

# *Collection description: surveying the landscape*

*New Directions in Metadata  
OCLC/SCURL Pre-IFLA Conference,  
Edinburgh, 15-16 August 2002*

**Pete Johnston**  
**UKOLN, University of Bath**  
**Bath, BA2 7AY**

cd-focus@ukoln.ac.uk  
<http://www.ukoln.ac.uk/>

---

UKOLN is supported by:

**JISC**

**re:source**



## ***Collection description: surveying the landscape***

- Collections & collection description
- Approaches to collection description
- The RSLP collection description model and schema
- Collection-level description in practice
- Issues, challenges.....

# *The resource disclosure/ discovery context*

- Library context
  - Full Disclosure : retrospective conversion, cataloguing
  - RSLP : disclosure of/access to research collections, collaborative management
  - RSLG
- HE/FE context
  - JISC Information Environment : more seamless discovery of/access to distributed resources
- Cross-domain context
  - Resource : Framework for Collections Management
  - NOF-digitise : digital content creation

# *The resource disclosure/ discovery context*

- Broader resource discovery context
  - user wants information relevant to **task/activity**
  - **content providers** exposing content through multiple services, channels
  - **service providers** “surfacing” content from multiple sources
- Technological context
  - Open Archives Initiative Protocol for Metadata Harvesting Release 2.0 (stable)
  - Web services
  - XML everywhere....
- Access, integration, sharing, collaboration....

***Collections,  
collection description &  
collection-level description***

# *What is a collection?*

- Collection
  - “an aggregation of items”
- Aggregations of, e.g.
  - natural objects: fossils, mineral samples...
  - created objects: artefacts, documents, records...
  - digital resources: documents, images, multimedia objects, data, software...
  - digital surrogates of physical objects: documents, images...
  - metadata: catalogue records, item descriptions, collection-level descriptions (!)...

# *What is a collection?*

- Various criteria for aggregation, e.g.
  - By location
  - By type/form of item
  - By provenance of item
  - By source/ownership of item
  - By nature of item content
  - ....
- Permanent, temporary
- Discrete, distributed
- Collections created with intent/purpose
  - In library, collection development policy

# *What is collection description?*

- Michael Heaney, *An Analytical Model of Collections and their Catalogues*
- Hierarchic
  - info about collection as whole, and about items (and relationships between items and whole)
- Analytic
  - info about items in collection
- Indexing
  - info derived from items in collection
- Unitary
  - info about collection as whole, not about items
  - “collection-**level** description”

## *Why collection-level description?*

- Enable collection provider to
  - manage collections
    - in collaboration with other providers
  - disclose information about collections
    - overview of otherwise uncatalogued items
    - summary where item-level detail inappropriate/unavailable

# *Why collection-level description?*

- Enable user to
  - discover/locate collections
    - physical/digital
  - select collections to explore/search on basis of summary description
    - physical/digital
  - compare collections as broadly similar objects where items heterogeneous
- Enable software agents to
  - select digital metadata collections to search on behalf of user e.g. on basis of profile/preferences
  - perform searches across multiple digital metadata collections

# ***Approaches to collection-level description***

# CLDs in archives

- “Collections” defined by **provenance** of (unique) items
  - records of organisation or individual
  - principle that value of individual record derives from context, relationships
- Archival description
  - emphasis on “multi-level” resource description
  - well-established standards for unitary and hierarchical CD
  - ISAD(G), EAD
- Established services: NRA, Archives Hub, A2A, SCAN etc

# *CLDs in museums*

- Focus on description of (unique) object
  - management more than discovery?
- But notion of “collection” is used
  - collection management
  - collection assessment/mapping
- Various criteria
  - type/form of item
  - subject
  - ownership/source
- Some CLD e.g. guides to holdings
  - little standardisation?
  - some use of EAD, Dublin Core
  - Resource Framework for Collections Mgt, Cornucopia, 24-hr Museum, SRA projects

# *CLDs in libraries*

- Focus on description of (non-unique) item
  - well-established standards (MARC, AACR2)
  - emphasis on discovery
- Until recently, collection-level description relatively informal, unstructured
- Collections defined by
  - location
  - subject
    - items potentially dispersed
    - complex relationships
- Standards
  - some use of MARC for CLD (especially in USA)
  - adoption of RSLP CD schema in RSLP programme



# *CLDs for digital resources*

- Some description of aggregates of resources
  - use of general metadata schemas (e.g. DC, GILS)
  - application-specific, protocol-specific approaches
- Evolution of approaches to creating digital collections
  - “proof of concept” (technological focus?)
  - greater attention to custodianship, use
  - focus on integration, reuse, interoperability, sustainability

– **(Cole 2002, Besser 2002)**
- Growing interest in collection-level metadata
  - IMLS Guidelines “Collections Principle” No. 2
  - JISC IE, NOF-digitise (more later!)

