MY DOCUMENT

CURATION FOR PARTICIPATION

An
EIGHT-STEP GUIDE
to curating open scholarly content

Kyle Rother, Sarah Goodier, Lena Nyahodza
and Laura Czerniewicz

CONTENTS

CONTENTS

INTRODUCTION

WHO IS THIS GUIDE FOR?*

This guide is for anyone who finds themselves curating scholarly and educational content at UCT, even if you don't know that it is called curation. You may be a postgraduate student tasked to put a professor's research output online, or you may be a communications officer whose tasks have now extended to curating content. You may be a webmaster who finds that you are asked to upload all kinds of resources. You may be employed part-time, or on an ad-hoc basis. You probably won't have a qualification in digital curation or library science. You will probably have little to no experience in this type of work. But you find yourself in this position. So, what needs to be done?

*Please note that since this guide was conceived mainly for use within UCT, many of the resources and links to subscription services may not work if you are accessing them from outside of the UCT network.

CURATION FOR PARTICIPATION

What is curation? And why is participation so central?

Putting content online is not merely a bureaucratic or administrative matter. Certainly it needs to be done carefully, and requires a meticulous approach, but the reasons for doing it are not trivial: they are central to the scholarly enterprise. Why? The reasons are twofold: preservation and discoverability.

Preservation. Scholars need a record of their work. In a digital age, ironically, scholars may lose control of their own work, with their output stored in the cloud, handed over to publishers, lost on discs or mislaid on old computers. Their institutions can play an important role in preserving scholarly content safely in a secure, central and trusted place, for future use.

Discoverability. How scholarly content is curated is absolutely critical to the chances of it being found, shared and cited. Studies make it clear that online searches are the main way that content is found;[1] it is rare for a URL to be used or even known.

In the digital world in which scholarship takes place, having content found online is one of the ways that scholars become part of academic and disciplinary networks. This is why good curation really matters.

What does 'curation' mean? Curation has become a popularly used term, often referring to the selection and filtering of content through social media tools such as Scoopit. The focus of this guide is not so much on the selection aspects of curation, and there is not likely to be much filtering. It is about all the rest of what needs to be done.

This guide is about open scholarly content, which means it is about making the actual content available freely and legally and not just a description of it. It applies to all types of content in a range of formats. The objective is the description and display

of content, either individually or as a collection. How to do this properly is the focus of the guide.

ABOUT THE GUIDE

The guide takes you through 8 steps:

Step 1: Identifying content
Step 2: Sourcing content
Step 3: The legal stuff
Step 4: Digitising content
Step 5: Describing the content
Step 6: Uploading the content
Step 7: Telling people about the content and tracking impact
Step 8: Curation as an ongoing process

Each step may include reference to specific resources which are intended to offer direction in your work, as well as various people who may be able to offer more guidance. At the end of the guide you will find a glossary of important terms you will need to know, as well as a list of all the policies cited and some others of general interest and applicability.

STEP 1

IDENTIFYING CONTENT

How much content will you be curating? Which content types will you be working with? These decisions are unlikely to be yours. You will need to work with whoever is employing you to make your brief very clear. You will need answers to the following questions so that you don't drown!

WHICH CONTENT?

Although this guide is intended for working with scholarly content, that term is somewhat ambiguous and difficult to define. Is it output in the form of peer-reviewed journal or monograph publications? Draft submissions on policy or legislation which is informed by research activities? Resources which are generated as part of the teaching process? All of these constitute outputs of the scholarly process, and should be treated equally.

You will be guided by the brief you get from your employer, or if they are themselves unsure, you should work together to decide on the scope, bearing in mind the intended outcome of the curatorial exercise. What is the focus of the unit or department, and what are the main types of outputs it produces? Who is most likely to use them, and for what? These are some questions to ask when prioritising the order of work. Sometimes the creation of teaching resources forms a side activity, and is more appropriately handled by someone else. Remember, your job is to curate, not create!

WHICH TIME PERIOD?

For administrative purposes, it is best to limit yourself to outputs produced during the course of work at UCT. You may be implementing UCT's Open Access Policy which has the specific purpose of making available content produced at or by UCT. This also reduces the complexity of shared copyright and licensing, but more about that later. So even if an academic had a long and illustrious career before coming to UCT, you should be quite strict about only working with content they have produced since arriving at the university (unless of course your brief and the budget allows otherwise). Once they have seen the benefit of the curatorial exercise you perform on this content, they may be motivated to do the same kind of thing with the rest of their output!

Some academics have decades of employment at UCT and you and your employer will have to make choices. If you want to curate the work of a whole group of people, you may decide on a shorter time period to deal with at first. And of course, the further back in time you go the less likely it will be that the content will be in a digital format, which adds another layer of complexity to your task.

WHICH PEOPLE?

Obviously you will only be working with UCT staff, however, even within this relatively small community (in terms of the global network of scholars) there is still a variety of different people. You may be working with a seasoned veteran researcher with an extensive body of work, one who has no interest in curation; or you may be working with a relative newcomer, with comparatively little to share, but who really understands the value of openness, and is eager to participate. You may encounter a good deal of resistance, and lack of helpfulness. Once the value of the curation becomes clear and the academics see the results, they are likely to become enthusiastic. It is a fair assumption that all academics want their work to be found and to be read — both prerequisites for citations!

WHO TO ASK

- Your employer or Advisory Group
- UCT's Library and Information Studies Centre offers training in Digital Curation, as well as courses in Research Data Management
- Other people involved in similar exercises at UCT (you may be surprised at how many there are): since this is a new field, everyone is learning from everyone else, and it is good to have a Community of Practice within which to share your tips and challenges

STEP 2

SOURCING CONTENT

Academics are busy people — finding time to meet with them can be difficult, so it may be easiest to start creating a list of outputs on your own. Online databases are a great source of publication information.

