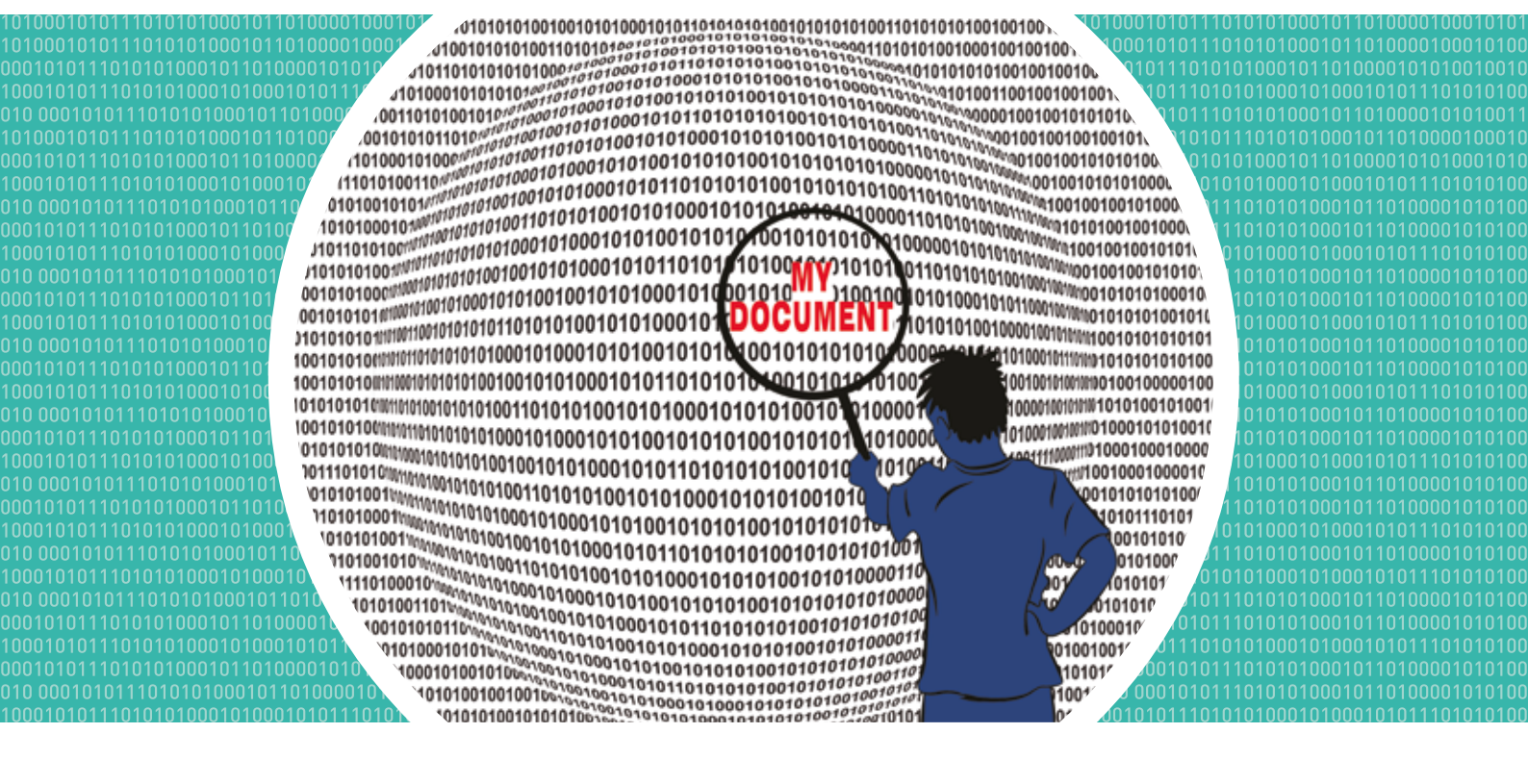


OpenUCT Guide

open.uct.ac.za



CURATION FOR PARTICIPATION

An
EIGHT-STEP GUIDE
to curating open scholarly content

Kyle Rother, Sarah Goodier, Lena Nyahodza
and Laura Czerniewicz

Cover image by Stacey Stent

Kyle Rother, Sarah Goodier, Lena Nyahodza and Laura Czerniewicz
OpenUCT Initiative
University of Cape Town
Private Bag, Rondebosch, Cape Town 7700, South Africa
First edition August 2014
This [second] edition June 2015

This work is licensed under a Creative Commons Attribution-Share
Alike 4.0 International Licence

Available at: <http://open.uct.ac.za/handle/11427/8431>

Contact us: kyle.rother@uct.ac.za, sarah.goodier@uct.ac.za,
lena.nyahodza@uct.ac.za, laura.czerniewicz@uct.ac.za

ACKNOWLEDGEMENTS

OpenUCT would like to thank Richard Higgs, whose interesting points and insightful comments helped shape and improve this guide. Thanks also to COMPRESS.dsl for design, editing and layout. Grateful acknowledgements are due to the Carnegie Corporation of New York for financial support in producing this resource.

DISCLAIMER

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



CONTENTS

Introduction	3
Who is this guide for?	3
Curation for participation	3
About the guide	4
STEP 1 Identifying content	5
Which content?	5
Which time period?	6
Which people?	6
Who to ask	7
STEP 2 Sourcing content	8
Looking on the web	8
Looking at UCT	9
Looking elsewhere	10
Contacting academics	10
Re-use	10
Who to ask	10
STEP 3 Understanding the legal stuff	11
Copyright	11
Acceptable versions	12
Third-party content	12
Choosing a licence	14
Who to ask	16
Useful resources	16
STEP 4 Digitising content	17
Digitising texts	17
Who to ask	21
Useful resources	21

STEP 5 Describing the content	22
Metadata.....	22
File names.....	23
Metadata for different content types.....	24
Working with batches	25
Metadata for harvesting	26
Maximising discoverability.....	27
Who to ask.....	27
STEP 6 Uploading content	28
Where to put stuff	26
Getting it out there.....	29
Who to ask.....	29
Useful resources	29
STEP 7 Alerting people and tracking impact	31
Telling people about the content.....	31
Tracking impact	32
Useful resources	34
STEP 8 Curation as an ongoing process	35
Establishing a record	35
Curation is iterative	35
Developing a culture of curation.....	36
Policies	37
Glossary	38
References	40

INTRODUCTION

WHO IS THIS GUIDE FOR?

