The African Elephant: a digital collection of anatomical sketches as part of the University of Pretoria's Institutional Repository - a case study

Authors:

Amelia Breytenbach : Metadata specialist
Department of Library Services
University of Pretoria, South Africa
Email: amelia.breytenbach@up.ac.za

Ria Groenewald : Digitization Coordinator
Department of Library Services
University of Pretoria, South Africa
Email: ria.groenewald@up.ac.za

Abstract

Purpose – Although several collections have been digitized and made available in the University of Pretoria's Institutional Repository, a pilot study has not been done to measure the project management and workflow. The collections available in the repository at the time of this project were all long-term projects. There was a need to identify a project small enough to conform to normal project management requirements to use as an example to establish the planning and workflow of future projects. The purpose was to determine the outcome and quality of the final web-ready institutional repository product against specific digitization project goals.

Design/methodology/approach – A collection of anatomical sketches in the custody of the Faculty of Veterinary Science, Department of Anatomy and Physiology was identified as a possible collection that could comply with the above criteria. The different sketches in the Elephant collection could be digitized in phases, making it an ideal project for future comparison. In each phase a number of tasks were identified which the various role players should complete during the workflow process. Each phase would be compared to the previous completed phases to measure the outcomes and progress made in quality and time. Through successful interaction and collaboration between the Library and the Department of Anatomy and Physiology during the digitization process, valuable tacit knowledge could be preserved for future use in the field of Veterinary Science.

Findings – The completed project delivered on key areas such as the electronic availability of the collection through metadata description. Basic preservation of the physical collection was undertaken as necessary and the physical as well as the digital collections were archived for future use. The conclusion will describe the lessons learned and how it can be applied in future projects to the advantage of the institution.

Practical implications – The paper provides a very useful case study for other academic libraries who want to develop their own digital collections.

Originality/value – This paper offers practical help to libraries starting with digitization. It supplies valuable information for project management, planning of workflow and estimate time frames for completing a specific task in the digitization process.

Article Type: Case study

1. INTRODUCTION

The University of Pretoria, Faculty of Veterinary Science (hereafter *the Faculty*) has access to very rich and diverse heritage and learning resources. Since these are the only resources in South Africa that are also widely used in the rest of Africa, it is of the utmost importance to safeguard them and enhance their accessibility. The physical deterioration as well as the loss of information regarding these resources is being accelerated as there are no clear guidelines on the preservation of such items at the University.

The establishment of an open access repository at the University of Pretoria paved the way for the digitization of these valuable materials. The UPSpace interface provides for easy decentralized self-archiving by Faculty, and organizes the documents in logical, easily retrievable fashion. The repository uses DSpace™ software, developed jointly by the Massachusetts Institute of Technology (MIT) Libraries and Hewlett-Packard Labs. The software complies with the Open Archives Initiative (OAI), thus allowing items to be easily discovered by web search engines, services and indexing tools.

A project proposal was written by the Veterinary Science Library (hereafter *the Library*) digitization task team and selection criteria for digitization projects in the Faculty were documented. The way to ensure open access to the physical veterinary paper-based resources was now defined and could be implemented by the digitization of the collections.

The Faculty's Department of Anatomy and Physiology (hereafter *the Department*) has a collection of anatomical sketches made by Ms Christine Seegers in the early 1990's. Ms Seegers is one of the few trained biomedical artists in South Africa and her work is unique in the veterinary field. The sketches are artwork that will increase in value as time goes by and are valuable for teaching, learning and research.

Projects should comply with the strategy of the University and the Library and should be measurable. By using the selection criteria the Library could identify several valuable collections to be digitized. Three of the selection criteria points indicated clearly that the project falls in the criteria scope for digitization.

(a) Enhancement of the accessibility of the original paper-based resources.
(b) Metadata could be used to provide a better understanding and recognition of the original works in the Faculty and to make them electronically accessible.
(c) The resources are suitable and would enrich research, teaching and learning.

Key areas in which the project will add value for the Faculty are:

(a) Providing open access to valuable resources that are otherwise inaccessible for outside users.
(b) Supporting of lifelong learning through technology and improved learning initiatives.
(c) Possibility of earning a third stream income by selling of high resolution digital images.
(d) Capturing of tacit knowledge about resources that will otherwise be lost forever.
(e) The project can serve as an example of a knowledge management system at the University of Pretoria.

