

# Data Seal of Approval

Dissemination, assessment &  
procedures

in 30 slides

Dr. Henk Harmsen

Deputy director DANS

[henk.harmsen@dans.knaw.nl](mailto:henk.harmsen@dans.knaw.nl)

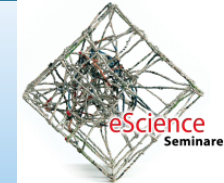
[www.dans.knaw.nl](http://www.dans.knaw.nl)





# MPG eScience Seminar 2009: Repository Systems

<http://www.datasealofapproval.org>



integrity

Omroep.nl | Nederland 1 | Nederland 2 | Nederland 3 | Radio 1 | Radio 2 | 3FM | Radio 4

Radio 747

omroep.nl de publieke omroep op internet  
over publieke omroep | contact | help

woensdag 24 mei 2006/08:55  
snel naar een omroepstele | kies een omroep

actueel **web** tv radio teletekst

**Het WK komt eraan!**  
De NOS WK-site staat online. Hier vind je informatie over het speelschema, de spelers, het laatste nieuws en een compleet video en audio-archief. Daarnaast kan je oefenen op vrije trappen en doelschieten. Misschien beland je zo zelf wel op het WK!  
WK-site

Creëer je eigen eilandengroep  
Westbroek gaat kamperen

**uitzending gemist**  
uitzendinggemist.nl

**Belgische Kortfilms**  
Een special met mooie korte films van onze zuiderburen  
bekijk uitzending

Journal  
Onderweg naar morgen  
De Lama's

**gids** tv radio

nu op tv | alle zenders | Complete gids

Ned1	08:40	Goedemorgen Nederland
Ned2	07:00	NOS-Journaal
Ned3	08:53	Tom en het aardbeienbammetje met honing
RTL 4	06:45	RTL Nieuws
RTL 5	06:30	Best of shopping
SBS 6	08:00	Lois & Clark: The new adventures of Superman
Net 5	08:25	Children's ward
RTL 7	08:00	RTL Nieuws

**live webkanalen**

Nederland4 | radiocast.nl

Webcastio | Webtv

**Prijsvraag**  
Bekijk het filmpje op Sterren.nl en beantwoord de vraag die Wolter Kroes daarin stelt. De prijs? Dat verklaart Wolter in het filmpje. Lees meer

**Weblog**  
Een tip: ga eens naar de kerk als je in alle rust naar een wedstrijd wilt kijken! Juichen

**themasites**

**NOS Journaal**  
Israël geeft Hamas een half jaar

- Referendum in Leeuwarden over plein
- SCP: gehandicapten willen meer hulp
- Nieuwsoverzicht

**NOS Sport**  
Miami wint eerste duel in Detroit

- Duits tennisteam klopt Argentinië
- EiffelTowers een zege van titel

**Wetenschap**  
22 mei: Nóg sneller warm

- 22 mei: Noorderlicht nieuwsbrief
- 22 mei: Wetenzap

**Geschiedenis**  
18 mei: "Little Richard: de gevestigde orde snakt naar adem"

- 18 mei: "Jerry Lee Lewis: omdat hij gemeen en onuitstaanbaar was"
- 18 mei: Geschiedenis op TV en radio

**Cinema.nl**  
Bombast en bordkartonnen personages

- Horrorplot met gaming-jus
- Huwelijk sneeft in stijlenbotsing

**3voor12**  
missAntarctica gooit na koud drie jaar handdoek in de ring

- Schrijver Fred de Vries was geen funpunk
- Malkovich schrijft ook Pophistorie

**Consument**

**zoek** vul een zoekwoord in | Ga

meest gezocht: uitzending gemist, spuiten en slikken, dit was het nieuws, songfestival, koefnoen, boer zoekt vrouw

vind uw website: Aangedaan, Aangespoeld, Adres onbekend, Afslag UMC, Ali G in de USAii, Ali in the Game

sorteer: alfabetisch, net/zend, omroep, rubrieker

**mijn regio** selecteer uw regio

**nieuwsbrief**  
vul uw emailadres in | abonneer mij

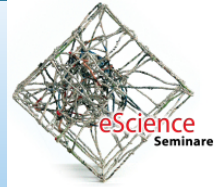
Meld u aan voor de wekelijkse Omroep.nl nieuwsbrief! | bekijk nieuwsbrief

