I’ve called this presentation "putting meat on the bones" because it is about how respond to the challenge of enriching metadata and engaging with our collections when we are under constraints in resourcing. There is really no point in doing lots of digitisation if it doesn't include the aim of engaging communities and wider audiences, and if the metadata is not up to scratch, people will not engage, due to an inability to search and browse.
I work in a Cultural/Heritage environment, and Edinburgh, as you probably know, is something of an arts and historical mecca. I could talk at length about its inspiring the writings of Sir Walter Scott and Sir Robert Louis Stevenson, or that it spawned painters Sir Henry Raeburn and Alexander Nasmyth, or that it is called home by the likes of JK Rowling and James Bond (indeed, there is a picture of Sean Connery from his early career in our collections which, in the interests of good taste, I have decided not to include here). I could talk about its many or famous theatres, art galleries and museums, or that it hosts the enormous annual international arts festival. I could, but it's not strictly what I'm here to talk about.
But being in Edinburgh, we are trusted, prestigious and a magnet to donors. The Special collections department has 400,000 rare books, buildings full of archives and 60km of manuscripts. We have enviable museums, most notably Musical Instruments, Art, Geology and Anatomy.
And one of the most accessible ways of finding our special collections is through our Image Collections. Now, these are not only images of our collected objects, but of born digital material too, and one of my roles for the last few years has involved getting as much of it as possible online through the LUNA Imaging platform and devising workflows to make life easier for our Digital Imaging Unit. Note that the person who put this page together is subconsciously obsessed with skeletons. I’m not sure he’s aware that he put three on this screen. I looked at this and thought I could maybe run with that...
LUNA is a commercial product admittedly, but our close relationship with the support area has allowed us to influence how it is shaped- which offers high-res zooming through JP2K files. We have only 50,000 images and page-turnable books in our main site, but we have a secondary site with a lot of content that just isn't ours... It's an image management tool as well as a viewer, and is central to everything we do.
Our Digitisation team really consists of two people and their camera. We can't begin to guess how many images equates to a fully digitised library, but it would take several lifetimes at the current rate. There was talk of them working shifts to ensure the camera is constantly making images. They do an incredible job. The difficulties don't end there though- the way that public sector educational purse strings are pulled, there is often a better chance of getting a high-end camera in place, rather than spending money on somebody to catalogue the items, and currently there is nobody to do that. This means we need to get hold of metadata- to put meat on the bones of the images, make them searchable and engageable- we need to be resourceful and creative. I'd like to take you through our cataloguer-free process (which has CSV at its heart, just in case you're wondering where I'm going with this).
Let's take as an example our Incunabula collection (here's a good page: German Bible, 1483)- Incunabula are the earliest printed books- prior to 1501. People get very excited about these kinds of pictures- I think it's the slightly gruesome, violent content, viewed through medieval eyes, which were fairly politically incorrect.
Now, this might be digitised because someone in the department feels it's historically significant, and the whole thing should be digitised, but maybe not! Often what will happen is that a visitor/reader will request its digitisation because they've found it, liked it, and are willing to pay for the work. Great. They fill in a form, which, to my utter chagrin, is still paper-based (there has been a review of process going on for some time- I've built them an online process, but they're not sure they're ready to implement! Things do move quite slowly in the special collections world). This will give us SOME identification metadata, perhaps even some searchable tagging, but we can't really hassle the user for academic level information, so we must take what we're given- or knock it back if they haven't said what they've seen. Indeed, we have improved the quality of their form with a view to improved discovery, but we also have an enormous backlog which isn't going to be re-opened.
The request then goes down to the photography department. They take a high-res picture of the item, and this then passes through Photoshop for cropping etc, the images are archived (at such high resolution, each picture and its derivatives take up approximately 0.5Gb server space) and the metadata is entered to- yes-
the Photographer's Excel spreadsheet- again, we've built them a web-based alternative, but they don't want to let Excel go, let's be honest, there a lot of things that Excel does well and quickly here that would actually be really hard to replicate; although they do accept that the concept of one user in at a time, and various quirks can be challenging). This has a macro underlying it which does all the bits and pieces that they need to allow this data to seamlessly flow into the LUNA system.
