



Running Hydra in a small shop

Three case studies

Richard Green, University of Hull
Adam Wead, Penn State University

with help from

Chris Awre, University of Hull
Simon Lamb, University of Hull
Steven Ng, Temple University

What's the problem?



- True or false?
- Running and maintaining a Fedora-based digital repository is a daunting task
- It necessarily requires a significant team of developers and support staff

What's the answer?



- False.
- This presentation will look at the experiences of three different organisations which have used the Hydra digital repository solution to create Fedora-based repositories
- Although apparently very different in size and budget, these three organisations have in common that, with only a single software developer, they have developed and/or are developing a Fedora-based repository with all the benefits that brings

The three case studies



- The Chinese Historical Society of Southern California
 - Small non-profit organisation with three staff



- The Rock and Roll Hall of Fame and Museum, Cleveland, Ohio
 - Medium non-profit organisation with approximately 120 staff



- The University of Hull, UK
 - Medium sized UK university. Approximately 20,000 students and 2,500 staff (including 1000 academics) across two campuses

How was Hydra “sold” to the institution?



- Steven: “It was an easy sell because using an open source solution we would not have to pay maintenance or license fees. We could focus on obtaining grants to cover servers and related hardware. In addition, I presented a quick demo of a customized Curate server themed for our organization and seeded it with images from a recent CHSSC event. Finally, there are few other historical societies in Southern California who were addressing digital preservation. This would give us the prestige of being a pioneer in the community.”



- Adam: “It wasn’t a hard sell because none of my colleagues understood what I did, so I could sell them a potato and tell them it would solve all RockHall’s problems. However, that did put a lot of pressure on me to find something that worked. In the process of looking at other solutions, none of them covered all our needs and the ones that came close were all very expensive. So in that way, it was an easy sell.”



Hull is a founder member of the Hydra Project and in that sense it was a “no-brainer”. That said, we’d repeat the Hydra mantra that it is “a collective solution with breadth and depth that exceeds the capacity of any individual institution to create, maintain or enhance on its own.” We knew what we wanted, from previous experience with Fedora behind a different front end; Hydra delivers that and, now, much more.

How many developers?



- One. (Steven) A board member/volunteer developer. Installing digital repository software.



- One. (Adam) Employed by the Library and Archives. Responsible for all systems, tech support, and library technology

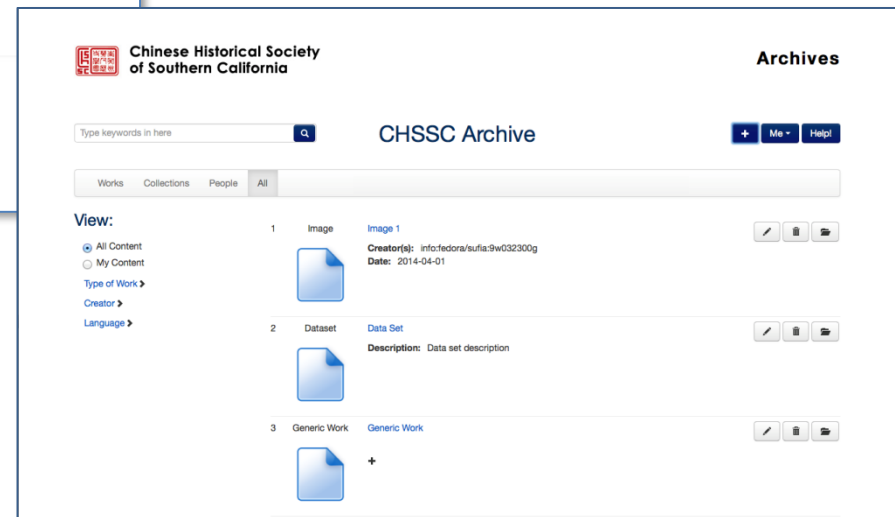
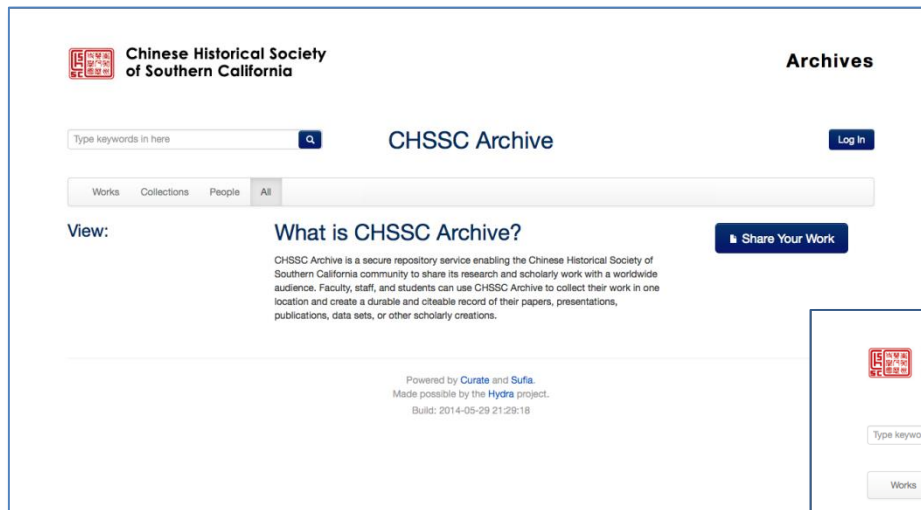


