



Data Archiving and Networked Services



Cost and benefits of digital preservation in a federated data infrastructure:

Who pays what and who gets what in The Netherlands?



Peter Doorn
Director, DANS



EUDAT Conference
Cost and funding models

EUDAT



Rome, 28-30 October 2013

DANS is an institute of KNAW en NWO

Driven by data



Contents

- Cost models
- The value of research data
- About DANS and our services
- Cost models and projects at DANS
- From theory to practice
- The collaborative/federated data infrastructure
- Front and Back Office (FOBO)
- Costs and benefits in the FOBO model
- Business model for the BO functions



„DE COST GAET VOOR DE BAET UYT.”



Outlay must precede returns or Costs come before profit or No pain, no gain

18th Century “Bureau for Trade Information” next to Stock Exchange, Amsterdam (now a coffee shop)

DANS

Driven by data



Digital preservation costing initiatives

UNIVERSITY OF LEEDS



- LIFE 1, 2 and 3. Projects to explore digital preservation costing, and develop costing models.
- Cost Model for Digital Preservation (CMDP): Project at the Royal Danish Library and the Danish National Archives to develop a new cost model. Currently covers Planning, Migrations and Ingest
- Keeping Research Data Safe 1 and 2 (KRDS): Cost model and benefits analysis for preserving research data
- Presto Prime cost model for digital storage
- Cost Estimation Toolkit (CET): Data centre costing model and toolkit, from NASA Goddard
- Cost Model for Small Scale Automated Digital Preservation Archives (Strodl and Rauber)
- APARSEN Project activity focused on digital preservation costing
- EPSRC and JISC study on Cost analysis of cloud computing for research
- Cost forecasting model for new digitization projects (Excel and web tool under development) (Karim Boughida, Martha Whittaker, Linda Colet, Dan Chudnov)
- DP4lib business and cost model for a digital preservation service
- DANS Costs of Digital Archiving Volume 2 Project, focusing on preservation and dissemination of research datasets
- Blue Ribbon Task Force on Sustainable Digital Preservation and Access
- Economic Sustainability Reference Model
- ENSURE Project - Enabling kNowledge Sustainability Usability and Recovery for Economic value
- Cost Model for Electronic Health Records (Bote, Fernandez-Feijoo, and Ruizb)
- 4C. EU funded project on costing. Due to commence in 2012. Led by JISC
- <http://wiki.opf-labs.org/display/CDP/Home>
- [An extended blog-rant on why this typifies a big #fail for our community](#)



Paul Wheatley

“Digital Preservation Cost Modelling: Where did it all go wrong?”

<http://www.openplanetsfoundation.org/blogs/2012-06-29-digital-preservation-cost-modelling-where-did-it-all-go-wrong>

DANS

Driven by data



So many cost models and approaches...

- Most preservation activities (for research data) are publicly funded: subsidized organizations working for subsidized clients
- Open data <?> Valorization
- Preservation does not come alone: providing access, projects, ...
- Which activities (personnel costs) to include in cost calculations? What you look at is what you see!
- Costs and funding of hardware (storage and servers) and software (development of archiving systems) appear to vary a lot
- Data preservation and access is a moving target
- One model fits all?



The value of data

- Hard to quantify: investment, depreciation, added value...
- Not for profit, but for scientific progress
- Valorization: value of data increases by re-use
- Limits to growth: free data storage and access model is a dead end; how to sustain the success of the operation?
increasing data volumes lead to increasing costs
- Archiving services
 - charge re-use of data: <-> open access
 - charge deposit of data: ± gold open access
- Treat commercial customers differently?



Data Archiving and Networked Services in The Netherlands

Created in 2005, combining the forces of:

- Steinmetz Archive for the social sciences (1964)
- Netherlands Historical Data Archive (1989)
- Scientific Statistical Agency (1994)
- Electronic Depot for Netherlands Archaeology (2004)



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



Driven by data



DANS services

Data Seal
of
Approval

DANS Home Browse Advanced search P.K. Doorn My requests

Search My datasets New deposit

New deposit

Select the discipline of your dataset below. Please read the instructions first, then start the deposit procedure.

