

BARN RAISING IN A VIRTUAL WORLD . . .

Allan Bell, University of British Columbia

Alex Garnett, Simon Fraser University

Carla Graebner, Simon Fraser University

Geoff Harder, University of Alberta



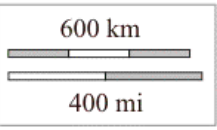
...HOW INNOVATIVE APPROACHES IN FUNDING LED TO AN ARCHITECTURE WE CAN ALL USE

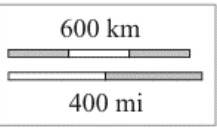
- About us
- History of collaboration
- Common infrastructure
- Our projects



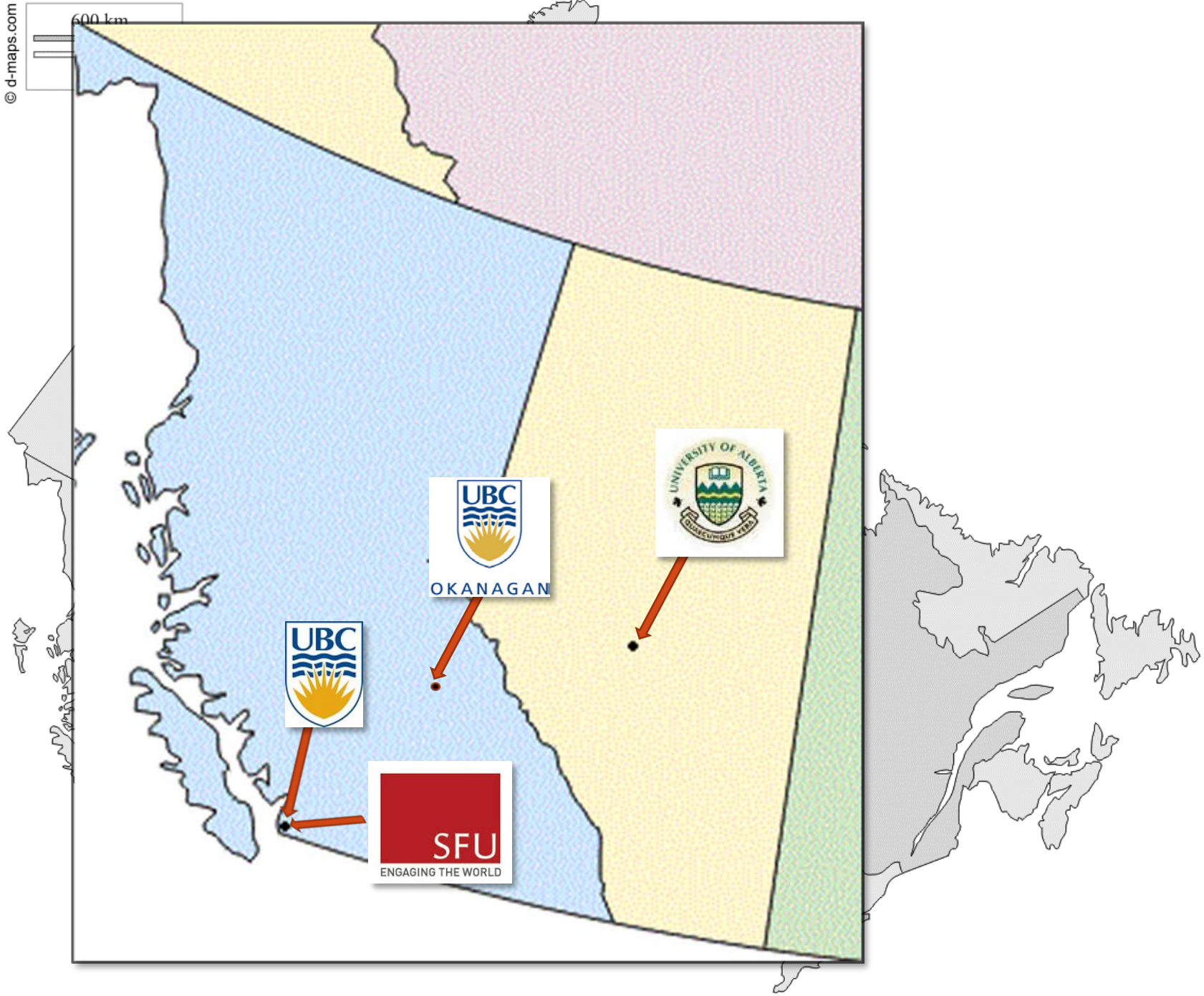
“Barn raising - the farm of Alex. McLaren, near Pense.”
Ref Number R-B12988 , Item Date 1912.
Saskatchewan Archives.







600 km



BARN RAISING

- **Laying in the lumber and the materials**
- **Making the plans**
- **Clearing the ground**
- **Tradesmen are hired**
- **Topping out**
- **Identifying local resources**
- **Project planning**
- **Laying the groundwork**
- **Deploy the talent**
- **Topping out**



IDENTIFYING LOCAL RESOURCES

- **Informal**
 - Communities of practice / common interest
 - SLAIS
 - Data Camp, CODE4LIB BC
 - Professional associations
 - Geography
- **Formal**
 - Canadian Association of Research Libraries
 - Council of Pacific and Prairie University Libraries
 - British Columbia Research Libraries Group



PROJECT PLANNING

Unique competencies / contributions / expertise

- COPPUL Digital Preservation Working Group
- COPPUL PLN
- Canadian Government Information PLN
- BCRLG Government Documents Digitization
- ABACUS/Dspace migrating to Dataverse
- Data Curation
- ?



THE TALENT



Identify local expertise

- Identify local expertise
- Develop partnerships
- Choose the vendor/developer
- Communicate objectives, parameters and time frame



LAYING THE GROUND

Environmental Scan

- Institutional priorities
- Existing resources
- Potential partnerships

Open Source solutions

- Community oriented
- Responsive to funded projects
- Made available to the community under creative commons licensing upon completion
- Collaborative continuum