2. THE PROJECT

The project was a pilot project and it was small enough to measure the time and outcome of each task. It was possible to divide the workflow of the project into different sections with a definite start and end for each task in a section.

It was decided not to attach a specific time frame to the project. As this project was regarded as a pilot project for similar future digitization projects the team decided to keep the workflow as open as possible. The available resources, budget and time used in the normal daily work of the Library would have to be used for the project. No additional resources or funds could be allocated to the project, therefore only materials and methods that were absolutely necessary for the successful completion of the project were to be used.

Conservation and preservation are not formally applied by the University on resources of this kind. The Library personnel decided to include a basic conservation and preservation component into its digitization program. After digitization the originals will most likely be stored, and not be handled for some time. The Library personnel hope that by applying basic preservation, future deterioration of the original resources can be

prevented. Preservation methods are expensive and the budget constraints forced the team to follow the project plan conscientiously without any deviation.

The project team agreed to complete the work in the shortest possible period. This decision was based on the fact that one can only measure and learn from knowledge gained, positive as well as negative, with a hundred percent dedication from the team members.

2.1 Collaboration between the two parties involved

A meeting was held between the head of the Department and personnel from the Library (two persons involved with digitization and metadata and the leader of the Library). The Department was willing to cooperate and ascertain the condition of the collection. The Department had to reassure itself of the completeness of the collection as it consists of several separate, smaller collections of sketches i.e. sketches of the anatomy of the African elephant, African buffalo, horse, cat and dog, and had been stored for some time. The personnel involved in the digitization process decided to treat each group of collections as smaller projects that would form part of the main project: *Christine Seegers biomedical illustrations.*

To ensure the successful interaction and collaboration between the Library and the Faculty the following mutual understanding was reached regarding the digitization process.

The Department agreed to supply the following:

(a) The title of the sketch in a descriptive format i.e. *Dorsal or ventral view of elephant skull.*
(b) At least two descriptive keywords that would be of search value to the user and help the metadata specialist assigning controlled subject headings.
(c) File naming for the elephant sketches should be according to the documented file naming convention of the digitization team. The Department decided to use the abbreviation “oli_“ as a prefix. It represents the first three alphabetical letters of the word *olifant*, Afrikaans for *elephant.t*
(d) All the other information should be in English and in terms that would be understood by both casual and informed users.
(e) The information to be written on the back of the sketch canvas in a soft pencil (acceptable for conservation purposes).

The Library agreed to the following:

(a) The sketches to be personally handled by the team members, including the transport thereof to the main campus where the DigiBook scanner is housed, and back to the Department.
(b) Basic conservation methods to be applied to the sketches i.e.
    - cleaning of canvas area (very dusty)
    - wrapping of each canvas separately with acid free tissue paper after scanning.
(c) An inventory to be compiled by the project team with the following information:
    - Particulars of canvasses taken to the Library
        - date when canvasses are taken to the Library
        - filenames and numbers written on the back of each canvas
        - title of canvas
        - date when canvasses are taken back to Department
    - Final signature by Department to indicate that they received the canvasses back in the state agreed upon.
(d) Copies of the scanned images in JPEG, and PDF format to be handed back to the Department on DVD’s (supplied by the Department).
(e) Uploading to the digital repository, UPSpace by the project team with security settings stipulated by the Department.
(f) After completion of the project a meeting to be arranged with the head of the Department where the final quality control and the testing of the security settings are to be done.

2.2 Planning and workflow

After copyright clearance with the Department and finalizing the documentation for the first project, *The skull and vertebrae of the elephant,* the 47 canvasses containing 51 sketches were collected by the project team. A timetable was compiled with the following markers:

(a) Date of receipt of sketches
(b) Scanning date
(c) Beginning and ending dates of digital conversion process
(d) Beginning and ending dates of final quality control and metadata
(e) Completion date of project in Library
(f) Final approval by Department

A detailed workflow and itinerary was established to identify the order of the tasks to ensure a quality end product. A diagram of the workflow is available in Addendum A.

3. TASKS, ROLE PLAYERS AND TIME SPENT

The following is a description of the different tasks carried out during the project, the team members involved, and the estimated time spent on the task.

