The Handle System

Ulrich Schwardmann

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
(GWDG)

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

Metadata and Persistent Identifiers for
Social and Economic Data
Berlin, 7. May 2012

The Handle System

Content

about us
EPIC
GWDG
Handle System
Scientific Data
Metadata
PID+Metadata
Handle+Metadata

1. about us
   - EPIC
   - GWDG
2. Handle System
3. Scientific Data
   - Metadata
   - PID+Metadata
   - Handle+Metadata

EPIC

EPIC Partners:

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

SARA
Reken- en Netwerkdiensten

CSC
IT Center for Science

Mission: ... setting up and maintaining a joint service with high availibility for registering, storing and resolving persistent identifiers based on handles for the scientific community on a non-profit basis. (MoU)
Candidates: DKRZ, CINECA, CINES, STFC, BSC, INGV, DESY, PSNC
more ...
http://pidconsortium.eu

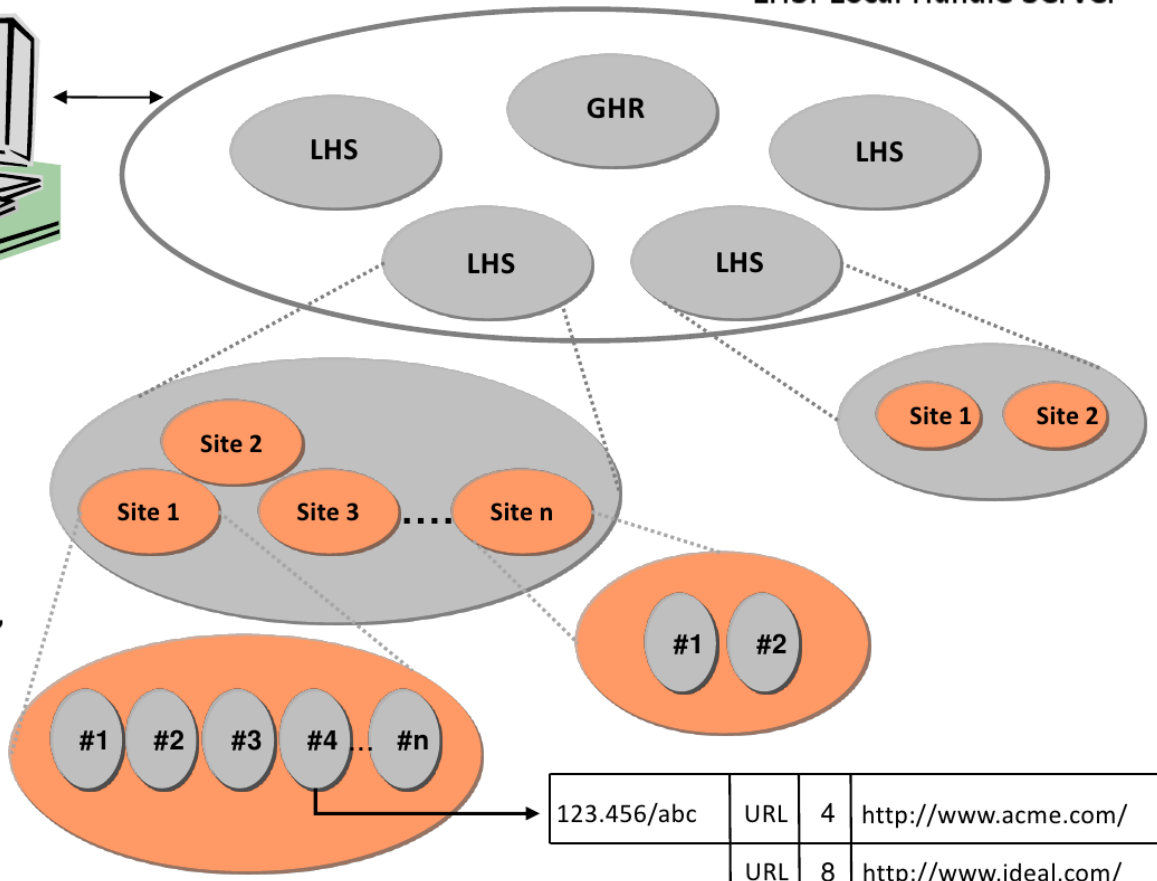
GWDG

- GWDG is a corporate facility of the Max-Planck-Gesellschaft and the
- Georg-August University of Göttingen.
- for both it is IT competence center
- and operates a computer center for all local institutes
- GWDG was founded in 1970 as company.
- is located in Göttingen
- It operates on a non-profit principle
- 25,000 users, 1000 HPC users
- Staff: about 80 employees
- several funded eScience projects

Handle System

Handle Resolution

GHR: Global Handle Registry
LHS: Local Handle Server



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.