OPEN SOURCE SOLUTIONS

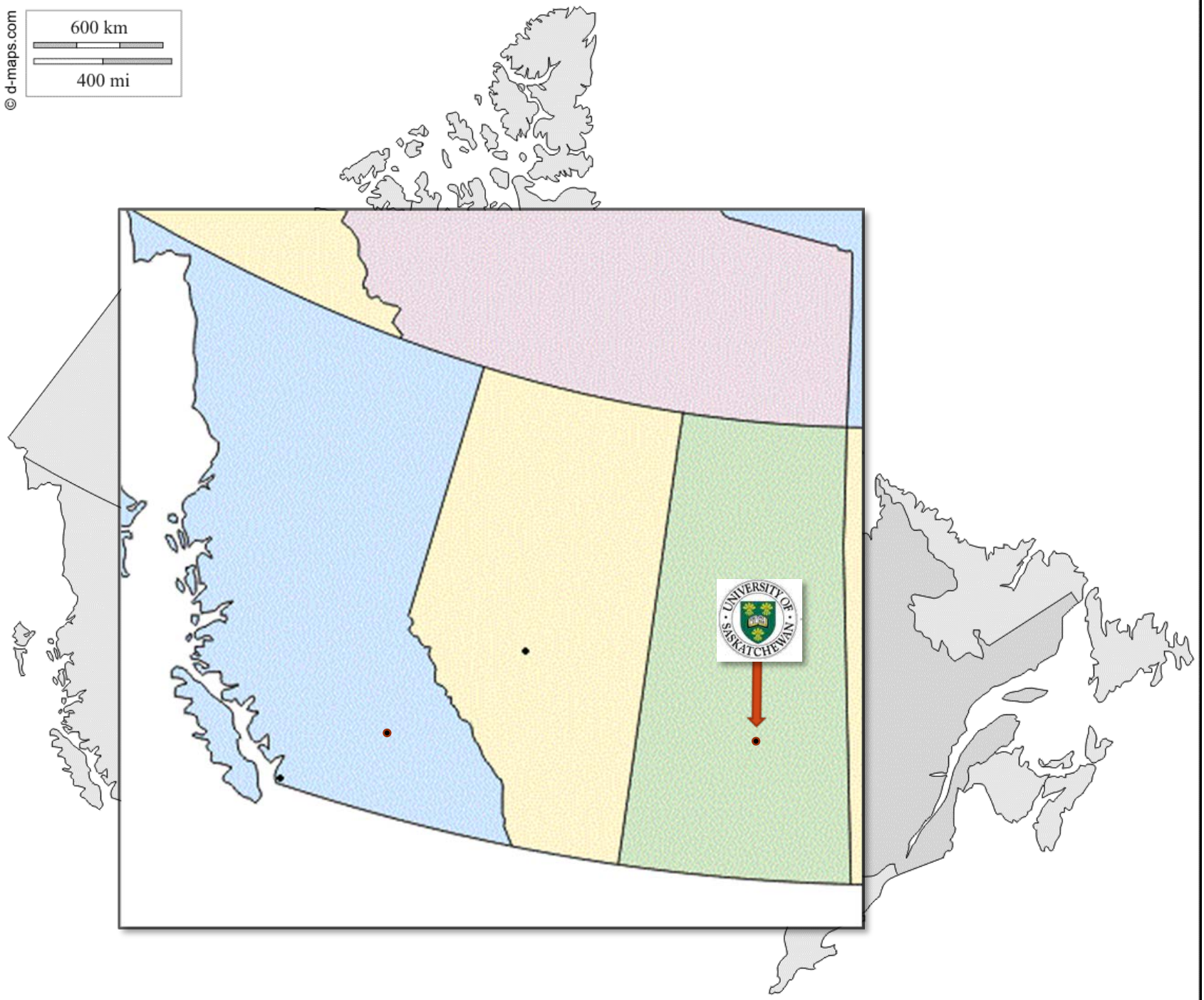
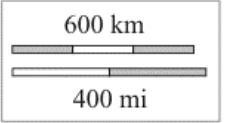
Archivematica

- Simon Fraser University Library: Research Data Repository [SFU RaDaR]
- Simon Fraser University Archives
- University of British Columbia: Digital preservation program
- University of Alberta :Canadian Polar Data Network
- University of Saskatchewan

Islandora

- Simon Fraser University Library RaDar
- University of Alberta
- University of Saskatchewan





SIMON FRASER UNIVERSITY

SFU RaDaR

To undertake a 'Proof of Concept' project which will identify the technological infrastructure, policies and procedures, and costs associated with the development of a Research Data Management Repository for research data to ensure its security and integrity over time



SIMON FRASER UNIVERSITY

Two separate but related customizations:

- **Automated Ingest** of multiple files via a scriptable workflow API.
- **A Storage API** to allow Archivematica to use non-filesystem storage (e.g. NFS mounts or Fedora Commons) for objects.



AUTOMATED INGEST

- Archivematica is designed to be *actively used* by depositors, curators, and archivists.
- This doesn't necessarily scale for us -- we want to use it at the tail (preservation) end of existing repository services for electronic theses and local research data.
- These objects are homogeneous and can receive the same normalization treatment.



Any

Keyword



Show files?

or

Any

Keyword



[Add New](#)

Submission Information Package

UUID

Ingest start time

Submission Information Package	UUID	Ingest start time	
Memphis_Jack_files	7de5a829-69e2-4a96-9c97-af5acdb87dfc	2013-04-29 14:26	
▶ Micro-service: Normalize			
Job: Normalize [?]		Awaiting decision	Actions
Job: Resume after normalization file identification tool selected.		Completed successfully	
Job: Set SIP to normalize with FITS-Droid file identification.		Completed successfully	
Job: Select format identification tool		Completed successfully	
Job: Set resume link after tool selected.		Completed successfully	
Job: Find options to normalize as		Completed successfully	
Job: Move to workFlowDecisions-createDip directory		Completed successfully	
Job: Grant normalization options for no pre-existing DIP		Completed successfully	
Job: Set remove preservation and access normalized files to renormalize link.		Completed successfully	
Job: Check for Access directory		Completed successfully	
Job: Check for Service directory		Completed successfully	
Job: Identify manually normalized files		Completed successfully	
▶ Micro-service: Clean up names			
▶ Micro-service: Remove cache files			
▶ Micro-service: Include default SIP processingMCP.xml			
▶ Micro-service: Rename SIP directory with SIP UUID			
▶ Micro-service: Verify transfer compliance			
▶ Micro-service: Verify SIP compliance			
▶ Micro-service: Approve SIP creation			

[Ingest](#) / [Cat_videos](#) / [Metadata](#) / Add

Metadata

Cat_videos

Applies to

Metadata can be added at the SIP/AIP level only

Title

Creator

Subject

Description

Publisher

Contributor

Date

Use ISO 8061 (YYYY-MM-DD or YYYY-MM-DD/YYYY-MM-DD)

Type

Format

File name	File format	Preservation normalization attempted	Preservation normalization failed	Already in preservation format	Access normalization attempted	Access normalization failed	Already in access format
799px-Euroleague-LE_Roma_vs_Toulouse_IC-27	PUID="fmt/116", FormatName="Windows Bitmap", FormatVersion="3.0", FormatRisk="", Extension="dib,bmp,"	Yes	No	No	No	No	No
799px-Euroleague-LE_Roma_vs_Toulouse_IC-27	image/bmp	No	No	No	Yes	No	No
BBhelmet	PUID="fmt/19", FormatName="Acrobat PDF 1.5 - Portable Document Format", FormatVersion="1.5", FormatRisk="", Extension="pdf,"	Yes	No	Yes	Yes	No	No
BBhelmet	application/pdf	No	No	No	No	No	Yes
G31DS	image/tiff	No	No	Yes	Yes	No	No
G31DS	fmt/9	No	No	Yes	Yes	No	No
G31DS	fmt/10	No	No	Yes	No	No	No
G31DS	fmt/8	No	No	Yes	No	No	No
G31DS	fmt/7	Yes	No	Yes	No	No	No
lion	PUID="fmt/91", FormatName="Scalable Vector Graphics", FormatVersion="1.0", FormatRisk="", Extension="svg,"	No	No	Yes	No	No	No

