

EPIC services, an overview

2nd Eudat conference, 2013

Ulrich Schwardmann

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
(GWGD)

Am Fassberg, 37077 Göttingen
ulrich.schwardmann [at] gwgd.de

30 October 2013, Rome



EPIC

Ulrich
Schwardmann

- is dedicated to provide persistent identifier (PID) services
- main scope is European scientific and cultural heritage communities
- is at the moment a consortium of four major European scientific computing centers
CSC, DKRZ, GWGD, SURF-SARA

EPIC –
Consortium

EPIC Services

Infrastructure
Types
DOI integration
Support

PIDs 4
eResearch

all partners

- have a solid backing of national funding authorities
- have long experience in providing reliable, safe and secure services
- have technical sustainability
- have a structure that can act, if necessary, as a company
- have the ability to provide SLAs
- are involved in several big eScience projects
- have signed a MoU to provide a PId system for the scientific community
- are looking for more partners to strengthen the scope and reliability



EPIC

Ulrich
Schwardmann

EPIC –
Consortium

EPIC Services

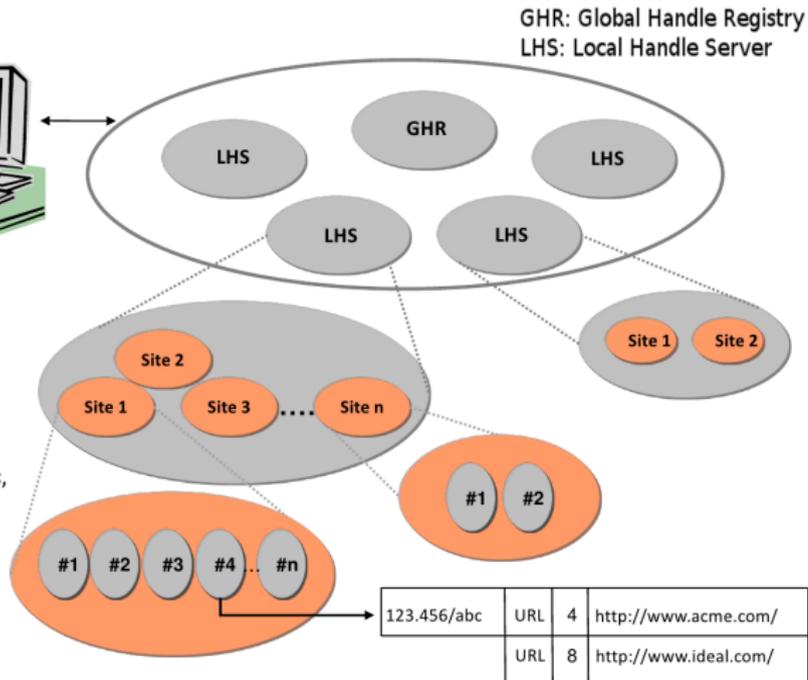
Infrastructure
Types
DOI integration
Support

PIDs 4
eResearch

Handle Resolution



The Handle System is a collection of handle services, each of which consists of one or more replicated sites, each of which may have one or more servers.



EPIC

Ulrich
Schwardmann

EPIC –
Consortium

EPIC Services

Infrastructure
Types
DOI integration
Support

PIDs 4
eResearch

- Highly redundant resolution system
 - LHS (local handle server) resolves for one prefix
 - Secondary LHS is mirror of Primary
 - Answer given by the fastest LHS
- Single point of failure: GHR?
 - GHR is a cluster of servers
 - Secondary GHR at GWWDG and in China
- Handle proxies for resolution of PIDs



EPIC

Ulrich
Schwardmann

EPIC –
Consortium

EPIC Services

Infrastructure
Types
DOI integration
Support

PIDs 4
eResearch

- API service
 - the service for minting PIDs
 - the partners run API services for several prefixes
 - different profiles and policies for minting PIDs can be given on a prefix level
- resolutions service
 - each partner provides a resolver (primary LHS) for the prefixes on service
 - each partner mirrors primary LHS of the others



EPIC

Ulrich
Schwardmann

EPIC –
Consortium

EPIC Services

Infrastructure
Types
DOI integration
Support

PIDs 4
eResearch

- global handle registry
 - database of the LHS
 - Primary GHR at CNRI in USA
 - mirrors running at GWWDG on behalf of EPIC and one in China
- handle resolution proxy
 - first proxy outside USA was recently implemented at GWWDG
 - part of a resolution network
- monitoring of these services

