

Open Standards-Based Development for Discovery & Interoperability

Tiffany Chan   | tjychan@uvic.ca

Senior Developer | University of Victoria Libraries 

15 June 2023 | Open Repositories 2023

Link to slides: uviclibrary.github.io/OR2023Pres



Speaker notes

- We are a mid-size university by Canadian standards
- We use several platforms to deposit and disseminate our stuff, including DSpace, Internet Archive, Dataverse and others
- One of those platforms, which I'm going to describe in detail is called Vault

Link to slides: uviclibrary.github.io/OR2023Pres

University of Victoria ("UVic") Libraries

~20,000 students, 900 full-time faculty, 150 library staff

Multiple digital platforms for our collections

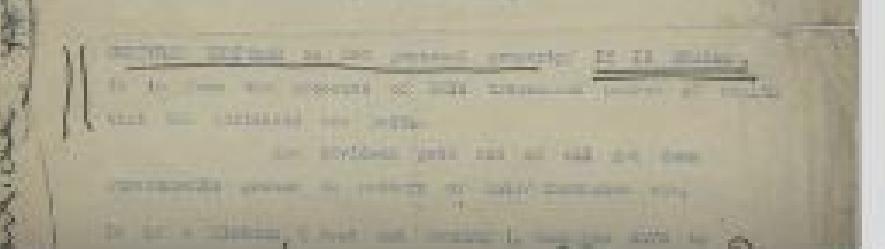
Speaker notes

- We call it Vault because it mostly holds digitized objects from our Special Collections and Archives (typically our most unique and rare materials)
- Also houses materials from faculty and community partnerships, such as audio recordings or oral histories
- Vault is built on an open-source software called Hyku, which is maintained by the Samvera community
- At present, we have 94 collections with over 12000 works in them. Those works are comprised of over 98000 files; about 70% of works were previously in ContentDM
- To give you an idea of staffing levels, we have 2 developers who spend ~80% of their time on Vault and 2-3 metadata staff who spend 5-30% of their time creating or editing metadata

F.T. Prince Collection

Frank Templeton Prince (1912-2003) was an English poet and academic. The collection consists...

Ezra Pound Collection



Vault

Holds digitized materials from Special Collections & Archives, faculty/community partnerships

Customized instance of **Hyku (Samvera)**, migrated from ContentDM starting in 2018

90+ collections, 12000+ works, 98000+ files

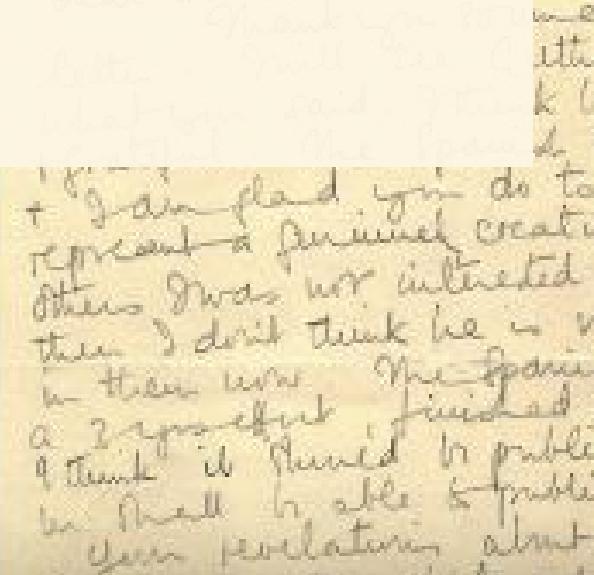
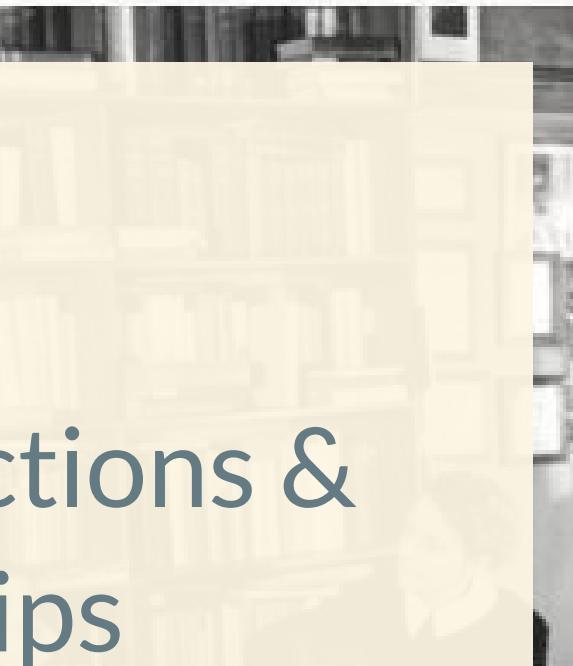
2 developers, 2-3 metadata staff working directly in/with Vault (among other projects)



