B2Share

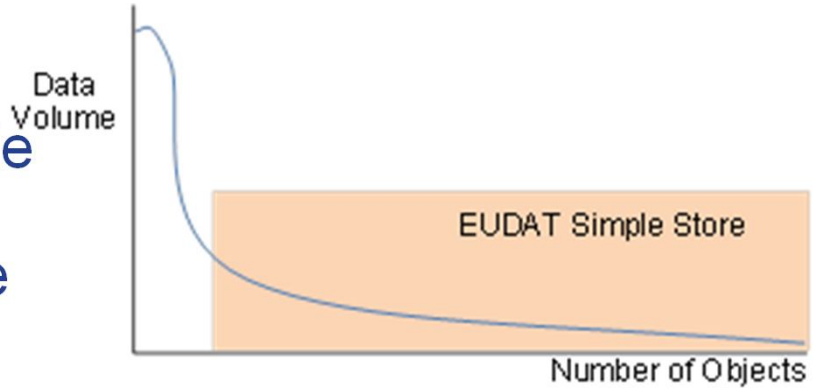
Easy data deposit & share

Ilja Livenson
PDC/SNIC
October 28, 2013

EUDAT

Motivation

- Address the “long tail” of “small” data and the researchers/citizen scientists creating/manipulating them
- EUDAT will offer a simple storage service - **B2Share** to
  - Offer a service to *everyone* with a data
  - Get a wide visibility in the research world
  - Address the many researchers and citizen scientists and not just the “few” organized in large communities
  - Make data accessible, shareable and persistent
  - Offer a strong European service

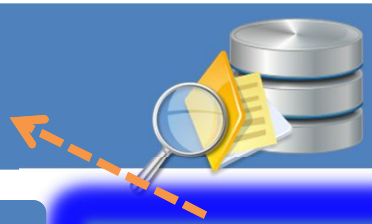


Approach

- Start simple
- A web service with a simple deposit flow
- Integration with EUDAT services behind the curtains
- Open-source and evolving based on user feedback
- Hosted by EUDAT partners

Metadata Catalogue

Aggregated EUDAT metadata domain.
Data inventory



Safe Replication

Data curation and access optimization

Data Staging

Simple Store

Researcher data store (simple upload, share and access)