3.1 Create collection on the repository

As UPSpace is the preferred platform to house the collection, the project team consisting of the collection administrator, metadata specialist and digitization coordinator compiled an UPSpace project outline for uploading the collection to the repository. In this outline the collection administrator firstly informed the UPSpace manager about the new collection and gave a short description of the main collection and sub-collections to establish the hierarchical structure of the collection in UPSpace.

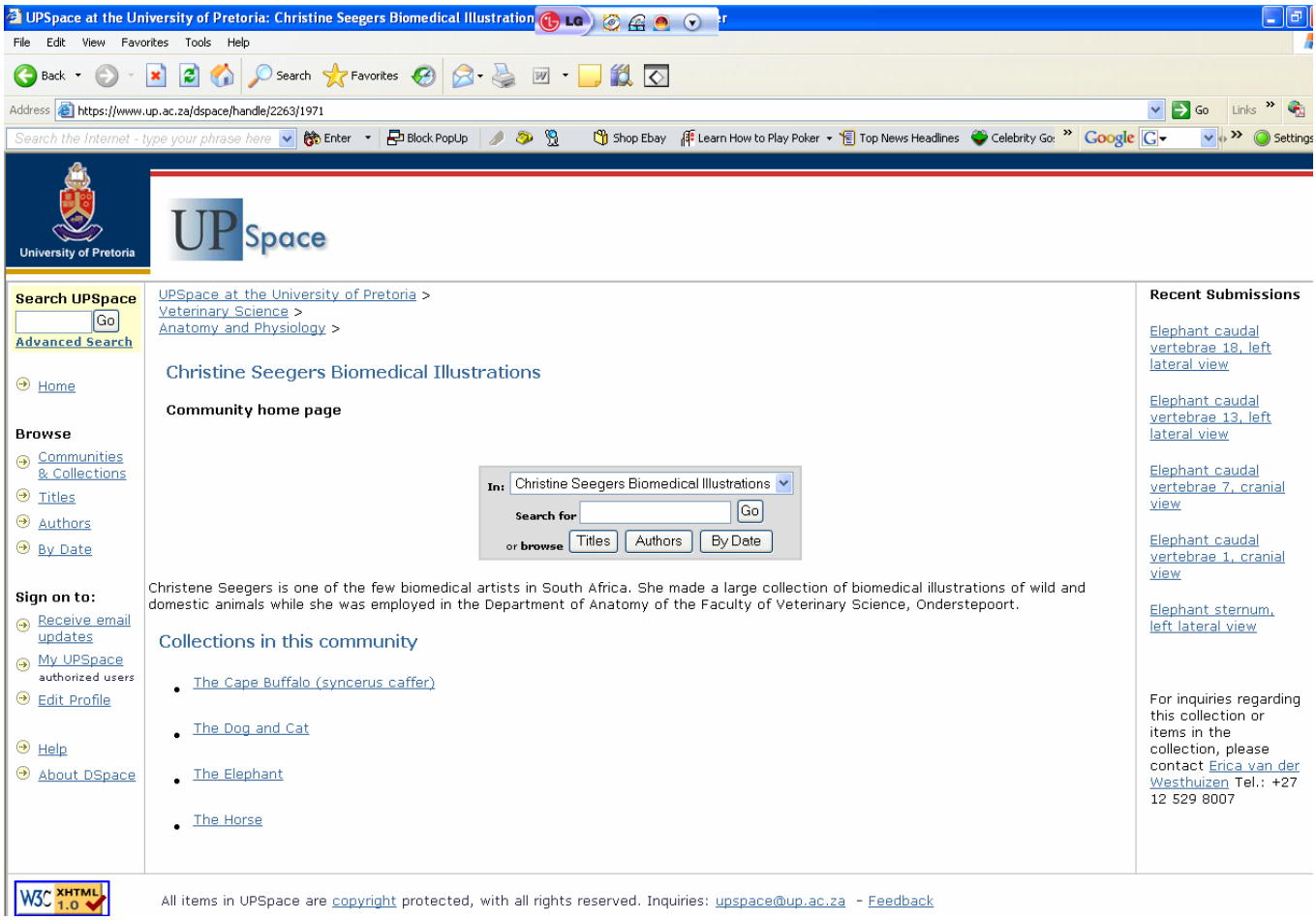


Fig.1: Hierarchical structure of the collection in UPSpace

A policy document was compiled to describe the acceptable formats for uploading, the criteria the digitized items must adhere to, and the persons involved in each part of the workflow. The UPSpace manager used this policy document to assign the necessary authorization for the various role players for uploading in UPSpace. Hereafter the collection administrator, metadata editor and submitter verified the collection and sub-collections created on UPSpace and did the necessary corrections. The metadata editor did the planning of the metadata elements to describe the sketches and to create a metadata template beforehand for the collection. The template was created with the metadata elements that are generic to all the items or sketches to be submitted to the specific collection. The technical metadata as well as the copyright statement were determined and these form part of the template.