This guide is for you if you find yourself curating scholarly and educational content at UCT, even if you don't know what curation is all about. You might be a postgraduate student who is putting a professor's research output online, or you might be a communications officer whose tasks have now extended to arranging and presenting your unit's content.* You might be a webmaster who finds that you are asked to upload all kinds of resources. You might be employed part-time, or on an ad-hoc basis. You probably don't have a qualification in digital curation or library science, and might have little to no experience in this type of work. But you find yourself having to figure this out. So what needs to be done?

CURATION FOR PARTICIPATION

What is curation? And why is participation so central?

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

Preservation. Scholars need to keep a record of their work. In our digital age, scholars sometimes lose track of copies of their work — with their output stored in the cloud, handed over to publishers, lost on discs or mislaid on old computers. Universities play an important role in preserving scholarly content in a secure, central and trustworthy place for future use.

Discoverability. How scholarly content is curated is critical to it being found, shared and cited. Studies make it clear that online searches are now the main way that scholars locate research studies and data; URLs are rarely known or used.¹

* Please note that since this guide was conceived mainly for use within UCT, many of the links to subscription services and university resources are best accessed from within UCT's network.

In our digital world, making output available online is one of the primary ways that scholars participate in academic and disciplinary networks. This is why curation really matters, and why it is vital that curation is done well.

What does 'curation' mean? In popular lingo, curation sometimes refers to the selection and filtering of content through social media tools like Scoop.it. The focus of this guide is not so much on the selection or filtering of content because researchers themselves need to make those decisions. But the guide should help you work out how best to arrange and present the scholarly output that you have been asked to put online.

This guide is specifically about *open* scholarly content, which means it is about making actual content (and not just a description of it) freely and legally available, and this applies to all types of content and a range of formats.

ABOUT THE GUIDE

The guide takes you through 8 steps:

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

Most steps include references to specific resources and units at UCT that should be able to give you some help and support. You will find a glossary of useful terms, as well as a list of all the policies cited, and some of general interest and applicability, at the end of the document.

STEP 1

IDENTIFYING CONTENT

How much material will you be curating? Which types of files will you be working with? These decisions might not be yours so you will need to work with your manager to make sure your brief is very clear. You will need answers to the following important questions so that you don't flounder or do the wrong thing!

WHICH CONTENT?

Scholarly output can be difficult to define. Is it peer-reviewed journal articles, monographs or chapters in books? Does it include submissions on proposed state policies or legislation? How about teaching resources and slides for conference presentations? All of these constitute outputs of the scholarly process, and should be considered as possible items for curation.

Be guided by your brief and work with your manager to decide on the intended outcomes of the curatorial exercise. This will guide how much and what kinds of material should be included. Make sure you are clear on the primary focus of the individual, unit or department you are working for, and the main outputs they produce. Try to work out who is most likely to use the resources you make available, and what for. The answers to these questions will help you prioritise what to focus on. Remember, your job is to curate, not create!

You may find that someone within the unit already has a strong online presence and knows how they want their output curated. This person could be a useful guide as you set about curating the work of others in the unit, so you should always be receptive to their input, and try to ensure that approach followed is consistent across the unit.

WHICH TIME PERIOD?

For administrative purposes, it is best to limit yourself to outputs produced by staff during the course of their work or studies at UCT. UCT's **Open Access Policy** has the specific purpose of making content available that has been produced by researchers and others at the university. This reduces the complexity of shared copyright and licensing (**see Step 3**). So even if an academic has had a long and illustrious career spanning several other institutions, be quite strict about working with what they have produced since arriving at UCT (unless your brief and the budget allows otherwise). Once they see the benefit of the curatorial exercise, they might be motivated to do the same kind of thing with the rest of their output!

For those academics who have worked at UCT for decades, choices may have to be made about what outputs to select or prioritise. The older the material, the less likely it is to be in a digital format, and this could add another layer of complexity to your task (**see Step 4**).



WHICH PEOPLE?

Obviously you will be working only with UCT staff and some postgraduate students. Although this is a relatively small community (in terms of the global network of scholars), you can expect to encounter a variety of people. You might be reporting to a veteran researcher with an extensive body of work who has no interest in curation, or you might be working with a relative newcomer, who has comparatively little to share, but really understands the value of open access, and is eager to contribute. You might encounter a good deal of resistance or apathy. But once the value of the curation becomes clear and

academics experience the benefits, most become more enthusiastic. It is a fair assumption that all academics want their work to be found, read and cited as this has a direct impact on their reputations.

WHO TO ASK

- Your employer or advisory group.
- UCT's Library and Information Studies Centre, which offers training in digital curation, as well as courses on research-data management.
- Other people involved in similar work at UCT (you might be surprised at how many there are). Since this is a new field, everyone is learning from everyone else. You could try speaking to people in similar positions to yours in other units to share tips and challenges, or you could contact someone in UCT Libraries' Access and Visibility Services for guidance and advice.

STEP 2

SOURCING CONTENT

Academics are busy people – finding time to meet with them can be difficult. If they can't make time to discuss their own curatorial efforts, it might be a good idea to start creating a list of outputs that you can ask them to double check and add to if necessary.

LOOKING ON THE WEB

Online databases are a great source of publication information. 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 to your curation list. Searching for an academic by name can provide results if the name is fairly unique. Adding 'UCT' to your query will help you narrow down the search and get more relevant results.

Many academics have at least one profile that includes a list of their publications. Look for these profiles on the relevant departmental website and on academic networking sites, such as **Academia.edu** and **ResearchGate**. Some also include their publications on professional sites such as **LinkedIn**. If you are lucky, you might find some of the academics



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

you are assisting have a Google Scholar profile — these often contain a fairly comprehensive list of publications that are already online and searchable by Google; look up names on Google Scholar to find out who has a profile.

Bibliographic databases and indexing services (e.g. **Scopus**, **Web of Knowledge**) can also help you locate research material by particular academics. These databases are available via **UCT Libraries** subscriptions to these services and many of these allow you to search by author and affiliation.