LOOKING ON THE WEB

A basic **Google** search (for a broad range of outputs) and a **Google Scholar** search (for the more scholarly works) will help you add items (and objects) to your curation list. Searching for the academic by name can be useful if their name is quite unique; however, you might also want to add 'UCT' to your query to narrow down the search and get the most relevant results.

Have a look for their researcher profiles (e.g. on the department website as well as sites like **Academia.edu** and **ResearchGate**) — many academics have at least one profile and list their publications on it. Some also include their publications on professional sites like **LinkedIn**. And if you are really lucky, you might find academics with a Google Scholar profile; look up their names on Google Scholar to find out.

Magnifying glass image by Tall Chris available under a CC-BY licence.

Searching **bibliographic databases** and indexing services (e.g. **Scopus**, **Web of Knowledge**) can also help you locate items by an academic. These databases are available via **UCT Libraries** subscriptions to these services and in many of these you can search by author and affiliation.

You can also search on the web by particular formats and types of content. Use **Google Advanced Search** and choose from the 'file type:' options to search for PDFs, presentations, etc.

LOOKING AT UCT

The latest **UCT Annual Research Report** is a good source for research outputs in the narrower sense — see the Research Report guide for what is included. Departmental/ project websites can also be a good source of information regarding publication lists as well as other types of outputs.

UCT uses a web-based information management system — the Integrated Research Management Applications (IRMA) — for collecting data for the institution's Annual Research Report and the publication count. You may be able to extract data on accredited journal articles and other outputs from this system.

Some academics at UCT make available some of their teaching resources on the **OpenUCT repository** (http://open.uct.ac.za). There is a fine line between a teaching resource and an academic resource, so it is worth checking here too.

LOOKING ELSEWHERE

Many academics contribute to national news through opinion pieces or interviews, and you may be able to find articles and contributions through newspaper archives, many of which are now online.

Some academic disciplines rely more heavily on publication in monographs, i.e. books and book chapters. It may be useful to do a search of the **UCT Libraries catalogue**, and also more widely on **SABINET**, the South African catalogue which contains bibliographic information for all books published in South Africa.

Research institutes at or linked to UCT may have their own collections and lists of publications. It may be useful to look at their websites.

CONTACTING ACADEMICS

Once you have a list of outputs, you can contact the relevant academic/s. Starting with an email/phone call requesting a meeting to discuss the content curation activities will help to get the ball rolling. You can ask the academic to have a look at the list you have put together and add any outputs that you may have missed.

Some of the content will be easily obtainable from the online and database searching discussed above. Other content may be sourced from the academic that you're curating.

RE-USE

Obviously, it would not be in your best interests to re-capture and re-file content that already exists in an open repository elsewhere. Open journals, for example, already have authorised copies published on their websites or through open content aggregators, together with all the required metadata. Do not reinvent the wheel: rather link to a resource that already exists somewhere, using metadata harvesting and the DOI standard (if the resources have DOI numbers associated with them).

WHO TO ASK

To inquire about training on IRMA, you can contact **rea-irma@uct.ac.za**. Ask UCT librarians for help in searching the catalogues and databases.

STEP 3

THE LEGAL STUFF

Before you do anything else, you have to know what the legal status of the content is, what you can legally share, and/or on what basis the scholar wishes to share their content.

COPYRIGHT

The scholarly content you are dealing with is likely to be the product of a single person or a small group, in the form of a published research article or paper, or a chapter published in a book. If the article has been published in a commercial, peer-reviewed journal, copyright will often rest with the publisher of that journal. In this case, you will have to check what the conditions are for sharing content from that journal (see 'Acceptable versions' on **page 9**).

Copyright in content that is a product of work conducted at UCT, and published under the auspices of UCT, will usually rest with the author/s or with the unit or department in which they work. This means you will have to get permission from the author/s or their department, but presumably you will have done this as part of the process of sourcing the content. Even so, you will have to explicitly state that you have permission to be sharing the content, and there are a number of options for doing this (see 'Choosing a licence' on **page 11**).

If the work was partly or wholly funded by extra-institutional sources, then copyright may rest jointly with the author/s or their department (as above) and with the external funder. This is a slightly more complicated situation, as permission will have to be sought from the funder, and sometimes these bodies have quite specific conditions around the sharing of content. However, these should be established by any contracts agreed at the outset of the research process, and stipulated as such in writing.

ACCEPTABLE VERSIONS

In the case of commercially published content, copyright is a major issue. Most publishers are, unfortunately, fairly conservative about allowing the sharing of content

in which they hold copyright. This is usually in the form of a final, edited and laid-out published version of the work as it appears in a journal or monograph. Some publishers allow for fully open access publication, in which case they will make their own final versions available, or allow them to be shared.

Many other publishers also make provision for authors to share their work in one form or another after a set embargo period has elapsed. These will usually be the author's own final draft or manuscript as it stands before (pre-print) or after (post-print) peer-review, but before final edits and publication. If it is your good fortune to have been provided with these, then your job is slightly simpler. If, however, the academics with whom you are working do not have pre-prints or post-prints to share, you can try encourage them to cultivate the habit of keeping versions of their work. This is a good practice both from an academic as well as a curatorial perspective, as it establishes a legacy of work for an individual or for a department. But more on this later.

If you do have pre-prints or post-prints, then you can often share these quite openly, sometimes after an embargo period stipulated by the publisher, and provided that you link the copy you have shared to the publisher's own final version. Much more information on conditions for sharing can be found on the **SHERPA/ROMEO** site. Here you can search for a journal title or ISSN, and check the open access provisions of the publisher of that journal.

THIRD PARTY CONTENT

If the content contains copyrighted materials or resources which have been sourced from a third party, then either permission will have to be sought to share those materials or resources, or they will have to be replaced with alternatives which are legally shareable. An example of this might be if you want to share a lecture or presentation slide show which contains pictures that have a copyright designation to someone who is not the person from whom you sourced the content.

