

Endangered African Languages Featured in a Digital Collection: The Case of the #Khomani San | Hugh Brody Collection

Kerry Jones, Sanjin Muftic

Director, African Tongue, Linguistics Consultancy, Cape Town, South Africa
Postdoctoral Research Fellow, English Language and Linguistics, Rhodes University, Makhanda, South Africa
Research Associate, Department of General Linguistics, Stellenbosch University, Stellenbosch, South Africa;
Digital Scholarship Specialist, Digital Library Services, University of Cape Town, South Africa
jonesleekerry@gmail.com, sanjin.muftic@uct.ac.za

Abstract

The #Khomani San | Hugh Brody Collection features the voices and history of indigenous hunter gatherer descendants in three endangered languages namely, N|uu, Kora and Khoekhoe as well as a regional dialect of Afrikaans. A large component of this collection is audio-visual (legacy media) recordings of interviews conducted with members of the community by Hugh Brody and his colleagues between 1997 and 2012, referring as far back as the 1800s. The Digital Library Services team at the University of Cape Town aim to showcase the collection digitally on the UCT-wide Digital Collections platform, Ibali which runs on Omeka-S. In this paper we highlight the importance of such a collection in the context of South Africa, and the ethical steps that were taken to ensure the respect of the #Khomani San as their stories get uploaded onto a repository and become accessible to all. We will also feature some of the completed collection on Ibali and guide the reader through the organisation of the collection on the Omeka-S backend. Finally, we will outline our development process, from digitisation to repository publishing as well as present some of the challenges in data clean-up, the curation of legacy media, multi-lingual support, and site organisation.

Keywords: endangered African languages, N|uu, Kora, Khoekhoe, digital curation, online showcasing, heritage knowledge, ethics of repositories

1. Introduction

Language endangerment and language loss is a worldwide phenomenon and the African context is no exception to this loss of linguistic diversity. As a result, the scramble to identify, document and preserve indigenous languages using digital technology has gained traction. Despite vast research in the field of language vitality, “relatively little is known about Africa’s endangered languages” (Kandybowicz & Torrence, 2017). In Southern Africa, identification and documentation of endangered languages in collaboration with indigenous communities has only fairly recently begun. Records of such efforts are dispersed in various locations around the world. In the instance of the #Khomani San | Hugh Brody Collection, South Africa is fortunate to have secured the return of this valuable collection. Upon its return, we were then faced with the challenge of making the contents of the collection freely available and not reserved for the select few.

The potential solution to this problem was to host the collection on a digital platform. The Digital Library Services (DLS) at the University of Cape Town (UCT) suggested the use of their Digital Collections platform, Ibali, which runs on Omeka-S and could accommodate open access. The digital curation process required a number of different steps, from setting up a workflow, to organising the collection data within particular schemas. Once the data was organised within the predetermined schemas, the data could then be hosted on the platform in a structured manner. In this paper we discuss the different phases of collection development with a particular focus on the processes undertaken to publish the collection through an

institutional repository. Throughout these processes we encountered many questions, challenges, debates and potential solutions revolving around how to best digitally curate this collection in an ethical and user-friendly way. This collection serves as an important example of representing indigenous knowledge via a digital platform. Furthermore, it describes a methodology to decolonise the archive while maintaining international standards.

By curating an output that is accessible to speakers of endangered languages, the South African public and academics, we created a more inclusive online environment for understanding our historical and contemporary South African context. It must however be noted that lack of access to a reliable internet connection and electricity does preclude many South Africans from being able to readily access online content. Households in South Africa with the least access to communication media are from the Northern Cape (10.3%), which is also the province where the majority of N|uu, Kora and Khoekhoe speakers live (Statistics South Africa, 2018, 36). Such challenges are overcome through access to resources via state and privately funded libraries and computer centres.