You can also search the web for particular file formats and content types. Use **Google Advanced Search** and choose from the ‘file type’ options to search for PDFs, presentations, and so on.

LOOKING AT UCT

UCT’s latest Annual Research Report is a good source for lists of research outputs in the narrower sense. UCT uses a web-based information management system — known as IRMA (Integrated Research Management Application) — for collating data for the Annual Research Report and the annual publication count. You might be able to extract data about accredited journal articles, books or book chapters from this system (see the ‘Who to Ask’ section on **page 10**).

Some academics make teaching resources available on the **OpenUCT** repository. There is a fine line between teaching resources and academic resources, so it is worth checking these too.

Some departmental/project websites can be a good source of information about publications and other types of academic outputs. Research institutes at (or linked to) UCT might have their own collections and lists of publications, so you might find it useful to check their websites too.

LOOKING ELSEWHERE

Many academics contribute to national media debates by writing opinion pieces or giving interviews. You might find articles and video footage in the archives of media organisations, most of which are now online.

Some academic disciplines rely more heavily the publication of monographs, edited collections, and book chapters. It may be useful to search the **UCT Libraries catalogue**, as well as **SABINET**, a catalogue that contains bibliographic information about all books published in South Africa.

CONTACTING ACADEMICS

Once you have a list of outputs, you can contact the relevant academic/s again. Get the ball rolling with an email or a phone call requesting a meeting to discuss your curation activities. You can ask the academic to look at the list you have compiled and add any outputs that you might have missed.

RE-USE

Obviously, it is not sensible to recapture and reload output that already exists in an open repository elsewhere. Open journals, for example, publish authorised copies on their websites or through open-content aggregators, along with all the necessary metadata. Do not reinvent the wheel: rather find out if creating simple links to resources that already exist (using metadata harvesting and DOI numbers) will enable you to sufficiently fulfil the aim of your curation project, (for more on metadata harvesting and DOI numbers see **Steps 5 and 6**).

WHO TO ASK

To inquire about training on IRMA, you can contact rea-irma@uct.ac.za.

You can ask UCT librarians for help when searching catalogues and databases.

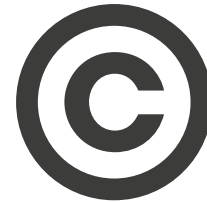
STEP 3

UNDERSTANDING THE LEGAL STUFF

Before you do anything further, you have to find the legal status of each output, that is — what you can legally share, and/or on what terms the output can be shared.

COPYRIGHT

Copyright obtains automatically in any work. The only way it can be waived is if authors expressly decide to do so, and state that they do not wish to retain any of their rights as copyright holder.



Most scholarly content is produced by a single person or a small group, and is often published as an article in a journal, a chapter in a book, or as a conference paper. If published in a commercial, peer-reviewed publication, copyright will often rest with the publisher. This means that you will have to find out under what conditions if any, the publisher will allow the material to be shared (see ‘Acceptable versions’ on **page 12**).

Copyright related to research work conducted and published while at UCT, usually rests with the author/s or with the unit or department in which they work. This means you will have to get permission to share the material from the author/s or their department. Even if you had already obtained this permission when sourcing the content, you still have to explicitly state that you have permission to share it, and on what terms (see ‘Choosing a licence’ on **page 15**). This information is usually included in metadata, or on a cover sheet added to a document (for more about metadata, see **Step 5**).

If the work was partly or wholly funded by extra-institutional sources, copyright may rest jointly with the author/s or their department *and* the external funder. This is a slightly more complicated situation as permission has to be sought from the funder too. Some of these organisations have quite specific conditions around the sharing of content. However, these

conditions should have been stipulated in writing in the contract signed at the outset of the research process, so the relevant conditions should not be too difficult to find.

ACCEPTABLE VERSIONS

For commercially published content, copyright is a major issue. The version publishers hold copyright in is usually the final, edited and typeset version of the work as it appears in a journal or book. Most publishers are, unfortunately, fairly conservative about allowing content in which they hold copyright to be shared. However, some publishers support open-access publication, in which case they make their final versions available, or allow these to be shared.

Still other publishers make provision for authors to share their work after an embargo period has elapsed. In this case, the version that is shared is usually the author's final draft before or after the peer-review process (pre-print), but before the final editing and typesetting has been done (post-print). If you have access to either pre-print or post-print versions of documents, your job will be slightly simpler. If, however, the academics with whom you are working do not have pre-prints or post-prints for you, try to encourage them to cultivate the habit of keeping at least pre-print versions of their work. This is a good practice both from an academic as well as a curatorial perspective, as it helps establish a legacy of work for an individual and for the department they work for.

If you have pre-prints or post-prints, you can often share these quite openly (after checking that any embargo period that might have been stipulated by the publisher has elapsed), provided that you link the version you have shared to the publisher's final version as well. 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 various journal publishers.

THIRD-PARTY CONTENT

If the content includes copyright material sourced from a third party, then either permission for sharing those materials has to be sought, or alternatives that are legally shareable will have to be found or created and inserted instead.

For example, when putting together a presentation, academics often lift photographs or other material (tables, diagrams, etc.) off the internet for illustrative purposes without checking whether any copyright regulations apply. This should be fine, as long as the lecturer uses the images for one presentation only, and provided the presentation is done for instructional or educational purposes and not for commercial gain. However, if the academic then wants to share the presentation with others, even just internally at UCT (e.g. on **Vula**), copyright infringement is possible if the images are not legally open resources.

When in doubt, it is wise to seek permission. If the copyright holder is contactable, and willing to share, they can give you permission to use their material, but ask for this in writing for your records. If they are unwilling to share, you have to remove the material. In this case, you might need to find appropriate replacements that are licensed for sharing (see more about licensing on **page 14**), or you can ask someone to create a similar image from scratch. Either way, consult the person who created the presentation, as they will be able to say which aspects of the image are essential to their argument, thus potentially saving you a lot of effort. Note also that several websites contain fully shareable images and content (see the 'Useful resources' section on **page 17**).



CC-BY image by Libby Levi for opensource.com available under a CC-BY licence.

