

SAPCON CONFERENCE: Preservation on a shoe-string
William Humphreys Art Gallery, Kimberley.
25-28 July 2006.

PREVENTION : THE ONLY CURE –
PLANNING A SPECIAL COLLECTIONS IMAGE DATABASE

Mimi Seyffert, University of Stellenbosch libraries, Special Collections.

My topic for today is Prevention: the only cure – planning a Special Collections Image database. I want to talk to you about our current situation concerning preservation and the solutions we are planning for the future.

Firstly though, I would like to sketch some background on our setup at the university library. I always find it quite useful at events like these to find out what colleagues in different parts of the country in different institutions are doing and how they are going about the business of collecting and preserving, so I hope that the information I'm sharing with you today might be similarly useful.

The JS Gericke library is the main library of the University of Stellenbosch, situated underground at the heart of the campus. We have been at these current premises for 23 years and luckily the library was designed with the Special Collections section in mind, resulting in the good physical layout we enjoy today, including a strong room and compact storage facilities. Unfortunately the growth and expansion of the section could not be foreseen, especially as Special Collections are sometimes viewed as stagnant, so we are running out of space fast.

The section consists of three subdivisions, namely Africana, Rare Books and a manuscripts section. These three sub-sections function together quite well as a unit, in many cases providing the user with the primary resources on the one side and the secondary and supporting resources easily accessible in the same vicinity. We are currently five staff members, most of us doing work in all three sections simultaneously.

Our Africana section consists of about 20 000 items, mainly books, ranging from a 1556 Leo Africanus edition to Madam and Eve. We collect works on Sub-Saharan Africa – history, natural history, art, politics & government, culture etc. and we also have a few specific areas of focus namely works on Stellenbosch and the Western Cape, Anglo Boer War material and South African art. Books are consulted in the Africana room where the staff member on duty will get the required books from the shelves for the user. Our most valuable items are safely secured in locked glass cabinets, free from dust, fingers and other nightmares.

In our Rare books section we keep everything that is too large, small, old, heavy, controversial, expensive or just weird, to be kept on the open shelves of the library. It is a real treasure trove of curiosities in print and boasts a very controversial and expensive recent addition in the form of a 45 kg, 50 x 50 cm, 800 page tribute to Mohammed Ali, bound in Cadillac pink. Seriously though, the section is used extensively by art students, as it houses a large collection of expensive art books, which takes quite a beating through the year.

Our manuscript section houses approximately 370 collections, ranging from large collections spanning 15 meters shelf space to smaller ones, being a single volume or even a page. We have collections mainly of Afrikaans literary figures, artists, political figures, journalists and academics. Some examples of our bigger collections are those of CJ Langenhoven, NP van

Wyk Louw, Maggie Loubser, Sheila Cussons. Items in collections are individually numbered and important collections are described in detailed catalogues, providing the researcher with summaries of every single document in a collection, as well as an extensive index. This reduces unnecessary handling of material, as the user can decide from the catalogue exactly which documents he or she might want to look at. Sorted collections are housed in acid free, boxes and folders, an expensive, but necessary means of preservation.

All three subsections are protected by CO₂ gas in case of fire, as opposed to water as in the rest of the library. We're supposed to house our collections at a constant temperature of between 19 and 21°C and between 48 and 50% humidity, but I'm afraid, I don't always trust the air conditioning. No material leaves the section and staff members try to keep a constant eye on the handling of material by users. Photocopies and other means of reproduction are allowed in most instances.

Now, that all seems pretty good, on the surface anyway. So, where do the problems creep in. The library service has been in a very lucky position for the last forty odd years to have the services of a full time restorer and bookbinder, as well as a University bindery, and our Special Collections section has been one of the primary beneficiaries of these services. Those days, however seems to be numbered and the future of our restoration laboratory and the skills employed there hang in the balance. We are so used to relying heavily on the skills of our restorers to mend and protect our material, that it seems almost unthinkable not to have them around to just run to quickly in a case of emergency, for example to rebind a volume for exhibition or to restore and encapsulate a torn document or to dish out some much needed advice. So now we have to start imagining how we can get by without these luxuries and focusing on what we can do ourselves, lacking the skill, but with the tools available to us to protect our collections and prevent further damage, especially to our extensively used materials.