- One. (Simon) Employed by central IT - probably 0.8 of time on Hydra. Occasional support with UI (etc) work from a colleague

What does the Hydra “head” do?



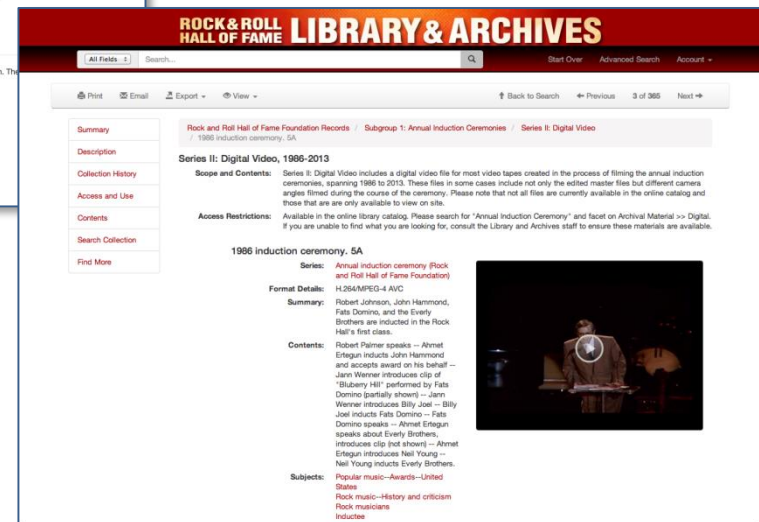
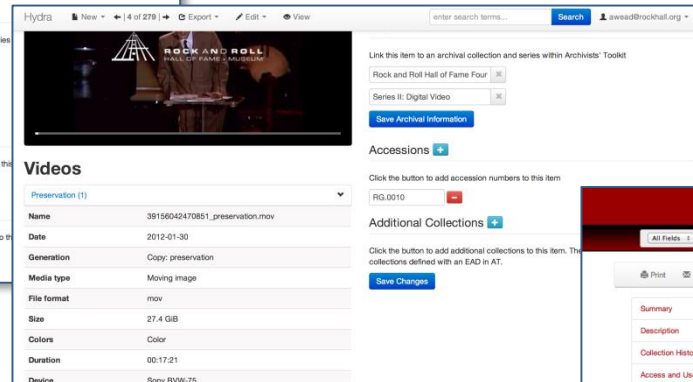
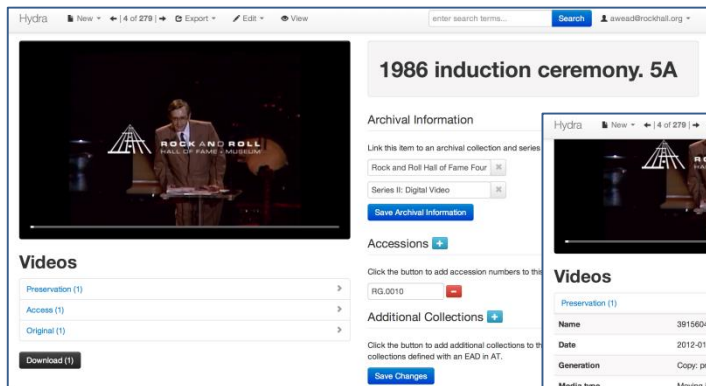
- CHSSC’s Digital Repository is for all of its digital content including images, audio recordings of oral histories, videos, and datasets.



What does the Hydra “head” do?



- Privately accessible application for ingesting, describing, and preserving video. Metadata is exported for inclusion within our archival finding aids.



What does the Hydra “head” do?



Hydra provides the University's Digital Repository - potential home to all digital content including theses, journal articles, minutes, datasets, images, multimedia etc. Full workflow from create, through QA, to search/discover/download. Multi-layered security.