CHOOSING A LICENCE

Once you have established the copyright status of the content, and know what you can and cannot share, you have a number of options. If traditional copyright law (all rights reserved) applies, this continues unaltered. Alternatively, the author might wish to apply an open licence to the content, stipulating conditions for its usage.

All rights reserved

In the case of fully copyrighted material, content can still be open for viewing and use, but the conditions of its re-use are highly limited. For instance, work licensed under most copyright law can only be copied and distributed with permission from the copyright holder. While the practice known as 'fair dealing' does help, it limits others to quoting only a small proportion of the work. So, if someone wants to share an entire piece of research, the original author and/or copyright holder have to be contacted to give their express consent. This is often time consuming, and it can be costly. A preferable alternative is that copyright be maintained, while licences are applied that stipulate conditions for a work's use and re-use.

Licensing

The most established and popular kinds of licence are the **Creative Commons** (CC) licences. These are open licences that can be applied to all kinds of work, and explicitly indicate the conditions under which copyright material can be used. Note that CC licences are not a substitute or a replacement for copyright, and they are based on copyright law. The licences simply remove the need for others to seek permission to use material by specifying varying kinds permissions in advance, and making it clear what the conditions related to these permissions are.

A copyrighted work can have a CC licence that extends certain rights to the user, but retains others exclusively for 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 in advance to anyone wishing to quote from, adapt or modify the work, as long as the source is acknowledged, and the resulting work is also shared. This is not so far from the existing practices of academia. Similarly, if an author prefers not to retain any rights as copyright holder, a CC-0 or 'CC-Zero' licence can be applied.



Creative Commons music image by opensource.com available under a CC-BY 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 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

See the OpenUCT Guide on *Open Access Content Licensing: A Three-Step Guide for Academics*.

OpenUCT also has 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 for you and for academics who might want to know more about open licences and how to apply them. Click here for 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 the file format and condition of the content.

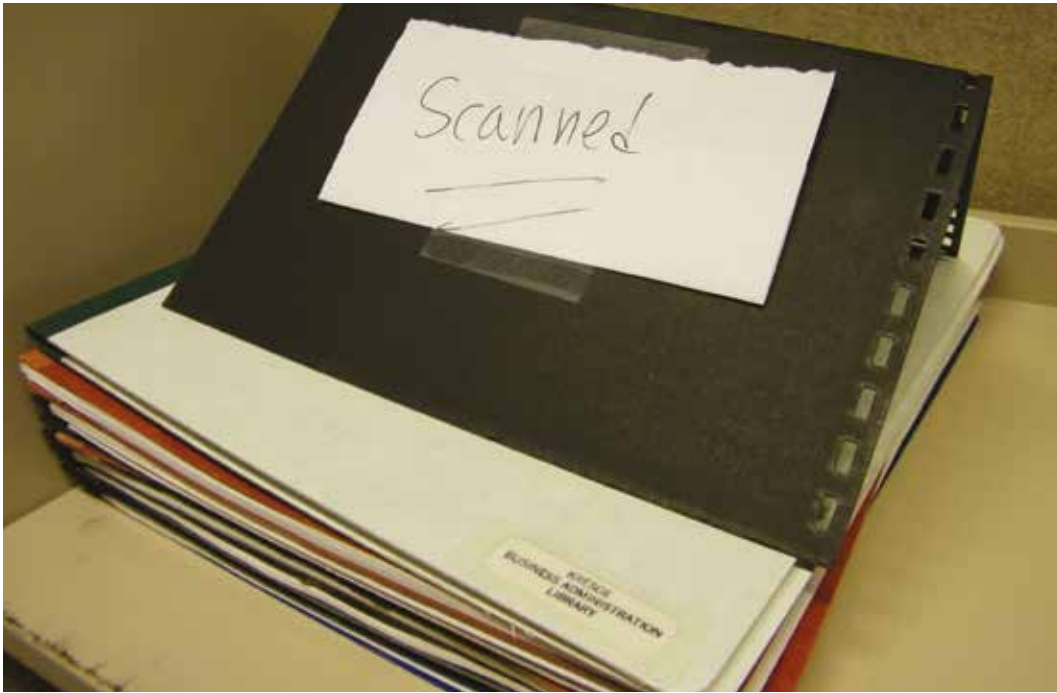
DIGITISING TEXTS

Apart from different kinds of digital files, you might be asked to include legacy documents that exist in hard copy only. If you have access to a scanner, digitising should be fairly straightforward.

You have two options when digitising documents. You can either create an image file (like a photograph of the document), so that the digital copy is a static picture of the original. Alternatively, you can scan the document using optical character recognition (OCR) software. Not everyone has access to this software, but if you do, you can create a more flexible rich-text file that can be edited, searched and copied, etc.

If you are simply creating an image, accepted practice is to scan it in grayscale at 300dpi. This is preferable to scanning in bitonal black and white as the resulting images appear clearer, and if the quality of the original document is poor, greyscale scanning picks up more details. Choosing greyscale must be done via the scanner's software prior to scanning. The scanner will also usually provide several file-format options that you can use to save the image (e.g.: .tiff, .jpg, .png, and .pdf). Accepted best practice is to use TIFF files for archival masters of images, particularly if you are scanning photographs. For documents, JPEG or PDF files are preferable.

If you decide to go the OCR route, the scanner either scans the original document directly and allows you to save the document as a rich text file, or it converts an existing image file into a rich text format (a .rtf file). Rich text files allow users to manipulate the text (by



This pile done! image by Corey Seeman available under a CC-BY-NC-SA licence.

searching, copying, correcting etc.). They also allow search engines to crawl through the actual document and index its contents. For this reason, rich text files are preferable and recommended over image scans for certain kinds of documents. However, as mentioned, not everyone has access to OCR software, and although the software programs are becoming more stable, 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 PDF, a variety of software packages exist to convert PDFs to Word or rich text documents. Most of these packages, such as Adobe Acrobat (not the same as Acrobat Reader), 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.