Sometimes a lecturer, when putting together a presentation, will take images or other resources (tables, diagrams, charts, etc.) from the web for illustrative purposes without checking whether they are in copyright at all, are available under open licences or are in the public domain. This is fine, as long as the lecturer is only using the images in that presentation, and for instructional or educational purposes and not sharing them online even just internally at UCT (e.g. on **Vula**). However, if the academic then wants to share that presentation with others, there is the possibility of copyright infringement on those images or resources if they are not legally open resources. If the copyright holder is contactable, and is willing to share the material, then they can give you permission to do so — best to get this in writing for your records. If, however, they are not willing to share, you will need to remove their material from the content. In this case you may need to substitute that material, either with appropriate replacements which are licensed for sharing (more about licensing on **page 12**), or you can try to reproduce the originals. This should be done in consultation with the person who created the content, as they may be able to give you input on what is essential for their purposes, and perhaps save you a lot of effort.

There are a number of ways to source openly shareable content, from filtered Google searches, to whole sites devoted solely to open content. See the 'Useful resources' section on **page 12**.

CHOOSING A LICENCE

Once you have established the copyright status of the content, and whether you can share it, there are a number of options for releasing it. All rights reserved copyright may apply, and this will continue unaltered. Alternatively the author may wish to apply an open licence to the content, stipulating the conditions for its usage.

Licensing

In the case of fully copyrighted material, the content can still be open to be viewed and used, but the conditions on re-use are highly limited. For instance, a copyrighted work can only be used fully with permission from the copyright holder. This means that anytime someone wants to re-use a piece of research beyond the limits of what fair dealing permits, they will need to contact the author and/or copyright holder to get their express consent. This is a cumbersome procedure. Preferably, the copyright can be maintained, but conditions for use and re-use can be stipulated by applying a licence to specify permitted uses.

The most established and popular kinds of open licence are the Creative Commons (CC) licences (see https://creativecommons.org/). These are open licences which can be applied to any kind of work, and indicate explicitly the conditions under which copyrighted material can be used. Note that CC licences are not a substitute or a replacement for copyright; they are actually based on copyright law. They specify licence permissions of varying kinds in advance, making it clear what the conditions of use are. Copyright obtains automatically in any work. The only way it can be waived is if an author expressly decides to do so and states that as the copyright holder they do not wish to retain any of their rights; they would then apply a CC-0 or 'CC-zero' licence.

A copyrighted work can have a CC licence that extends certain rights to the user, but leaves others exclusively to the copyright holder. For example, a very common CC licence is the CC-BY-SA, which means that anyone is free to use and re-use a piece of work, as long as the author (or copyright holder) is acknowledged (BY standing for 'attribution'), and any resultant works are released under a similar licence themselves (SA standing for 'share alike'). Thus, permission has been granted to quote from, adapt or modify a work, as long as the source is acknowledged, and the resultant work is also shared. This is not so far from the existing practices of academia in any case.

WHO TO ASK

The **IP Unit** in the Faculty of Law here at UCT is home to experts in open licences as well as copyright generally.

The University's **Research Contracts and Intellectual Property Services** (RCIPS) is an office dedicated to handling copyright clearance, trademark and patent issues, and they can provide much more guidance on UCT's IP policies and obtaining clearance to share work.

USEFUL RESOURCES/LINKS

The OpenUCT Initiative provides a useful resource on '**Finding Open Stuff**' if you need to replace copyrighted material with legally open material.

The **Creative Commons** website is an excellent resource both for you and for academics who want to know more about open licences and how to apply them.

Find here a brief introduction to **copyright** and **Creative Commons** concepts.

For open licensed images: **Wikimedia Commons** (images normally carry BY-SA Creative Commons licences).

STEP 4

DIGITISING CONTENT

Once you have collected your content, and established its copyright status, you need to prepare it for sharing. The process involved will vary depending on what format and condition the content is in.

DIGITISING TEXTS

As well as different kinds of content, you may be asked to curate legacy documents which only exist in hard copy. This will involve an additional layer of work to digitise these documents. If you have access to a scanner, digitising should be a fairly straightforward exercise.

You have two options in digitising documents: you can either create an image file of the document, meaning that the digital copy will simply be a static picture of the original; alternatively, you can create a rich text copy which you will be able to edit and manipulate, but this will involve the use of optical character recognition (OCR) software, which not everyone has access to.
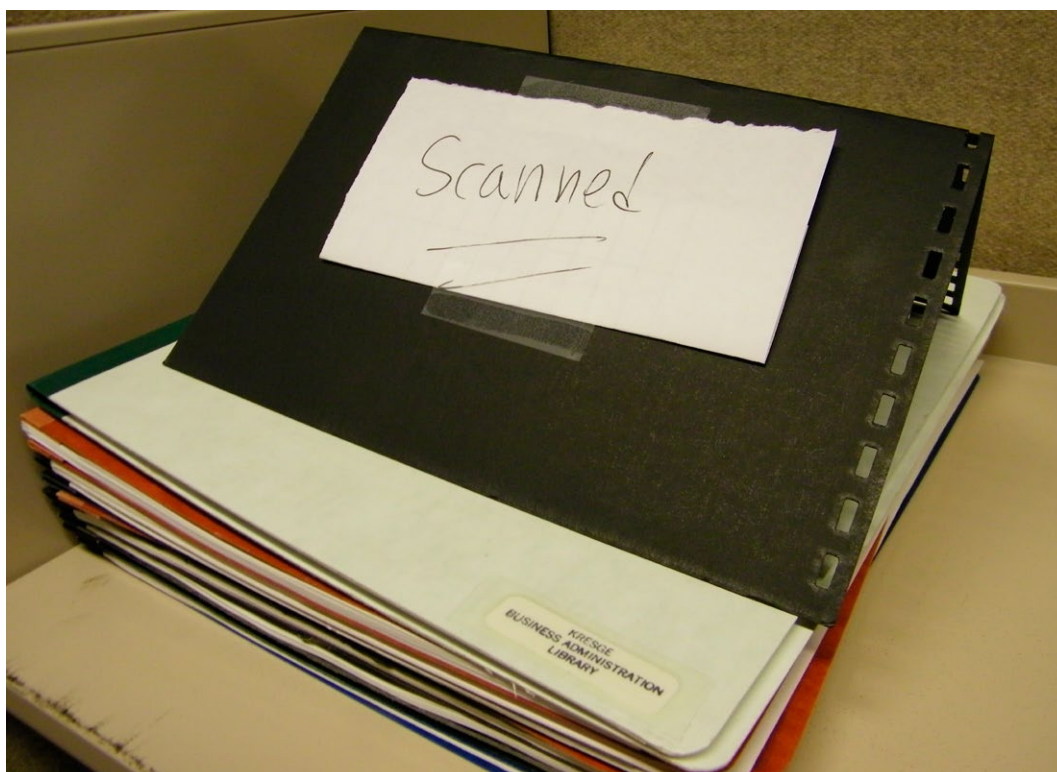
This pile done! image by Corey Seeman available under a CC-BY-NC-SA licence.

