

# Federations and Hierarchies - Networked DLs



hussein suleman  
uct cs honours 2005

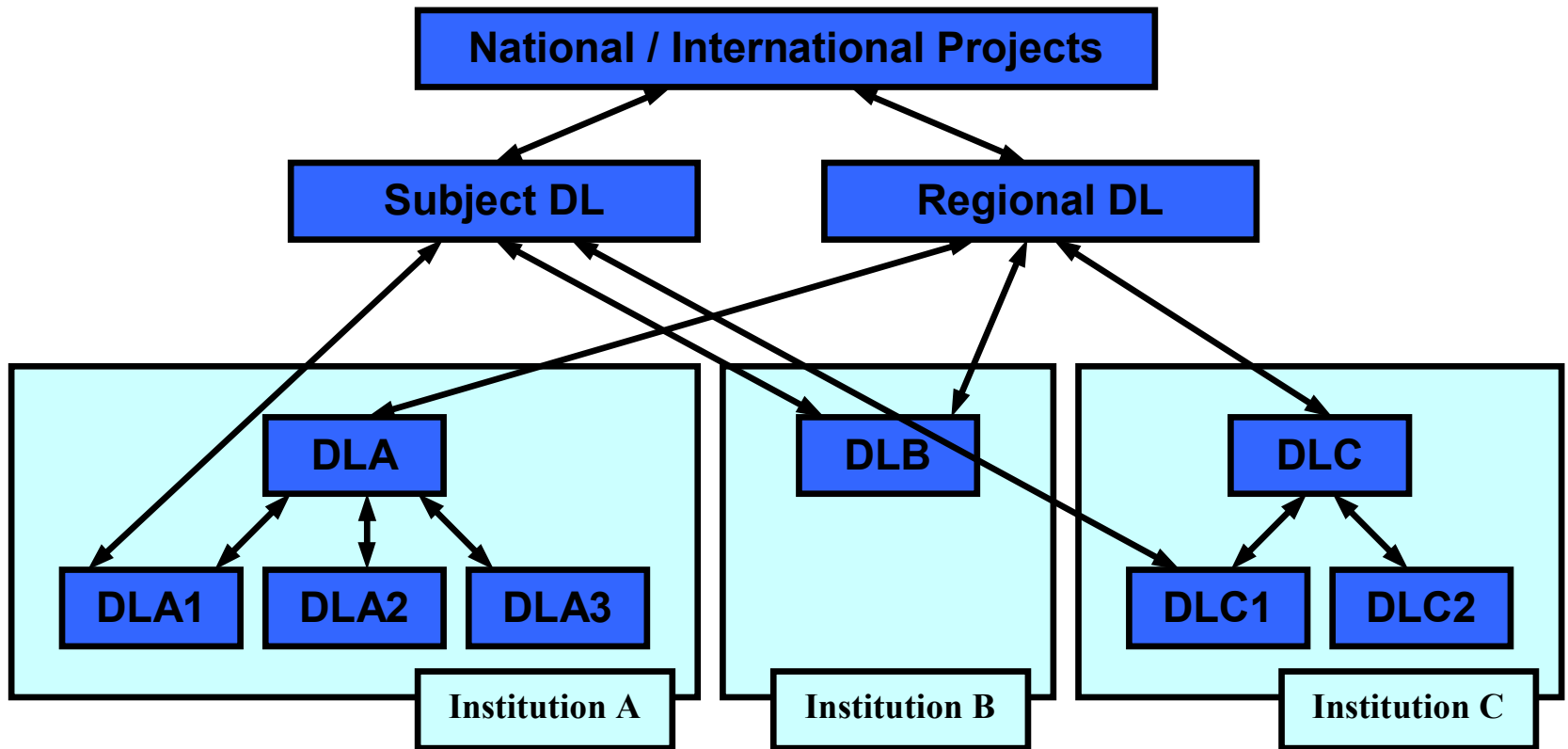
# Design Decisions

---

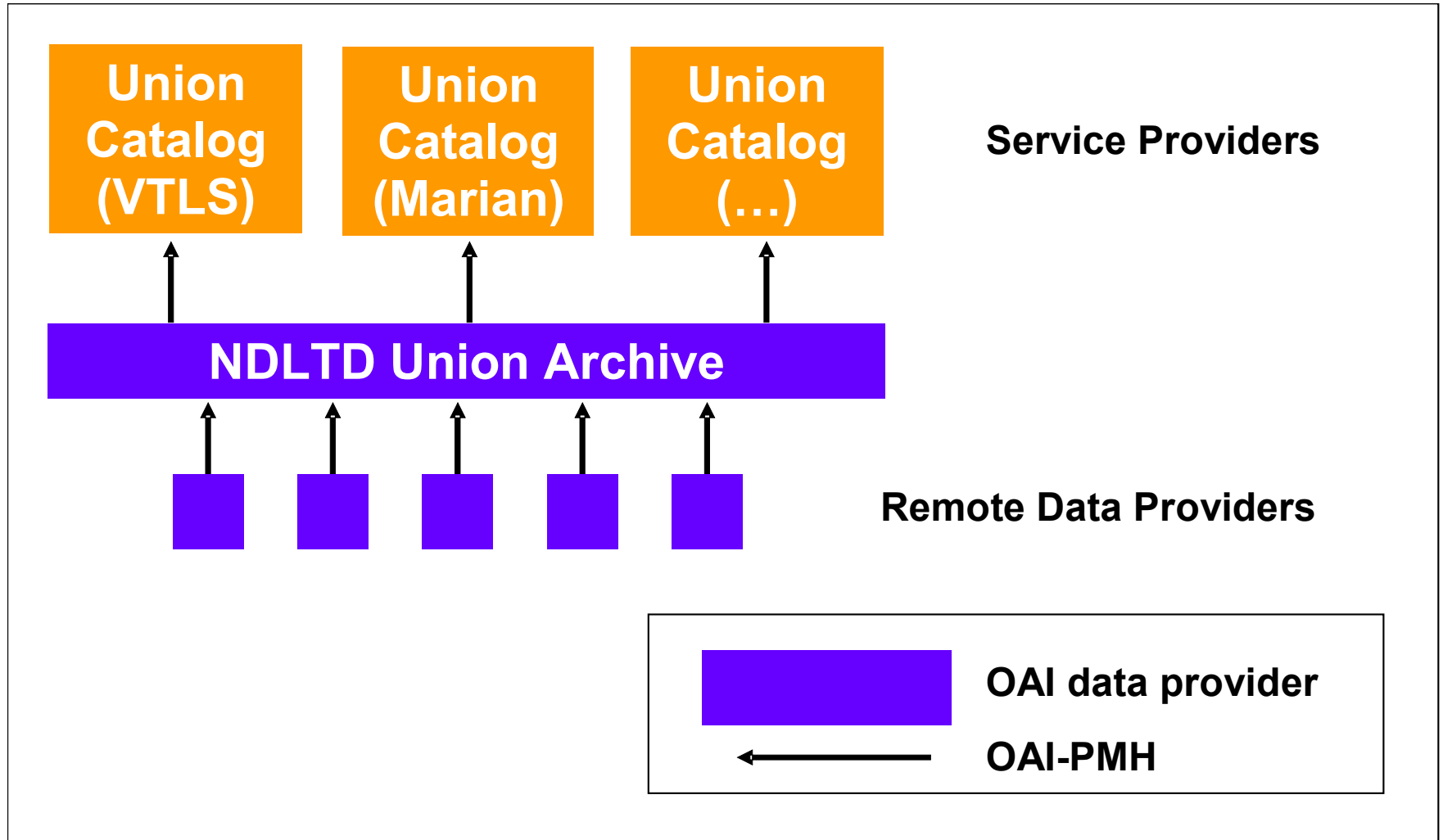
- Metadata standard(s).
- Metadata standard(s) semantics.
- Metadata encoding(s).
- Communication principles/mechanisms:
  - federation and/or harvesting
  - synchronous and/or asynchronous
  - stateless and/or stateful
- Privacy, security, etc.

