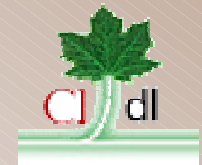# Interoperability and Web Services
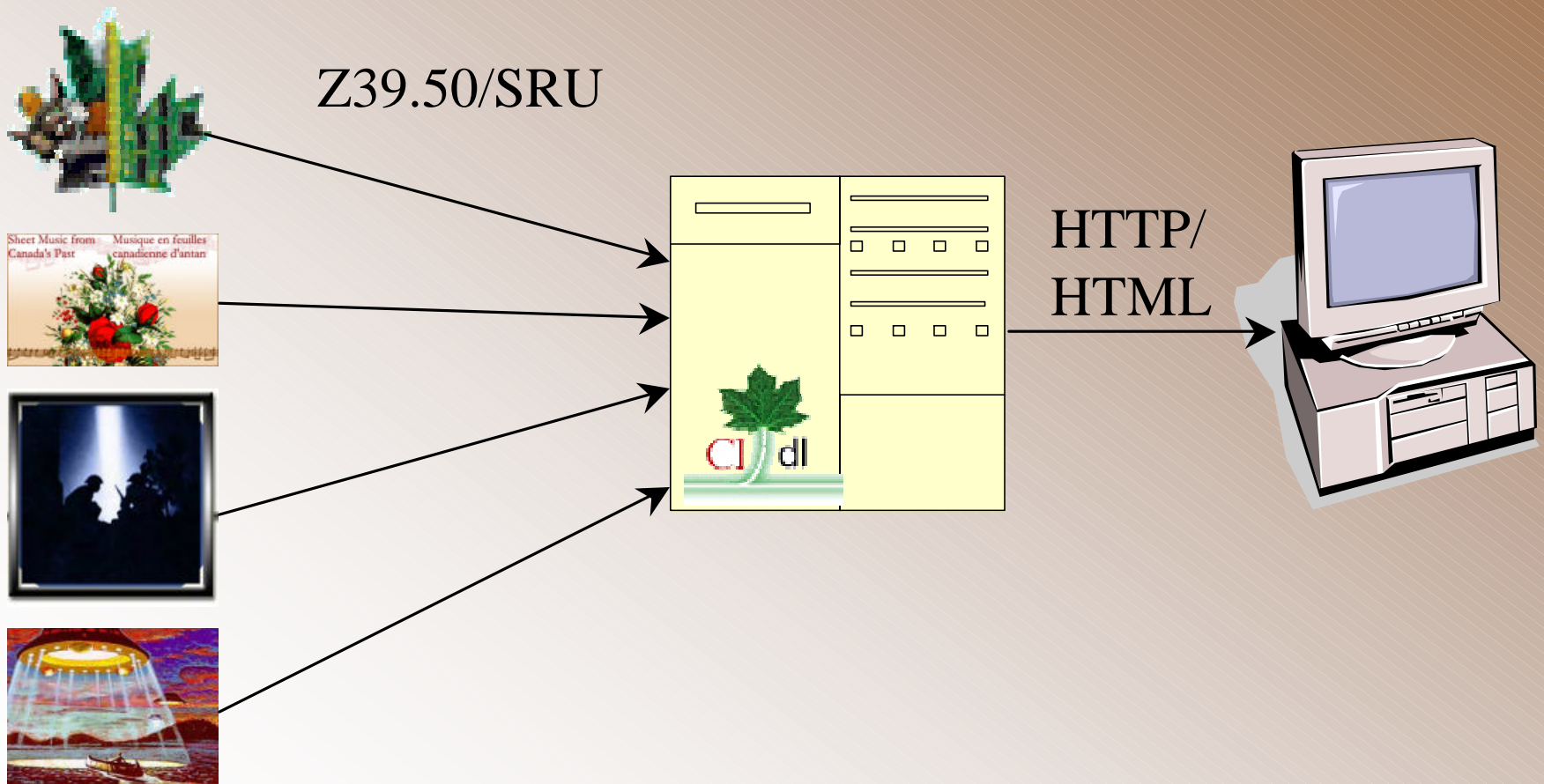
T. Hickey

Interoperability of What?

