



## DASISH workshop on Persistent Identifiers

Ulrich Schwardmann

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen  
(GWDOG)

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

08 December 2014, Cologne



123.456/abc

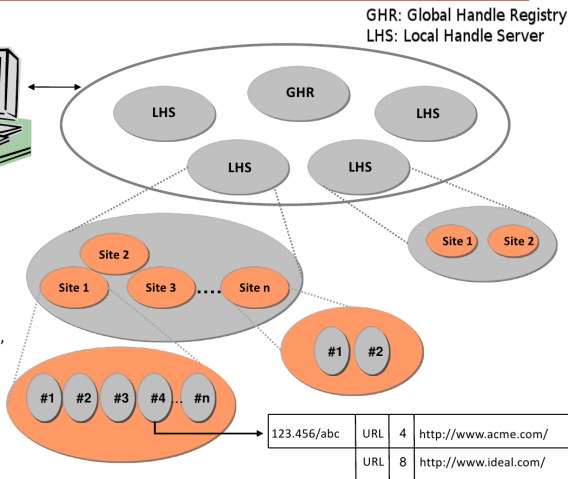
are strings consisting of a prefix (123.456) and a suffix (abc), separated by a slash ( / )

- a prefix is typically owned by some local organisation
- a prefix consists of digits and dots
- a suffix is an alphanumerical string
- additional suffix policies are in the responsibility of the prefix owner

## 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 services,  
an overview



Ulrich  
Schwardmann

The Handle  
System®

Infrastructure  
DONA

EPIC

EPIC –  
Consortium

EPIC  
Infrastructure  
Services

Types at EPIC

Templates

DOI integration  
Support

PIDs for  
eResearch

Corporation for National Research Initiatives



## ■ Local Handle Server (LHS)

- LHS resolves for one prefix
- Secondary LHS is mirror of Primary LHS
- Answer given by the fastest LHS

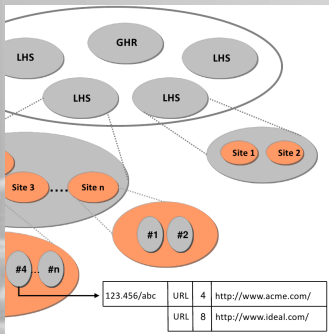
## ■ Global Handle Registry (GHR)

- maps prefixes to LHS
- GHR is a cluster of servers

## ■ Handle proxy (cluster) for resolution of PIDs

<http://hdl.handle.net/123.456/abc>

## ■ highly redundant resolution system



# DONA

## Digital Object Numbering Authority



- the administrative responsibility of the Handle System is currently handed over to an **international organisation**
- Digital Object Numbering Authority (**DONA**):
  - is a swiss foundation
  - under the auspices of the International Telecommunication Union (ITU),
    - the United Nations specialized agency for information and communication technologies
- the assignment of prefixes is controlled by DONA
- Multi-primary Prefix Administrators (**MPAs**) are responsible for prefix namespaces
- DOI fits into this framework (for the namespace 10.xxx),
  - but there are still bilateral contracts between DOI and CNRI.

EPIC services,  
an overview



Ulrich  
Schwardmann

The Handle  
System®

Infrastructure  
DONA

EPIC

EPIC –  
Consortium

EPIC  
Infrastructure  
Services

Types at EPIC

Templates

DOI integration

Support

PIDs for  
eResearch

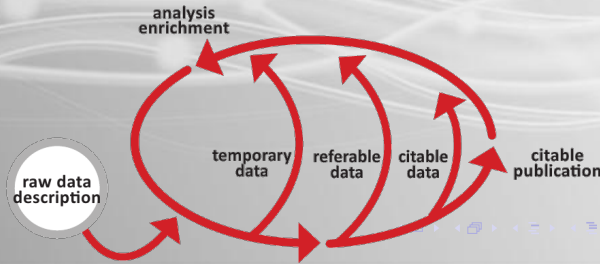


- DONA has a steering committee consisting of representatives from stakeholders worldwide
- DONA is responsible for
  - the global technical infrastructure,
  - formulation of technical policies
  - and the framework under which the MPAs operate,
  - an essential part of which is the central MPA Coordination Group.
- The MPAs are responsible for the day-to-day operations of the infrastructure.
  - An MPA must pay an annual fee to DONA to obtain the right to issue new prefixes in return.
  - The initial group of designated MPAs is very limited
  - current DONA policies are to maintain a rough balance between continents.