Over the last couple of years our section has grown extensively and our areas of focus has also changed to keep up with the times. The demand for information in digital form increased steadily. Now, if you are sitting on material which is not available anywhere else in the world and you get a request for information in digital format you cannot in the twenty-first century reply, sorry, you have to fly all the way from Europe to come and consult our catalogue to see if we might have what you're interested and then we might make some photocopies for you, charged at 34 and a half cents per copy! Students, our primary users, have of course also started demanding scans instead of photocopies, especially art students. I ventured to the art department a while ago to do some investigating and was quite astonished to see the advanced equipment that especially the graphic design students work with, and now have a much better understanding of why the students need digital information and why a black and white photocopy just won't do. And as I said, if we have the only copy of a certain image, we cannot deny our primary users this just because we do not have the means to supply it to them, in the format they require.

About five or six years ago, we acquired a scanner and have been scanning on demand ever since. Now this is fine for supplying the user with the requested information in digital format, but I found myself duplicating work, scanning the same images year after year. When I saw this pattern forming I did obviously try to save these images on my computer and on CD's, but one's desktop can only handle so many image files, and the ever growing stack of CD's on my already cluttered desk, was starting to irritate me. I realised that I was not working according to any kind of intelligent system and it was costing me time and some of the material was showing signs of wear and tear from unnecessary handling. If I could tell the art students to browse through a digital database or if I could access the photograph of the 1905

Victoria college teaching staff instead of scanning it at least 10 times for each university department who celebrates a centenary and need that picture for their publication, it would make my life much easier and probably indirectly assist the preservation of unique material. So obviously the idea of a digital image database came to mind.

Now we've been introduced to the idea of digitisation as a means of assisting preservation some time ago and I at a stage I did a feasibility study for digitising some of our material. It was however decided then not to follow that route. In the following years the subject came up again and I wanted, for example, to use already scanned material for a virtual exhibition. Somehow I kept bumping into brick walls and could not find support from the IT department or even colleagues. When I broached it again last year, I was asked, rather out of the blue, but haven't you heard of DSpace? Don't you know that we're already implementing DSpace for an E-theses project? Well, no, I didn't, but thanks anyway. So, DSpace was introduced to me by accident. The library decided to embark on an e-thesis project in conjunction with the University's IT department and DSpace was decided on as the software to use.

DSpace is an open source digital repository system which captures, stores, indexes, preserves, and distributes digital research material. It looked as if it could work for what I have been planning in my head – that is to have a system where I could digitize an image or other information once, describe it, store it on a server and retrieve it when I need to. Luckily the library had acquired a new server for the E-Theses project, so that expense was already covered. DSpace being open source software means that it can be downloaded without any cost.

Now that we had that in place, I could start planning in more detail what I wanted to achieve with this digital project. I realised from the start that we did not have any extra funds to start a project which involved outsourcing any functions and we could not really spare any other personnel for the project as ours is an already understaffed section of the library with a great deal of daily user traffic and mountains of indexing work, especially in the manuscripts department. This means that I had to think carefully about how I could streamline the work to fit into our daily routine.

The first thing that was important to me was to have the structure of the digital database logically set up to reflect our section and the somewhat confusing substructures within it. It was important that I communicated these ideas to the IT person who was the unlucky one to be nominated to help me. What I saw in my head sometimes proved to be quite difficult to convey to my IT colleague, but we eventually managed to sort something out.

The other factor was determining which material to add to the database. As I mentioned I already had stacks of digital information. This could be added immediately and described. Many of these images were however not scanned at 600 dpi, which is the best practice when it comes to digital archival masters, but rather at 300 dpi and in JPEG format, rather than TIFF. I made the decision to add these anyway, since the work has been done and the quality of the images were good enough to use in print. Most of these images were scanned for publishers and already used in print.

The condition of the material also has to be taken into account. Sometimes badly damaged and worn items can be digitally reformatted and made available to the user. We've recently worked with a collection of valuable old pamphlets which are printed on very acidic paper and literally crumbles under your fingers. Our bookbinder took these apart to see what he could save, but many of them are beyond repair. These can now be scanned in, using Optical

Character Recognition (OCR) software, which will also make them fully word searchable, adding value to the items.

Now I also had to start thinking about frequently used material. Concerning the art books it was quite easy. I know the yearly topics of the students and decided when a student asked for one or two images from a certain catalogue, I might as well scan the rest too, because experience had taught me that the rest of the class would be around soon enough for the same type of material. I would then put a notice on the book that students can consult me for digital images from the catalogue in question. I would not be comfortable with publishing it to the web at this stage because of copyright issues. Students can then look at the digital images and I can e-mail the images directly to them or copy it to a CD, obviously for research purposes only. It then takes the place of the photocopy.

