



@



**A repository for Audio and Video**

**Caroline Webb**, Repository Developer  
The Library, Glasgow Caledonian University

# Outline

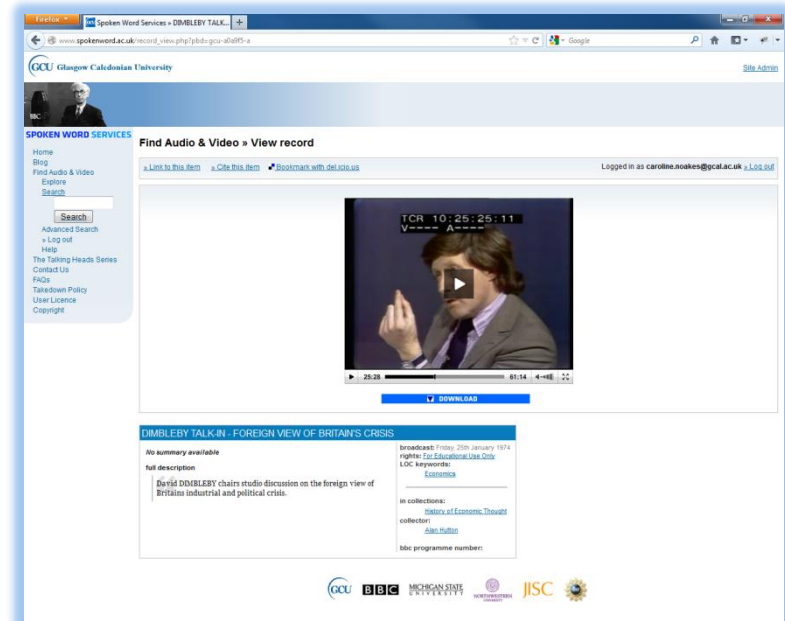
- A bit of History
- Why we chose Hydra
- Where are we now?
- Techy stuff
- Living with Hydra
- Future plans

# Our Repository: Spoken Word Services

- JISC/NSF funded project 2003-2008
  - Digital Libraries in the Classroom
  - Collaboration between GCU, Michigan State, Northwestern and the BBC
- Agreement with the BBC allows us to make audio/video from their archives available for educational use.
- Online since 2006
- 240 Hours of video, 530 Hours of audio
- Service now maintained by Digital Development Team with the Library

# Spoken Word Services

- REPOS repository software developed by Michigan State University
- User interface developed in-house (php)



[www.spokenword.ac.uk](http://www.spokenword.ac.uk)

# It works, but....

- Limited documentation for REPOS software => difficult to manage and develop
- REPOS software incompatible with newer versions of PHP => prevents server upgrades and poses maintenance problems
- Poor Data Model
  - No support for relationships between objects
  - Non-standard metadata
- Not compatible with web services – no APIs

# Fedora to the rescue

- Highly flexible, scalable and configurable
- Any type and format of content can be stored, metadata and relationships stored as XML
- All management and access tasks possible through web service API interfaces
- Open source
- Good user community

.....but what about a user interface?

# Here comes Hydra...



- Flexible framework – easy to build something that's right for you and modify when your needs change
- Very active community – continuous development of the framework components
  - UK user-base (though admittedly small at the moment)
- Ruby + Rails = rapid development, TDD = increased robustness.
- **The best bit – Already some good Hydra-Heads out there**

# Here comes Hydra...



## Hydra@Hull

- User-friendly Search/Browse ✓
- Access to content restricted by role ✓
- Integrated content management ✓
- Handle Audio/Video content ✗