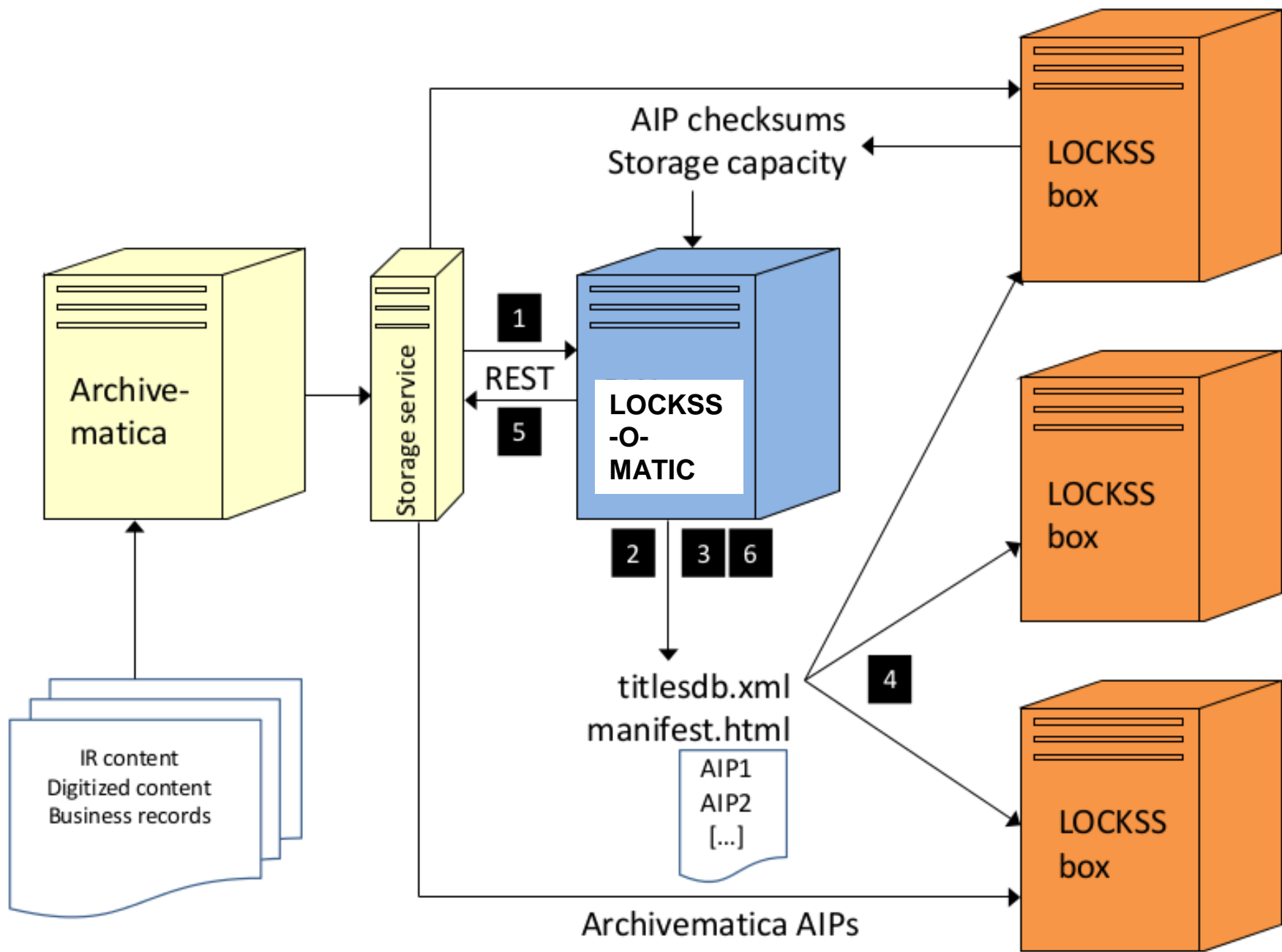
Showing 1-10 of 21

Next

STORAGE API

- We're configuring Archivematica to use a **local LOCKSS** network for storage.
- Archivematica is very good at file-level and metadata preservation, but it doesn't do much for long-term *bit-level* preservation (beyond checksums).
- Our use of LOCKSS (which stores multiple copies of each file and verifies them against one another) with the new Storage API makes up for this.

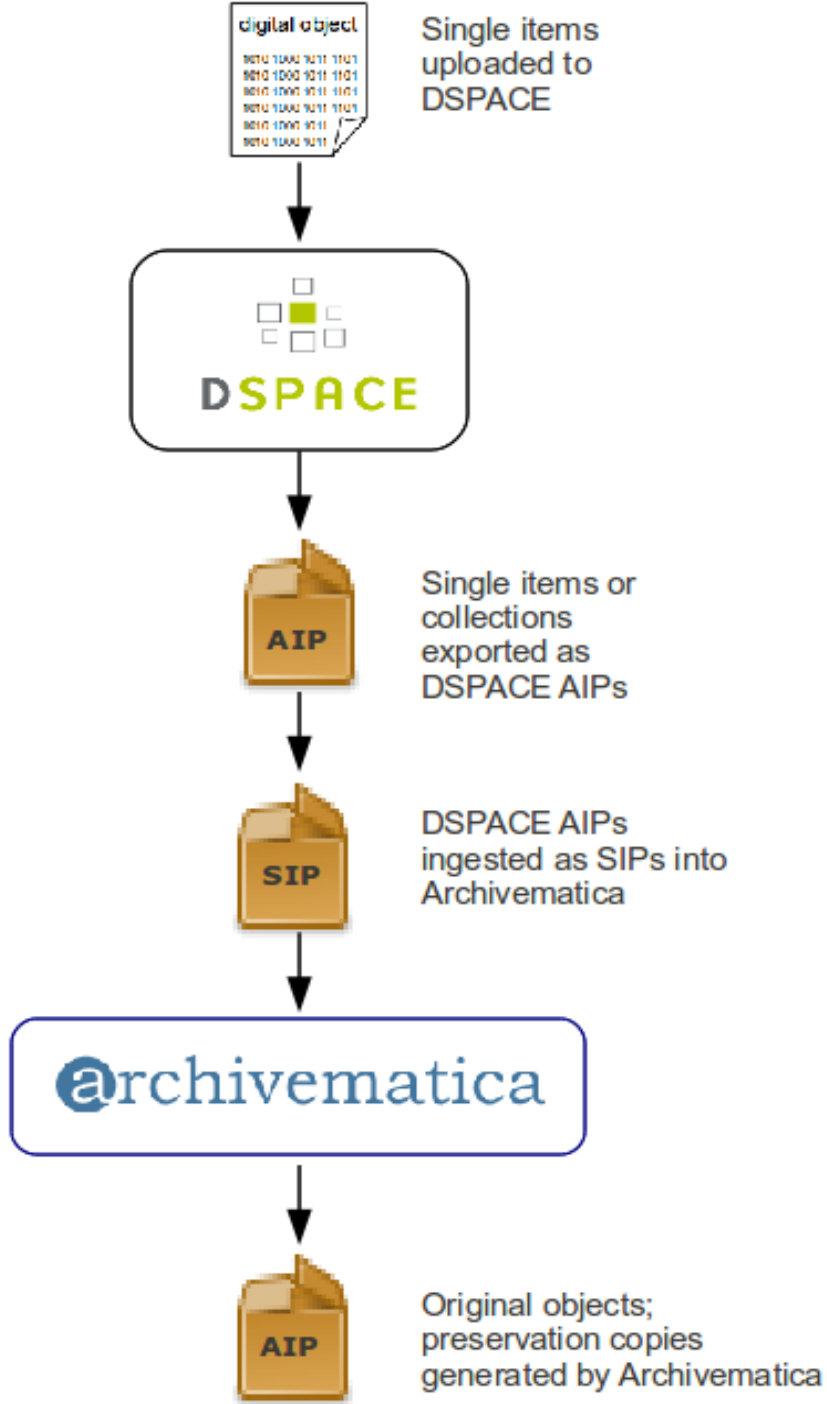




THE UNIVERSITY OF BRITISH COLUMBIA

- **Digital Preservation Program**
 - New and legacy born digital archival material
 - cIRcle, DSpace-based repository
 - Digitized collections in CONTENTdm
 - Websites
 - Research data
- **COPPUL Archivematica as a Service and EduCloud**
- **Centre for Hip Health and Mobility Research Data Management Pilot**
- **Research Data Management Training**
- **Research Engagement and Consultation Hub: REaCH-IT**