HIV In My Day



Herbert Geddes



Herbert Read Collection

ake Seel
lection

Bombs

victoria
rary Local
ollection

ake Seel
lection

Bombs

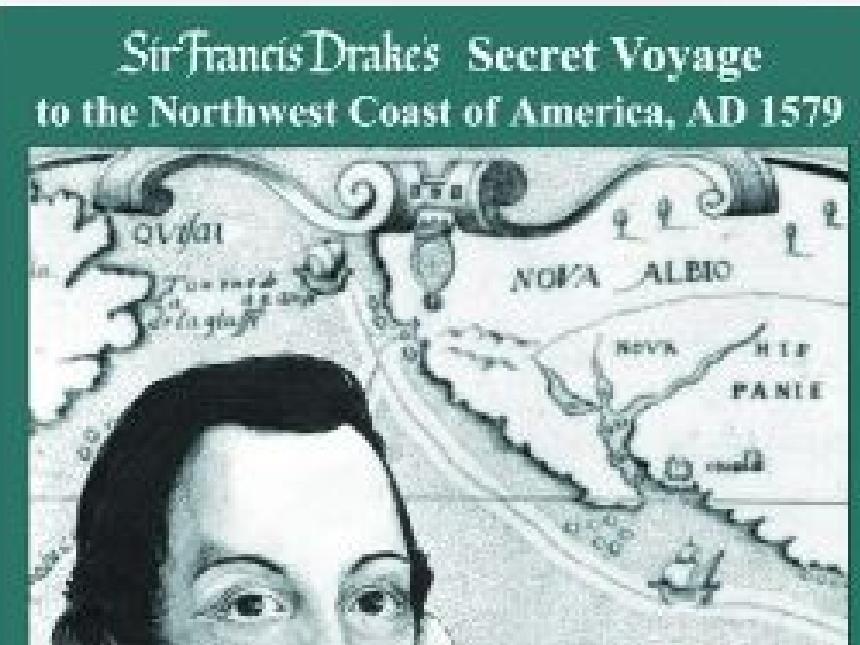
victoria
rary Local
ollection

Ezra Pound Collection



F.T. Prince Collection

Frank Templeton Prince (1912-2003) was an English poet and academic. The collection consists...



