

**OR2013** Charlottetown, PEI, 9 July 2013 Anne Ward, Independent Consultant Sharon Farnel, University of Alberta

### Using Linked Data to Integrate Resources from Cultural Heritage Institutions across Canada





- "Out of the Trenches: A Linked Open Data Project"
  - Project Background
  - Approach
  - Visualizing the Data
  - Lessons Learned
- "Out of the Trenches" and Over the Top
  - New Recruits / Resources
  - Automation: Mapping / Linking
  - The Next Campaign: Opportunities or Challenges?
- Conclusions

## Agenda



# "Out of the Trenches: A Linked Open Data Project"

Video, Report, and RDF Data http://www.canadiana.ca/en/pcdhn-lod

![](_page_2_Picture_3.jpeg)

## Project Background

![](_page_3_Picture_1.jpeg)

- (PCDHN) is a consortium of Canadian cultural heritage institutions that includes universities, libraries and archives
- A proposal was made to the network in 2011 to "develop a national model for linking our metadata in order to improve the discoverability of Canada's documentary heritage collections and to provide Canadians seamless and integrated access to this content"
- A subset of PCDHN partners agreed to participate/provide First World War resource metadata and undertake a proof-of-concept using linked open data

![](_page_3_Picture_6.jpeg)

## Approach

- Metadata records (e.g., MARC, Excel spreadsheets, Dublin Core RDF, MODS XML) provided by partners were analyzed
- Metadata published as RDF by other organizations was reviewed (including Bibliothèque nationale de France (BnF) who included some Resource Description and Access (RDA) elements in their RDF mappings)
- A preliminary set of target ontologies was identified, including commonly used ontologies such as FOAF and DC (for discovery by a broader set of users), as well as ontologies better suited to the actual content / type of resource described (e.g., RDA for published materials such as music recordings and music sheets, OAI/ORE for aggregated resources (for more specificity))
- Example mappings were created and ratified by/with partners ... using a classification model informed by FRBR/RDA, CIDOC, ISAD/ISAAR ...

![](_page_4_Picture_6.jpeg)

## Approach: The Model

![](_page_5_Figure_1.jpeg)

![](_page_6_Picture_0.jpeg)

• ... the ratified ontologies (element sets) and vocabularies identified by the partners were applied to the model and used to develop the mappings for individual metadata records ...

## Approach

![](_page_6_Picture_4.jpeg)

### Approach: Applying Ontologies and Vocabularies

properties using published ontologies (element sets) including RDA, Dublin Core, Friend-of-a-friend (FOAF), event ontology, time ontology, etc.

> properties using published ontologies (element sets) for "concepts" - SKOS (Simple **Knowledge Organization System**)

value of "place" properties using published URIs including www.geonames.org or geographic based ontologies (element sets) where detailed information is required

![](_page_7_Figure_4.jpeg)

![](_page_7_Picture_7.jpeg)

## Approach: Mapping Metadata, Creating "Authorities" and Linking

- resource)
- conformed to the model (e.g., for person, organization, event, resource)
- headings that do not currently exist in RDF form)
- used to create links between them using the relationships provided by the organization)

 Conscious decision was made to include the mapping of individual metadata elements to multiple RDF elements (e.g., "rdaGrp2:preferredNameForThePerson" and "foaf:name") to address broader community use (i.e., through more commonly understood/used ontologies) and specificity (e.g., RDA for specifying preferred and alternative names, IBSD for more explicit characterization of the content of a

Authority data (RDF files) was created for the "things" in the metadata records that

• For some concepts, PCDHN authority data was created (e.g., for french subject

 PCDHN URIs were assigned to the unique "things" represented in the metadata and ontologies (e.g., "dc:creator" / "rdaRole:author" for linking a resource to the person /

![](_page_8_Picture_10.jpeg)

## Approach: Mapping Metadata, Creating "Authorities" and Linking

- vocabularies (such as LCSH and FAS RDF for resources
- ... a conscious decision was also made to include multiple authorities for a given concept (to achieve maximum discoverability) as well as retain its textual value (for use in an application to showcase the data)
- Common subjects (e.g., all First World War subject headings) were injected into the RDF for the resources
- The content of the resource was also characterized using vocabularies such as ISBD, LCSH, TGM, and FAST using the corresponding URIs (linking to RDF datasets in the LOD cloud)
- Mappings (mapping rules) for each individual set of metadata records were developed along with mappings of textual values to vocabularies (for subject and content type)
  (see Appendix E in the report at PCDHN<sup>®</sup>PPoof-of-concept Final Report)

vocabularies (such as LCSH and FAST) and the corresponding URIs used in the

![](_page_9_Picture_7.jpeg)

