NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

MathArc

Cornell University

Co-operating Preservation Archives
Sharing Collections Among Dissimilar OAIS Repositories

William Kehoe, Adam Smith, Marcy Rosenkrantz
Cornell University Library

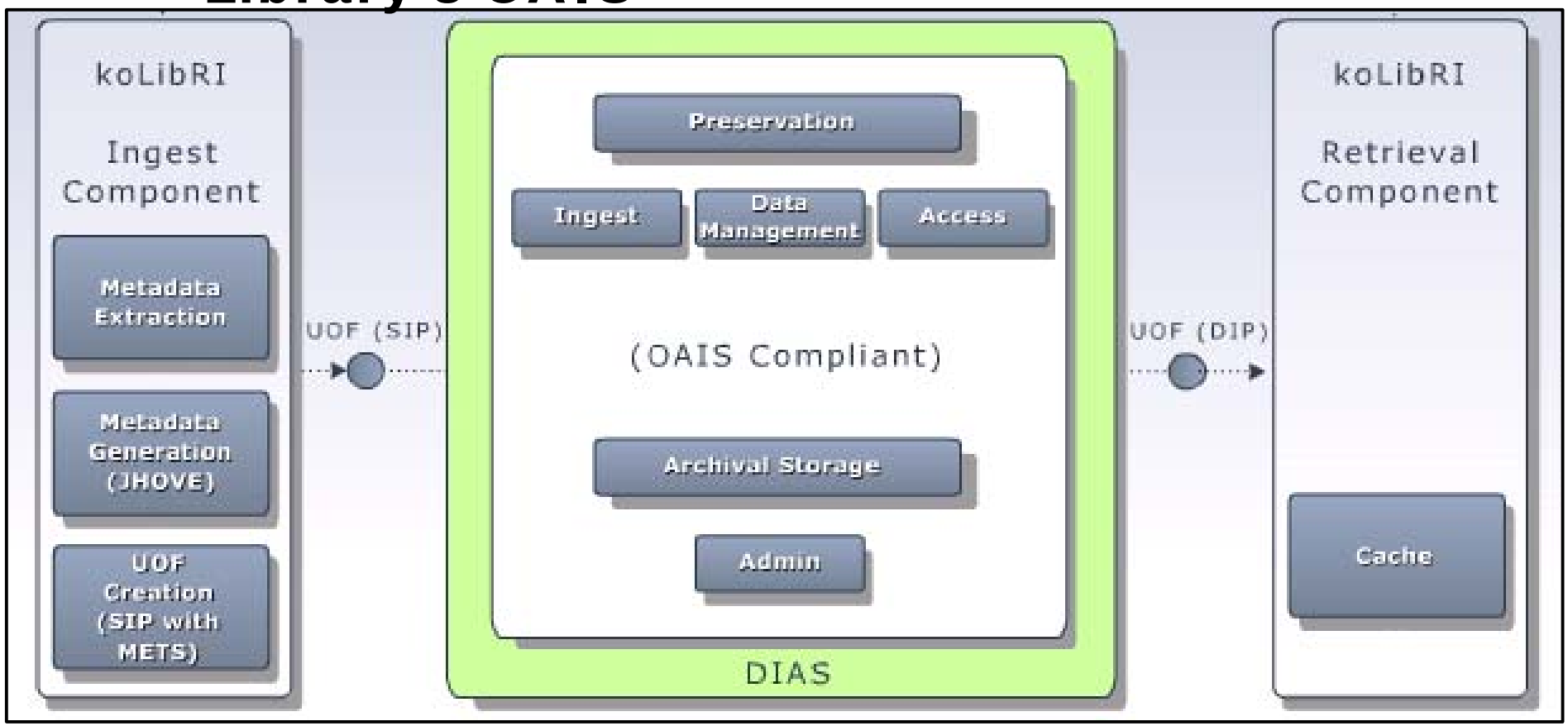
Markus Enders, Olaf Brandt
State and University Library Göttingen

DLF Fall 2006 Forum
November 10, 2006

OAIS types

- Stand-alones
- Similar systems sharing content
- Dissimilar systems sharing content

Göttingen State and University Library's OAIS



IBM-Netherlands/IBM-Germany's DIAS system provides the back end.

