Handles in EUDAT

Willem Elbers

The Language Archive

MPI for Psycholinguistics, Netherlands

SEVENTH FRAMEWORK
PROGRAMME

Tools
Projects

The
Language
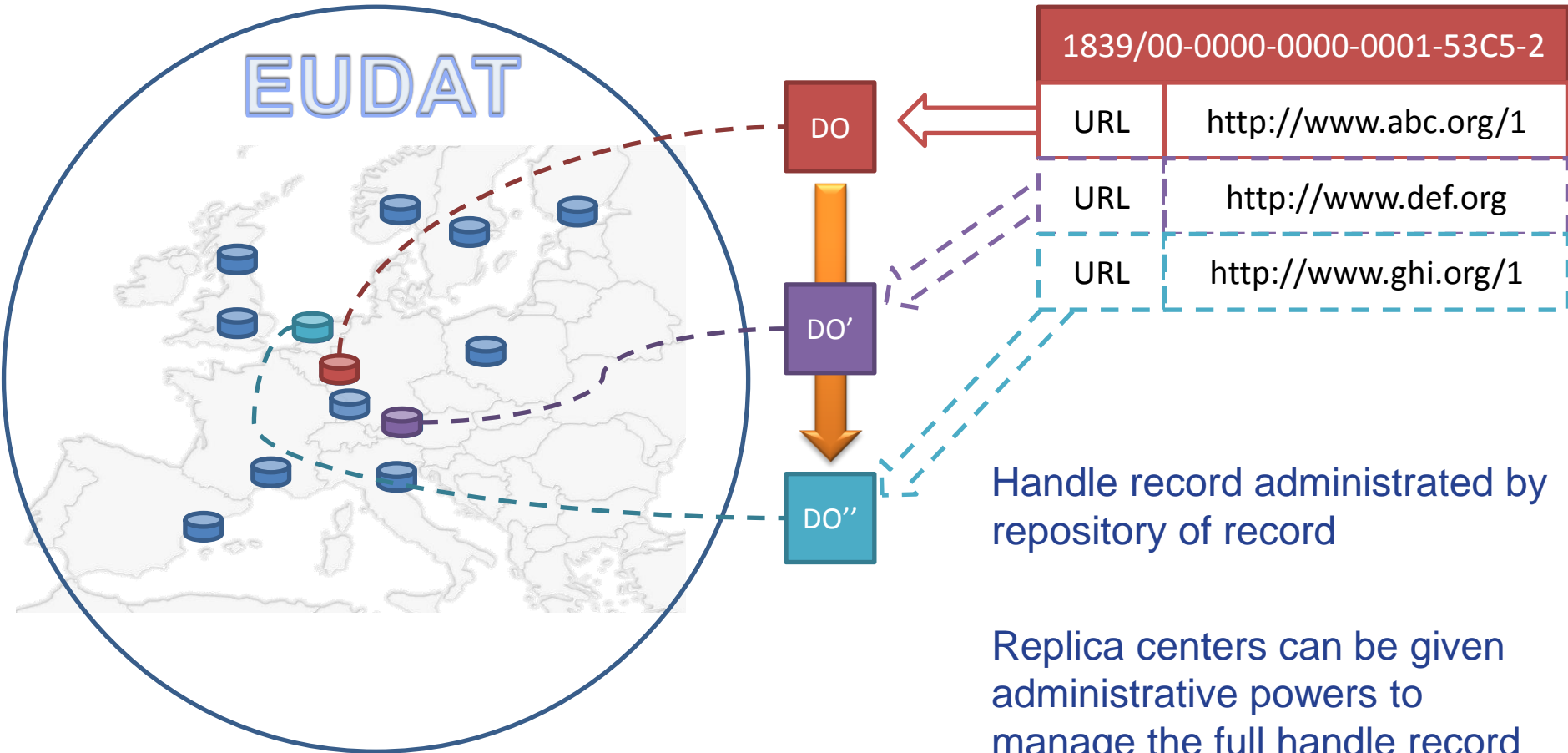Archive

Language
Data

Experts
Collaboration

EUDAT

Outline

- EUDAT PID requirements
- The problem area explained
- EUDAT solution
- EPIC client script
- iRODS integration
- Conclusion

Requirements

- The EUDAT data domain handles registered data
  - Each digital object must have a persistent identifier

- This persistent identifier is used for:
  - Replica identification
  - Identification of the repository of record
  - Querying of additional information
    - Checksum (time stamped)

- Actionable URLs

- Handle system used within EUDAT

B2SAFE workflow

1839/00-0000-0000-0001-53C5-2	
URL	http://www.abc.org/1
URL	http://www.def.org
URL	http://www.ghi.org/1

DO

DO'

DO''

Handle record administrated by repository of record

Replica centers can be given administrative powers to manage the full handle record → Alternative needed!