When it comes to material from our manuscript department, I'm definitely not at this stage embarking on any major projects concerning the scanning of total collections. It would be a waste of time and money as those documents are housed safely and described in detail, as I explained, to minimise handling. Also, it is not our idea to duplicate the work we already do manually, electronically, or to replace any of the work already done. We want to use technology to add value to our collections.

Concerning manuscript collections, it is difficult to judge which collections are frequently used. Once every ten years as opposed to never in a 100 years can count for frequent use in the world of manuscript collections. There are however some collections which we know from experience are really frequently used. For identifying these I had to call in the help of my colleagues, who have been providing information service in our section for many years. One such a collection we focused on is our general collection of University of Stellenbosch photographs and I'm pretty sure that in 12 years time when the University celebrates it's centenary, these images are again going to be in great demand, as they have been previously with the centenary of the Victoria College, forerunner of the University of Stellenbosch.

I've been scanning these photographs at 600 dpi and storing them directly on the server. I try to set aside a few hours everyday to do this. I have decided to describe these photographs as comprehensively as I can and to make sure that all the metadata is correct. I do not want a case of garbage in, garbage out. This can be quite time consuming, as in many instances I have to do a bit of extra research to verify some information regarding the photographs, but I'm sure somewhere in the future someone will appreciate the effort.

I might also just mention here that we use the Dublin core metadata element set in DSpace and I must admit that for a librarian, it remains difficult to think in terms of keywords as opposed to Library of Congress subject headings and name authorities which will forever haunt you, so I have decided to use these where appropriate and try to work according to some standard, even if it makes things just easier for myself. With the kind of material we have it is after all important to distinguish between the hundreds of Malans, Mullers and Malherbes and Wilcocks the building or the Wilcocks the man.

Another part of the metadata description which is important in our case is to state the source, and specifically the collection the document comes from. The distinct individual number of the document has to be reflected in the record and I need this field to be searchable. Thus far we have not yet overcome this obstacle.

Now what I've done up to now I still regard as experimental and realise that I still have a long way to go. But through this exercise and actually making a start, some things have become apparent. We still have a long way to go in terms of writing policies and copyright statements. I do feel however that we have moved forward a bit and have something to work with now and at least a space on a server for storing digital images and other information.

I'm not going to stand here and discuss the ins and outs of DSpace, as its website is rather self explanatory and offers a lot of information and for those interested I urge you to have a good look. I will however share with you my personal verdict on DSpace. Firstly, I find it very user-unfriendly. Maybe not so much from the end-user's perspective, but definitely from an administrator's perspective. The fact that we need a programmer to set things up for us, has definitely been a big drawback. There has also apparently been problems with creating privileges and authorizations for the different collections and sub communities which one creates in DSpace. DSpace is set up that one can register users as administrators, submitters or readers. This affects the availability of records loaded under a certain set of authorizations or rules.

When it comes to loading images and metadata, I find it a cumbersome process and I'm not sure to what extent this can be streamlined, but I don't think it will make a huge difference.

The search capabilities have also been a bit disappointing. To my knowledge, some improvements can be made, but it is as if the system does not always respond the way I want to and seems to be a bit unsophisticated. It also does not allow for diacritics. On the other hand though, I realise that one should use the system for what it was designed for and perhaps change your thinking, and not try to make it something it cannot be.

As with most other open source software, there are continually appearing new versions, which of course means updates, and one has to be aware of what's happening. This means forming part of the "DSpace community" and using the information and forums they have on their website to your advantage.

Overall, I think that the content management side of DSpace should be in the hands of a librarian, not necessarily an IT specialist. Also, most of us are not completely stupid and I think the days of thinking of Special Collections librarians as "technologically challenged" should be long gone.

I also just want to add quickly that Greenstone, which we looked at last year, seemed to me more user-friendly and easier to operate. It is said, however, that Greenstone is more ideally suited for smaller collections. On the other hand, Greenstone has better searching capabilities and you do not need a programmer to set things up for you. Our IT department feels however that DSpace is better for larger projects and has better possibilities in terms of publishing to the web and running on networks.

In conclusion then, sometimes we try to do everything so perfectly and more often than not, especially when you work in bigger environments where there is a committee for everything, it leads to forever reinventing the wheel. Instead, maybe we should just go out there and do something tangible, even if it is not at first as sophisticated as you want it to be. You learn as you go along, but you need something to work with, and this is what I feel I have achieved at this stage in the process.