DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

User Requirements for PID Service Providers: Survey Results from DASISH WP 5.2

Catharina Wasner
GESIS – Leibniz Institute for the Social Sciences

DASISH Workshop on Persistent Identifiers
December 8-9
Cologne

SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Overview

1. Introduction: DASISH
2. DASISH WP 5.2
3. Results from the survey on the community level
4. Results from the survey on the data centre level
5. Conlusion: DASISH Requirements for PID Services

SEVENTH FRAMEWORK PROGRAMME

DASISH – Data Service Infrastructure for the Social Sciences and Humanities

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

DASISH

- DASISH is a cluster project for 5 ESFRI (European Strategy Forum on Research Infrastructures) Roadmap infrastructures in SSH – Social Sciences and Humanities
- 18 partners in 13 European countries
- The project has received the financial support of the European Union's Seventh Framework Programme
- It started in January 2012 and ends in December 2014
- It focusses on common activities across disciplines and infrastructures

SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Common Solutions to Common Problems

Goal:

enhance visibility and re-usability of digital resources, tools and services

Areas of possible synergies:

- data quality
- data archiving
- data access
- legal and ethics aspects
- educational activities

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

DASISH WP 5.2: Robust PID Service

- From the DoW:
  "The output of this task will result in reports that may lead to upgraded PID services fulfilling the DASISH requirements and that will foster the use of PIDs.
  The procedure will be as follows:
  - (1) DASISH will make a comparison of three relevant PID services (DataCite, EPIC, PersID).
  - (2) DASISH will work out requirements for such PID services based on the current usage and plans in the participating communities.
  - (3) A report will summarize the results of this work and DASISH will create PR documents and education material that help promoting the use of PID services and that may lead to adaptations of the existing services.
  - (4) Training courses will be offered that describe how to register and use PIDs."
- Partners:
  - SND – Swedish National Data Service: task coordination
  - GESIS – Leibniz Institute for the Social Sciences
  - DANS – Data Archiving and Networked Services
  - MPI-PL – Max Planck Institute for Psycholinguistics
  - UGOE – Göttingen State and University Library
  - UCPH – Centre for Language Technology University of Copenhagen

SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

DASISH WP 5.2: Robust PID Service

- From the DoW:
  "The output of this task will result in reports that may lead to upgraded PID services fulfilling the DASISH requirements and that will foster the use of PIDs.
  The procedure will be as follows:
  - (1) DASISH will make a comparison of three relevant PID services (DataCite, EPIC, PersID).
  - **(2) DASISH will work out requirements for such PID services based on the current usage and plans in the participating communities.**
  - (3) A report will summarize the results of this work and DASISH will create PR documents and education material that help promoting the use of PID services and that may lead to adaptations of the existing services.
  - (4) Training courses will be offered that describe how to register and use PIDs."
- Partners:
  - SND – Swedish National Data Service: task coordination
  - GESIS – Leibniz Institute for the Social Sciences
  - DANS – Data Archiving and Networked Services
  - MPI-PL – Max Planck Institute for Psycholinguistics
  - UGOE – Göttingen State and University Library
  - UCPH – Centre for Language Technology University of Copenhagen

SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

DASISH Requirements for PID Service Providers

- Definition PID Service Provider:
  - service that provides at the very least PID registration and resolving
  - relevant within DASISH are e.g. DataCite, EPIC, national URN:NBN services/URN:NBN Cluster project
- Do requirements differ between the DASISH communities?
- Is it possible to arrive at a general widely accepted list of requirements?
- Survey in Nov/Dec 2012 conducted within DASISH
  1. on the level of communities
  2. on the level of data centres

SEVENTH FRAMEWORK PROGRAMME

DASISH DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Questionnaire for the communities, Question 1/3: Are there any policies/routines for PID services at the community level?

- CESSDA no
- CLARIN yes
- DARIAH no
- ESS no
- SHARE no (but in progress)

SEVENTH FRAMEWORK PROGRAMME

DASISH DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Questionnaire for the communities, Question 2/3:
Approximate how many/what percentage of the data centres in the community use PID services.

- CLARIN: ~82%
  (mostly used Handles via EPIC; own Handles, URN:NBN)
- CESSDA: ~29%
  (mostly used DOIs; URN:NBN)
- DARIAH: not specified
  (ARK, DOI, Handle, Crossref, EPIC, and a PID used by the SUDOC reference registry)

SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Questionnaire for the communities, Question 3/3: Are there any additional requirements on the PID services at the community level?

- In 2009 CLARIN published requirements for PID services.
- The other communities have not produced such a document yet.

SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Responses from the data centres

The questionnaire was sent to all CESSDA and CLARIN members and 10 DARIAH members who were rated as archives.