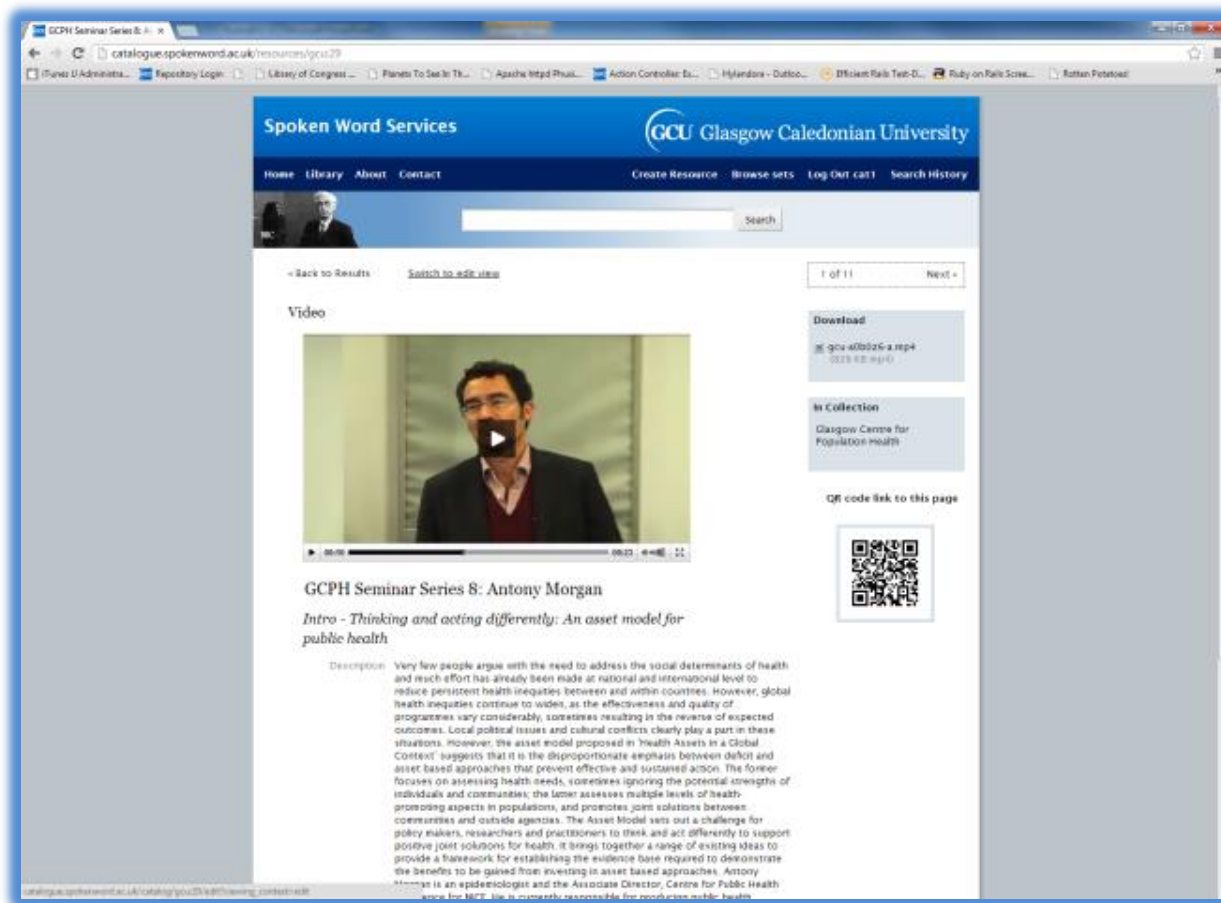


there

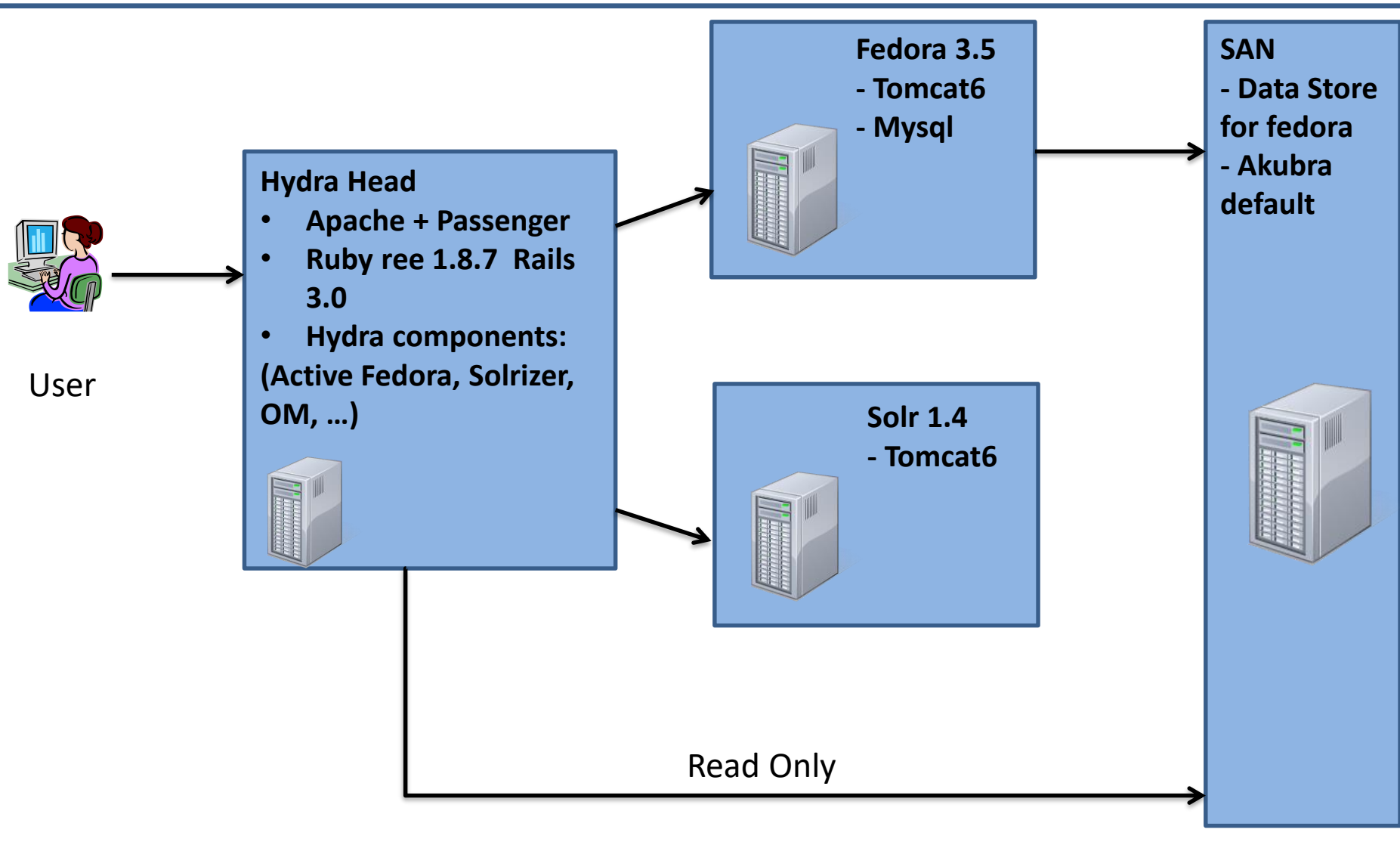
# Where are we now?

- Started Hydra development Jan 2012
- Added in Audio/Video capabilities
  - Inline playing + download
  - Uploading audio/video content +associated metadata
  - Some automated metadata extraction
- Customised display
- Demo server up and running

# catalogue.spokenword.ac.uk



# Techy stuff: Architecture



# Techy stuff - Serving Audio/Video

- Use Progressive Download/Pseudo-Streaming
  - Historically had problems with streaming (browser compatibility, network issues)
- Files stored as Managed Content in Fedora
- Fedora does not support HTTP range requests
  - Prevents seeking through a file that is not fully downloaded
- Let Hydra-head access Fedora file store directly (read only)
- Hydra-head checks Authorization
- Uses X-SENDFILE to send media to browser

# Techy stuff – Media player

- We use JW Player
  - Good cross-browser/device support
  - Auto-detects HTML5 or Flash as needed
  - Free for non-commercial use
  - Support for MP4, FLV, WebM, AAC, MP3, Vorbis...
- We use MP4 (m4v and m4a)
- Additionally offer MP3 for audio download

