

CARLETON UNIVERSITY'S "BETTER RECORD" PROJECT: SHARING PROGRESS AND FUTURE GOALS

Several thin, white, parallel lines of varying lengths and angles are positioned in the bottom right corner of the slide, creating a sense of movement and design.

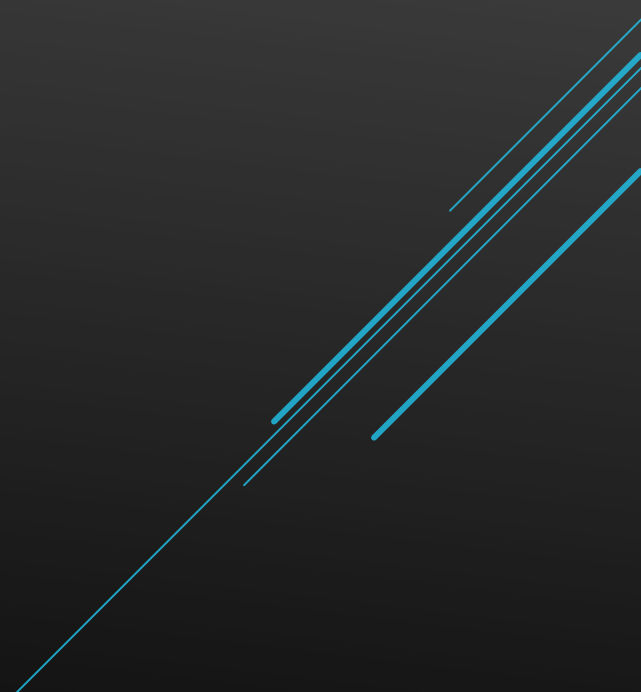
Kevin Burke

Carleton University

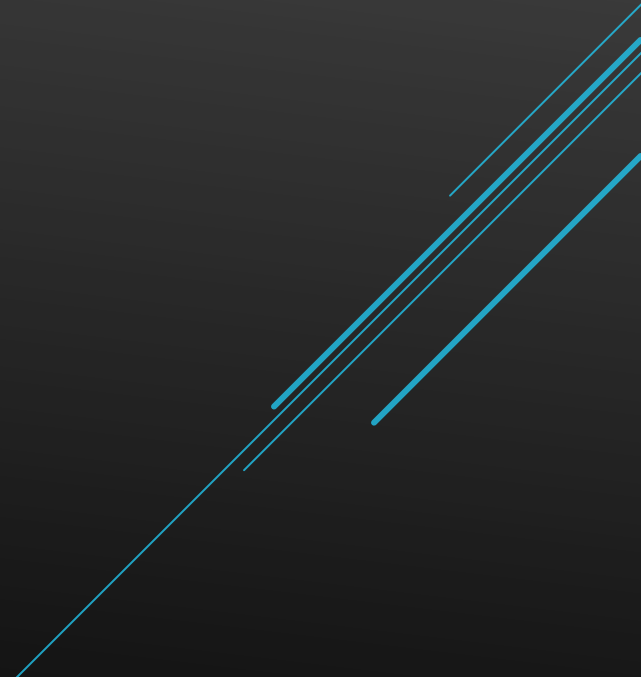
July 8, 2025

LAND ACKNOWLEDGEMENT

I am joining today from the unceded territory of the Algonquin Anishinàbeg nation, in what is currently known as Ottawa, Canada. I am a settler on this territory, with family roots in Great Britain and Ireland. I have had the privilege to live and work uninvited on this territory as well as the unceded lands of the Musqueam, Squamish, Tsleil-Waututh, and Mi'kmaq.