**help**  
Naar de Help-pagina's  
Vragen over Uitzending Gemist  
Vragen over Omroep.nl/Gids

**instellingen**  
Lettergrootte: klein | middel | groot

**services**

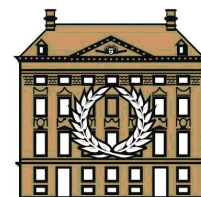
authenticity



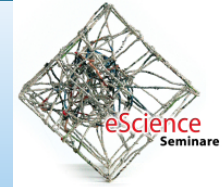
# Data Archiving & Networked Services

## DANS

- an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), also supported by the Netherlands Organization for Scientific Research (NWO).
- since its establishment in 2005, DANS has been taking care of storage and continuous accessibility of research data in the social sciences and humanities.

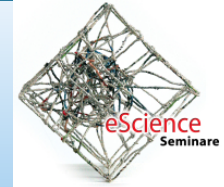






# TASK

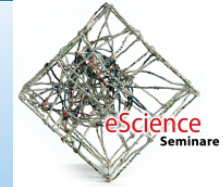
- Blueprint DANS 2004:
- Make Dutch data-archives '*DANS proof*'.  
*Archives will be judged on the base of 'DANS standards' and eventually they will be upgraded according these norms.*
- *If necessary or desirable DANS will take initiatives to make existing data-infrastructure 'DANS proof'*



# Quality criteria research data

## Research data

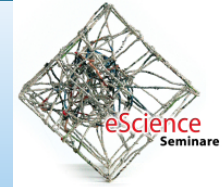
1. can be found on the internet
2. are accessible
3. are in a usable format available
4. are trustworthy
5. can be referred to



## Reinventing the wheel?

- *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive - NESTOR*
- *Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist* - Research Library Group (RLG) & OCLC/NARA- 84 criteria
- *Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)* - Digital Curation Centre (DCC) and Digital Preservation Europe (DPE)
- Digital Preservation Management (DPM) framework 2.0 ICPSR ...





## Back to the basics

DANS standards <-> Trusted Digital Repositories

DANS-proof <-> Data Seal of Approval

repository <-> stakeholders

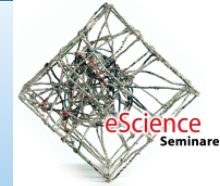
certification <-> guidelines

audit <-> self-assessment

control <-> trust

details <-> overview

central <-> user-community



# Stakeholders

- funding agencies
- data producers
- data consumers
- data repositories

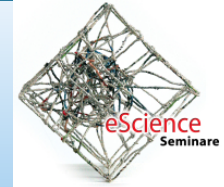






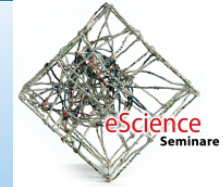
## Roles

- Funding agencies: *oblige seal of approval*
- Data producers: *produce excellent data*
- Data consumers: *use data in a trustworthy way*
- Data repositories: *responsible for long term preservation and distribution of research data*



# Responsibilities

- The *data producer* is responsible for the quality of the digital research data.
- The *data repository* is responsible for the quality of the storage and the availability of the data: data management.
- The *data consumer* is responsible for the quality of use of digital research data.

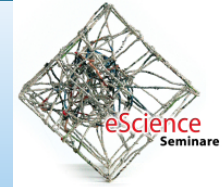


# The data producer (1-3)

## *Guidelines relating to the quality of digital research data*

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.

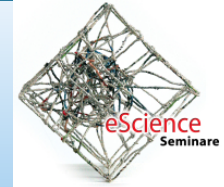




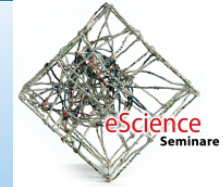
## The data repository (4-13)

### *Guidelines relating to the quality of the data repository*

4. The *data repository* has an **explicit mission** in the area of digital archiving and promulgates 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 to 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.

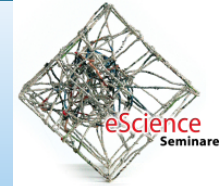


# The data consumer (14-16)

*Guidelines relating to the quality of the use of digital research data*

14. The *data consumer* must **comply 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 licences** of the data repository regarding the use of the research data.