Example Collections

Sylvester Collection



HIV In My Day

Collection



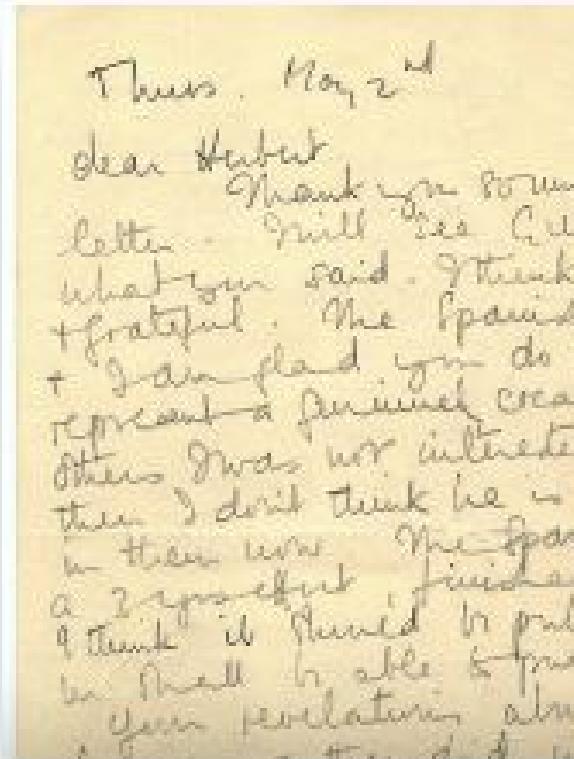
Herbert Geddes



FTM



Gisèle Freund Photographs of James Joyce in F



Herbert Read Collection

Speaker notes
No notes on this slide.

Sir William Orpen to Lady Beatrice (Or-21)

works.