Handle Protocol Services



EPIC

Ulrich
Schwardmann

- Types
- Templates

EPIC –
Consortium

EPIC Services

Infrastructure

Types

DOI integration

Support

PIDs 4
eResearch



EPIC

Ulrich
Schwardmann

EPIC –
Consortium

EPIC Services

Infrastructure
Types
DOI integration
Support

PIDs 4
eResearch

- what are types?
- types are *parameters* associated with a PID
- which types should be provided? =====
 - discussion in RDA WG *PID info types*
- minimal necessary information about PID needed:
 - **URL**
 - pointer to the resource
 - **PID owner**
 - responsible for consistency
 - to contact in case of discrepancies
 - **expiration date**
 - determines timeline of curation of the PID
 - not timeline of PID or its resolution
 - to allow removal of data object

Types Provided in API version 1



EPIC

Ulrich
Schwardmann

EPIC –
Consortium

EPIC Services

Infrastructure
Types
DOI integration
Support

PIDs 4
eResearch

- mandatory
 - URL
 - PID owner
 - author, title, creator
 - publication and expiration date
- not mandatory
 - meta data URL
 - checksum (MD5,SHA-1)
 - file size

- depends on the community context
- types and the services to make use of it shouldn't be dependent on complex technologies
- Example: authorization
 - whether authorization is necessary to access digital object makes sense
 - specific authorization information not useful: needs (complex) authorization technology

usefull additional types



EPIC

Ulrich
Schwardmann

EPIC –
Consortium

EPIC Services

Infrastructure
Types
DOI integration
Support

PIDs 4
eResearch

- mandatory:
 - **URL,**
 - **PID owner,**
 - **expiration date**
- technical:
 - **publication date,**
 - **checksum,**
 - file size,
 - accessibility (authorization)
- content:
 - **meta data URL,**
 - author, title, creator,
 - previous version

- rationale: templates are used to adapt PIDs to the *correct* granularity
 - *correct* has to be defined by the use case and community
- templates are used to instrument **services** on data objects also together with the handle **resolution**
- the template implementation in the handle system is simply a rewrite rule
- delimiter and replacement is configurable on a prefix level
- example (delimiter is: @):
 - 11858/00-ZZZZ-0000-0001-CCD1-4@aaa=bbb&ccc=ddd
- translates into:
 - <http://wwwuser.gwdg.de/~tkalman/downloads/formtest.php?aaa=bbb&ccc=ddd>



Ulrich
Schwardmann

EPIC –
Consortium

EPIC Services

Infrastructure
Types
DOI integration
Support

PIDs 4
eResearch

- reasons:
 - EPIC PIDs are mainly intended for data reference
 - citability with EPIC PIDs is possible
 - but DataCite PIDs have the branding for citability
 - there is a user request for a seamless transition
- agreement between DataCite, TIB and EPIC
 - there is a special DOI prefix (provided by TIB), that is used
 - the EPIC PID (prefix/suffix) becomes suffix of this DOI prefix.
 - GWWDG became therefore member of DataCite on behalf of EPIC



EPIC

Ulrich
Schwardmann

EPIC –
Consortium

EPIC Services

Infrastructure
Types
DOI integration
Support

PIDs 4
eResearch

- **user management** for API access
- **consulting**
 - use of the API
 - installation and maintenance of the API software (limited)
 - training on request
- **user requests**
 - RfC scheme is established for API version 2
 - technical board of EPIC
 - evaluates, prioritizes and schedules RfC
 - solution is done on a best effort basis by partners
- API service **VM creation** and **hosting** (in preparation)



EPIC

Ulrich
Schwardmann

EPIC –
Consortium

EPIC Services

Infrastructure
Types
DOI integration
Support

PIDs 4
eResearch

■ customer

- decision about validity timeline
 - no expiration date means infinit validity
- PID consistency control
- PID maintainance
- a policy for handover of responsibility for PID maintainance in case of discontinuity of responsibility

■ provider

- persistency of PIDs
- resolvability of PIDs (EPIC MoU)
- maintainability of PIDs (EPIC MoU, API maintainance)
- services for consistancy control
- a policy for maintaince of expired PIDs

Thanks for your attention



EPIC

Ulrich
Schwardmann

<http://pidconsortium.eu>

Visit this afternoon the
EPIC User Forum
Rome, today 30.10.2013, 14:30-18:00

EPIC –
Consortium

EPIC Services

Infrastructure
Types
DOI integration
Support

PIDs 4
eResearch

Questions ??