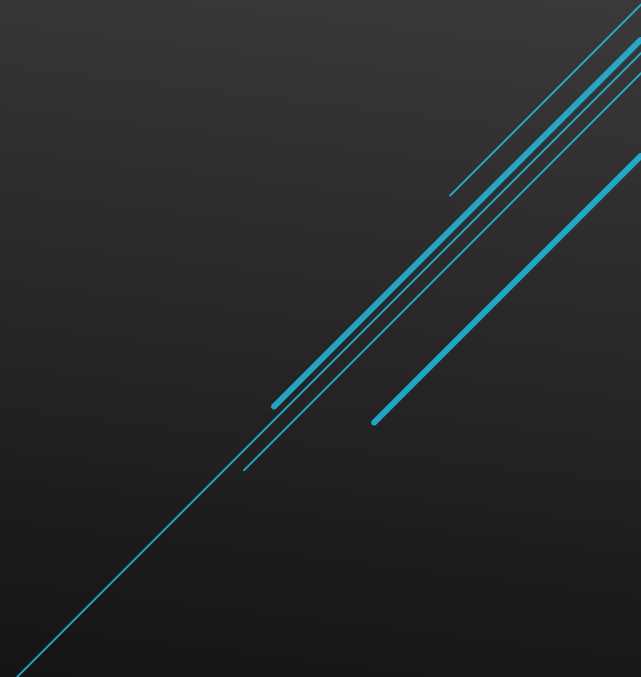


OVERVIEW

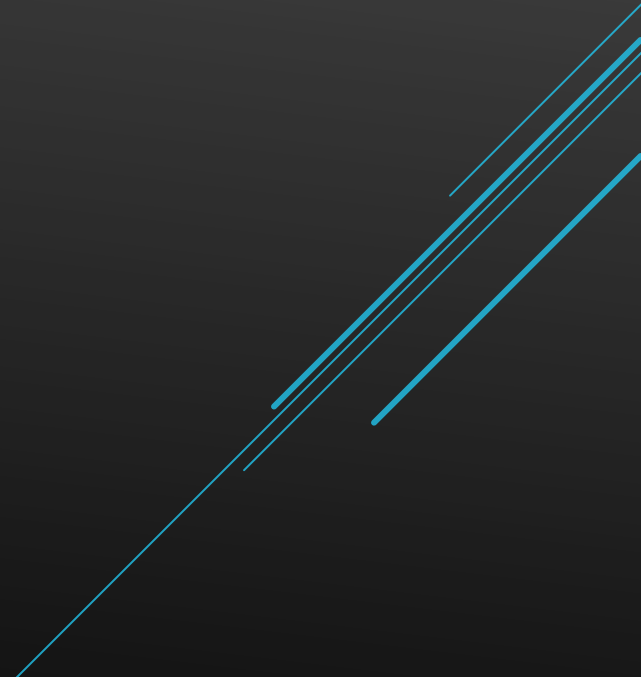
- ▶ Context
 - ▶ Phase 1 Workflow and Results
 - ▶ Phase 2 Plans and Related Work
 - ▶ Questions
- 
- A series of parallel teal lines of varying lengths and slopes, located in the bottom right corner of the slide, creating a modern, abstract graphic element.

CONTEXT

LEGACY DATA

- ▶ Many legacy records dating back to locally-developed LSP (CUBE)
 - ▶ Records built on local practice, not compatible with current standards
 - ▶ Dubbed the “690 Records” since subject headings were in 690 field
 - ▶ Held back from Network Zone (NZ) and WorldCat
 - ▶ Existing OCLC numbers moved from 035 to 095
 - ▶ Target of several manual cleanup projects 2020-2024
- 
- Several parallel teal lines of varying lengths and slopes are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

SITUATION IN 2024

- ▶ ~100K legacy records remaining
 - ▶ Manual cleanup underway, but not feasible
 - ▶ Creation of metadata analyst role created capacity for batch processing
 - ▶ Batch processing project begun in August 2024 with phase 1 completed in December 2024
- 
- Several parallel teal lines of varying lengths and slopes are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

