A Field Manual for Cultural Heritage Informatics

The X-Lab, Carleton University

2022-02-15

Contents

Chapter 1

Welcome

The XLab – The Cultural Heritage Informatics Collaboratory – represents both a space and a series of relationships within and without the University. It is a kind of transdisciplinary 'skunkworks' for fostering encounters with and between cultural heritage and digital media and computation. As a skunkworks, or a space for trying non-traditional or imaginative new solutions, it aims to bring together tenacious tinkerers, who are willing to experiment, to wonder, to try, and to iterate – to try, try again. These creative engagements sometimes will not fit within existing disciplinary modes of thought or administrative structures. We accept and celebrate this – the XLab values process and relationships, grounded in lived situations.

The digital era is, in some ways, an era that has a renewed focus on orality, on the transmission of knowledge through personal relationships. When we think of the 'field' for cultural heritage informatics, we acknowledge that we are dealing with 'belongings', not objects; that cultural heritage is imbued with meanings, and that knowledge holders are everyone from children to Elders, academics to artists, and that knowledge and knowledge holders are found everywhere: there are no simple binaries. Thus we see cultural heritage informatics embedded in networks of relationships, in flows of knowledge and ideas. The work of the XLab is to promote and understand these flows and relationships. The fieldwork of cultural heritage informatics can be situated in understanding the metadata of these flows.

The idea of 'cultural heritage informatics' can be understood as actions in the context of relationships. Such relationships have to be fostered and built on mutual trust. These actions involve thinking through new protocols for how we work with differing communities, respecting the knowledge and sovereignty of the communities, with authentic reciprocal community engagement. It will require the creation of differentiated pathways of access that respect that not all cultural heritage knowledge is meant for everyone.

The actions of cultural heritage informatics might run the gamut from digitization and ontologies and description, to the devising of protocols and platforms for the sharing of cultural materials, to performance and storytelling cultural heritage through immersive technologies. The actions of cultural heritage informatics might be in creating the necessary metadata to link repositories of knowledge together to generate new knowledge, new relationships. It might mean the hard work of curation and restoration to decolonize collections of cultural heritage dispersed across Western museums. The actions of cultural heritage informatics, whatever they may be, are grounded in our relationships with the cultures and communities with whom we work.

Chapter 2

Audience

- discussion here of who we are writing for; our imagined audience and their level of digital literacy
- that is to say: a small scale, probably volunteer run organization with minimal access to computing resources; that is, a couple of personal computers, some smartphones, ideally a high-speed internet connection, and while webspace-of-one's-own would be good, we'll use open resources (vercel, heroku, colab, figshare, zenodo, binder, gh-pages, etc) to get things up and running

2.1 How to use this book

- how code blocks work etc

2.2 Setting Up Your Machine

- since we're imagining small scale organizations, we're also imagining the use of common low-cost PCs
- this will align our goals/approaches with dh mincomp type stuff
- but sometimes we'll recommend things like Binder or Reclaim Hosting, especially since Tim Sherratt has figured out how to deploy his glam workbench to reclaim cloud. The problem there is to show people how to make THAT work for their own data
- installation of anaconda/miniconda; installation of R and R Studio
- command line basics

2.3 Basic Principles

- future proofing

- minimal computing
- Indigenous Data Sovereignty & Protocols
- Open Access and Sensible Limitations
- personal and community safety

Chapter 3

Digitization

Digitization does not equal 'preservation'. But it can mean that your materials find a broader audience. It can mean that you have a better sense of what materials you're responsible for, the gaps in your materials, and give you an indication of what areas you should be putting your energies into. Digitization can mean that your materials become available for research. It might even mean that your collection can be linked into other collections, so that a better, truer, picture emerges. But first things first: how does a small organization get its materials into a digital format?

- basic low cost digitization
- data repositories available - zenodo, figshare, others - and why a small org should use them
- a workflow: digitization - metadata creation - repositories - public facing - internal/external research

3.1 Photo Management

- digitization
- Tropy for research photo management
- cross reference to collectionsbuilder for putting stuff online/making it findable
- IIF standards stuff

3.2 Image-to-Text

- flatbed or phone
- treat as image, or treat as text?
- tabula, other OCR options
  - OCR in bulk using R & Tesseract

3.3 3d Photogrammetry

- assuming conventional cameras or phones
- meshroom - via app, or via commandline/google collab
- hosting such models
- metadata & London Charter
- photogrammetry of objects
- photogrammetry of spaces

Chapter 4

A Gentle Introduction to Digital Literacy

- digital literacy and data literacy are not the same thing
- not going to 'teach' python or R, but rather show working blocks of code and how to reuse/repurpose for immediate tasks
- will link out to appropriate other tutorials as necessary
- the basic goal is always minimal amount to get things going

4.1 Introduction to Python

- things like creating new environments for particular tasks
  - packages and where to find them
  -

4.2 Introduction to R & R Studio

- things like creating new projects for particular tasks
  - packages and where to find them
  -

4.3 Version Control

- go with gitlab maybe rather than github?

Chapter 5

Tutorials

5.1 Tables to Databases

- arranging data in tables
- arranging data in graphs
- setting up a basic database of either kind
  - sqlite, for local stuff? people have VIEWS on that
  - graph basics are probably too much, for the intended audience, but maybe there's something gentle out there
- data management workflows

5.2 Cleaning Up Messy Data with OpenRefine

5.3 Publishing Data with Datasette

- putting data online with Datasette
- useful plugins for Datasette
- using the desktop version of Datasette
- putting Datasette online but behind a password

5.4 Working With Oral Interviews

- automatic transcription with Mozilla DeepSpeech

5.5 Working with Other Data Sources

- accessing databases using Python, R
- scraping various sites

5.6 Working with Images

- PixPlot for exploring a collection

Chapter 6

Public Facing Work

6.1 Static Websites With Hugo

- your organization needs a basic, fast, secure website that is aesthetically pleasing

6.2 Building a Collections Website with CollectionBuilder

- you have wonderful materials. Showcase them.

6.3 Omeka for Exhibitions

- Tell stories about your materials, pulling them together from your collection's materials

6.4 Visual Storytelling with Mural

Chapter 7

Templates & Other Useful Materials

On this page will be links to templates for wrangling metadata, etc.

- Long-form storytelling

7.1 Exhibits with Wax

- why and when you might use Wax

7.2 Mukurtu

- Differing communities have differing protocols about access to cultural heritage information. Mukurtu is built with this in mind.

7.3 Bots

- Some simple bots that can make your materials more accessible

7.4 Interactive (Non-)Fiction

- Tell interactive stories using Twine or Ink or Ren'Py

Chapter 8

Final Words

We have finished a nice book.