



# *Data Archiving and Networked Services*



**and  
the**



**Peter Doorn**

Data Archiving and Networked Services

EUDAT Conference

Trust & Certification Session

Amsterdam, 24-25 September 2014



## EUDAT Services & Support

- EUDAT's mission is to design, develop, implement and offer "Common Data Services" as they have been introduced in the "[Riding the Wave report](#)" to all interested researchers and research communities.
- These common data services:
  - must be relevant to several communities;
  - will be available at European level;
  - need to be characterised by a high degree of openness:

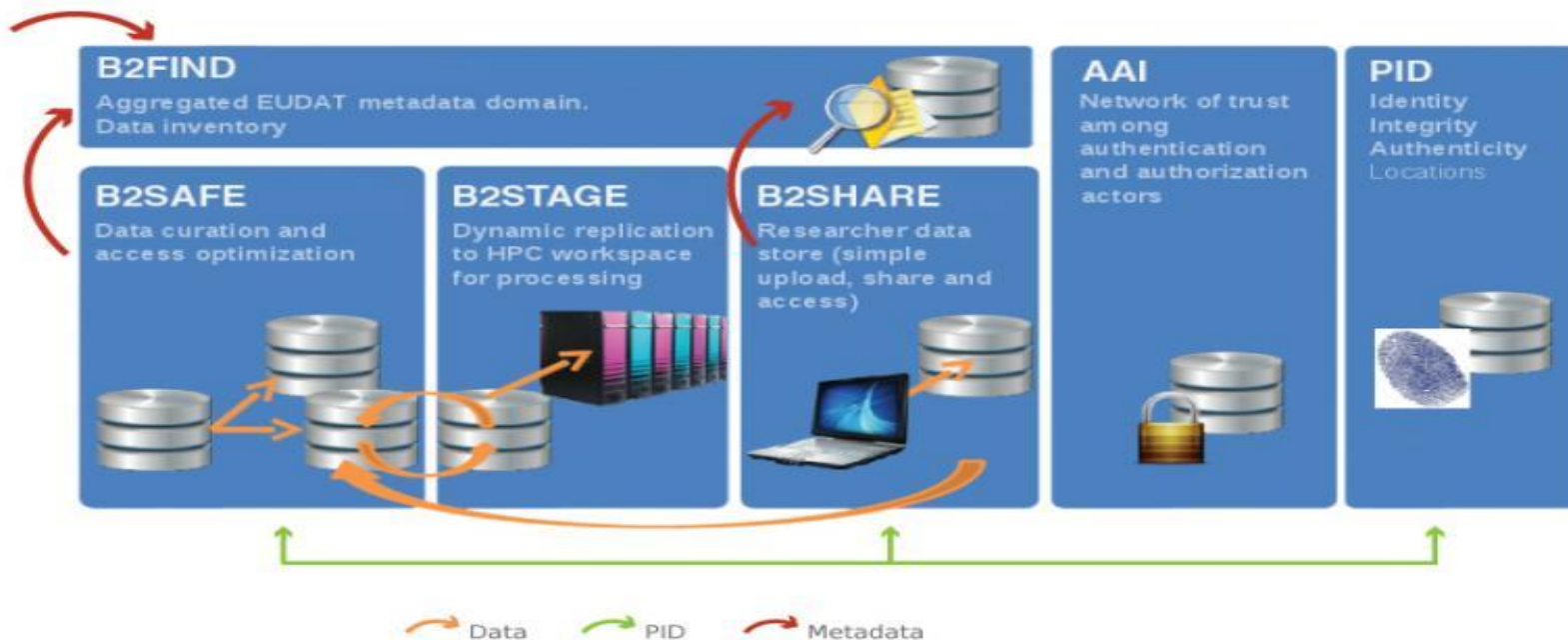
- (1) Open Access should be the default principle;

**And must be Trustworthy**

- (3) Flexible to allow new communities to be integrated



## Service Roadmap: Phase 1 Services (Oct. 2011 – Oct. 2013)



## Phase 2 Services (Dec. 2011 – Feb. 2014)

- (1) *B2DROP* to synchronize file systems with a central store
- (2) *B2NOTE* offering semantic referencing and annotation

## Phase 3 Services (Dec. 2011 – Feb. 2014)

data lifecycle support, support for dynamic data, workflow execution environments and additional semantic services



# Creating Trust

## Part of new proposal EUDAT2 on next phase of services:

More attention for non-technical work & “humanware”:

- data policies: management, preservation, deposit, access
  - training, support and advice
  - ***trust & certification***
- 
- Services do not run without people supporting them
  - Policies decide how services are run
  - “A policy is an assertion or assurance that is enforced about a collection or a dataset” (RDA Working Group on Practical Policies)
    - Computer actionable policies: enforce management, automate administrative tasks, validate assessment criteria, and automate scientific analyses...
    - Human actionable policies: data depositors, data users, data stewards...