The total time for completion of these tasks was 8 hours. All extra administrative time regarding tasks concerning the repository is included in these 8 hours.

3.2 Digitization process

3.2.1 The digitization coordinator was responsible for the creation of a database where the digitization process was to be noted in detail. All extra administrative time regarding the digitization process is included in this time frame. The creation, typing and editing of the database took approximately 10 hours.

Elephant	Date scan	Size of original canvas (cm) (w x h)	dpi	scanned sized (w x h) - KB	%	Date Conv	format	(w) x (h) pixels	ppi	Photo shop 8 work on image	format	size in pixels (w x h)	General info (Dimensions w x h)	ppi	Save for web	Download information	Adobe Acrobat	Password	Description
oli_001	16-Mar-07	35.5 x 25.5	610	8520 x 6400 pixels 156M	100	24-03-07	jpeg 1.26M	840 x 525	300	Mode: RGB 8-bits channels Gray Scale Clean canvas: white area Crop Rotate canvas: 0.4 mm - cw	gif	550 x 344	9.31 cm x 5.83 cm 100% dither; selective pallett, 256 colours	150	✓	87.93 k 32 sec @ 28.8 kbps	Standard Size 86 kb		oli_001: Medial view of elephant skull, juvenile
oli_002	16-03-07	35.5 x 25.5	610	8424 x 6352 pixels 153.1M	100	24-03-07	jpeg 1.32 M	840 x 548	300	Mode: RGB 8-bits channels Rotate canvas: 0.4 mm - cw Crop Clean canvas: white area	gif	550 x 359	9.31 cm x 6.08 cm 100% dither; selective pallett, 256 colours	150	✓	83.28 k 31 sec @ 28.8 kbps	6.0 Standard		oli_002: Lateral view of elephant skull, juvenile
oli_003	16-03-07	35.5 x 26.0	610	8640 x 6048 pixels 149.5M	100	24-03-07	jpeg 1.55 M	840 x 638	300	Mode: RGB 8-bits channels Crop Clean canvas: white area	gif	550 x 421	9.31 cm x 6.08 cm 100% dither; selective pallett, 256 colours	150	✓	111.8 k 41 sec @ 28.8 kbps	6.0 Standard		oli_003: Caudal view of elephant skull, juvenile
oli_004	16-03-07	41.0 x 28.0	610	8640 x 6656 pixels 164.6 M	100	24-03-07	jpeg 1.55 M	840 x 643	300	Mode: RGB 8-bits channels Crop Clean canvas: white area	gif	550 x 421	9.31 cm x 6.08 cm 100% dither; selective pallett, 256 colours	150	✓	107.5 k 39 sec @ 28.8 kbps	6.0 Standard		oli_004: Ventral view of elephant skull, juvenile
oli_005	16-03-07	35.5 x 26.0	610	6224 x 7728 pixels	100	06-04-07	jpeg	840 x 916	300	Crop Mode: RGB 8-bits channels Clean canvas: white area Rotate 90° CW	gif	550 x 420	9.31 cm x 7.12 cm 100% dither; selective pallett, 256 colours	150	✓	101.2 k 37 sec @ 28.8 kbps	6.0 Standard Size 103 kb		oli_005: Dorsal view of elephant skull, juvenile

Fig.2: Example of the database

3.2.2 The scanning process was the responsibility of the digitization coordinator. The DigiBook10000RGB was used for the digitization of the canvasses. Different camera settings were needed when canvas sizes varies too much. Each canvas was measured and the information typed into the database. Scanning was done between to 600 - 800 dpi in full colour and saved to TIFF format.