123.456/abc	URL	4	http://www.acme.com/
	URL	8	http://www.ideal.com/

Handle System

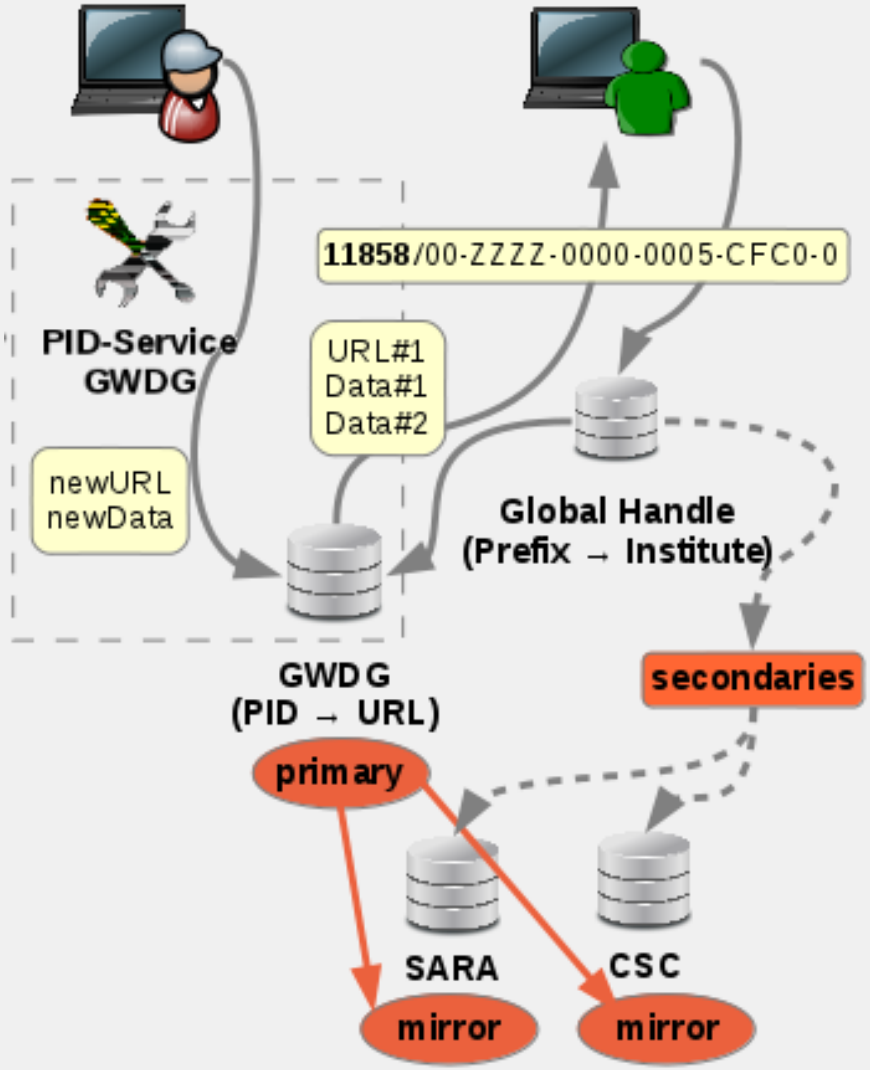
- Highly redundant resolution system
  - LHS resolves for one prefix
  - Secondary LHS is mirror of Primary
  - Answer given by the fastest LHS
- Single point of failure: GHR?
  - GHR is a cluster of servers
  - Secondary GHR at GWDG and in China (and Australia in near future)

Handle System

Quality and Trust

- uniqueness has to be guarantueed by the LHS
  - as the other trust issues too
- EPIC is aimed to build an umbrella for handle LHS
  - with common software bases
  - common quality of service
  - mutual backup of services