***The RSLP  
Collection Description  
Model and Schema***

# Research Support Libraries Programme

- Support for academic researchers
  - disclosure of collections
  - discovery of/access to collections
  - collaborative management of collections
- Collections in RSLP
  - projects **describing** primarily collections of physical items (library/archive)
  - projects also **describing** digital catalogues (which describe physical items)
    - collections of metadata records
  - projects **creating** new digital collections of collection descriptions
    - collections of metadata records

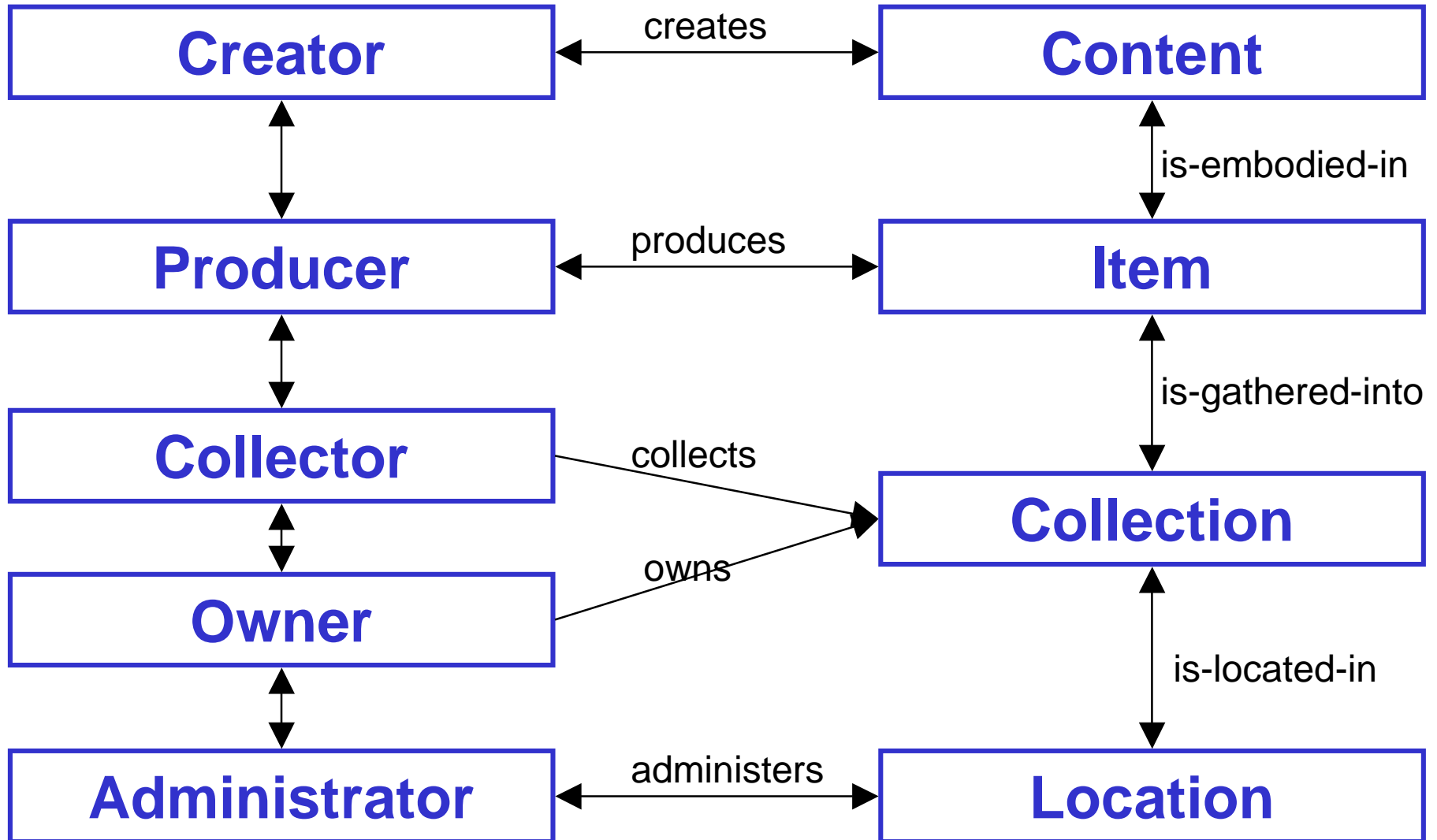
# *RSLP Collection Description project*

- Project
  - funded by RSLP, with support from OCLC, Sept 1999 – Sept 2000
  - collaboration between
    - Michael Heaney (Oxford)
    - Andy Powell, Michael Day (UKOLN)
- Aims
  - means of consistent collection description in RSLP
  - minimise duplication of effort in projects
  - simple, high-level description
  - aligned with other metadata work

