

OAIS Practices Archive & Audit Store

Jake Cowton
CERN

2nd June 2014

Recap: Persistent Identifiers



- Long-lasting references to digital objects
- Citability & traceability throughout lifecycle
- ARK, DOI, Handle, PURL
- PI → Resolver → Final Location
- DataCite client

DOI – Digital Object Identifier



- Reserve a new DOI locally
- Assign it to a local record
- Save the information in the local PI store
- Register the DOI & metadata in DataCite
- Update the local record

What is an OAIS?



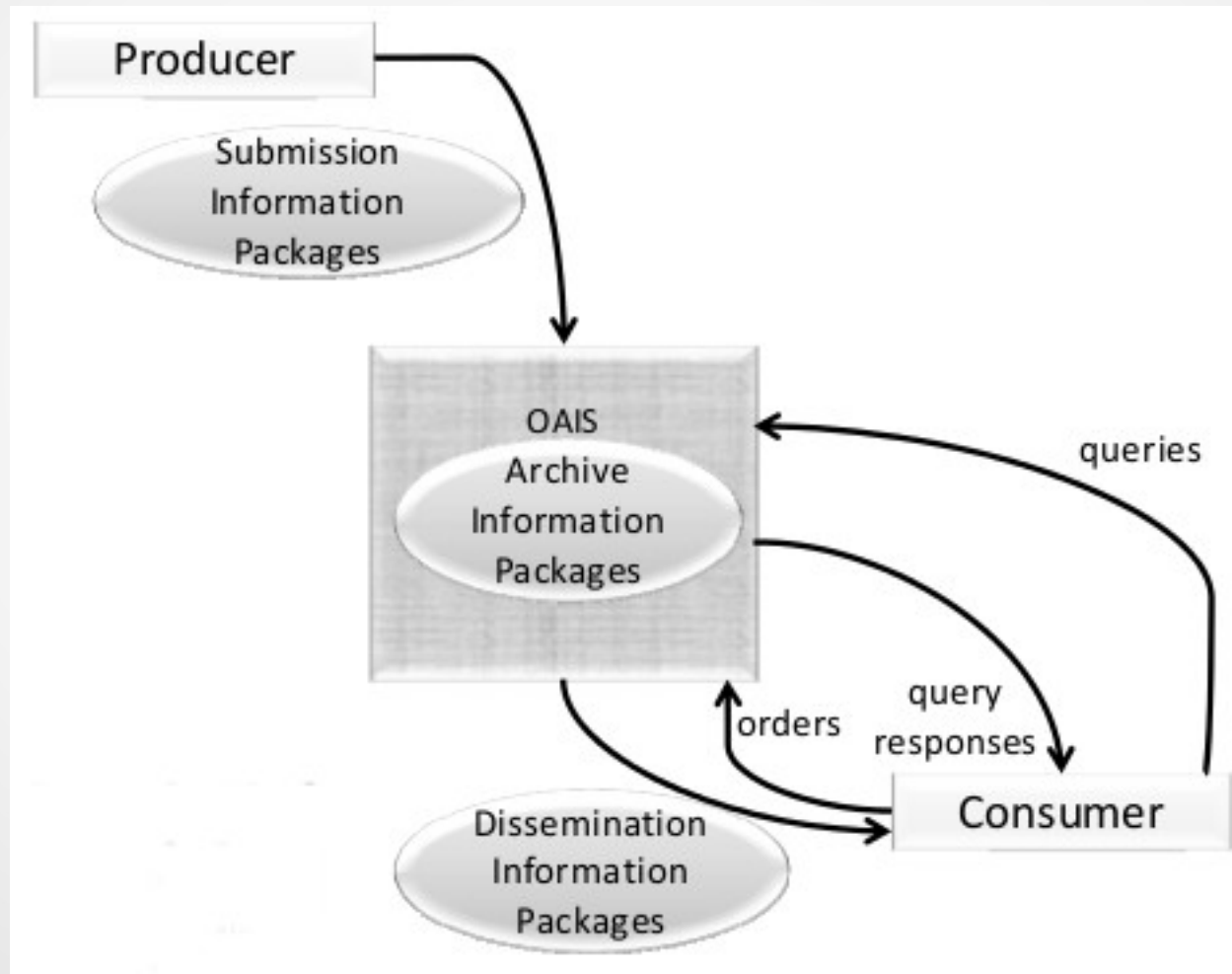
“An archive that has accepted the responsibility to preserve information and make it available for a designated community.”

