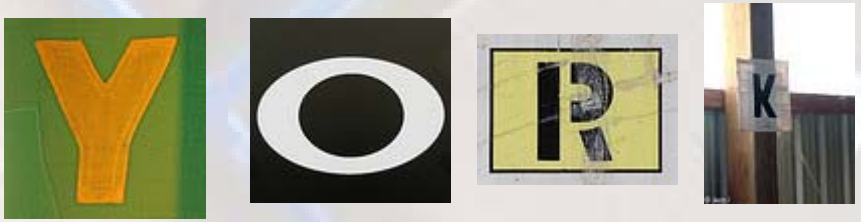
Towards a …

for …



Julie Allinson

E-Science Collaborative Workshop, Sheffield, 13 November 2008

Images from flickr,
credits available on request

THE UNIVERSITY of York

Digital Library Project

- Project started in August 2007
- runs for 3yrs
- at that point it becomes a service
- JISC provided funding for the first 17 months, for the SAFIR project (Sound, Archive, Film, Image Repository)
- we have committed to having a system in place, with pilot content (from History of Art) and metadata at the end December 2008

THE UNIVERSITY of York

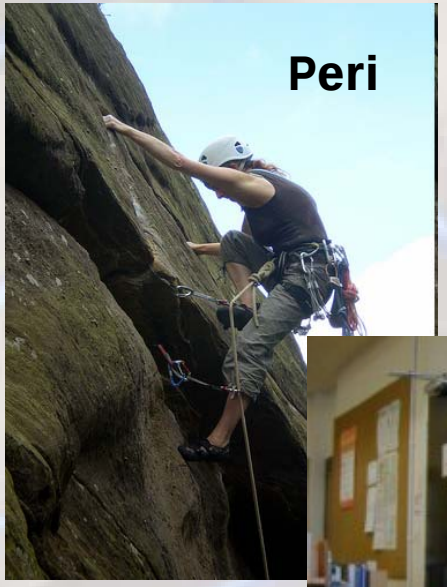
What are we doing?

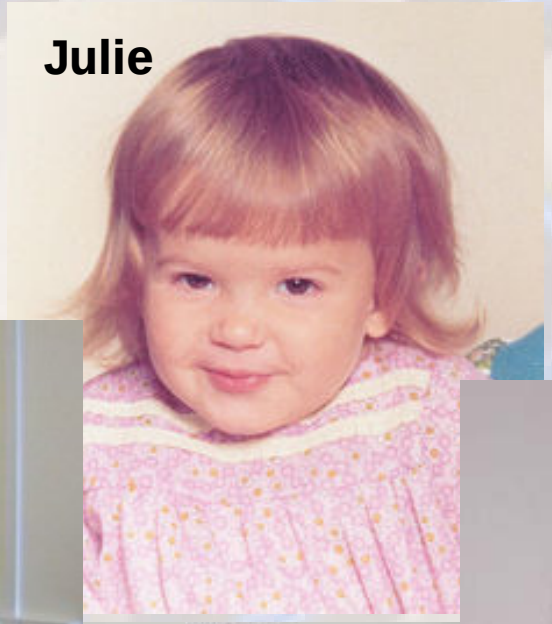
We're building a University-wide Digital Library service for **multimedia research** resources. An infrastructure, not a complete solution.

Step-by-step:

- user and functional requirements analysis
- software selection
- establish policies, metadata profiles, resource creation guidelines, copyright clearance procedures etc.
- implement software with interoperability and access control
- assess copyright restrictions on identified resources for inclusion in the repository
- add resources to the repository with metadata

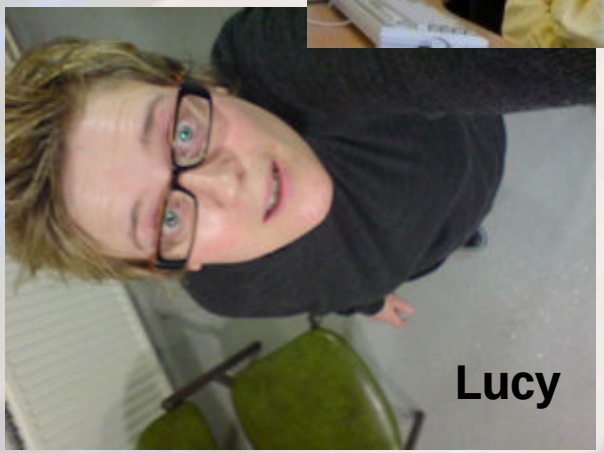
Who? Meet the Team

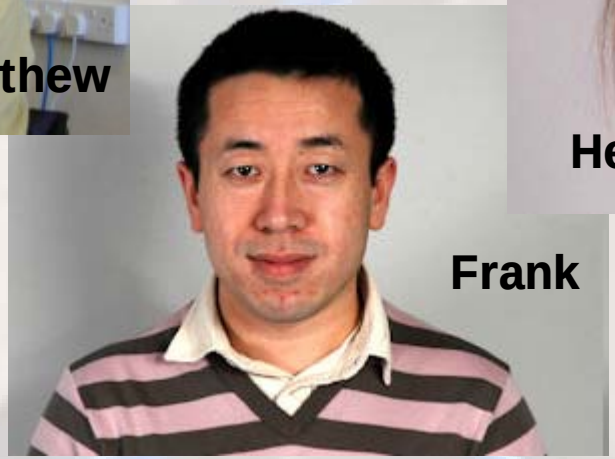

Peri


Julie


Matthew


Helen


Lucy


Frank

THE UNIVERSITY of York

Who else …

- Elizabeth Harbord (Project Director)

Project Team

- Wayne Britcliffe
- Anthony Leonard

Steering Group

- John Local, Julian Richards, Stephen Town

Academic Advisory Group

- Subject representatives and others

What do our users want?