Cornell University Library's OAIS

Data management
- MySQL

Access
- Django

Ingest
- Java
- METS
- PREMIS
- OAI_DC
- OAI-PMH
- JHOVE
- XMLBeans
- iBatis

Archival Storage: aDORe

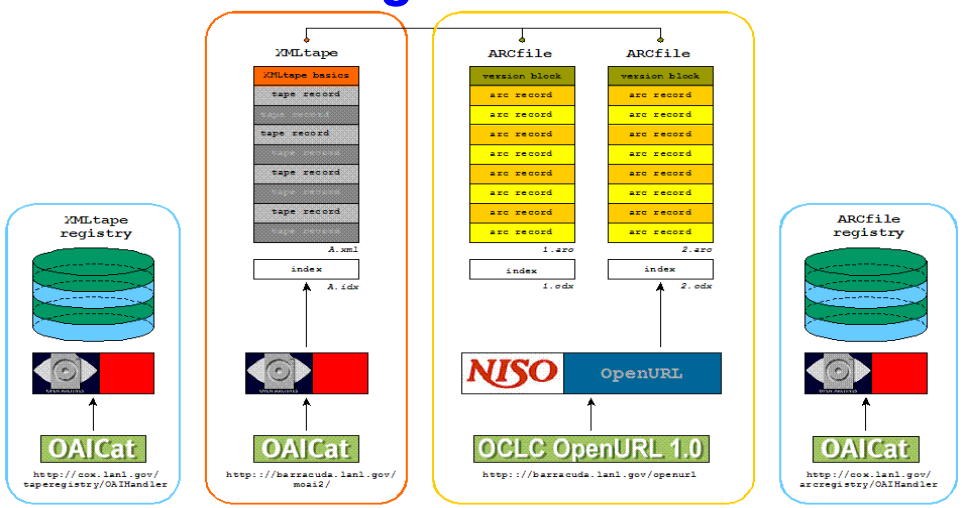
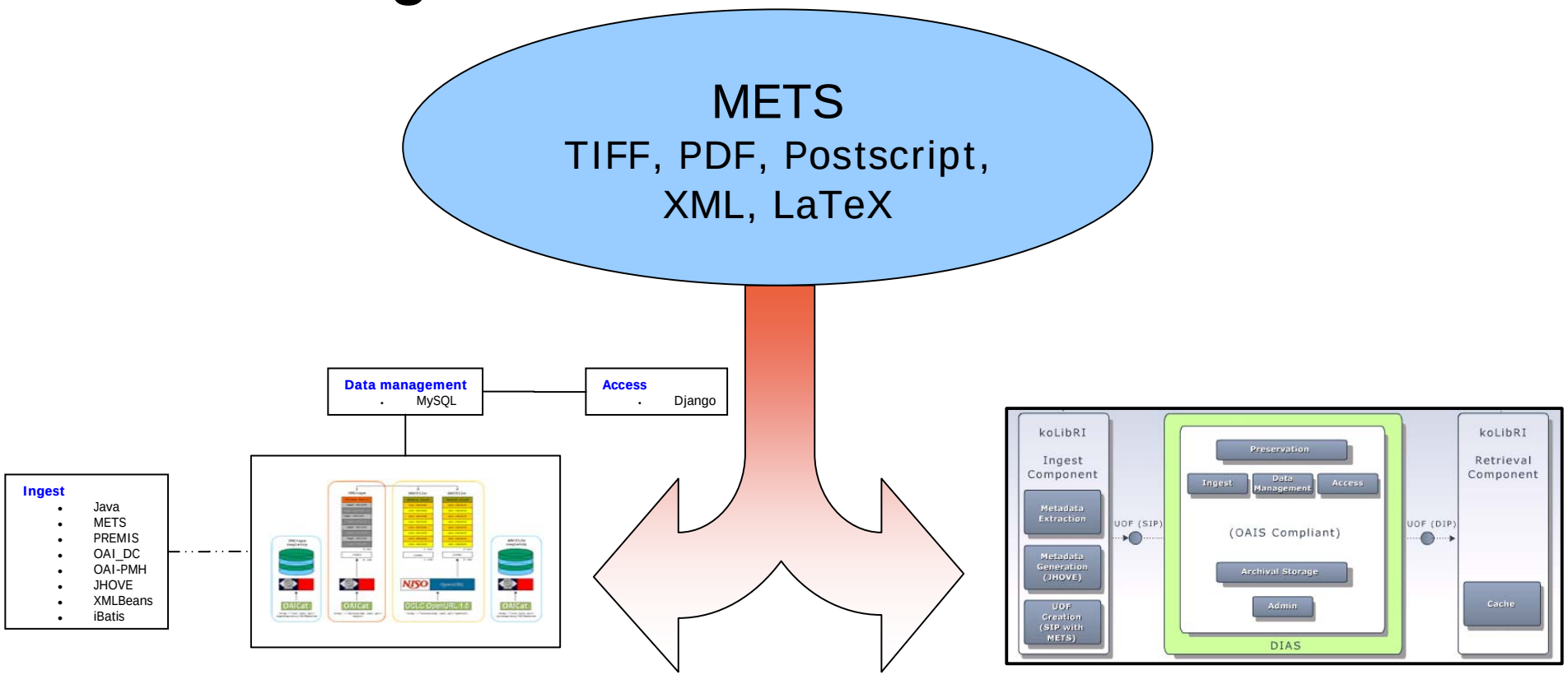


Image by Jeroen Bekaert. Copyright Los Alamos National Laboratory. Used with permission.