The derivation of the images could not be done on the scanner's hard drive and had to be transported to other computers for this purpose. To transport the files between computers we decided to copy the images to an external hard drive. This proved to be a better choice in this specific case because of the size of the original images and the unavailability of space on the Library's servers. The scanning process described above took an estimated 6.5 hours.

3.2.3 The derivation of the original scanned images took approximately 127.5 hours and was done by the digitization coordinator. An average of 2 – 3 hours was spent per sketch (51 sketches x 2.5 hours)

Tasks involved were -

(a) Cleaning and conversion of scanned images
    - Image converted to 300 dpi, gray scale (RGB 8-bits)
    - Cleaning, rotation and cropping (where necessary)
    - Saving in JPEG, 300 dpi for Department and creating thumbnails for database from JPEG's
    - Saving web-ready images in GIF 150 dpi
(b) Creating watermark in MSWord conversion to PDF
(c) Creating final PDF document for uploading to UPSpace repository

3.2.4 For storage the scanned and derivative information was written on DVDs:

(a) 1 x set with original archival (TIFF) images
(b) 3 x sets with JPEG formats of all images
(c) 3 x sets with PDF format of all images plus database

One set of DVDs went to the Department and one set to the Veterinary Science Library. The DVDs containing the original archival (TIFF) images as well as a complete set of JPEG and PDF images are to be stored in the digitization office. Quality control is to be done on a regular basis on these DVD's. A special archival server was acquired after this project had been completed. The archival images are now also stored on that server.

The time spent on the above was 2 hours and was the responsibility of the digitization coordinator.

3.3 Preservation of canvasses

As previously mentioned only basic preservation was done. The canvasses were rubbed by grated PVC free, elastic rubber and cleaned with a lint free cloth and a soft brush. Thereafter each canvas was wrapped in acid free soft tissue paper to prevent dust and mould damage to the canvasses. This has to be done carefully and the airflow between the tissue paper and canvas removed as far as possible. The wrapping was unfortunately done more than once and valuable time wasted in the process.

An average of 45 minutes was spent per sketch. The responsibility for the task was allocated to the digitization coordinator and the total time spent was approximately 38.25 hours.

