

CNI Briefing

OCUL Artificial Intelligence & Machine Learning (AI/ML)

Jacqueline Whyte Appleby

Catherine Steeves

August 2024

Ontario Council of University Libraries

- OCUL founded in 1967
- OCUL Scholars Portal est. 2002
 - a shared technology infrastructure to provide common platforms for digital content and services to support research

ABOUT SCHOLARS PORTAL

- Scholars Portal is OCUL's shared digital services and infrastructure arm
- Based at the University of Toronto, 30 staff (librarians, programmers, systems staff)
- Via service agreements SP is increasingly extending to institutions across Canada
- Establishing and supporting national platforms



Key Current State Metrics – SP Services

183 / 49

Number of VM instances /
Number of Physical Servers

Breakdown by Platform Type

- Windows - 37
- Linux - 146

Breakdown by VM Specification

- Small - 50
- Medium - 80
- Large - 58

5-7 years

Average Lifecycle of Equipment

- Servers – 5 years
- Storage – 5 years
- Network Equipment – 5 years
- Non-prod infrastructure can be extended up to 7 years on an as-needed basis

1

Number of Data Centres

- ~40% of the data centre floor space may be available for additional equipment
- 4 nodes at other universities for OLRC
- No secondary site

1 PB*

Collective Storage

- 72% of the provisioned storage capacity is currently utilized
- ~450TB of backup is stored on physical tape drives, replicated across two locations

Breakdown by Storage Type

	Provisioned Size	Consumed
Direct Attached Storage	224 TB	118 TB
Network File Storage	700 TB	542 TB
Tape Storage	1,000 TB	443 TB

10%-20%

Typical Growth Rate per year

- 10% - 20% - Server VMs
- 15% - 20% - Storage
- 10% - Network Traffic

1 PB/year

Network Traffic

- 487 TB/yr – Traffic inbound to UTL
- 537 TB/yr – Traffic outbound from UTL
- Network traffic is common across UTL & SP



* Note: Operational storage is a 1/2 or 1/3 factor of raw physical storage due to data replication included in the storage architectures and protocols (e.g. RAID10/6 converts to 1/2 while OpenStack Swift converts to 1/3)

AI/ML Explosion in Public Consciousness

Generative AI has **exceptional transformational power** due to its:

- Scope of impact and power as a catalyst for innovation
- Ability to automate complex tasks that traditionally required human intelligence
- Ability to make sense of vast amounts of information and data and generate insights
- Ability to adapt and learn resulting in continuous improvement
- Capacity to enhance human capabilities and augment creativity in art and design
- Accessibility – availability and rate of adoption

Fall 2023 - OCUL AI/ML Task Force

- Mark Asberg, Queen's University
- Talia Chung, University of Ottawa
- Scott Gillies, Wilfrid Laurier University
- Vivian Lewis, McMaster University
- Mark Robertson, TMU
- Catherine Steeves, Western U (Chair)
- Amy Greenberg, OCUL
- Kate Davis, Scholars Portal
- Michael Ridley, University of Guelph

Engagement

BUILDING A COMMUNITY AND FOSTERING COLLABORATION:

Consultations – began spring/summer 2023

OCUL Directors support in principle and start-up funds – Fall 2023

National community of practice – launched January 2024

Hackfest (McMaster) – February 2024

Interim Report and National Summit – March 2024, Final R. May'24

Building connections – ongoing: CARL, CRKN, iSchools, CUCCIO, Internet Archive, Alliance, LAC, other regional consortia...

OCUL ML/AI Task Force Goals

- Provide a compelling case for collaborative action
- Identify relevant use-cases, issues, and challenges
- Present strategic action plan to build capacity to explore and implement AI/ML

March 2024 - Interim Report

“... the premise of the report and strategy is that machine learning will have a **transformational impact on research libraries** that is **best understood, managed, and realized if OCU Libraries work together** on a select number of core initiatives.”

