USING THE DSA TO BENCHMARK AND GUIDE TRUST WITHIN CESSDA

HERVÉ L’HOURS
PRESERVATION PLANNING MANAGER
UK DATA ARCHIVE
UNIVERSITY OF ESSEX

IASSIST 2014
5 June 2014
THE UK DATA ARCHIVE

We acquire, curate and provide access to the UK's largest collection of social and economic data.
THE UK DATA SERVICE

• quality social research data

• a unified point of access to data from ESDS, Census Programme, Secure Data Service and others
CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES

CESSDA
CESSDA

- European infrastructure
- informal organisation since 1976
- now a legal entity hosted in Bergen, Norway
- 13 members:
  Austria, Czech Republic, Denmark, Finland, France, Germany, Lithuania, Netherlands, Norway, Slovenia, Sweden, Switzerland, United Kingdom
- each with a designated Service Provider (SP)
Immediate goals of the process were to:

- address alignment with CESSDA Statutes
- consider the roles of:
  - trust
  - Trusted Digital Repositories status
CESSDA: Trust Relationships

Producer -> Repository 1 -> Consumer

Traditional Data Lifecycle
CESSDA: Trust Relationships

[Diagram showing the relationship between Consumer, CESSDA Portal, Producer, Repository 1, and Consumer. The diagram illustrates the traditional data lifecycle.]
Statutes signed by all Service Providers:

- Data Documentation Initiative (DDI) metadata compliance
- common single sign-on user authentication system
- harvesting of all resource discovery metadata
- data downloadable through common data gateways;
- maintain local language(s) within the multi-lingual thesaurus
CESSDA: OBLIGATIONS

Statutes signed by all Service Providers:

- OAIS compliance and trusted digital repository status
- cross-national data harmonisation activities
- cross-national question bank
- mentor support for:
  - Associate Members
  - countries with immature and fragile national infrastructures
- Common Data Access and Dissemination Policy
What is a TDR?

- mission to provide reliable, long-term access to managed digital resources to its designated community, now and into the future
- constant monitoring, planning, and maintenance
- understand threats to and risks within its systems
- regular cycle of audit and/or certification
AUDIT AND CERTIFICATION OF TRUSTED DIGITAL REPOSITORIES

Data Seal of Approval DSA
- international community consortium
- 16 guidelines, trust-based self-assessment and peer-review

DIN 31644: Criteria for Trustworthy Digital Archives
- NESTOR German Competence Network for Digital Preservation (Working Group on Certification)
- 34 criteria

ISO 16363: Audit and Certification of Trustworthy Digital Repositories
- successor to RAC/TRAC
- over 100 metrics, full external audit process
THE CERTIFICATION FRAMEWORK

Memorandum of Understanding for a Three-tiered Framework

- **Formal certification**: DSA + full external audit and certification based on ISO 16363* or DIN 31644**
- **Extended certification**: DSA + structured, externally reviewed and publicly available self-audit based on ISO 16363* or DIN 31644**
- **Basic certification**: Data Seal of Approval (DSA)

Information and Documentation - Criteria for Trustworthy Digital Archives

http://www.trusteddigitalrepository.eu

*ISO 16363 - Audit and Certification of Trustworthy Digital Repositories
**DIN 31644 - Information and Documentation - Criteria for Trustworthy Digital Archives
Trust: “Reliance on and confidence in the truth, worth, reliability of a person or thing”

Trust is transitive

Transparency: “Minimum degree of disclosure to which agreements, dealings, practices, and transactions are open to all for verification”

Community: The Data Seal of Approval was established by a number of institutions committed to the long-term archiving of data. By assigning the seal, the DSA community seeks to guarantee the durability of the data concerned, but also to promote the goal of durable archiving in general.
The Data Seal of Approval is granted to repositories that are committed to archiving and providing access to data in a sustainable way.