Role for DANS



# Survey of 30 Institutions for Highest Priority Policies

Based on RDA Working Group: Practical  
Policy, Rainer Stotzka, Reagan Moore,  
Presentation for Plenary 3  
([https://www.rd-  
alliance.org/filedepot?cid=104&fid=466](https://www.rd-alliance.org/filedepot?cid=104&fid=466))

Policy	Importance
Integrity	217
Preservation	150
Access control	126
Provenance	108
Data Management plans	99
Publication	75
Replication	66
Data staging	52
Federation	37
Metadata sharing	23
Regulatory	16
Collection properties	7
Identifiers	7
Data sharing	7
Versioning	7
Licensing	6
Format	6
Data Life Cycle	6
Arrangement	5
Processing	5



# Trust in research data

- Trust is at the very heart of storing and sharing data
- Trust involves:
  - Data creators
  - Data users
  - Data repositories
  - Funders



Trust comes on foot, but leaves on horseback



# What is trust built on?

- Dedicate yourself (mission statement)
- Do what you promise (stable, sincere and competent reputation)
- Be transparent (peer review, get certified)



"They don't trust each other to share research."





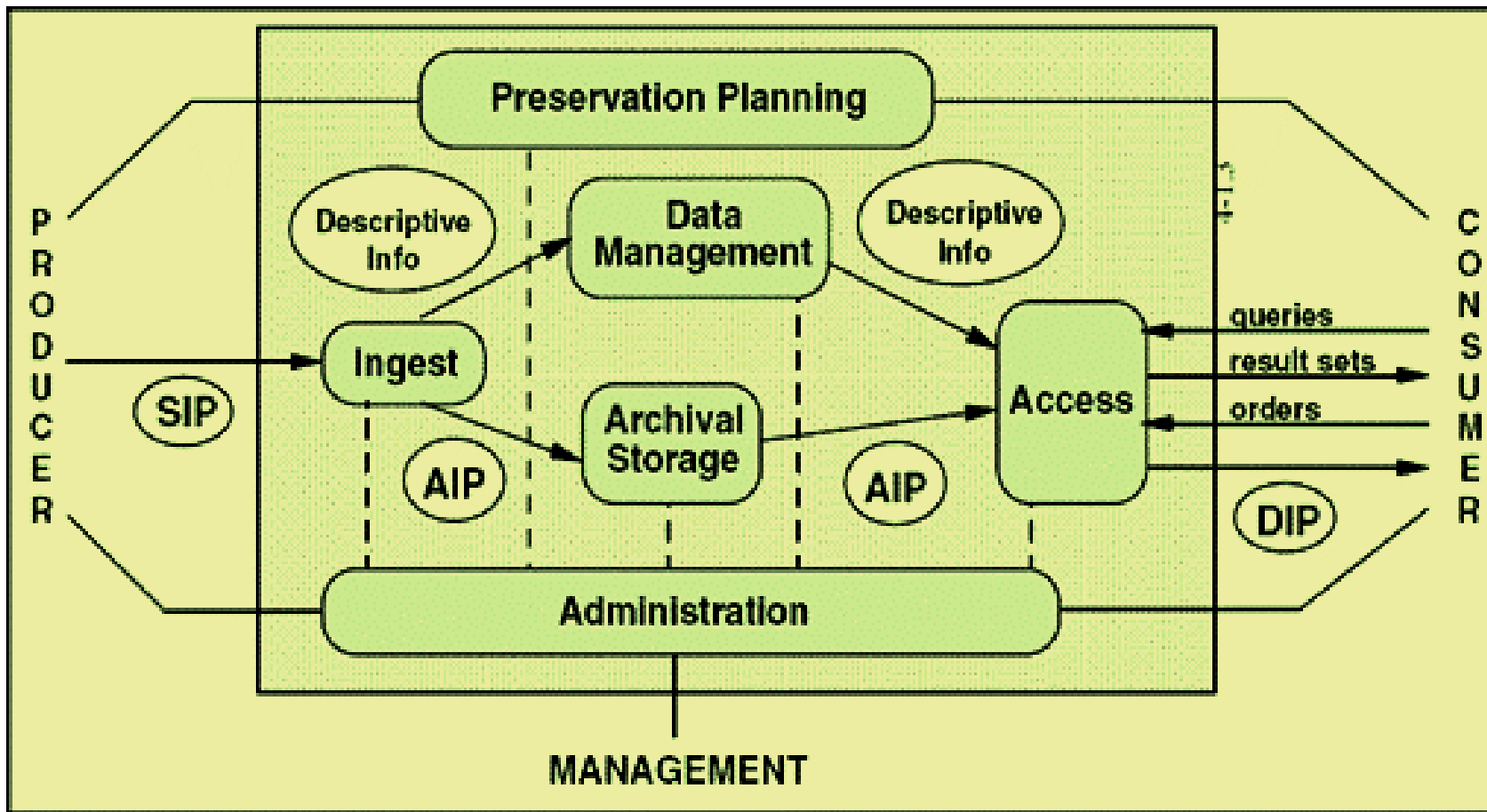
# The need for trusted digital repositories