The screenshots illustrate the Hydra Connect workflow:

- Create Resource:** A form for uploading digital content, including fields for Author, Resource, and Rights.
- Results:** A search results page displaying a list of research papers, such as "Proteomic identification of putative biomarkers of neoadjuvant chemotherapy resistance in luminal (ER+) breast cancer".
- Assets:** A detailed view of a specific dataset, including a map and download links for various formats (e.g., PDF, CSV, XLS).

But Hydra uses Ruby? I don't.



- No prior experience with digital repository software but Steven had experience with small scale Ruby on Rails development.



- Adam had no prior experience with Ruby on Rails. The desire to use Fedora as the repository backend drove the RockHall to look for suitable front-end solutions, leading them to Hydra, and Ruby came with it!



- Prior to Hydra, Hull had no Ruby expertise. A commercial company was used to kick start work on the Hydra repository and part of the package was to train the in-house developer. Simon rapidly learned Ruby which is now his preferred development environment.

How did the Ruby expertise develop?



- Largely self-taught but with a Computer Science major as part of a degree (10 years before Ruby!)... Steven has extensive software development experience in aerospace and in commercial handheld software



- Adam learnt on his own! He had Perl, PHP and some Java prior, so “picking up a new language was easy”



- As noted on the previous slide, on the job training, but also books, screencasts, and looking at others' code

How much “local” code did you write?



- There are no gems specific to CHSSC beyond the core gems that would be required to deploy a vanilla Curate/Hydramata server. We have customized CSS to match the existing chssc.org style and color scheme.



- Gems specific to the RockHall installation were related to EAD (Encoded Archival Description) in Solr, and PBCore metadata schemas. Beyond that, a bit of workflow code for ingestion, and UI work for user experience.



- In addition to the core Hydra gems (specified in *the* Hydra gem) Hull has an additional, locally written one which provides all the customisation for local workflow, different display formats for different content etc.

Isn't Hydra's approach of "test-driven development" hard?



- Steven had some experience with TDD in working with other Ruby on Rails shops in the Los Angeles area.



- Adam said: Steep learning curve, but well worth it. Having test coverage ensures you're not going crazy, which is paramount in a one-person shop.



- Simon took a while to get his head round test-driven development but now finds that it saves time in the long run.

What about support?



- The community of Hydra developers was, and still is, an essential element in the process



- The community of Hydra developers was, and still is, an essential element in the process



- The community of Hydra developers was, and still is, an essential element in the process

What about documentation?



- Steven initially started with the “Dive Deep into Hydra” but found it only took him so far. When he installed Curate last September, documentation was more specific and he was able to get a Curate server up and running. *(Curate is an early “solution bundle” and is the basis for the upcoming Hydramata)*



- Adam started with the earliest release of Hydra, which had almost no documentation. Lots of head banging, but it’s improved greatly, mostly because the project has better articulated goals. Of course, it still needs work.



- When Simon started, Hydra’s documentation was not wonderful and a lot of information needed to be gleaned via the mailing lists and/or the IRC channel. It is much better now than it was. (Thanks in no small part to considerable community input.)

Deployment



- Steven is developing a Chef cookbook to aid his deployment.



- Adam says deployment is not simple, mostly because he did not invest time into things like Capistrano, or Puppet, or any other tools to aid in this effort.



- Simon says, now that it is more mature, the Hydra stack is reasonably straightforward to deploy. Hull manually deploys at present, and would benefit from the growing trend to utilise tools like Capistrano and Chef for deployment and server commissioning.

Is it all worth it?



- Still in development, but CHSSC believe it will be worth it. One volunteer has the risk of a single point of failure. Steven has just moved onto a new job and can only put in a little time to continue the work. It's difficult to find other volunteer Ruby developers local to CHSSC who share the vision for digital preservation to continue the work.



- Yes: everything RockHall needed, at a cost they could afford. Very high staff cost, but nothing in life is truly free. Adam has recently taken a post elsewhere; it appears as though his successor will not be a developer, and instead will concentrate on systems and digitization projects. Development will be done ad hoc, on a project-based timeline. RockHall are still intent on using Hydra; however, Adam fears they underestimate the benefits of a full-time, dedicated developer.



- Yes: highly customised product matching local needs closely but with the support of a very knowledgeable virtual community to provide advice and development. Straightforward to evolve as needs evolve. A one man shop is potentially a SPOF which, from the institution's point of view, could be bad news - see above. (We've chained Simon to his desk...)



projecthydra.org

Richard Green, University of Hull (r.green@hull.ac.uk)

Chris Awre, University of Hull (c.awre@hull.ac.uk)

Simon Lamb, University of Hull (s.lamb@hull.ac.uk)

Hydra implementation: hydra.hull.ac.uk

Steven Ng, Temple University (steven@nomadicoder.com)

Hydra implementation not accessible to public

Adam Wead, Penn State University (amsterdamos@gmail.com)

Hydra implementation not accessible to public