CURATION AND DISSEMINATION

Willmers, Michelle;King, Thomas;

;

Curation and dissemination

by Michelle Willmers and Thomas King

Summary

The Research on Open Educational Resources for Development (ROER4D) project was a four-year (2013–2017), large-scale networked project which set out to contribute a Global South research perspective on how open educational resources can help to improve access, enhance quality and reduce the cost of education in the Global South. The project engaged a total of 103 researchers in 18 sub-projects across 21 countries from South America, Sub-Saharan Africa and Asia, coordinated by Network Hub teams at the University of Cape Town (UCT) and Wawasan Open University.

This chapter forms part of a project activity toolkit, which is comprised of five documents outlining activities associated with each of the ROER4D UCT Network Hub pillars of project management activity: networking, evaluation, communications, research capacity building, and curation and dissemination (C&D). It is hoped that these chapters will be of practical use to other projects attempting to integrate any of these activities in their operational strategies.

While it has a particularly close link to the project's research capacity building and communications functions, the processes entailed in C&D activity pertained to and supported all aspects of project activity in that they were designed to facilitate the secure storage and findability of content (both internally and externally) and promote collaboration.

This document outlines the project's strategic approach in terms of positioning itself as publisher in the context of an open research framework. It highlights the workflows and processes associated with two key aspects of its C&D activity: publication of a peer-reviewed edited volume and the project's open data initiative.

Acronyms and abbreviations

CC	Creative Commons
C&D	curation and dissemination
DMP	data management plan
OER	open educational resources
IDRC	International Development Research Centre
MGC	Memorandum of Grant Conditions
PI	Principal Investigator
ROER4D	Research on Open Educational Resources for Development

Introduction

The Research on Open Educational Resources for Development (ROER4D) project was a large-scale, networked project which ran from 2013 to 2017 with the aim of providing evidence-based research on the use and impact of Open Educational Resources (OER) from 21 countries in South America, Sub-Saharan Africa and Asia. Capturing, archiving, describing, publishing and communicating research outputs arising from ROER4D sub-projects was a crucial component of the project's ambition to establish an empirical baseline of OER research in the Global South – and for this research to be reusable in the future.

The curation process – the systematic organisation, description and preservation of resources so that they are searchable and accessible, both within and beyond the network – was crucial in order to facilitate access to resources as well as collaboration around content creation and data analysis. Traditionally seen as an add-on or luxury, large-scale research projects are increasingly allocating resources to professionalising the curation of their content, including undertaking publishing functions that were traditionally the role of commercial publishers, and embarking on sophisticated communication and dissemination strategies in order to broaden networks, increase citation and gather evidence of uptake.

Initially, ROER4D curatorial activity was subordinated under the project's communications portfolio. In 2015, it became clear that a coordinated approach was required in order to organise the large amount of written outputs and data produced by its 18 sub-projects, and to fulfil the project's open research ambitions. The open research ambition also required resourcing for dissemination activity to supplement and leverage the curatorial work. Curation and dissemination (C&D) was therefore elevated to its own Network Hub portfolio with a focus on continuing existing curatorial activity and establishing a set of parameters for how content could be optimally shared, with whom, and under what circumstances. In this sense, the project's C&D strategy arose directly from the project's strategic approach to openness and sharing.

The project's C&D focus also arose from the "D" in ROER4D: Development. The project operated on the assumption that OER have a role to play in alleviating societal development pressures associated with the lack of access to quality education. Within this context, it was considered crucial that findings arising from ROER4D research be as accessible and widely shared as possible so that they could reach a broad range of communities and audiences beyond the confines of traditional scholar-to-scholar communication.

The "Development" component within the project title also related to the project's focus on research capacity building, and the development of its researcher network. Operating on the assumption that most scholars (in any geographic setting) require assistance in distilling, packaging and presenting their research for external consumption, the C&D strategy aided in research capacity building and the professionalisation of researchers' curatorial activity by establishing a layer of publishing infrastructure within the project to support and work collaboratively with researchers to refine the structure and presentation of their content for sharing.

The ROER4D C&D strategy was not an attempt to replace the established trajectory of article and chapter publication with traditional publishers in the wake of a research process, but was instead an attempt to supplement researchers' individual curation and publication strategies with a coordinated approach around releasing a centralised,

cohesive collection of content arising from the project. It was also an attempt to instil openness in the conception and design of research outputs, and to further develop the culture of openness and sharing that existed amongst sub-project researchers.

Positioning the project as publisher

In order to boost the visibility and reach of its outputs, and optimally service the project's open research ambitions, ROER4D positioned itself as a publisher in order to promote its research in the international arena while retaining control over release timeframes and processes.

Significant resourcing was devoted to review and publishing expertise within the project in order to undertake rigorous editorial processes, adopt a professional approach to presentation and sharing of content (including microdata), and utilise a range of electronic formats and platforms for content preservation and dissemination. To this end, the project appointed a Curation and Publishing Manager with a background in academic publishing and scholarly communication, as well as a Project Curator with curatorial and copyright management expertise.

The project explicitly adopted an open research methodology in terms of its approach towards information sharing, bringing together aspects of open access, OER, open research and open data practice. Within this context, the C&D strategy enabled the UCT Network Hub, the project's central coordinating entity, to establish itself as an open publishing partner to its network of researchers, supplementing their OER research efforts with a strategic open approach to sharing that research.