# General Model

---



# NDLTD 1/2

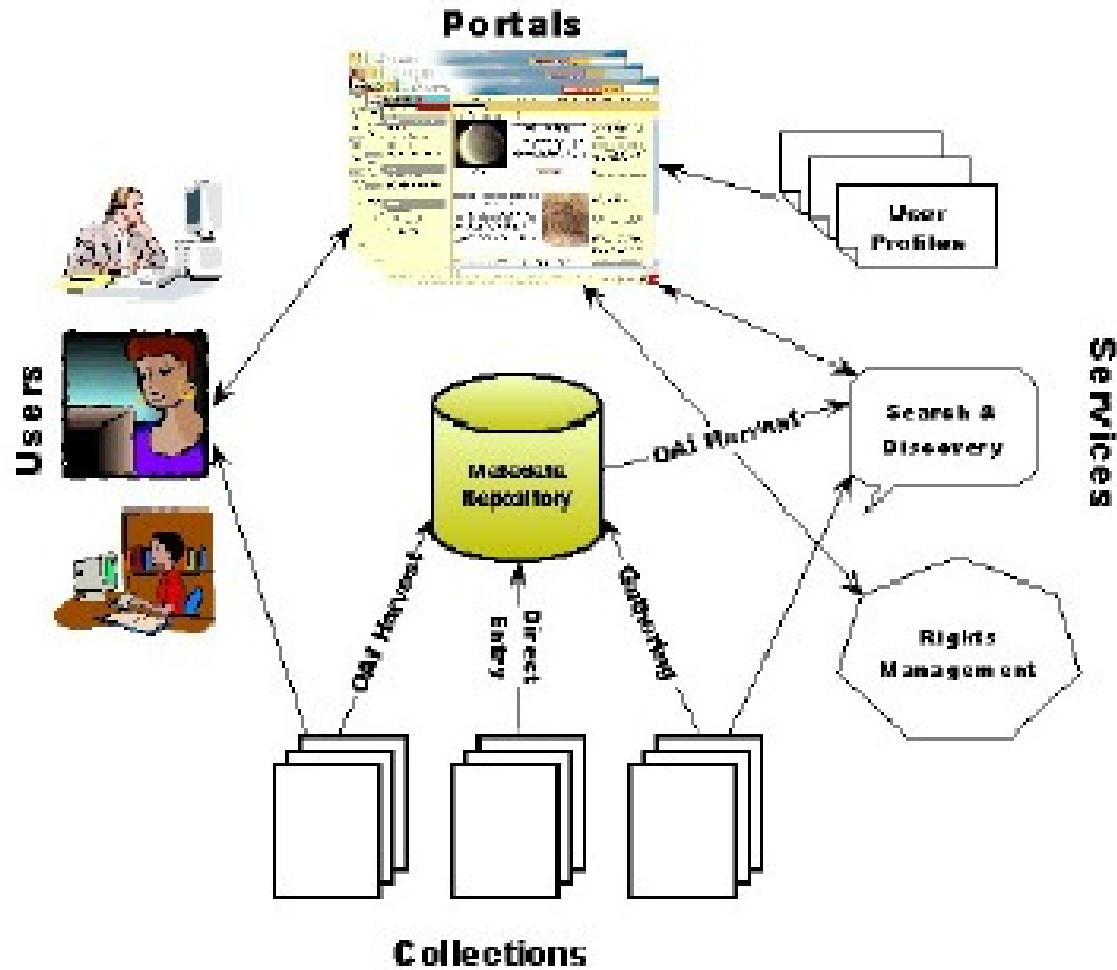


# NDLTD 2/2

---

- Multiple metadata formats: MARC, DC, ETDMS.
- Central metadata archive maintained by OCLC – metadata is harvested, gathered and submitted “manually”.
- Member sites are either single institutions (e.g., UHK) or cooperatives (e.g., PhysNet).
- Parallel services provided by:
  - MARIAN search system (VT)
  - Virtua (VTLS) – <http://www.vtls.com/ndltd>

# NSDL 1/2



Lagoze, et. al. (2002) Core Services in the Architecture of the National Digital Library for Science Education (NSDL), JCDL 2002.

# NSDL 2/2

---

- 8 metadata formats, including Qualified-DC and IMS Metadata Set.
- Central metadata archive maintained by Cornell University/US-NSF – metadata is harvested, gathered and submitted “manually”.
- Member sites are mostly subject-based digital archives (e.g., CITIDEL for computer science).
- Piecewise services provided by:
  - NSDL Core Integration Team
  - NSDL Service Partners

# RePEc

---

- Research Papers in Economics
- “Many archives == => One dataset == => Many services”
- Uses a standard (pre-XML) metadata format and static files on anonymous FTP servers (simplicity!).
- Central but mirrored metadata collection.



# Federated Resources: Appropriate Copies

---

- When multiple archives contain a copy of a resource, which one should users be directed to?
- Issues
  - cost, network speed
- SFX is a context-sensitive linking system where users are directed to local or accessible copies of resources.
  - Transparent access to appropriate copies of resources!

# Federated Service Suites:

## OpenURL

---

- Extension of SFX ideas from “appropriate copies” to “appropriate services”.
- Context-sensitive reference linking to services associated with resources.
- Based on a specific format for URLs (hence the name).
  
- Example:
  - [http://sfxserv.rug.ac.be:8888/rug?id=doi:10.1007/978-1-4020-2888-9\\_10](http://sfxserv.rug.ac.be:8888/rug?id=doi:10.1007/978-1-4020-2888-9_10)
  - (from DLib paper on OpenURL)

# Federated Metadata and DeDuping

---

- How do we reconcile multiple records in an archive that describe the same resource?
  - Keep multiple metadata records?
  - Delete least descriptive / oldest records?
  - Link records?
  
- Protocol solution: OAI has a container for record provenance
  
- Metadata solution: DC has a relation field but IMS has a relation category of fields!

# Federated Search

---

- Z39.50 / ZING
- Dienst / OpenDLib
- SDLIP (Simple Digital Library Interoperability Protocol)
  - Defines simple HTTP/CORBA-based interfaces for remote searching (esp. wrapping existing search systems)
  - Stateless/ful a/synchronous operation
  - 3 sub-interfaces
    - source metadata
    - search
    - result access

# Federated Authentication

---

- Shibboleth is a networked authentication/authorisation model.
  - Single sign-on model.
  - Access without identity!
    - Only by attributes e.g., student at UCT.
  - Privacy is user-controlled – users can decide how much information to pass on to third parties.
  - Trust relationships are established among collaborating systems out-of-band.
  - SAML records are exchanged to specify attributes (and club membership) of users.
- Security Assertion Markup Language (SAML) is an XML-based security assertion language.