- In 1996 the [Consultative Committee for Space Data Systems](#) established a task force that developed [OAIS](#) (Open Archival Information System, accepted as [ISO 14721](#) in 2002), a high-level model for the operation of digital archives.
- Independent auditing was deemed necessary to certify OAIS-compliance and thus engender trust.
- Development of OAIS auditing metrics began in 2003 and resulted in Trustworthy Repositories Audit & Certification (TRAC 2007).
- TRAC is the basis of the Trusted Digital Repository (TDR) document that was accepted as ISO 16363 in 2012.

[http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=56510](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56510)



# OAIS – Open Archival Information System



SIP = Submission Information Package

AIP = Archival Information Package

DIP = Dissemination Information Package



# ESFRI Research Infrastructures and Trust



**DARIAH**  
Digital Research Infrastructure  
for the Arts and Humanities

## Compliance as a DARIAH Trusted Digital Repository

The reliability and trustworthiness of digital preservation programmes are of utmost importance to all users, providers and managers of digital data. The preservation of

Organisations which join DARIAH or begin engaging in repository type activities will be expected to meet certain minimum standards. This will enable DARIAH to vet organisations and recommend practices to be followed. All organisations should meet policy requirements in order to be counted as Trusted.



**CESSDA**  
social science data

metadata    accessing data  
data harmonisation    ethics  
data catalogue    confidentiality  
data preservation

**CLARIN ERIC**  
Common Language Resources and Technology Infrastructure



## Requirements for CLARIN Centres

“Centres need to have a proper and clearly specified repository system and participate in a quality assessment procedure as proposed by the Data Seal of Approval or MOIMS-RAC approaches”

## Building Trust: CESSDA Self-Assessment Project

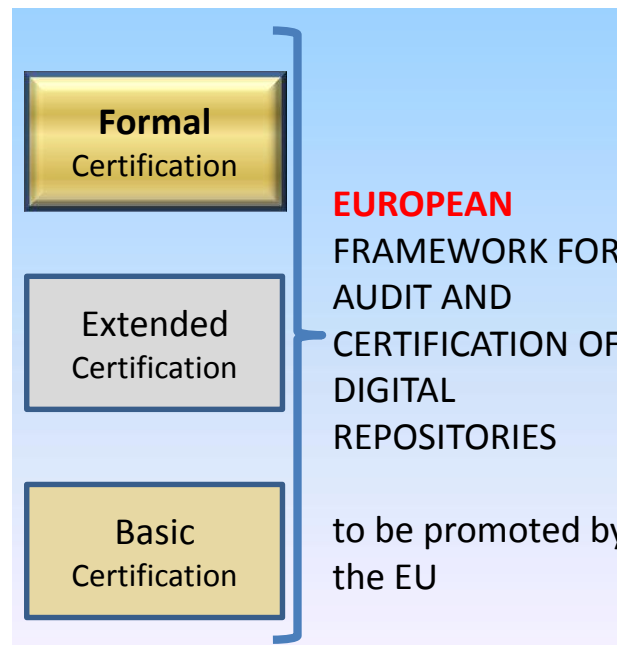
Participants from fifteen CESSDA member organisations discussed the CESSDA-ERIC requirements and agreed upon using the Data Seal of Approval (DSA) guidelines as a tool to gain information on the level of their conformance with the DSA and the CESSDA-ERIC requirements.