Navigating challenges around language diversity, varying geographical context, differing research cultures and divergent stakeholder focus, the project explored the complex dynamics associated with facilitating multilingual content development and sharing processes, while navigating the "conundrum of open" in terms of when and how content should be shared. The ROER4D C&D strategy supported internal document sharing within the network via collaborative platforms, as well as open online sharing of completed outputs with the public via websites, publishing platforms and repositories. Decisions about *what* to make open were determined in collaboration with researchers, and were contingent on contractual frameworks as well as ethical and consent factors. The decision around *when* to make content open was also taken in consultation with researchers, and was largely informed by researchers' individual publication ambitions and timeframes associated with UCT Network Hub quality assurance processes.

Overall, the ROER4D C&D strategy was informed by a four-pillared approach to openness (King, Hodgkinson-Williams, Willmers & Walji, 2016), namely: the development of a shared open culture (ideology of openness); opening up the research process (operational openness); producing outputs in open formats and using open platforms where possible (technical openness); and ensuring that at all forms of openness conformed with policy, grant and ethical provisions (legal openness). These aspects of openness were synergistic and mutually constitutive, and collectively manifested in the ROER4D C&D model.

The broad aims of the C&D strategy were as follows:

- To address infrastructure deficits and provide content management solutions (including content hosting) in a research community with uneven institutional support and capacity challenges.

- To ensure that the ROER4D legacy is freely accessible for long-term reuse in line with international curatorial and publishing standards.
- To provide an archival and publishing service to sub-project researchers and the UCT Network Hub.
- To support the Principal Investigator (PI) and Deputy PI as well as sub-project researchers in editorial and research capacity development processes.
- To complement the UCT Network Hub communications efforts in an integrated communications/dissemination strategy.
- To facilitate and evolve the project's open research approach.

Within this context, the principle areas of C&D activity entailed:

- Curatorial support in the management and organisation of content produced by the UCT Network Hub and sub-projects.
- Establishment and maintenance of various forms of e-infrastructure required for preservation and dissemination services.
- Editorial development and support in research reporting and online publication.
- Production and publication of a final ROER4D edited volume.
- Data publication support to sub-project researchers through the project's open data initiative.

Within this overall strategy, openness was conceived as a means to participation, impact and preserving legacy.

Contractual framework and stakeholder analysis

An important first phase in the project's role of publisher was the assessment of the contractual framework. The overarching Memorandum of Grant Conditions (MGC), sub-grant agreements, organisational policy frameworks, as well as ethics approval and consent processes all needed to be assessed in order to gain a sense of the UCT Network Hub's and researchers' rights, limitations and commitments in terms of content management and publication activity. MGCs will often make stipulations around privacy requirements that must be adhered to (particularly in terms of data transfer and publication), and copyright or licensing provisions are also typically defined within this framework.

As a first step in articulating a C&D strategy, the ROER4D Curation and Publishing Manager undertook a detailed analysis of the MGC and sub-grant agreements to establish the parameters in which publishing activity was to take place, paying attention to the stipulations of the funder as well as the various sub-project host entities. One sub-project host, for instance, insisted on adherence to their organisational policy of publishing under a Creative Commons Attribution-ShareAlike (CC BY-SA) licence, rather than the project's CC BY default. This stipulation was integrated into the publishing strategy for that sub-project's outputs.

The organisational policy frameworks of participating researchers were important to consider; these typically included those institutions' intellectual property policies, policies governing activity related to institutional repository and other public platform sharing, and research data management policies or guidelines. In the ROER4D context, institutional and funder requirements were carefully brokered by contracts officers and captured in sub-grant agreements.

Once the contractual framework pertaining to C&D activity was made explicit, a needs analysis was undertaken to establish the focus of C&D activity and identify where resources should be focused. This needs analysis took into consideration the needs and desires of sub-project researchers, funders, host organisations and associated stakeholders. It also identified who the key envisioned audiences for the research outputs were, and which formats or output genres were most conducive to engaging these groups.

In articulating a strategy, it was also important for the UCT Network Hub to consider the infrastructure (hardware, software, connectivity, etc.) required to meet project objectives, and to establish or share infrastructure in cases where there were deficits. In cases where there was an absence of institutional curatorial systems, third-party repositories and sharing platforms such as Zenodo and Figshare were utilised. The project also purchased storage space in a secure UCT-hosted archive in which project outputs and documentation were stored for a period of at least five years after project completion. Determining levels of internet connectivity and digital proficiency amongst researchers were key considerations in terms of defining a workable archiving and preservation solution.

In the ROER4D project, researchers were placed at the centre of publishing activity and their partnership and collaboration with the UCT Network Hub was crucial in terms of publishing project outputs. The curatorial process was crucial in order to facilitate collaboration around content creation and sharing. Significant resources were also devoted to working with researchers in a developmental editing process and in providing other forms of support in the manuscript development process (such as data presentation and reference management) in order to build capacity and address quality assurance concerns.

The edited volume publishing process

Within the context of the ROER4D publishing process, certain publication components such as blogs and announcements (shared via the project website and social media platforms) were typically more rapid and informal in terms of release, while more formal scholarly outputs and data publication (typically shared via repositories and data portals) required longer timeframes and more extensive investment. That said, all content released by the project underwent quality assurance processes which addressed features of substance and presentation.

The primary output which arose from the ROER4D publishing process was the peer-reviewed edited volume, *Adoption and Impact of OER in the Global South* (Hodgkinson-Williams & Arinto, 2017). This work, comprised of 16 chapters with seven accompanying open datasets, was co-published with African Minds open access publishers and Canada's International Development Research Centre (IDRC). The content development and peer review processes entailed in this publishing process were undertaken by the UCT Network Hub in collaboration with African Minds and participating researchers. Traditionally, this might be regarded as a "self-publishing" approach, but the ROER4D publishing model was in fact a hybrid of traditional publishing mechanisms (as manifest in the co-publishing arrangement with external publishing partners) and an open research approach (manifest in collaborative peer review and an advance online release strategy).