Why OAIS?



- Interoperability
- Expectations
- Format
- Long-term preservation

The Model in Brief



- Preservation Description Information (PDI)
- Content Information
- Packaging Information
- Package Description

The OAIS Reference Model

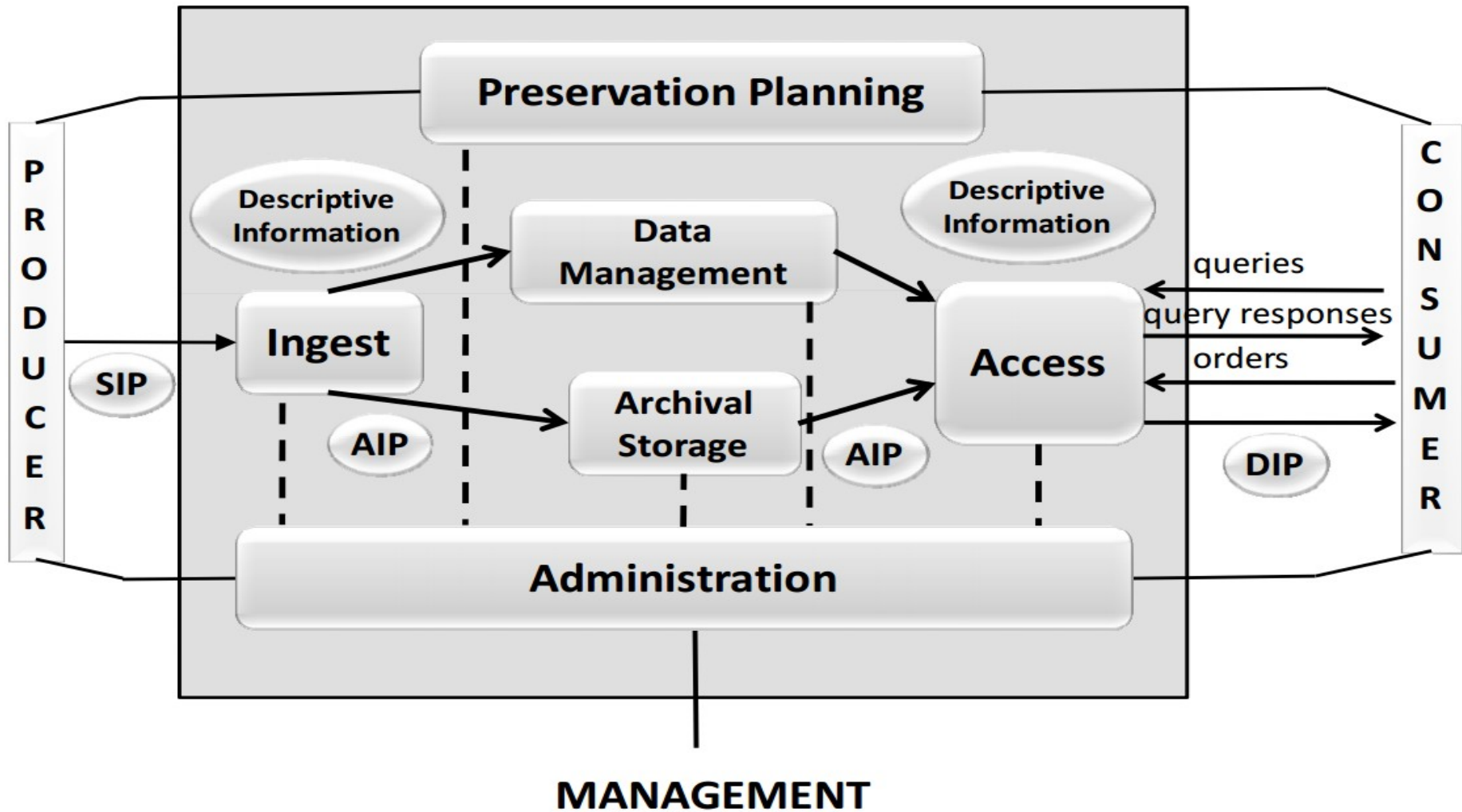


Figure 4-1: OAIS Functional Entities

- A new module for Invenio
 - Disaster Recovery
 - BagIt
 - Daemon usage
- 1) Collect all files for a given record
 - 2) Convert to BagIt
 - 3) Compress
 - 4) Write to any type of storage

Developed by the Library of Congress
(National Library of USA)

“A hierarchical file packaging format
for storage and transfer of arbitrary
digital content.”

Example Bag



```
myfirstbag/
|
|  manifest-md5.txt
|    (49afbd86a1ca9f34b677a3f09655eae9 data/27613-h/images/q172.png)
|    (408ad21d50cef31da4df6d9ed81b01a7 data/27613-h/images/q172.txt)
|
|  bagit.txt
|    (BagIt-version: 0.96 )
|    (Tag-File-Character-Encoding: UTF-8 )
|
|  \--- data/
|    |
|    |  27613-h/images/q172.png
|    |    (... image bytes ...)
|    |
|    |  27613-h/images/q172.txt
|    |    (... OCR text ...)
|    |
|    |  ....
```