2. Ethical Considerations and Community Collaboration

The choice to house and curate the collection in South Africa was an ethical decision made by Hugh Brody in collaboration with Open Channels, a charity organisation based in the United Kingdom, the #Khomani San Community and the South African San Institute (SASI), a local non-government organisation. By physically hosting

the collection locally, this makes its contents more accessible to South Africans as opposed to having the collection hosted at an overseas institution. International travel is financially out of reach to the majority of the South African population, most especially minority groups such as the descendants of the #Khomani San. The collaboration between all interested parties and UCT was expressed in a Gift Agreement outlining the contents of the collection. As part of this agreement, UCT committed to processing, cataloguing and ensuring the collection was freely accessible. This collaboration and contractual agreement became the launching pad for future collaborations and work to come.

3. Linguistic Context of the Featured Languages in South Africa

In Southern Africa there are three language families that were previously grouped together and known as Khoesan languages (Heine & Honken, 2010). Today, Khoesan languages are more accurately described as Ju, Khoe and Tuu languages which is in accordance with the Comparative Method in linguistics (Güldemann, 2008, 123). All three of these language families are endangered and are associated with traditional hunter gatherers and pastoralists, also known as San and Khoe respectively. At the time of their initial documentation there were approximately 35 languages known to science from the Ju, Khoe and Tuu language families (Voßen, 2013). Since then, several of these languages have gone extinct, with only 13 remaining, of which 6 are spoken in South Africa (Jones, 2019).

Ju, Khoe and Tuu languages and their click sounds are a hallmark of Southern African linguistic and cultural heritage. Yet efforts made in research, development and preservation of Khoe and San heritage often falls short when it comes to meaningful community collaboration and accessibility to the content produced. Accountability and accessibility to historical and contemporary research pertaining to Khoe and San peoples is paramount to equity and democratisation of knowledge dissemination in Southern Africa. Khoe and San peoples have a documented history fraught with conflict, dispossession, and identity and language loss resulting in today's context as being minority marginalised groups scattered across Southern Africa.

The #Khomani San | Hugh Brody Collection documents the stories of the #Khomani people over the last 100 years entailing detailed accounts of linguistic and cultural genocide. Through a process of a cultural audit lead by Nigel Crawhall in collaboration with the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the SASI (Crawhall, 2001) 24 remaining speakers of N|uu made themselves known. N|uu was previously thought to be extinct and belongs to the Tuu or

more specifically the !Ui-Taa language family, our most fragile of the three families. Today, only three mother tongue speakers of N|uu remain and it is therefore classified as a severely endangered language.

Upon further engagement with community members of the #Khomani San it was revealed that some could speak Kora also known as !Ora as a second language. Recordings and transcriptions of such examples can be found in the #Khomani San | Hugh Brody Collection. Today Kora is considered moribund (du Plessis, 2018). It was also evident that speakers of N|uu had undergone language shifts to Afrikaans and Khoekhoe. Afrikaans is the historically dominant language during the apartheid era in South Africa and Namibia. Whereas, Khoekhoe also known as Khoekhoegowab, is the lingua franca of the Kalahari with approximately only 2000 speakers remaining in South Africa (Witzlack-Makarevich, 2006, 12).

The utilisation of data from the #Khomani San | Hugh Brody Collection was instrumental in several successful land claims in the Northern Cape, South Africa. It is therefore not only a historical and linguistic record of South African history but a unique collection of evidence that resulted in restitution for the #Khomani San.

4. Samples from the Collection

What is unique about the #Khomani San | Hugh Brody Collection is that it is based in southern Africa and was actively created with the #Khomani San community members and researchers not only during the creation of the data but the processing of it too. The #Khomani San | Hugh Brody Collection comprises mainly of audio-visual (legacy media) recordings of interviews conducted with members of the community by Hugh Brody and his team between 1997 and 2012. The initial project was motivated by the plight of the community who had been dispossessed from their ancestral land. Through genealogical mapping and detailed interviews, the team was able to support a successful land claim on behalf of the community in 1999, when President Mbeki personally visited to return the land to its rightful owners. Subsequent successful land claims in the area were to follow. Fieldwork continued until 2012, after which the collection was deposited in trust to UCT at Special Collections, UCT Libraries in 2013 (BVF-41 Project Plan). The collection consists of various data types such as transcripts, videos, audio clips, maps and images. Below are some examples of transcript files, photographs and maps that illustrate this highly collaborative data production process.

