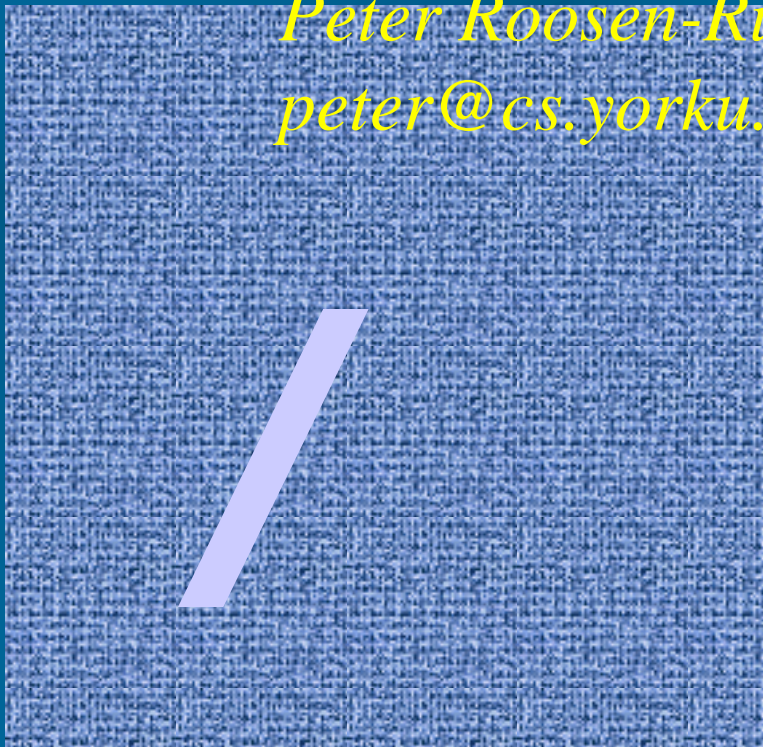


# the ONDISC project

*Peter Roosen-Runge*  
*peter@cs.yorku.ca*



*presented at the  
institute for research in  
learning technology*

December 6, 2001

# onlinedistributedcontent

- ◆ funded by Ontario Ministry of Energy, Science and Technology, 2000-2002
- ◆ servers & development staff at Sheridan College
- ◆ executive director: *Paul Hoffert*

# OnDisC members

- ◆ **CBC**
- ◆ **TVOntario**
- ◆ **CHIN**
- ◆ **Canadian Music Centre**
- ◆ **CCCA**
- ◆ **Encyclopedia of Music in Canada**
- ◆ **Magic Lantern**
- ◆ **Irwin Publishing**
- ◆ **Sheridan College**
- ◆ **Canadian Film Centre**
- ◆ **DCMS**
- ◆ **RightsMarket**
- ◆ **SOCAN**
- ◆ **SGI**
- ◆ **(York University)**
- ◆ **(University of Toronto)**
- ◆ **(DXNet)**

28/11/01

# the goal

- ◆ to enable instructors to create **multimedia “course kits”**
  - analogous to *text kits, anthologies*
- ◆ increased access to licensed content
  - analogous to CanCopy clearance
  - someone pays
- ◆ university may pay for license
  - analogy to *online journals*

# OnDisC system

provides --

- ◆ student & faculty access
- ◆ rights clearances
- ◆ technology for delivery system
- ◆ authorization & tracking
- ◆ user privacy
- ◆ viable business models

# rationale

- ◆ benefits of learning from wide range of content formats
- ◆ costs reduced through
  - online distribution
    - user may pay client-side costs
  - more appropriate “granularity”
  - cheaper long-term storage

# problems

## ◆ chicken & egg issues

- provider side: dubious or unproven business models
- instructor side: how to get to a critical mass?

# more problems

- ◆ metadata issues

- everybody's data is different, often inadequate for users

- ◆ no institutional infrastructure (except for library?) for participation

# the value of free content

- ◆ *"What's holding (the sector) back is the value of content . . . People can get so much content for free on the Web, you really need to have something pretty extraordinary that they need to be able to charge them. That's an issue people have been scratching their heads (about) since the dawn of the Web."* (Nielsen)
- ◆ publishers can't compete just by owning stuff
  - but no infrastructure to compete on quality and scope

# who are the 'real' users?

## ◆ instructors

- “coerce” students to access course content
- create demand based on what they teach
- require institution to license journals, databases. etc.