Archaeology Instructions ▶ Start deposit	History Instructions ▶ Start deposit	Social and behavioural sciences Instructions ▶ Start deposit
Life science and medicine Instructions ▶ Start deposit	Language and literature Instructions ▶ Start deposit	All other disciplines Instructions ▶ Start deposit

The Data Seal of Approval Guidelines

1. The data producer deposits the research data in a data repository with sufficient information for others to assess the scientific and scholarly quality of the research data.
2. The data producer provides the research data in formats recommended by the data repository.
3. The data producer provides the research data together with the metadata requested by the data repository.
4. The data repository has an explicit mission in the area of digital archiving and promotes it.
5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.
6. The data repository applies documented processes and procedures for managing data storage.
7. The data repository has a plan for long-term preservation of its digital assets.
8. Archiving takes place according to explicit workflows across the data life cycle.
9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.
10. The data repository enables the users to utilize the research data and refer to them.
11. The data repository ensures the integrity of the digital objects and the metadata.
12. The data repository ensures the authenticity of the digital objects and the metadata.
13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.
14. The data consumer complies with access regulations set by the data repository.
15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in higher education and research for the exchange and proper use of knowledge and information.
16. The data consumer respects the applicable licenses of the data repository regarding the use of the research data.

EASY: Electronic Archiving System for self-deposit



Persistent Identifier
URN:NBN resolver

NARCIS HOME OVER NARCIS ENGLISH

De toegang tot wetenschappelijke informatie in Nederland

Aanmelden gegevens NARCIS

Zoeken... **ZOEKEN**

804.510 PUBLICATIONS	28.594 DATASETS	59.217 ONDERZOEK	50.513 PERSONEN	2.895 ORGANISATIES	1.733 beta versie VERRIJKTE PUBLICATIONS
-------------------------	--------------------	---------------------	--------------------	-----------------------	---

NARCIS: Gateway to scholarly information In the Netherlands



Driven by data



Early approaches to covering the costs of data archiving and access

DANS Predecessors (1990s – 2005):

- “Data marketing” project of Historical Data Archive to promote re-use
- Subscription system by Steinmetz Archive (for social sciences)
- Research Funding Agency contract with Statistics Netherlands (CBS) and other govt. organisations:
 - yearly payment of K€ 450
 - subscription by faculties at reduced rate or “pay per dataset”
 - DANS made access free in 2005 and re-negotiated CBS-contract in 2010



Cost projects at DANS

Anna Palaialogk (2008/9)

Activity Based Costing Model (ABC)

- Improving tactical and strategic decision-making
- Understand the use of scarce organizational resources in various business activities



Zuleica Arias (2011)

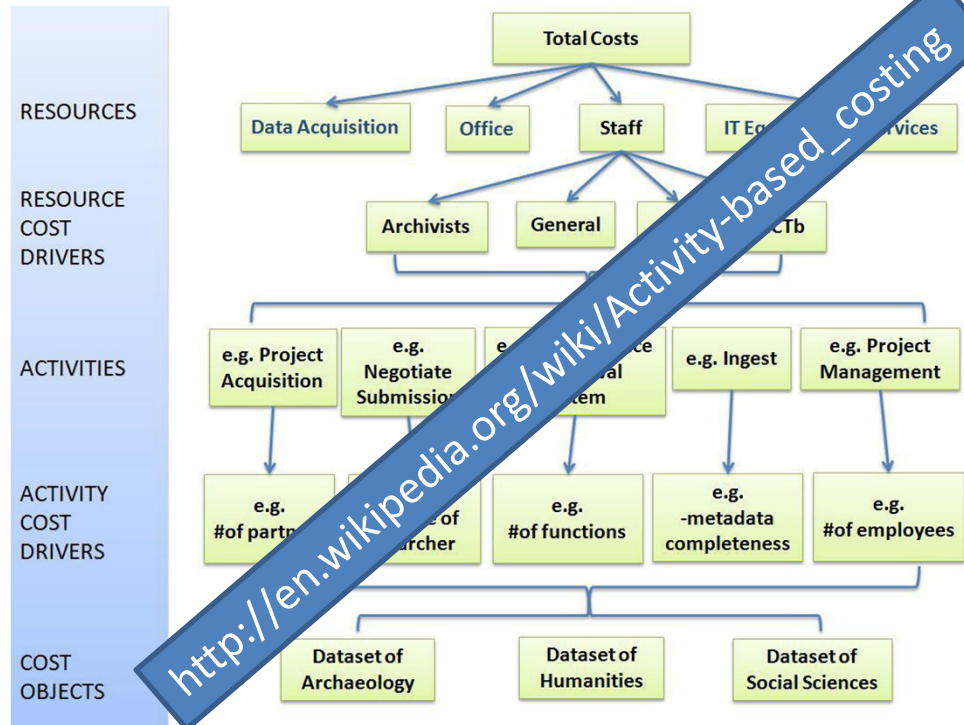
Balanced Scorecard (BSC)