4.1 Transcript files

The transcript files in the collection are based on over 128 hours of film footage of which more than 30 hours include speech that has been transcribed to date. The transcriptions are verbatim, and colour coded in either N|uu (green), Kora

(orange), Khoekhoe (red) or Afrikaans (blue) and translated into English (black), including timing notations. The transcript files are colour coded to assist the reader in easily identifying the different languages transcribed and translated in each transcript file. The process of creating the transcript files was highly collaborative including linguists specialising in each of the languages found in the collection and hands-on verification with mother tongue speakers throughout the transcription and translation process. The challenge with working with a newly developed orthography for N|uu in the field of linguistics, resulted in the development of a multilingual dictionary featuring N|uu with accompanying translations into Afrikaans, Khoekhoe and English for over 1400 lexical entries (Sands et al., 2006).

Table 1 is an example from transcript 1998_01-01 as spoken by Katriena |Una Kassie Rooi where she mixes Afrikaans and N|uu when explaining who she is and where she comes from. For the purposes of a black and white publication the Afrikaans in the transcription column is in roman text and the N|uu in italics.

Time code: 00:00:05	
English translation	Transcription
Una: Yes, I am now a Bushman. I am a pure Bushman. My father was a Bushman. My mother was a Bushman. My father's father was a Bushman. That is the reason that I am a Bushman. I am Una. My father was #Han. My grandfather was old Hans Kassie. My father was Tities Kassie. He was old Hans Kassie's child. I am now Una who is the daughter of old Tities Kassie.	Una: Ja, <i>ng ke nou n Saasi</i> . Ek is n regte <i>Saasi</i> . <i>Ng ainki ke n Saasi</i> . <i>Ng xainki a ng Saasi</i> . <i>Ng ainki ke ng Saasi</i> . <i>Ke a gao ke ng Saasi</i> . <i>Ng ke ng Una</i> . <i>Ng ainki a n #Han</i> . <i>Ng oupa ng ou Hans Kassie</i> . Dis nou my pa is Tities Kassie. Dis ou Hans Kassie se kind. Ek is nou <i> Una</i> , wat ou Tities Kassie se dogter.

Table 1: Excerpt from transcript file 1998_01-01 in Afrikaans and N|uu translated into English

Table 2 is an example from the transcript 2001_01-04 as spoken by Anna Swarts as she tells a traditional story in Khoekhoe about the Jackal and the Hyena. She tells of a time before there was order in the universe and animals were humans.

Time code: 00:00:11	
English translation	Transcription
Anna: Thanks, my brother. My brother, it was like this.	Anna: Aio ti !gâ. Ti !gâ, nēti i ge ge ī i. !Nā, kō, #hīras tsi

When the jackal and the hyena were human beings... the jackal was actually the clever one. Then he said: "Oh, we are dying of hunger. So, what should we do?" The Hyena responded: "Do like this, do like this. You are the man, isn't it? So you should make a plan." Then the jackal went back and laid down in the road.	!gīrab tsīra ge khoe io, os ge... !Gīrab ge hūka ge gā-aisa ge ūhā i. Ob ge lība "Ēse, !ās xam ta lō. Om nī mati dī?" (#Hīras:...) Nē, nēti dī, nēti dī re, sats kom a aoreo, xuige satsa lawe-e dī re." Ob ge !gīriba oa tsi ge daob !nā sī ge !goe.
---	---

Table 2: Excerpt from transcript file 2001_01-04, traditional folklore told in Khoekhoe and translated into English

4.2 Photographs

Photographs from the collection range from the early 1900s to 1999. Those from the beginning of the 20th century were selected by researchers in the 1990s from UCT's archives, copied and taken along to the southern Kalahari. When interviewing community members, researchers would show the old photographs from UCT and explain that they were trying to learn more about the original people from the area. During this process, Elsie Vaalbooi recognised a photograph of herself (Figure 1) and another of her mother (Figure 2) taken circa 1911. This was the first time she had seen these photographs since they were taken. Elsie was one of the last speakers of N|uu and today her son, Petrus Vaalbooi (Figure 3), is the traditional leader of the #Khomani San.