Handle System



The Handle System

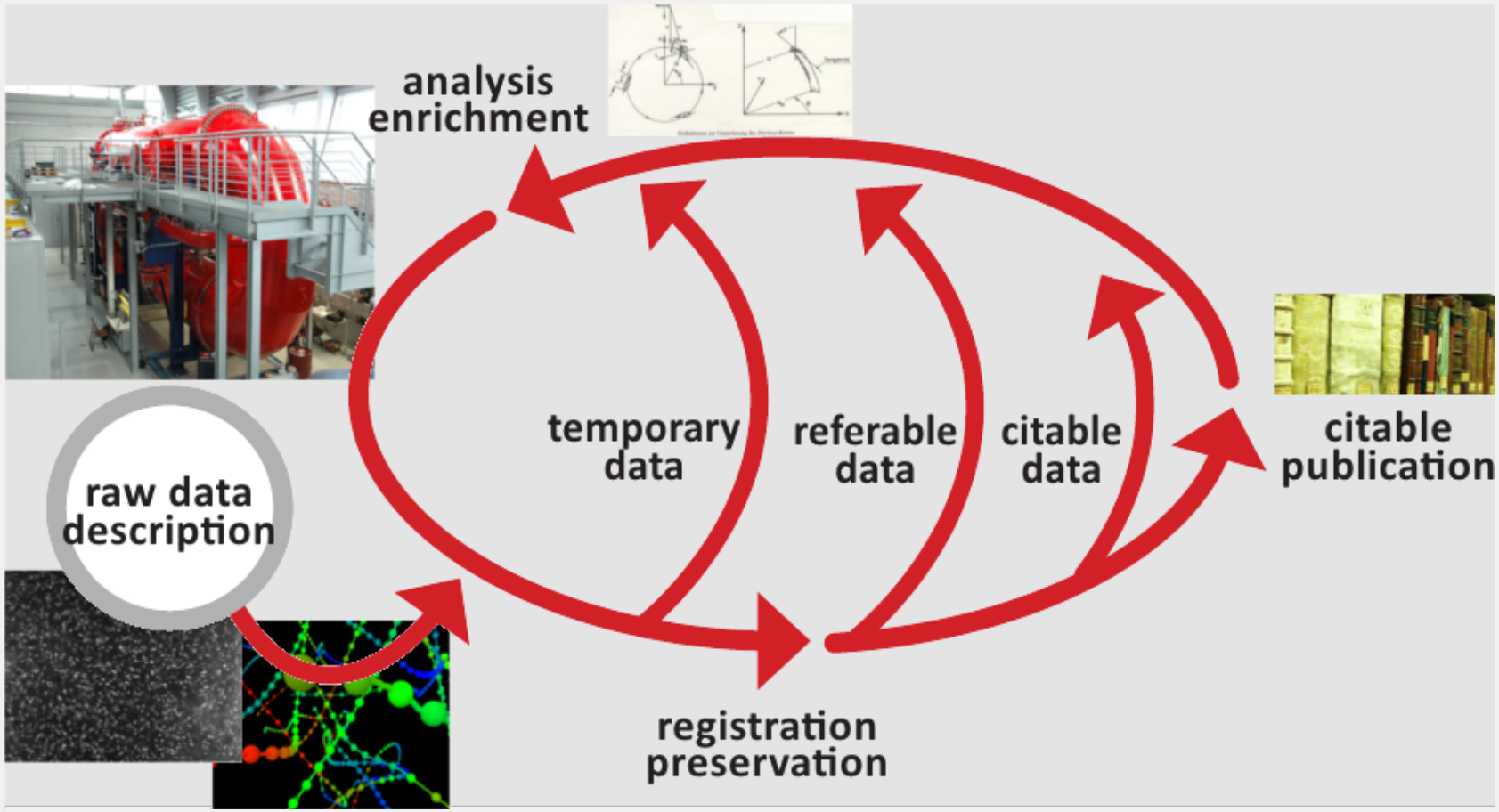
Ulrich Schwardmann

about us
EPIC
GWDG

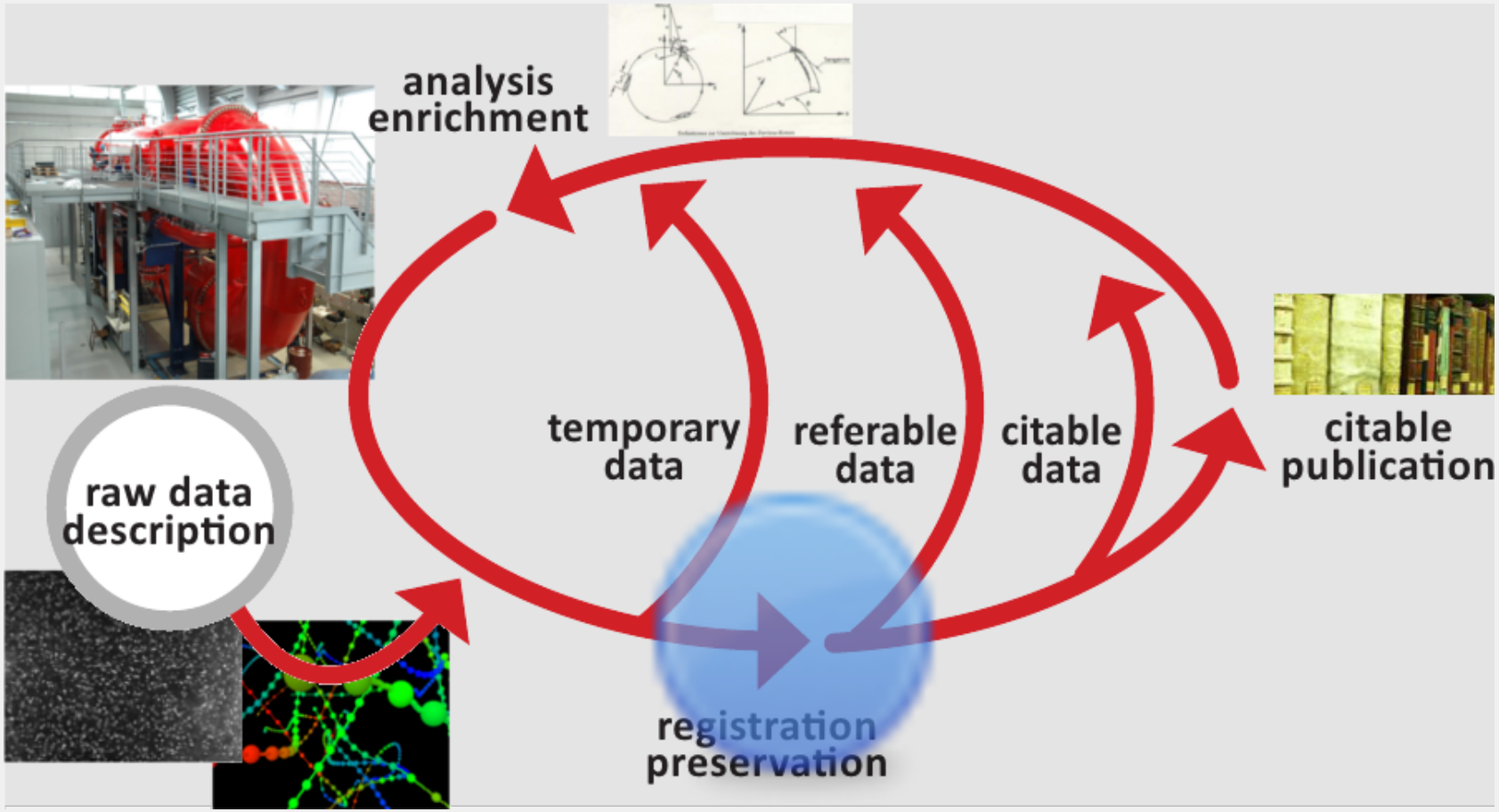
Handle System

Scientific Data
Metadata
PID+Metadata
Handle+Metadata

The Lifecycle of Scientific Data



The Lifecycle of Scientific Data

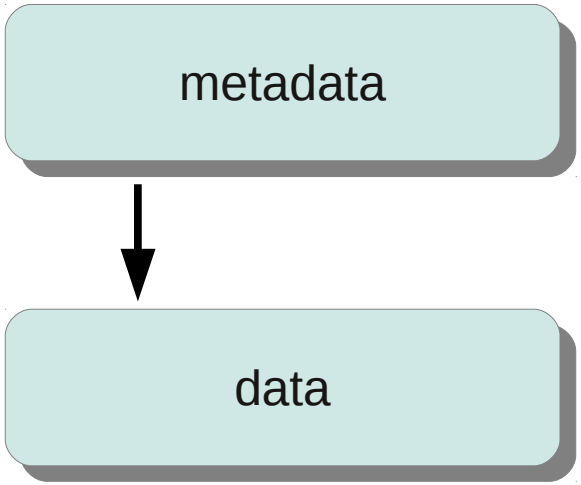


The Lifecycle of Scientific Data