Feature

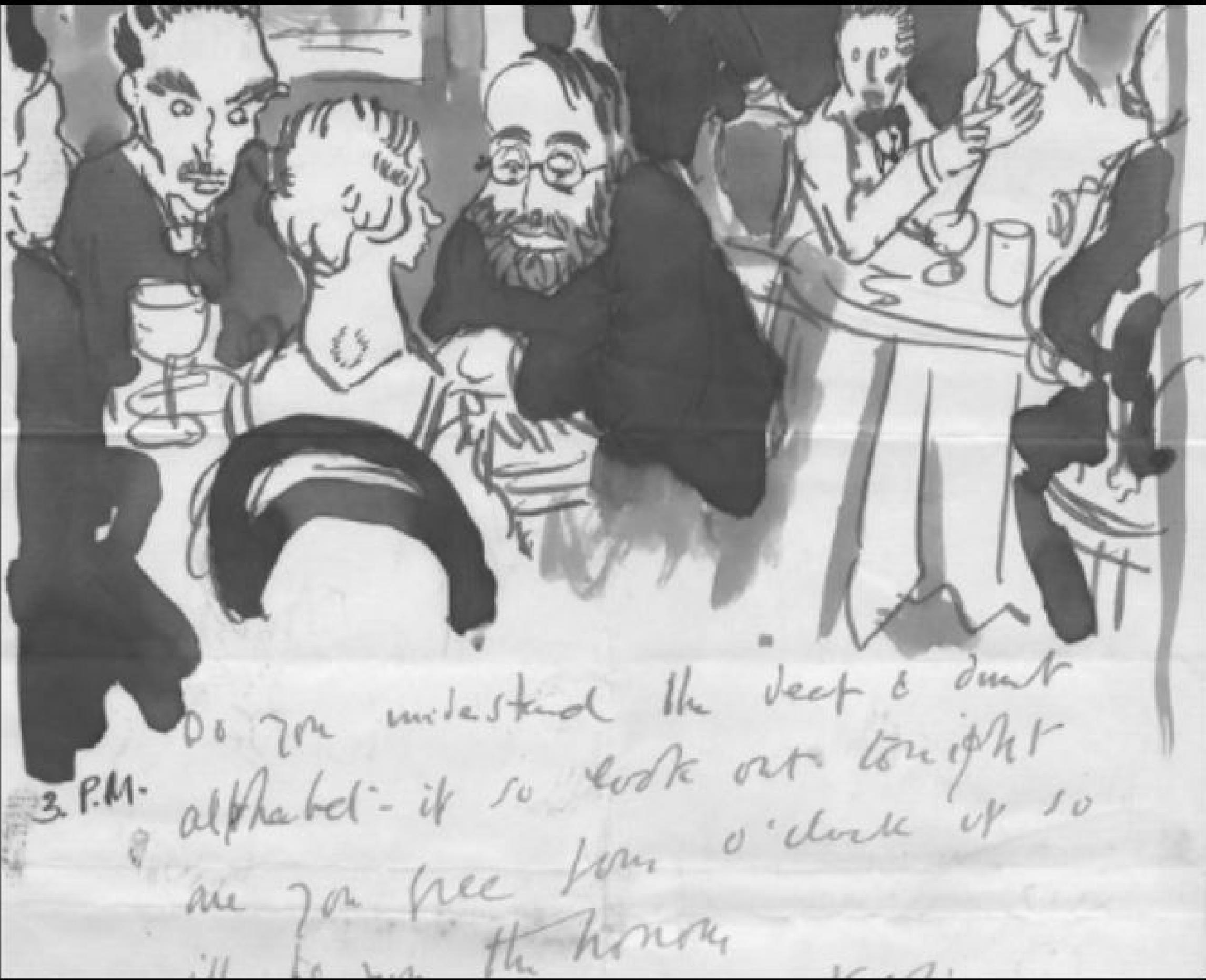
Analytics

Speaker notes

- We have several collections of correspondence (letters) in Vault, many from the early 1900s that have been scanned and uploaded as high-resolution images
- IIIF viewer pictured here

Add to collection Edit **Sir William Orpen Illustrated Letter**

Sir William Orpen to Lady Beatrice Glenavy (Or-21)



Mathew Ko Colour Films: Victoria's Chinatown and Region



Speaker notes

- Mathew Ko Colour Films collection is a collection of home movies from the early 1950s
- Many of our collections, like this one, depict local places or events
- When we were choosing a platform, we wanted one that could handle a diverse range of media and file types
- Examples of things we do NOT put in Vault: theses & dissertations (they go in DSpace), datasets (Dataverse)

Speaker notes

- Why did we choose Samvera for these types of collections? We wanted a platform that supported openness and interoperability
- For example, Hyku uses IIIF, which is a set of web standards and APIs for disseminating & accessing images
- Hyku creates IIIF manifests automatically when ingesting images and provides a IIIF server as well as image viewer
- Hyku also uses Fedora as its underlying repository system, which has built-in support for Linked Data, an important factor for our institution

Why Migrate? Support for Open Standards

IIIF: automatic manifest generation, image server ([Riiif](#)),
and viewer ([Universal Viewer](#))

Linked Data Platform with [Fedora](#)

- Hyku also lets us use different controlled vocabularies to make our data more interoperable (i.e. able to be understood by other systems that use or are aware of those vocabularies)
- Controlled Vocabulary: "vocabulary" as in a system for describing things (as when cataloguing), "controlled" because it's a pre-determined, normalized list of values
- This is a table showing all our metadata fields that use controlled vocabularies, along with the specific vocabulary that we use
- As you can see, most of our fields use OCLC FAST, except for our rights and genre fields
- One of our biggest, most resource-intensive challenges during migration was not just dumping data out of ContentDM, but also reconciling that data with a new Vault data model
- This often meant converting textual values into unique identifiers (URIs)
- * Advance slide *
- For more details on how we implemented and migrated to FAST, you can read an article that my colleagues, Dean Seeman, Karen Dykes, and I wrote on our implementation of FAST. I've linked to it on this slide.

Metadata: Linked Data (URI) Fields

Field Name(s)	Controlled Vocabulary
Provider, Creator, Contributor, Subject, Geographic Coverage, Physical Repository	(OCLC) FAST
Genre	Getty Art and Architecture Thesaurus
Rights Statement	rightsstatements.org , Creative Commons

* See Figure 3 in [our recent article](#) for a more detailed table and discussion of metadata

- This is just an example of what the metadata looks like as indexed in Vault
- For every controlled vocabulary field, we have both the URI and the human-readable label saved in a corresponding "label field"