May 2024 - Final Report

GROUNDING IN PRINCIPLES FOR "RESPONSIBLE AI" :

- Centre equity, diversity and inclusion
- Enhance accessibility
- Human-in-the-loop development
- Ensure transparency and explainability
- Integrate open-source tools and solutions

May 2024 - Final Report

“Our strategy is as follows:

1. Select AI/ML use cases that can be implemented as operational projects for the benefit of library users, library staff or both.
2. Implement these projects as consortial, collaborative initiatives.
3. Strive to maximize the availability of tools, services, and outcomes to support both English and French speaking learners and researchers.
4. Use these projects and related work to build capacity in libraries regarding AI/ML awareness, knowledge, and technical skills.
5. Enable transformational change to further the mission of libraries.”

OCUL AI/ML Program Recommendations

- Program of five projects
- Resource requirements
 - Identify and apply for funding opportunities
 - Two years of OCUL New Initiative Funds
 - Three key positions (one sponsored by UT CIO)
 - Local compute
- Governance and PM structure

Operational Projects

- Audio to text pipeline
- Virtual reference
- Government documents ML/RAG model
- Accessible Content E-Portal (ACE) document remediation
- Capacity building

Projects: Audio-to-text pipeline (Whisper)

The image shows a screenshot of a file manager window on the left and a text editor window on the right. The file manager window displays a directory structure for 'mnt/temp-globus' with various subdirectories and files. The text editor window shows the content of the file 'sp_days_2024_day_1.txt', which contains a transcript of a presentation.

File Manager Window:

- 172.27.110.74 – SFTP
- wangr@172.27.110.74
- /mnt/temp-globus
- Filename
- alberta
- algoma
- brock
- carleton
- guelph
- input
- options.txt
- output
- sp_days_2024_day_1
 - sp_days_2024_day_1.json
 - sp_days_2024_day_1.mp3
 - sp_days_2024_day_1.srt
 - sp_days_2024_day_1.tsv
 - sp_days_2024_day_1.txt
 - sp_days_2024_day_1.vtt
- sp_days_2024_day_2
- laurier
- ottawa
- queens
- toronto
- trent
- waterloo
- windsor
- wlu
- york

Text Editor Window:

sp_days_2024_day_1.txt

Hello everybody, welcome to Scholars Portal Days 2024, Collaborative Intelligence. My name is Sabina Pagato and I will be your host for today. So just a couple of housekeeping notes before we get started. The first thing is that this session is being recorded and the recording will be posted to the Scholars Portal YouTube channel after the events. If you want to see the slides that our presenters are using, you can use SpotDocs. All the slides from today are posted on the SpotDocs page, which we could drop that link in the chat. That would be amazing. We also have captions that are available today. So you can turn on captions using the Live captioning zoom bar. You'll notice there's a few different ways that you can interact today. You can use the Q&A module if you want to ask a question to the presenters. You can also view questions other people have asked the presenters and vote them up if it's a question that you also would like to know the answer to. You can use the chat to talk to fellow attendees. The default for the chat is to go to hosts and panelists. So if you want to chat with everybody, you can just change that in the drop down menu to everyone and chat about the program today. We do ask you to be considerate and respectful. Disrespectful behavior will not be tolerated and you may be kicked out of the meeting if you engage in disrespectful behavior. We also wanted to mention that we do have a hashtag if you want to chat about today on social media. That is at SPdays24. So before we get going, we just wanted to quickly acknowledge the land that we're joining you from today. We want to acknowledge the land on which Scholars Portal operates and on which most of our staff live and work. For thousands of years, it's been the traditional land of the Huron-Wendat, the Seneca, the Mississaugas of the Credit. Today this meeting place is still the home to many Indigenous people from across Turtle Island and we're grateful to have the opportunity to live and work on this land. We know we have people joining us from across Ontario today, even from across Canada. So we invite you to reflect on the land you're joining us from today. If you wish, you can add your own land acknowledgement into the chat. We will also be having a break later this afternoon where we share some images of what spring looks like in the different places where we're joining from today to have a bit of an opportunity to reflect on our own individual relationships with the land that we live on. So here's a quick overview of today's agenda. As I mentioned earlier, our theme is collaborative intelligence and today, day one, is really