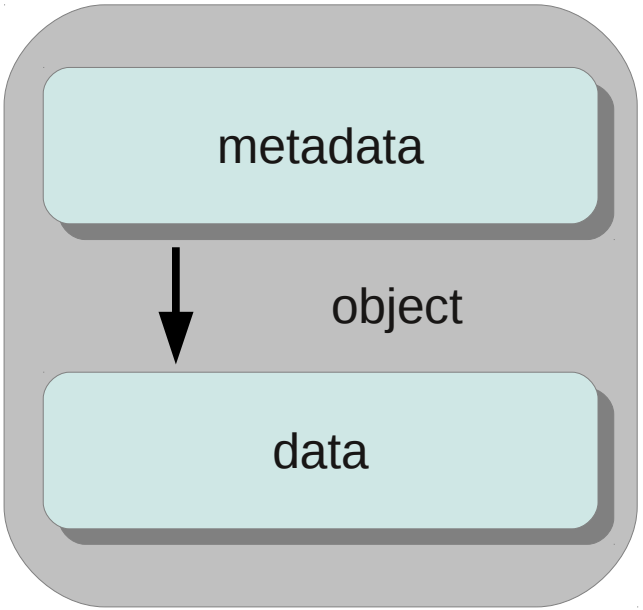
- scientific data, that is not cited, may (must?) be removed after a while
  - due to data lifecycle management to fullfil good scientific practices
  - because storing data becomes more and more expensive
- remember: persistence in PID means persistence of the identifier, not necessarily of the data

