

# Collection Description, Dublin Core, and the NISO Metasearch Initiative

Using the IESR: What's in it for you?  
London  
Wednesday 12 January 2005

**Pete Johnston**  
Research Officer, UKOLN, University of Bath  
p.johnston@ukoln.ac.uk

UKOLN is supported by:

**JISC**



Museums, Libraries and  
Archives Council



UNIVERSITY OF  
**BATH**

[www.bath.ac.uk](http://www.bath.ac.uk)



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# Collection Description, Dublin Core & NISO Metasearch

- The NISO Metasearch Initiative
- Metasearch, Collections & Services
- Collection Description
  - Service Description



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# The discovery problem

- User wants to access/use items from multiple content providers
- User has to discover, access and interact with multiple presentation services
  - different user interfaces
  - different authentication/access requirements
  - manual merge of human-readable results (HTML)



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# The Metasearch solution

- The provision of "Metasearch" services that
  - enable user to search across the metadata databases of multiple content providers from a single interface
  - manage multiple result sets and present to user
  - manage authentication/access
  - real-time cross search and/or harvesting
- Seamless (to the user) discovery of and access to heterogeneous, distributed resources!



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# Metasearch services work, but...

- User requires resources from more content providers
  - Limited use of standards-based search interfaces
  - Many proprietary APIs
  - Some "screen scraping" (parsing of HTML)
- Problems for service providers
  - fragile, susceptible to changes by content provider
  - labour-intensive, scalability issues
  - duplication of effort (e.g. info about services)
- Problems for content providers
  - efficiency/effectiveness of search
  - access management, logging etc
  - branding/IPR/presentation of results



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# The NISO Metasearch Initiative

- Response to content/service provider concerns
- Bring together content providers, system vendors, library service providers, standards developers
  - "To identify, develop, and frame the standards and other common understandings that are needed to enable an efficient and robust information environment"
- Task Group 1: Access Management
- Task Group 2: Collection (& Service) Description
- Task Group 3: Search & Retrieval



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# Metasearch, Collections and Services



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# Functional Model: "Surveying the landscape"

- Agent
  - "Enters" information landscape
    - A "default" set of collections, based on information about the agent
  - "Surveys" landscape
    - Modifies landscape by adding/removing collections, based on information about the collections
  - "Discovers" items of interest within collections
    - "Drills down" into selected collections