The process of publishing the ROER4D edited volume was comprised of five main components: content development, peer review, production, advance online release, and consolidated release and dissemination.

Content development

A central feature of the ROER4D research agenda was the expectation that sub-projects would produce interim and final research reports summarising their research processes and findings. In line with the project's open research approach, there was initially an ambition to make these reports openly available online. It soon became clear, however, that the ROER4D research reporting process was a first step in the content development process that was aimed at profiling OER research from the Global South and that more could be gained from working with authors to distil their findings and develop research reports into publishable, peer reviewed chapters.

It was the Network Hub's belief that this need for guidance and support from a publisher or other supportive entity before making content open was not solely endemic to the ROER4D researcher community, but that most academics require some assistance in the "packaging" and presentation of their work – particularly when there is a diverse academic and non-academic audience being targeted. The Publishing and Curation Manager in conjunction with the PI and Deputy PI therefore played a central role in working with researchers to ensure cohesion in the expression of their argument, refining claims to corresponding data and ensuring that protocols of scholarly expression were adhered to. The process followed was similar to the publishing process referred to as "developmental editing" (Norton, 2009), and was accompanied by detailed editorial assistance in terms of proofreading and reference checking.

While the need for this kind of support is commonplace, there was a particular imperative around this support in terms of the Global South researcher context – largely because of the desire for this research to be able to compete for the attention of an international audience, which, among other things, meant publishing in English when English was the third or fourth language of many researchers in the ROER4D cohort. It was also important to acknowledge that different geographical regions had differing scholarly conventions in terms of how an argument was presented and substantiated.

The developmental editing process was therefore a central quality assurance mechanism as well a means to obtain some level of coherence and uniformity across all 18 sub-projects' outputs. This work was undertaken with an acknowledgement of the problematic and deeply hegemonised notions of "quality", "professionalisation" and "academic convention" which are rooted in a Western, Anglocentric scholarly paradigm. The content development process was therefore focused on allowing the voice of the author to resound clearly in their work while adhering to established (Western) norms and conventions of presenting scholarly work. It was approached with respect and sensitivity, always acknowledging the author as the expert in their context while allowing for valuable research capacity building to take place in order to upskill researchers in terms of participating in a professionalised publishing process.

The content development process was iterative and guided researchers through the process of moving from an interim to final research report, ultimately manifesting in a chapter draft which would constitute that sub-project's contribution to the edited volume. Within this process, the Publishing and Curation Manager edited reports and chapter drafts as a first layer, addressing aspects of grammar, coherence and citation; while the PI and Deputy PI undertook a second layer of mentorship and content development, addressing theoretical and conceptual factors. Queries arising

from this process were shared with the relevant research teams in Google Docs in order to promote dialogue around the issues raised, and Skype calls were conducted to discuss high-level issues and next steps in the editorial development process.

The content development process was arduous and time consuming for both researchers and members the Network Hub, but was ultimately seen as a valuable mechanism for not only developing the quality of the content eventually shared openly online, but also for boosting the confidence of researchers from developing country contexts who felt challenged in terms of their participation in English-dominated scholarly environments. The content development process culminated in a chapter manuscript draft, which was peer reviewed for publication.

Peer review

Peer review is widely viewed as a central means of quality assurance. Many institutions and research organisations demand evidence of an independently administered peer review process in order to recognise publication in formal appraisal systems, and publication of peer-reviewed work is a major aspiration in most academics' career trajectories. It has, however, also been heavily critiqued and is viewed by many as being "badly in need of repair".[1]

The ROER4D developmental editing process provided an opportunity to explore the interrelationship between quality assurance and research capacity development. The project approach to peer review of its edited volume was an extension of this desire to have a safe, transparent and collaborative process through which the quality of content could be improved by close scrutiny of experts in the field. It was also necessary to have a dynamic system which could be staggered in terms of the editorial development process (not all chapters were ready for peer review at the same time) and the varying degrees of revision required to different chapters. The project therefore administered peer review of the edited volume chapters as part of its centralised C&D strategy, collaborating with African Minds publishers on mechanisms to demonstrate the rigorousness of the process.

Each chapter was reviewed by two reviewers identified by the Publishing and Curation Manager in collaboration with the PIs and the lead authors of each chapter. Authors were included in all correspondence with reviewers and the process was framed as a conversation, with various documents and tracking systems established to record the comments from reviewers (which, in many cases, were extensive) along with the authors' responses and subsequent edits to their chapters. These tracking mechanisms were necessary in order to demonstrate the credibility and thoroughness of the review process. They were not made public, but are available for scrutiny by external parties should any queries around the review process arise.

The peer review process constituted a third layer of quality assurance within the edited volume publishing process. In addition to the formal peer review process by two external reviewers, many authors also had draft chapters reviewed by colleagues within their own organisations and made refinements to the chapters at each stage of feedback. This multi-layered process was valuable not only in terms of addressing cogency of argument and validity of claims, but also in refining aspects of presentation (such as cutting back on figures in cases where readers complained of "data fatigue").

---

1   http://theconversation.com/the-peer-review-system-for-academic-papers-is-badly-in-need-of-repair-72669

While these multiple rounds of refinement were a challenge for authors – particularly as the publishing work took place outside of the formal grant period in which they were compensated for their time – the process was invaluable, not only in terms of boosting the rigorousness of the work presented, but also in terms of research capacity building.