PHASE 1: RECORDS WITH OCLC NUMBERS

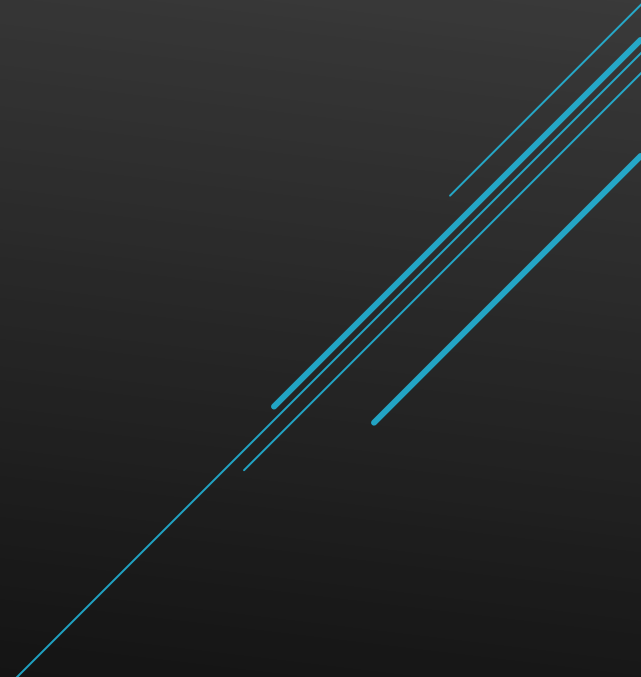
WORKFLOW OVERVIEW

- ▶ Export records and collect existing OCLC numbers
- ▶ Search WorldCat for matching records via Z39.50
- ▶ Clean WorldCat records and extract updated/improved records
- ▶ Merge Carleton and WorldCat records
- ▶ Check confidence of matches and review manually
- ▶ Load records into the Institution Zone (IZ) and share with the NZ
- ▶ Deduplicate records in the NZ

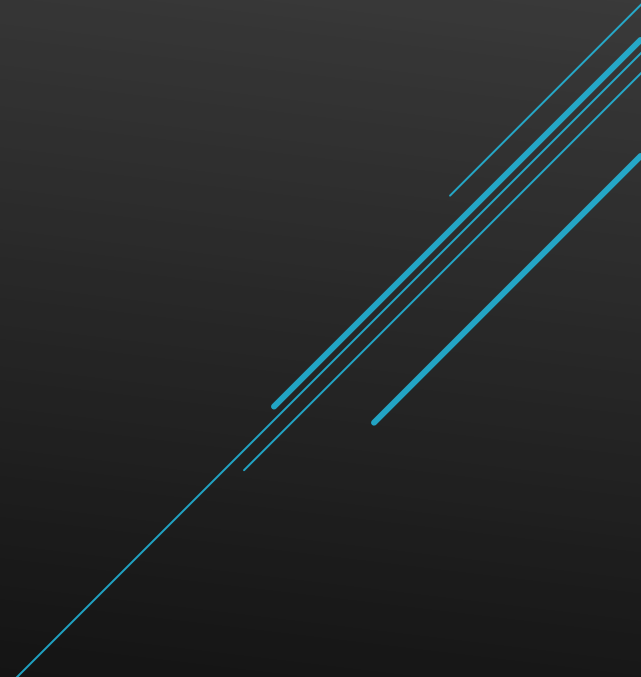
Happen in Alma

EXPORT RECORDS AND COLLECT EXISTING OCLC NUMBERS

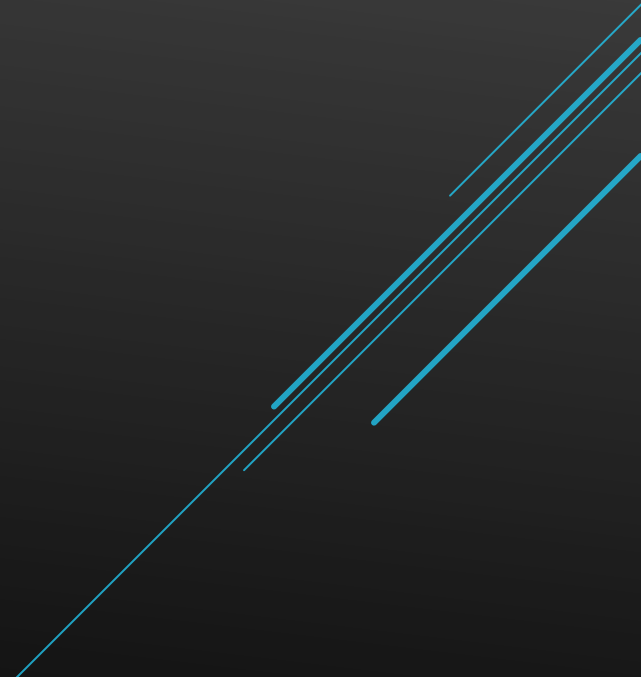
- ▶ Records cleaned in MarcEdit (removing any non-OCLC control numbers from 035) and valid OCLC numbers extracted
- ▶ Relies on existence of OCLC numbers in records for matching with WorldCat (~60K records)



SEARCH WORLDCAT FOR MATCHING RECORDS VIA Z39.50

- ▶ Using extracted OCLC numbers, WorldCat searched with MarcEdit's Z39.50/SRU utility
 - ▶ Used all OCLC numbers (including duplicates through valid/cancelled numbers)
 - ▶ Records retrieved saved
- 
- A series of parallel teal lines of varying lengths and orientations, located in the bottom right corner of the slide, creating a modern, abstract graphic element.