OAI-PMH metadata and file exchange



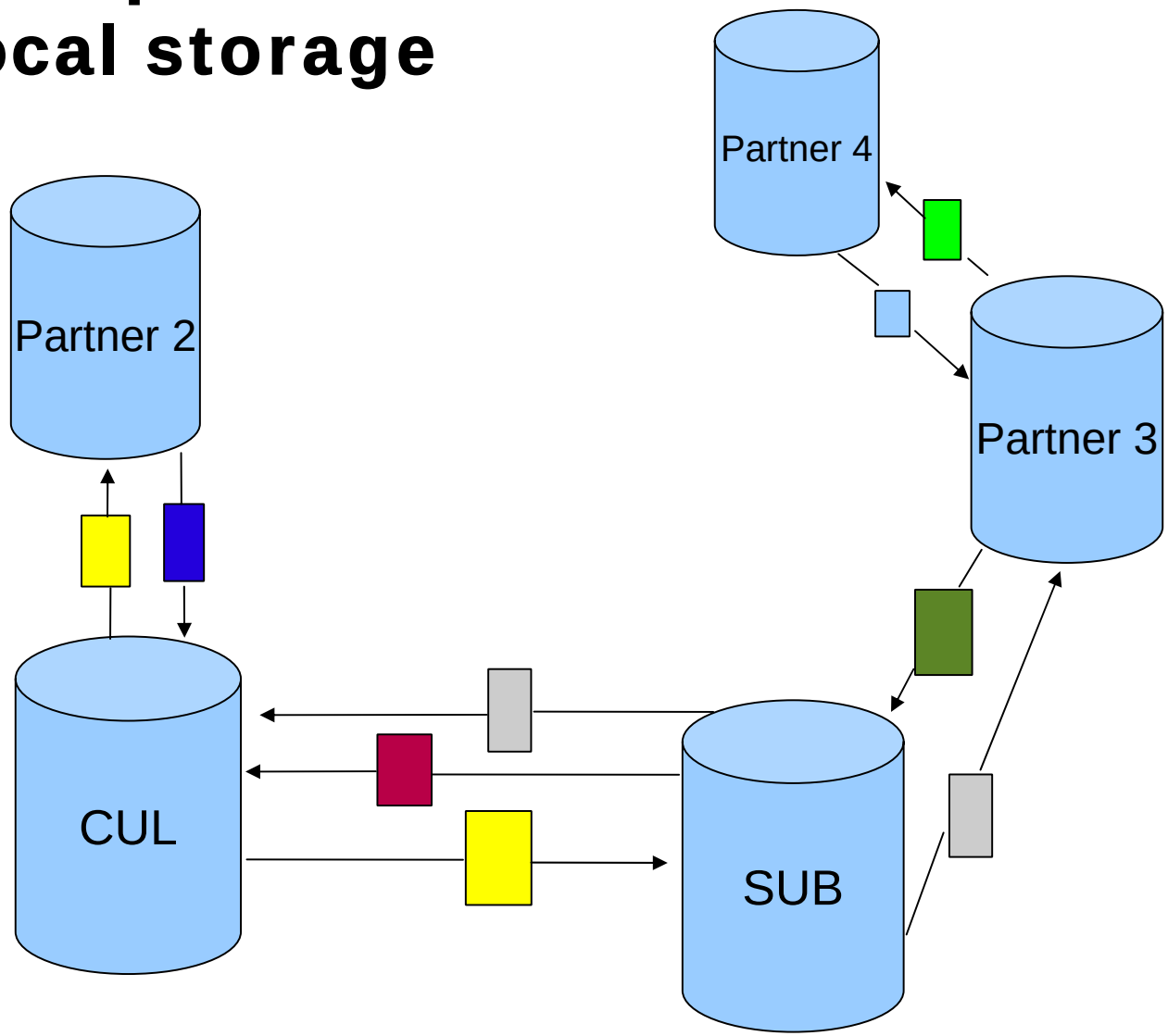
A Trade-off

Base64-encoded packages make audits simpler.