The "conversation" aspect of the peer review process was more successful in some cases than others, but there were instances of vibrant interaction between authors and reviewers where finer points of refinement as well as high-level issues pertaining to the discipline were discussed. In some cases, reviewers also indicated that the work reviewed (and the interaction with the authors) fed into the research work in which they were engaged.

The ROER4D peer review process was necessarily staggered in that all draft chapters were not simultaneously sent out for review. This enabled the UCT Network Hub to run multiple aspects of the publishing process concurrently, sending certain chapters out for review while others were still being written – a strategy which proved crucial in terms of managing a publishing process under the restrictive timeframes of the overarching grant period.

Every aspect of the quality assurance process, from developmental editing to peer review review, was undertaken in collaboration and consultation with authors at a time when both they and the ROER4D edited volume editors considered the content ready for review. This level of flexibility and responsiveness to the project context may have been a challenge for an external publisher to navigate, and would have significantly extended the length of time required to administer the review process. The development aspect of the review process may also not have been as effective if the process was undertaken by a third party in a blind peer review process.

Production

The ROER4D project was determined that the edited volume should be a quality product, both in terms of substance and presentation. This was deemed crucial in terms of amplifying the Global South empirical contribution and the volume serving as an output that could advance the work and reputation of participating researchers.

This presentation layer, along with the need for endorsement of the peer review process, was one of the main reasons why the project entered into a co-publishing relationship with African Minds. African Minds is an open access, not-for-profit publisher, which publishes predominantly in the social sciences. Its authors are typically African academics or organisations, and it offers innovative approaches to those "frustrated by a lack of support from traditional publishers or by their anachronistic approach to making research available".[2] The publisher ethos aligned closely with the ROER4D open approach and there were clear synergies in terms of the mutual desire to innovate in a more inclusive, progressive publishing approach.

African Minds is a member of the Open Access Scholarly Publishers' Association and operates on a model where the publishing cost is partially or fully subsidised by authors, projects or their sponsors. Publication decisions are taken by an editorial review board and it operates on a cost-recovery model where prices are set as low as possible in order to ensure maximum reach and visibility in an international marketplace. In

---

2   http://www.africanminds.co.za/

addition to production and editorial services, it also provides distribution, print-on-demand and e-book production services through third-party vendors. Its international distribution service, which includes libraries in its customer base, was a crucial factor for ROER4D in terms of obtaining maximum international reach and visibility of the edited volume title. African Minds therefore provided text design, typesetting and proofreading, as well as peer-review certification and distribution services to ROER4D.

Midway into the negotiations with African Minds, the IDRC expressed a desire to come on board as a third co-publisher.[3] This was viewed as being highly beneficial to the project in terms of the title being included on the IDRC list and carrying the funder's imprint. There was also considerable benefit in terms of the additional distribution of print copies, which the IDRC undertakes to its own mailing list. A cost-sharing arrangement was therefore entered into and a tri-party co-publishing arrangement was forged.

Within this tri-party arrangement, the project also had the benefit of the IDRC Publisher's expertise. Workflows were expanded to incorporate the IDRC as a third stakeholder in the publishing process and the IDRC Publisher provided advice and guidance on all aspects of the publishing process, ranging from text design to licensing issues and distribution. Most crucially, she also provided an additional quality assurance layer, vetting each of the chapters for publication and providing detailed editorial feedback to boost the overall quality and cohesiveness of the volume.

Advance online release

Time lag is a challenge in any book production process. Given the staggered nature of the ROER4D sub-grant periods, some sub-projects were wrapping up on research processes when others were just getting into instrument design and data collection. There was, however, an ambition to capture all the sub-project outputs in a single collection in order to amplify the voice and message from the Global South. The thought of having chapters "gathering dust on a shelf" while others were being finalised was problematic for the project. In order to counter this, the project agreed with co-publishing partners that it would pursue an advance online release strategy.

In an open research context, advance online release could be taken to refer to the release of work-in-progress. This was not the ROER4D strategy. As already stated, the project felt it important to administer the requisite quality assurance processes prior to making content openly available. The project therefore released peer reviewed, typeset and proofread chapters that were vetted by all co-publishing partners as they became available over a five-month period leading up to the publication of the full volume.

Citation of advance online material can be tricky in a formal publishing process as it introduces the potential for confusion around the final version of record and how to cite the content. In the ROER4D context, there were no changes to text between interim and final release and all interim and final chapters carried details on how to cite the work along with unique digital object identifiers pointing to interim and final versions, which could be pre-reserved via the Zenodo content sharing platform (see Figure 1 for an example of an example of an advance online release citation guide). In this sense, it

---

3   The IDRC has a long-standing tradition of functioning as a publishing partner to its research projects, in many cases partnering with other publishers. For a full list of its publications, see: https://www.idrc.ca/en/resources/publications.

was more like a journal article "post-print" sharing process, although chapters were finalised in terms of editorial and typesetting work by this stage.

The advance online release strategy was an important mechanism for enabling authors and the UCT Network Hub to promote chapters while the full volume was being prepared for publication.

How to cite this resource:

Cox, G. & Trotter, H. (2017). Factors shaping lectures' adoption of OER at three South African universities. In C. Hodgkinson-Williams & P. B. Arinto (Eds.) *Adoption and Impact of OER in the Global South*. Chapter 9 advance publication. DOI: 10.5281/zenodo.154562

Corresponding author: Glenda Cox <glenda.cox@uct.ac.za>

Figure 1: **Example of advance online release citation guide in edited volume chapters**

Consolidated release and dissemination

The full ROER4D edited volume was published in December 2017. An open access full-colour version was made available online in PDF, while a black-and-white print version was available for sale via print-on-demand through the African Minds website[4] and Amazon.com[5]. The decision to make both online and print versions available was prompted by findings from ROER4D research which indicated that many teachers and academics in developing countries (particularly those operating under connectivity constraints) still felt it crucial to have access to print materials or digital content that could be accessed offline.