Translates an organization's mission and existing business strategy into a limited number of specific strategic objectives that can be linked and measured operationally

Driven by data

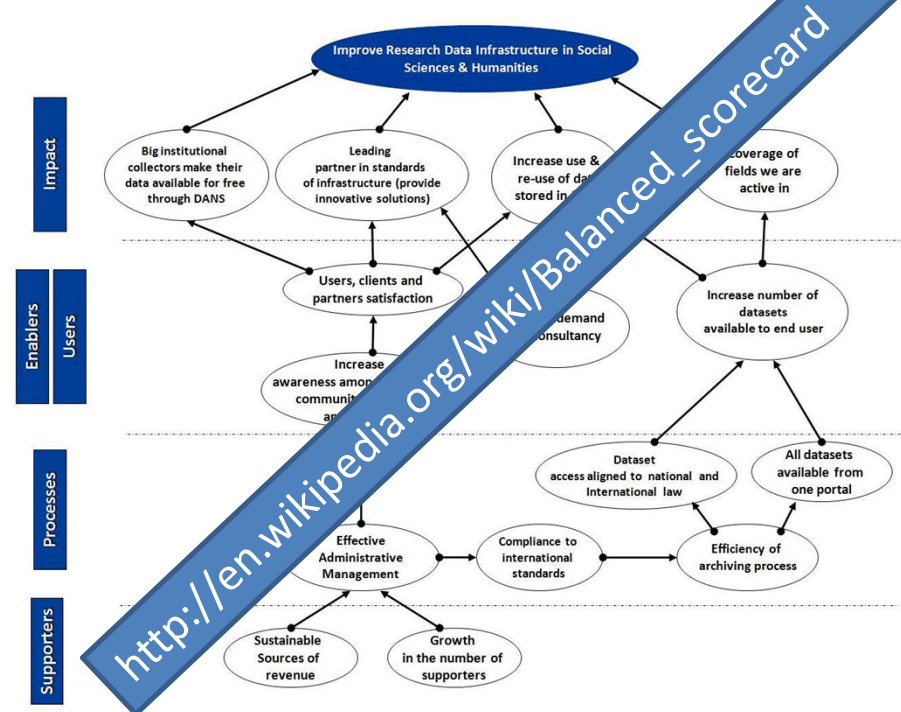


Activity Based Costing Model (ABC) Balanced Scorecard (BSC)



http://en.wikipedia.org/wiki/Activity-based_costing

Based on Cooper and Kaplan (1988)



http://en.wikipedia.org/wiki/Balanced_scorecard

Based on Kaplan and Norton (1997)

For more information see: Anna S. Palaiologk, Anastasios A. Economides, Heiko D. Tjalsma, Laurents B. Sesink (2012), 'An activity-based costing model for long-term preservation and dissemination of digital research data: the case of DANS', in: *International Journal on Digital Libraries*, Sept. 2012, 12:4, p. 195-214.
<http://link.springer.com/article/10.1007%2Fs00799-012-0092-1>



Driven by data



Table 4 The most labour- and cost-intensive activities in DANS

Activity	Labour intensity (%)	Cost intensity (%)
Preservation	17.02	29.83
Development of archival system	14.03	14.43
Project management	10.42	7.85
Administrative support	8.20	7.44
Improvement of dataset presentation and access	7.83	7.08
General management	7.07	6.24

Source: Palaiologk et al. 2012, p. 208



Current Involvement of DANS in European Projects on Costs



Heiko Tjalsma

- APARSEN Network of Excellence
 - Work package on cost/ benefit data collection and modelling
- 4C Project
 - Collaboration to Clarify the Costs of Curation





APARSEN Work package on cost modelling



Final ***Report on testing of cost models and further analysis of cost parameters*** available

Main elements:

- Analysis of the mapping of cost parameters to the ISO 16363 Standard
- Gap analysis: Testing, by comparing, three cost models (DANS, DP4lib and LIFE3)
- Review of costs in relation to benefits

