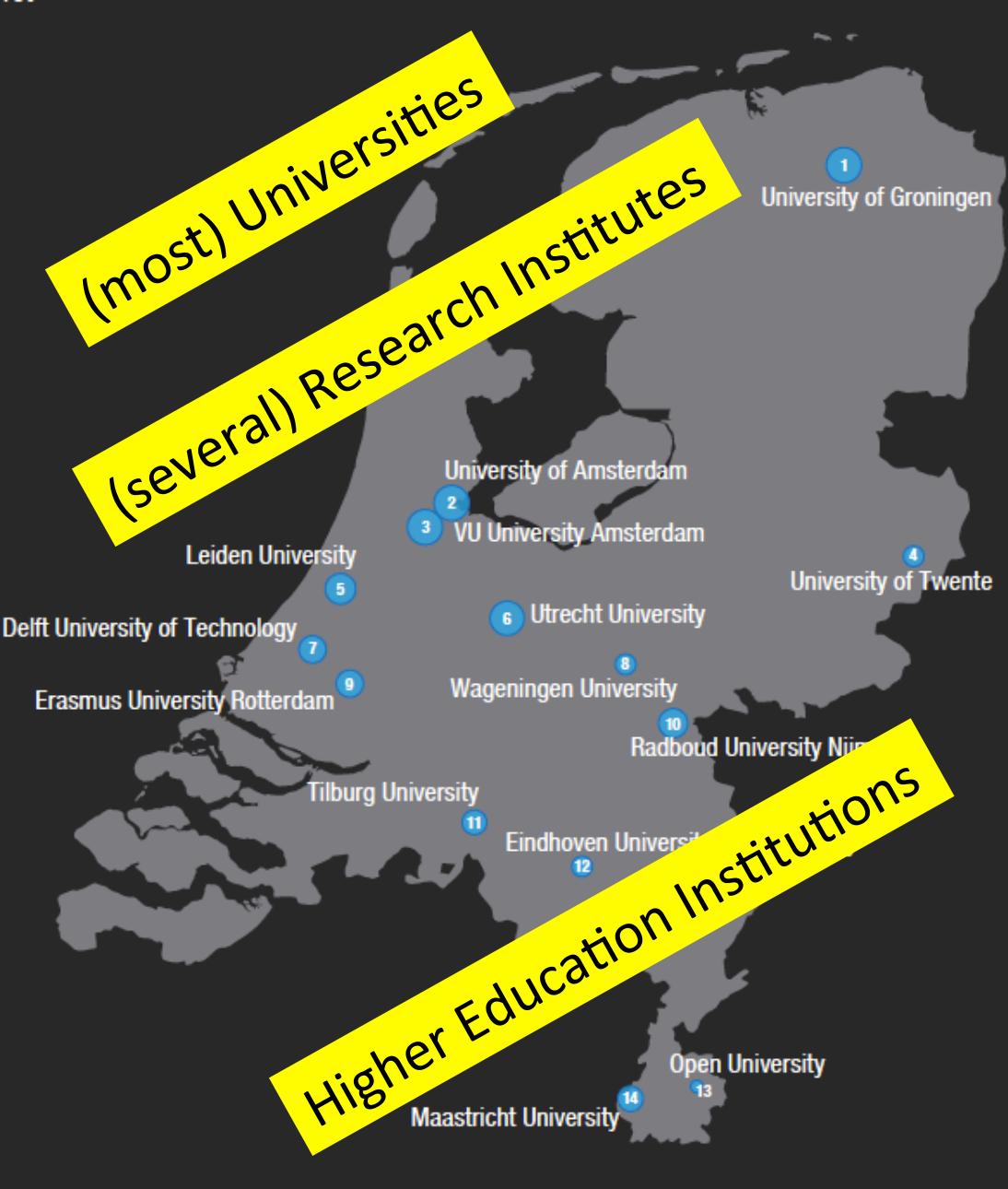
Research Data Journal
for the Humanities
and Social Sciences



<http://www.brill.com/rdj>

Why DataverseNL?