URIs and Labels on the Backend

```
"creator": ["http://id.worldcat.org/fast/788816"],  
"creator_label": ["Orpen, William, Sir, 1878-1931"],  
"contributor": ["http://id.worldcat.org/fast/1050534"],  
"contributor_label": ["Glenavy, Beatrice Moss Campbell, Baroness, 1883-1970"],  
"subject": ["http://id.worldcat.org/fast/1050534",  
            "http://id.worldcat.org/fast/78887",  
            "http://id.worldcat.org/fast/2012667",  
            "http://id.worldcat.org/fast/1050534"],  
"subject_label": ["Painters--Correspondence",  
                 "Orpen, William, Sir, 1878-1931",  
                 "Glenavy, Beatrice Moss Campbell, Baroness, 1883-1970",  
                 "Painters--Biography"]
```

Labels in the Interface

Creator

Orpen, William, Sir, 1878-1931

Contributor

Glenavy, Beatrice Moss Campbell, Baroness, 1883-1970

Subject

Painters--Correspondence

Orpen, William, Sir, 1878-1931

Glenavy, Beatrice Moss Campbell, Baroness, 1883-1970

Painters--Biography

How can people or machines access our linked data?

1. Go to the page for any work
2. If there is a question mark in the URL, delete that and everything after it
3. Add slash manifest to the end to see the manifest, or add .ttl to the end to see turtle (RDF)

Accessing Linked Data via the IIIF Manifest or RDF/TTL

Go the page for any work and add /manifest.json or .ttl
to the end of the URL

Example:

vault.library.uvic.ca/concern/generic_works/1a597e67-17a8
-4c97-9c63-c835a41f5474 [.ttl](#)

- The other important consideration is that Vault/Hyku is open-source software, meaning anyone can use, examine, alter and redistribute its source code (definition from ibm.com)
- This lets us build custom features that we use in production

Why Migrate to Open-Source Software

More flexibility and control over how we describe, organize, and present our objects and collections

Customized features and interface tools

```
18
19
20
21
22
23
24
25
26
27
```

This repository provides you a development environment

```
</p>
<div class="height-30p"></div>
<div class="d-flex">
  <a href="/" class="btn btn-primary font-weight-bold">
    <a href="/" class="d-flex align-items-center gap-2">
      
      <span class="ml-2" style="font-size: 0.8em;">View GitHub Repository
    </a>
  </div>
```

Speaker notes

- One of these custom features is support for EDTF, or Extended Date/Time Format, which was a request from our metadata department
- EDTF uses special symbols to express complexity in dates that you couldn't get with basic ISO formats
 - * *Advance slides* *
- The goal was to save and index EDTF date strings and then display humanized versions on the front end

Extended Date/Time Format

EDTF: a way to express nuanced dates (using symbols) in a way that basic ISO formats cannot

Examples: 1912~ means "approximately 1912"

.../1000 means "Before 1000 (AD)"

- To give an example, here is what the data looks like when our metadata staff enters it into Vault...
- * *Advance slide* *
- ...and here is what the date looks like in our indexing software. We use the [edtf-ruby gem](#), developed by Sylvester Keil and others, to parse the EDTF date string
- Once parsed, the date is saved in 2 additional formats: 2) the earliest possible date is saved for sorting items; 3) a list of numbers/integers that correspond to each possible year.

EDTF Date Indexing

Dates are entered using EDTF notation...
And indexed in multiple formats

```
"date_created": ["1476/1500~"],  
"year_sort": "1476-01-01T00:00:00Z",  
"year_range": [1476, 1477, 1478, 1479 ... 1500]
```

- * Scroll to Date Created field and highlight text *
- Here's what EDTF looks like in the interface, where we use the human-readable date, which are generated by the [edtf-humanize](#) gem by Duke Libraries.
- The date links to a search for all items created in the same year or year range using the [Blacklight Range Limit](#) gem.
- * Click link *
- This draws from the list of years generated from the EDTF date.

[Home](#)

Leaf from a Medieval Book of Hours, Fragment Public

[Analytics](#)[Download PDF transcription](#)

Description and transcription completed by Guifen Jiang, as part of work for a manuscript studies class.

Audiuit dominus et misertus... Holocaustum et pro peccato non.