Maintaining Data Objects

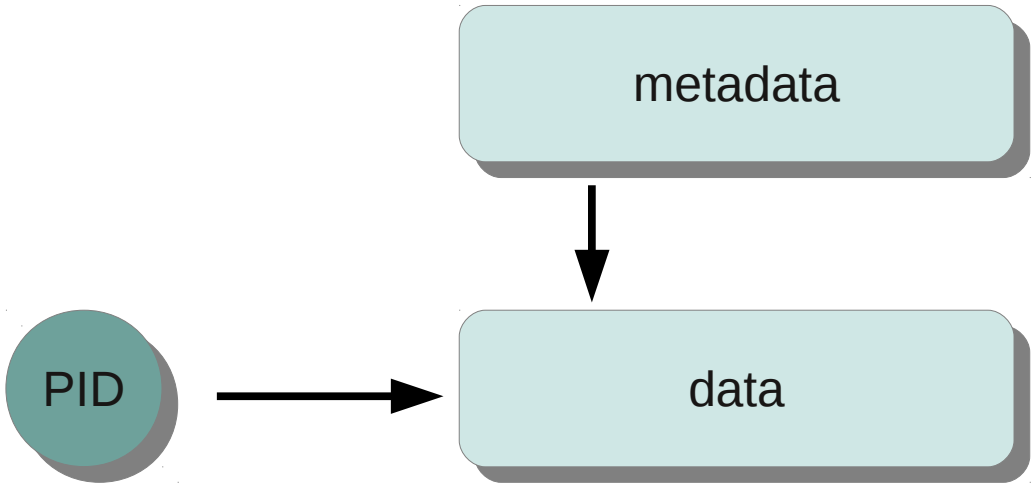
Maintaining Data Objects



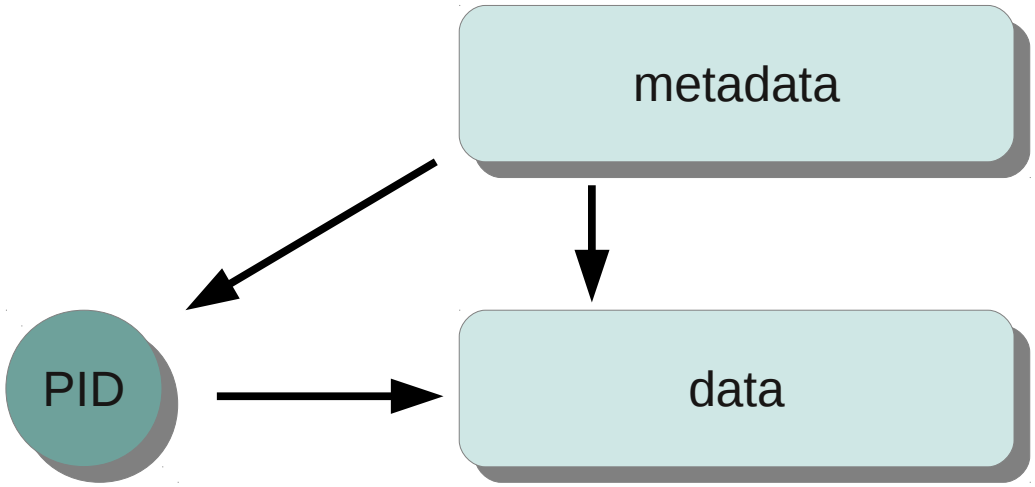
Maintaining Data Objects



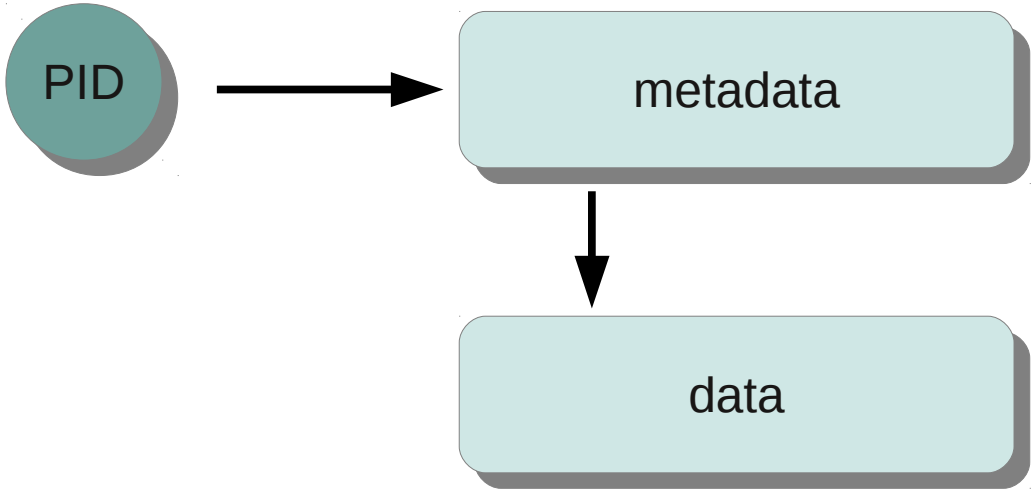
Maintaining Data Objects



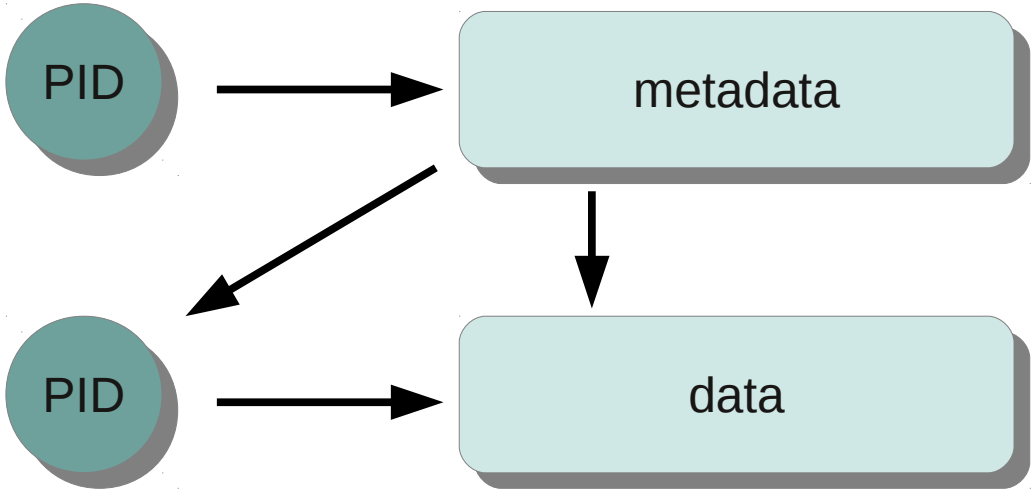
Maintaining Data Objects



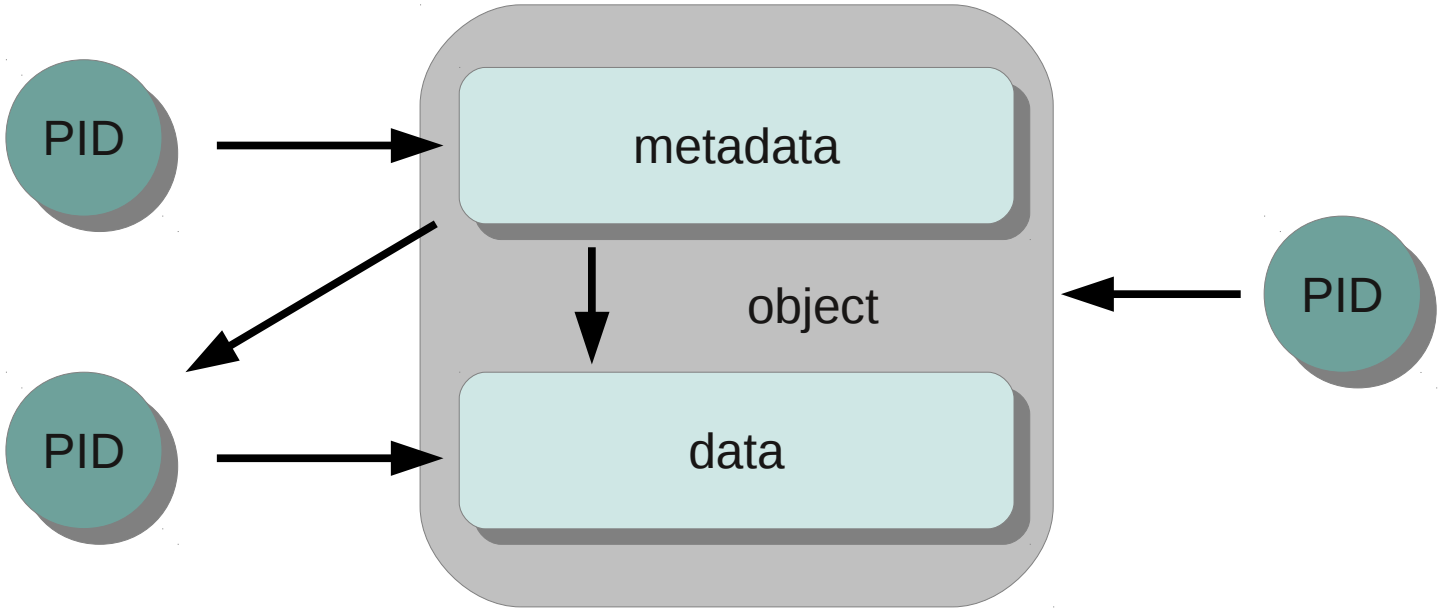
Maintaining Data Objects



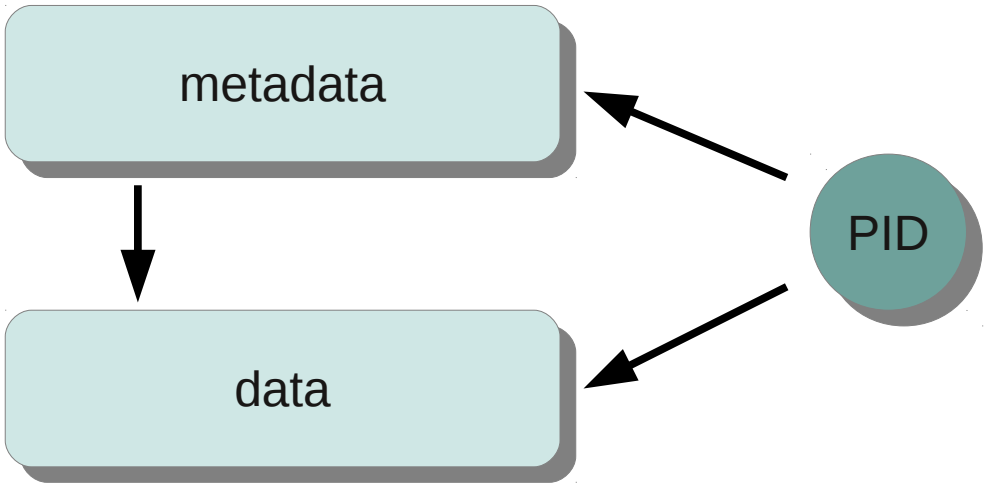
Maintaining Data Objects



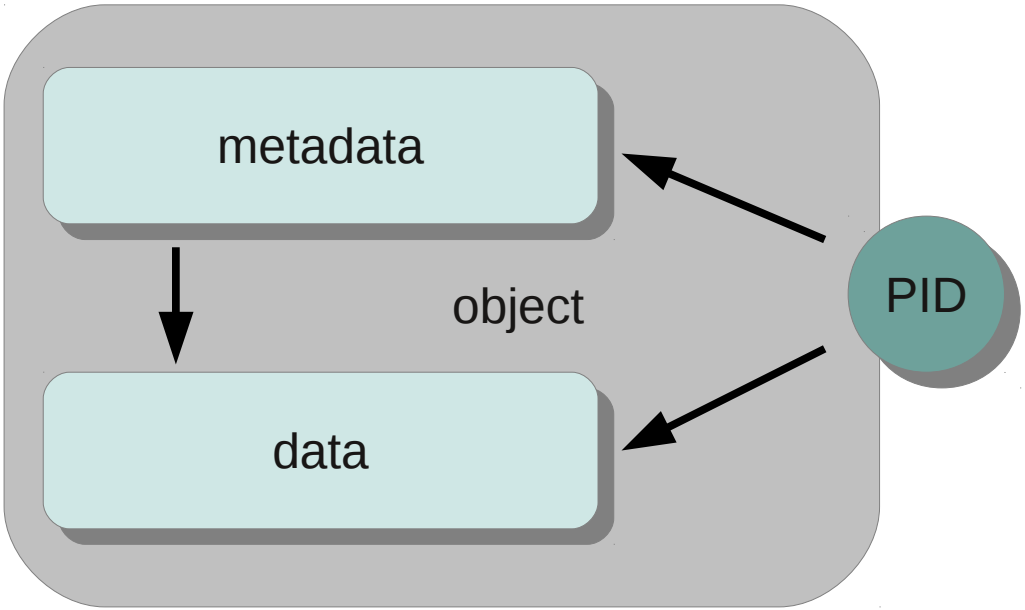
Maintaining Data Objects



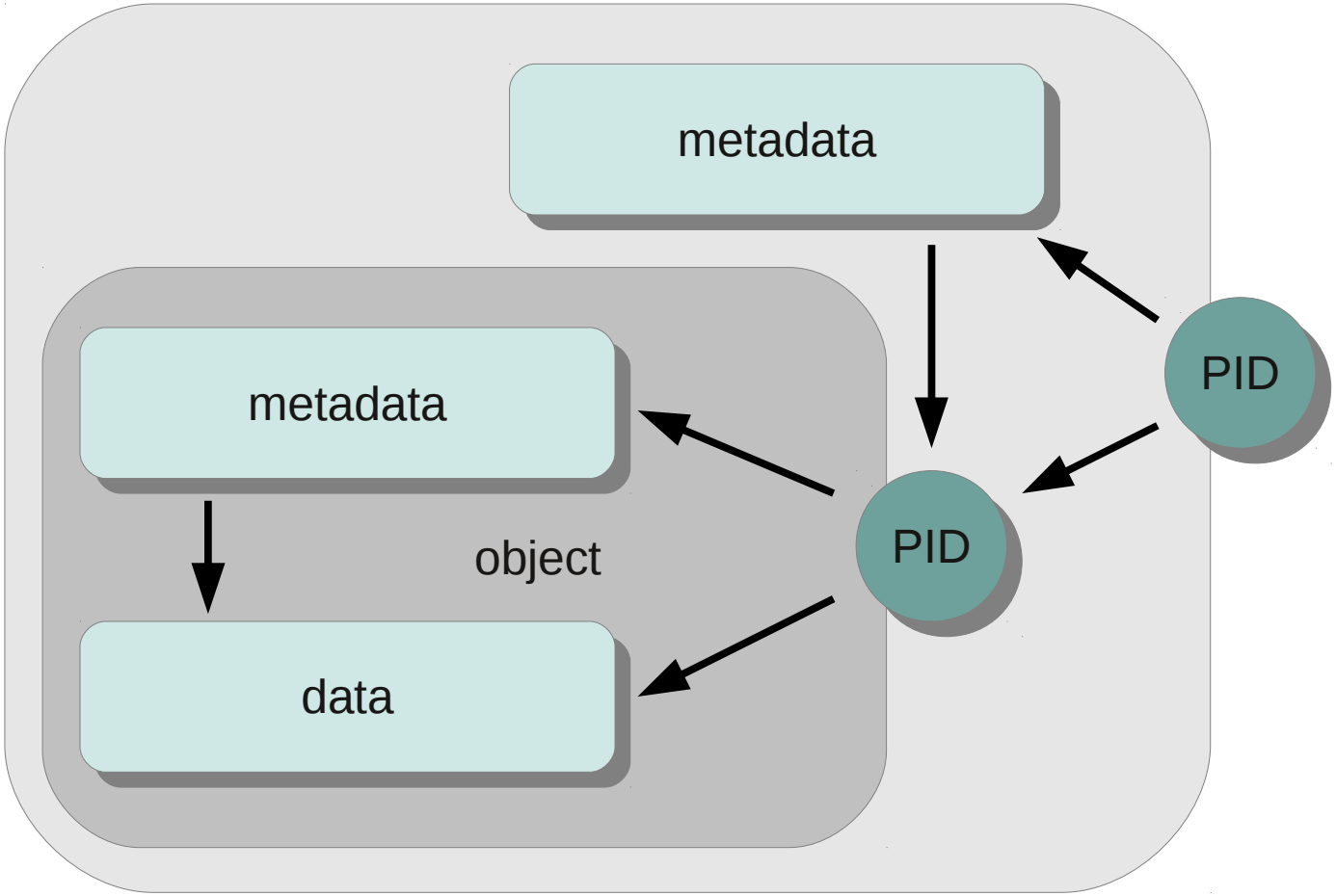
Maintaining Data Objects



Maintaining Data Objects



Maintaining Data Objects



Handle, Data and Metadata

- what can the handle system do, to refer to both: data and metadata?
  - the handle PID is originally designed for one reference: it resolves to the URL
- but:
  - handle uses a database to store the URL for a PID
  - this database can contain other fields too
  - and the LHS can be asked to resolve these other fields

Handle, Data and Metadata

Examples of other PID fields:

- expiration date
  - because scientific data gets outdated, the PID must contain an expiration date