Proofreading and cleaning

After creating a rich text document via a PDF converter or OCR software on a scanner, it is vital to make time to carefully proofread and clean the converted documents against the original, checking if:

- any text is missing
- headings and paragraphs are still properly positioned
- tables and figures are converted and properly positioned
- text from tables is not presented as plain text
- content in tables and figures has not been compromised (no missing text)
- any letters or words have been corrupted
- text spacing has been correctly retained

Tables sometimes end up being located at the end of a scanned document; it is therefore important to look at the entire file before starting the cleaning process.

1. Start by checking whether the integrity of text, tables and figures is intact. If tables and figures are displaced, restore them by copying and pasting. Avoid cutting because if the pasting fails, you might lose content and have to rescan. If figures and tables are corrupted and cannot be manipulated, consider copying these from the original document one at a time, or try converting the document using different software. Save your document often while you check.
2. Take note of corrupted letters and symbols that might crop up. These might be difficult to spot, so you need to scrutinise the document thoroughly or risk leaving errors in the file that will compromise the integrity of the work. If possible, run a spell check, and save often as you make corrections. See an example of corrupted text on page 21:

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 promole full paricipalion on society and the economy through equilable and meaningful access too education; (2) to pursue the language policy most supporlive of general conceplual growth amongst learners, and hence to eslablsh addilive mullilingualism as and approach to language in education; (3) to promole 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. Remove extra paragraph breaks and check line breaks: in the conversion from Word to PDF, the software often inserts paragraph breaks where line breaks appear in the original, leaving odd gaps in sentences. Check also that hyphens that might have been needed at the ends of lines in the original, don't now appear where they are no longer needed. Save the document again.
4. Insert page numbers.
5. Save the document in the required format for online sharing.

6. If you are considering sharing the final product of the work online, insert a page for the relevant metadata along, with a statement on the provenance of the scan (which gives information about where the original can be found, who scanned the file and when, and what settings were used, and so on).

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**, a special division within UCT Libraries that is responsible for digitising the university's archival collections.



Is licensing really the most important question for OER? image by Caroline Madigan for opensource.com available under a CC-BY-SA licence.

USEFUL RESOURCES

The United States' Library of Congress has some useful resources on archiving at <http://digitalpreservation.gov/personalarchiving/> including a document entitled **Scanning Your Personal Collections**.

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. Good descriptions make the difference between simply making content available, and enabling others to find and use it.

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 is contained in a metadata record, attached to the file it describes, and is used by search engines to gather information about the file’s contents. Metadata records are also usually displayed in a kind of directory that viewers use to help them decide whether an item is of interest or value to them. For this reason, it is essential to ensure that your metadata is clear and concise, consistent and intelligible.

There are a number of internationally accepted standards for generating metadata. Probably the most common of these is the **Dublin Core Metadata Initiative (Dublin Core)** for short). Dublin Core separates metadata into discrete elements or fields that represent key aspects of any document, image or dataset, such as its title (the ‘title’ field), 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 as much and as often as desired, and combined into records of varying levels of complexity or depth; Dublin Core is not prescriptive.

UCT's Information and Communication Technology Services and UCT Libraries have jointly formed a Metadata Working Group that has developed a set of guidelines on generating metadata for different types of content (see 'Who to ask' on **page 28**). These guidelines (based on the Dublin Core) specify the basic and mandatory metadata fields for UCT.

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 relevant file extension (e.g. .pdf, .jpg, .tiff, .xls, .doc, etc.) should be included.

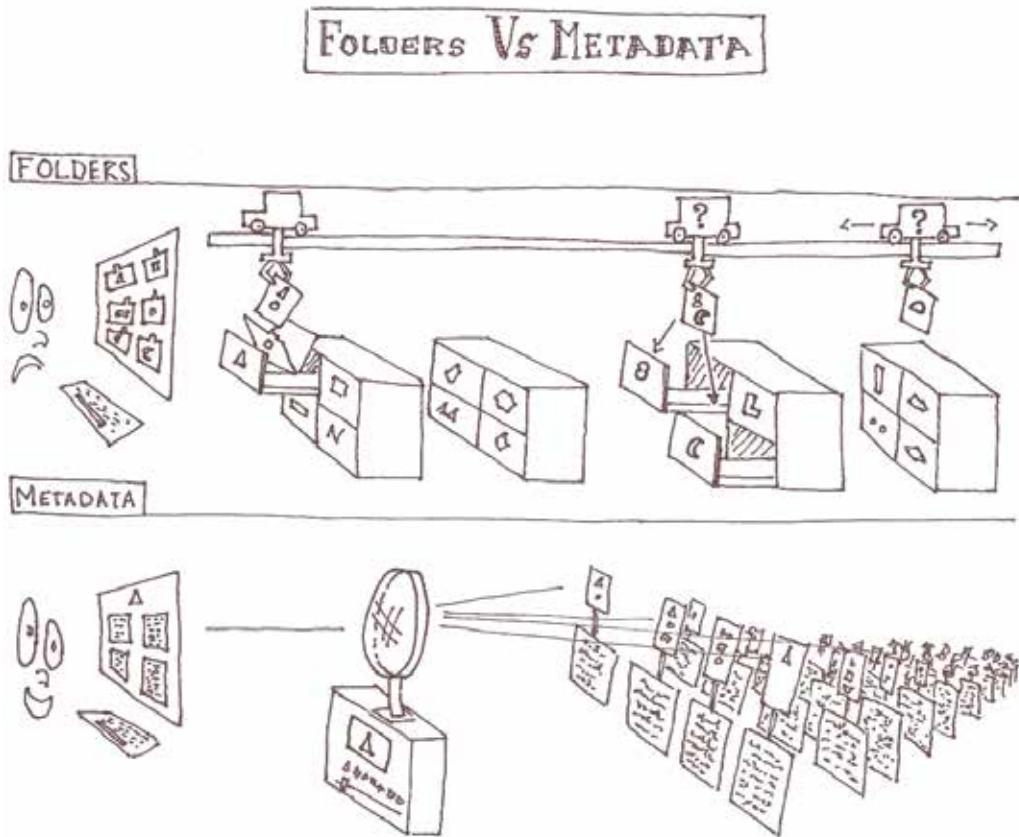
Try to use only lowercase, unless it's for an acronym, and do not use the following characters in file names:

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

Also avoid using parentheses (brackets) and spaces in file names. Underscores (_) and plus signs (+) can be used.