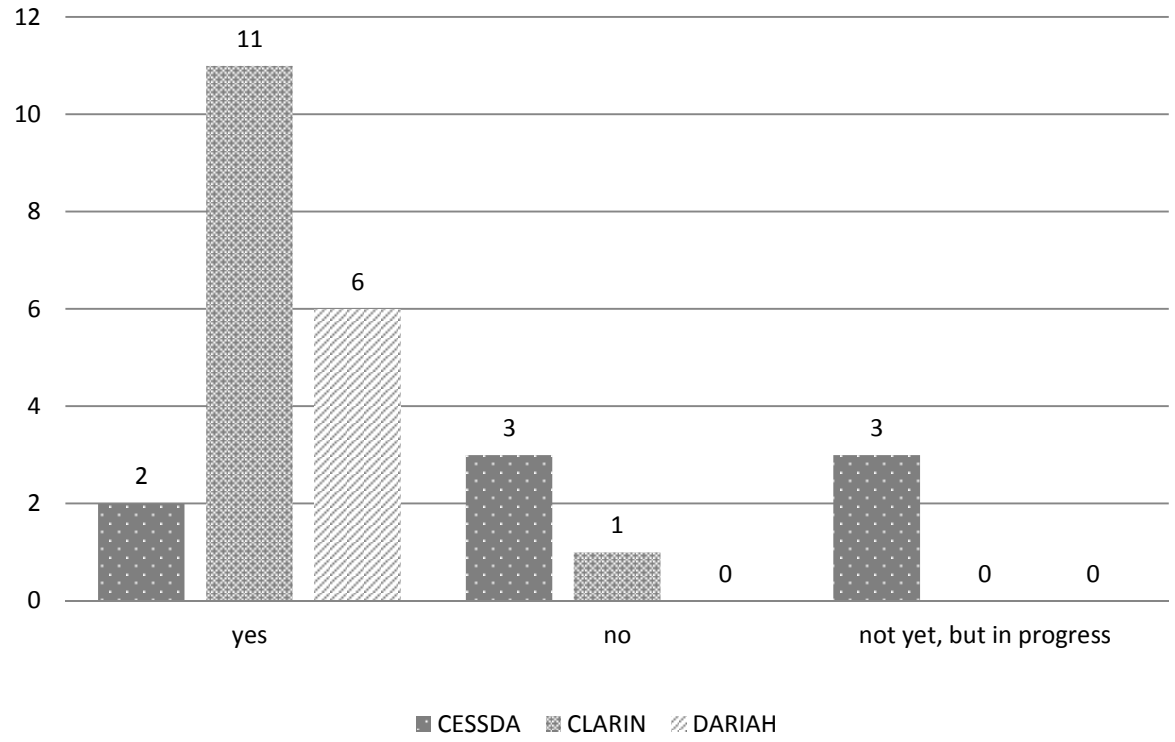
- CESSDA 8/21
- CLARIN 12/20
- DARIAH 6/10 (67)

SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Questionnaire for the data centres, Question 1/7: Are there any policies and/or routines for the use of PID services at the data centre?