- The problem
  - Decentralized digitization projects
    - Various standards and procedures
- Desired:
  - Uniform access
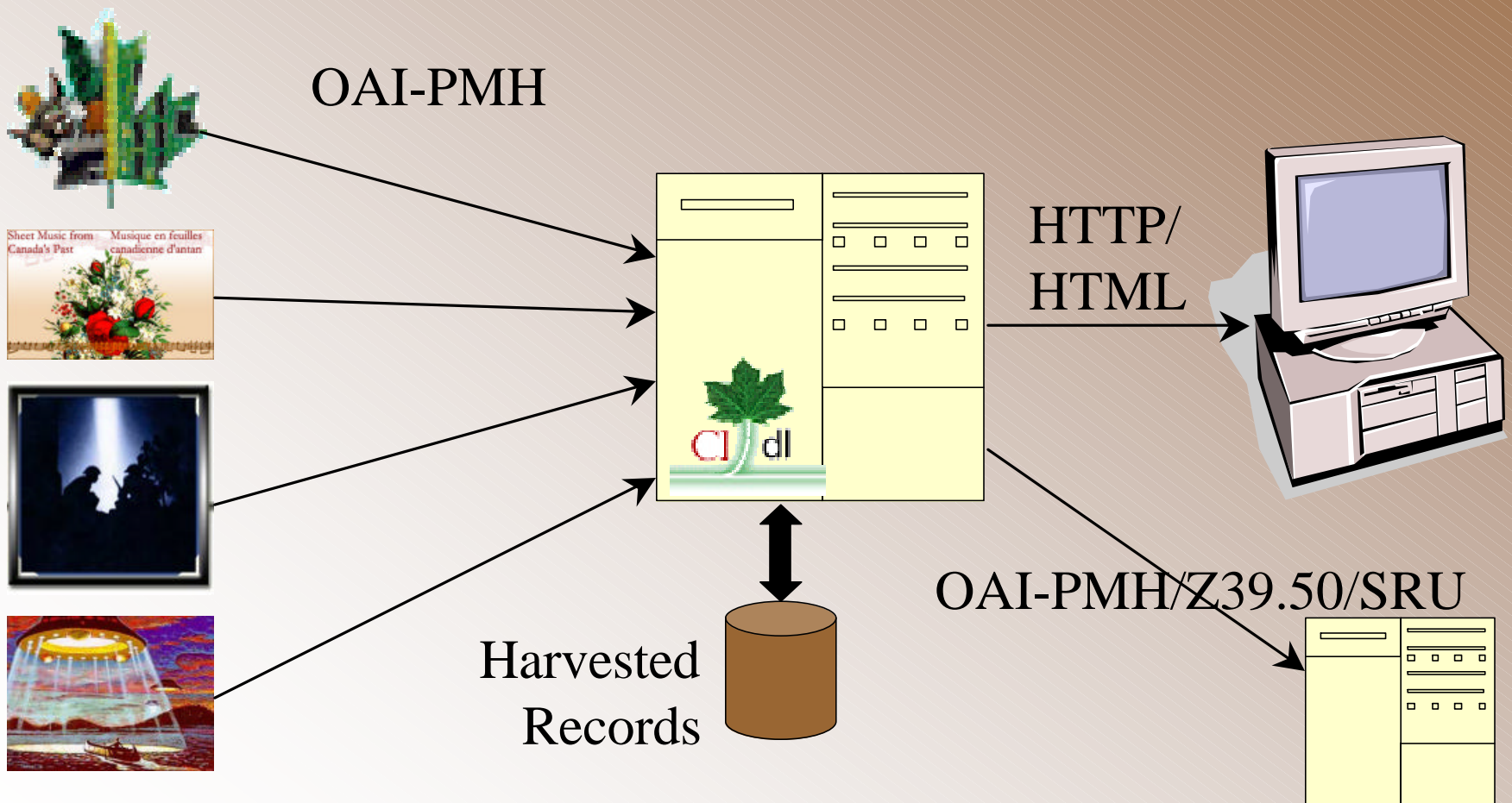 - Wide use

Two Major Option

- Federated search of distributed systems
- Centralized search of harvested metadata

Federated Search Option

Z39.50/SRU

HTTP/
HTML

Aggregation Option

OAI-PMH

HTTP/
HTML

OAI-PMH/Z39.50/SRU

Harvested
Records

Interoperability Technologies

- Z39.50/SRU/SRW
- HTML & Web
- OAI-PMH
- OpenURL, Proprietary solutions

What are Web Services?

- Machine-to-machine communication
- Run over standard Web protocols
  - XML syntax, HTTP packaging
- Several approaches
  - REST
    - Representational State Transfer
  - SOAP
    - Simple Object Access Protocol

Z39.50 ?   SRU/SRW

- Search and retrieve
  - SRU does it with URL's
  - SRW does it with SOAP

- Designed for interactive sessions
  - Widespread use
  - Records can be written as well as read
  - Many interoperability issues are now being dealt with

http://oclc.org/SRW/WorldCat?query=title+exact+"Hunting%20Dogs"

Straight HTML

- Expose each record as an HTML page
- Create a 'tree' above them
- Search via Google

OAI Protocol

- Uses HTTP
- Fairly simple URLs to
  - Identify the server
  - List formats, record sets
  - Get records (can specify date modified)
  - Has *flow control* so that large sets can be managed

http://www.openarchives.org/

OAI Protocol (continued)