Book of Hours: Use of Paris, Matins of the Office of the Dead, Third Nocturn.

Codicology

Single sheet of very thin parchment, creamy and well-preserved. Text is written in one column of 18 lines, in black ink, with burnished gold initials and red and blue splashing of capital letters. Dimensions are 160x110mm, with text block measuring 105x90mm, including the floral border on the recto (105x20 mm). Reddish plummet ruling visible.

Paleography and Decoration

One scribe for main text throughout, writing Latin in a Gothic Northern textualis, c. 1475–1500. A second hand is responsible for rubrication and decoration. Three-coloured initials (red, blue, and gold). Floral border in the right margin on the recto is decorated with strawberry and green leaves on a gold background, with blue and gold curly penwork on the upper and lower part of the pattern. Two types of line fillers are used: a blue and red bar, and a flower.

Textual Remarks

Leaf contains part of Matins of the Office of the Dead, Third Nocturn, according to the Use of Paris. Contents may be compared to ff. 174v3–176r5 of a full Use of Paris Book of the Hours housed at Tufts University (Medford, Tisch Library, Tufts University, MS 2, available online at: <https://dl.tufts.edu/concern/pdfs/r781wt689>).

In Collection:

[Medieval and Early Modern Manuscripts Collection](#)

Subject

[Prayers and devotions](#)
[Manuscripts, Medieval](#)
[Manuscripts, Latin \(Medieval and modern\)](#)

Language

lat

Identifier

Shelf Mark: Frigm.Brown.Lat.8

Date created

1476 to approximately 1500

Welcome to Vault, the digital asset management system used by UVic Libraries to store and provide access to digitized materials both from the rare and unique holdings in Special Collections and University Archives and from UVic Libraries' project partners.

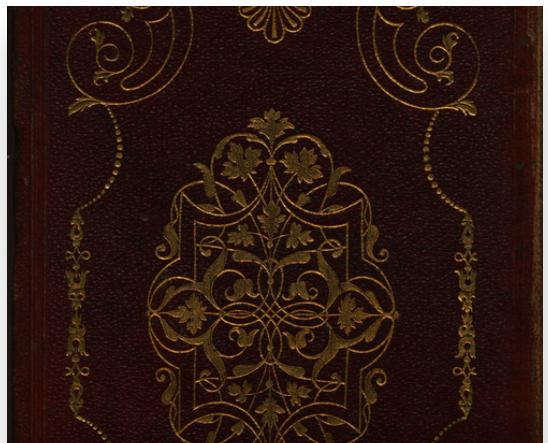
Click on a collection below to get started or use the search bar to search across collections. For additional guidelines on searching Vault, see the [Help section](#).

- We've also heavily customized our home page. Here, users can quickly run faceted searches by genre, time period (again, EDTF dates), popular subjects, or (geographic) place.
- We also have the ability to upload collection thumbnails or select them from works in the collection
- Collections are also listed alphabetically under "All Collections"
- For staff, we also display recently created works or collections

Browse by Genre

[Genre](#) [Time Period](#) [Subject](#) [Place](#)
[Diaries](#) [Historical Maps](#) [Letters](#) [Photographs](#) [Periodicals](#) [Serials](#) [Sound Recordings](#) [Video Recordings](#) [More »](#)

Featured Collections

[All Collections](#) [Featured Collections](#) [Featured Works](#) [Recent Collections](#) [Recent Works](#)


19th Century
Periodicals



Victoria Feminist
Newsletters



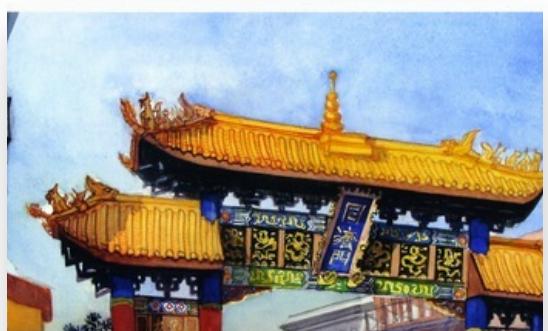
Hudson's Bay
Company Maps



Transvestia

Homepage Facets

Allows users to filter using broad categories (vault.library.uvic.ca). Users can also browse collections and works visually.