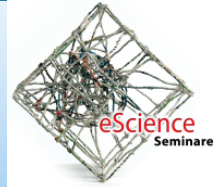




# Subcontracting

Indeed a repository, but not an archive, still in for the data seal of approval?

- Implement eleven guidelines: 1, 2, 3, 5, 9, 10, 11, 12, 14, 15 en 16.
- Let five guidelines (4, 6, 7, 8 and 13) be executed by another DSA archive.



## Subcontract: 4, 6, 7, 8 and 13

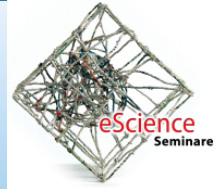
- 4 explicit mission
- 6 guarantee for quality management
- 7 long-term plan for durable archiving.
- 8 archiving according to predetermined criteria
- 13 technical infrastructure international standards compatible



© Original Artist  
Reproduction rights obtainable from  
[www.CartoonStock.com](http://www.CartoonStock.com)



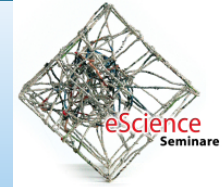
"I subcontracted the platypuses."



## Unique selling points (1)

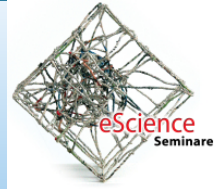
- DSA is oriented toward **research data**, not primarily toward publications.
- DSA not only pays attention to the archiving institution, but also to the data producer and the data consumer (**shared responsibility**).
- DSA is not in conflict with for example TRAC, but is rather **a step toward** it.
- DSA does not choose standardization but opts for '**trust**' (like custom of peer review in the scientific world).





## Unique selling points (2)

- DSA also focuses on **smaller organizations**.
- DSA is relatively **light** and therefore easy to implement.
- **Openness, dynamics** and **speed** are possible in the actual implementation.
- DSA is formulated as **points of attention, not as solutions**.
- DSA offers possibilities for **subcontracting archiving** and still meet the requirements of the DSA.



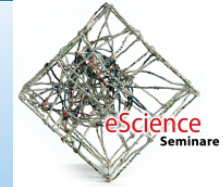
## Audit & Certification: don'ts and dos

### Problems:

- Complex
- Standardization / static
- Expensive
- Time-intensive
- Translucent / opaque

### Solutions:

- Simple
- Good practices / dynamic
- Peer-review / trust
- Light
- Transparent / open

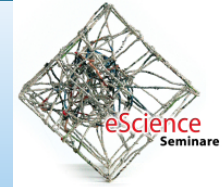


# Quality of the data repository

- First step to Awareness.
- **Data Seal of Approval Assessment (DSAA).**
- Building blocks: **keystone...**
- After the awareness: DRAMBORA, TRAC





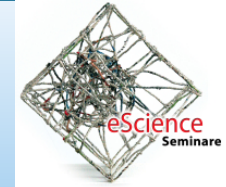


# Quality of the data repository

What does it mean:

## **Data Seal of Approval Assessment?**

- A document, which can be found on your repository page, where you describe how you implemented the **16** DSA guidelines
- An assignment of the DSA logo by the DSAA international board (DSAA-IB)



# An example of DSAA

## **11. The data repository ensures the integrity of the digital objects and the metadata**

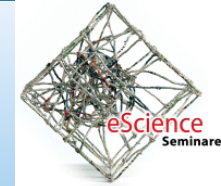
*This guideline relates to the information contained in the digital objects and metadata and whether it is complete, whether all changes are logged and whether intermediate versions are present in the archive.*

- Does the repository utilise checksums? What type? How are they monitored?
- How is the availability of data monitored?
- Does the repository deal with multiple versions of the data? If so, how?

## **12. The data repository ensures the authenticity of the digital objects and the metadata**

*This guideline refers to the relationship between the original data and that disseminated, and whether or not existing relationships between datasets and/or metadata are maintained.*

- What is the repository's strategy for changes? Are data producers made aware of this strategy?
- How is versioning handled?
- Does the repository maintain provenance data and related audit trails?
- Does the repository maintain links to metadata and other datasets, and if so how?
- How are the essential properties of different versions of the same file compared?
- Does the repository check the identities of depositors?