# *A Model for Collections*

- Michael Heaney, *An Analytical Model of Collections and their Catalogues*
  - Implementation independent
- Based primarily on a library and archival view of ‘collection’...
  - but intended to be applicable across wide range of collection types
- Identifies
  - Entities
  - Attributes/properties of entities
  - Relationships between entities
- Functionally (IFLA FRBR) concerned with :
  - Finding (access points for discovery)
  - Identifying (sufficient for user interpretation)

# *RSLP Model (simplified view)*



# *Entities in RSLP Model*

- **Content**
  - an intellectual creation, without reference to an instantiation
- **Item**
  - a concrete realisation (physical/digital) of Content
- **Collection**
  - an aggregation of items
  - implementer to make pragmatic decisions about level of aggregation
    - “functional granularity”

# *Entities in RSLP Model*

- Location
  - “the physical or digital place where a collection is held”
  - physical location - e.g. a library or museum
    - postal address
  - digital location i.e a network service
    - URL of Web site, Z39.50 target details
  - Note: physical item at single location, but digital item available through multiple services

# *Entities in RSLP Model*

- Agent
  - “a person or corporate body related to the collection or location”
  - initiates actions
  - has rights, controls access/use
- Collector
  - gathers items into a collection
- Owner
  - has legal possession of a collection
- Administrator
  - has responsibility for the physical or electronic environment in which a collection is held