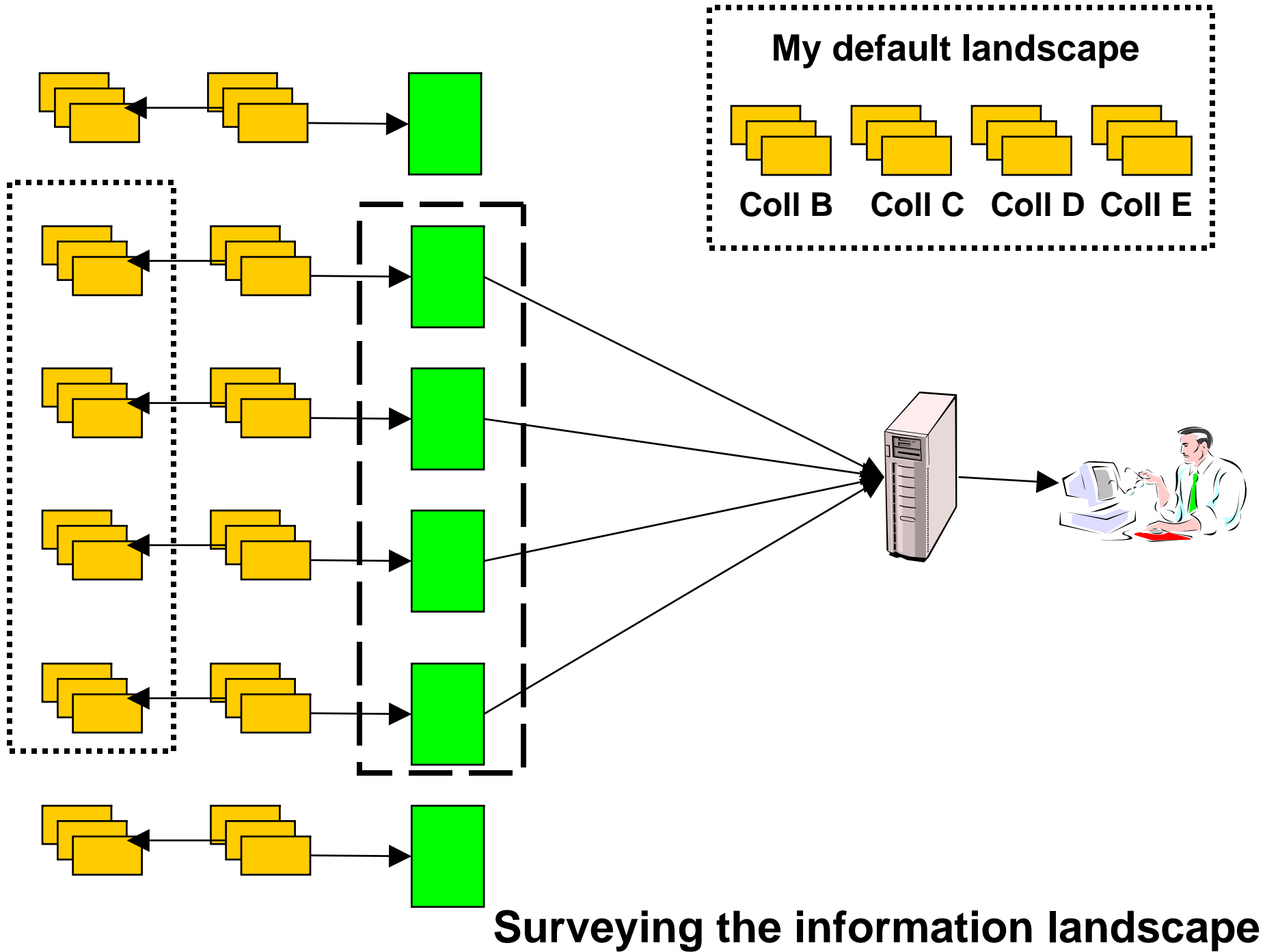
<http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/functional-model/>