- Growing requirements for data managements by Research Funders & Institutions
- Institutions want to provide a system to support RDM and data sharing for their staff
- Changing role of Research/University libraries: finding new roles in information/data services
- Alternatives for institutions:
 1. Develop local data services including data repository → Leiden University
 2. Use a commercial supplier of repository services/systems → University of Amsterdam
 3. Use existing national/international data repository services → Continuation of existing practices
 4. “Shared Services” in a public cooperation → Most Dutch universities → DataverseNL



DataVERSENL partners (15)

4TU. CENTRE FOR
RESEARCH DATA



VU VRIJE
UNIVERSITEIT
AMSTERDAM



TILBURG UNIVERSITY



Maastricht
University



NIOO

avans
hogeschool

DANS
Windesheim

Several more on the way...

Cooperative Model

- Collaboration agreement + S.L.A.
- Shared costs (service membership + storage)
- Advisory Board consisting of partner representatives decides on the general policy and direction
- Administrators committee to discuss technical and functional issues
- DANS provides and manages the system & storage, organizes the meetings (back-office)
- Research institutions run their own RDM-support for their staff (front-office)
- Every partner is responsible for their own content
- Work in progress to transfer data from Dataverse to certified long-term storage (DANS EASY or 4TU.Data Centre)

Advantages for partners

- Efficient: low costs for partners
- Build common RDM support experience
- Advisory Board and Administrators Committee provide platform for exchange and learning from each other
- One instance: search across partner holdings
- Dataverse (developed at Harvard IQSS) is one of the most used data repository systems
 - Re3Data: 43 occurrences, see: <http://www.re3data.org/>
 - Dataverse.org: 23 installations
- Dataverse is Open Source

World-Wide Dataverse Installations



Some peculiarities of DataverseNL

- Dataverse is essentially a rigorous repository/archiving system, it is not as free & flexible as a “Dropbox” or VRE
 - but this exactly helps to guarantee the integrity of data during research; in a VRE changes to the data can go undocumented and unnoticed
 - Track data provenance
- Why is DataverseNL a solution for RDM during research and for short-term archiving?
 - Partners are responsible for the content and RDM policies → DANS cannot guarantee the long-term preservation
 - Data during research is often changed and does not always comply with long-term preservation criteria (e.g. privacy)

DataVERSE advantages and added functions

- Excellent integration in institutional websites
- Federated Login for Dutch universities
- Reporting tool on content: dashboard for administrators
- Interactive data exploration and analysis tools, visualisation and geographic mapping of data
- Metadata formats for various disciplines (social sciences, humanities, astronomy, bio-sciences) compliant with DDI
- Compliant with data citation and data publishing guidelines
- Integration with e-journals (OJS)
- Privacy tools for sharing sensitive data (DataTags)
- Dataverses can reflect the structure of the organisation (University, Faculty, Department, Research Group, Project)
- Work on implementation of FAIR principles in progress
- "Big data" solutions

DataVERSENL integrated in University websites

VU, Amsterdam

**VU
VU
UNIVERSITY
AMSTERDAM** VUNET

Search VUnet

| SERVICES | CURRENT | ORGANIZATION | THEMES |

ARCHIVING RESEARCH DATA

Archiving new dataset

Choose one of these options to archive your dataset.

WIZARD 

Help in choosing a data repository

- By answering a number of questions, you will quickly determine which data storage solution is most suitable for you.