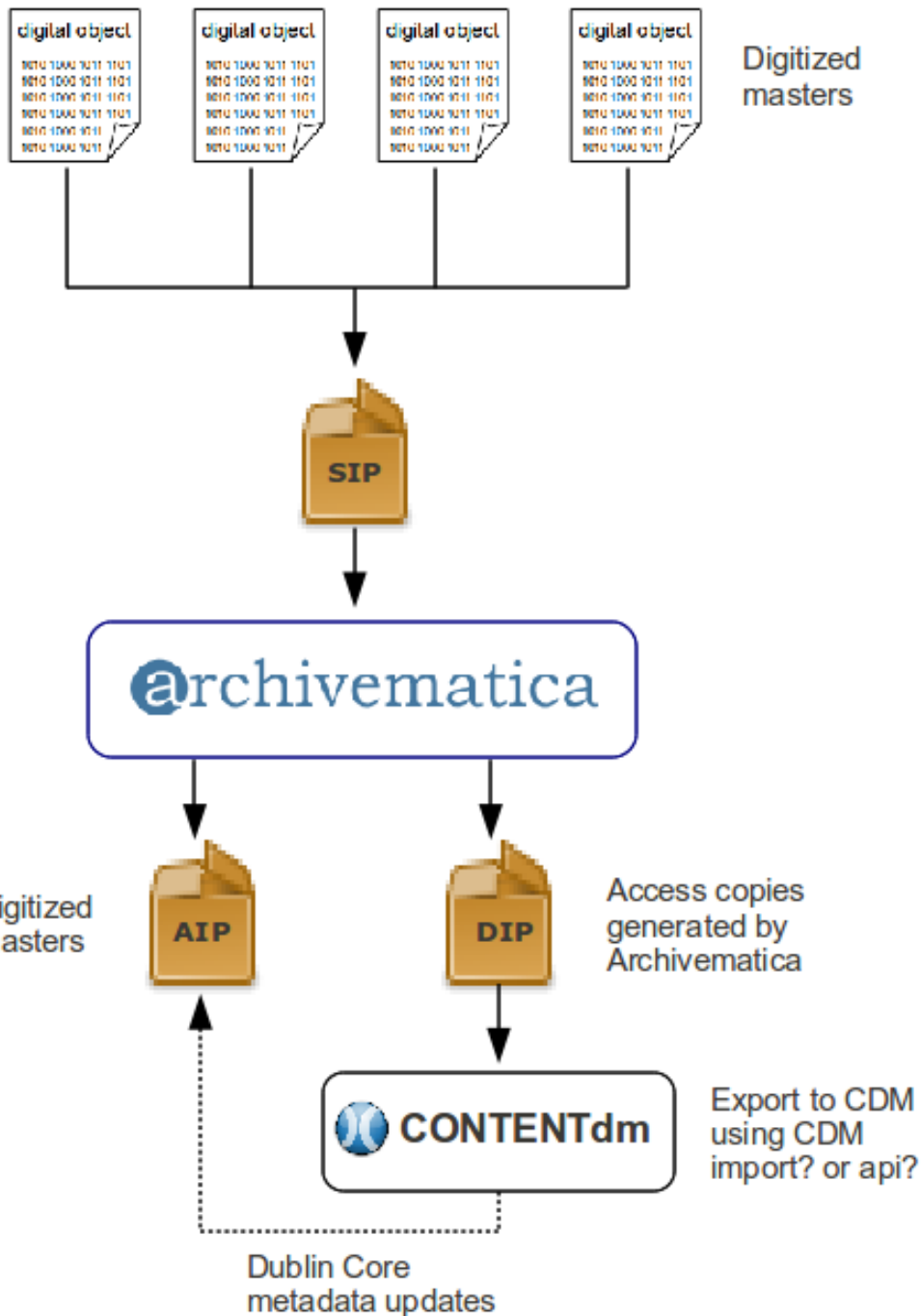


DIGITAL PRESERVATION PROGRAM

cIRcle (Dspace)

- Archivemata receives submissions from DSpace
- Also have Archivemata to DSpace workflow





DIGITAL PRESERVATION PROGRAM

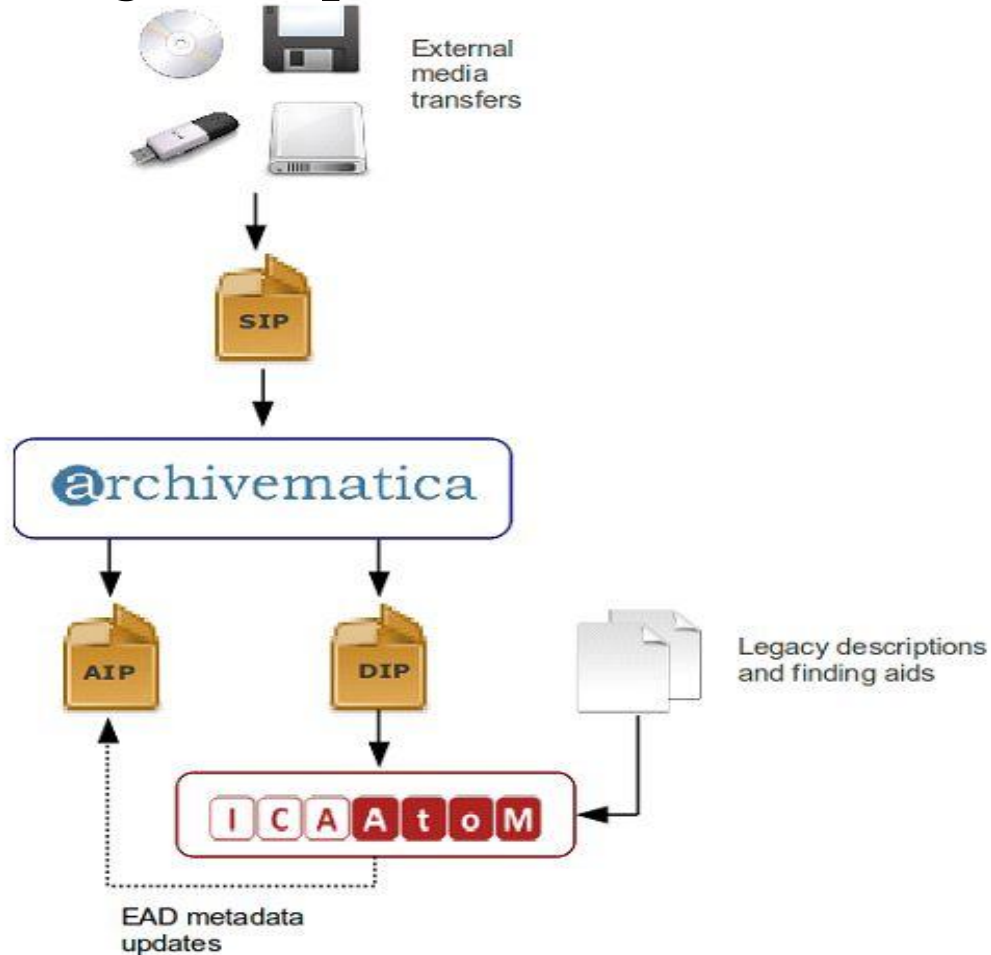
CONTENTdm

- Master files uploaded to Archivemata
- Archivemata produces access versions and pushes to CONTENTdm



DIGITAL PRESERVATION PROGRAM

RBSC/UA born-digital acquisition workflow

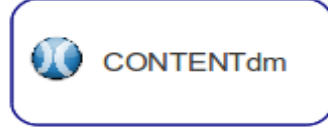




Public access provided through a unified discovery interface integrating Archivemática index and UBC Library document viewer application



- UBC faculty & students
- CHRP contributors
- Other



automated OAI-PMH updates



DIP reconfigured for upload to CONTENTdm



- UBC librarians & archivists
- Digital Initiatives, RBSC, Archives



- External media acquisitions
- Born-digital network transfers
- Web archiving harvests



processing workstations



digitization workstations



UBC SAN & backup



LOCKSS server



ARCHIVEMATICA-AS-A-SERVICE

Council of Prairie and Pacific University Libraries (COPPUL)- Digital Preservation Working Group