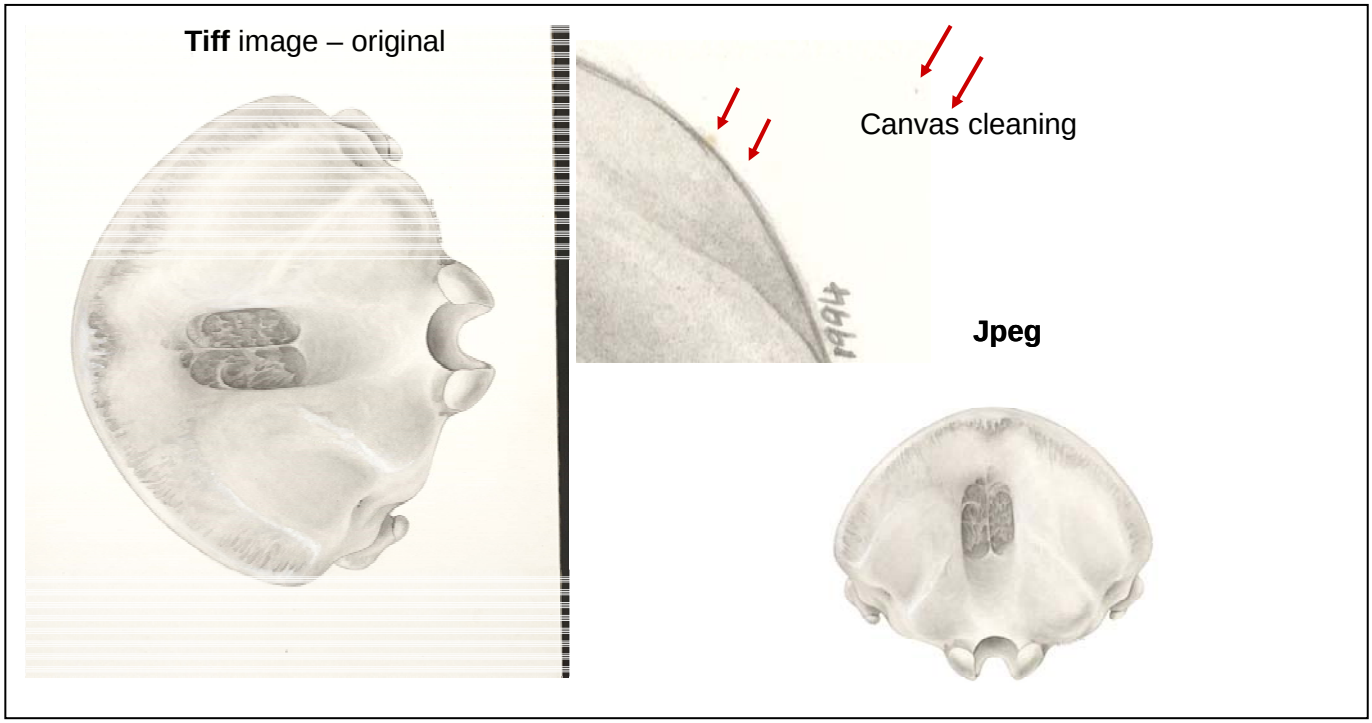


Fig 3: Example of the original scanned image, an enlarged view of a section from the TIFF image and the final display in JPEG format

3.4 Uploading to UPSpace

Before submitting the collection to UPSpace, quality control was done on the images. The first level of submission to UPSpace with baseline metadata, supplied by the Department, was done by the digitization coordinator. This task took an estimated 10 hours.

3.5 Final quality control, metadata and added value

To promote a good digital collection, a final quality control was done by the metadata editor of the collection. This task took a total time of 14 hours to complete. Quality control of the metadata was done on all the elements submitted with the item. Various tasks were part of this process:

(a) The physical sketches were unwrapped a second time to check if baseline metadata was correctly submitted to UPSpace. The baseline metadata was written on the back of each sketch.
(b) An abstract was written for each sketch to further describe the sketches e.g. what the sketches are about, the medium used for the artwork and the actual sizes of the canvasses.
(c) The metadata editor checked for consistency in assigning of the free language keywords to ensure precision of retrieval. A list of free language keywords, used in the collection, was also composed to be uploaded at the end of the project.
(d) Additional keywords were added from CAB (Commonwealth Agricultural Bureaux) Abstracts and Zoological Record thesauri.
(e) The metadata editor also assigned LC (Library of Congress) subject headings to allow a higher degree of accuracy in retrieval. This added value is important for the metadata in a qualified Dublin Core standard that is used in UPSpace.
(f) Correction of spelling mistakes and punctuation was done to make sure the metadata is filed correctly in UPSpace.
(g) Control of consistency between item and bit stream that was uploaded ensured that the information was correctly connected.

After the quality control was done for each item, the item was approved and committed to the archive for uploading in UPSpace and for availability on the web through open access. The metadata editor also checked the database on Excel for inconsistencies. The time involved in this process was included in the databases' time above.

As this uploading of the first part of the collection was a learning process and pilot study, the metadata editor could adapt the metadata template in UPSpace for future submissions of items to the Elephant collection

A last and final quality control was done by the metadata editor and digitization coordinator on the collection in UPSpace as well as the Excel database. Final corrections and additional data were added where needed. The time for this process was included in the metadata uploading and administrative timing assigned for the database

A final discussion was held with the head of the Department and the canvasses and DVD's were handed over to the Department for safekeeping.

4. CONCLUSION

4.1 Lessons learned

Lessons learned from the project will help to implement further refinements in future projects in order to enhance the workflow and delivery time.

(a) The canvasses should be cleaned after the conversion of the scanned images to minimize handling thereof.
(b) Wrapping of the canvasses should be done after the metadata editor's final quality control. The canvasses can be loosely covered with tissue paper and properly wrapped up at this stage.
(c) During the course of the work it was realized that additional elements would have made the template more complete and this would have saved time during the uploading process.
(d) Thumbnails should be created during the conversion process and not afterwards. The thumbnails should rather be done at 150 dpi instead of 100 dpi. These thumbnails should also in future be uploaded to UPSpace in a JPEG bit stream.
(e) The project consists of artwork and detail is of the essence. Extra time should be provided for the conversion process; even the signature had to be cleaned to make it clearer.
(f) Staff should be allocated to the project. Team members had to fit this project into their normal work program.
(g) The PDF software should be set up for a certain view percentage to enable the end user to see the sketch without having to set the zoom effect – additional knowledge is needed for this.
(h) Transportation of the collection between the libraries (main campus where the scanner is situated and the branch Library at the Faculty) was a stumbling block.