The Photographers here update the spreadsheet with the data they've been given. I dread to think how long even doing this from a paper-based form takes - I try not to get involved with that bit. We've tried to simplify the system for them as much as possible though - eg using shorthand, such as a RIGHTS CODE, which can be extrapolated out from a couple of characters, to tell us the rights status of the original work, the rights status of the reproduction, and the creative commons licence status. They also make use of Excel filtering and auto-complete to ensure that the Shelfmarks and Author forms that they put in are consistent. The Author name will get tossed around into various formats for display etc. Of course, inherent machine-generated data, such as EXIF, is already embedded in the image, and LUNA can read and display that, to allow the technical status of the image to be seen.
One of the bits of required functionality from the macro was to embed descriptive/identifying metadata into the image, which means that, should we lose its link with the system, we still know what it is. This uses a command line utility called EXIV2, which allows you to 'squirt' XMP and IPTC data into the item. There is absolutely no support available for doing this in Windows, so we are pretty much on our own calling this from Excel.
We then generate CSVs for upload into the LUNA system. There are about 20 LUNA collections which could potentially be updated here, and resultantly, we need to create a separate CSV per collection. You can see that the code to generate the CSV gets pretty unwieldy, but this is due to a huge metadata template (largely based on the VRA standard). Managing that is a bit of a challenge, but it’s now quite robust. It does mean a bit of hard-coding though, as the system expects a specific name for each CSV.
So, the data gets into LUNA. Great- but it does result in something of a skeleton record (apologies). It's good for identification, if you know what you're looking for, but would you believe there is nothing on that record to actually tell you there's a skeleton in it? If you did a search for skeleton, that would not come up. Therefore, we need to find other ways of enriching the metadata.
Inevitably, as is the way these days, we went to the crowd. We developed some fairly low-brow metadata "games"- we emphasise the low-brow nature by making them look like they're from the 1970s (you could call them that, but possibly wouldn't), with a scoreboard, voting modules and blatant bribery to get people to play them. They were, amazingly, successful. We've got 10,000 tags across 4,000 images back into LUNA from our first few goes at this, and we have another 20,000 at various stages of the moderation/voting process to make them acceptable for upload. We've found that we'll only get this data if we hold an event- while they are online and available, we don't yet have the community engagement to find that people are adding to our stuff without us pushing them. We've also engaged with crowdsourcing specialists Tiltfactor and Zooniverse to look at other ways of doing this.
Once we've got some tags, we're then in a position to push this garnered material back into LUNA. We do this by taking an extract from the MySQL database that underpins the game, converting into CSV, and, through Authenticated REST API programming, push it back into the LUNA system. It's fairly robust, and we do it ad-hoc whenever we have enough approved data to justify running it.
Technically, that covers most of the basic workflow and systems we have in place to get metadata onto our images. But there is more we can do. As we’ve seen from perhaps higher profile institutions, engagement with your collections and getting the community familiar with you and your content is invaluable. To that end, we do try to promote these things as much as we can. We built a Flickr API (and you can garner tags from that of course), and the Flickr collection gets more hits than our main site. The BBC/Public Cataloguing Foundation photographed a lot of our items, and we got those onto their ArtUK platform. And then we make our data available, through OAI-PMH (this is a picture of that, which may make no sense otherwise!), for whoever wants it. Some people do- from Informatics students building Apps to Artists who have used Google Analytics data in conjunction with images and metadata to make some interesting things.
Europeana have made use of the OAI-PMH accessibility to get about 15,000 of our images into their cloud platform, which obviously raises the profile, and with the APIs developed there, we're hopeful there will be more potential to enrich metadata. It's a pretty big pond though...
Another thing we can do to raise our profile as a statement, but also in allowing content to be accessed is to make images OPEN. Insofar as we can, we're aiming to get images CC-BY by default, and allow their download resolution to be the biggest possible. There are huge swathes of images which we know are out of copyright and that we can apply this to, so the idea is just to do it. Selling the images online has never been a particularly lucrative income stream, and we want people to engage and use them.
To that end, we’re going open-­‐IIIF. This framework promotes interoperability-­‐the idea is that an image is only ever physically hosted once, and that the user can access it through a URL. Because of this, you would think there’d be a lot of restriction on what you’d want to do-­‐but the URL is tremendously clever, and allows you to format your image in all sorts of ways.
To Recap

- Let reader do the work
- Robust workflow
- EXIF data
- The crowd can help
- Get community engagement:
  - Make data available
  - Make images open
Links

Image Collections:  http://images.is.ed.ac.uk
Collections Repo:  https://collections.ed.ac.uk
Metadata Games:  https://librarylabs.ed.ac.uk

Contact:  lddt@mlist.is.ed.ac.uk, scott.renton@ed.ac.uk
@u0eldd, @scoatch