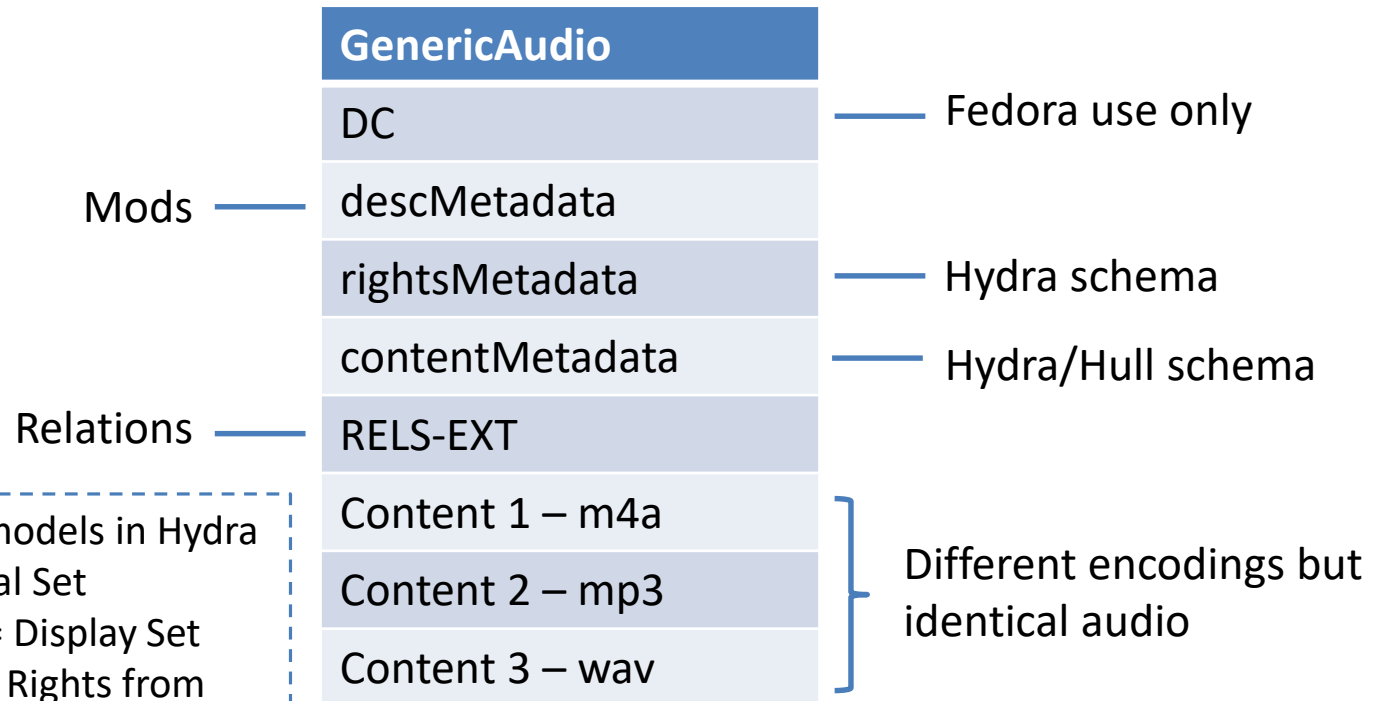
# Techy stuff: Metadata Extraction

- Automatically extract metadata from media files on upload
  - Frame Size
  - Duration
  - Codecs
- RVideo ruby gem used to hook into Ffmpeg
  - FFmpeg free audio/video conversion tool
- Hope to extend so will auto transcode to mp4/mp3 on upload
  - Will probably need to be asynchronous,- JMS? Apache Camel?

RVideo: <http://rvideo.rubyforge.org/>

FFmpeg: <http://ffmpeg.org/>

# Techy Stuff– Compound Content Models



GenericVideo model similar

# Living with Hydra



**“It took a while to get to know Hydra but now we’re best friends.”**



Steep learning curve – especially if new to the whole stack.

Documentation can be patchy

Found Ruby/Rails training essential

Can sometimes feel hard to keep up.



Flexible nature of framework  
Easy to adapt to new content types as needed

Interactive tutorial, documentation improving  
Active mailing list always happy to answer questions

Allows fast development  
Test Driven Development ensures robust code

Very active development of Hydra components

Hydra users share their code freely (Github) – no need to reinvent the wheel

# What next?

- Integration with University authentication system - LDAP
- Models/Views for BBC objects
- Automated transcoding
- More content types (theses, images...)
  - Split into 'GCUStore' with separate search/view interface for Spoken Word Services

# Contacts and Websites

[caroline.webb@gcu.ac.uk](mailto:caroline.webb@gcu.ac.uk)

Spoken Word Services – [www.spokenword.ac.uk](http://www.spokenword.ac.uk)

Hydra Head – [catalogue.spokenword.ac.uk](http://catalogue.spokenword.ac.uk)

Github - [github.com/SpokenWordServices/Hydra-GCUStore](https://github.com/SpokenWordServices/Hydra-GCUStore)