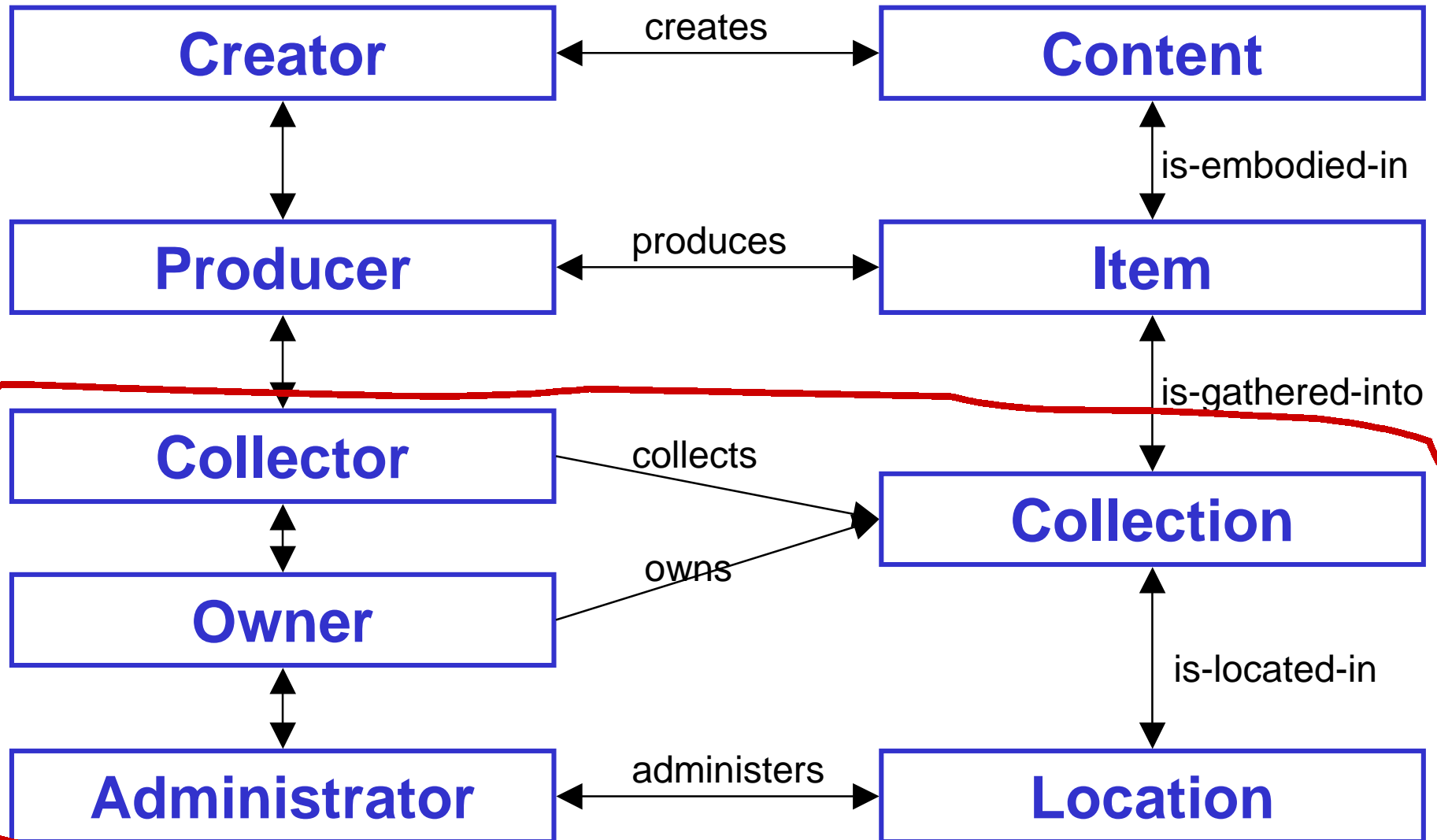
# *External relationships*

- Collection to collection
  - Has-Part, Has-Complement
  - Has-Association
  - Has-Version
- Collection to publication
  - Has-Publication
- Collection to collection description
  - Is-Described-By

# *A Schema for Collection Description*

- RSLP Collection Description Schema
  - structured set of metadata attributes
  - simple description of **subset** of entities in Heaney model
    - Collection
    - Location
    - Agents
  - attributes based on Dublin Core Element Set (including refinements) where possible
- RSLP CD schema supports creation of “unitary” collection description
- RSLP CD instance
  - set of linked descriptions of several resources

# *RSLP Schema as subset of model*



# *A Schema for Collection Description*

- Can be expressed using RDF model
  - RDF/XML syntax for serialising descriptions
- Associated list of collection types
  - terms split into four areas
    - type, curatorial environment, content, policy/usage
- Not a substitute for existing richer schemas for CLD
  - a means of creating simple, high-level descriptions for wide range of collections
- Implemented by several RSLP projects
  - Subject-based or regional databases of CLDs

# ***Collection-level description in practice***

# *Collection-level description & Full Disclosure*

- Full Disclosure Study 1999
  - emphasis on item-level description
  - recommends creating library “collections register”
- Full Disclosure Prioritisation Study 2002
  - CLD as “essential first step in identifying priorities for more detailed retrospective conversion, cataloguing and documentation work” (sect 2.8)
- CLD not replacing item-level description
  - But pragmatic means for mapping resources, informing prioritisation

# *Collection-level description & NOF-digitise*

- NOF-digitise
  - £50m content creation programme
  - Supporting strategy for social inclusion
  - Digitised objects
  - Learning materials
- Projects creating digital collections
- CLD as mechanism for disclosure/discovery
  - NOF-digi technical standards recommend use of RSLP CD Schema
  - Use in NOF-digitise portal
  - (Re) use in other services



# *Collection-level description & the JISC Information Environment*

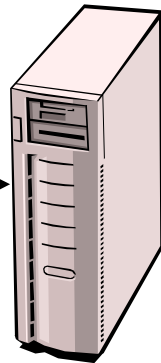
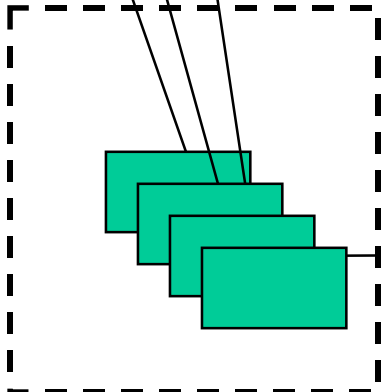
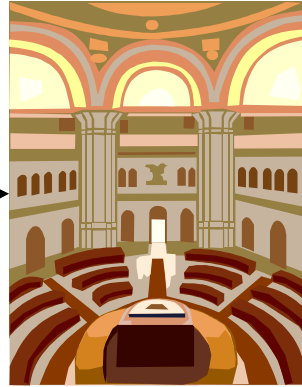
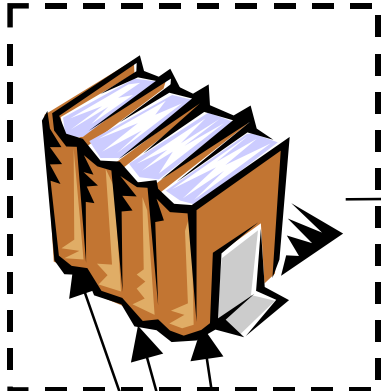
- Content made available as collections
  - various content providers
- Physical collections
  - of physical resources (e.g. books, journals)
- Digital collections
  - of digital resources (texts, images, multimedia objects, software, datasets, “learning objects” etc)
  - of digital metadata records
    - describing **physical items, digital items, physical collections**
    - metadata record contains **identifier/locator** of resource
- Users access content through services