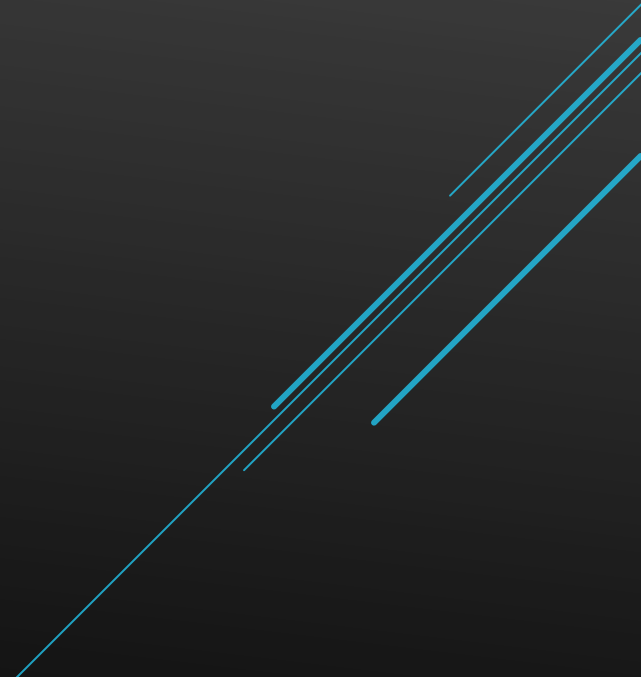
CLEAN WORLDCAT RECORDS AND EXTRACT UPDATED RECORDS

- ▶ Records from WorldCat deduped using MarcEdit's deduplication tool
 - ▶ Subject headings cleaned up to match local practice
 - ▶ Followed CF Acceptable Subject Thesauri policy from 2019
 - ▶ Kept LCSH, LCGFT, RVM, FAST, MeSH, etc.
 - ▶ Extracted all the records with valid 6XX fields
 - ▶ These records were improved either through OCLC's merging process or other institutions cleaning them up
 - ▶ Records without valid headings added were discarded
- 
- Several parallel teal lines of varying lengths and orientations are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

MERGE CARLETON AND WORLDCAT RECORDS

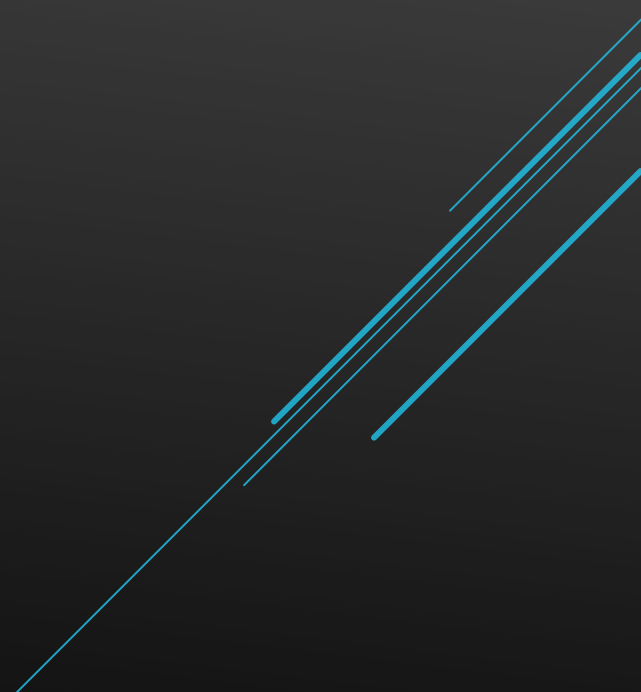
- ▶ Used MarcEdit merge tool to copy Carleton University local fields over to WorldCat records
 - ▶ Matched on OCLC #
 - ▶ Retained Carleton's 001 for matching when loading into Alma
- ▶ Extracted records with improvements (not all records matched)
- ▶ Did final (basic) cleaning of improved records in MarcEdit
 - ▶ Fixed indicators where \ had been swapped for 0 and vice versa
 - ▶ Sorted fields, deleted blank lines, etc.