The collection of photographs encompasses many themes such as: individual and family portraits; indigenous fauna and flora used by the #Khomani San; culturally or historically important places; language work (as seen in Figure 4); and physical evidence of occupation in the Kalahari Transfrontier Park by the #Khomani San before their eviction.

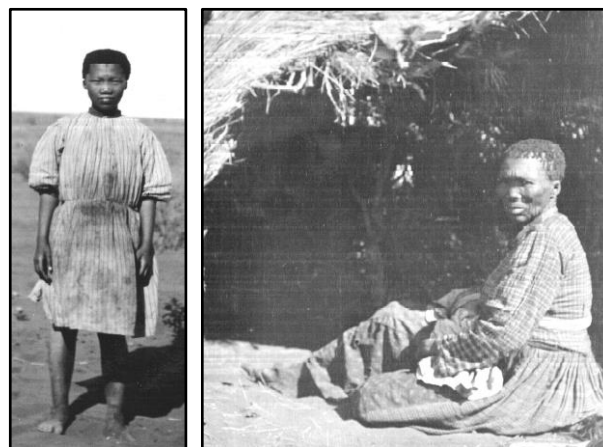


Figure 1 & 2: Photograph of Elsie Vaalbooi (standing) and her mother Marie ||Qoesi (sitting) circa 1911.



Figure 3: Photograph of Elsie Vaalbooi and her son Petrus Vaalbooi taken shortly before Elsie died in 1997.



Figure 4: Linguist Nigel Crawhall (sunglasses around his neck), working with Dawid Kruiper (hand pointing over page) and family on trilingual wordlists

4.3 Maps

Several detailed maps were created in collaboration with community members, linguists and cartographers to visually represent the #Khomani San genealogy and movements of different family members over time. The GIS data collected provided the necessary input to produce

accurate maps representing place names in Khoekhoe, Afrikaans and English as seen in Figure 5. The variety of Afrikaans spoken by the #Khomani San community members is unique to the area and provided great insight into the original place names of the southern Kalahari.

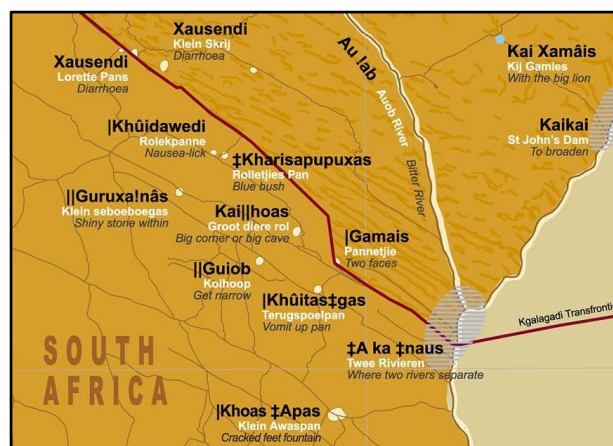


Figure 5: Map of traditional place names in Khoekhoe, Afrikaans and English within and surrounding the Kgalagadi Transfrontier Park

5. Organisation of the Collection on the Omeka-S Backend

The collection has gone through several phases of development, including the ongoing addition of detailed transcripts of the legacy media by specialist linguistics company African Tongue (Director, Kerry Jones). Elements of the collection were previously published online through an Islandora (Fedora Commons) website which provided basic archival description of selected media. However, the website did not provide a way of navigating the collection beyond those descriptions and therefore had to be revised. Consequently, the decision was made to publish the collection through UCT's institution wide collection repository, Ibali, powered by Omeka-S.