Our users are both the creators/depositors and consumers of content

- they want to put stuff in …
- … easily
- and get stuff out again
- … easily
- to have access in various places (classroom, laptop, library, home PC)
- re-use it in new ways
- teach with it, research with it
- make it available to students
- control access in various ways
- maybe even sell stuff
- without unnecessary barriers to use
- and with the promise of safe, secure and well-managed storage

What do our users want? (2)

- they want different kinds of metadata and classification
  - a fine art image is not a musical performance is not a movie is not a transcription of an interview is not an archival document …
  - history of art is not archaeology is not music is not linguistics is not theatre …
- describing different kinds of things

What kinds of resources?

- Images – 2D, 3D and 4D in a range of file formats and sizes
- Digital audio files, including musical performance and broadcast materials
- Digital video and film, including performance and broadcast material
- Transcriptions
- Digitised text and manuscripts
- Web resources
- Presentational resources
- Datasets, for example statistical, experimental or analysis data
- Collections and aggregations combining any of the above types
- Archival finding aids

The content is out there

- History of Art slides and digital images
- Archaeology slide collection
- King's Manor slide collections
- Firthian archives in Language & Linguistics
- Child development video and audio
- Theatre performances from TFTV
- Commercial music, video, film
- More ...

What does this mean?

- We need a system that is
  - Flexible, customisable, future-proof
  - able to handle different types of data
  - and different types of metadata
  - able to control access
  - able to accept deposit in various ways
  - and provide access natively and via tools
  - handle very large files (.wav, video)
  - and complex objects
  - and can integrate with York infrastructure

Integrating approaches ...

White Rose Research Online – the place for research papers and scholarly texts – shared with Leeds and Sheffield

The Yorkshare Virtual Learning Environment

More integration …

- Single Sign On and Federated Access Management
- Student information management systems
- Library catalogue
- Specialist tools: image viewers, audio analysis, streaming services
- Interoperability/integration beyond York
- … and more

Scope is important : what's out?

- Research publications or any similar materials -> White Rose Research Online
- Courses and course materials/ learning objects -> Yorkshare
- Current web pages -> YorkWeb / Web CMS
- Corporate and administrative records, management information -> YIMS
- People information and identity management
- Mutable information, word docs, emails etc.
- Collaborative project working and collaboration tools
- Large datasets and complex scientific data -> data archives etc. (at the moment?)
- Mass digitisation projects, running and managing them at least, storing the outputs is in scope

Where are we now?

- Functional requirements
- Policies and guidelines
- Software Evaluation / Recommendation
  - Fedora (Open Source digital library architecture) & Muradora (Open Source interface)
- Data migration
- Custom metadata profiles and creation
- Customised interfaces

Open Source architecture

- Fedora Commons –
  - flexible digital library architecture
  - everything is an object described in XML
  - it's just a storage layer
  - different options for storing content
  - open APIs for ingest and access etc.
- Muradora
  - open source front-end
  - access control potential

Some benefits of a digital library

- Opening up existing resources to a wider audience, promoting use and re-use
- Making inaccessible resources accessible
- Bringing access to all, across disciplines
- Improving practice in creation and use
- Offering a secure, safe home, taking the management headache away from Departments
- Increasing opportunities for collaboration
- Expanding the library role into digital curation

Some challenges

- layers of expertise, interest and engagement differ hugely
  - how do we engage users effectively
- copyright and controlling access
  - how do we make the best use of resources legally and educate others to do the same
- no clear divide between research and teaching
  - how do we avoid confusing users with decisions about where to put stuff?
  - how do we embed our service into researcher & academic workflows
- managing expectations
  - how do we strike a balance between what is possible and what is wanted, now

Digital library philosophies

- Openness rocks
  - use open standards and open access as much as possible to allow transparency of information flow between machines and humans
- Confusion confuses
  - let administrators and machines make as many decisions as possible
- Reinvention is pointless
  - if there is a tool that already does a job well, use it, don't duplicate it
- Balance and flexibility in all things
  - aim to strike a balance between offering mass appeal and individual tailored services

Is this relevant to ‘Data’?

- Current focus is on humanities subjects, specifically History of Art … they have ‘data’
- Fedora offers a flexible infrastructure,
- which can store and manage any type of ‘data’
- and has APIs which can be opened up to technical staff
- future research proposals and projects can make use of this new infrastructure
- we already recognise that research data is valuable, and have started auditing collections
- part of our remit is ‘data curation’, we just need to work out what we can and should do, at a digital library, library and institutional level


Questions?