EUDAT Handle Record Administration

Each handle record is managed in a separate administrative domain!

EUDAT Handle records

- **RoR record (www.example1.com)**

123/abc			
	URL	1	http://www.example1.com/1
	URL	2	http://hdl.handle.net/**456/abc**
	CHECKSUM	10	3801e6832b73e0fc4d39b3630712cd33
	XYZ	N	…

- **Replica record (www.example2.com)**

123/abc			
	URL	1	http://www.example2.com/xyz/1
	URL	2	http://hdl.handle.net/789/abc
	ROR	3	http://hdl.handle.net/**123/abc**
	CHECKSUM	10	3801e6832b73e0fc4d39b3630712cd33
	XYZ	n	…

Demo

- Repository of record handle:
    - 1839/00-0000-0000-0000-0690-E

- Replica records:
    - 11096/0004f3e2-c18e-11e2-8120-00163ef6845e
    - 11112/7ae8c49e-c297-11e2-a94a-14feb57d1251

Pros and Cons

- Pros
  - Flexible
  - Direct link from any replica handle record to the RoR handle record
  - Works out of the box when resolving handles via the http proxy (10320/loc field)

- Cons
  - Finding all replicas is an expensive operation (walk the tree)
  - PIDs for replicas are publicly resolvable! It is up to the community to decide how to use them
  - Software required to be aware of this approach if you want to do more than just resolving

EUDAT

Joining the EUDAT CDI

- EUDAT using the EPIC handle service within the CDI
  - EUDAT centers are required to use EPIC

- Communities joining the EUDAT CDI have three options:
  - Register a handle prefix and let EPIC operate the handle server
  - Register a handle prefix and operate the handle server
    - Must take care of availability!
  - Use a different PID technology and take all responsibility for operating it

EPIC

- EUDAT is using the epic v2 api hosted at SurfSara
  - https://epic.sara.nl/v2/handles/

- A handle prefix must be registered with CNRI before you can use this service

- The EPIC API offers a HTTP-REST interface
  - Search handles with GET request
  - Update handles with UPDATE request
  - Create handles with POST request
  - The API supports both text and json content types

iRODS integration

- EUDAT offers integration of the EPIC API into iRODS via a python script.
  - This comes out of the box with the B2SAFE service

- The script takes your credentials as input
  - Supplied on the command line
  - Stored in a configuration file (iRODS or local fs)

- The script supports the following actions
  - Searching
  - Resolving
  - Creation
  - Modification

Facts

- Currently over 8 million registered handles for EUDAT in total

Prefix	Organization	#handles
10916	SurfSara	32
11095	*MPI-Pl*	*0*
11096	RZG	72586
11097	Juelich	30782
11100	CINECA	7641425
11112	SurfSara	50761
11113	CSC	27289
11140	SNIC	8217

Prefix	Organization	#handles
11161	EPCC	79
11202	Sigma	6
11099	*INGV*	*7*
11101	BSC	0
11114	*LifeWatch*	*0*
11137	CINES	91
11225	PSNC	30
11230	*KNMI*	*0*

Conclusion

- Approach in production for B2SAFE
  - Administrative issues are solved
  - flexible enough to incorporate the integrity checking policies

- Improvements considered to the linked structure to overcome replication discovery performance issues

- Ongoing work to use the linked structure to present
  - Overview of a communities replicas
  - Overview of a replica health status

Questions

Thank you for your attention

iRODS integration

- The script is stored in the iRODS command directory

- Can be invoked via the iRODS execute command rule

10320/loc field

Use EPIC to run the handle server

Operate your own handle server