# Certification of digital repositories



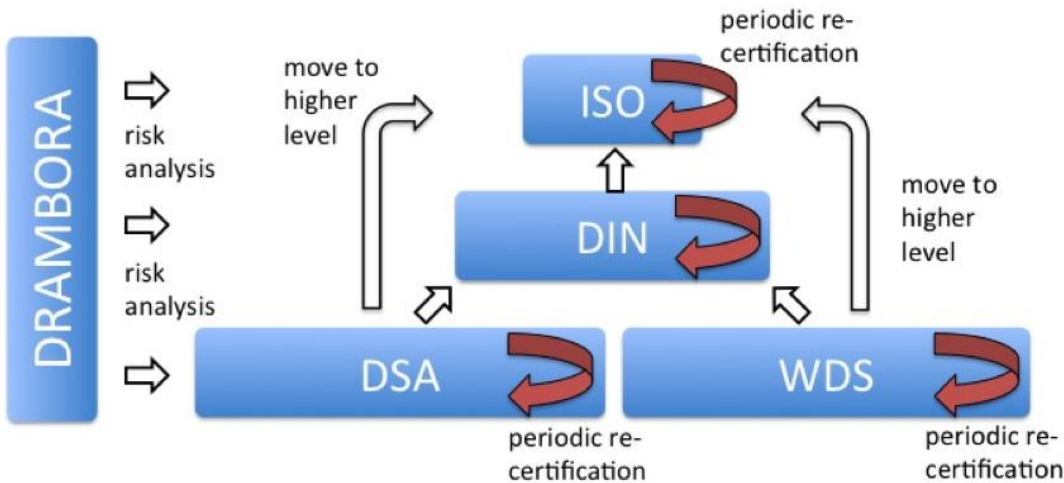
- International framework
- 3 standards
- 3 levels (basic, extended, formal)





# CTRUST: Certification of digital repositories

Digital Repository Audit Method Based On Risk Assessment



- 10 partners, many associated partners
- Uniting all important players & initiatives

- International framework
- 4 standards, 3 levels (basic, extended, formal)



WORLD DATA SYSTEM



*Driven by data*



Criterion/Metric	DSA	WDS	nestor/Seal/DIN 31644	ISO 16363
1	-	X	-	-
2	X	X	X	X
3	X	-	X	X
4	-	-	X	X
5	-	-	-	X
.???	.???	.???	.???	.???
Certification costs	free	membership	€500	variable fee
No. of metrics	16	17	34	109
Time needed (estimate)*	2 weeks	2 weeks	1,5-3 months	3-6 months
Focus/accents/ orientation	smaller repositories, humanities and social sciences	life/earth/ spatial sciences	German national/state libraries, archives, universities	bigger/ national /business/ science heritage archives



# Certification Standards: Data Seal of Approval (DSA)

- DANS initiative (2005/6)
- International Board
- 16 guidelines
- Self assessment
- Transparency
- 35 seals awarded since 2010



Data producers are responsible for the quality of research data, repositories for storage and long-term access, and users for correct use of data

The research data:

- can be **found** on the Internet
- are **accessible** (clear rights and licenses)
- are in a usable format (**i**nteroperable)
- are **reliable**
- can be **referred to** (persistent identifier)



# On-going work on Trust

- Work Package on “Trust” within APARSEN project:  
<http://www.alliancepermanentaccess.org/>
- Research Data Alliance Interest Group on Certification:  
<https://www2.rd-alliance.org/internal-groups/rdawds-certification-digital-repositories-ig.html>



*Driven by data*

**Thank you for your attention**

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

[www.narcis.nl](http://www.narcis.nl)

[peter.doorn@dans.knaw.nl](mailto:peter.doorn@dans.knaw.nl)