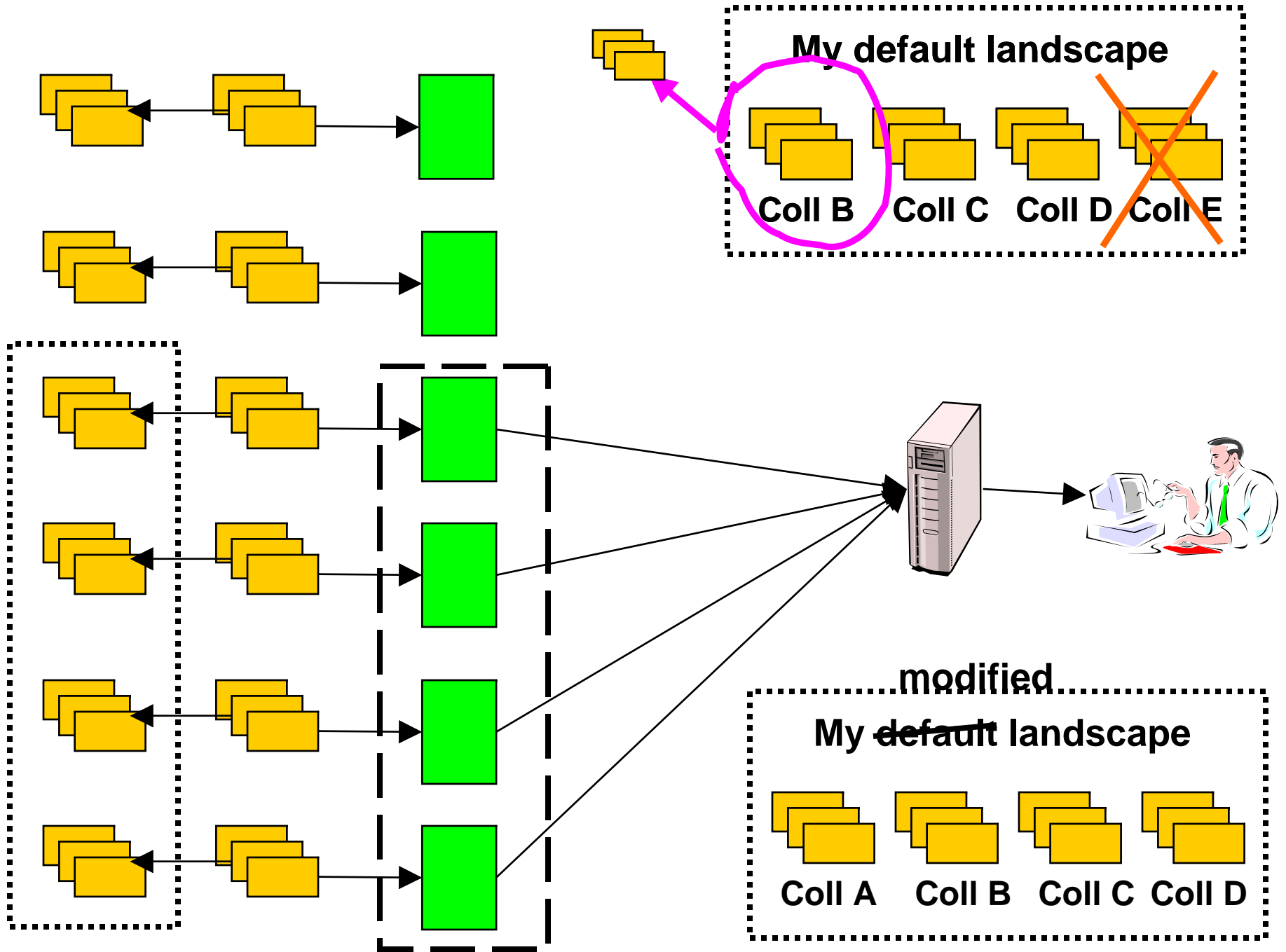
<http://ccinterop.cdjr.strath.ac.uk/>



UKOLN

A centre of expertise in digital information management





# Functional Requirements

- Allow an agent to
  - **Discover** collections of potential interest
  - **Identify** a collection
  - **Select** one or more collections from amongst a number of discovered collections
  - **Identify** the services that provide access to the collection
  
  - **Select** a service with which to interact
  - **Interact** with service
    - Subject to "knowledge" of interface semantics