Project: Virtual reference

- Development partner with LibraryH3lp
- Enhancement to library staff offerings
- Advancing research/analysis of transcripts



Project: Government documents metadata enhancement

INTERNET ARCHIVE WEB BOOKS VIDEO AUDIO SOFTWARE IMAGES JACQUELINE WHYTE AP... UPLOAD Search

ABOUT BLOG PROJECTS HELP DONATE CONTACT JOBS VOLUNTEER PEOPLE

University of Toronto Government Publications Digitization Collection

These documents have been digitized as part of the University of Toronto's project to fully digitize its holdings of Canadian federal and Ontario provincial government publications. At this time, these documents are best searched using a full-text search. Additional descriptive metadata work is ongoing. Three key things to note about this collection:

More...

COLLECTION ABOUT

50,467 Results

Search Search this collection [Advanced Search](#)

Search metadata
 Search text contents

Sort by: Weekly views Title Date published Creator

Title	Creator	Type	Views
31761114660335			3,029
The official motorcycle handbook	Ontario. Ministry of Transportation		437
TRADE OF CANADA WITH THE UNITED STATES	CANADA. DOMINION BUREAU OF STATISTICS		835
Telephone directory: Ottawa-Gatineau		2002	588
31761114660384			184
SUMMARY OF TRADE OF CANADA	CANADA. DOMINION BUREAU OF STATISTICS		822
31761114667769			218
EVOLUTION OF SOCIO-TECHNICAL SYSTEMS - A CONCEPTUAL FRAME...	Trist, E. L. (Eric Lansdown)	1981	290
FINANCING OF MOTOR VEHICLE SALES	CANADA. DOMINION BUREAU OF STATISTICS		274

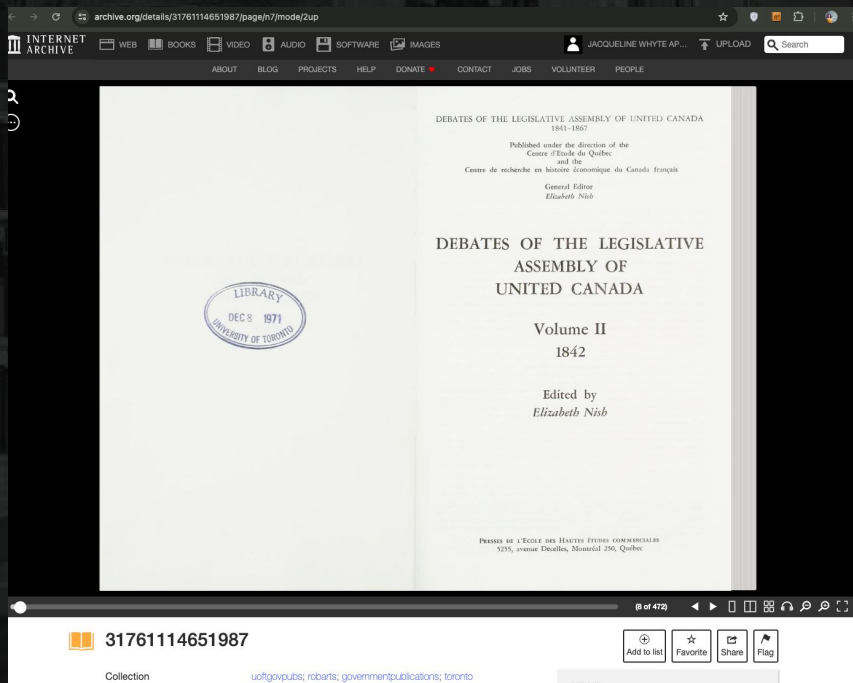
Part Of: University of Toronto - Robarts Library, Canadian Government Publications Portal, toronto texts

Media Type: texts 50,467

Availability: Always Available 7,786

Year: 1988 974, 1990 671, 1987 623, 1989 585, 1991 529, 1944 516

Project: Government documents metadata enhancement



Using RAG to
add metadata
& build links
between data
sets.

