

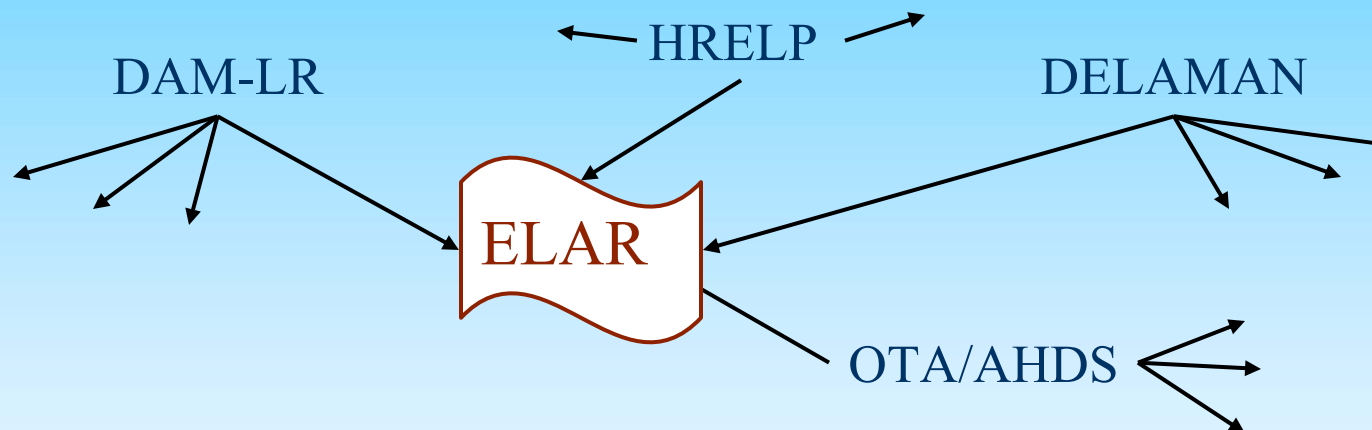


# DAM-LR as a Language Archives Federation

David Nathan

Endangered Languages Archive,  
SOAS, University of London

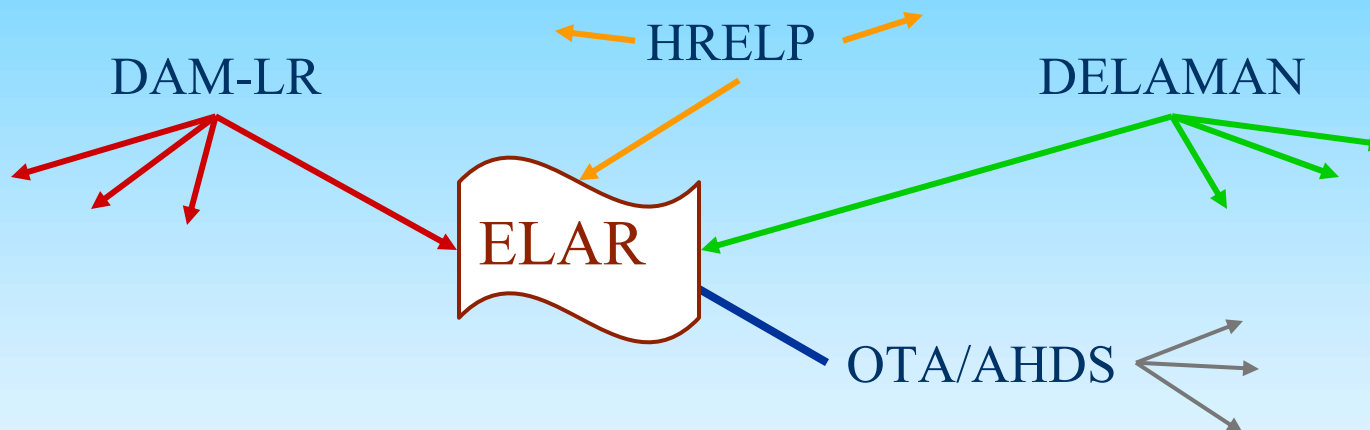
# Introduction



We find ourselves in relationships, networks, memberships of varying

- ❖ configurations
- ❖ activities, obligations, advantages, lifetimes

# Introduction



We find ourselves in relationships, networks, memberships of varying

- ❖ configurations
- ❖ activities, obligations, advantages, lifetimes

# Introduction



## ❖ Types

- ❖ subprojects
- ❖ memberships
- ❖ project partners
- ❖ commercial
- ❖ shared interests

## ❖ Other

- ❖ open/closed memberships
- ❖ shared or distinct client groups

# Why “federation”?



- ❖ term arises frequently
- ❖ has useful technical counterpart
- ❖ incentives and opportunities

# FederationT



- ❖ in the IT/information retrieval/provider environment
- ❖ why? value to users - need less knowledge to access more resources
- ❖ multiple entities provide a single and unified service
- ❖ associations of enterprises that come together to exchange information about users and resources to enable collaborations and transactions: “authenticate locally, act federally” (K Klingenstein, U Boulder,

[www.educause.edu/ir/library/powerpoint/LIVE038.pps](http://www.educause.edu/ir/library/powerpoint/LIVE038.pps))

# FederationT



## ❖ social model

- ❖ Loose governance, shared mission for collective good (eg facilitate research)
- ❖ Does not entail financial obligations