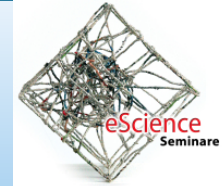


## Different solutions?

- LAMUS
- ORACLE
- File System
- FEDORA
- eSciDoc
- iRODS
- Zenity

...as long as they are transparent!

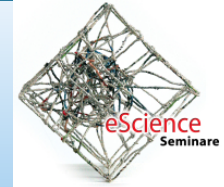




## DSAA – Editorial Board

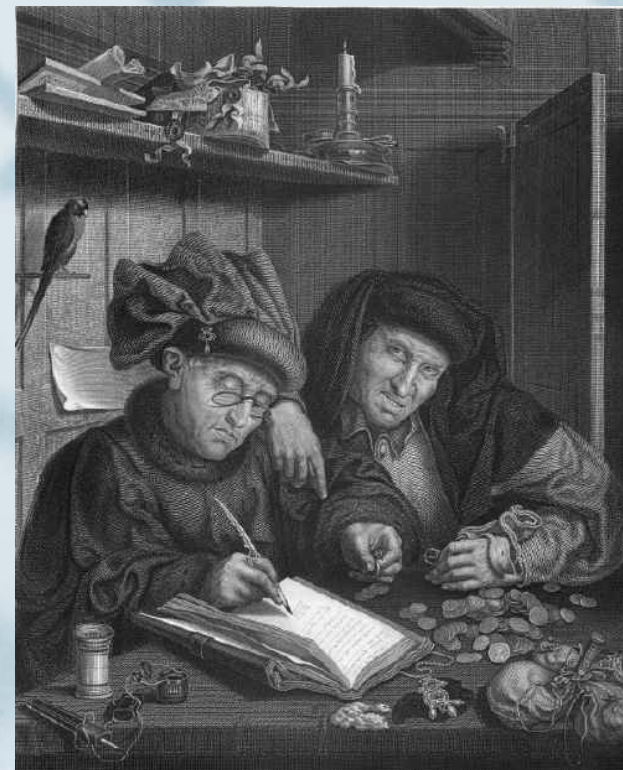
Henk Harmsen (NL, Chair), Paul Trilsbeek (NL, MPI/CLARIN), Natascha Schumann (D, DNB/NESTOR), Olivier Rouchon (F, CINES), Mary Vardigan (USA, ICPSR), Lisa de Leeuw (NL, Secretary), Hans Pfeiffenberger (D, AWI, IPYDLE), Matthew Woolard (UK, UKDA/CESSDA), Laurents Sesink (NL, DANS/DARIAH).



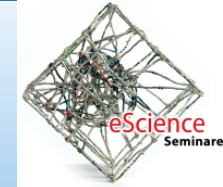


# The board as an editorial board

- Search, assess and teach new reviewers
- Bright expertise on different domains
- Responsible for the DSA guidelines
- Responsible for the DSAA guidelines and implementation
- Input specific 'infrastructural' requirements