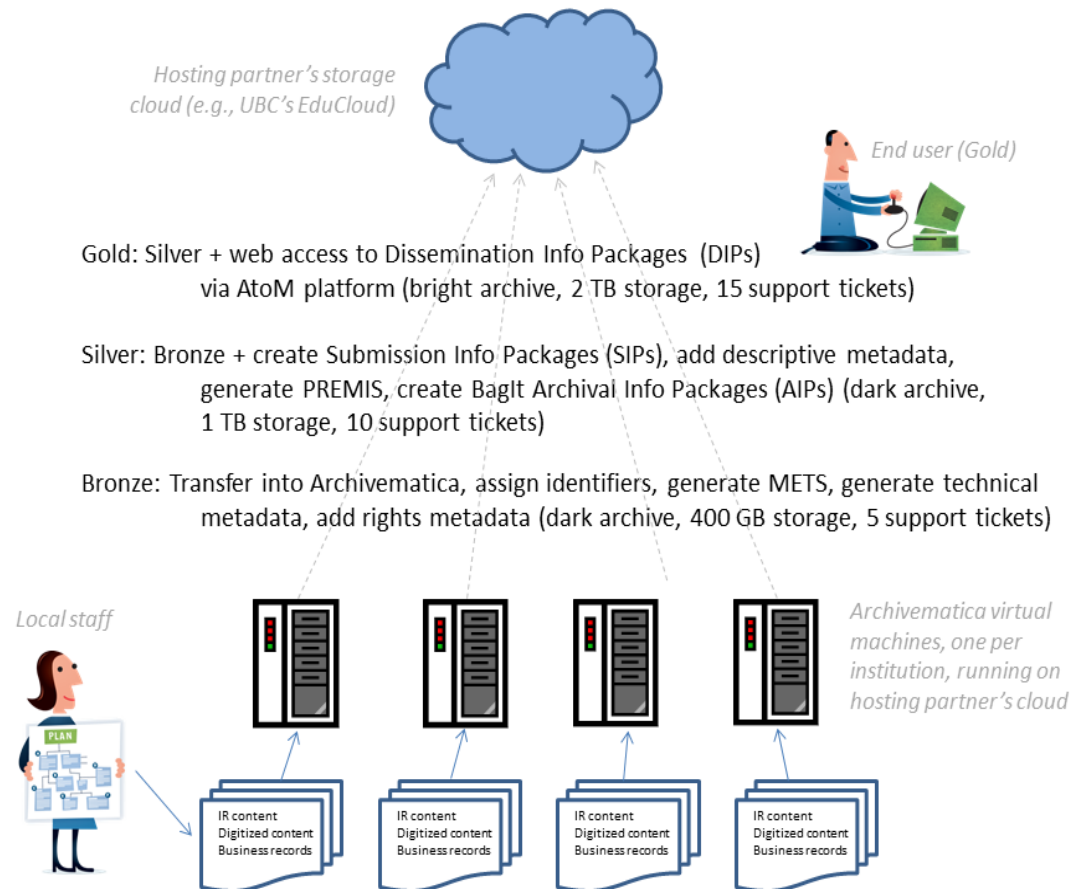
Univ of Victoria (gold)

Univ of Saskatchewan (gold)

Vancouver Island Univ (silver)

Univ of Lethbridge (bronze)

Mt. Royal Univ (bronze)



CENTRE FOR HIP HEALTH AND MOBILITY PILOT

CENTRE
for HIP Health
and Mobility

CHHM HR Flash

September 10, 2013

Welcome to Kristof Kessler!

We would like to warmly welcome Kristof Kessler who joins CHHM as lead for development of a research data management & curation strategy for the Centre. This position is jointly funded by CHHM (through the CFI Innovation Fund), UBC's Office of Research Services and the Central Library.

The University has identified support for research data management and curation as a key priority and is working with other universities in Western Canada and across the country to develop strategies to address this issue. The Centre was selected as a key partner in this project at UBC due to the breadth and scope of research data that is acquired/generated, analysed, stored and disseminated by CHHM researchers.

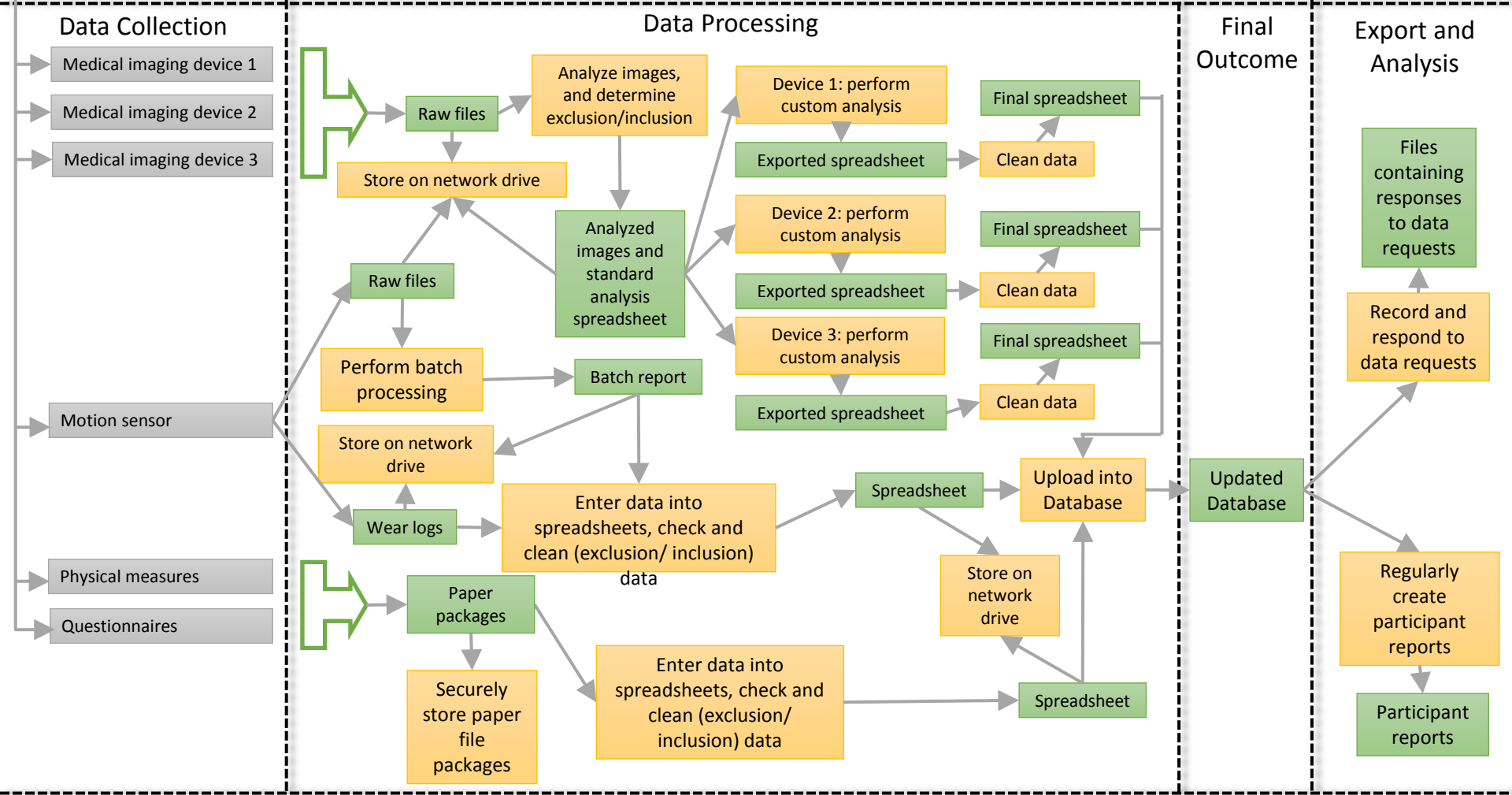
Over the next few months Kristof will carry out a full needs assessment and develop a plan for data acquisition, management and curation at the Centre. Kristof will be working closely with team members from UBC Research IT and the Library to ensure that information obtained from the Centre's needs assessment provides the framework for implementing a data management and curation strategy for the larger university research community.