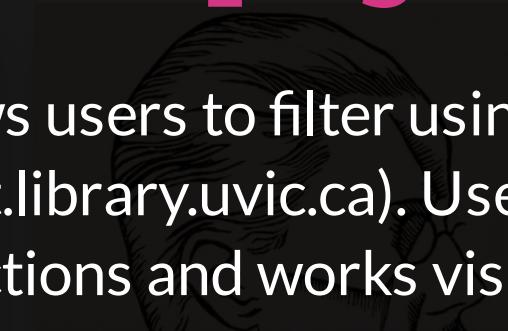
Victoria's Chinatown, a gateway to the past and present of Chinese Canadians



Community Mapping Project



E.M. Forster Collection



Modernism and Post-War Literature Collection



University
of Victoria

Quicklinks

[Hours and Locations](#)
[Main Library Site](#)
[Digital Exhibits](#)
[Terms of Use /](#)

Contact

[Contact Form](#)
[250-721-6673](#)
[778-718-5745](#)
[More Library Contact Info](#)

[Home](#)

Interview with Nicki Ward Public

[Analytics](#)[Download video](#) | [Download transcript \(PDF\)](#)

Interactive Video Transcripts

Nicki Ward discusses her involvement in organizing and participating in the first Toronto Trans March, as well as running for political office.

Narrator Name: Nicki Ward. Interviewer Name: Evan Taylor. Pronouns: She.

Trans Activism Oral History Project

Able Player uses .vtt files (Web VTT format) to create interactive transcripts for videos when available.

Subject
Ward, Nicki

Publisher
TheArquives
Transgender Archives

Language
eng

Keyword
Graphic novel
Activism
Health care

C-279

Trans Memorial

ID

Trans March

Police

C-16

Bill C-389

Toronto

Political Activism

Speaker notes

- Rather than use the default video viewer that comes with Hyrax/Hyku, we use an accessible audio/video player called [AblePlayer](#), created by Terril Thompson.
- AblePlayer can automatically display subtitles and transcripts from a specific file format called WebVTT or web video text tracks.
- With this player, users can click on specific points in the transcript to skip to a specific timestamp (*demo*).

- Vault's code is open-source and is available on Github
- For local testing and development, we use a software called [Docker Desktop](#) (we don't run Docker in production, although other institutions do)
- In general, we find this more approachable than installing and troubleshooting all of Hyku's dependencies individually
- It can even run on a laptop (albeit slowly), depending on the laptop

Running the Code

Code: github.com/UVicLibrary/Vault

Switch to the `docker_multitenant` branch

Instructions: run Vault locally with Docker

Speaker notes

- None of the software customizations I've just shown you would be possible with proprietary software like ContentDM.
- However, there are some things to consider when considering switching to open-source software like Hyku
 - * Advance slide *
- Although the application itself is free to use, maintaining and customizing it is not; it's a question of where you want to spend the money (product vs. people/labour)
- The major advantage of open-source is the freedom to modify (creative control)
 - * Advance slide *
- Local open-source development also allows developers like me to be more responsive to requests or bugs; priorities are determined locally by people doing the work rather than a vendor
- This encourages more communication between units
 - * Advance slide *
- Hannah Frost points out that code customizations inevitably create "complications in the upgrade path." Any major upgrade to the underlying code base requires a lot of work to sort through and resolve conflicts

On locally developed, open-source software

Not free (\$) but freedom: labour vs. product

More responsive, shorter development cycles

Hannah Frost: tension between customization/sustainability

Thanks!

Tiffany Chan   | tjychan@uvic.ca

Link to slides: uviclibrary.github.io/OR2023Pres

Slides made with [reveal.js](#)