If you are simply creating an image file of a document, an accepted practice is to scan it in grayscale at 300dpi. This is preferable to scanning in bitonal black-and-white, as the resulting images will appear clearer, and if the quality of the original document is poor, bitonal scanning may not pick up all details. These options can be set in the scanner's software prior to scanning. You will also usually have several options of file format to save the image into, for instance .tiff, .jpg, .png, and .pdf. Accepted best practice is to use the TIFF format for archival masters of images, but this usually applies more to photographs. For documents, you should consider using JPEG or PDF.

OCR software works either by scanning the original document directly into the software, which will generate a rich text file, or by running an existing image file through software to convert it into a rich text format. Rich text documents are much more flexible, and allow users to manipulate the text (search, copy and paste, etc.). They also allow search engines to crawl through the actual document and index its contents. For this reason, rich text scans are preferable and recommended, however, as mentioned, not everyone has access to OCR software, and although they have become much more stable in recent years, OCR programs are notoriously difficult to work with, especially if the original documents are in poor condition.

Should you wish to interact with or manipulate the contents of a scanned PDF, a variety of software packages exist to convert PDFs to Word or rich text documents. Some, such as Adobe, need a licence (which the department you are working for might have); others allow a certain number of documents to be converted as trial versions; others are free software. Sadly, the technical conversion is the least of your challenges.

After conversion you will need to proof read the converted documents against the original, checking if:

- there is any text missing
- headings and paragraphs are still properly positioned
- all tables and figures are converted and properly positioned
- tables and figures are separate and not copied on top of each other
- text from tables is not presented as plain text
- content in tables and figures has not been compromised (no missing text)
- there are any corrupted letters or words
- text spacing has been correctly retained

Tables might not be in the correct position, but located at the end of the document; it is therefore important to look at the entire document before starting the cleaning process.

Cleaning

1. Start by checking whether the integrity of text, tables and figures has been retained. If tables and figures are displaced, restore them by copying and pasting, and avoid cutting if possible. Copying and pasting works out better because if the pasting of any element fails, content may be lost or corrupted in the process of trying to restore it. If figures and tables are corrupted and cannot be manipulated, consider copying these individually from the original document, or try converting the document using different software.

2. Take note of corrupted letters and symbols that are likely to feature in the text. These may be difficult to notice and you need to scrutinise the document thoroughly, otherwise you might leave content in that will compromise the integrity of the final product. See an example of corrupted text on **page 15**:

A PARAGRAPH WITH CORRUPTED CHARACTERS

7KH_HYDOXDWLRQ_RI_WKH_QHZ_/L(3_LQ_/LPSRSR_3URYLQFH_
As already mentioned, LiEP was adopted by government in July 1997 to achieve the following aims: (1) to promote full participation in society and the economy through equitable and meaningful access to education; (2) to pursue the language policy most supportive of general conceptual growth amongst learners, and hence to establish additive multilingualism as an approach to language in education; (3) to promote and **GHYHORS_DOO_RI¿FLDO_ ODQJXDJHV_____WR_VXSSRUW_WKH_WHDFKLQJ_DQG_OHDUQLQJ_RI_ DOO_RWKHU_**

THE CORRECT VERSION OF THE PARAGRAPH ABOVE

The evaluation of the new LiEP in Limpopo Province
As already mentioned, LiEP was adopted by government in July 1997 to achieve the following aims: (1) to promote full participation in society and the economy through equitable and meaningful access to education; (2) to pursue the language policy most supportive of general conceptual growth amongst learners, and hence to establish additive multilingualism as an approach to language in education; (3) to promote and develop all official languages; (4) to support the learning and teaching of all other

---

3. Clean every sentence by restoring proper spacing, since most Word documents converted from PDFs split sentences half way, leaving wider gaps at the edge of each sentence.

4. Insert page numbers.

5. Save the document in the required format.

6. If are you considering sharing the final product of the work broadly or online, you may want to insert a cover page where you add the relevant metadata along with a statement on the provenance of the document.

WHO TO ASK

If you are unsure of how best to handle a specific type of content, you can seek guidance from the **Digitisation Unit**. This is a special division within UCT Libraries responsible for digitising the university's archival collections. They also provide a digitisation service to faculty and staff of the university.