## ◆ how can/should instructors participate in system management and licensing issues?

# it can't be done yet(?)

- ◆ librarian reactions (UofT, McMaster)
  - can't support multimedia
  - no sound cards in the PCs
  - no bandwidth
- ◆ but it *is* being done by individual instructors
  - with free Web content in dedicated labs & on home PCs!
  - (the future is already here; it just isn't evenly distributed)

# untapped funding

- ◆ new \$ available for participating educational institutions from Heritage, CANARIE
- ◆ faculty researchers essential for OnDisC systems and content to match institutional needs

# instructor-generated markup

- ◆ comments from a UCB faculty focus group:

*Participants were insistent that they must have the ability to add their scholarly comments to the descriptive data provided with images.*

*"The academic process is the generation of metadata. That's what the academic exercise is."*

# symmetric system

- ◆ institutions want to be publishers
  - some incentive for instructors (e. g. xx% of royalties)
- ◆ instructors can help create the catalogue by supplementing the content provider's minimal metadata

# where is the content?

- ◆ in principle: anywhere on the Internet
- ◆ in practise, we have been asked to host content:
  - commercial providers have a little
  - public sector providers (CHIN, NLC) have too much content! (e. g. >100K images)  
*different solution needed*

# inter- vs intra-net?

- ◆ commercial providers leery of Internet-based distribution
  - Internet seen as intrinsically open
- ◆ institutional intranets raise provider comfort level
- ◆ *superdistribution* removes some concerns about unauthorized distribution over public Internet

# a catalogue item

<item>

<type>image</type><format>image/jpeg</format>

<description>"A View of Mile End is represented here. . . ". Sepia watercolour on paper.</description>

<title>View Rear Mile End, Montreal</title><date>1831</date>

<subject>"Landscape, Mile End cityscape, Montreal"</subject>

<creator>Duncan, James</creator>

<key>MCCDM686</key>

<thumbnail><http://www.mcgill.ca/mccord/images/ListView/4-017.gif></thumbnail>

<locator><http://www.mcgill.ca/mccord/images/ObjView/4-017.jpg></identifier>

<publisher>McCord Museum of Canadian History</publisher>

<rights>CHIN</rights>

</item>

28/11/01

# what it looks like

- ◆ <http://www.ondisc.ca:8091/cocoon/samples/xsp/catalogSearch.xml>

# retrieving an image record

Top

Bottom

Previous 20

Next 20



Object Name: Game , Card , Box  
Object Type: Card  
Manufacturer: De La Rue & Co.  
Earliest Production Date: 1880  
Latest Production Date: 1885 C  
Material: wood, paper  
Accession Number: PE975.1.2694a-bn  
Institution: Point Ellice House  
[More information](#)

For © contact  
Point Ellice House.  
All rights reserved.

# scraping HTML source

```
<TD ALIGN="LEFT" VALIGN = "TOP" WIDTH=40%>
<FONT FACE="Arial" SIZE=2>Object Name:</FONT>
</TD>
<TD VALIGN = "TOP" ALIGN = "LEFT" WIDTH=60%><B>
<FONT COLOR="#333399" FACE="Arial" SIZE=2>
  †Game†, †Card†,Box</FONT>
</B></TD></TR> <TR><TD ALIGN="LEFT" VALIGN = "TOP"
  WIDTH=40%>
<FONT FACE="Arial" SIZE=2>Object Type:</FONT>
<FONT COLOR="#333399" FACE="Arial" SIZE=2>
  Card</FONT>
</TD>
```

# Archives as open Web services

- ◆ we want to provide access through our interface but use CHIN's search engine
- ◆ choices --
  - 'scrape' the HTML
  - have CHIN provide a 'Web service' - use HTTP to send a query and retrieve an XML file
- ◆ Open Archive Initiative for Metadata Harvesting (OAI)