BUT

They make larger transport packages.

Partnerships: Reciprocal storage



MathArc Protocol: Purpose

Brief recap:

- **Goal:** The MathArc protocol enables different OAIS implementations to automatically ingest each others digital objects.
- MathArc is used by repositories that have agreed to partner for the purpose of sharing custodial responsibilities for those digital objects.
- The primary custodian that owns the collections being harvested controls the access and rights of each partner to those collections.

MathArc Protocol: Registry

- A **registry**, implemented as XML files stored on a private LOCKSS network, stores partners, rights and local processing information:
  - A global partners.xml file describes each partner, administrative details, and access mechanisms.
  - Files specific to each partner describe:
    - what collections that partner owns, and the rights of the other partners to harvest those collections
    - metadata describing how to harvest collections from other partners

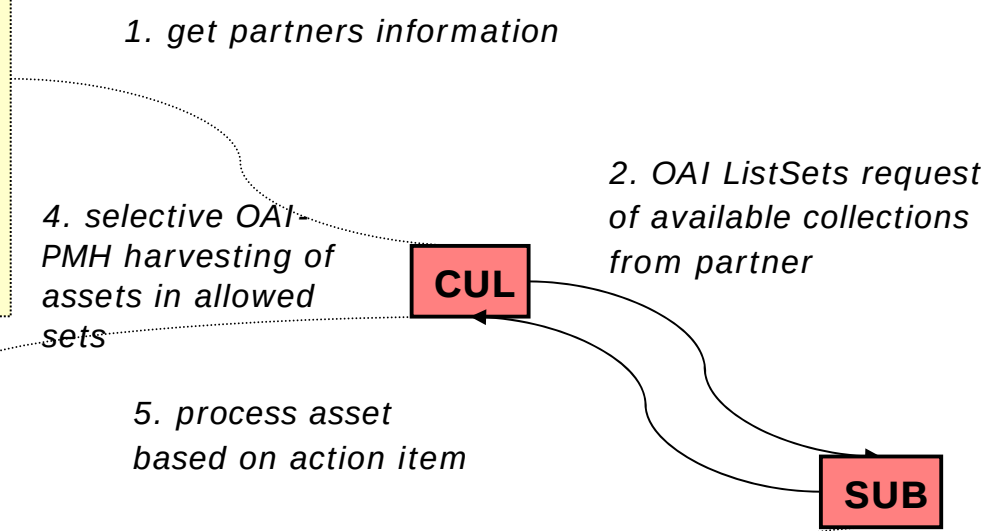
MathArc Protocol: Harvesting

- The registry drives protocol requests and responses:
  - protocol implemented around standard OAI-PMH harvesting
  - METS is the digital object metadata format
  - embedded PREMIS metadata within METS contains versioning information, which triggers custom ingest functions

Harvesting Göttingen (SUB) assets from Cornell (CUL)

(partners.xml from private LOCKSS network)

```
<partner>
   <identifierPrefix>SUB</identifierPrefix>
   <name>Niedersächsische Staats- und
Universitätsbibliothek</name>
   <service_provider_IP>
      134.76.176.172
   </service_provider_IP>
   <data_provider_URL>
      http://gdz-srv3.sub.uni-goettingen.de:8081/matharc/
   </data_provider_URL>
</partner>
```



(cul.xml from private LOCKSS network)

```
<harvestedSets>
    <harvestedSet status="active" id="SUB:Mathematica">
      <latestHarvestDate partner_id="SUB">
                  2007-08-01T12:00:00+01:00
      </latestHarvestDate>
      <action_item eventType="migration">
                  ingestAsset
      </action_item>
      <action_item eventType="updateAssetMetadata">
                  ingestAsset
      </action_item>
      <action_item eventType="replacement">
                  ingestAsset
      </action_item>
      <action_item eventType="inconsistencyDiscovered">
                  verifyAsset
      </action_item>
  </harvestedSet>
</harvestedSets>
```

(sub.xml from private LOCKSS network)

```
<accessinformation partner="SUB">
   <ownedSets>
      <ownedSet id="SUB:Mathematica">
         <allowed_partners>
            <partner_identifierPrefix status="active"
shareable="yes">
               CUL
        </partner_identifierPrefix>
         </allowed_partners>
      </ownedSet>
   </ownedSets>
</accessinformation>
```

MathArc Characteristics

- A protocol for exchanging and managing complex digital objects among custodial partners
- Preserves digital objects, not access systems
- Supports asymmetric distribution of collections among partners

http://www.library.cornell.edu/dlit/MathArc