USEFUL RESOURCES

Scanning Your Personal Collections, http://www.digitalpreservation.gov/personalarchiving/documents/scanning_collections.pdf

STEP 5

DESCRIBING THE CONTENT

It is almost impossible to emphasise strongly enough how important it is to properly describe the content you are sharing. This will mean the difference between simply making your content available online, and actually allowing it to be found and used.
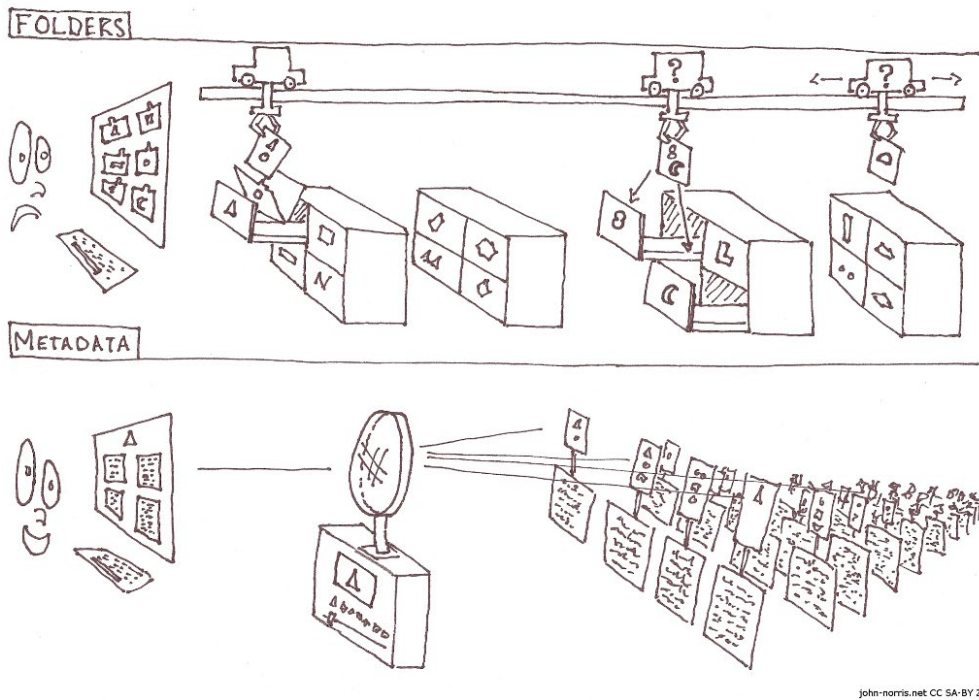
METADATA

As a curator, your most powerful tool is metadata. Metadata is essentially 'data about data'; in other words, it is information that does not actually form part of the content itself, but describes the content. Metadata will be contained in a metadata record, which is attached to the file it describes, and can be used by search engines to gain some information about the file's contents. This metadata record will also usually be displayed in some kind of repository, and viewed by users to make a decision about whether an item might be of interest or value to them. For this reason, it is essential to ensure that the metadata is clear and concise, consistent and intelligible.

There are a number of standards for generating metadata that are accepted internationally, probably the most common of these being the **Dublin Core Metadata Initiative** (DCMI), or **Dublin Core** for short. Dublin Core separates metadata into a number of discrete elements or fields which represent certain key aspects of a document, image or piece of data, such as its title (the 'title' element), the name of the person who created it (the 'creator' field), and the date it was created or published (the 'date' field). As can be seen, these are — for the most part — fairly self-explanatory, and that is the beauty of Dublin Core: it is a standard that is designed to be used by almost anyone, with almost any level of experience or expertise, to describe almost anything. Furthermore, these elements can be used *ad libitum*, and combined into records of varying levels of complexity or depth; Dublin Core is not prescriptive.

UCT Information and Communication Technology Services (ICTS) and UCT Libraries have jointly formed a Metadata Working Group (MWG) that has developed a set of guidelines on generating metadata for different types of content (see 'Who to ask' on **page 20**). These guidelines (based on DCMI) specify the metadata elements or fields that are considered to be mandatory, or a minimum for purposes of identification, at UCT.

Folders vs Metadata image by John Norris available under a CC-BY-SA licence.

FILE NAMES

File naming is an often overlooked but important part of the curation process, as it allows you to quickly identify the documents you are working with. If files are named in a consistent manner, you will be able to keep track of large batches of content. A good idea is to include the name of the department or unit with which you are working in the file name in some way. For example, if you were working with a batch of annual position papers produced between 2006 and 2008 by the Academic Development Programme (ADP) within the Centre for Higher Education Development (CHED), you might name them as follows:

> CHED_ADP_positionpaper_2006.pdf
> CHED_ADP_positionpaper_2007.pdf
> CHED_ADP_positionpaper_2008.pdf, etc.

This gives each file a context, and gives you some idea of what each file is without having to actually open the document and read the title. Note that the file extension (e.g. .pdf, .jpg, .tiff, .xls, .doc, etc.) should appear at the end.

Try to use only lowercase, unless it's for an acronym. Do not use the following characters in file names:

<	(less than)	**>**	(greater than)
:	(colon)	**"**	(double quote)
/	(forward slash)	****	(backslash)
|	(vertical bar or pipe)	**?**	(question mark)
*****	(asterisk)		

Also try to avoid using parentheses and spaces in file names. Try using underscores (_) and/or plus symbols (+) instead.

METADATA FOR DIFFERENT CONTENT TYPES

Different content types may have different metadata requirements. For instance, licensing provisions for journal articles and book chapters may be different from those of open educational resources. The UCT MWG has developed templates for different content types which make it much easier for you to prepare content for deposit into the UCT default repository, and which can be used for all content types. See the 'Who to ask' section on **page 20**.

WORKING WITH BATCHES

If you have been tasked with curating the content of an entire unit or department, you will likely be working with a large amount of material. If you don't have a clear process or structure, it is easy to lose track of what you are doing. It is a good idea to keep content grouped together by type, and this is where a consistent file naming protocol will prove useful.