  - or the PID must point to the metadata, that contains the expiration date
  - in any case one needs to get more information, then just the URL of the object
  - expir.date can be part of the metadata if and only if PID points to metadata or to both, data and metadata
- checksum
  - for integrity check it makes sense to maintain a checksum
  - checksum can be part of the metadata if and only if PID points to metadata or to both, data and metadata
- metadata
  - must be an additional field if one also wants the PID to refer to the data directly

the current EPIC solution

11858/00-ZZZZ-0000-0005-BD17-A
Prefix / Suffix

Create Verbose Handle

URL: http://www.gwdg.de/~parallel/parallelrechner/Hardware_Ueberblick.pdf
Filesize: (in bytes)
Checksum: (optional, see below)
Checksum is write-once, lower case hex: 'md5:32digits' or 'sha1:40digits'
Title: Parallelrechner Hardware Ueberblick der GWDG
Author(s): Ulrich Schwardmann
Publication date: 2010-11-12 (see below)
Expiry date: (optional, see below)
Dates can be in format yyyy, yyyy-mm-dd or yyyy-mm-dd hh:mm:ss
Metadata URL: http://www.gwdg.de/~parallel/parallelrechner/Hardware_Ueberblick.xml
Suffix: (user defined, optional)
Confirm in XML: (default: HTML)
Enter data above and Create Verbose Handle

XML-Output of Handle content


```
−<pidserviceresponse>
  <action>view</action>
 −<Handle>
    <pid>11858/00-ZZZZ-0000-0005-BD17-A</pid>
   −<url>
      http://www.gwdg.de/~parallel/parallelrechner/Hardware_Ueberblick.pdf
    </url>
    <size>-1</size>
    <pubdate>2010-11-12</pubdate>
    <title>Parallelrechner Hardware Ueberblick der GWDG</title>
    <authors>Ulrich Schwardmann</authors>
   −<metadata_url>
      http://www.gwdg.de/~parallel/parallelrechner/Hardware_Ueberblick.xml
    </metadata_url>
    <creator>demo2</creator>
  </Handle>
 −<User>
    <uid canCreate="no" canModify="no">anonymous</uid>
  </User>
</pidserviceresponse>
```

the current EPIC solution

11858/00-ZZZZ-0000-0005-BD17-A
Prefix / Suffix

Create Verbose Handle

URL: http://www.gwdg.de/~parallel/parallelrechner/Hardware_Ueberblick.pdf
Filesize: (in bytes)
Checksum: (optional, see below)
Checksum is write-once, lower case hex: 'md5:32digits' or 'sha1:40digits'