Throughout the ROER4D project period, the C&D strategy was complemented by the project's communications strategy, which was aimed at boosting the profile of the ROER4D network and the visibility of its outputs. ROER4D made extensive use of social media channels, and operated on a twofold approach in which both the Network Hub and individual researchers interacted with a wide range of audiences through a range of channels.

Upon publication of the consolidated volume, the project embarked on a communications and marketing campaign, leveraging the networks and communication channels established in the course of the project in order to boost the reach and impact of the research findings. Communications activity included dissemination via blogs, posts to social media channels, conference presentations and articles in

---

4   http://www.africanminds.co.za/dd-product/adoption-and-impact-of-oer-in-the-global-south/
5   https://www.amazon.com/Adoption-impact-OER-Global-South/dp/1928331483/ref=sr_1_1?ie=UTF8&qid=1
     518091576&sr=8-1&keywords=adoption+and+impact+of+oer+in+the+global+south

the press. There was also a focused research-to-policy strand of activity focusing on leveraging findings in a policy context.

An overview of the edited volume publishing process is presented in Figure 2.
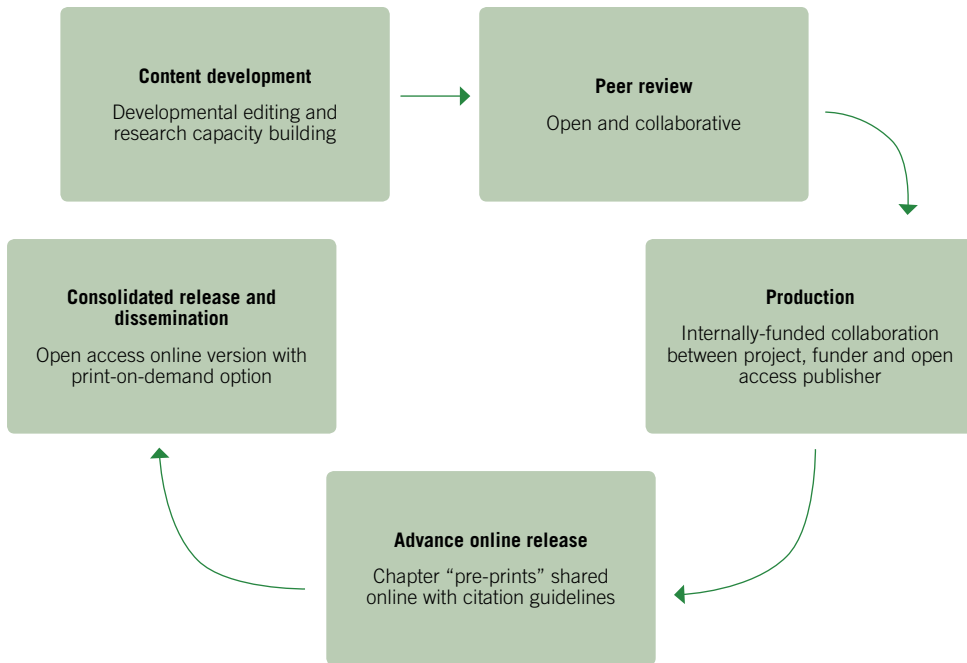
Content development

Developmental editing and research capacity building

Peer review

Open and collaborative

Consolidated release and dissemination

Open access online version with print-on-demand option

Production

Internally-funded collaboration between project, funder and open access publisher

Advance online release

Chapter "pre-prints" shared online with citation guidelines

Figure 2: Overview of the ROER4D edited volume publishing process

The ROER4D open data initiative

ROER4D has been recognised for its innovative approach to open data publication and in 2018 received the Open Education Consortium's Open Resources, Tools and Practices Award in the "Open Data" category.[6]

Seven of the chapters in the edited volume were published with accompanying open datasets. Six of these were published by the C&D team on the DataFirst Data Portal[7], while one (associated with Chapter 5 in the volume, "Co-creation of OER by teachers and teacher educators in Colombia") was published in Spanish by the organisational host for that particular sub-project.[8] An eighth open dataset, arising from the meta-synthesis, was published by the C&D team on the DataFirst Data Portal after the release of the edited volume.

Open data publishing was a central component of the ROER4D open research approach and the C&D strategy. The ROER4D open data initiative was envisaged as performing two principal functions:

1. Enhancing research quality by curating data produced by sub-project researchers, ensuring that it is appropriately presented, described and accessible for reuse.

---

6   http://www.oeconsortium.org/2018/03/oec-announces-the-2018-oe-award-winners-of-open-resources-tools-practices/

7   https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/ROER4D

8   https://karisma.org.co/cokrea/?p=1007

2. Supporting the development of an empirical baseline of OER research by accompanying published research outputs with the data that supports the findings presented in the sub-project studies.

Participation in the open data initiative was voluntary for sub-project researchers, as open data publication was not mandated in the original project scoping or contractual agreements. While the curatorial services provided by the project were a relatively uncontentious area, open data publishing was a novel concept for many researchers and there was a certain amount of reluctance in terms of the substantial investment required in terms of the time and expertise needed in order to prepare datasets for publication.

The data curation and publication process was viewed by the UCT Network Hub as part its synergistic open strategy, in the sense that the open research approach adopted by the project takes the position that all open activities have the potential to enhance rigour.[9] Given the imperative to produce an empirical baseline of OER research in the Global South, data publication was also viewed as a key component in promoting long-term reuse and meta-level synthesis of ROER4D research, both within the project and by third parties.

