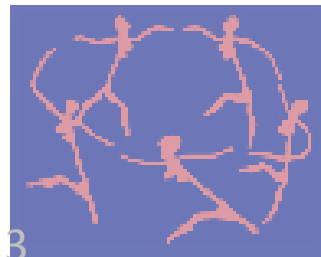


DASISH

Digital Services Infrastructure for Social Sciences and Humanities

Daan Broeder

TLA - MPI for Psycholinguistics





DASISH Origin

FP7 Capacities Work Programme: Infrastructures

INFRA-2011-2.3.1: Implementation of common solutions for a cluster of ESFRI infrastructures in the field of "Social Sciences and Humanities".

A project under this topic should implement harmonised solutions for the ESFRI Infrastructures in the field of Social Science and Humanities on issues like, for example *metadata frameworks, registries, single-sign-on systems and permanent identifiers.*

- 18 partners from 10 EU countries + Norway
- 5 ESFRI infrastructures: CESSDA, CLARIN, DARIAH, ESS and SHARE
- DASISH budget: 6ME -> 700PMs
- Started January 2012

- Already existing collaborations
 - Many partners are part of multiple research infrastructures
 - ESFRI projects merging e.g.
 - In AT common CLARIN & DARIAH project
 - In NL new CLARIN + DARIAH -> CLARIAH proposal

- **DASISH brings together 5 ESFRI infrastructures by focusing on common activities across disciplines and infrastructures**
- **DASISH aims to provide solutions to common problems that will also strengthen international collaboration**

CESSDA Council of European Social Science Data Archives

CLARIN Common Language Resources and Technology Infrastructure

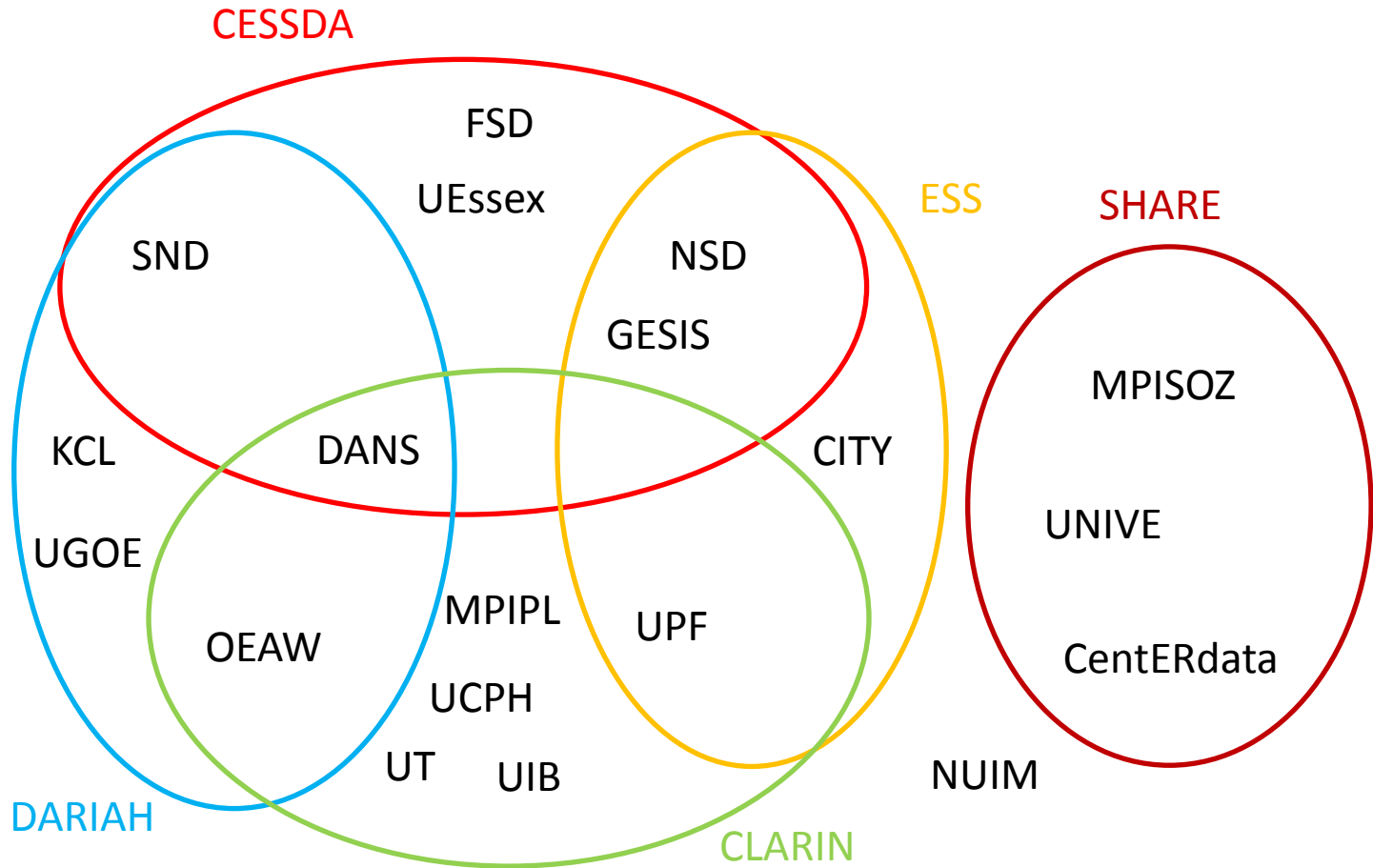
DARIAH Digital Research Infrastructure for the Arts and Humanities