- XML version of Dublin Core is required
- Other metadata formats possible (wrapped in XML)
- Typical uses:
  - Publish metadata for a special collection
  - Use to keep two catalogs synchronized

Synchronized Metadata

- Distribute creation
- Distribute responsibility
- Centralize searches
  - Avoid federated searching
- Propagate changes
  - Cataloguing
  - Authorities
- Share data

State of OAI-PMH

- Version 2.0 as of 2002-06-14

- Migration Document

- Implementation Guidelines
  - Repository Implementers
  - Harvester Implementers
  - Aggregators, Caches, and Proxies

State of OAI (continued)

- Hundreds of repositories
- Dozens of aggregators
- Quickly moving to 2.0
- Forming the basis for many services
  - ePrints UK
  - NDLTD
- Some interesting aggregation services

Metadata is Small

- Moving typical collections at T1 speeds
  - 10,000   ~10 seconds
  - 100,000   ~2 minutes
  - 1,000,000   ~15 minutes
  - 50,000,000 ~15 hours
- WorldCat plus authority files
  - Equivalent to 3 DVD movies
  - Have a copy on my PC

Distributing Creation

- Creation is a communal activity
- Doesn't mean everyone has to be on a central system
- Need access to tools
  - Authority files
  - Analysis
  - Transformations
- Needs some central coordination

Access to Tools

- Web Services
  - XML/SOAP
  - Web infrastructure
- Proper 'granularity'
  - How often accessed
  - Whether interactive

Distribution of Responsibility

- OAI provides the mechanism
  - Changes can be detected
  - OAI is completely 'pull'
- Still need agreements
- Issues in how corrections are made

Centralized Searches

- Federated searches aren't worth it
- Cheaper and easier to consolidate metadata
- Easier to ensure uniformity
  - E.g. being able to find documents
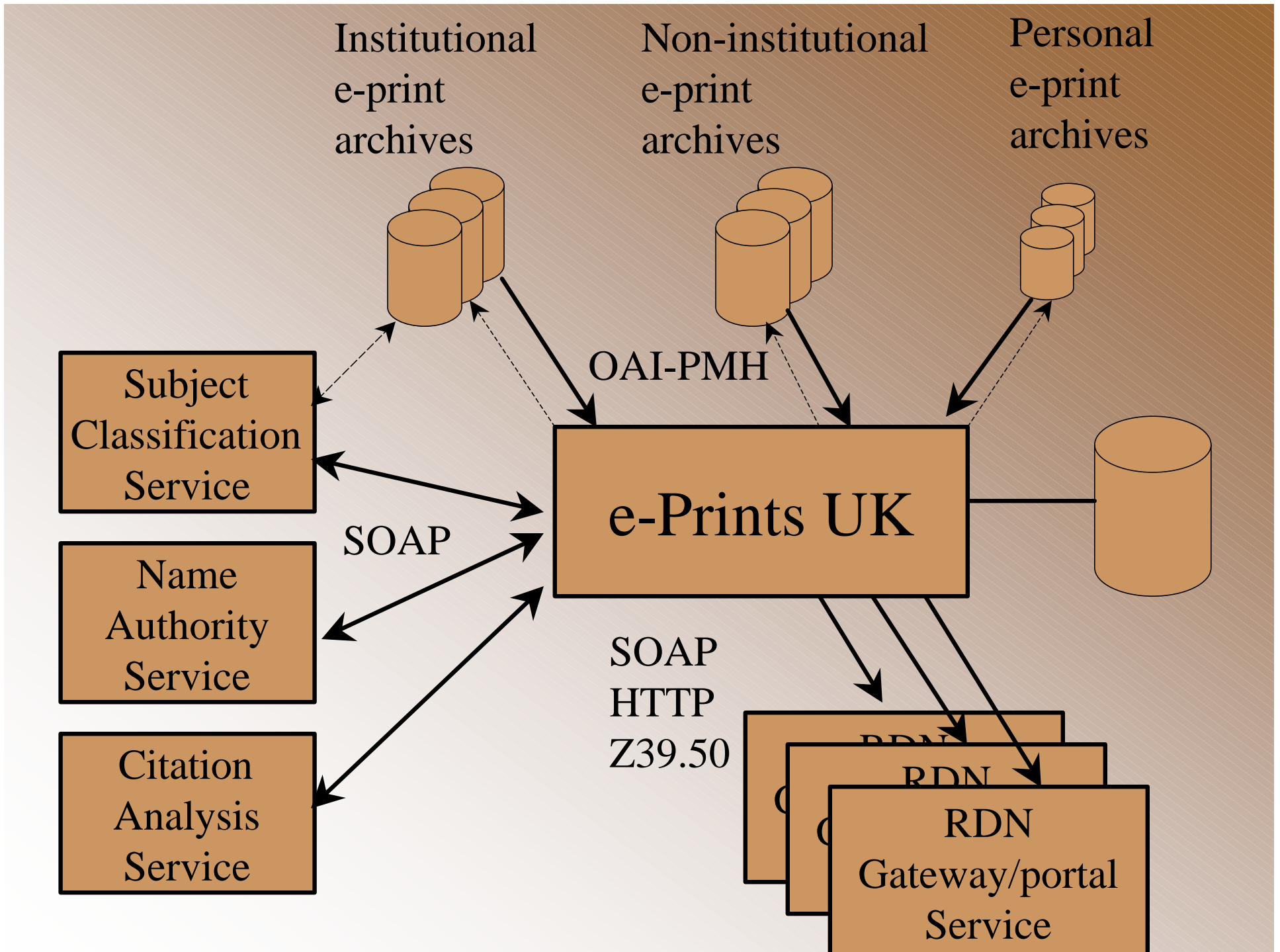- Technical point of view: replicate