Project X

Consent

Create spreadsheet templates



Project Details:
 Longitudinal study.
 Number of participants unknown.
 Participants generally suffer from some form of more or less benevolent mutation.
 Data is collected in various locations around the world, but mainly focused on the Eastern United States.

- For each Instrument:**
- Name/ID/Room/Location
 - Internal/external data source?
 - Activity performed
 - Settings
 - Operator(s)
 - Security
 - Use Frequency/Duration

- For each Outcome:**
- Types of Files/Documents
 - Number and Size of Files
 - Storage Location
 - Description/Organization
 - Security and Backup
 - Data/File Sharing
 - Need for Confidentiality, Integrity and Availability

- For each Processing Step:**
- Workstation/Laptop/ID
 - Software
 - Activity performed
 - Security and Backup
 - Team Members involved
 - Use Frequency/Duration

RESEARCH DATA MANAGEMENT TRAINING

- **Setting the context: UBC Library's involvement in data management activities (January 2014)**
- **Data 101: introduction to research data and data management for Librarians (February 2014)**
- **Storing and working with data (March 2014)**
- **Data Summit (April/May 2014)**




DEVELOPMENT OF A RESEARCH IT SUPPORT PROGRAM AT UBC

A centralized/distributed research IT support program is urgently needed to advance and coordinate the complex and specialized technology-intensive needs of the research community, address granting agencies requirements to maximize investments, mitigate risks and costs to the institution, and support an environment of research excellence at UBC.

**Research Engagement and Consultation
Hub: REaCH-IT**



UNIVERSITY OF ALBERTA

- Support for a growing research data management, curation, and preservation program - situated within a strong agenda to support digital scholarship
- Open and transparent toolkits
 - Modular, not monolithic
 - Trust what we can see, not what we've been told
- Funding – planning assumption: no new money! 



PARTNERSHIP

- **Archivematica contract and partnership**
 - An opportunity to study the issues related to the design of Archival Information Packages (AIPs) for research data sets
 - An opportunity for the development of consistent, shareable AIPs for curated research datasets
 - An opportunity to better understand data formats - automated and manual normalization workflows as part of FPR development
 - An opportunity to build further footings for ongoing and new work with regional, national and international partners like UBC and SFU
- **Use Cases:**
Canadian Polar Data Network



Image: Cow's head in a bucket. c1909. <http://peel.library.ualberta.ca/postcards/PC003821.html>



USE CASE: CPDN


HOME ABOUT SERVICES DEPOSIT SECURITY & INFRASTRUCTURE MEMBERSHIP COLLABORATIONS RESOURCES

DATA MANAGEMENT PLANNING

DIGITAL PRESERVATION

DISCOVERY

DISSEMINATION

 Canadian Polar Data Network
CPDN

Welcome

The Canadian Polar Data Network aims to develop a research data management infrastructure for the most recent [International Polar Year](#) program. We have significant experience with developing a distributed preservation backbone, curating data across the research lifecycle, producing standards-based research metadata, contributing to data and metadata standards, working collaboratively with multiple institutions and across sectors to preserve data, sharing knowledge and aligning with communities of practice. We encourage and welcome new collaborations with researchers and partners.

If you are interested in depositing data, see the [Deposit](#) page.

If you are interested in becoming a Network partner, see the [Membership](#) page.

Latest Updates

May 2013:

The CPDN is pleased to announce its completion of *The Canadian High Arctic Research Station (CHARS): Data and Information Management for Science and Technology* report, which outlines requirements, expectations, and goals for [CHARS](#) with respect to Arctic data and information management. This substantial document will aid the CHARS program as they move forward with planning.

January 2013:

Final versions of the [Data Deposit Agreement](#) and [Governance Charter](#) have been uploaded to the [Resources](#) page. The Governance Charter has also been summarized and split between the [Purpose](#), [History](#), [Principles](#), and [Governance](#) pages of this website for your convenience.

RDM planning
policy development
packaging
access
preservation
discovery

Collaborations



The [International Polar Year \(IPY\) 2012 Conference](#) drew international attention to social and economic issues. The IPY 2012 Conference contributes to the translation for action that will influence global decisions, policies and outcomes.



The [Sustained Arctic Observing Networks Initiating Group \(SAON IG\)](#) was formed to term Arctic-wide observing activities that provide free, open, and timely access to added services and provide societal benefits.



The [International Council for Science World Data System \(ICSU WDS\)](#) concept aims globally interoperable distributed data system that incorporates emerging technologies worldwide 'community of excellence' for scientific data, with searchable common stewardship and provision of quality-assessed data and data services to the international community.

BEAUFORT REGIONAL ENVIRONMENTAL ASSESSMENT

The [Beaufort Regional Environmental Assessment \(BREA\)](#) is a four year (2011 to 2015) environmental and socio-economic research to assist in preparing all parties, including to new investments in oil and gas in the Beaufort Sea. The proposal was initiated by the Region, territorial and federal governments, the oil and gas private sector, and academic institutions.