**Collection  
description**

**Service  
description**



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# Collections and Services

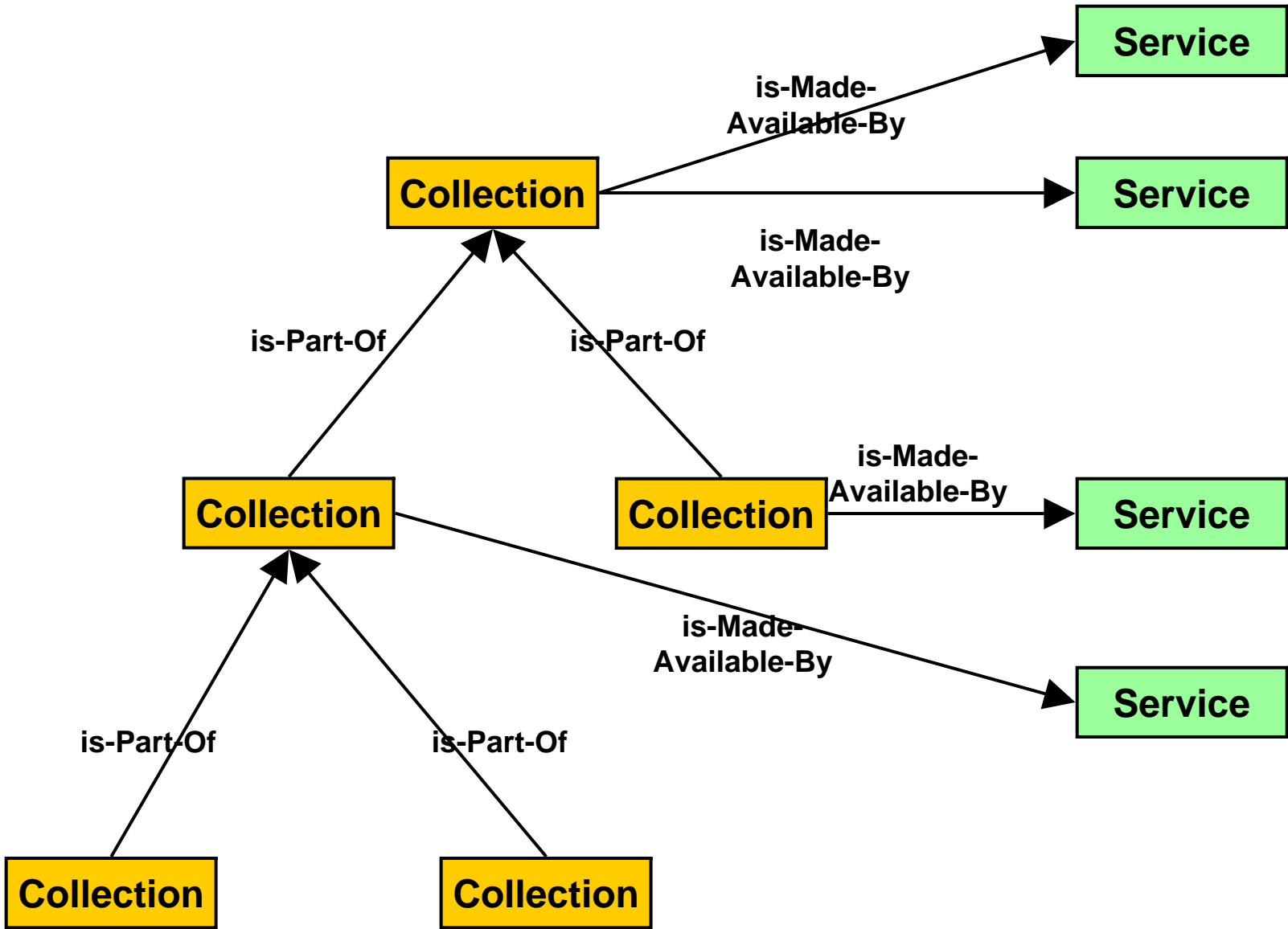
- A **collection** *is-made-available-by* zero or more **services**
- A **service** *makes-available* exactly one **collection**
- A **collection** *is-part-of* zero or more (super-) **collections** (parent)
- A **collection** *has-parts* zero or more (sub-) **collections** (child)



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management



# Collection Description & Service Description

- NISO MI TG2 specifying metadata schemas for describing collections & services
  - Data model
  - Metadata semantics
  - Syntax(es) for representation and data exchange
  - Guidelines for use
- Should build on/reuse existing work where possible
- Make recommendations for future work
- N.B. NISO MI TG2 is **not**
  - building a service; or
  - specifying the architecture within which a service might operate
  - specifying the protocols for the exchange of collection/service metadata



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# Collection Description



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# DC Collection Description Working Group

- Established ~2001, active 2003 -
- Provide forum for sharing information
- Develop a DC Application Profile for collection-level description
  - Specification of how DC (and other) properties are used for describing collections
- Informed by experience of RSLP CD implementers and other CLD initiatives
  - RSLP projects, TEL, JISC IESR, IMLS, others