METADATA FOR DIFFERENT CONTENT TYPES

Different content types might have different metadata requirements. UCT's Metadata Working Group has developed templates for different content types that make it relatively easy to prepare content for deposit into UCT's default repository, **OpenUCT** (see the 'Who to ask' section on **page 28**).



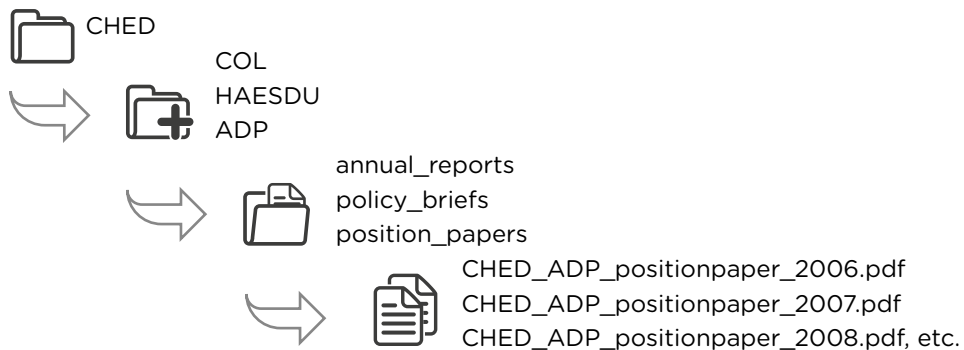
john-norris.net CC SA-BY 2.0

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

WORKING WITH BATCHES

If you have been tasked with curating the content of an entire unit or department, you are likely to be working with a large amount of material. Without a clear process or structure, it is easy to lose track of all the different documents. Consider keeping content grouped together in batches that have some link in terms of content. Creating a consistent file-naming protocol will prove useful.

A file-naming protocol goes hand in hand with a logical 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 mentioned above. Your files all contain the word 'positionpaper', so they should be placed together in a folder named 'position_papers'. That folder will sit alongside other folders that contain similar types of document (e.g. 'annual_reports', 'policy_briefs', etc.) inside a parent folder named 'ADP'. Then the ADP folder might be included in a further parent folder (perhaps named 'CHED'), and alongside folders for other units in CHED (e.g. 'HAESDU', 'COL', etc.). Eventually you will end up with a folder structure that looks something like this:



Folder structures can quickly gather layers and complexity, so try not to create too many layers, as this makes for cumbersome path names, and can 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, (three folders plus the file name). It is a good idea to try to keep your structure no more than three folders deep. It might also be sensible to develop a 'finding aid' — that is, a document that outlines your folder structure and file-naming conventions.

This will give you and future curators a snapshot of the scope and structure of the collection.

If you are creating metadata for 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 can be arranged in rows, and the metadata elements in columns, 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

One of the benefits of working in a spreadsheet is that the metadata for many of the items will be identical, so you can simply pull the same text into multiple cells. There are also benefits when it comes to uploading content, and using a spreadsheet is a great way of maintaining a full record of a unit or department's online output, but more on that later.

METADATA FOR HARVESTING

One of the most powerful applications of metadata is that it allows search engines to 'harvest' data from repositories, thus dramatically expanding the potential reach of content. Meticulously compiled metadata also helps to prevent duplication, in that an item need only be stored in a single repository but still be accessible to many other repositories all over the world. Making metadata harvest-friendly makes online content far more accessible,

thus potentially increasing its usage and enhancing citation rates and so enhancing UCT's institutional reputation. This is another good reason for sticking to international standards such as the Dublin Core.

MAXIMISING DISCOVERABILITY

Most open access repositories make their content available for indexing by search engines such as Yahoo and Google. Through clever curation and the consistent use of metadata, you help to ensure the indexing of your content by these search engines. Relevant keywords added to your metadata will help the search engines sort your material appropriately, and help to make your content appear higher up in result lists. This is part of a process called search-engine optimisation. Few of us look past the first page of search results, so the higher up in the list your content appears, the more likely they are to be clicked on, read and cited.²

Subject headings, controlled vocabularies and keyword lists

To help ensure that metadata is created in ways that maximise the harvesting and sharing of resources, most academic disciplines have developed and agreed on predefined subject headings and keywords that are accepted across their fields. Also known as open metadata standards, these are lists of common terms or controlled vocabularies used by institutions working in the same discipline. Many disciplines also have their own open repositories, and it is worth finding out if these use controlled vocabularies and/or have created any open metadata standards. You can email the repository manager to find out about this, usually via the 'contact us' section.

WHO TO ASK

UCT's Metadata Working Group is responsible for implementing the university's Metadata and Information-Architecture Policy. Read the **UCT metadata entry guidelines** before you start.

STEP 6

UPLOADING CONTENT

You've arrived at possibly the most important part of curating: sharing the stuff. As essential as all the preceding steps are to the process, if it isn't online, it isn't easily accessible.

WHERE TO PUT STUFF

UCT policy supports the open sharing of content generated by university community. The university has established a number of portals via which material can be shared, and any appropriate portal is acceptable. The default repository, **OpenUCT**, is designed to accommodate a wide range of content types and formats, and to expose these as widely as possible. The OpenUCT portal serves several functions: it preserves and records; it provides a 'shop front' through which anyone can see and access what the university is producing; and it is designed to be harvestable by search engines, ensuring that the content reaches the widest possible global audience.

Your unit or department might have a website of its own and prefer to host its own content. This is fine, and the work you have done will still prove valuable to the uploading process. The website might have an existing 'publications' section, or you might have to create one. Depending on your experience and the scope you have been given, you might have to consult the site's webmaster about this.

Even when sharing content on a unit or department's website, it might still best to describe and upload material to the OpenUCT repository, and then create links to the unit website. This is because the OpenUCT repository is intended to fulfil an archival function, which means that its content will be appropriately preserved and maintained indefinitely. So, if the unit or department you are working in ever merges with another or decides to shut down its website, the content will still remain accessible via the OpenUCT platform. Using OpenUCT in this way also has the practical advantage of saving space on the departmental



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

website. In addition, each item loaded on the OpenUCT repository is automatically allocated a permanent digital object identifier (DOI number); these numbers are becoming increasingly useful for tracking purposes.