<http://www.alliancepermanentaccess.org/wp-content/uploads/2012/04/D32.2-Report-on-Testing-of-Cost-Models-and-Further-Analysis-of-Cost-Parameters.pdf>

DANS

Driven by data



4C Project: Collaboration to Clarify the Costs of Curation



- Started February 2013, runs until February 2015
- Coordinated programme of outreach and engagement
- Identifying research and analysing requirements
- Assessment of gaps in the current provision of tools, frameworks and models
- Various stakeholder engagement and dissemination events (focus groups, workshops, conference)
- Establishment of an international *Curation Costs Exchange framework*

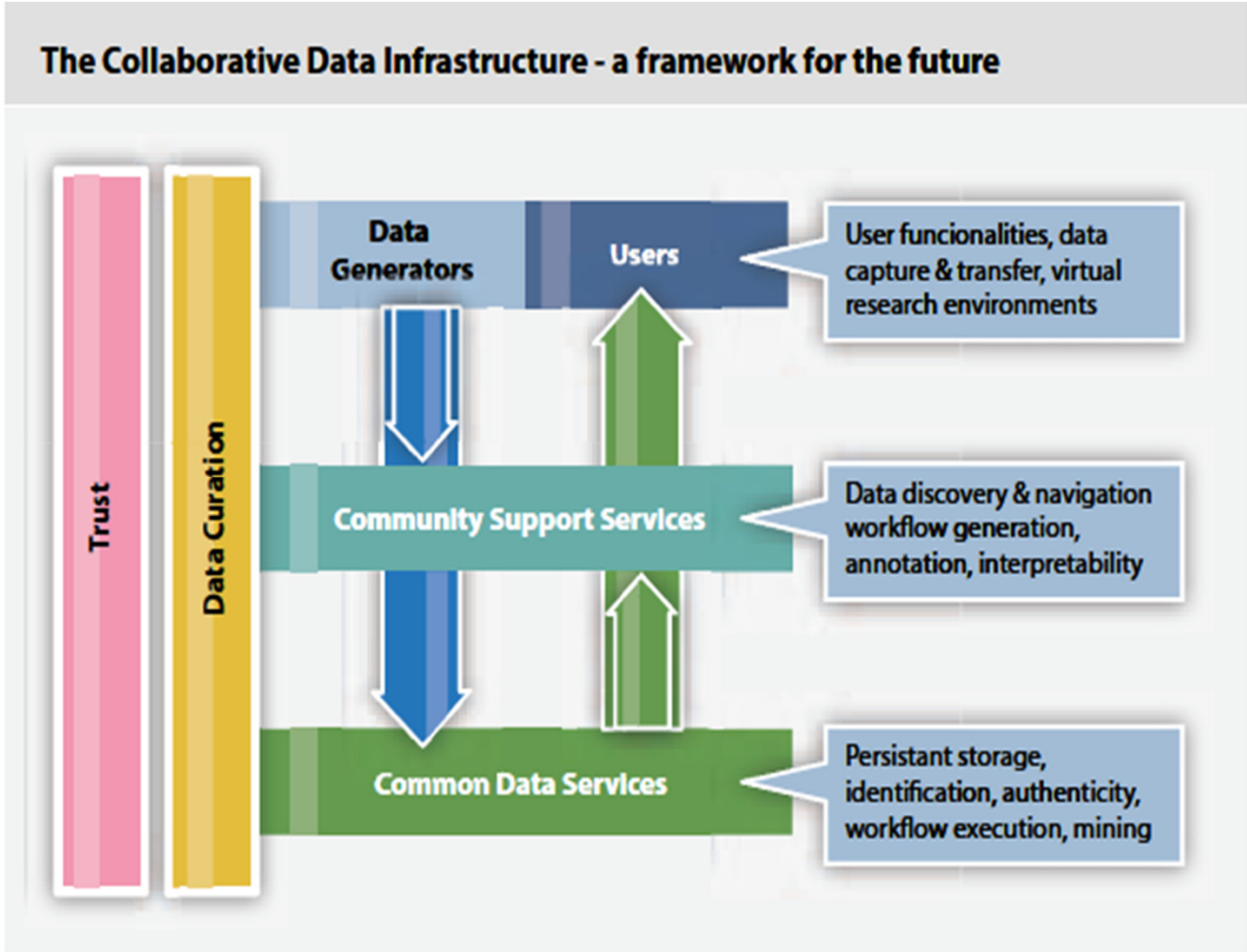


From theory to practice

- Models are one thing... what about the real world?
- It is not just about the costs, but also about the benefits
- Charging (y)our customers:
 - who? → data depositors or data users? Individuals or institutions?
 - how much? → fixed or variable costs?
 - for what? → storage? backups? services? metadata / data documentation?
 - for how long? → projects have temporary funding!