A file naming protocol goes hand-in-hand with a folder structure. Obviously, content of the same type can be grouped together into folders. For example, say you're working on the ADP position paper series from above. Your files all contain the 'positionpaper' tag, so they will all go together in a folder named 'position_papers'. That folder will go with all the other content type folders (e.g. 'annual_reports', 'policy_briefs', etc.) in a parent folder named 'ADP'. You also might not only be working with the ADP, so that folder would go in a further parent folder (perhaps named 'CHED') with all the other named unit folders (e.g. 'HAESDU', 'COL', etc.). Eventually you will end up with a folder structure that looks something like this:

```
CHED
  ↳ COL
    HAESDU
    ADP
      ↳ annual_reports
        policy_briefs
        position_papers
          ↳ CHED_ADP_positionpaper_2006.pdf
            CHED_ADP_positionpaper_2007.pdf
            CHED_ADP_positionpaper_2008.pdf, etc.
```

You can see how quickly a folder structure gathers layers and complexity, so try not to bury things too deeply, as this will make for cumbersome path names, which may make uploading and linking content complicated. From the above example, the path name for the 2006 position paper would be 'CHED\ADP\position_papers\CHED_ADP_positionpaper_2006.pdf'. This is already four layers, which is three folders plus the file name, and you should aim to keep your folder structure no more than three folders deep. It may also be a good idea to develop a 'finding aid', a document which outlines your folder structure and file naming conventions. This will allow you to get a snapshot of the scope and structure of the collection.

If you are describing and applying metadata to large batches of content, a good strategy is to work in a spreadsheet at first. It is good practice to work with one spreadsheet for each type of content, and keep this in the folder with the files. The

items would be arranged into rows, and the metadata elements would be organised into columns. Thus, for example, your spreadsheet might look something like this:

FILE NAME	TITLE	CREATOR	DATE
CHED_ADP_ positionpaper_2006.pdf	Example title No. 1	Academic Development Programme	2006
CHED_ADP_ positionpaper_2007.pdf	Example title No. 2	Academic Development Programme	2007
CHED_ADP_ positionpaper_2008.pdf	Example title No. 3	Academic Development Programme	2008

You can see the value of working in a spreadsheet, as much of the metadata for many of the items will be identical, so you can simply pull values into multiple cells. There are also benefits when it comes to uploading content, and it is the best way of maintaining a consistent record of a unit or department's output, but more on that later.

SUBJECT HEADINGS, CONTROLLED VOCABULARIES AND KEYWORD LISTS

Depending on the academic field in which you are working, there are generally subject headings and keyword lists that are useful in providing you with a predefined set of terms that are understood by popular convention in the discipline. Many disciplines have their own open repositories, and it is worthwhile finding out if they use existing controlled vocabularies and any open metadata standards.

METADATA FOR HARVESTING

One of the most powerful applications for metadata is the ability to 'harvest' between repositories, thus dramatically expanding the potential reach of content, as well as preventing re-work and re-capture by storing an item only once in one repository, but making it accessible through many repositories anywhere in the world. Making your metadata harvesting-friendly helps you to get the most out of the content, which is another good reason for looking up relevant disciplinary controlled vocabularies and sticking to standards like Dublin Core.

WHO TO ASK

The Metadata Working Group is responsible for implementing the Metadata and Information Architecture Policy which is jointly owned by UCT Libraries and ICTS. You can access the **MWG guidelines**.

STEP 6

UPLOADING THE CONTENT

You've arrived at possibly the most important part of curating to participate: sharing the stuff. As essential as all the preceding steps are to the process, if it's not online, no one can get to it.

WHERE TO PUT STUFF

UCT supports the open sharing of content generated by the work of the university community, in both policy and practice. There are a number of portals for doing this, and any appropriate portal is acceptable. The default repository, **OpenUCT** has been designed to accommodate a wide range of content types and formats, and to expose these as widely as possible. It serves several functions: it preserves and records; it is a 'shop front' where anyone can see and access what the university is producing; and it is designed for its content to be harvested and discoverable by search engines, aiming to ensure that the content reaches the widest possible global audience.

It may be that your unit or department has a website of their own, where they would prefer to host their content. This is fine, and the work that you have done will still prove valuable to the uploading process. The website may have an existing 'publications' section, or you may have to create a new one. Depending on the scope of your activities and experience, you may have to consult on this with the webmaster of the site.

Look out for the trees image by Libby Levi for opensource.com available under a CC-BY-SA licence.

Even when you are sharing content on a unit or department's website, the ideal would still be to describe and upload it to the university's repository, OpenUCT, and then link from there to the website. This is desirable, as the OpenUCT repository is

intended to fulfil an archival function, so the content hosted in it will be preserved and maintained appropriately. This means that should the unit or department you are working with ever dissolve, or the website be removed, their content will still be accessible. It is also practically useful as it saves space available for hosting content on the departmental website. Finally, the repository system gives each piece of content a permanent identifier (a digital object identifier, or DOI), which is incredibly useful in many ways, especially for tracking purposes.

There are also a number of disciplinary repositories such as **arXiv** (pronounced 'archive'), which is specifically for pre-prints of scientific articles in subjects such as mathematics and physics; and Research Papers in Economics (**RePEc**) for working papers and pre-prints on economics and related subjects.

Depending on the discipline you are working in, you will need to discuss with your employer whether to upload to the disciplinary repository and then link to OpenUCT as required by UCT's open access policy. Your employer may not necessarily be aware of the OpenUCT repository, and may have allocated budget for the creation of a new repository from scratch. If this is the case, we recommend that you (gently) remind them that there is already a content repository at UCT that fills most requirements, and that any budget they have for archiving and curation would be better allocated to content sourcing, metadata allocation and ingestion than to technology that already exists at UCT.

GETTING IT OUT THERE

OpenUCT has a self-deposit interface which guides users through the process of uploading their content. It is intended to be a self-managed process, and is designed to be as user-friendly as possible. If you are working with large batches of content, however, it may be advisable to speak to a technical advisor at UCT Libraries about doing a batch upload of content. This is where all the work you put into organising your content — naming files, developing a folder structure, and generating metadata spreadsheets — will pay dividends, because if all of this has been done well, the upload process should be painless.

Other repositories will have similar upload processes; however, you will have to consult their relevant user documentation for specific instructions.