As mentioned, a number of disciplinary repositories exist internationally. For example, **arXiv** (pronounced 'archive') 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. (For information about other disciplinary repositories, see 'Useful resources' on **page 30**.) Talk to your manager about whether you should upload your material to a disciplinary repository and then link to OpenUCT as required by UCT's **Open Access Policy**.

If your manager is unaware of the OpenUCT repository, and has allocated funds for the creation of a new repository, we recommend that you (gently) explain that OpenUCT fulfils most requirements, and that any available budget might be spent more effectively on content sourcing, metadata allocation and digitisation of archival material than on replicating technology that already exists.

GETTING IT OUT THERE

OpenUCT has a user-friendly interface to guide users when uploading content. However, if you are working with large batches of material, it might 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 off, because if all this has been done well, the upload process should be painless.

Most repositories have similar upload processes but it is always best to check their user documentation for specific instructions.

Uploading content directly to a website is slightly more complicated so, as mentioned, you might need to work with a webmaster to do this. However, all of the curatorial work you have done will make the process much simpler, and you can then also link this content to items that you have uploaded to OpenUCT.

WHO TO ASK

It is probably best to discuss the most appropriate portal with the head of the unit or department you are working with because strategic or political motives often drive this decision. But be sure to familiarise yourself with all of the options, and their various pros and cons, so that you can offer advice if needed. Once a decision has been made, you can either start uploading, or speak to someone in UCT Libraries or at the helpdesk run by UCT's Information and Communication Technology Services.

USEFUL RESOURCES

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

STEP 7

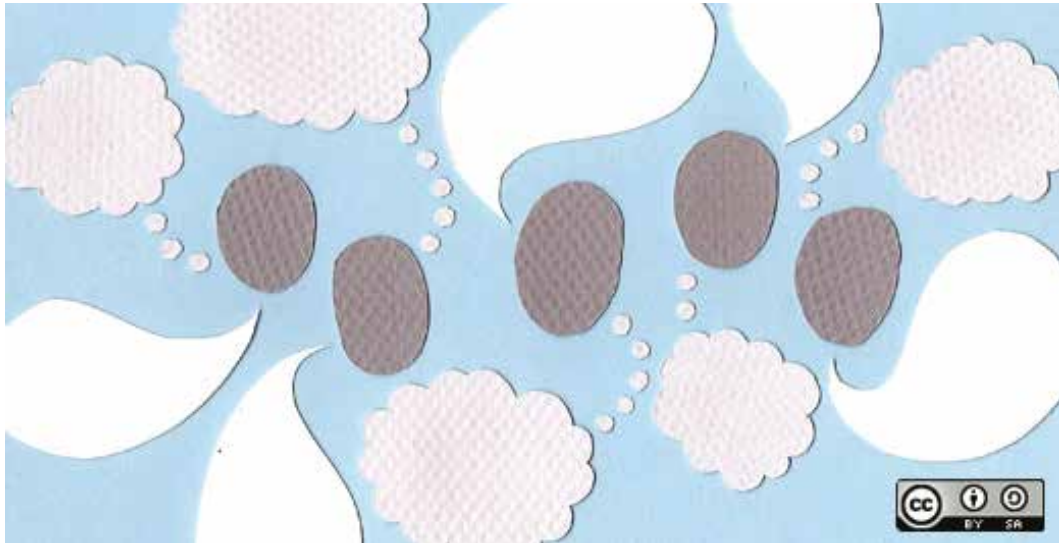
ALERTING PEOPLE AND TRACKING IMPACT

Just putting content online is like holding a party without sending out invitations – if no one knows about it, they're not likely to show up. But don't worry: this bit is easy and it can be fun. When you start telling people where to find your content, you might be surprised by who it reaches, who reaches out to you in return, and where your unit or department starts to make an impact.

TELLING PEOPLE ABOUT THE CONTENT

With online content multiplying constantly, hoping that someone will come across your material while browsing the internet is not the best strategy. Alerting people within your community and beyond to your online content, telling them what it covers, and giving them relevant links will help ensure that your content is more widely read.

Academics or the relevant communications officer should write short articles, blogs or e-Newsletters and circulate the links to the latest paper/s you've put online, so be sure to send them a note whenever new content is uploaded. Linking online papers to specific web pages or author profiles in academic networking sites such as [academia.edu](https://www.academia.edu) and [researchgate.net](https://www.researchgate.net) can also help to get outputs noticed by broader networks. Using social media to share links to online outputs can also help to increase awareness. In an experiment with her own work, one researcher tweeted and blogged about her outputs and tracked the response. Publicising her research using 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 Public channels such as social media can also help reach a broader audience than just circulating outputs via the core scholarly networks.



You're making decisions by consensus, but are you collaborating?

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

TRACKING IMPACT

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

Statistics

OpenUCT (and many other repositories) automatically collect statistics associated with each output. These include number of views, downloads, etc. and provide an indication of how many people are accessing each item. The basic statistics are often publicly accessible, and more detail might be available if you contact those responsible for the repository directly.

Analytics

In addition to the basic statistics, more detailed analytics (data provided by **Google Analytics**, for example) can track how much traffic an item or group of items is getting, but also where this is coming from — both geographically as well as which other sites are referring users to your material. So, for example, this helps to track how much traffic is coming via Twitter or academia.edu. Access to Google Analytics data is, however, restricted to authorised people, so you would have to ask for this data from a site owner or repository manager.

Altmetrics

Altmetrics (short for alternative metrics) provide an even more accurate picture of the reach and impact of academic outputs. Several service providers such as Altmetric.com, Plum Analytics, and ImpactStory.org. track, for example:

- views and downloads
- discussions and recommendations on science blogs, Wikipedia, Twitter, Facebook and other social media
- the number of times a work is saved on Mendeley, CiteULike and other bookmarking services
- citations in the scholarly literature, as tracked by Web of Science, Scopus, CrossRef and others

Taken together, statistics, analytics and altmetric data can help to inform you about where 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 impact data that will be most useful to your department. This considered approach will help to yield results that are meaningful in your context. Drilling into this data might also help you find interested people and collaborators where you wouldn't have known to look before.