UKOLN

<http://dublincore.org/groups/collections/>

A centre of expertise in digital information management

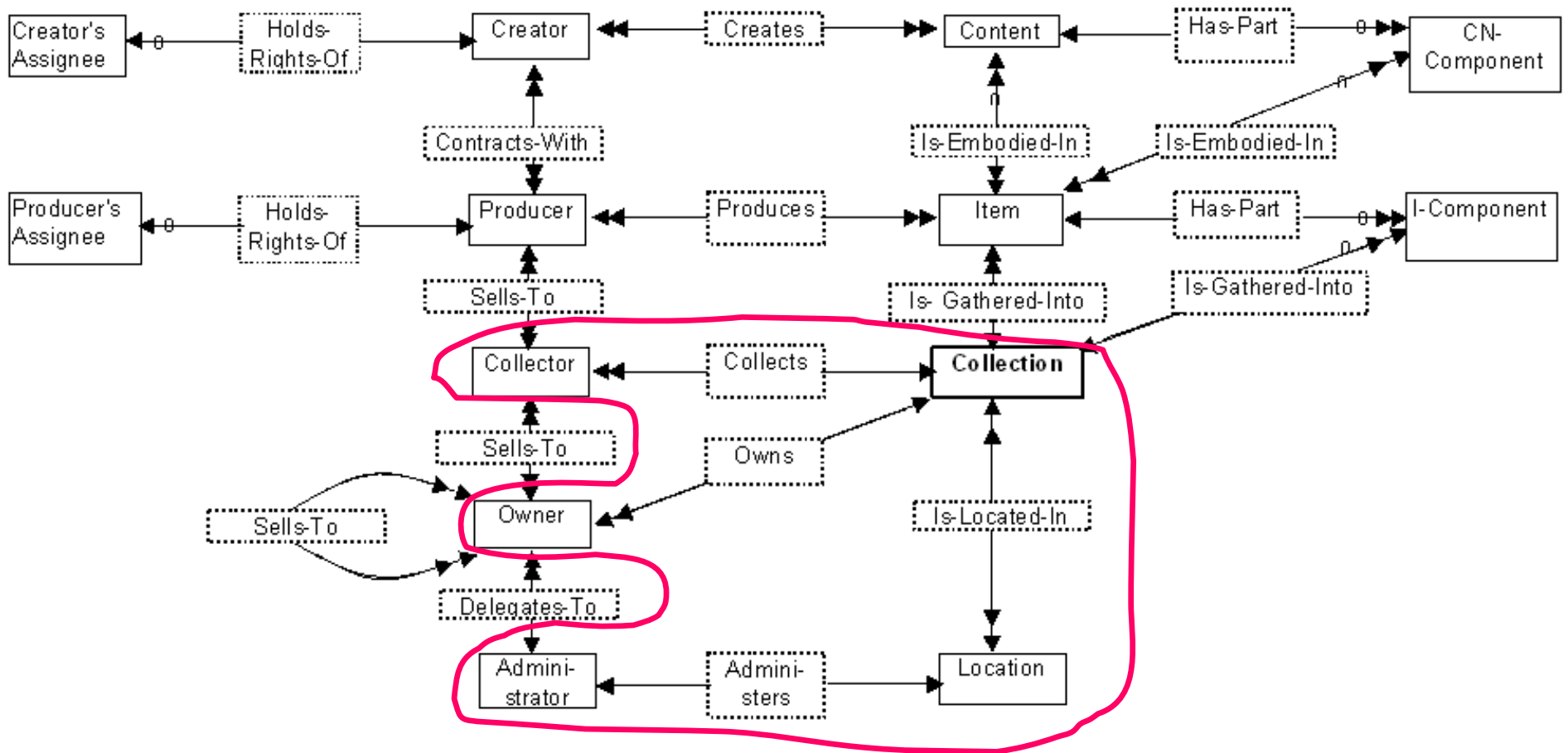
# RSLP Collection Description

- Entity-Relation **model** (Michael Heaney, University of Oxford)
  - Implementation independent
  - Intended to be applicable to wide range of collections
  - Informed by IFLA FRBR approach as well as existing descriptive standards
- DC-based **metadata schema** (Andy Powell, UKOLN)
  - Expresses subset of RSLP model
  - Simplification of model
- Significant influence on other initiatives
  - **But** concerns over status, ownership, visibility, persistence, maintenance, etc