Project: ACE summaries & remediation

Preface	ix
Abbreviations	xi
Some Notes on the Notes and the Translation	xiii

Introduction

Why Read This Book (and This Introduction)?
Myth and Lit 101

TABLE DES MATIÈRES

La percée des <i>Recherches logiques</i> et la phénoménologie: une introduction par Denis Fisette.....	9
Le général et l'abstrait: sur la maturation des <i>Recherches logiques</i> de Husserl par Robert Brisart.....	23
Carnap, Husserl, Euclid et l'idée d'une géométrie matérielle critique de Twardowski.....	41
ncipe du contexte chez Bolzano et Husserl chez Bolzano.....	93
» et essence dans les <i>Recherches Logiques</i>	111
	125

Acknowledgments ix

INTRODUCTION 1

CHAPTER 1. Words as Bricks and Pages as Mortar
Building the Lettered Barriada 25

CHAPTER 2. The Workshop Is Our Homeland
Global Communities, Local Exclusions 54

CHAPTER 3. In the Margins of the Margin

Chapter 3: Comparing Sequences	51
3.1 Dissimilarity Measures to Compare Sequences	52
3.2 Alignment Techniques	53
3.2.1 <i>Optimal Matching</i>	53
3.2.2 <i>Assigning Costs to the Alignment Operations</i>	56
3.2.3 <i>Critiques of Classical OM</i>	61
3.3 Alignment-Based Extensions of OM	64
3.4 Nonalignment Techniques	73
3.5 Comparing Dissimilarity Matrices	74
3.6 Comparing Sequences of Different Length	77
3.7 Beyond the Standard Full-Sample Pairwise Sequence Comparison	78

Chapter 4: Identifying Groups in Data: Analyses Based On Dissimilarities Between Sequences	83
4.1 Clustering Sequences to Uncover Typologies	83
4.1.1 <i>The Rationale Behind Clustering Sequences</i>	86
4.1.2 <i>Crisp (or Hard) Clustering Algorithms</i>	88
4.1.3 <i>Partitional Clustering</i>	91
4.1.4 <i>Using Cluster Quality Indices to Choose the Number of Clusters</i>	
4.2 Illustrative Application	
4.2.1 <i>Hierarchical Clustering: Ward's Linkage</i>	
4.2.2 <i>Partitional Clustering: Partitioning Around Medoids</i>	