An Example: ePrints UK

- Records harvested from archives
- Web services to enhance metadata

JISC

http://www.rdn.ac.uk/projects/eprints-uk/

Institutional e-print archives

Non-institutional e-print archives

Personal e-print archives

OAI-PMH

Subject Classification Service

Name Authority Service

Citation Analysis Service

SOAP

e-Prints UK

SOAP
HTTP
Z39.50

RDN

RDN

RDN Gateway/portal Service

Example: XTCat

- Experimental Theses Catalogue
- Union catalogue of thesis and dissertation metadata

http://alcme.oclc.org

XTCat (continued)

- Records extracted from WorldCat
- Available via OAI-PMH
  - Separate 'set' for electronic theses
- We harvest from other repositories
- Integrate with existing records
- Allow re-harvesting (e.g. by VTLS)

XTCat (continued)

- Have 'seed' data
  - WorldCat Theses and Dissertation records
- Have an organization
  - NDLTD encouraging OAI
- Problems
  - Reliable servers
  - Reliable harvester
  - Duplications
  - Multiple formats

Lessons from Z39.50

- Protocols aren't magic
  - Need to worry about:
    - Formats
    - Indexes
    - Contents of fields
    - Normalization
    - Language

Title: Bill's first log cabin

Contributor: Robert E. Gard

Contents: Photograph [1], black and white, 14 x 9 cm and glass slide, 9 x 9 cm. From William W. Parker, Waskatenau, Alberta. "We Remember" Collection, [19--].

Date Created: [19--]

Genre: Other

Coverage Time: [19--]

Language: English

Source Collection: Alberta Folklore and Local History Collection

Source Series: Photographs

Accession No: 96-93-958

Rights Access: Open

Creators or Authors: Parker, William W.

Subject Headings: Log cabins - Alberta – Photographs

Photographs

Title: Interior of observing station, view 1

Published: [1884]

Physical Description: 1 photograph : b&w ; 10 x 12 cm

Scope and Content: Item is a photograph of the interior of the wooden cabin used as an observing station and living quarters, probably at Ashe Inlet. The photograph shows the stove and a wooden table with two shelves mounted on the wall above. The photograph may have been taken by W.A. Ashe, observer at Ashe Inlet during the Hudson Bay Expedition of 1884.

Source of Title: Title based on contents of photograph.

Inscriptions: "IIX"

Location: Ms. Coll. 310, box 15, folder 2

ID Number: P10026

General Note: All the supplies needed to construct the station buildings were carried on the S.S. Neptune, along with fuel and other provisions sufficient to maintain the stations over the winter, until they could be re-supplied the following summer.

Forms part of the file "Hudson Bay Expedition photo album".

Subjects: Cabins
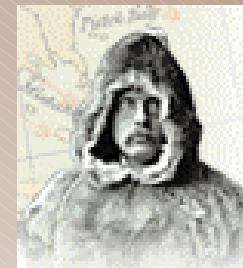
Some Givens

- XML
  - Standardizes character set
  - Standardizes record format
- Z39.50 ?   SRU/SRW
- FTP ?   OAI

- Interoperability is still hard
- Digital preservation is harder still

Questions?

Alberta Folklore and Local History
Collection

University of Alberta Libraries
A division of Learning Systems

Proprietary Solutions

- Many times cheapest
- Costs faced up front
- Possibly less flexibility
- Probably less open
- Not as interesting
- Attractive when problem is well understood

OpenURL

- URL based

- Currently being extended

- Basic operation:
  - Send a coded citation
  - Magic happens

- Not interactive

- Aimed at 'appropriate copy problem'

http://sfx.bath.ac.uk/sfxmenu?genre=book&isbn=1234-5678