UNIVERSITY OF ALBERTA
LIBRARIES

 Scholars Portal
A Service of the Ontario Council of University Libraries



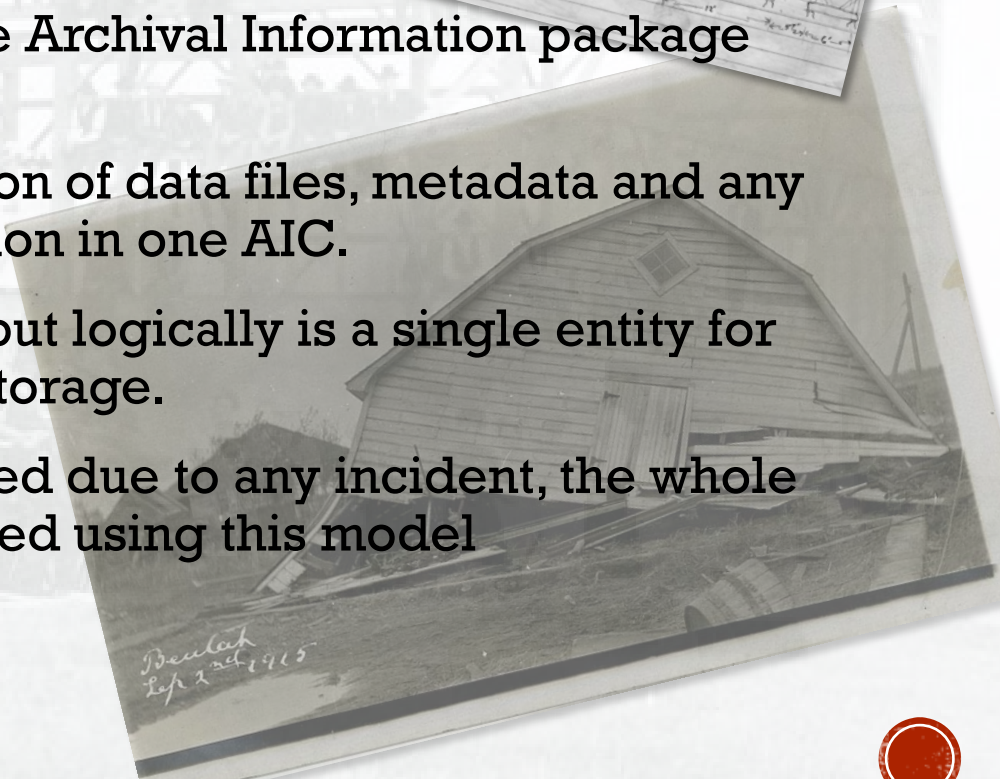
BUILDING UP WITH ARCHIVEMATICA

- **Format policy registry - automated and manual normalization**
 - Research data doesn't always play nice
- **Should an AIP be kept small, pairing a single data file with its metadata?**
 - If a set of AIPs are part of a larger aggregation, for example, a "dataset", "study" or "project," should a separate AIP exist to serve as the parent of a group of related AIPs? Or should the parent information be bundled within each offspring AIP?
- **Should a single AIP consist of all files in a "dataset"?**
 - Different definitions of what equates to a dataset. A related set of files serving some defined purpose; act as a unit.
- **Should a single AIP exist for a study or a project?**
 - Is this the right scope?
- **How important is it to be consistent in the design of AIPs?**
 - Is it acceptable to produce AIPs on the basis of all three of the above structures?
- **How much should today's physical archival storage limitations play in determining the design of AIPs?**
 - For example, if the maximum storage block is three terabytes, how much does this contribute to the decision around smaller versus larger AIPs? The maximum block size certainly limits the size of a single AIP, but should it also contribute to an overall principle around the design of AIPs.

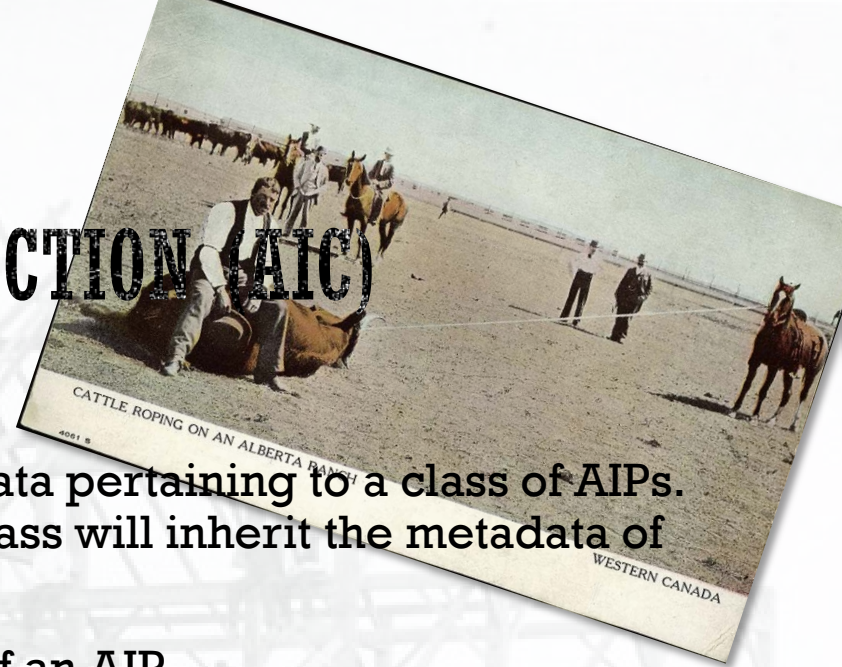


STICK FRAMING

- OAIS Reference Model
- Make use of the concept of Archival Information Collection (AIC) in combination with the Archival Information package (AIP)
- Model facilitates encapsulation of data files, metadata and any other necessary documentation in one AIC.
- An AIC can have many AIPs but logically is a single entity for transmission or for archival storage.
- If parts of an AIC are separated due to any incident, the whole structure can be re-constructed using this model



ARCHIVAL INFORMATION COLLECTION (AIC)

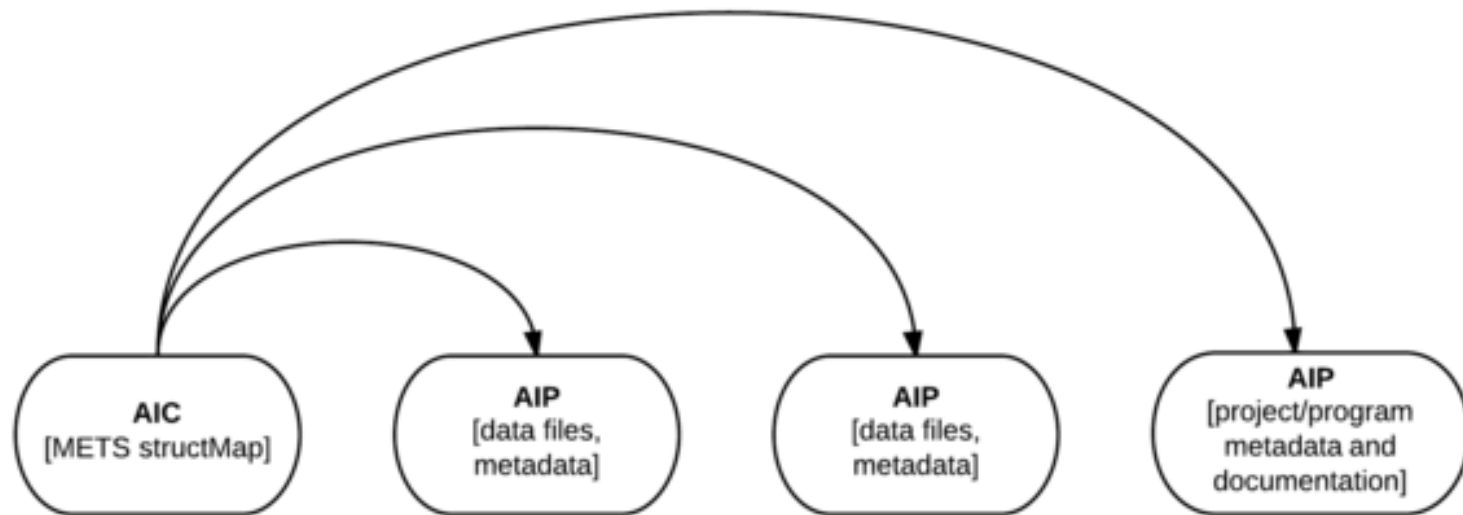


- An AIC is a type of AIP that contains metadata pertaining to a class of AIPs. All AIPs that are members of a particular class will inherit the metadata of that AIC(kept in a documentation AIP).
- The structure of an AIC is identical to that of an AIP.
- The use of AICs offers the following advantages:
 - the type and amount of information contained in an AIC is discretionary;
 - large volumes of metadata can be accommodated easily;
 - the information contained in an AIC is preserved in the same archive as its AIPs;
 - certain types of metadata for an entire class of AIPs need only be stored once;
 - a new AIP can be added without significant changes to the overall structure.
- Any orphaned AIP can easily be linked backed to its parent