xi	LIST OF MAPS	xiii	LIST OF TABLES
xv	ACKNOWLEDGMENTS		

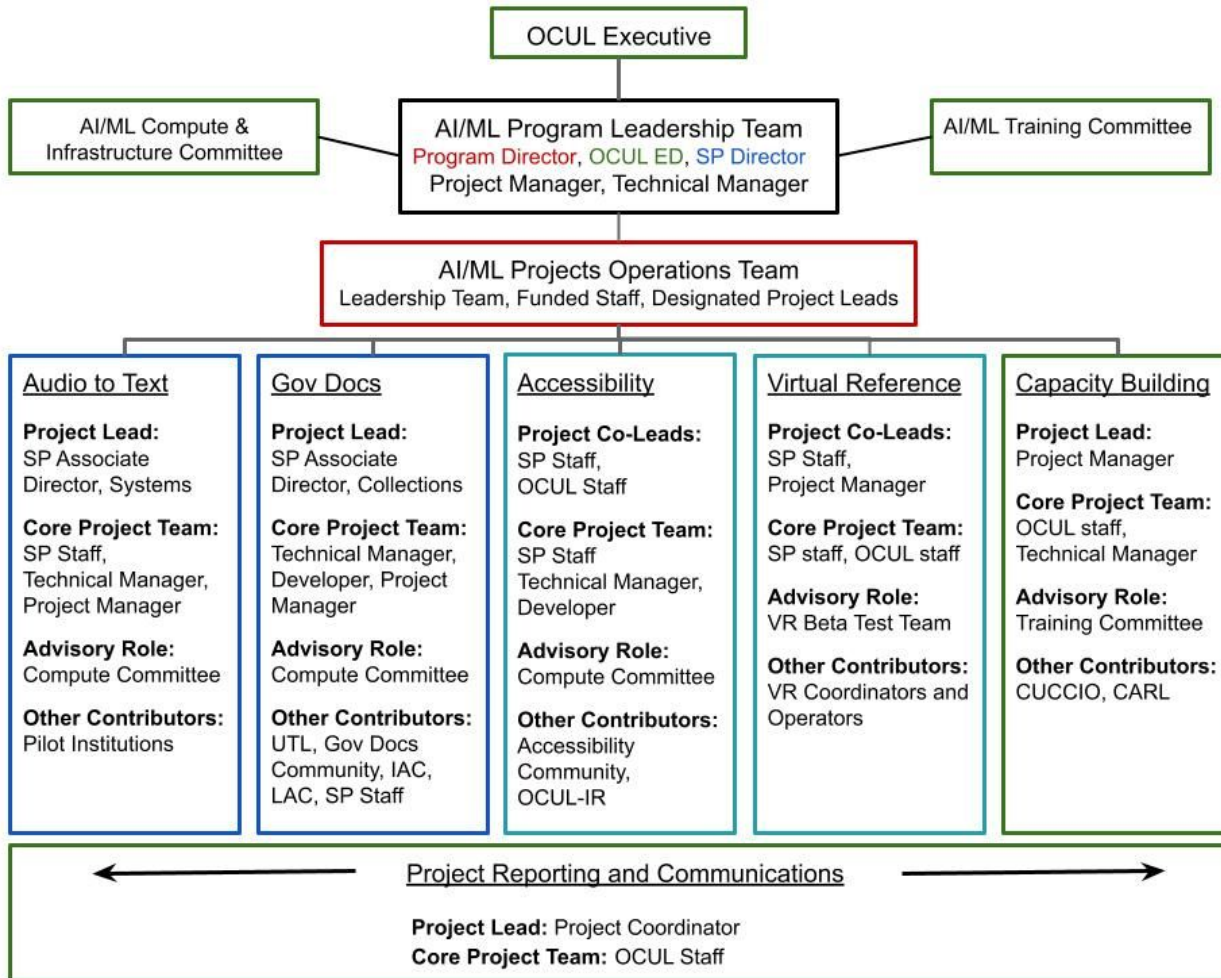
i	Introduction: Black Political Thought as Shaped in the South
19	1 Flight: Movement Matters
44	2 The Emancipation Circuit: A Road Map
80	3 Virginia: Assembly
109	4 North Carolina: Custody
133	5 South Carolina: Majority
165	6 Georgia: Mobilization
196	7 Florida: Faction

Project: Capacity building

- Technical training, enterprise strategy, policy and advocacy
- Full gamot - awareness, AI literacy, AI/ML technical skills & development expertise
- Opportunity for collaboration and partnership - CUCCIO, CARL, CRKN, Compute Ontario.....

Program Leadership Team

- Program Director, C. Steeves
- Technical Manager, P. Botha
- Project Manager, TBD
- Developer, TBD
- OCUL Executive Director, A. Greenberg
- Scholars Portal Director, K. Davis



Next Steps

- Program plan
- Recruitment
- Project charters
- Convene project(s) team
- Establishment of advisory groups
- Staggered project implementation
2024/25

Links

Interim Report:

<https://ocul.on.ca/aiml-interim-report-released>

Final Report and Strategy:

<https://ocul.on.ca/ai-machine-learning-final-report-strategy>

Job posting, Program Manager:

<https://ocul.on.ca/job-ai-machine-learning-program-manager>

For upcoming job postings and AI/ML Program updates
watch: <https://ocul.on.ca/news>

Thank you!

csteeves@uwo.ca

jacqueline@scholarsportal.info