## ❖ commercial model

- ❖ Provide services of measurable value, usually for financial exchange

## ❖ hybrid model

- ❖ eg Publisher provides academic access to journals

## ... Hybrid model cont'd



- ❖ basic issue is that stakeholding including access rights are separated from the actual data provider that user interacts with.
- ❖ therefore need ways of implementing security, compliance, accountability etc, all packaged as “trust”

# FederationT



- ❖ in such systems (using LDAP, Shibboleth etc), *federation* is the mechanism by which *trust* is implemented - technical details in other talks
- ❖ however, terms *federation*, *trust* are used in special ways, hidden by the particular environments in which such systems are established (e.g. universities, libraries etc)

# Brief history



- ❖ OCLC (Ohio) 1967 - share authority files across libraries
  - ❖ issues: standards, metadata, technology, membership etc
  - ❖ now 9,000 members, 65 million records
- ❖ mid 1990s - federated databases widespread
- ❖ Dublin Core “networked discovery and retrieval [via] *federating* disparate databases” (Lynch 1998)



❖ central concept is *unification*:

❖ goals

❖ semantics

❖ standards (syntax)

❖ services

❖ (see Google etc as weak federating platform)



# Standards



- ❖ syntax: assisted by mature nature of XML, OAI, OLAC, IMDI, harvester protocols etc

# Semantics



- ❖ semantic interoperability is hard to attain (Colomb)
  - ❖ DC (/OLAC) are lowest common denominator
  - ❖ inherent nature of language resources (languages/concepts/approaches)
  - ❖ specialisations of archives
    - ❖ granularity
    - ❖ policies, e.g. protocol services
  - ❖ Schüller's "spotting" problem - 90% of resources are scattered

# Opportunities



- ❖ build on work so far, e.g.:
  - ❖ OLAC
  - ❖ GOLD
- ❖ DAM-LR partners have clear goals, clients, audiences, collection policies, and research strengths
- ❖ seek linguists' participation and investment, e.g. via agenda for recognition of intellectual value of data
- ❖ note: Depositor focus transcends unification goal
  - ❖ information managers rather than information seekers
  - ❖ see *Live Archives*

# Barriers



- ❖ Google's strengths and limitations reflect tension between unity vs. agency among info providers
- ❖ Colomb: semantic heterogeneity:
  - ❖ many sources - categories, interpretations, goals, time
  - ❖ expensive to resolve
- ❖ we lack the grounded categories of libraries and commerce
- ❖ linguistics seems to *guarantee* non-interoperability!

## ... as organisational issues



- ❖ FederationsA (*agency*)
- ❖ core issues
  - ❖ membership
  - ❖ cf. DC as lowest-common-denominator strategy
  - ❖ change
  - ❖ our field e.g. documentary linguistics
  - ❖ technologies
- ❖ need forum of negotiation, not only interoperable terms
- ❖ cf. FederationsT as broadcast of technical standards

# FederationsA to take account of:

- ❖ A. Decisions
- ❖ B. Operations
- ❖ C. Resources
- ❖ D. Services
- ❖ E. Rights and obligations

# A. Decisions



- ❖ identity, representation
- ❖ decision making
- ❖ managing relationships
- ❖ structure for managing the above

## B. Operations



- ❖ level of interdependence
  - ❖ peripheral (eg work together on standards, e.g. OLAC)
  - ❖ utility (e.g. add capabilities, eg DAM-LR)
  - ❖ process (e.g. core functions, e.g. ELAR/OTA)

## C. Resources



- ❖ resource sharing/interchange
- ❖ do activities have overt costs/value?
- ❖ need clear and binding agreements
- ❖ legislative requirements and liabilities
- ❖ utility (e.g. add capabilities, eg DAM-LR)

## D. Services



- ❖ “Federation services” (Dspace)
- ❖ e.g. support interoperability
  - ❖ creates a topology (ie. implies non-interoperability beyond federal boundaries)
- ❖ but there are many other possible services:
  - ❖ ingestion
  - ❖ training
  - ❖ publishing
  - ❖ fundraising

## E. Rights and obligations



- ❖ “Common destiny”:
- ❖ FederationA: A body which self-declares both its constituency and the willingness of its members to create, commit to, and be bound by its ongoing collective operations, decisions, and destiny

## Some issues are hidden



- ❖ where resources are not consumed or can be freely allocated
- ❖ agencies are acting consistent with “public good”
- ❖ areas where boundaries are unclear or unimportant

# Ethical and legal aspects



- ❖ Sensitivities of much of our data has implications for A, B, C, D, E
- ❖ ELAR - emphasis on collecting and implementing sensitivities and restrictions on access
- ❖ “Lossy merge” of such protocol data is not feasible

# Legal aspects



- ❖ The unification of presentation also means that there needs to be a relevant entity responsible for:
  - ❖ compliance with data protection and privacy laws
  - ❖ informing and enforcing member conformance
  - ❖ representing agencies in any legal process
  - ❖ dealing with trans-jurisdiction issues
  - ❖ reporting

# Conclusion



- ❖ FederationT requires FederationA:
  - ❖ to meet specific needs of domain and its users
  - ❖ to meet ethical and legal obligations
  - ❖ for direction and sustainability
  - ❖ to inform its data and process models
- ❖ FederationA requires FederationT:
  - ❖ to meet service goals
  - ❖ to manage its collaboration
  - ❖ to use resources effectively
  - ❖ to keep humanities technologically informed