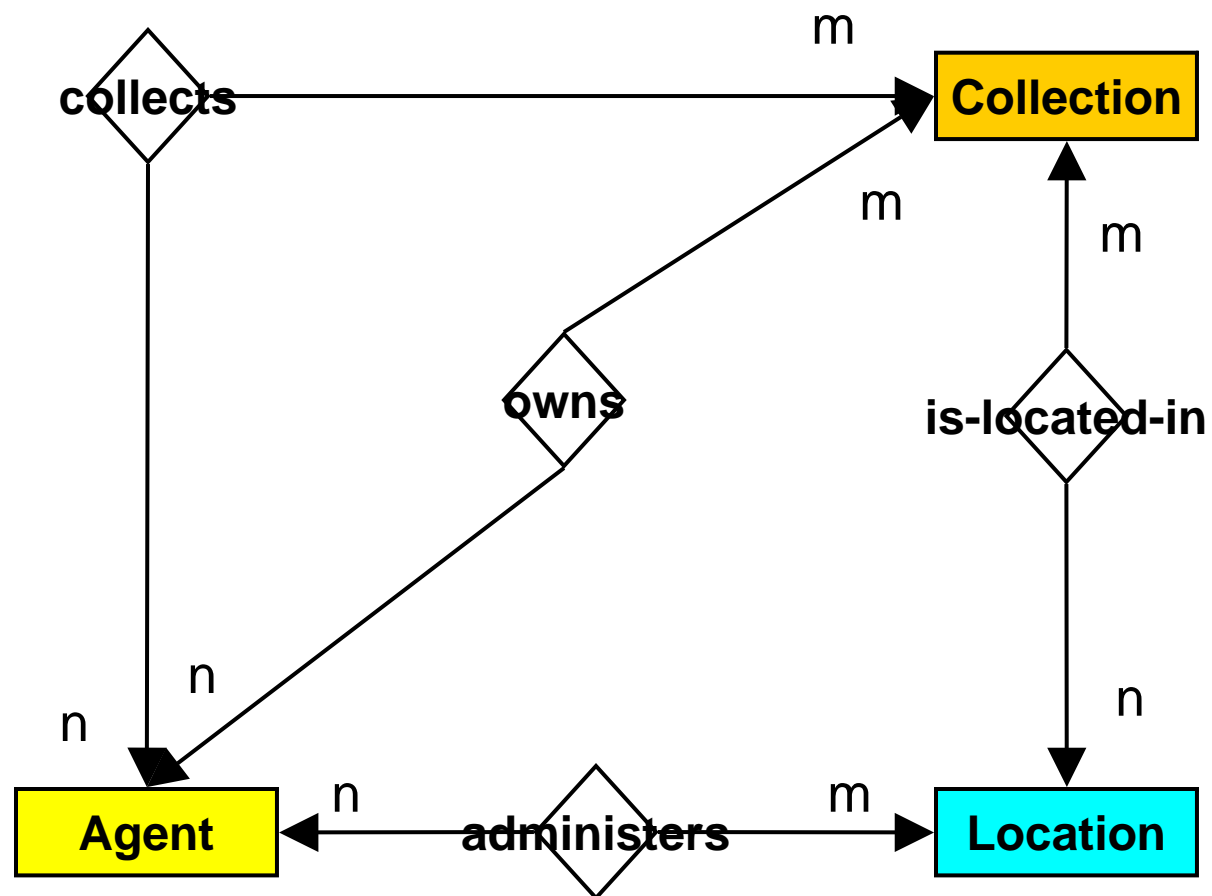


<http://www.ukoln.ac.uk/metadata/rslp/>

# RSLP CD Model & Schema



# RSLP