Title: Parallelrechner Hardware Ueberblick der GWDG
Author(s): Ulrich Schwardmann
Publication date: 2010-11-12 (see below)
Expiry date: (optional, see below)
Dates can be in format yyyy, yyyy-mm-dd or yyyy-mm-dd hh:mm:ss
Metadata URL: http://www.gwdg.de/~parallel/parallelrechner/Hardware_Ueberblick.xml
Suffix: (user defined, optional)
Confirm in XML: (default: HTML)
Enter data above and Create Verbose Handle

XML-Output of Handle content


```
-<pidserviceresponse>
  <action>view</action>
  -<Handle>
    <pid>11858/00-ZZZZ-0000-0005-BD17-A</pid>
    -<url>
      http://www.gwdg.de/~parallel/parallelrechner/Hardware_Ueberblick.pdf
    </url>
    <size>-1</size>
    <pubdate>2010-11-12</pubdate>
    <title>Parallelrechner Hardware Ueberblick der GWDG</title>
    <authors>Ulrich Schwardmann</authors>
    -<metadata_url>
      http://www.gwdg.de/~parallel/parallelrechner/Hardware_Ueberblick.xml
    </metadata_url>
    <creator>demo2</creator>
  </Handle>
  -<User>
    <uid canCreate="no" canModify="no">anonymous</uid>
  </User>
</pidserviceresponse>
```

Thanks for your attention

Questions ???