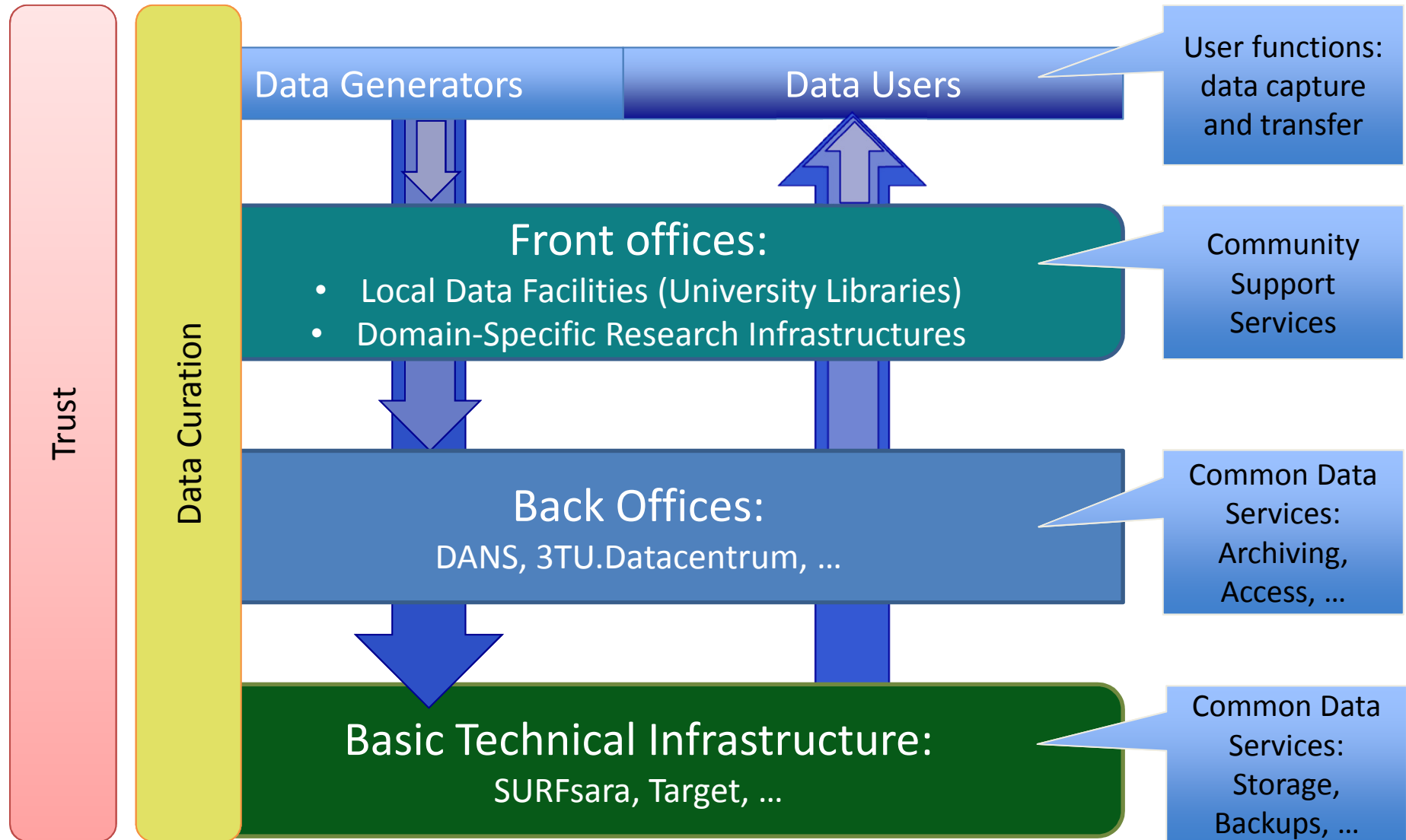


New approach of sharing costs and benefits of data preservation & access: policy context