CD Schema



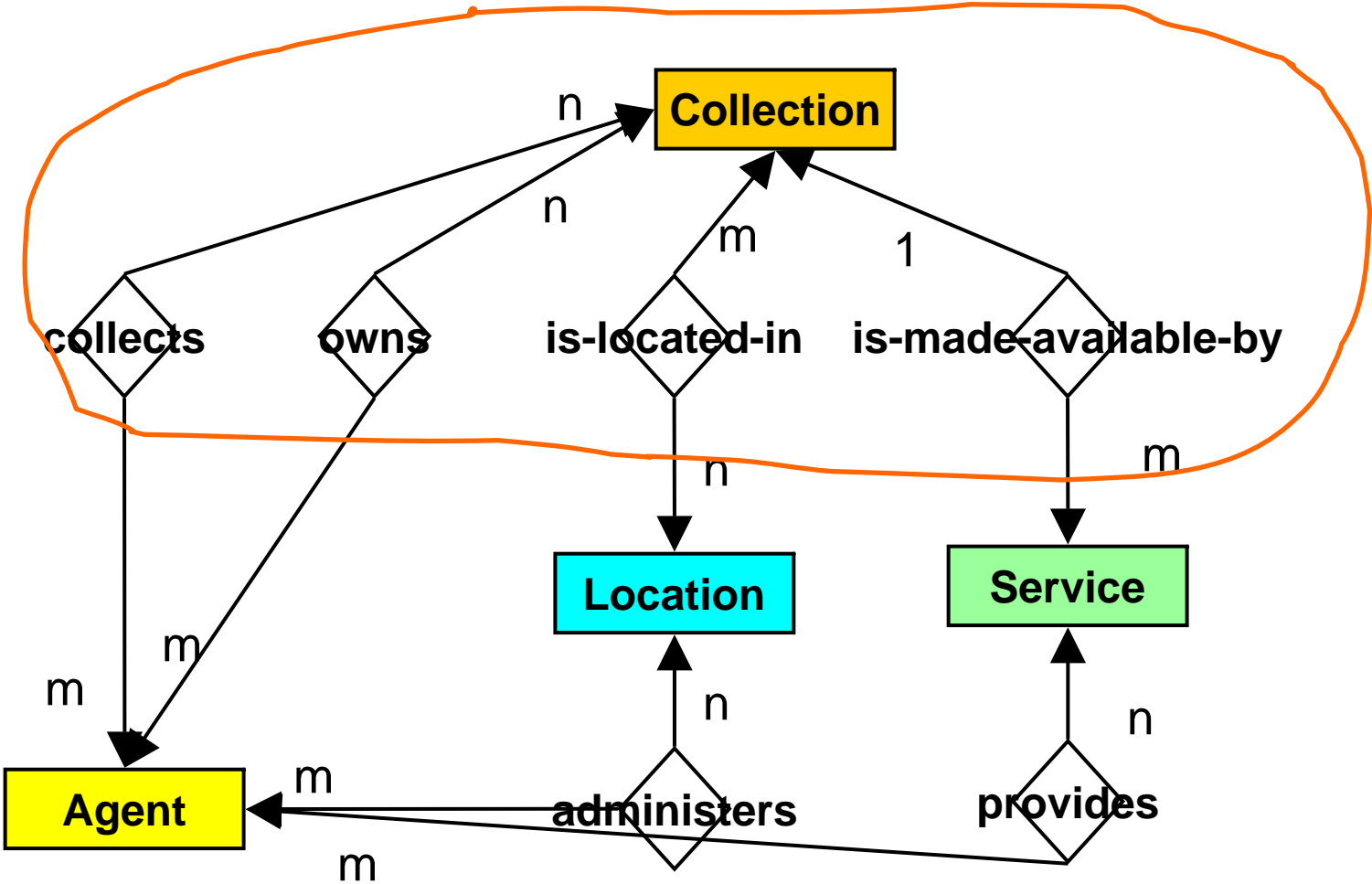
# DC Collection Description Application Profile (DC CD AP)