- **Data Producers**
  - *Assurance* of reliable data storage

- **Funding Bodies**
  - *Confidence* that data is available for reuse

- **Data Consumers**
  - *Enables* assessment of repositories
The data are:

- available on the Internet
- accessible
  - while taking into account relevant legislation with regard to personal information and intellectual property of the data.
- usable (file formats)
- reliable
- citable (can be referred to)
## Compliance Levels

<table>
<thead>
<tr>
<th>Level</th>
<th>Compliance Level Definition</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Not Applicable</td>
<td>Provide an explanation</td>
</tr>
<tr>
<td>1</td>
<td>We have not considered this yet</td>
<td>Provide an explanation</td>
</tr>
<tr>
<td>2</td>
<td>We have a theoretical concept</td>
<td>Provide a URL for the initiation document.</td>
</tr>
<tr>
<td>3</td>
<td>We are in the implementation phase.</td>
<td>Provide a URL for the definition document.</td>
</tr>
<tr>
<td>4</td>
<td>This guideline has been fully implemented for the needs of our repository</td>
<td>Provide a URL for the definition document.</td>
</tr>
</tbody>
</table>
Data producers

- deposit quality and compliance
- recommended formats
- requested metadata
Data repositories

- explicit mission
- legal/contract compliance
- data storage processes
- preservation plan
- explicit data lifecycle workflows
- responsibility for access
Data repositories

- discovery, use and reference
- object/metadata integrity
- object/metadata authenticity
- technical infrastructure standards
Data consumers

- compliance with access criteria
- proper use and exchange of information
- usage licence compliance
CESSDA TRUST PROCESS: METHODOLOGY

- mapping obligations to DSA Guidelines (where possible)
- initial workshop
  - understanding of trust requirements and the DSA
- conformity assessment
  - self-assessment, peer review and gap analysis
- follow-on workshop:
  - results of process
  - consider next actions for obligations and TDR status
Informal application of the DSA process, no formal DSA Board involvement and no DSA assigned

- Service Providers conduct self-assessment
- anonymous peer-review
- analysis and repeat grading of assessments by ‘Expert Group’
  
  Mari Kleemola, FSD
  Herve L’Hours, UK Data Service
  Natascha Schumann, GESIS
  Heiko Tjalsma, DANS

- Oversight from Matthew Woollard, Director, UK Data Archive and UK Data Service
CESSDA TRUST PROCESS: METHODOLOGY

- generic feedback provided to all Service Providers
- gap analysis of trust requirements
- all anonymous
  - though including feedback to DSA Board
A (partial) success:

- DSA a valuable basis for open information sharing and discussion but…
- challenges when integrating granular, specific obligation into higher level DSA Guidelines
- discussion naturally divided between the two areas
- likely to become less of an issue as service providers become more familiar with the DSA
CESSDA TRUST PROCESS: **OUTCOMES**

- useful (but separate) discussion on meeting obligations
- good overall conformance across CESSDA
- issues with providing evidence for internal repository processes
  - we all focus on the externally facing (producer/consumer) evidence
CESSDA TRUST PROCESS: ACTIONS

DSA as a basis for future CESSDA Trust status:

• aligned CESSDA evidence (easier to maintain and compare)
• initial confidentiality within CESSDA
• a good basis for communications around trust issues for members

Issues to consider:

• adopting a version
• changes to the criteria beyond CESSDA control
• staggered adoption
• influencing the criteria
CESSDA TRUST PROCESS: ACTIONS

Common approach to:

- designated community
- formal identification of depositors
- acceptable formats and format risk
- alignment with DDI and deposit metadata
- mission statements
- registry of legislation
- licence harmonisation
- common data storage descriptions
- versioning
- lifecycle/business process descriptions
- critical preservation events to be captured
Communities of Trust

CESSDA Portal

Consumers

Repository 1
Repository 2
Repository 3
Repository N

Producers
## Compliance Levels

<table>
<thead>
<tr>
<th>Level</th>
<th>Compliance Level Definition</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Not Applicable</td>
<td>Provide an explanation</td>
</tr>
<tr>
<td>1</td>
<td>We have not considered this yet</td>
<td>Provide an explanation</td>
</tr>
<tr>
<td>2</td>
<td>We have a theoretical concept</td>
<td>Provide a URL for the initiation document.</td>
</tr>
<tr>
<td>3</td>
<td>We are in the implementation phase.</td>
<td>Provide a URL for the definition document.</td>
</tr>
<tr>
<td>4</td>
<td>This guideline has been fully implemented for the needs of our repository</td>
<td>Provide a URL for the definition document.</td>
</tr>
</tbody>
</table>