Data management principles

As already mentioned, C&D activity overall (and open data publication, in particular) was not part of the original project formulation, and was implemented retrospectively in terms of sub-project research processes (in many cases after research was completed). There was therefore no explicit data management planning in the project scoping process. That said, project leadership did have the foresight to reference data publication in the original Subgrant Agreement between the project host (UCT) and sub-project host organisations. Clause 13.2 states:

> *... the Parties acknowledge that the sharing of data and the dissemination of the research results advance the state of knowledge in the field [and] is in the common interest of both Parties. Accordingly, the Parties agree to share data and co-operate with each other and to publish jointly when appropriate.*

Data sharing was seen as a natural extension of open content sharing, and was regulated by the same copyright and licensing provisions that pertained to other outputs. Data management planning took place retrospectively in late 2015 when the project's open data initiative got underway. Given the idiosyncrasies of the various sub-project contexts and differing nature of the studies, the project adopted a bifurcated approach to data management planning: a data management plan (DMP) was articulated at Network Hub level to capture high-level, coordinating principles; and sub-project-specific DMPs were developed for individual sub-project datasets when authors indicated an interest in publication.

The project adopted the Digital Curation Centre DMP template[10], with some modifications, and conceived of the DMP as a key mechanism to facilitate data publication and engage

---

9   For a more detailed exploration of the relationship between open data practice and rigour in the ROER4D research process, see the blog post on "The ROER4D open data initiative: Some lessons learned": http://roer4d.org/3575.

10 http://www.dcc.ac.uk/resources/data-management-plans

authors in thinking about key aspects of data sharing activity. In retrospect, the C&D team would recommend expanding the scope of data management planning to not only address external data sharing, but also to establish the norms and conventions through which sub-project data are shared internally. Within a large network project context, it is important not only to establish the protocols for open data publication, but also to make explicit the principles governing sharing data between sub-projects as well as sub-project data sharing with the Network Hub and PIs. Figure 3 provides an overview of the ROER4D data flow from sub-project to Network Hub to publication, differentiating between internal (in many cases, restricted) data exchange and external (open) data publication.
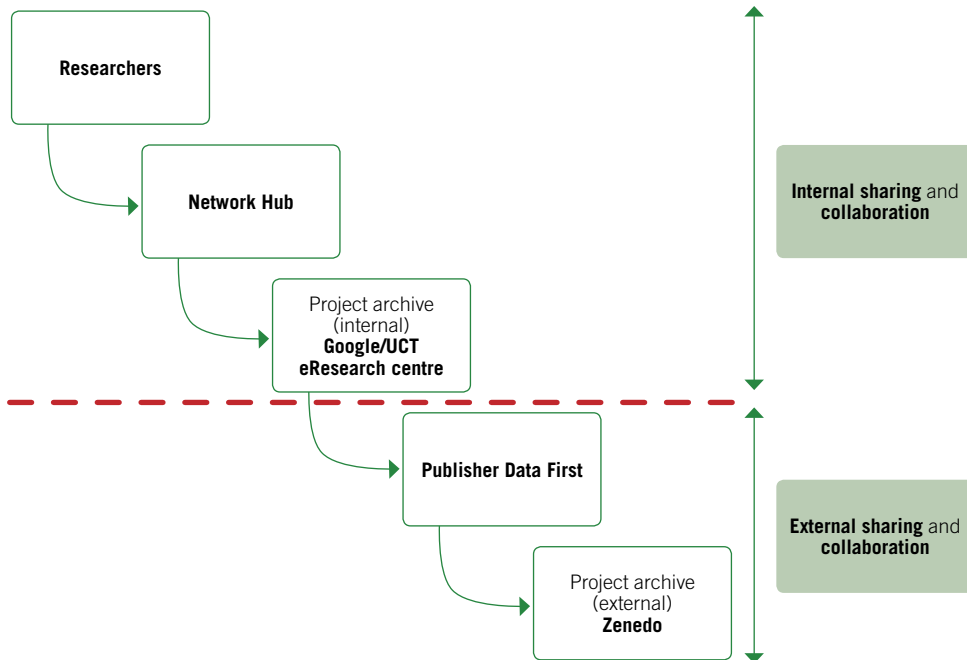

Figure 3: Overview of the ROER4D project data flow

Data interrogation

Given the contractual provisions and qualitative or mixed-methods nature of much of the ROER4D sub-project datasets, the C&D team articulated a data interrogation approach that balanced ethical considerations with the commitment to publishing valuable data for reuse.

Data interrogation began with a review of the sub-project ethics and consent processes, with a particular focus on whether the wording of the consent form or process precluded open data publication. The use of phrases such as "data will be used for research processes only" or "data will be destroyed after the project ends", which are typically copy-and-paste remnants from previous studies, were problematic in the ROER4D context and prevented data-sharing in certain sub-projects. Thorough interrogation of the language used and the processes followed in gathering consent was therefore crucial when undertaking data publication retrospectively. Only once consent processes were investigated could a decision be made on what data was legally and ethically shareable within a specific sub-project's context.

The contractual provisions of the ROER4D project clearly stated that at no point could "raw" sub-project data be shared with the Network Hub. The first step in the de-identification process was therefore undertaken by sub-project researchers. This usually consisted of a high-level process in which direct identifiers (names, email addresses and identification numbers) were removed from a dataset.