## Approach: Visualizing the ata

- The partners also decided to "visualize" the data in the RDF files generated from the mappings rather than simply implement another "search" application
- Inspired by initiatives such as:

  - Tim Wray's Canvas (<u>http://timwray.net/2011/12/canvas/</u>) The Real Face of White Australia (<u>http://invisibleaustralians.org/faces/</u>) Mildenhall's Canberra (<u>http://mildenhall.moadoph.gov.au/</u>) and
  - applying ideas from <a href="http://discontents.com.au/shoebox/every-story-has-a-">http://discontents.com.au/shoebox/every-story-has-a-</a> beginning
- the partners elected to tell the story of an individual soldier ... and along the way ...

![](_page_10_Picture_9.jpeg)

## Visualizing the Data

- ... dimensions for the visualization were derived from the characteristics of the resources/entities described, including:
  - the type of entity (e.g., person, event, organization)
  - their relationships (e.g., subject of, affiliations to organizations, participation in events)
  - the type of resource (e.g., films, photographs, postcards, posters, songs, newspapers)

![](_page_11_Picture_5.jpeg)

![](_page_11_Picture_6.jpeg)

World War, 1914-1918, Panoramas

![](_page_11_Picture_9.jpeg)

World War, 1914-1918,

1914-1918,

rld War, 1914-1918, Events

![](_page_11_Picture_12.jpeg)

World War, 1914-1918, Newspapers

![](_page_11_Picture_14.jpeg)

Nº 12

![](_page_11_Picture_16.jpeg)

# Visualizing the Data

- The scope of the visualization was limited to:
  - exploring two dimensions: a single soldier's story (Mike Foxhead), and the events of the First World War
  - using the links to resources within the PCDHN only (although links to external resources are provided (e.g., links to published vocabularies such as OCLC Fast and LCSH, and authorities such as VIAF))
  - querying data along a limited set of predefined paths

![](_page_12_Picture_5.jpeg)

№ 13

![](_page_12_Picture_7.jpeg)

# Visualizing the Data: The

authority data (RDF for Mike) a depiction biographical

information

![](_page_13_Picture_2.jpeg)

Name: Date of birth: Place of birth Date of death: Place of death

Affiliated with:

Regimental Number: Occupation

### His Life with the CEF

of Death of Death

![](_page_13_Picture_9.jpeg)

blaces associated with the perso (place of birth, death)

"geonames" used for places and provide maps

> "dc:subject" used for resources linked to places

![](_page_13_Picture_14.jpeg)

# Visualizing the Data: The

"dc:subject" used for resources linked to Mike thumbnails specified in RDF or generated from source resource

"dc:subject" used for resources linked to the group / organization

![](_page_14_Picture_3.jpeg)

"event:isAgentIn" used for events Mike participated in "dc:subject" used for resources linked to the event

group/organization Mike was associated with ("rdaGrp2:affiliation")

![](_page_14_Picture_8.jpeg)

# Visualizing the Data: An Event

### uthority data for an event

"event:isAgentln" used for identifying who participated in the event

![](_page_15_Picture_3.jpeg)

![](_page_15_Figure_4.jpeg)

place associated with the event

Geonames" used for place and provide map "dc:subject" used for resources linked to places links to other resources for the event

![](_page_15_Picture_8.jpeg)

## Visualizing the Data: An Event

### "dc:subject" used for resources linked to the event

![](_page_16_Picture_2.jpeg)

№ 17

![](_page_16_Picture_4.jpeg)

### Visualizing the Data: All Resources Josiah Pinne

![](_page_17_Picture_1.jpeg)

- ... and of course as many in the cultural heritage community have realized, "RDF and LOD are an elegant approach for integrating resource discovery across different domains, institutions, and services"
- ... RDF and LOD "enable web users and third party organizations to integrate PCDHN resources with their own resources to create their own "stories" and "virtual exhibitions""
- ... RDF and LOD "effectively remove the constraints of existing web approaches in which the paths followed by web users are explicitly defined by the organization hosting the metadata/resources"

![](_page_18_Picture_5.jpeg)

- A large and rich set of RDF ontologies / element sets exists today
- Multiple ontologies / element sets can be used together: no longer must resource descriptions be constrained to a "one-size-fits-all<sup>1</sup>" nor to a "onevocabulary-to-rule-them-all<sup>2</sup>" model
- Through the work of the community, equivalences between ontologies / element sets can be expected in the future, for example by linking terms in different element sets ... ... as well as between vocabularies (FAST and LCSH, LCSH and Rameau / German subject headings)

![](_page_19_Figure_8.jpeg)

