

Web / storage options for ROE History Project

Andy Lawrence 2024-06-14

This is a skim over the various options. If we narrow things down we can do more investigation.

We will need storage for a variety of assets - videos, audio recordings, images, digitisations, and our own documents and notes.

Ideally these assets should be stored in some kind of organised database rather than a jumble of flat files - i.e. a Digital Asset Management System, (DAMS) which enables us to search and browse, and add new things. This also means that ideally we should associate standardised metadata with each asset.

We also will likely want some kind of public interface, i.e. web pages, to showcase ROE history. Again, to make this easy for us, and to enable shared editing, we need a Content Management System (CMS).

1. Requirements

1.1 Size

Simple documents are of the order of tens of kB each, so this is not the main driver of storage requirements. Digitised images are of the order of a few MB. If we allow say 10MB per image, then 100GB will hold 10,000 images. Video recordings can potentially be far larger. For example an MP4 recording I made for an online course that was 21min long, at resolution 1280x720, came in at 697MB. So an hour's worth of an oral history recording could be say 2GB. However, if we are happy with only modest quality, I found, using the dark arts of ffmpeg, I could compress the size by roughly a factor of 16. So to store 100 interviews at full resolution may be 200GB, but storing at low-medium quality maybe only 10-20 GB.

My conclusion is that we probably want to ensure storage of the order hundreds of GB , but we don't need PB scale storage.

1.2 Useability

We will soon have a multiplicity of documents and videos, so I think key requirements are (a) shared space that we can all add to, view, and edit; (b) some kind of organised and searchable DAMS ather than just collecting files in folders; and (c) a decision on a standard set of metadata that we record with each file. However, this may take a little while to get going, so we might set up some kind interim storage that is at least shared. Likewise for the public interface, we need a CMS that makes it easy for any of us to edit the web pages.

1.3 Simplicity

We are not going to hire a software developer or system manager to run things for us (although we might ask for help in initial installation.) So for both the asset management system and the web management system we need (i) a "turnkey" system, i.e. one that is ready out of the box, and (ii) a system with an easy to use interface, and (iii) a system that provides some level of long term security and backup. Also, it would make sense to go with methods that some of us, or other local staff, already have experience with; and/or methods that are already used by local partners, such as NMS and the University Library and Archives.

1.4 Expandability / future proofing

In the first instance, as we concentrate on the oral history project, we want to keep things especially simple. But it could be wise to take decisions that will suit possible more general future work on ROE archives, and from my point of view at least, a solution that can be used for both

ROE and JCMB archives. We also want to bear in mind that eventually we will have a new Librarian/Archivist, who may well have their own ideas!

2. Storage options

2.1 ROE-IfA data storage

We could easily ask for hundreds of GB storage on the Cuillin cluster. We would need to back it up ourselves, but its not too hard to set up. It would be easy to have shared access to simple flat files. There is no pre-installed DAMS, but we would be allowed to install our own.

In principle we could buy our own workstation/PC but this would commit us to some installation, maintenance, and replacement.

2.2 ROE-STFC data storage

I have no idea what options might be, going the STFC route.

2.3 University data storage

The University runs a centralised DataStore that is Windows based. Every staff member has a standard allocation of 500GB and group sharing is possible. It should be possible to work with University IS to set up a special project with dedicated storage, and maybe even get assistance with installing software. But we would all need University accounts I think.

2.4 Commercial storage

Cloud storage typically has a monthly cost. For example, a standard Google account has 15GB for free; 200GB is £2.49/month, and 2TB us £7.99/month. Web hosting options (see below) typically include some storage alongside, but of a relatively modest size, of the order 25GB.

3. Public (web) interface options

3.1 ROE-based personal web pages

Jason maintains the www.roe.ac.uk web pages, but its not obvious how to share editing, and it is overdue an overhaul anyway. On the Linux network, everybody has a "public_html" directory on spider, which makes personal web pages possible by manually writing html and css. For example, John Peacock's page is quite nice at

<https://www.roe.ac.uk/~jap/>

But there is no content management system, so it is quite hard work to maintain, and hard to make it really nice - for example John's page doesn't work properly on a mobile device.

Every Cuillin account also has a "public_html" area, and this likewise makes public web pages possible - for example

<https://cuillin.roe.ac.uk/~jharno/SLICS/>

There is no CMS pre-installed, but in principle we could install our own version of Wordpress.

3.2 STFC web pages

ATC has STFC hosted web pages at

<https://www.ukatc.stfc.ac.uk/Pages/home.aspx#>

I have no idea how this works or if it is a possible option to create a special page for this project.