ESS European Social Survey

SHARE Survey of Health, Ageing and Retirement in Europe

- SHARE, CLARIN and CESSDA have now established ERICs or equivalent organizations
- CESSDA, CLARIN and DARIAH are centre network oriented
- ESS and SHARE are (from the data management perspective) single centre
- So a natural split in DASISH activities between:
 - SHARE + ESS very much interested in improving quality of data and data acquisition tools
 - CLARIN, DARIAH and CESSDA interested in optimizing data management & archiving

Consortium II



- CRISP, ENVRI, BioMed, DASISH
- It is about sharing services, finding common solutions
- Clusters have no proper sustainability
- They are dependent on participating research infrastructures & ERICs
- Cluster project role in infra services: Investigate, recommend, brokering
 - Commercial service
 - share an ESFRI project specific one
 - Use a generic EU infrastructure service (EUDAT)
 - ...



DASISH Challenges

- Need to create common infrastructure not just strengthen community specific ones
- Traditions vary considerably
 - Between SS at one side and the Humanities. But also within the Humanities
 - Some collaborations/communities have a long & rich history
 - Others are fairly new (as an infrastructure)
 - Organizational models and complexity varies and impacts preferences for solutions
 - Differences wrt. understanding IT issues
- Language and terminology vary even more so as past discussions learned us

- PID: investigate requirements and service offerings
- Discovery: common metadata catalogue
- Metadata quality recommendations
-
- AAI Goal: Trust federation allowing SSO and single identity
 - Many users -> maintaining separate user store considered prohibitive
 - Limited complexity for the user, e.g. no user certificate manipulation

Perceived AAI solutions (at that time)

- *Federated Identity Management offered by the national IDFs and use of SAML2 should be sufficient if ...*
- *EU inter-federation eduGain works*
- *All home-organizations join and the correct user attributes are released*

- **CESSDA:**
 - CESSDA old existing collaboration, already AAI experiments with different ‘federated’ AAI systems
 - CESSDA SSO project allowing access over two IDFs to CESSDA portal implementing a VOSP solution for extra user attributes
- **CLARIN:**
 - Operational multi-IDF multi-SP federation
 - CLARIN homeless IdP
 - Runs own service provider federation
- **DARIAH:**
 - youngest one with plans and ongoing preparations for FIM, unsure yet of organization type
 - Robust homeless IdP solution

DASISH DOW: Create an SSH trust federation for SSO and Single Identity based on SAML2 ...

- using & collaborating with NRENs, Terena/eduGain if a workable solution exists
 - Opt-in policy per IdP is the biggest problem
 - CLARIN experiences are here indicative
 - Still when solved it is the solution of choice
- alternative is creating a proper SSH federation along the lines of CLARIN
- ... or expansion of CLARIN SPF



DASISH requirements

Problems will remain:

- Insufficient release of attributes
 - Need for flexible community solutions augmenting user attribute sets
- Need for Identity stores for the homeless
 - Community managed?
 - ... or one large for EU research?
- If we want general application of FIM also for sensitive data we will need multiple levels of assurance

Ongoing work

- Investigation of the CLARIN experience and setup is important
 - Status of attribute release by EU home organizations
 - Acceptability of the CLARIN SPF agreement for others
 - centers from other ESFRI projects
 - ESFRI project federation solutions should not be exclusive
 - some centers participate in multiple ESFRI projects and would naturally belong to multiple federations
 - Possible basis for a new SSH federation
- Keep monitoring what is the status of eduGain efforts
- Be aware of danger of incoherent messages to the home-organization
 - Linkup to CLARIN SPF
 - Linkup to eduGain
 - Linkup to SSH fed

Thank you for your attention

1	Management	UGOT		44
2	Architecture and Quality assessment	DANS	<ul style="list-style-type: none"> • Liaise with other e-infra initiatives • Get requirements for a ref. architecture • Assess results 	83
3	Data Quality	MPG-MEA	Improve EU wide survey quality: terminology, translation, vocabulary normalizations	200
4	Archiving	NSD	<ul style="list-style-type: none"> • State of preservation in SSH • Assessment of deposit services recommendations, negotiations. • Deposit service convergence 	67
5	Data Access and Enrichment	MPG-PL	<ul style="list-style-type: none"> • Federated Identity • PID requirements • Metadata quality improvement • Joint metadata domain • Workflow use cases • Annotation framework 	171
6	Legal and Ethical Issues	MPG-MEA	<ul style="list-style-type: none"> • identify legal and ethical issues wrt. current and new SSH data types resulting from the integration, linking and archiving • Legal & ethical VCC 	68
7	Education and Training	UGOE	Training modules, workshop program	56
8	Dissemination	UCPH	Communication strategy and means	34