Uploading content directly to a website is slightly more complicated and, as mentioned, you would probably need to work with the webmaster to do this. However, all of the curatorial work up to this point will make the process much simpler. You can then link this content to records in OpenUCT.

WHO TO ASK

It would probably be best to discuss the most appropriate portal for sharing content with the head of the unit or department you are working with, as there will usually be strategic or political motives behind this decision. But be sure to familiarise yourself with all of the options, and their various pros and cons, so that you can best advise. Once a decision has been made, you can either go ahead and start uploading, or speak to someone in the relevant organisation, such as UCT Libraries or ICTS.

USEFUL RESOURCES

The subject repositories named above are only a few of the many open access repositories around the world. The Directory of Open Access Repositories (**OpenDOAR**) is a centralised, worldwide database of these.

STEP 7

TELLING PEOPLE ABOUT THE CONTENT AND TRACKING IMPACT

Having content up is like having a party without invitations — if no one knows about it, they're not likely to just show up. But don't worry: this bit is easy and it can be fun. When you start showing people where to find your content online, you may well be surprised by whom it can reach, who reaches out to you because of it, and where it starts to make an impact.

MAXIMISING DISCOVERABILITY

Most open access repositories should make their content available for indexing by search engines such as Yahoo and Google Scholar. Through your curation process and your consistent use of metadata, you've already helped to ensure the indexing of your content in search engines once it is added to a repository. For example, by adding relevant keywords, when people search for resources in your specified subject areas, your content has a better chance of appearing higher in the search results — this is part of a process called search engine optimisation (SEO). As few people look past the first page of search results,[2] the higher up your results appear, the more likely they are to be clicked on and possibly read and cited.

TELLING PEOPLE ABOUT THE CONTENT

With the amount of content available online increasing all the time, hoping that someone will come across your content on their own while browsing the internet might not be the best strategy. Getting the message out about your content being online, what it is about and where to find it as well as sharing it within your community and beyond can also help it to be found and read.
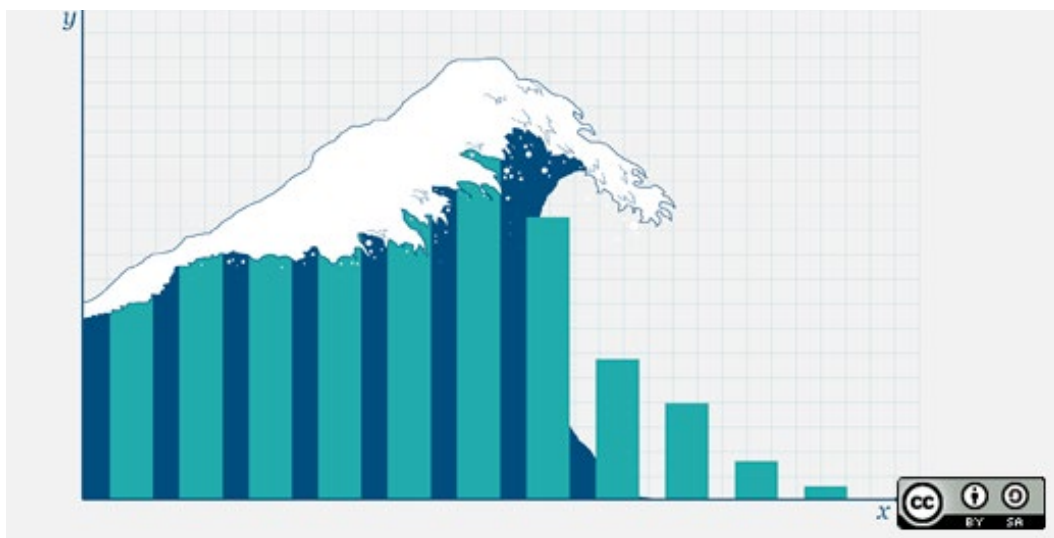
Academics or the communication officer should write short articles or blog about the latest paper(s) you've put up online and circulate the repository links to the latest openly available outputs on the departmental site and in e-Newsletters. Linking the papers in the repository to academic networking sites like **academia.edu** and **researchgate.net** can also help to get the outputs noticed by the right networks. Using social media to share the link to your outputs online can also be beneficial: in an experiment with her own outputs, one researcher tweeted and blogged about her outputs to see what would happen. Publicising her research using these social media made a big impact on how much it was accessed and downloaded: 'The papers that were tweeted and blogged had at least more than 11 times the number of downloads than their sibling paper, which was left to its own devices in the institutional repository'.[3] Using more public channels, like social media, can also help your work to reach a broader audience than you might do by just circulating outputs in your core network.

TRACKING IMPACT

Curating, sharing and promoting your outputs are all well and good, but how do you know if it is working? How can you tell if these activities are helping your work to reach more people, make an impact and possibly get more citations down the line? And how can you report on this impact?

Statistics

OpenUCT (and many other repositories) have statistics associated with each output. These show views, downloads, etc. and can give an indication of how many people are accessing the content. Several of the more basic statistics may be publicly accessible, with more details available from the repository, where possible.

Analytics

In addition to the basic statistics for resources within the repository, more detailed analytics (for example, data provided by **Google Analytics**) can show how much traffic an item or group of items is getting, as well as where this is coming from — both geographically as well as through which other sites via referrals. So, for example, you'd be able to track how much traffic is coming from Twitter or academia.edu. Google Analytics data is, however, restricted to the site owner and other authorised people.

Altmetrics (alternative metrics)

Altmetrics can provide a much more immediate picture of the reach and impact of your outputs than you could get from citations, which can take years to start accumulating, and often only apply to academic works such as journal articles. There are several altmetrics service providers such as **Altmetric.com**, **Plum Analytics**, and **ImpactStory.org**.

Taken together, these statistics, analytics and altmetric data can help to inform you about where your work is going, who's reading it and what networks you're reaching. Thinking upfront about what matters to you (and your funders) in terms of impact can help you to focus on the type of data that will be most useful to you. This considered approach will help to yield results that are meaningful in your context. Drilling into this data may also help you find interested people and collaborators where you wouldn't have known to look before.