USEFUL RESOURCES

OpenUCT has published two further guides like this one that are both very useful; see *Open Access Content Licensing: A Three-Step Guide for Academics* and *Measuring Impact: A Five-Step Guide for Scholarly Units*.

The UK-based NGO, Jisc, has published several guides on resource discovery, including *Make Your Digital Resources Easier to Discover*.

If your academic colleagues wish to know more about tracking their impact, you can refer them to another OpenUCT Guide in this series, *Academics' Online Presence: A Four-Step Guide to Taking Control of your Visibility*.



STEP 8

CURATION AS AN ONGOING PROCESS

Curation doesn't just happen once. It is a continual process that should ideally form part of other processes. Although you might have come into this project without much idea of what to do, you have hopefully begun to understand the value of curation, and developed some ideas about how to do it. Here's how to keep going.

ESTABLISHING A RECORD

You might have come into an environment where curation was not a priority, or not even considered. You might have been faced with identifying a bewildering array of different content types of, extracting copies from intractable scholars, making sense of how it fits together, establishing whether and how you're allowed to share it, collating it for sharing, finding somewhere to put it, and then telling people about it so that it gets seen, read and used. You might 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 achieved a great deal, and helped to establish a record that 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 might have established order where previously there was none, or helped to build a collection of brand new content, the initial push has to be maintained or chaos will return. Now that you have developed some processes, you need to make sure that those curatorial processes are integrated into the administration of a unit, department, or an individual scholar's way of working.



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.

Every UCT academic is required to complete a self-review process each year, using an evaluation form (HR174) on which they list the teaching and research activities they have conducted throughout the year. This offers a convenient way of identifying research publications and teaching materials for sharing. Now, thanks to UCT's **Open Access Policy**, they are also required to submit copies of their research outputs for sharing. Using the annual evaluation forms as a guide, you can work with academics to ensure that their content is being shared appropriately and in fulfilment of the university's mandate.

In this way, you can help to provide a structure for the unit or department to continue curating its content. This will have long-term benefits, and help the unit to perform the reporting duties required by the university and its funders.

DEVELOPING A CULTURE OF CURATION

In time and with encouragement, a culture of curation can develop. If academics experience the benefits that can derive from good curation, they might be more inclined to do it or support it more consistently. In explaining the benefits, stress how important it is for academics to be aware of copyright agreements they sign and to retain pre-print and post-print versions of their work. Demonstrate the impact of greater visibility, and the benefits to individuals and the department of open content.

POLICIES

All of UCT's active policies can be found at <http://uct.ac.za/about/policies/>; those listed below have been referenced in this guide, and that you should be familiar with:

- The **Open Access Policy** has the specific purpose of making content produced at or by UCT openly available. It officially sanctions and encourages the sharing of content. While it stipulates that content must be shared through an appropriate portal, it does not prescribe which one, leaving the choice of portal up to individuals.
- The **Metadata and Information Architecture Policy** provides guidance on how to generate metadata for all kinds of resources at UCT.
- The **Intellectual Property Policy** provides guidance on all of the university's decisions around intellectual property. It makes explicit provision for the sharing of content under open licences.
- The **Web Content Management Policy** contains the minimum standards and guidelines to be followed by all official UCT websites to ensure the consistency, currency, accuracy and reliability of information published online.
- The **Teaching and Learning Strategy** officially acknowledges the importance of developing and sharing open educational resources.

GLOSSARY

- altmetrics** a range of metrics used for measuring academic impact that go beyond the scope of traditional citation counts and h-indexes
- CC** Creative Commons, a type of open licence widely used in conjunction with copyright to give users the right to use material in certain ways without having to ask permission first
- copyright** a legal right that allows an author or creator of a resource to control how the intellectual property they generate is stored and distributed
- curation** the act and process of collecting, describing, and providing access to resources, either physical or digital
- DOI** digital object identifier, a type of identifier that always refers to the same online resource
- fair dealing** permits the use of a limited portion of copyright materials in certain circumstances - including for educational purposes - provided that the source is attributed
- harvesting** a regular and automated process of collecting metadata descriptions from different sources to create useful aggregations of metadata and related services (see **ANDS**)
- impact** in this context, impact refers to all of the various effects that research products and educational resources — in their many forms — 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
- metadata standards**
are sets of metadata terms which are commonly used by many institutions working in the same discipline to ensure that metadata is created in ways that maximise sharing and harvesting
- OA** open access, a movement that promotes free access to research for academics and the public

- OCR** optical character recognition, software for converting hard copies into editable digital files
- open educational resources** teaching and learning materials specifically generated or modified for sharing openly and online
- 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 that allows for the storage and access of digital resources; repositories can be 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 and 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:
- 2 Beitzel, SM, Jensen, EC, Chowdhury, A, Frieder, O and 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.
- 3 Terras, M. 2012. **The verdict: Is blogging or tweeting about research papers worth it?** The Impact Blog: London School of Economics and Political Science, 19 April 2012.

CURATION FOR PARTICIPATION: AN EIGHT-STEP GUIDE TO CURATING OPEN SCHOLARLY CONTENT

This guide is intended to introduce first time curators to key concepts in the practice of digital curation. Presented in eight steps, it will take you through the process of curating research and teaching and learning resources from start to finish and beyond. From identifying and sourcing content, to understanding legal concepts like copyright and licencing; digitising texts, describing resources with metadata, and sharing them online; tracking impact and ensuring that curation becomes a continuing practice. This guide is rich with resources and pointers on where to find more information. Although it was written mainly for the context at UCT, the concepts and skills are common to curatorial practice everywhere.

Part of a series:

Measuring Impact: A five-step guide for scholarly units

<http://open.uct.ac.za/handle/11427/12936>

Open Content Licensing: A three-step guide for academics

<http://open.uct.ac.za/handle/11427/12937>

Academics' Online Presence: A four-step guide to taking control of your visibility

<http://open.uct.ac.za/handle/11427/2652>

Curation for Participation: An eight-step guide to curating open scholarly content

<http://open.uct.ac.za/handle/11427/8431>