4.2 Measurements of success

4.2.1 Project success – (time and budget)
The time frame developed during the project can be used for future projects and estimated costs can now be budgeted for in the project planning of the remainder of the collection.

4.2.2 Learning success – (knowledge gained)
The workflow process is documented and note taken of the strengths and weaknesses of the project. Successes and stumbling blocks were identified and can be adapted for future projects.

4.2.3 Audience success – (marketing and feedback)
Marketing of the digital collection showed an increasing interest in the collection from students and lecturers of the Faculty and also from outside users. Marketing was done during Faculty Day to showcase the collection on UPSpace, a short notice was written for the Faculty Newsletter and a permanent display in the Library form part of the marketing process. The response to the Elephant collection led other personnel to start the uploading of their own collections to the repository.

4.3 Benefits of project

By completing and the documenting the workflow of the project several benefits were gained by the Library and the Department. The project falls in the category of the libraries' strategic plan of becoming an eScholarship environment. The eScholarship program enables the low-cost publication and widespread distribution of the materials resulting from research and teaching at the University of Pretoria. The Elephant collection as part of the eScholarship program is now an open-access infrastructure that offers the Department direct control over the creation and dissemination of the full range of their research output in their field of veterinary anatomy and physiology.

Through the interaction and collaboration with the Department the Library showed that it could play an important role as a facilitator in lifelong learning and conservation of Faculty output. Through digitization of the Elephant collection the preservation of this scarce and unique resource was ensured for future use. Marketing of the collection nationally as well as internationally enhance public knowledge about and information source might otherwise have been lost. The collection also showcases the research output of the Faculty of Veterinary Science and creates a general awareness of our own heritage and indigenous knowledge.

After the completion of the project the team felt that they had gained much experience and knowledge during the project. The project was regarded as an experiment and it was handled as such without wasting unnecessary time and resources. The team became more familiar with the functions of Dspace™ during the project. A template had never before been created in the institutional repository. This function is now widely used in the repository for projects where generic data is involved.

Technical metadata was documented and written in the descriptive metadata fields of Dspace™. This metadata is now compulsory for all future digitization projects uploaded to the UPSpace repository. The database compiled during the project now serves as an example for all new projects and is used as a concept document. Fields in the database are adjusted for new projects according to a project's particular needs. The team completed the project with very few resources. The challenge of making a success of a new Library e-strategy and delivering a project that can be used as an example in the future, kept the team motivated.

The collection can be viewed at https://www.up.ac.za/dspace/handle/2263/1971

Bibliography

TASI: technical advisory service for images 2007, Bristol, viewed 17 October 2007, <http://www.tasi.ac.uk/index.html>

Dublin Core Metadata Initiative: making it easier to find information 2007, viewed 21 November 2007, <http://dublincore.org/>

