Digital Humanities Application Development in the Cloud

John Mulligan, Ph.D.
Rice University (Houston)
Center for Research Computing (CRC)
jcm10@rice.edu
Working in the Cloud

The virtualization of IT labor
The availability of scalable resources promises:

• Scalability through networking
  • Not requests for but self-provisioning of resources
• Fungibility through virtualization
  • Raw resources can be repurposed to any specific end
An ideology critique of the cloud

• The acceleration of research
• Resources on demand
• The fetishization of computing:
  • Not automation
  • But the invisibilization of labor
Automation *and* fetishization (a double-edged sword)

[T]he technologies and techniques of IT are not necessarily fused to the doxa of postindustrialism (restructuring, lean production, and all the rest of the new management dicta). As a mode of development, IT also generates what amounts to a semi-autonomous doxa (a *belief* in information or in technology), akin to the faith that a rural laborer in revolutionary France, about to burn the house of the nearest aristocrat, might have expressed by picking up a pitchfork and saying, “*This is what I believe.*”

Alan Liu, *Laws of Cool*, p. 40.
Use-Case

Developing Custom Humanities Software as a Service in the Cloud
Zotero lock-in

Your personal research assistant

Zotero is a free, easy-to-use tool to help you collect, organize, cite, and share research.

Download

Available for Mac, Windows, and Linux

New: Just need to create a quick bibliography? Try ZoteroBib.
Tropy lock-in

https://docs.tropy.org/using-tropy/add_files

https://forums.tropy.org/t/exporting-as-csv-file/365/3
Google Drive custom metadata (future)

https://support.google.com/a/answer/9292382?hl=en
Local vs. Networked Image Archiving Apps

Locally-hosted image interface & background scripts

- Custom database
- Image store
- Plugins, e.g., OMEKA export

Scalable CentOS VM(s):
- Custom database
- IIIF Loris image server
- Python Flask server
- Background scripts

Scalable networked image store

“Scalable” plugins, e.g.,
- HTC for OCR
- Google Drive Sync

Web interface
Prototype interface screenshots
Advantages of assuming scalable networked storage:

• No hard limit on archive size
• Backups/migration easier to automate and scale (multiple locations/platforms)
• Batch operations scalable (affordances for supercomputing)
Reflection on Methods

From jump-starting to incubating
Jump-starting vs. incubating

• Jump-start
  • Upsides:
    • Knowledge-sharing
    • PI experience
    • Seed funding
  • Downsides
    • Redundancy in labor (too many cooks)
    • Learning curve

• Incubation
  • Upsides:
    • Division of labor & domain expertise
    • Rapid prototyping
    • Institutional embedding
  • Downsides
    • Redundancy in software (reinventing the wheel)
    • Rapid obsolescence
Digital Humanities Application Development in the Cloud

John Mulligan, Ph.D.
Rice University (Houston)
Center for Research Computing (CRC)
jc10@rice.edu