If the consent process did allow for open sharing and the first-level de-identification was conducted, the C&D team requested a copy of the data files and associated research instruments, which were stored in the secure project archive. The C&D team then performed a thorough interrogation of the data and began organising it into a form that a third-party reader could interpret. Data files supplied by researchers often came in the form of a "data dump", characterised by a lack of structure, irregular and cryptic file-naming conventions, and other opaque organising principles. The C&D team undertook a process of sorting the data into a structure more easily understood by external users, in the process renaming files, usually according to their collection method (survey, interview, etc.) and their function (whether instrument or microdata). During this process, microdata were also converted to a range of formats in order to maximise their potential for reuse. Researchers were contacted to ensure that the data files did not contain any unshareable third-party copyright, such as extant data or visualisations included in the collection.

Once the files were reorganised, the microdata (focus group and interview transcripts, survey data, institutional record data, etc.) were mapped to the instrumentation to ensure that all microdata were supported by research instruments, and vice versa. If at this point data or instrumentation was absent from the collection, researchers were asked to provide them, or to provide a rationale for their absence.

The next phase involved the conceptualisation of the dataset. This process entailed a data-vetting process and decision-making on which elements of the collected set to include or exclude. The preference overall was to publish as much data as legally and ethically possible, even if not all collected data was used in a sub-projects' findings and analyses. All such data vetting was performed in partnership with the contributing researchers. Once an agreement was reached regarding which data would be shared, the de-identification process began.

De-identification principles

De-identification is the revision, redaction or elision of information within microdata that can be used to identify research participants and, in special cases, referents. It is a crucial security mechanism in terms of protecting the identity of research participants and the integrity of the researcher. As the "long tail" of open data sharing is predicated on the diverse ways in which external users can analyse data, de-identification should remove as little as is ethical in order to ensure the richness of the dataset. The purpose of de-identification is to preserve confidentiality or anonymity, not to compromise the quality of the prepared data through over-zealous redaction.

The ROER4D de-identification approach was based on two fundamental principles: a commitment to consensual, ethical and legal open practice, and a commitment to sharing comprehensive datasets that were well described and facilitated reuse. The ethical commitment, expressed as "First do no harm", was based on adherence to the ethical and consent aspects applicable to the research and data-sharing processes. In terms of de-identification, this involved removing as much as was needed to protect the

identity of the research participants. This was balanced by the need to produce a dataset that was seen to have value and utility to an external audience.

Mixed-methods approaches dominated the ROER4D research context, with the majority of the 18 sub-projects using some combination of qualitative and quantitative approaches for data collection. Most educational data publishers appear to work primarily with quantitative data (though some, such as the UK Data Service[11], do curate qualitative data). ROER4D's data publisher, DataFirst, primarily curates socio-economic quantitative data, such as South Africa's National Income Dynamics Survey[12], and their curatorial and metadata approaches have been designed for large-scale quantitative studies and largely deal with data in tabular form – prior to the fourth quarter of 2016, the platform would only accept data if it was formatted in one of the accepted tabular formats (CSV, XLS, SAS, SPSS or STATA).

The mixed-methods nature of much of the ROER4D data was one of the primary challenges associated with preparing ROER4D data for publication. Many of the established protocols and platforms for data sharing have traditionally been focused on quantitative data. As such, quantitative de-identification methods and metadata schema appear to be more established than their qualitative counterparts.

With qualitative data, the de-identification process is based on the meticulous, thorough and repeated reading of the data to determine what connections can be made between utterances that would enable identification of respondents or other individuals. Within the ROER4D context, there was no clear distinction between data cleaning, validation and de-identification processes in data processing activity; the repeated reading of the text uncovered all manner of issues – from typos and irregular use of clarifying questions to missing data elements. The re-reading process built familiarity with the dataset, which was particularly useful for identifying the less obvious secondary identifiers that, in conjunction with other data, could be used to identify a research participant. The data interrogation process is illustrated in Figure 4.
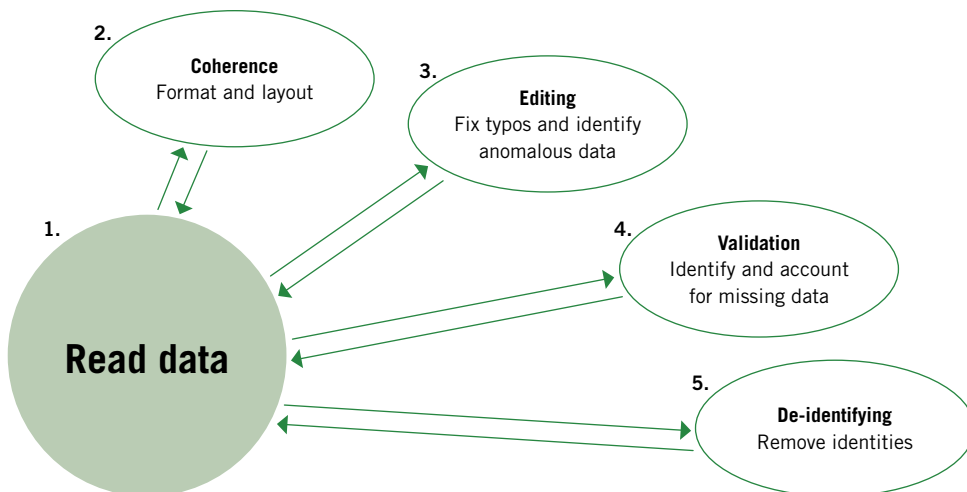

Figure 4: The ROER4D data interrogation and de-identification process

---

11   https://www.ukdataservice.ac.uk/get-data/key-data/qualitative-and-mixed-methods-data
12   http://www.nids.uct.ac.za/

Once potentially disclosive information was identified, the C&D team employed either *redaction* (replacing the content with "[redacted]" or a similar marker) or *revision* (restructuring sentences to de-identify participants or referents). In order to preserve the integrity of the dataset, the C&D team adopted redaction as the primary de-identification technique, as it was felt to be the most transparent means of ensuring anonymity/confidentiality by indicating where de-identification had occurred. Revision was used when dealing with non-English phrases and idioms in English-language data and in cases where full redaction would compromise the quality of the data.