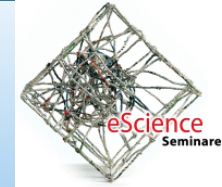




# Organisational roadmap

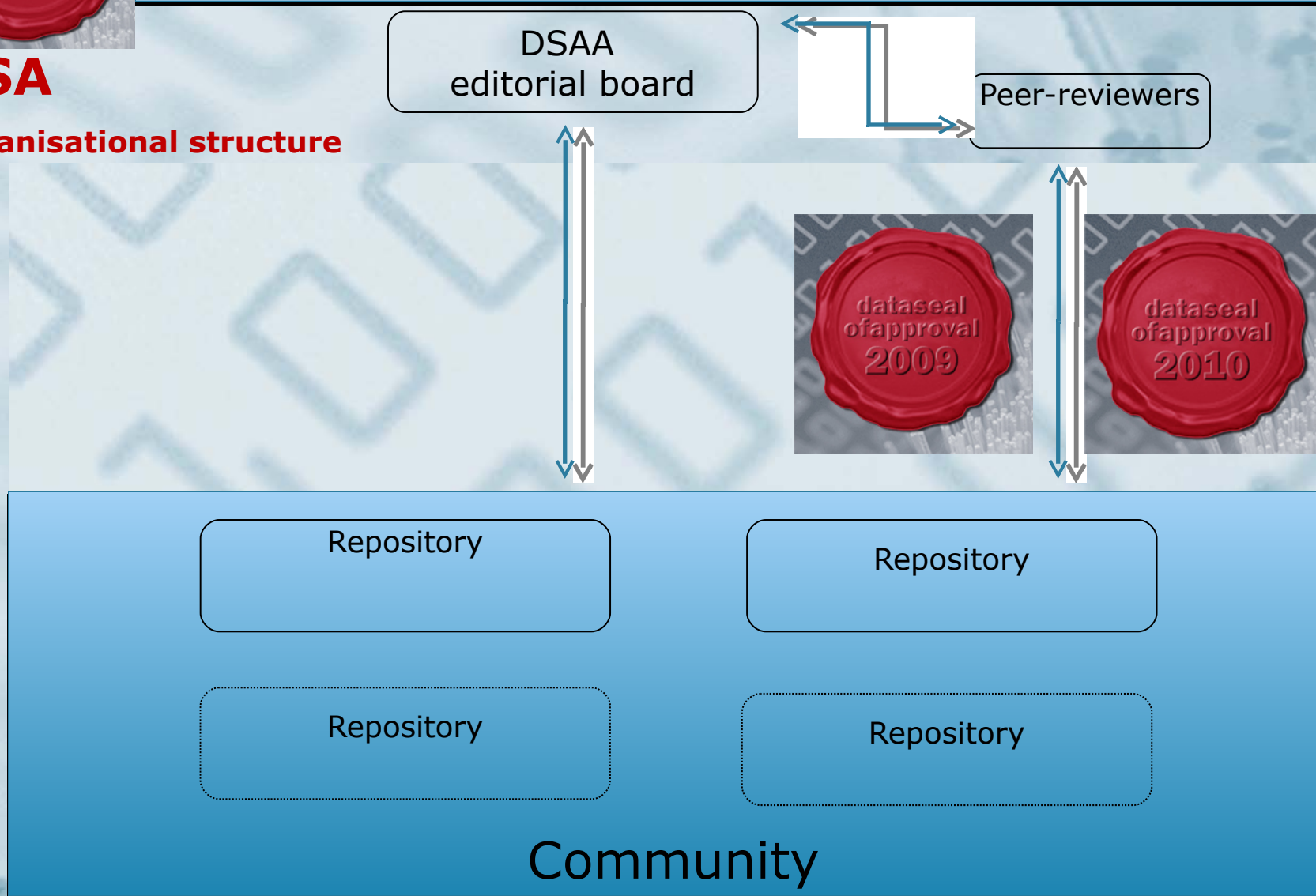
	<b>SUPERVISION</b>	<b>FUNDING</b>
2005	Data Archiving & Networked Services (DANS)	Data Archiving & Networked Services (DANS)
April 2009	Data Seal of Approval Assessment-Editorial Board (DSAA-EB)	Data Seal of Approval Assessment-Editorial Board (DSAA-EB)
2011	DSA Community (association)	EC and infrastructural projects (Clarin, Cessda, Dariah)

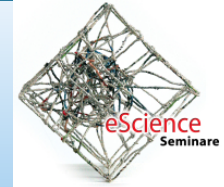




## DSA

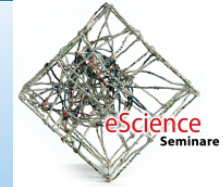
### Organisational structure





## Step-by-step (1) DONE!

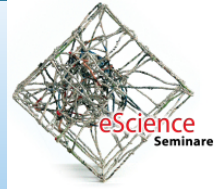
- European workshop May 2009 (hosted by the EC), domain map ready, handover DSA responsibility to the DSAA-IB (may 2009)
- Self-assessment of the repositories of the ISAA-EB members according to the **16** guidelines
- On the agenda of international projects:
  - Planets, DRIVER II (via SURF), DARIAH, CLARIN, CESSDA, Europeana, EC...
- Funding agencies
- Digital coalitions like DCC and NCDD



## Step-by-step (2) to do

- Independent external Assessment of the ISAA-EB
- Dialogue with Nestor, Drambora and TRAC in cooperation with the EC
- Assessments of other repositories:
  - 2009: 25
  - 2010: 50
- Set up network peer-reviewers
- Build a community
- Form an association





# Data Seal of Approval:

Thank you!

[www.datasealofapproval.org](http://www.datasealofapproval.org)