The OAIS Reference Model

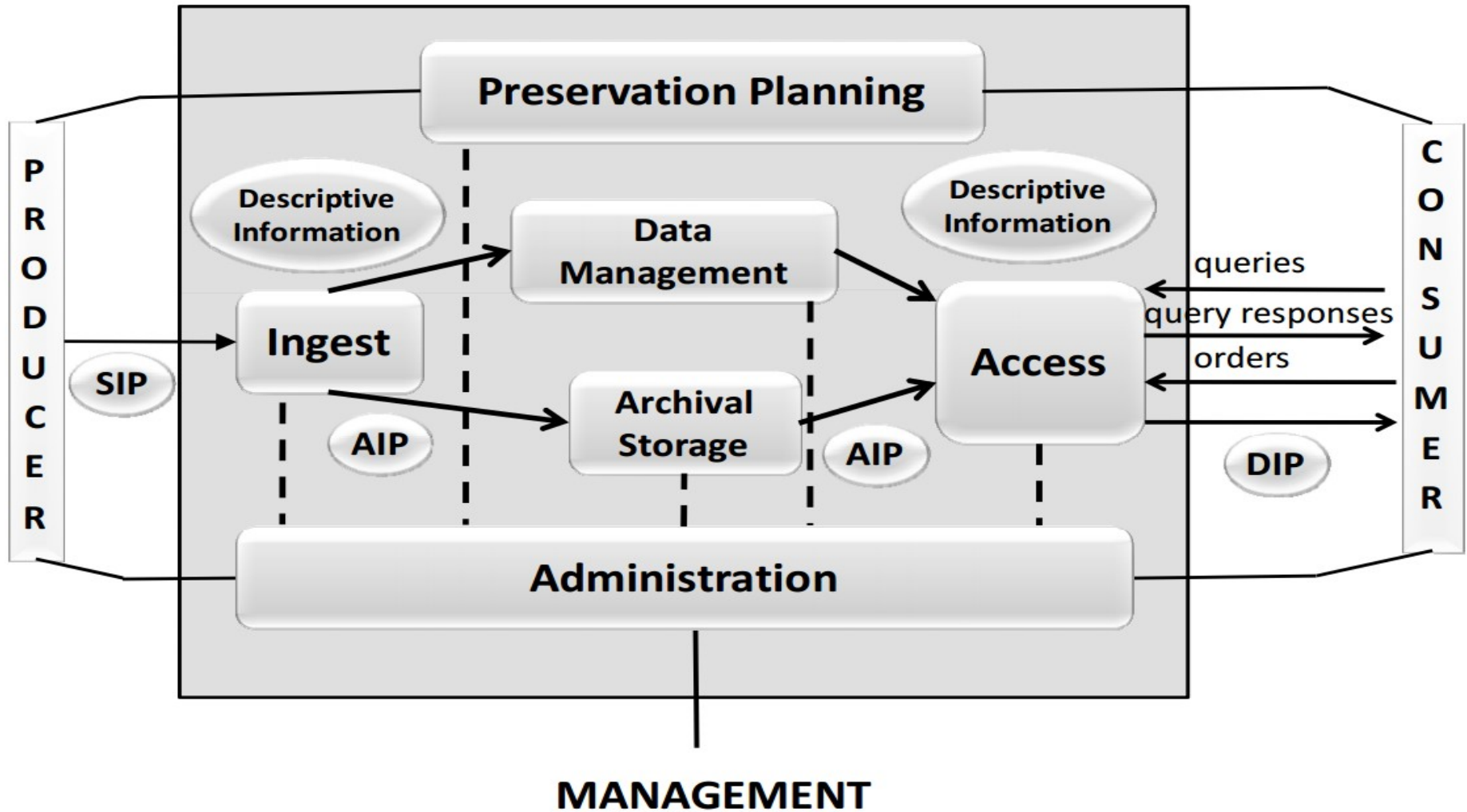


Figure 4-1: OAIS Functional Entities

Audit Store

Why create an Audit Store?



- This is a central auditing system
- For all data

added	2014-02-28	Search for direct production of charginos and neutralinos in events with three leptons and missing transverse momentum in $\sqrt{s} = 8\text{TeV}$ pp collisions with the ATLAS detector - ATLAS Collaboration (Aad, Georges et al.) arXiv:1402.7029 [hep-ex] CERN-PH-EP-2014-019
added	2014-02-27	Fourth generation searches at ATLAS - ATLAS Collaboration (Nektarijevic, Snezana for the collaboration) PoS HQL2012 (2012) 070
added	2014-02-26	Measurement of top quark pair production with the ATLAS detector - ATLAS Collaboration (Papadelis, Aras for the collaboration) PoS KRUGER2010 (2011) 016
added	2014-02-26	Measurement of the top-quark pair production cross section in ATLAS - ATLAS Collaboration (Costa, M.J. for the collaboration) PoS EPS-HEP2011 (2011) 350
added	2014-02-26	Measurement of the production of a W boson in association with a charm quark in pp collisions at $\sqrt{s}=7\text{ TeV}$ with the ATLAS detector - ATLAS Collaboration (Aad, Georges et al.) arXiv:1402.6263 [hep-ex] CERN-PH-EP-2014-007
removed	2013-11-05	The record with id 1193597 was deleted
added	2013-10-17	Measurement of the Top Quark Mass from $s\sqrt{=}7\text{ TeV}$ ATLAS Data using a 3-dimensional Template Fit - The ATLAS collaboration ATLAS-CONF-2013-046
added	2013-10-17	Search for direct top squark pair production in final states with two leptons in $s\sqrt{=}8\text{ TeV}$ pp collisions using 20fb^{-1} of ATLAS data. - The ATLAS collaboration ATLAS-CONF-2013-048