**Collection of physical items**

**Physical location**

**Physical service**

**Physical services make physical collections available at physical locations**

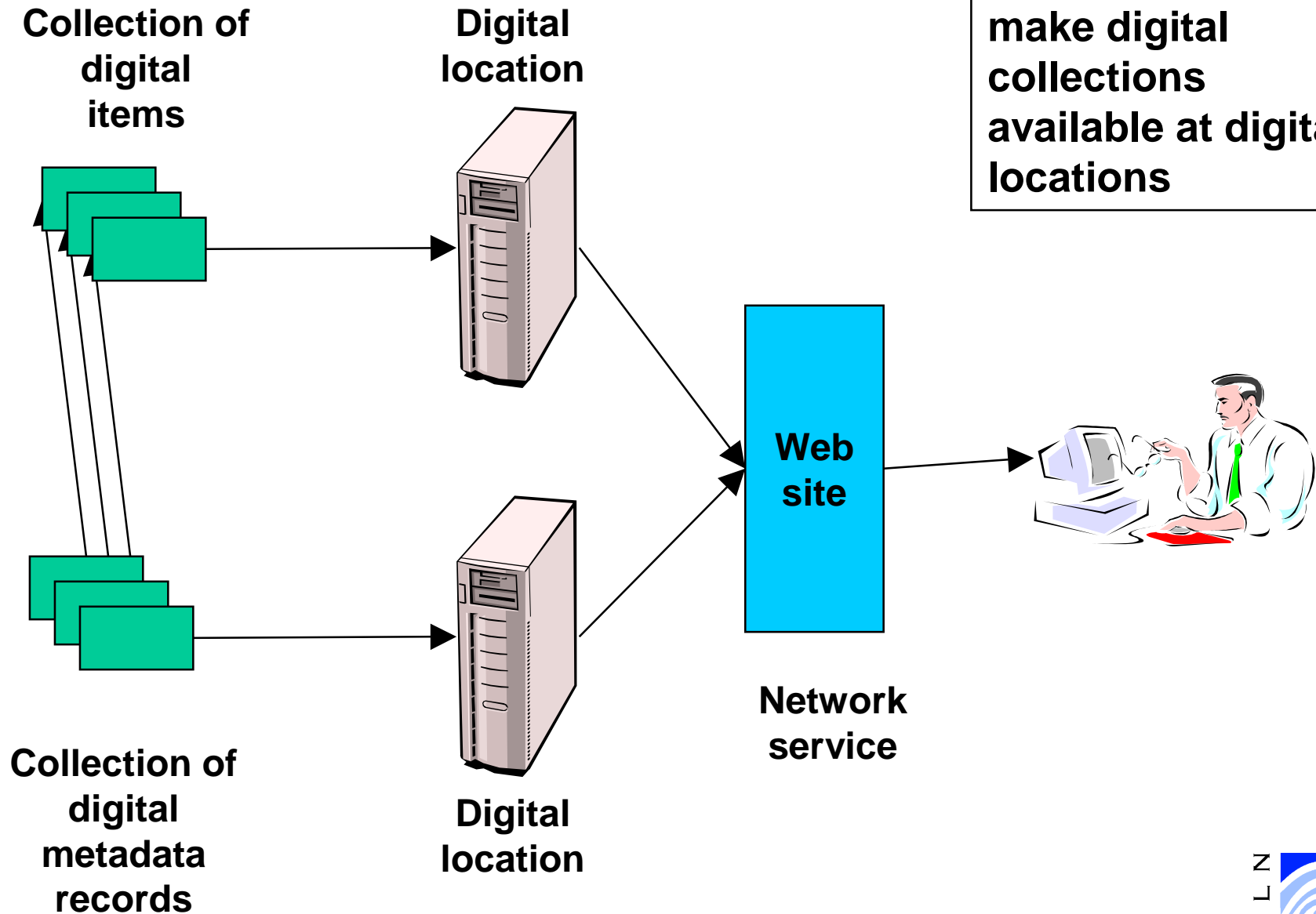


**Collection of digital metadata records**