CHECK CONFIDENCE OF MATCHES AND REVIEW MANUALLY

- ▶ All new records had some improvements, but not all were good to use
 - ▶ Some had data errors like missing 300 fields
 - ▶ Some had been changed to refer to different resources
 - ▶ Some mis-matches (likely due to truncated OCLC numbers in Carleton records)
 - ▶ Used a Python script to compare updated records with matching original Carleton record across key fields (1XX, 245, 250, 260/264, 300) and return a similarity score (0-100) for each pair
 - ▶ Low-confidence pairs (< 67% similarity) compared manually using another Python script to display records side-by-side
- 
- Several parallel teal lines of varying lengths and orientations are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

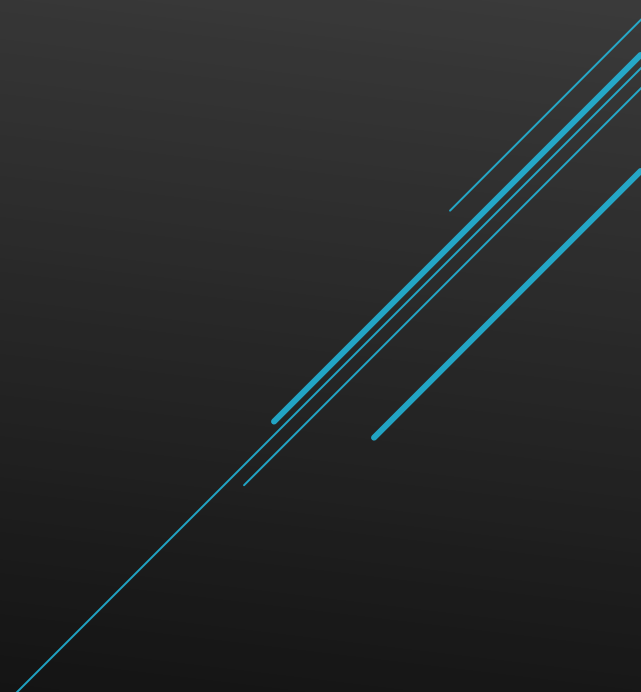
LOAD RECORDS INTO THE IZ AND SHARE WITH THE NZ

- ▶ Alma automatically merges records when sharing into the NZ, using existing NZ version as the preferred record
 - ▶ Incoming records generally more current than versions already in the NZ
- ▶ Needed to load records into the IZ first
 - ▶ Easy to ensure accurate matches, since we preserved Carleton's record IDs (001)
- ▶ Then shared into NZ using an unconfigured match profile to skip all matching
 - ▶ Created duplicate records with any previous versions in the NZ