DataverseNL 
For publishing on the internet
Not suitable for privacy or other sensitive information
Researchers decide whether access to dataset is public or limited
For publishing open data according to grant providers' regulations
Includes a link (persistent identifier) for data citations in publications
Retention period is at least 10 years
Bibliographical details (metadata) of VU researchers can be found in Metis
€3,50 per GB per year

DarkStor (< 50GB) 
For archiving sensitive data in a secured archive
Suitable for privacy information
Researchers can only access their data through an e-mail request

ArchStor (< 50GB) 
For archiving research data
Not suitable for privacy info
Researchers can only access request
Conform the VSNU code of reproducibility
No link (persistent identifier)
Retention period is 10 years
Bibliographical details (meta) be found in Metis

ArchStor (> 50GB) 
For a description, see Arch
For datasets larger than 50 GB, budget holder permission is required
Also see the manual "How do I compress my data?"
€0,22 per GB per year

University Library
A collection with over 800,000 books, journals, illustrations, early works and manuscripts. The combination with a voluminous collection of e-books and e-journals provides a treasure of scholarly information.

 [Opening hours →](#)

 [LIS Service Desk →](#)

DataVERSENL DataVERSE > Tilburg University DataVERSE

 [Tilburg School of Catholic Theology Dataverse](#)

 [Tilburg School of Social and Behavioral Sciences Dataverse](#)

Search this dataverse... Advanced Search

Filter Results

1 to 10 of 272 Results

 [Department of Practical Theology and Religious Studies Dataverse \(Tilburg University\)](#)
Dec 8, 2016 Tilburg School of Catholic Theology Dataverse

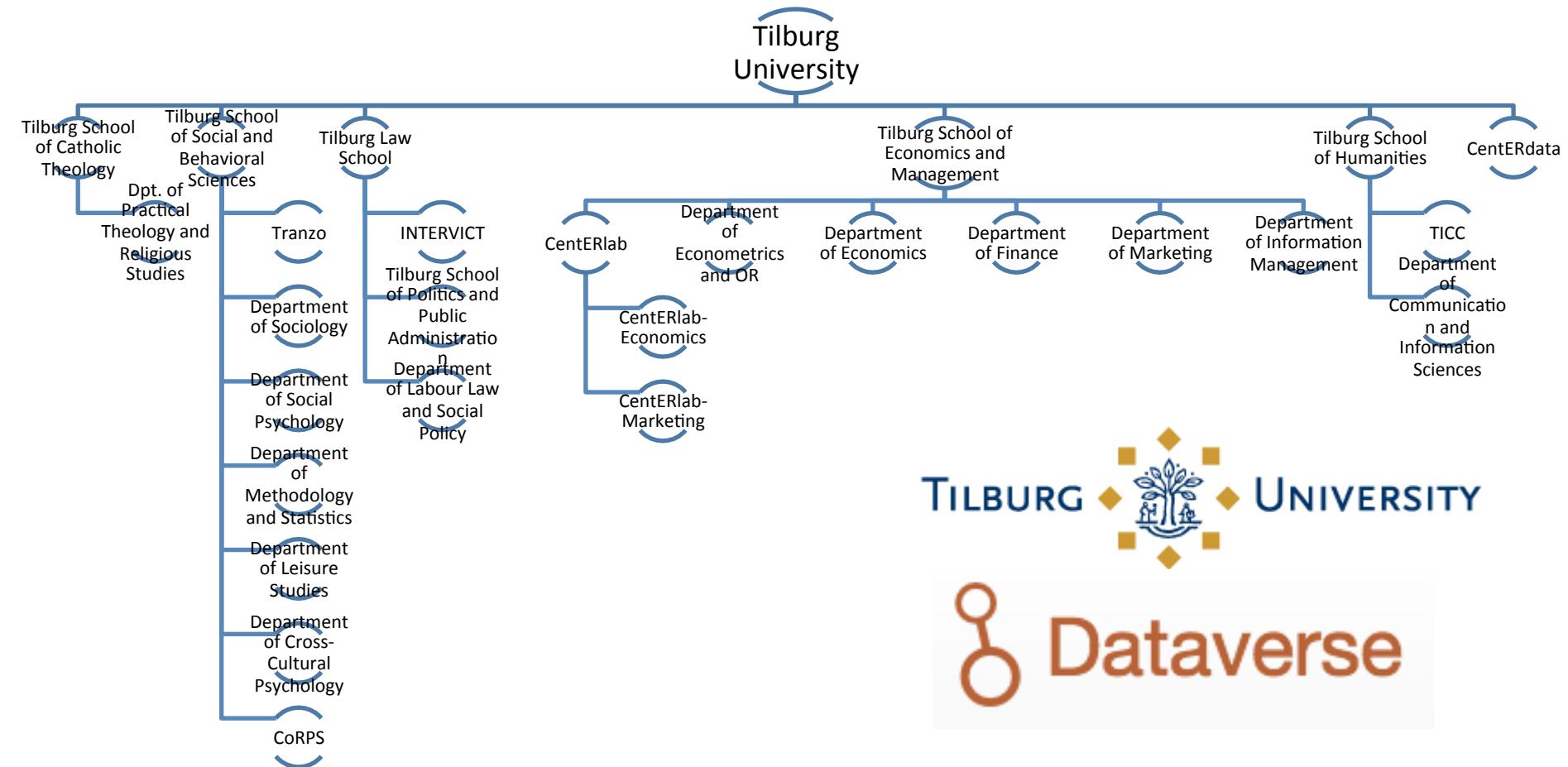
 [TIICC Dataverse \(Tilburg University\)](#)
Dec 8, 2016 Tilburg School of Humanities Dataverse