**Digital location**

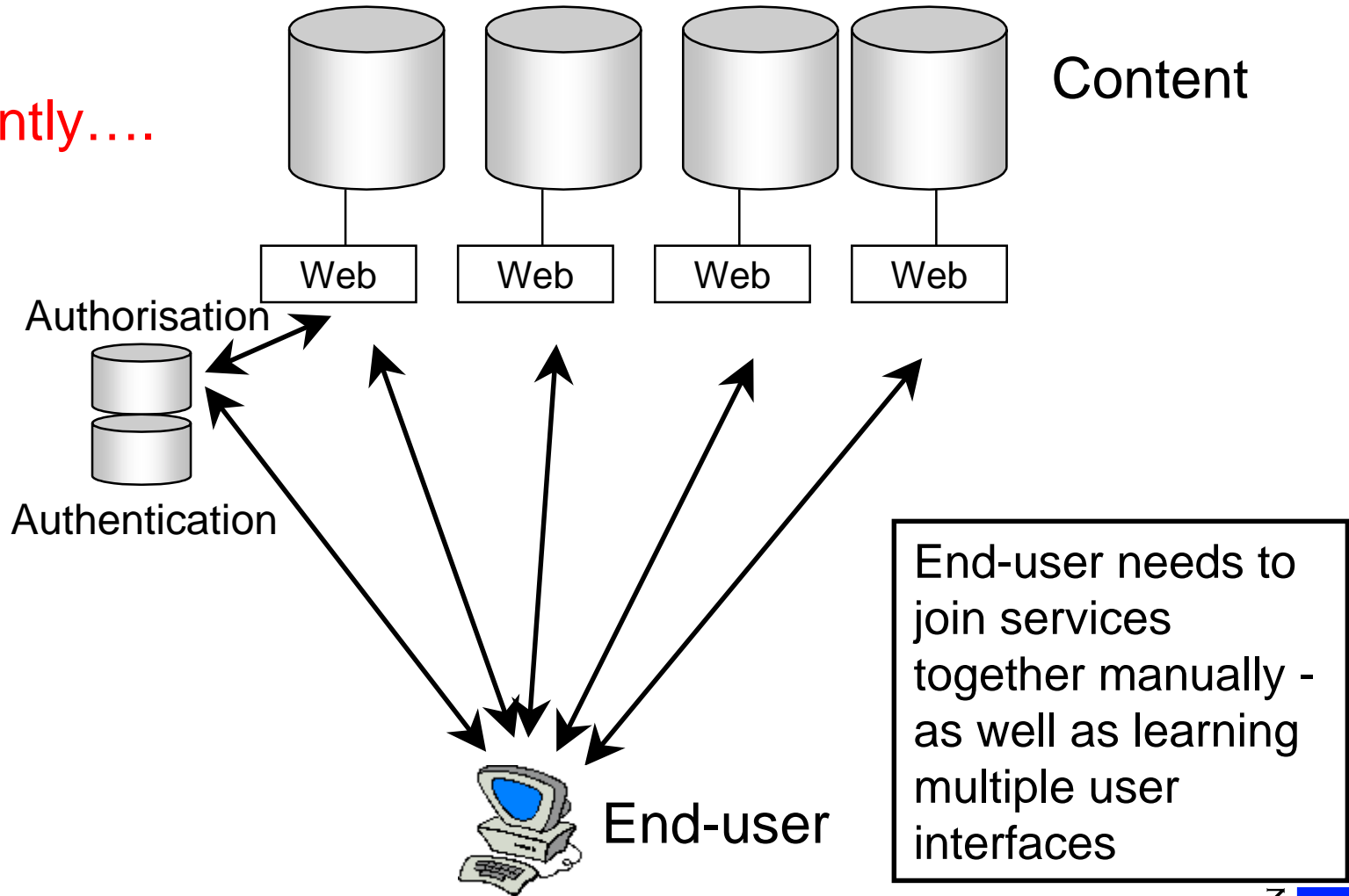
**Network service**

**Network services  
make digital  
collections  
available at digital  
locations**



# Using Collections in the JISC Information Environment

Currently....