- A "core" set of collection description properties
  - For **simple** collection-level descriptions
  - Suitable for a **broad range** of collections
  - Primarily to support **discovery** of collections
- Collection attributes of RSLP CD Schema as starting point
- DC CD AP building on Heaney E-R model
  - introduces Service as entity-type
  - describes Collection-Location, Collection-Service, Collection-Agent relationships
  - but excludes Location, Service, Agent description



<http://www.ukoln.ac.uk/metadata/dcmi/collection-ap-summary/>

# DC CD AP



# DC Collection Description Application Profile (DC CD AP)

- Current draft 2004-08-20 covers
  - Identification of collection
  - Content of items in collection
  - Form of items in collection
  - Process by which items gathered into collection
  - Ownership of collection
  - Rights of access to/use of collection
  - Location of collection
  - Services that provide access to collection
  - Relationships between collections



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# Current Issues for DC CD WG

- Clarify Collection-Location and Collection-Service relations
- Propose properties to DCMI Usage Board (or find alternative naming authority/maintenance agency)
- "One-to-one rule" issues
- DC Date WG progress on dates
- Develop supporting materials
- Clarify long-term administration of DC CD AP



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# Additional issues for NISO MI TG2

- NISO MI addressing (primarily) library service providers
- Clarify requirements for additional properties
  - Collection completeness?
  - Conspectus?
- Determine naming authority/maintenance agency for additional properties
- Linking collection and service description
- Decide on recommended syntax
- Develop supporting materials



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# Service Description



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# Service Description

- Led by Larry Dixson, Library of Congress
- (Informational) Service
  - Means of accessing collection
- Service description should include
  - Indication of protocol used
  - Access point for service
  - Authentication/authorisation information
  - Indication of operations/queries supported (where appropriate)
- Assumption that protocols themselves described elsewhere
- NISO MI evaluating use of Zeerex



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# Zeerex

- Z39.50-based specification
  - Based on earlier work, including Explain and Explain Lite
  - XML DTD
- Allows detailed description of Z39.50 and SRW/U services
  - Server info (host, port, database, authentication)
  - Database info (= ~ collection description, access conditions)
  - Index info (search/scan/sort indexes supported)
  - Record info (record syntax, element set name)
  - Schema info (XML Schema)



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# Zeerex and access protocols

- Sufficiently expressive/flexible to describe similar types of service?
  - Services that provide access to a “database”
- Access protocols in scope for NISO MI
  - Z39.50
  - SRW/SRU
  - OAI-PMH
  - HTTP
  - LDAP/X.500?
  - GRID metasearch?
  - GIS search facility?



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# Conclusions

- RSLP CD (and IESR) models/schema influence on other digital library initiatives
  - Collection is-made-available-by Service(s)
- DC CD AP as potential "core" collection description schema
- More complex picture re service description
  - Broad frameworks e.g. UDDI
  - Protocol-related specifications e.g. Zeerex
  - Pragmatic focus on subclasses of service?



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management

# Collection Description, Dublin Core, and the NISO Metasearch Initiative

Using the IESR: What's in it for you?  
London  
Wednesday 12 January 2005

**Pete Johnston**  
Research Officer, UKOLN, University of Bath  
p.johnston@ukoln.ac.uk

UKOLN is supported by:

**JISC**



Museums, Libraries and  
Archives Council



UNIVERSITY OF  
**BATH**

[www.bath.ac.uk](http://www.bath.ac.uk)



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

**UKOLN**

A centre of expertise in digital information management