WHO TO ASK

UCT Libraries can assist academics in starting to assess the impact and reach of their resources.

USEFUL RESOURCES

Academics' online presence: a four-step guide to taking control of your visibility
http://open.uct.ac.za/handle/11427/2652

Make your digital resources easier to discover, 2014
http://www.jisc.ac.uk/guides/make-your-digital-resources-easier-to-discover

STEP 8

CURATION AS AN ONGOING PROCESS

Curation doesn't just happen once. It is a continual process which should ideally be built into other processes. Although you may have come into the job without much idea of what to do, you will hopefully have begun to understand the value of curation, and developed some ideas about how to do it. Here's how to keep the ball rolling.

ESTABLISHING A RECORD

You may have come into an environment where curation had not been a high priority, or not even considered at all. You may have been faced with the prospect of identifying a bewildering array of different types of content, getting it from intractable scholars, making sense of it all and establishing whether or how you're allowed to share it, getting it ready for sharing, finding somewhere to put it, and making sure that people can find it, see it, and use it. You may have encountered new terms and novel concepts, performed new activities, and learned new skills. Talk about a baptism of fire! But if you've made it this far, you can be very proud of yourself. You have already achieved a great deal, and helped to establish a record which will hopefully continue to grow and develop. However, your work is far from over! Establishing a record is merely the first step in the curatorial endeavour.

CURATION IS ITERATIVE

Although you may have established order where there previously was none, or helped to build a collection of content, that initial push needs to be maintained or history will repeat itself. Now that you have developed some processes, you need to make sure that those processes continue. Integrating curatorial processes into the administration of a unit, department, or an individual is the best way of ensuring that they continue. Every academic at UCT is required to perform a process of self-review each year, completing an evaluation form (HR174) where they list their teaching and research activities conducted throughout the year. This offers a convenient way of identifying research publications and teaching materials for sharing. Now, thanks to the Open Access Policy, they are also required to submit copies of their research outputs for sharing. Using the evaluation forms as a guide, you can work with academics to make sure that their content is being shared appropriately to fulfil the university mandate.

Your work will also help to provide a structure for the unit or department to continue curating its content, and this structure will have benefits in the long term, helping the unit to perform reporting duties required by the university and funders.

DEVELOPING A CULTURE

Given enough time and constant encouragement, a culture of curation can develop. If academics can see all of the benefits that good curation brings, they may be more inclined to do it consistently. At all times during the process, stress how important it

is to keep records and retain versions of work. Demonstrate the impact of greater visibility, and the benefit to individual and departmental profiles of having openly available content.

An open source tutorial on an open source study on open source communities on open source... image by Meredith Atwater for opensource.com available under a CC-BY-SA licence.

POLICIES

All of UCT's active policies can be found **HERE**, but these are a few that have been referenced in this guide, and with which you should familiarise yourself:

- The **Open Access Policy** has the specific purpose of making available content produced at or by UCT. It officially sanctions and encourages the sharing of content, and it stipulates that content must be shared through an appropriate portal, but does not prescribe which one, leaving the choice of which portal up to individuals.

- The **Metadata and Information Architecture Policy** is intended to provide guidance on generating metadata for all kinds of resources at UCT.

- The **Intellectual Property Policy** provides guidance on all of the university's decisions around IP. It makes explicit provision for the sharing of content under open licences.

- The **Web Content Management Policy** contains the minimum standards and guidelines that need to be followed by an official UCT website to ensure the consistency, currency, accuracy and reliability of information published online.

- The **Teaching and Learning Strategy** officially acknowledges the importance of the development and sharing of Open Educational Resources.

GLOSSARY

altmetrics different types of metrics used for measuring impact beyond the traditional academic scope of citation counting and h-indexes

CC Creative Commons, a type of open licence widely used in conjunction with copyright to allow easy use of a resource

copyright a type of intellectual property right which allows an author or creator of a resource to control how the material they generate is used

curation the act and process of collecting, describing, and providing access to resources, either physical or digital

DOI digital object identifier, a type of persistent identifier which will always point to the same resource online

impact all of the various effects that research products and educational resources — in their many forms — can have throughout the cycle of research and teaching and learning

IP intellectual property, includes copyright, patents, and trademarks

metadata data about data, used to describe and locate resources online

OA Open Access, a movement which promotes free access to research, both for academics as well as for the public

OCR optical character recognition, software which is useful for converting physical documents into usable digital versions

OER Open Educational Resources, teaching and learning materials which have been generated or modified specifically for sharing in an open way

PDF portable document format, a file type commonly used for sharing documents and accepted as a standard because of its relative stability and accessibility

repository an online platform which allows for the storage and access of digital resources, they can be either institutionally operated or subject specific, closed or open

URL Universal Resource Locator, the address for a resource on the internet, such as a website or online database

REFERENCES

1. Connaway, LS, Dickey, TJ & Radford, ML. 2011. If it is too inconvenient I'm not going after it: convenience as a critical factor in information-seeking behaviors. *Library Information Science Research*, 33(3): 179–190. Elsevier Inc. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0740818811000375

2. Beitzel, SM, Jensen, EC, Chowdhury, A, Frieder, O & Grossman, D. 2007. Temporal analysis of a very large topically categorised Web query log. *Journal of the American Society for Information Science and Technology*, 58(2):166–178. Available at: http://onlinelibrary.wiley.com/doi/10.1002/asi.20464/full)

3. Terras, M. 2012. 'The verdict: is blogging or tweeting about research papers worth it?' Available at: http://blogs.lse.ac.uk/impactofsocialsciences/2012/04/19/blog-tweeting-papers-worth-it/

Cover image by Stacey Stent

ACKNOWLEDGEMENTS

DISCLAIMER
All suggestions given in this guide are to make you aware of the possibilities and are not necessarily endorsed by the University of Cape Town.