# *Using Collections in the JISC Information Environment*

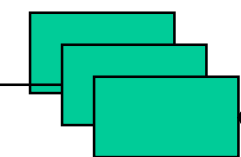
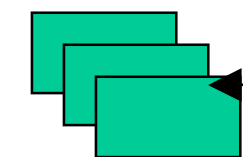
- HTML Web sites
  - Aimed at human reader not software tool
- Different user interfaces
- Different metadata schemas
- Researcher “joins up” services manually
  - Merging results requires manual copy/paste/edit
- The portal solution
  - Network service providing **single** point of access to **range** of heterogeneous network services
  - aim to be task/user-centred

# *The portal challenge*

- Content providers
  - disclose collections of **metadata records** about content through structured network **services**
- Portal needs
  - to find/identify **relevant** content collections
    - What collections available?
  - to access metadata records through **appropriate** structured network service
    - What network services available for collection?
- Service registry
  - Database of collection descriptions, service descriptions

# Collections, metadata, network services

Collection available via multiple network services



Collection of digital or physical items

Collection of digital metadata records

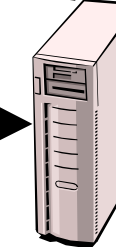
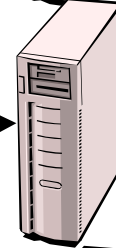
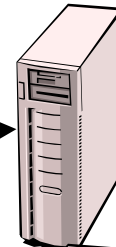
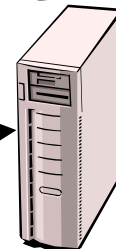
Web site