Data is stored at DataVERSENL DataVERSE. [Log In](#)

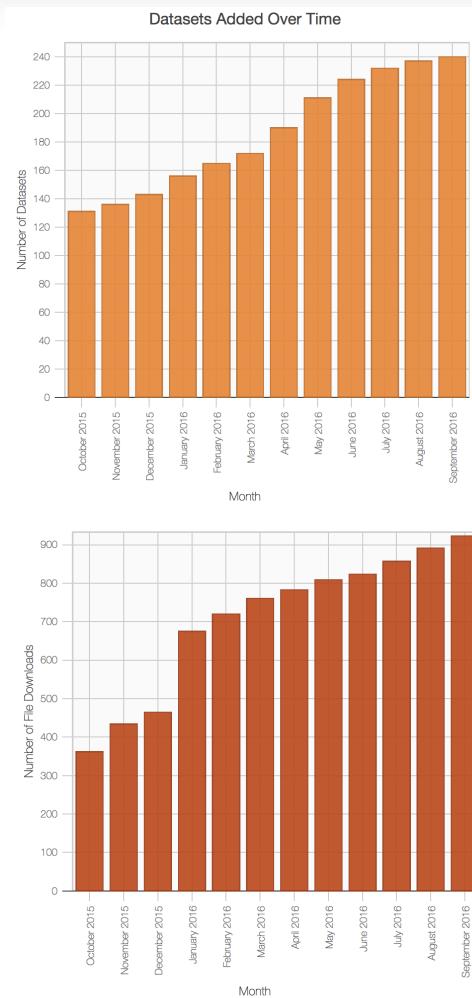
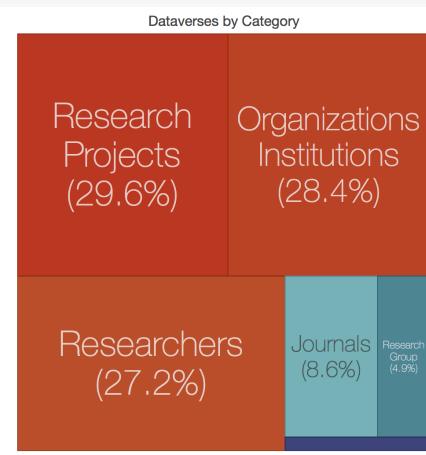
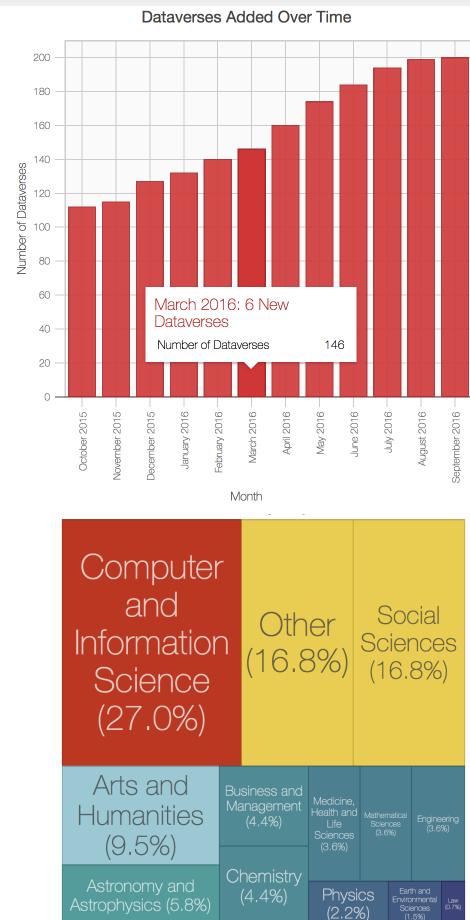
Powered by 