# EPIC: Consortium for a PID Infrastructure for eResearch



- is dedicated to provide persistent identifier (PID) services
- main scope are the scientific and cultural heritage communities:
- is a consortium of now six major European scientific computing centers *CSC (Finland)*, *DKRZ (Germany)*, *GR-Net(Greece)*, *GWDG (Germany)*, *PDC (Sweden)*, *SURF-SARA (Netherlands)*
- provides PIDs for digital objects already in an early state of scientific research



EPIC services,  
an overview



Ulrich  
Schwardmann

The Handle  
System®

Infrastructure  
DONA

EPIC

EPIC –  
Consortium

EPIC  
Infrastructure  
Services

Types at EPIC  
Templates  
DOI integration  
Support

PIDs for  
eResearch

# EPIC: Consortium for a PID Infrastructure for eResearch



- is dedicated to provide persistent identifier (PID) services
- main scope are the scientific and cultural heritage communities:
- is a consortium of now six major European scientific computing centers *CSC (Finland)*, *DKRZ (Germany)*, *GR-Net(Greece)*, *GWDG (Germany)*, *PDC (Sweden)*, *SURF-SARA (Netherlands)*
- provides PIDs for digital objects already in an early state of scientific research

EPIC services,  
an overview



Ulrich  
Schwardmann

The Handle  
System®

Infrastructure  
DONA

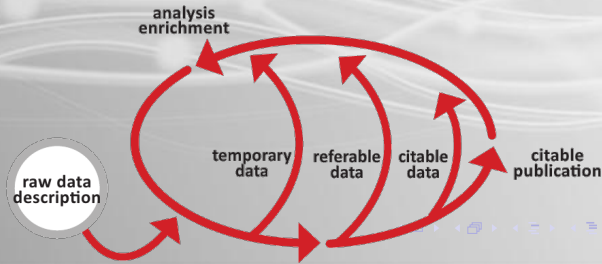
EPIC

EPIC –  
Consortium

EPIC  
Infrastructure  
Services

Types at EPIC  
Templates  
DOI integration  
Support

PIDs for  
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



- GWDDG has been designated as a European DONA MPA on behalf of EPIC
  - to represent EPIC's interests in the MPA Coordination Group
  - provide interaction with the DONA administrative board.
- As part of the MPA framework, GWDDG maintains the prefix namespace **21** on behalf of EPIC



- GHR mirror run by GWDOG on behalf of EPIC (first outside USA)
- handle resolution proxy
  - proxy server implemented at GWDOG (first outside USA)
  - as part of the global resolution network
- both improved significantly the average resolution time for European customers
- monitoring of these services

- about 31 million PIDs (Nov. 2014)
- about 30 prefixes
  - 5 prefixes with more than 2 million PIDs (8.5 million max.)
- 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
- Support of types and templates
- documentation and support

EPIC services,  
an overview



Ulrich  
Schwardmann

The Handle  
System®  
Infrastructure  
DONA

EPIC  
EPIC –  
Consortium  
EPIC  
Infrastructure  
Services  
Types at EPIC  
Templates  
DOI integration  
Support

PIDs for  
eResearch



- types are (additional) *parameters* associated with a PID
- the always mandatory type of a PID is the **URI**
  - pointer to the resource
- which (additional) types should be provided?
  - discussion in RDA WG *PID info types*
- necessary information about PID:
  - **PID owner**
    - responsible for consistency, to contact in case of discrepancies
  - **expiration date**
    - determines timeline for *curation* of the DO
    - not timeline of PID or its resolution



example policy for types provided in EPIC prefix 11858

- *mandatory*
  - URL
  - PID owner
- *optional*
  - author, title, creator
  - publication and expiration date
  - meta data URL
  - checksum (MD5,SHA-1)
  - file size



example policy for *checksum*: the integrity of DO

- **write multiple** by PID administrator
  - use case: **technical integrity** control
- **write once** by PID administrator
  - use case: **legal integrity** control



- see IETF RFC 6570
- 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>
- 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**





- 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
  - a special DOI prefix (provided by TIB) is used here
  - the EPIC PID (prefix/suffix) becomes suffix of this DOI prefix
  - GWDG became therefore member of DataCite on behalf of EPIC



- **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
- **in preparation:**
  - prefix administration
  - API service: **SaaS** and **hosting**

# Responsibilities concerning PIDs



EPIC services,  
an overview



Ulrich  
Schwardmann

The Handle  
System®  
Infrastructure  
DONA

EPIC  
EPIC –  
Consortium  
EPIC  
Infrastructure  
Services  
Types at EPIC  
Templates  
DOI integration  
Support

PIDs for  
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

# Towards a global resolution scheme



## current possibilities

- doi:10.1000/1
  - <http://hdl.handle.net/10.1000/1>
  - <https://nbn-resolving.org/doi:10.1000/1>
- hdl:21.21.2/123
  - <http://hdl.handle.net/21.21.2/123>
  - <https://nbn-resolving.org/hdl:21.21.2/123>
- urn:nbn:de:1111-20091210269
  - <https://nbn-resolving.org/urn:nbn:de:1111-20091210269>
  - <https://nbn-resolving.org/urn:nbn:de:1111-20091210269>
- ark:/1234/5678
  - <https://nbn-resolving.org/ark:/1234/5678>

EPIC services,  
an overview



Ulrich  
Schwardmann

The Handle  
System®  
Infrastructure  
DONA

EPIC  
EPIC –  
Consortium  
EPIC  
Infrastructure  
Services  
Templates at EPIC  
DOI integration  
Support  
PIDs for  
eResearch



# Thanks for your attention !

View also the EPIC Webpage:  
<http://pidconsortium.eu>