Audit Store

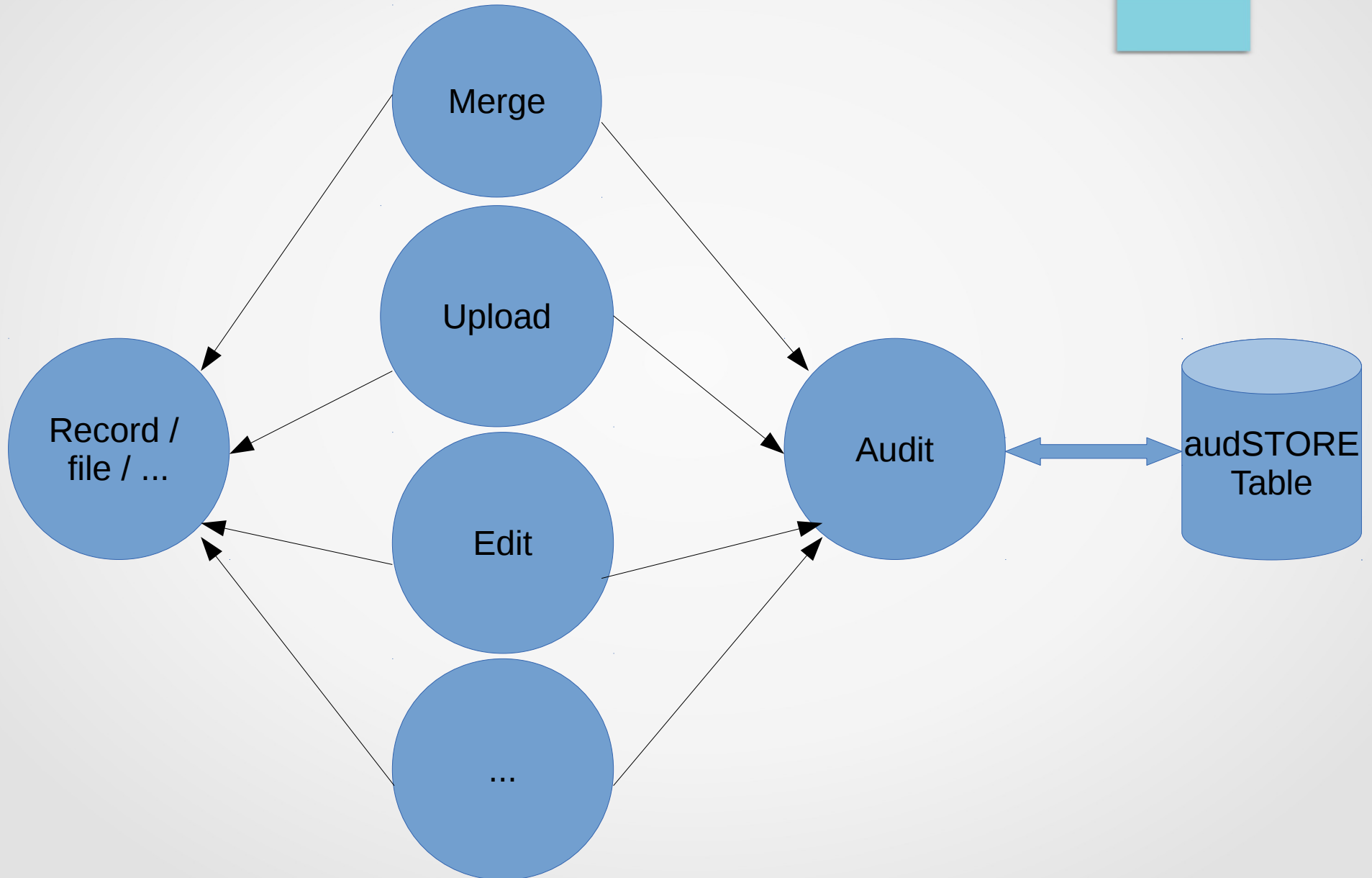


- 1 Table
- Storing what happens to things in the system
- Easier to follow the life-cycle of data

_id	object_type	object_id	action_user	action_date	action_name	action_args	action_result
146	record	123345	admin	2014-03-18 13:50:14	bibupload	*Some MARC data*	OK
147	record	432	admin	2014-03-18 13:52:38	bibedit	*Some changes*	OK

Plugs into any module

INVENIO



Conclusions



- Implemented ingestion storage – via BlogForever EU project
- Implemented DOI
- Implemented Archival Storage
- Implemented Auditing
- Full data tracking across life-cycle