Tilburg University

Structure of Tilburg University Dataverse

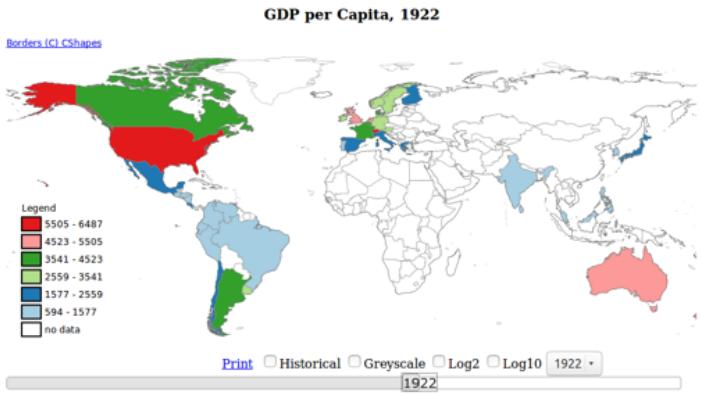
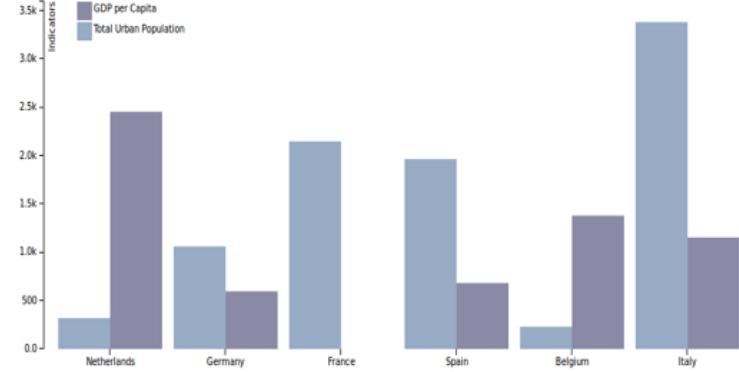
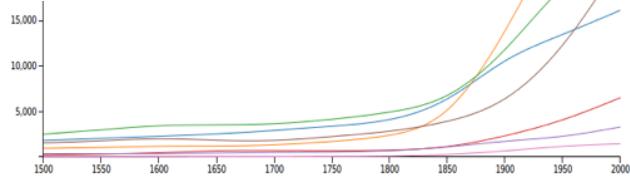