OAI repository

Z39.50 target

RSS channel

SOAP receiver



*Harvest via OAI-PMH*

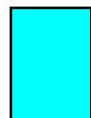
*Search/retrieve via Z39.50*

*Alert via RSS/HTTP*

*operations via SOAP*



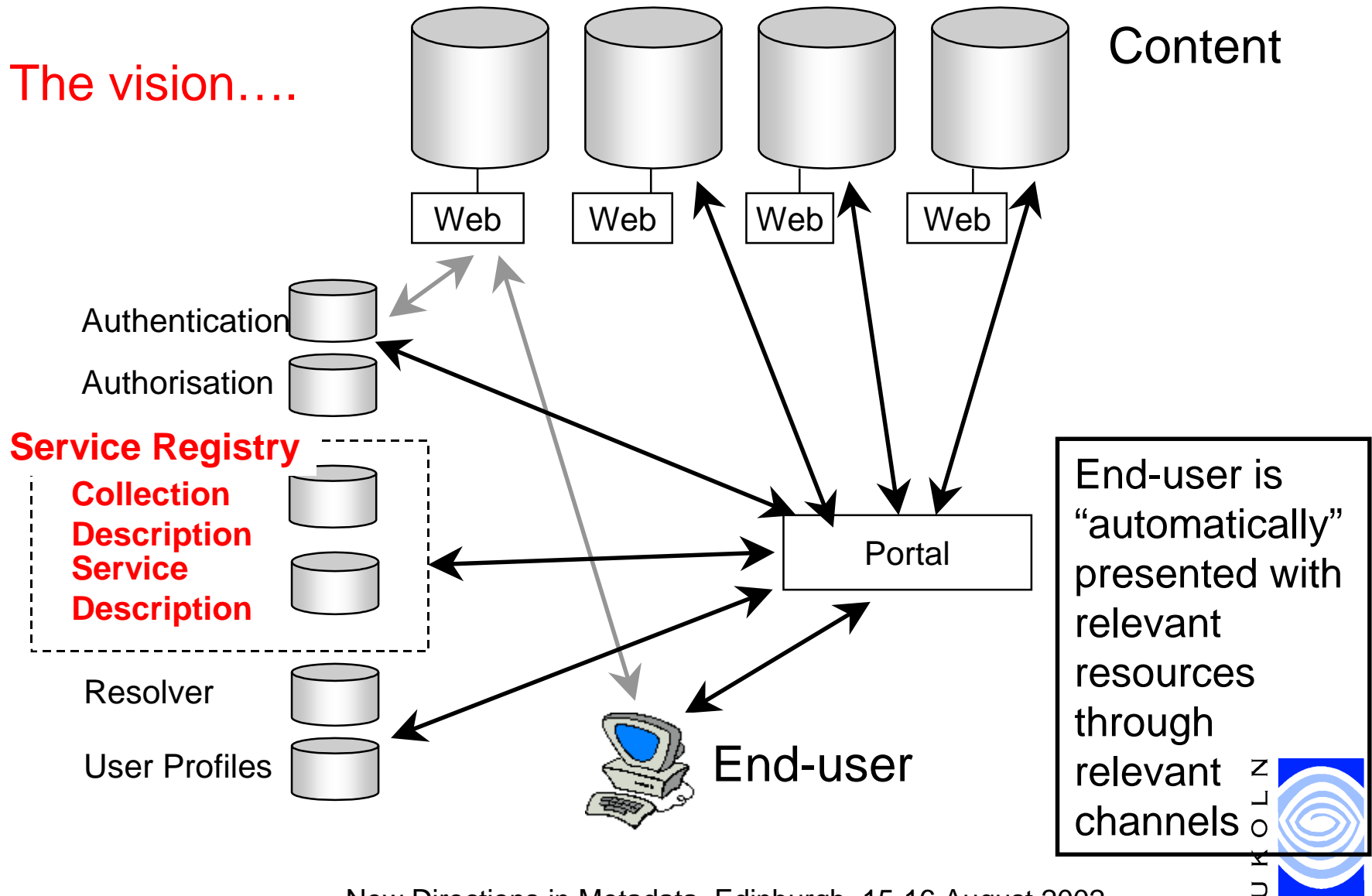
unstructured network service



structured network service

# The service registry in the Information Environment

The vision....



# *Surveying the landscape*

# Surveying the landscape

- CLDs as resource discovery metadata
  - “We’ve created this incredible constellation of collections, of pools of information.... And people can’t find which pool to look in”

(Lynch, 2002)
- CLDs support “**survey of information landscape**”
  - “to identify areas rather than specific features - to identify rainforest rather than to retrieve an analysis of the canopy fauna of the Amazon basin”

(Heaney, 2000)
- The “navigator” of the landscape may be a human researcher or a software tool

# *Issues, challenges....*

- CLD not a substitute for item-level description
  - complementing item-level discovery
  - enabling item-level discovery (JISC IE)
- What is a collection?
  - functional granularity, even-ness across domains?
- Collections, catalogues and services
- Collections and items
- Intra-domain, cross-domain discovery
  - access points for CLDs (subject, strength etc)
- CLD and (re)use: purpose, audience
- Rights issues: metadata records, resources
- Does CLD meet the requirements of
  - information managers? information users?

# *Acknowledgements*

UKOLN is funded by Resource: the Council for Museums, Archives and Libraries, the Joint Information Systems Committee (JISC) of the UK higher and further education funding councils, as well as by project funding from the JISC and the European Union.

UKOLN also receives support from the University of Bath where it is based.

<http://www.ukoln.ac.uk/>

# References

Andy Powell, *Collection Level Description: A Review of Existing Practice* (Aug 1999)

<<http://www.ukoln.ac.uk/metadata/cld/study/>>

Michael Heaney, *An Analytical Model of Collections and their Catalogues* (Jan 2000)

<<http://www.ukoln.ac.uk/metadata/rsdp/model/amcc-v31.pdf/>>

RSLP *Collection Description project* (Sep 1999)

<<http://www.ukoln.ac.uk/metadata/rsdp/>>

Pete Johnston and Bridget Robinson, *Collections and Collection Description* (Jan 2002)

<<http://www.ukoln.ac.uk/cd-focus/briefings/bp1/bp1.pdf/>>

Howard Besser, "The Next Stage: Moving from Isolated Digital Collections to Interoperable Digital Libraries", *First Monday*, Vol 7 No 6 (June 2002)

<[http://firstmonday.org/issues/issue7\\_6/besser/index.html](http://firstmonday.org/issues/issue7_6/besser/index.html)>

# References

Tim Cole, "Creating a Framework of Guidance for Building Good Digital Collections", *First Monday*, Vol 7 No 5 (May 2002)

<[http://firstmonday.org/issues/issue7\\_5/cole/index.html](http://firstmonday.org/issues/issue7_5/cole/index.html)>

Clifford Lynch, "Digital Collections, Digital Libraries, and the Digitization of Cultural Heritage Information", *First Monday*, Vol 7 No 5 (May 2002)

<[http://firstmonday.org/issues/issue7\\_5/lynch/index.html](http://firstmonday.org/issues/issue7_5/lynch/index.html)>

Digital Library Forum, *A Framework of Guidance for Building Good Digital Collections*. IMLS. (November 2001)

<[http://www.bl.uk/concord/pdf\\_files/fdigpriorityfinal.pdf/](http://www.bl.uk/concord/pdf_files/fdigpriorityfinal.pdf/)>

*Full Disclosure Prioritisation Study* (2002)

<[http://www.bl.uk/concord/pdf\\_files/fdigpriorityfinal.pdf/](http://www.bl.uk/concord/pdf_files/fdigpriorityfinal.pdf/)>

Andy Powell and Liz Lyon, *The JISC Information Environment Architecture*

<<http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/>>