AAI

Network of trust among authentication and authorization actors

PID

Identity
Integrity
Authenticity
Locations

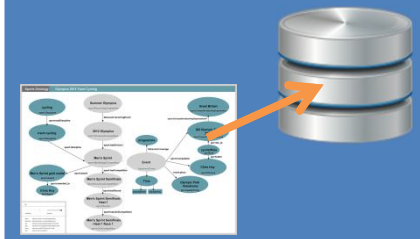
services to come

EUDAT Box

dropbox-like service
easy sharing
local synching

Semantic Anno

checking & referencing



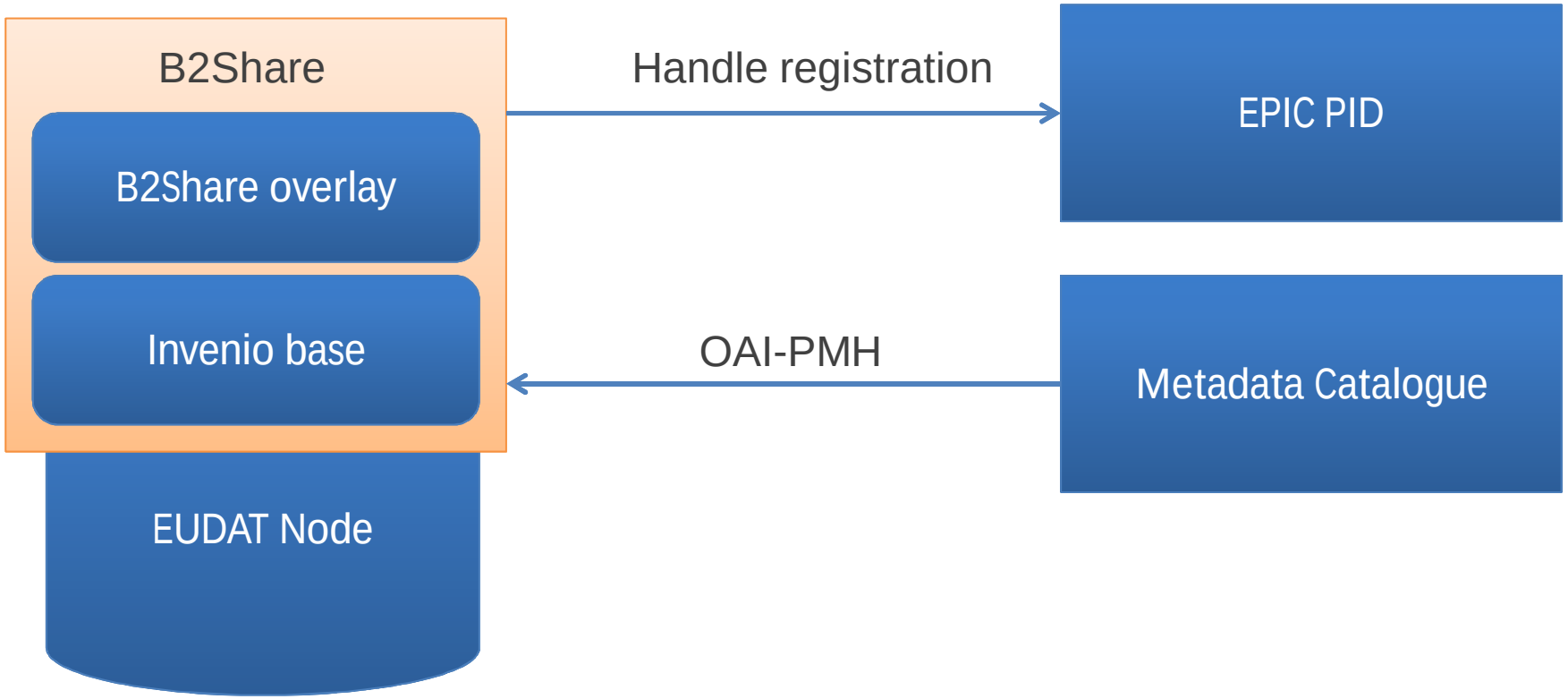
Dynamic Data

immediate handling

B2Share functionality

- Data upload
  - Metadata tagging
  - Parameterized by the scientific domain
- Search and retrieval
- Persistent references
- Metadata export

Architecture



Based on CERN's Invenio technology

- Production use in many different contexts
- Support from CERN's core team
- Synergies with OpenAIRE+

Invenio features

- navigable collection tree (regular, virtual)
- powerful search engine
  - combined metadata, reference and fulltext search
- flexible metadata (MARC, OA)
- handling any kind of document (multimedia)
- customizable input, formatting and linking
- personalization and collaborative features:
  - alerts, baskets, groups, reviews, comments
- internationalisation (28 languages)

http://simko.home.cern.ch/simko/talks/invenio-technology-openlab-aug-2011.pdf

EUDAT

B2Share deposit flow

Upload

- Select one or more files
- Click to start uploading

Domain

- Pick a scientific domain of submission
- Modifies the metadata enrichment form

Metadata

- Domain dependent
- Configurable to simplify adding new domains

EUDAT

B2Share security model

- All metadata is public
- Data can be marked as private – in this case accessing the data files will result in a 403 code

- More elaborate models will follow
  - E.g. sharing only among collaborators

EUDAT

Metadata Model

- All submission have a generic metadata associated with them
  - Title, description, tags, etc
  - DataCite-compliant
- Domain specific metadata is requested _only_ if a specific domain is selected
  - Areas of science: e.g. ecology, linguistics
  - Projects: e.g. DRIHM, CLARIN
- Input fields + validation are automatically generated from the metadata description

EUDAT

Integration with EUDAT services (existing)

- EPIC PID
  - On deposit EPIC service is contacted to register a UUID
  - UUID is embedded into the metadata record of the entry
- Metadata catalogue
  - B2Share exposes metadata via OAI-PMH protocol
  - Metadata catalogue consumes it and maps to its standard model

B2Share vs Zenodo

- Are based on the same framework (Invenio)
- B2Share is more ‘lightweight’
  - Leverages on the EUDAT services
- Zenodo targets ‘high-quality’ data, i.e. publications and associated data
- B2Share aims at offering a more generic submission mechanism for multiple scientific domains + EUDAT service integration

Technology

- Basis for the implementation: Invenio
  - Library software
  - Mostly CERN-based development
- A separate B2Share overlay
  - Installs on top of a working Invenio instance
  - Brings models, views and deposit logic
  - flask + jinja2 + sqlalchemy
  - https://github.com/SimpleStore/simplestore

EUDAT

Diapositiva 14

1 Ilja Livenson; 21/10/2012

Try out

- B2Share instance for trying out

https://ss-dev.pdc.kth.se

- Central instance will be available at:

https://b2share.eudat.eu

Adding a new domain

- Generic domain of submission is often not enough
- B2Share supports easy way for enabling additional scientific domains
- Hopefully engaged by a community

Defining metadata

- Define a domain description based on a template
- Add domain-specific fields

```
domain = ”Cats"
table_name = ‘cats_imags’
icon = ‘icon-cat’
fields = [
  {
   'name': ‘funlevel’,
   'display_text': “Fun level”,
   'col_type': db.String(256),
   'mandatory': True,
   'extra': False,
   ‘description’: ‘Increase in the happiness
            level caused by the image’
  }
]
```

Enabling the domain

$ # check enabled domains

$ grep CFG_SIMPLESTORE_DOMAINS invenio-local.conf

CFG_SIMPLESTORE_DOMAINS = generic, drihm, eudat, ecology

$ # append ‘cats’ to the list

$ sudo -u www-data /opt/invenio/bin/ \ inveniocfg --update-all

EUDAT

Future Work

- Integration with further EUDAT services
  - Ingestion into Safe replication
- Increased security/privacy model
- Clean API for batch uploads

THANK YOU!

ON BEHALF OF B2SHARE CONTRIBUTORS
ADRIAN MOUAT
HUW MORRIS
ILJA LIVENOSN

livenson@kth.se

EUDAT

Feedback

- We'd like to use this session also for getting immediate feedback ☺

1. Do you think that B2Share makes sense for your needs?

2. What would need to be changed to make the service more attractive?