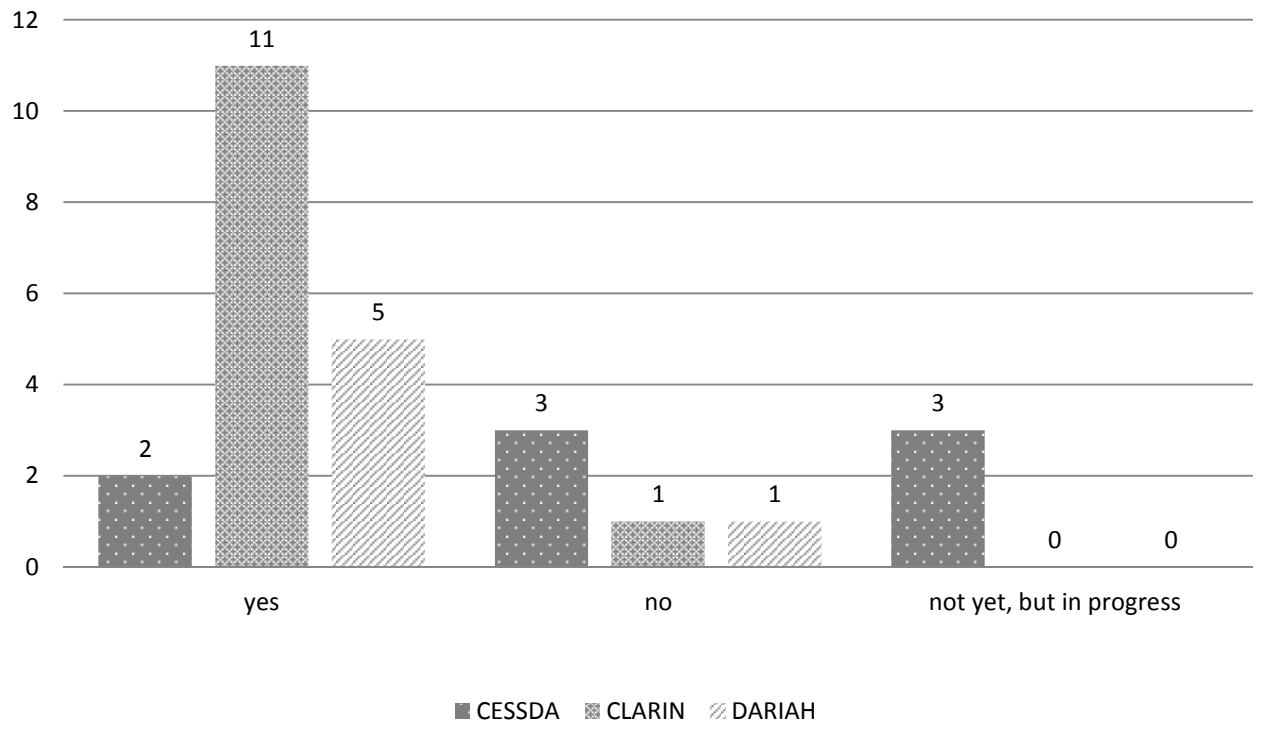


SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Questionnaire for the data centres, Question 2/7:
Does the data centre use PID services?



SEVENTH FRAMEWORK PROGRAMME

DASISH

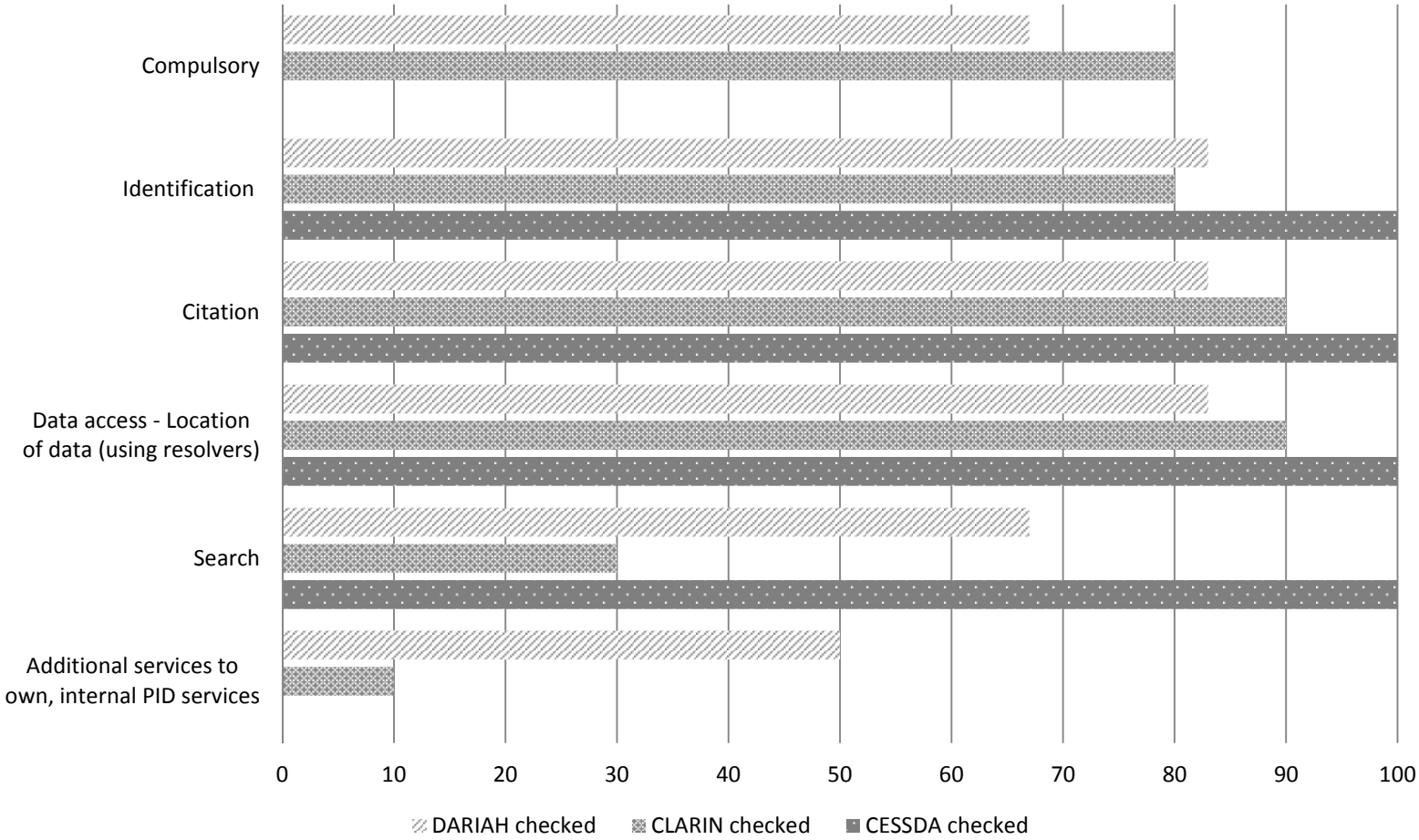
DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Questionnaire for the data centres, Question 3/7:

If the data center uses PID services: why does the data center use PID services? What purposes and expectations/requirements does the center have?

SEVENTH FRAMEWORK PROGRAMME

Distribution of the checked requirements amongst the communities (in %)



DASISH DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Questionnaire for the data centres, Question 4/7: If the data center does not use PID services: what is/are the main reason/-s?

a) ***If interested in using PID services; are there any obvious obstacles that prevent the data centre from using PID services***

   => 4 answers: lack of resources, waiting for emerging solutions and recommendations

b) ***If not interested in using PID services, why?***

   => All evaluated centres are interested in using PID services.

SEVENTH FRAMEWORK PROGRAMME

DASISH DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Questionnaire for the data centres, Question 5/7:
If the data center does not use PID services: what would it take for the data center to start using PID services?

=> 4 answers:

- Extended Services from the PID service providers: 2/4
- Education and training: 3/4
- Targeted/direct contact: 1/4
- Other specifications: centre needs some additional funding

SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Questionnaire for the data centres, Question 6/7:
What are the pros and cons with those PID Services that are currently recommended/used by the community/center?

Topics:

- Non-commercial services
- Relating metadata
- Availability of the PID service/dependence on the service provider
- Identification of fragments
- Different PIDs: need to limit the number

SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Questionnaire for the data centres, Question 7/7:
Other comments/requirements regarding PID services?

- Need of stable and reliable longterm national/international service providers to maintain the system/resolver
- Granularity, version managment
- Multiple PIDs: Make them interoperable
- Strong user platform
- Business model should be controlled by the scientific community
- Need for a meta-resolver

SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

1. DASISH requirements for a basic PID service

1. A PID registration and resolution service infrastructure has to be available under the responsibility of a reliable and long-term funded organisation, operating at a European or national scale.
2. It maintains the systems and offers services and support and is embedded in a European/national network.
3. It has a clear policy that describes the responsibilities of the different stakeholders.
4. It offers a minimum set of descriptive metadata: e.g. title, author, publisher, publication year, rights, PID.
5. The resolution technology has to be reliable, fast, persistent and scalable and is 24/7 available.
6. The business model has to be sustainable for all involved stakeholders. It is controlled by the scientific community.
7. It is available to centers and users EU wide.

SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

2. DASISH requirements for an extended PID service

1. The PID syntax and resolution mechanism of the PID service must accept the usage of version and fragment identifiers. The PID service provides support for the version and fragment management.
2. The PID service supports the traceability of research and efforts with regard to link literature, data and authors.
3. It provides different representations/formats of metadata associated with PIDs (content negotiation), and can ideally be assigned to authors and organisations.
4. The rights of an individual PID is owned by the author/organisation that produced the object to which the PID has been assigned.

SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

3. DASISH requirements for extra services

1. Education and Service for the data centres regarding PIDs in general are needed.
2. PID services within the ESFRIs have to be interoperable. Users should not be confronted with the PID diversity. To be a able to resolve all types of PIDs there should be a meta-resolver service that allows users to enter any type of PID and resolve it.

SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

DASISH WP 5.2 Deliverable

- Deliverable: D5.1a
- Deliverable Name: PID Services Report
- Responsible: Birger Jerlehag, UGOT
- Work Package Leader: Daan Broeder, MPG-TLA
- Contributing Partners and Editors: Timo Gnadt, UGOE; Arjan Hogenar, DANS; Bart Jongejan, UCPH; Merja Karjalainen, UGOT; Przemyslaw Lenkiewicz, MPG-TLA; Jens Ludwig, UGOE; Catharina Wasner, GESIS
- http://dasish.eu/publications/projectreports/DASISH-D5_1_BS_version141106.pdf

SEVENTH FRAMEWORK PROGRAMME

DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Thank you!

contact

Catharina Wasner

GESIS – Leibniz Institute for the Social Sciences

catharina.wasner@gesis.org

SEVENTH FRAMEWORK PROGRAMME