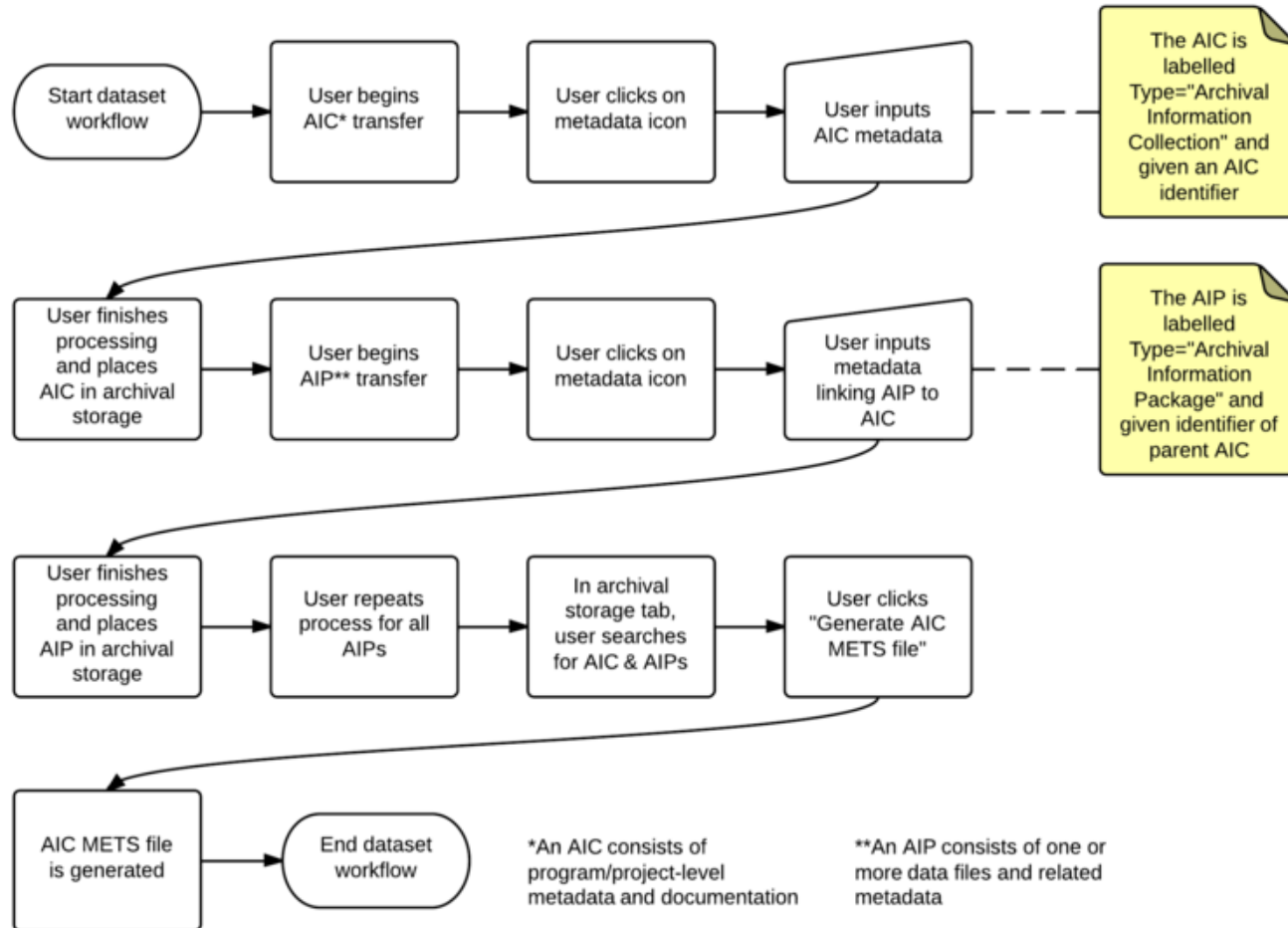
Image: Cattle Roping in Alberta. 1912. <http://peel.library.ualberta.ca/postcards/PC009286.html>



AIC OUTLINE



AIC CREATION WORKFLOW





RAISING THE ROOF

- Connect Archivemata up with existing CPDN workflows
- Connect Archivemata up with existing and new infrastructure and workflows – Fedora, OpenStack, Dataverse, etc.
 - Leverage Archivemata and ICA-AtoM for University Archives; handshake with Alfresco EDRMS
- Sustainability
 - Partner investment in common technologies will help ensure our tools remain healthy and strong
 - Growing expertise that can be shared between institutions to allow our brightest minds to do less of the same and more of what is really needed
 - Succession planning for the data and a bridge for easier sharing and collaboration within and outside of the library



TOPPING OUT

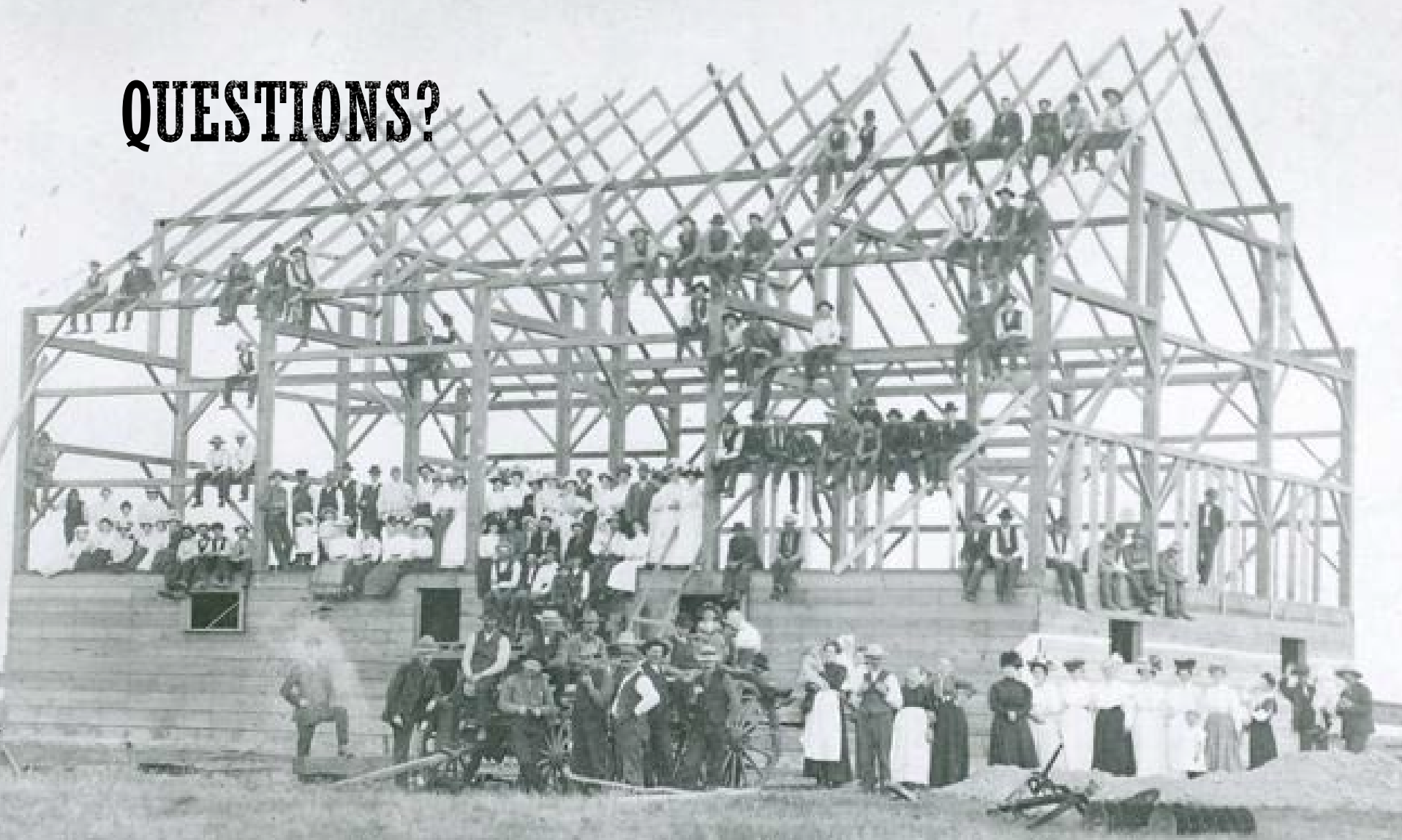


Martin Army Community Hospital "Topping Out" Ceremony

<http://www.flickr.com/photos/savannahcorps/8023265430/> Some rights Reserved.



QUESTIONS?



"Barn raising - the farm of Alex. McLaren, near Pense." Ref Number R-A4313 , Item Date 1912. Saskatchewan Archives.