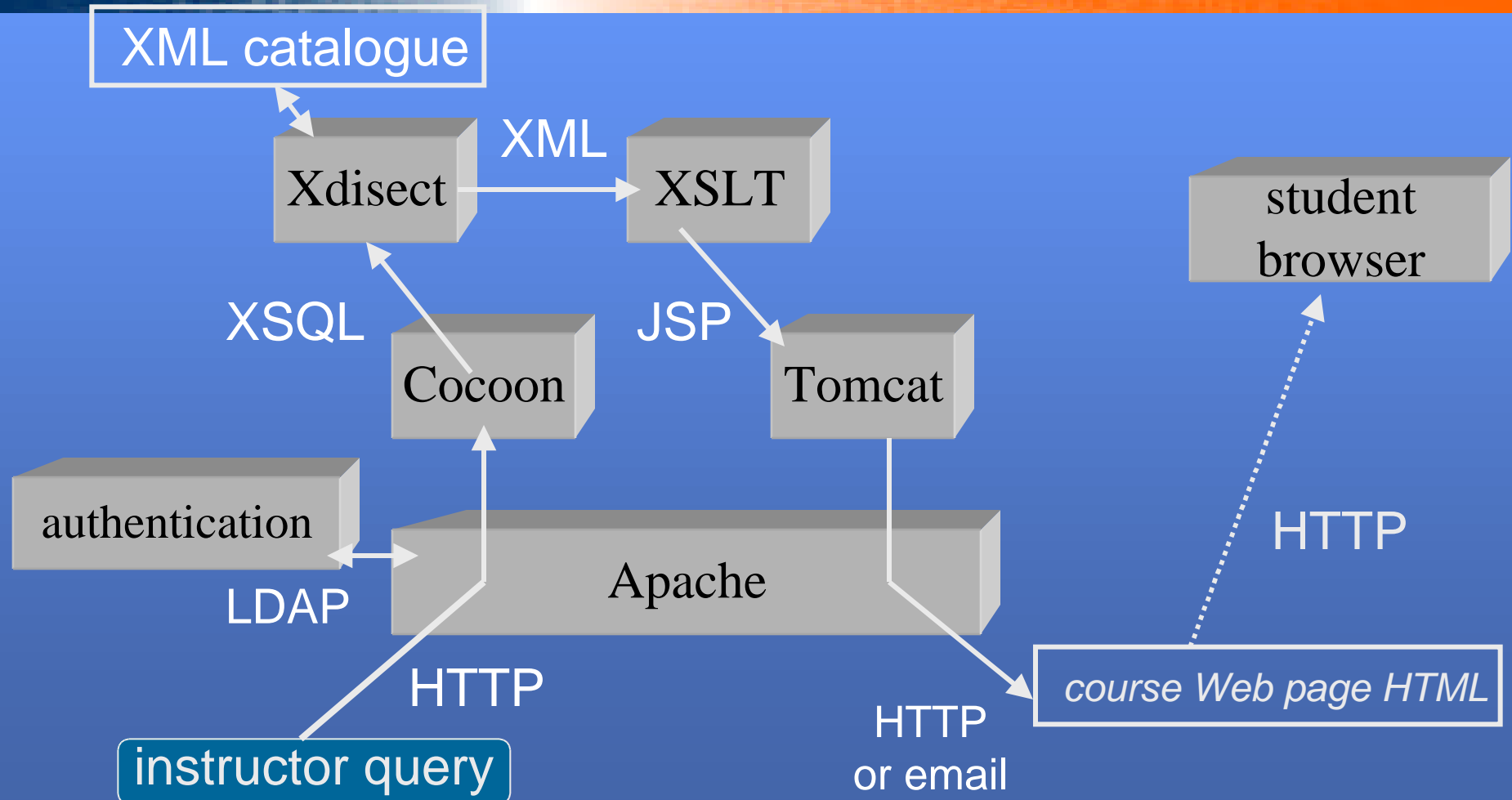
# Artefacts Canada XML data

```
<?xml version="1.0" encoding="ISO8859-1" ?>
<response>
<source>Canadian Heritage Information Network</source>
<header><record_number>1</record_number>
<total_records>2</total_records>
<next_record_url>http://daryl.chin.gc.ca/cgi-
  bin/human_robot?L=E&Idx=WAT&Image=X&M=
  2&Op=AND&Phrase=card%20game
</next_record_url></header>
<object_name>Game, Card, Box</object_name>
<object_type>Card</object_type>
```

# parallel work

- ◆ Gateways
  - e. g. GEM
- ◆ Learning object depositories
  - BELLE, MERLOT, SCORM
- ◆ Metadata standards
  - LOM, ISM, CanCore, etc. . .

“easily assembled at home”



# technology is not a problem

- ◆ technological hurdles of 3-4 years ago are evaporating in the new world of portals, XML, Web services, etc.
- ◆ no "show stoppers" either technologically, or in terms of costs
- ◆ it just gets better with each turn of the wheel of innovation

# inhibitors

- ◆ commercial publishers focus on rights rather than licenses
  - it's easier to say “No” than to ask “How?”
- ◆ industry does not yet have the technical experience or maturity to build distributed collaborative content delivery systems with consistent user-oriented metadata

# mediation

- ◆ the mediating role of instructors needs greater emphasis
- ◆ mediated systems are different from retail e-commerce!

## a Hayekian view

- ◆ self-organization (as in a free market) presupposes institutions.
- ◆ *“Centralization on one layer is a precondition for decentralization on the layers above it (as evidenced by the layers of Internet protocols - lower layers create ground rules for higher level layers).” (Phil Agre)*

we're not there yet

- ◆ we need the initiative of public sector providers (CHIN, CMC, NLC, TVO, CBC, etc) to create an environment with adequate delivery infrastructure and critical mass for licensed content
- ◆ we need more strategic participation (at the level of CITO's) at post-secondary institutions