Riding the Wave

The federated data infrastructure: a collaborative framework





Federated Data Infrastructure: policy considerations

- Data sharing gains importance
 - Funders demand OA to publicly funded research data, implying long-term preservation of these data in TDRs
 - Data fraud cases stimulated the development of data management policies in Dutch universities
- University libraries are looking for new roles in the digital age
- Federated model fits the scale of the country

<p>News</p> <p>nature International weekly journal of science</p> <h2>Report finds massive fraud at Dutch universities</h2> <p>Investigation claims dozens of social-psychology papers contain faked data.</p>	<p>PLoS one accelerating the publication of peer-reviewed science</p> <p>Home Browse Articles About For Readers For Authors and Reviewers</p> <p>RESEARCH ARTICLE OPEN ACCESS</p> <h3>How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data</h3>
--	---



Building a bridge Together



Supporting data curation through the front office / back office model

The data centers DANS and 3TU.Datacentrum aim to improve their services by creating a front office / back office model together with the universities. The recent cooperation between Leiden University Libraries and DANS and

3TU.Datacentrum provides the first use case of this new model. This case shows that by cooperating, we build an efficient and solid infrastructure that supports data sharing now and in the future.

We are complementary!

- Benefits for researchers**
 - + Researcher's **knowledge** at FO
 - + Researcher's **storage** of their data
 - + Researcher's **data management** in an early stage
- Benefits for front office (FO)**
 - + FO benefits from **data storage facilities** at BO
 - + FO benefits from BO's **knowledge** on data curation
 - + FO supports library's mission to be a **partner in research**
- Benefits for back office (BO)**
 - + BO benefits from **well trained FO contact persons** for researchers
 - + BO benefits from **direct contact** with their target group through FO
 - + BO **acquires more data** that are better prepared for archiving and reuse


 Universiteit Leiden
 Leiden University Libraries
 Leiden University Libraries
 Fleke Schouts
 f.schouts@library.leidenuniv.nl

3TU.Datacentrum
 3TU.Datacentrum
 Madelaine de Smaele
 m.m.a.desmaele@tudelft.nl


 Data Archiving and Networked Services
 DANS
 Ingrid Dillo
 ingrid.dillo@dans.knaw.nl



Roles and responsibilities (1): Front Office

- **Focus on information and awareness raising:**
 - Information portal research community
 - Awareness raising, support and training research community
 - Supporting VREs (research tools, data storage during research; Sharepoint, Dataverse, etc.; automatic ingest in TDR)
 - Liaising with back office
 - Data acquisition



Front Office



Roles and responsibilities (2): Back Office

Focus on expertise and long term storage:

- Providing expertise to the front offices: training courses for data librarians, consultancy, organize platform for exchange
- Providing specialist help and training courses to end users
- Expertise and innovation in the area of permanent storage, data management and re-use of data
- Hosting Dutch Dataverse?
- Long term preservation of data in a trusted digital repository



Back Office



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



DANS

Driven by data



Benefits for researchers

- + Researchers benefit from increased **data curation knowledge** at FO
- + Researchers have **better access** to sustainable storage of their data
- + Researchers **gain time** by engaging in data management in an early stage



Benefits for front office (FO)

- + FO benefits from **data storage facilities** at BO
- + FO benefits from BO's **knowledge** on data curation
- + FO supports library's mission to be a **partner in research**



Benefits for back office (BO)

- + BO benefits from **well trained FO contact persons** for researchers
- + BO benefits from **direct contact** with their target group through FO
- + BO **acquires more data** that are better prepared for archiving and reuse



Business model for implementing the FO-BO model

- Rolling out the model over all universities
- Creating technical infrastructure for automatic data ingest (building on Dutch Dataverse Network)
- Business model to cover the costs:
 - charging storage costs for institutional users (universities), archaeological companies, 'big' users (funded projects)
 - pay 5 times the annual fee and store "forever" (pension fund model)
 - pay for documentation costs in excess of one hour
 - pay for data management consultancy

DANS





Data Archiving and Networked Services

Thank you for your attention
and visit us at:

www.dans.knaw.nl

www.narcis.nl

<http://dataintelligence.3tu.nl/>



DANS is an institute of KNAW en NWO

Driven by data