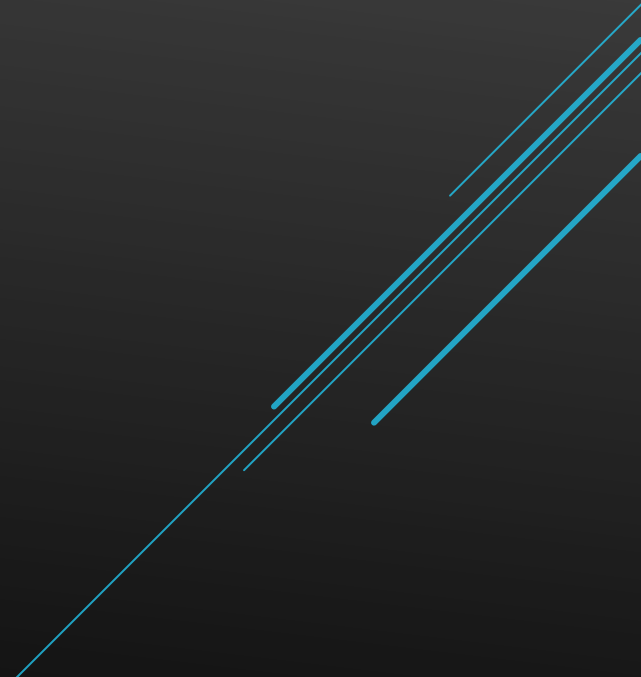


DEDUPLICATE RECORDS IN THE NZ

- ▶ Merged duplicate records in the NZ (identified by OCLC numbers) with updated Carleton records as the preferred version
- ▶ Required a similar method for checking the confidence of the matches
 - ▶ Some pre-existing records ambiguous about key details
 - ▶ Some pre-existing records with wrong OCLC numbers
- ▶ Low-confidence matches reviewed manually and referred to relevant institutions for correction

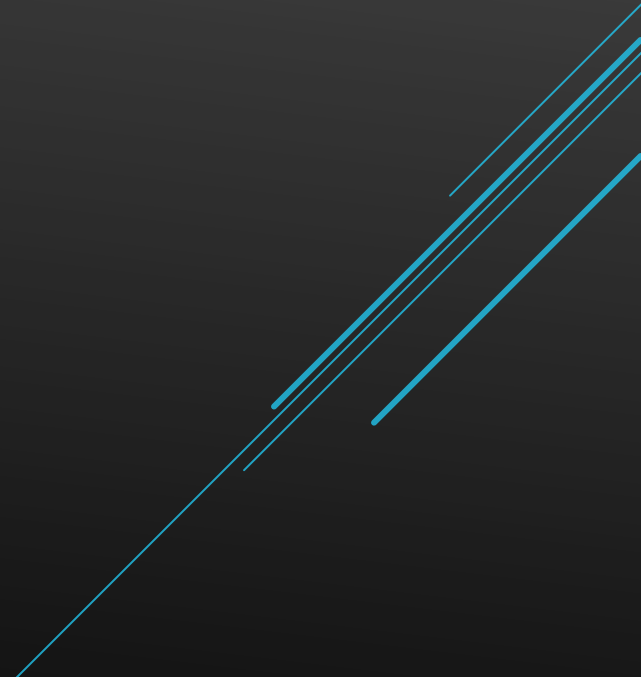


PHASE 1 RESULTS

- ▶ >48K records loaded into IZ and shared into NZ
 - ▶ ~24K records completely new to NZ, same amount pre-existing but merged and updated
 - ▶ >300K subject headings added to our catalogue
 - ▶ ~50K 690 records left to update...
- 
- Several parallel teal lines of varying lengths and slopes are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

PHASE 2 (ONGOING) AND RELATED PROJECTS

RECORDS WITHOUT OCLC NUMBERS

- ▶ ~50K “690 Records” left to update
 - ▶ Need to find other match points for searching WorldCat
 - ▶ Can use LCCN, but only present in ~20K records
 - ▶ Exploring other options for searching WorldCat
 - ▶ Extracting LCCN, ISBN, author, title, publisher, and/or publication date from records to build queries
 - ▶ Preliminary results very mixed...
- 
- Several parallel teal lines of varying lengths and slopes are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

RELATED PROJECTS

- ▶ Cleanup of other legacy records
 - ▶ Linking records to Network Zone without losing data
 - ▶ Implementing WorldCat Updates service for importing updated records
 - ▶ Running Streamlined Holdings Update to synchronize holdings with WorldCat
- 
- A series of parallel teal lines of varying lengths and orientations, located in the bottom right corner of the slide, creating a modern, abstract graphic element.



ACKNOWLEDGEMENTS

Many thanks to:

- ▶ Emma Cross, Cataloguing and Metadata Librarian at Carleton University
- ▶ Angela Hayward, Cataloguer/E-book Specialist at Carleton University
 - ▶ For sharing institutional background on 690 records!
- ▶ Jennifer Browning, Head of Cataloguing, Metadata, and Digitization at Carleton University
- ▶ Erin Bourgard, Acting Director of FIMS Graduate Library, University of Western Ontario
- ▶ Alex Fletcher, Omni Technology and Network Zone Specialist, OCUL
 - ▶ For helping develop the workflow, particularly the NZ portion!

THANK YOU!



Kevin.Burke@carleton.ca



github.com/kevintburke