3.3 University web pages

The School of Physics and Astronomy makes personal web pages easy to set up in a very similar manner to the roe/spider method, but once again with no associated CMS. On the other hand, the University's own web pages are assembled using a standard CMS - Drupal as the back end, with EdWeb as the front end. (This might be changing.) It is possible to set up organisation or project specific websites using the same technology, for which limited numbers of people can have editing rights. An example is the IfA website at

<https://ifa.roe.ac.uk/>

and the ScotDIST website which myself and Joe Zuntz set up a few years ago:

<https://www.scotdist.ac.uk/>

The standard EdWeb design is rather samey and blocky, but not bad. The editing interface is graphical and reasonably intuitive but a bit clunky. It would also have a bit of a fixed University look-and-feel, and I think everybody would need to have a University account.

Alternatively, it is quite easy to set up a Sharepoint site, but I think these only work internally.

3.4 Github pages

If you have a Github account (free) its not hard to set up a set of web pages. An example is a conference web page set up a few years ago by my Nic Ross, myself, and Alastair Bruce:

<https://quasarsincrisis.github.io//index>

3.5 Commercial web hosting

There are many commercial "make your own website" options that work straight out of the box. They often give very nice results. For example, the current ROE Trust web site uses Weebly :

<http://roetrust.weebly.com/>

This is very easy to edit and give nice results (Clive and myself both do it). We are on the cheapo free version, which has some limitations. But we also pay for the domain roetrust.org.uk through names.co.uk for £16/yr. The address

<https://roetrust.org.uk>

redirects to roetrust.weebly.com

My own personal website

<http://andyxlastro.me>

is run using Wordpress, on the webhosting service UK Cheapest. This provides the domain name (£13/yr), the web hosting, and the Wordpress software (currently at £47/yr but started at £22/yr...).

I think Wordpress gives the best combination of ease of use and attractiveness. In principle it might be possible to install it on an institutional machine, but this would involve some work or some help, and commitment to keeping it up to date.

4. Database / team interface options

4.1 ROE Media space

Jason maintains a system for storing and accessing media files at

<http://media.roe.ac.uk>

This uses an open source software package called ResourceSpace. This can store all of documents, images, videos, and audio files. It is moderately easy to browse. I don't know how easy it is to add material, or how much space is available, or how easy it would be to carve out a distinct presence as it were. But this may be the easiest route to get going.

4.2 ROE Library

The ROE library holdings are searchable using the "ROE Explore" system:

https://ed.primo.exlibrisgroup.com/discovery/search?vid=44UOE_INST:44UOE_ROE

Under the hood, this uses the ALMA Library Management system, with software called Primo as the front end. The data is actually handled and stored by Edin. University library. Although ALMA was originally designed as a Library Management System, it can also deal with more general digital collections. I guess this is also the system underlying the University Library's own search system DiscoverEd:

https://discovered.ed.ac.uk/discovery/search?vid=44UOE_INST:44UOE_VU2

4.3 University Dspace

The University also provides a digital repository service called "Collections Manager" using the open source Dspace repository software

<https://collectionsmanager.is.ed.ac.uk/>

Some of School of Physics and Astronomy photos have over some years been digitised and deposited there. The interface and the metadata are very poor - its hard to search. I don't think this is a problem with the software itself, but how it has been implemented. Its hard to tell who is running this. I don't what the relationship with University archives. I think it is more aimed at research groups cooking their own etc.

4.4 NMS

NMS have a very nice search interface for their collections.

<https://www.nms.ac.uk/explore-our-collections/search-our-collections/>

I don't know what the background technology is. If we hand things over to NMS, we know they are findable at least. If we have a public interface of our own with selected things showcased, then for example, we can link to information about items held by them, e.g.

<https://www.nms.ac.uk/explore-our-collections/collection-search-results/comparative-scale-case-thermometer/217225>

4.5 University Virtual Exhibitions

If we deposit things - physical or digital - with the University archives, there is a way to showcase them using the "Virtual Exhibitions" system. Here is an example:

<https://exhibitions.ed.ac.uk/exhibitions/library-lives>

5. Conclusions

Obviously we need to discuss this as a group, but my initial thoughts are

- (i) To get going asap on an interim flat file storage solution - either Cuillin or DataStore
- (ii) To set up a University EdWeb site as our showcase site
- (iii) To do more research on the best DAMS solution, but at the moment it looks to me like the best choice is either building on Jason's mediaspace or the ROE Explore system.