Real-time metrics: Dashboard



Data explorations and visualisations

Handle	Max	Mean	Min	N	Period	Std
Total Number of Goats	154000.0	1779.855	0.0	7037	1500-2010	6625.954
Cattle per Capita	50.0	0.452	0.0	7456	1500-2010	1.192
Goats per Capita	7.561	0.257	0.0	7037	1500-2010	0.119
Pigs per Capita	40.412	0.367	0.0	6841	1500-2010	0.398
Total Cattle	210384.0	3461.482	0.0	7457	1500-2010	8273.001



TwoRavens, which integrates with Zelig statistical framework)

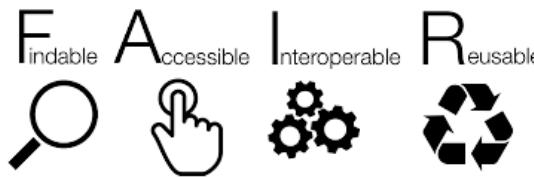
Work in progress on FAIR Data Assessment and GDPR compliant Data Tags



Level	DUA Agreement Method	Authentication	Transit	Storage
No Risk	None	None	Clear	Clear
Minimal	None	Email or OAuth	Clear	Clear
Shame	Click Through	Password	Encrypted	Clear
Civil Penalties	Sign	Password	Encrypted	Encrypted
Criminal Penalties	Sign	Two Factor	Encrypted	Encrypted
Max Control	Sign	Two Factor	Double Encryption	Double Encryption



F A I R
1 Archivist Assessment
3 User Reviews
16 Downloads



Disadvantages/weaknesses/challenges?

- Collaborative model is less business-like and straightforward:
 - Some institutions may prefer other business models:
 - outsourcing to a commercial provider
 - developing and managing in-house solutions
- How to serve best universities with different priorities and RDM policies with one system and under one agreement?
 - It is possible, but at a cost: it takes time to agree on common wishes (e.g. future functional requirements)
 - But: does it make sense that RDM policies and practices vary among institutions?
 - Some partners have stricter requirements for Service Level Agreements than others
 - Solution: incorporate in the business model → the higher the demands, the higher the membership fee

DataverseEU for new CESSDA partners?

CESSDA SAW “Strengthening and Widening” EU Project

The diagram features two main circular icons. The top icon is blue and contains the text "cessda SAW". Around it are flags of various European countries. A downward-pointing arrow from this icon is labeled "Strengthening and Widening". The bottom icon is black and contains the text "cessda". Around it are flags of different European countries. Below these icons is a horizontal row of flags representing the same countries. A diagonal banner across the bottom reads "Strengthening and Widening CESSDA". To the right of the icons is a map of Europe. The map shows color-coded regions: dark purple (top), medium purple (center), and light purple (bottom). Overlaid on the map are several circular icons representing project work packages (WP1 to WP5) and a legend. The legend includes:

- CESSDA SAW members (light purple)
- CESSDA SAW countries (medium purple)
- CESSDA = CESSDA SAW (dark purple)

Text on the map includes:

CESSDA SAW members

WP2 Dissemination & communication

WP3 Strengthening and widening through research & development

WP1 Project management & administration

WP4 Strengthening and widening through social media exchange

WP5 Strengthening and widening through pilot experiments

CESSDA – Strengthening and Widening
start: 01.08.2013 – end: 31.07.2017

All countries which are members of CESSDA are also members of CESSDA SAW and vice versa. All countries which are members of CESSDA SAW are also members of CESSDA.

CESSDA SAW
www.cessdasaw.eu
Project ID: 302077
Budget: 1.000.000 EUR
Duration: 01.08.2013 – 31.07.2017
Co-funded by the European Union
with additional funding from CESSDA's partners

Questions?

peter.doorn@dans.knaw.nl
www.dans.knaw.nl
<https://dataverse.nl/>



Watch our videos and webinars on YouTube

<https://www.youtube.com/user/DANSDataArchiving>