The unit of analysis within each project functioned as the limit of all de-identification processes; that is, if regional comparison formed a key part of the analysis, identifiers below the regional level could be removed without compromising the integrity of the data. The overarching principle of removing as few identifiers as possible to ensure the confidentiality or anonymity of respondents while preserving the integrity of the dataset guided all activity in this regard.

Other than the complexities of interrogating mixed-methods data, handling non-English data constituted another significant challenge. ROER4D sub-projects operated in a variety of linguistic contexts, and therefore generated large amounts of data in languages other than English. DataFirst, the project's data publisher, did however require that each published dataset include an English language component. While it is possible to translate quantitative datasets which are based on controlled vocabularies into English, the C&D team did not have the capacity to do so. The ROER4D collection on the DataFirst Data Portal contains one dataset with both English and Spanish text.

Data publication

The ROER4D data publication approach draws from Maxwell's assertion that the "qualitative–quantitative divide is a social construction that has been reified well beyond its legitimacy and usefulness" (2010, p.477). While that quote perhaps overstates the case, it does support the idea that qualitative data often contains an implicit quantitative or numerical component; thus the coding process does not have to be seen as an imposition from a quantitative paradigm, but as a legitimate quality-enhancement process. In published ROER4D datasets, the C&D team utilised a simple tabularisation of research questions and their attendant responses to fulfil this requirement to preserve the "process theory" (Maxwell, 2010, p.477) approach characteristic of qualitative data, and avoided trying to convert the transcripts into lists of variables. Serendipitously, the published sub-project 3[13] and sub-project 4[14] datasets used structured interview instruments to gather their qualitative component, which produced data that was relatively easily to represent tabularly as well as reproduce in a textual format (RTF).

DataFirst specialists reviewed the data once it was submitted by the C&D team and performed an additional quality assurance process. This took the form of identifying and querying apparently aberrant data entries (such as incorrect variables and disjuncts between the microdata and metadata), querying null fields (empty cells in quantitative

---

13   https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/578
14   https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/555

data files) and ensuring that the variable descriptions provided in the metadata matched those in the instruments. As the ROER4D de-identification and validation process focused largely on qualitative data, DataFirst's expertise in quantitative data management provided a valuable additional layer of quality assurance based on over two decades of quantitative data management.

As a final step in the data publication process, DataFirst converted the microdata files into CSV, SPSS, SAS and STATA formats in order to ensure optimal accessibility and reuse, and published the dataset. The published dataset (the microdata and associated instrumentation) was accompanied by comprehensive metadata, including a description of the research process and a detailed description of the de-identification process.

Conclusion

The ROER4D C&D approach was largely informed by the project's operational definition of open research: "conducting and sharing research in which a selection of research proposals, work-process documents, literature reviews, methodologies, research instruments, analytical frameworks, findings and/or data are intentionally shared on publicly accessible platforms in order for others to freely access, use, modify, and share them subject to measures that preserve ethical practice and legal provenance" (King et al., 2016, p.82).

Many proponents of the open research approach argue for an immediacy principle, whereby the time delays of traditional peer-reviewed publication are supplanted by a more rapid approach to content release and online audiences play a role in verification and quality assurance. The ROER4D C&D model presented a slightly more measured approach; one that was guided by the principle of doing no harm and ensuring that the interests of researchers and content authors always came first. Content was not always released as rapidly or as openly as originally envisioned, but the Network Hub repeatedly saw the benefits of a measured, strategic approach in terms of boosting rigour and building researcher confidence.

The UCT Network Hub has continued to publish conceptual tools, synthesis data, project guidelines and working documents after the project end date in a wide range of formats for a diverse audience base. Within these publishing processes, curation of content and editorial management continue to play a crucial role, as does the partnership of intermediaries and publishing partners. The project's C&D model does not purport to replace traditional systems of quality assurance and publication, but rather makes a start at innovating a new approach to curation and publication in which the project is a central, coordinating entity where dissemination mechanisms can develop in an agile, iterative fashion as the project context evolves.

C&D activity requires support, capacity development and resourcing. It also requires acknowledgement of the need for new curatorial and publishing skills to be incorporated in the research lifecycle in order for optimal reach and impact to be achieved. Ultimately, a professional approach towards C&D promotes a professional approach towards the research endeavour as a whole and is a crucial mechanism for researchers from marginalised communities to make their voice heard on the global stage.

References

Hodgkinson-Williams, C. & Arinto, P. B. (2017). *Adoption and impact of OER in the Global South*. Cape Town & Ottawa: African Minds, International Development Research Centre & Research on Open Educational Resources. Retrieved from https://doi.org/10.5281/zenodo.1005330

King, T., Hodgkinson-Williams, C., Willmers, M. & Walji, S. (2016). Dimensions of open research: Reflections on 'critical openness' in the ROER4D project. *Open Praxis, 8(2)*, 81–91. Retrieved from http://www.openpraxis.org/index.php/OpenPraxis/article/view/285

Maxwell, J. A. (2010). Using Numbers in Qualitative Research. *Qualitative Inquiry, 20(10)*, 1–8.

Norton, S. (2009). *Developmental editing: A handbook for freelancers, authors, and publishers*. Chicago: University of Chicago Press.