### Figure 3.1

Specific titles from RDA and FRBR can link to Dublin Core for more compatibility with more generalized metadata. Source: Karen Coyle, Chapter 3, Linked Data Tools: Connecting on the Web, Library Technology Reports, alatechsource.org, May/June 2012

<sup>1</sup> RDF enables "the replacement of attempts at one-size-fits-all schema, rules and other international/global standards with what might be termed an all-sizes-fit-one approach"

Source: Reconsidering Universal Bibliographic Control in Light of the Semantic Web, Gordon Dunsire, Diane Hillman and Jon Phipps, Journal of Library Metadata, 12:164-176 2012, © Taylor & Francis Group, LLC, DOI: 10.1080/19386389.2012.699831

<sup>2</sup> Term used by Owen Stephens in a e-mail exchange on the "Bibframe" Listserv

![](_page_19_Picture_15.jpeg)

- Some gaps exist, for example:
  - established as an RDF vocabulary (used as the "object" for a "predicate" of "rdf:type" for an archival resource)
  - examples in the vcard (<u>http://www.w3.org/2006/vcard/ns#</u>) ontology):

<rdaGrp2:preferredNameForThePerson> <rdf:Description>

<rdf:value>"name of the person"</rdf:value>

- <time:intervalStarts>"date"</time:intervalStarts>
- <time:intervalFinishes>"date"</time:intervalFinishes> </rdf:Description>

</rdaGrp2:preferredNameForThePerson>

• the "vocabulary" or "classification structure" for archival aggregations that is in use today (e.g., "collection", "fonds" / "series" / "sub-series", etc.) - these may be "rdaElement:systemOfOrganization") or as a class in an ontology (used as the

• addressing the "time" dimension of properties such as for the name of an individual without going into describing the events that led to name changes over time although this could be modelled by mapping the property as follows (borrowing from

![](_page_20_Picture_12.jpeg)

- can be added
- Element sets and vocabularies can be used in multiple contexts (rather than (rdaGrp2:affiliation) or profession (rdaGrp2:professionOrOccupation)
- tools to apply the rules on a larger scale

• Care needs to be applied to the selection of target ontologies (longevity, community support, stability), although mappings can be revised or "better" data statements

creating single purpose properties / vocabularies): for example, "rdaRole:author" can be used as a "predicate" to link an author and a resource, or as the "object" to the "rdaGrp2:professionOrOccupation" predicate; or an LCSH heading for "soldiers" can be used for the subject of a resource or as the value for a group affiliation

 Metadata mapping is essentially done once (although it may be updated as new information / better ontologies are identified): the most resource-intensive work is in defining the appropriate mapping rules - after that it is a matter of developing the

![](_page_21_Picture_8.jpeg)

- Integrating resources across separate and distinct collections is easy and no published on the web
- for large volumes
- Data queries can yield a very high volume of results (as demonstrated by the any duplication in results (for example, with resources fully described at an individual resource level as well as at the level of an aggregation, that share common characteristics such as subject)

negotiation is required due to the RDF model itself - everything is expressed in the same manner (triples) and using the data statements is a matter of understanding the intent of the ontologies used and the fact that these ontologies exist / are

Manually creating RDF for a small set of records is manageable but not sustainable

application when loading all resources for the First World War event): intelligent applications will be required to better organize results for users as well as address

№ 23

![](_page_22_Picture_8.jpeg)

## Lessons Learned: Project Recommendations

- Specific recommendations endorsed by the partners:
  - develop "visualizations" for other dimensions in the data, including "geographic" and "time" dimensions for events
  - "productionize" the visualization application
  - provide more intelligence in parsing the results
  - provide the ability to retrieve/filter by multiple dimensions
  - add additional resources from group members and interested parties
  - develop programmatic support for applying the mapping rules

... and as we will see next, the University of Alberta has taken up the challenge ...

![](_page_23_Picture_11.jpeg)

# "Out of the Trenches" and Over the Top

![](_page_24_Picture_1.jpeg)

# **UNIVERSITY OF ALBERTA**

№ 25

![](_page_24_Picture_4.jpeg)

## New Recruits / Resources

NONTH SEA

### MEMORIAL UNIVERSITY Memorial University of Newfoundland WWI Newfoundlandia

![](_page_25_Picture_2.jpeg)

### **EXPERIENCES IN FRANCE** BY WILFRED T. GRENFELL, M.D.

On Dr. Grenfell's return last autumn from Labrador, where for a quarter of a century he has been conducting a great work of civilization, he felt impelled, during a winter vacation from that work, to accept a position with the Harvard Surgical Unit, operating at the Twenty-Second Hospital of France, located at Boulogne. Here he remained for some months, and then, at the request of the English general at headquarters, he visited the northern armies in France in order to gain more direct experience in the first-line trenches. We print a partial account of his experiences in these two capacities .- THE EDITORS.