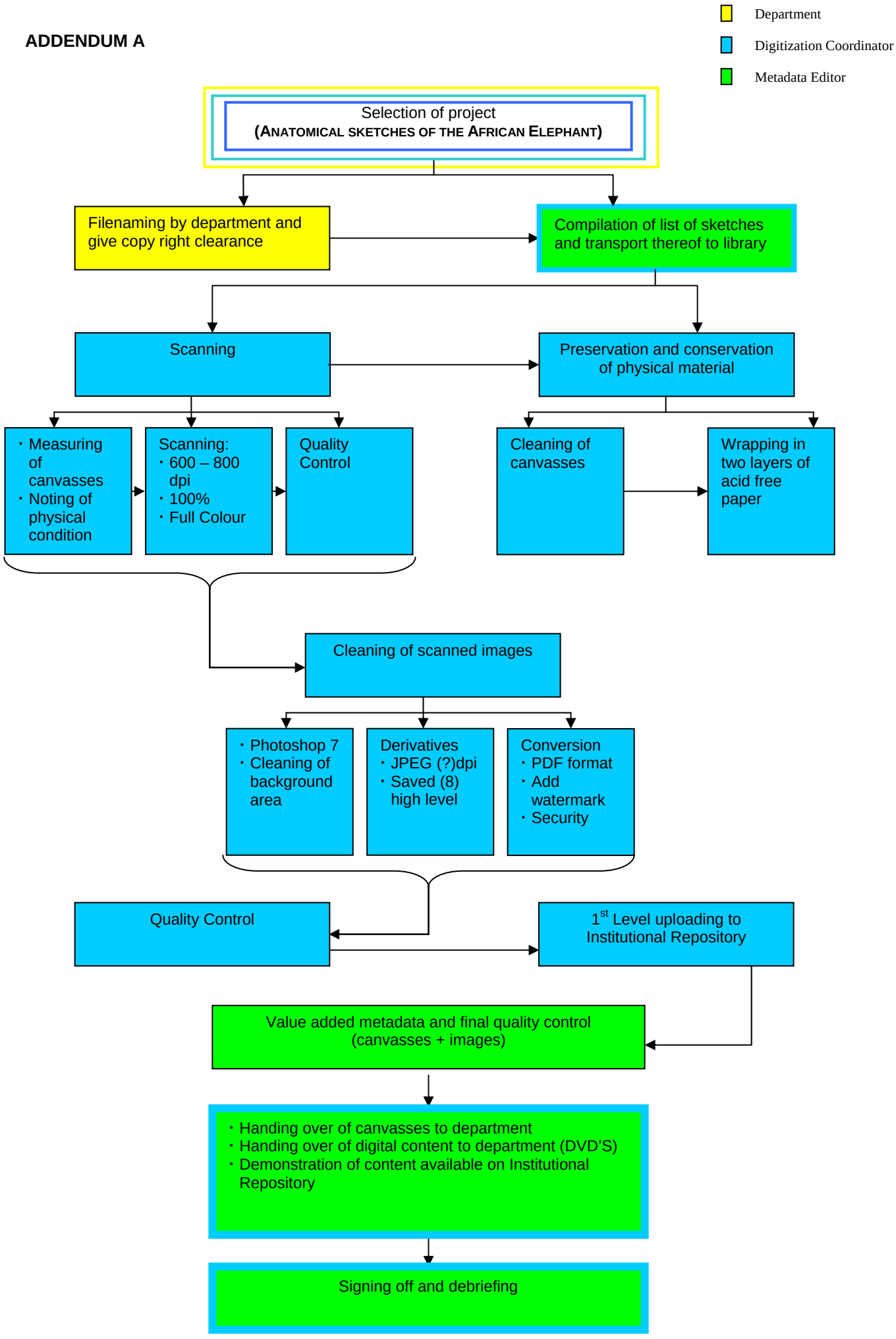
NISO Framework Advisory Group 2nd ed. (2004), *A framework of guidance for building good digital collections*, National Information Standards Organization, Bethesda, viewed 11 September 2007, <http://www.niso.org/framework/framework2.html>

Digital Imaging Project of South Africa (2007), *DISA guidelines for best practice,* viewed 1 November 2007, <http://aboutdisa.ukzn.ac.za/Guidelines/guideindex.html>

Tansley, R, Bass, M, Branschofsky, M, Carpenter, G, McClellan, G, Stuve, D (2007), *DSpace system documentation: contents for DSpace Version 1.4.2,* viewed 6 June 2007, <http://www.dspace.org/index.php?option=com_content&task=view&id=151&Itemid=116>

King's Digital Consultancy Services (2005), *Digital futures,* London, website viewed 23 November 2007, <http://www.digitalconsultancy.net/digifutures/>

ADDENDUM A


Department
Digitization Coordinator
Metadata Editor
Selection of project
(ANATOMICAL SKETCHES OF THE AFRICAN ELEPHANT)
Filenaming by department and give copy right clearance
Compilation of list of sketches and transport thereof to library
Scanning
Preservation and conservation of physical material
· Measuring of canvasses
· Noting of physical condition
Scanning:
· 600 – 800 dpi
· 100%
· Full Colour
Quality Control
Cleaning of canvasses
Wrapping in two layers of acid free paper
Cleaning of scanned images
· Photoshop 7
· Cleaning of background area
Derivatives
· JPEG (?)dpi
· Saved (8) high level
Conversion
· PDF format
· Add watermark
· Security
Quality Control
1st Level uploading to Institutional Repository
Value added metadata and final quality control (canvasses + images)
· Handing over of canvasses to department
· Handing over of digital content to department (DVD'S)
· Demonstration of content available on Institutional Repository
Signing off and debriefing