large a problem, and it is so difficult not to line. The menace of innumerable irresponsible be misunderstood, that I must begin by say- and half-qualified society would-be comforters ing that I speak only for myself, and am quite was only a passing phase. It was, however, aware that the same facts might lead other very clearly recognized. At Verdun, in people to act differently from the way I Russia, in Servia, and in Italy, and wherever should, just as in our surgical procedure big emergencies arise, the voluntary work of

my return from France you ask next advance in the Egyptian and Boer camme to give my deductions about the paigns. To-day the despised leech and barbergreat war from experience. It is so surgeon is a fully recognized officer of the the Red Cross Society is still of great value. What about prisoners?

l visited the families in France of many soners in Germany, and found everywhere same tale. Those prisoners were not e to live on the food that Germany proed them with. Hundreds of tons of bread I fat go through the mails every week. I self saw one little seamstress, herself an le from the conquered north of France, her ent husband a member of the French Fly-Corps, sending every week, from a house which she had been hospitably received, packages of food at a time out of her tiny nings to relations in German prisons. I ays found her cheerful and optimistic on ny delightful visits paid to the house. We w now that we did not understand the ench before. We English have learned only to appreciate them but to love them. have gauged them too much by the "Vie isienne" and the "Petit Journal Pour

A prominent Russian lady in London perted me to take a small share in her noble rk of an exactly similar character, only for ssians. Day after day and night after ht she and her daughter and close friends k up and despatch bundles of food to ividual prisoners in Germany. Among immense personal correspondence with m were most terribly pathetic and touching dents. It almost brought tears into one's s to read letters from two Russian gen-Is thanking her for having sent them cels of bread.

I had no opportunity of seeing German soners concentrated, but I saw wounded

my own dear mother I have received all your letters Johich you have put in with the others, and they dated from bie ) the Jan, ) the and gan 14th. I may say that there was a time this winter that I didn't get a letter from forme for a fortnight and I felt miseretry, and then just as we were going up the line they was three fareels came for me and seven letters. on ou fancé was you times cade in the other was housed samis gift of soits handerships, mits, and chocked Which I was very fleaned to get, and in the other one was writing paper from allie Kean. you can imagine how glad I wa. to get it all and then just after we got up the line the mail came and it was all for me every letter, How many do you think only Sixteen

Somewhy in trong

FACTS AND FANCIES. LOVINGLY DEDICATED TO THE BOYS OF THE NEWFOUNDLAND REGIMENT. AND TO EVERY WOMAN WHO HAS KNITTED A PAIR OF GREY SOCKS. By MRS, T. J. DULEY. Verses by Margaret Duley.

A PAIR

of

**GREY SOCKS** 

Nº 26

![](_page_25_Picture_13.jpeg)

# Automation: Mapping / Linking

- Investigating options for prog conversions
  - XSL stylesheets based on mappings to date
  - Existing generic conversion tools + some customization
  - Enhancement tools like <u>SILK</u>, <u>OpenRefine</u>

Investigating options for programmatic metadata mappings and

nappings to date tools + some customization K, OpenRefine

![](_page_26_Picture_8.jpeg)

# The Next Campaign: Opportunities or Challenges?

- New content, new partners
- Improved processes and workflows
- Enhanced application, new tools
- More exposure for the project and the data

### rs workflows w tools oject and the data

![](_page_27_Picture_7.jpeg)

## The Next Campaign: Opportunities or Challenges?

- And ...
  - Starting small is ok
  - Think curation
  - Focus on unique content

№ 29

![](_page_28_Picture_7.jpeg)

## Conclusions

- The future of resource description is not about making resources discoverable on the web, it is about "description that happens on, in, and as part of the web and the networked world we live in" (Library of Congress Bibliographic Framework Initiative) ...
  - ... it is about using semantic web technologies to describe resources ...
    - ... it is about linking resources to other resources (on the web) ...
      - ... the future of resource description is happening now ...
        - ... it is happening in many domains ...
- ... and it is not about adopting a "one-size-fits-all" model nor conforming to "onevocabulary-to-rule-them-all" ...
  - Nº 30
  - ... but about finding links between concepts expressed in ontologies / vocabularies and between instances of "things" expressed in these ontologies

![](_page_29_Picture_9.jpeg)

**Questions**?

Anne Ward eaward@computer.org

**Sharon Farnel** sharon.farnel@ualberta.ca

![](_page_30_Picture_4.jpeg)

№ 31

![](_page_30_Picture_6.jpeg)