The online exhibition tool Omeka-S, in conjunction with the IIF server, is widely used internationally by university libraries, and the majority of the Galleries, Libraries, Archives, and Museums (GLAM) sector. Omeka-S allows for standards-led organisation and exhibition of digital materials. As opposed to most other web applications, it is not just a front-end to make content look appealing. It is also supported by archival metadata-driven back-end processes, enabling the re-use of materials via many other discovery platforms (Omeka-S Proposal Project Plan). Therefore, it is a repository and not a website because the backend of repositories allows for the creation of metadata and the creation of robust schemas that capture consistent metadata for each of the uploaded items.

5.1 Data Curation: Digitisation to Repository Publishing

As the material in the collection is over 20 years old, and UCT had gone through a phase of changes in personnel handling the digitisation process, as well as changes in the technologies employed for digitisation, a re-evaluation of the overall context was needed. Any kind of workflow that would lead towards a published repository therefore required consolidation, or a period of curation of the contents of the collection in line with new technological advances.

The legacy media is split between tapes using Digital Video (DV) and DVCAM (Sony's version of Digital Video) formats, a subset of which has gone through the transcription process (the entire collection also includes audio tapes as well as VHS). Previously, there was also a process of digitisation which led to derivatives of these tapes being captured on optical and hard drives. All of the legacy media material on the various forms of storage needed to be reconciled, organised and catalogued. It also needed to be analysed, by identifying and organising the content that would lead back to the central aim of the repository - to preserve the histories and the languages of the #Khomani San.

With this in mind, the curation component of the digitisation process involved not only looking at the media from an archival point of view, but also from a more conceptual perspective in terms of what was going to be showcased and for whom. How were visitors to the site going to interact with the different media? Who were the visitors going to be and what would they be looking for? Are the visitors going to be members of the #Khomani San community, and if so, how could the repository be arranged to make it accessible for them to navigate. Taking these requirements into account, as well as organising the media data, a further step was taken to organise the content of the data. For example, the development of spreadsheets to capture information such as: the full names of #Khomani San community members, place names and their associated GIS data, specifically mentioned cultural ceremonies, and indigenous plants for their edible or medicinal properties. All these variables needed to be identified and described in order to curate the media on a website and allow for multiple exploration points of the collection. This organisation of the data led to the grouping of distinct themes within the collection:

- PeopleCommunity
- PeopleContributors
- Places
- CulturalPractices
- Plants

And within each theme, the metadata used to describe it would be unique. For example:

- PeopleCommunity
 - name_second_surname
 - name_first_surname
 - name_first
 - name_traditional
 - name_nickname
 - name_housename
 - name_other
 - gender
 - birth_place
 - birth_date
 - death_place
 - death_date
 - biography

5.2 Challenges in Data Clean-up

The starting point for the creation of linked data rested within the 200+ transcripts that had been developed thus far (covering only a fraction of the entire audio-visual archive). The transcripts followed a similar layout within a Microsoft Word Document, and included a cell which was used to identify keywords related to the media (video tape). The keywords could relate to individuals on the video, other people, cultural ceremonies, locations, etc.

The first step was to extract the keywords cell from each of the 200 documents and place them in a structured table. This table of keywords provided metadata for the creation of records in the repository. In this way, when a media clip, i.e. 1998-05_12 was uploaded, it would be possible to “tag” it with the appropriate keywords that had been identified in the transcript. Having these keywords as tags would enable exploration of the collection.

The process of automatic extraction of the keywords from the 200 documents was enabled through the use of Python programming language which has built-in libraries to access components of a Word document. The code created a table with a column for the “clip name” and another column for the “keywords”. With some data tools through the R programming language, it was possible to reframe the data in the format such as Table 3. In this instance, each unique keyword is in a new row and the columns alongside, list the names of the clips where those keywords are found.

However, with multiple individuals working on this collection over the years throughout several phases of the project to discover the data, the keywords became unavoidably inconsistent. While keywords were captured for every single tape, they were established by a number of different individuals without the building of a necessary controlled vocabulary. Therefore, multiple spelling variations arose, as well as more and more languages and their associated translations, resulting in a singular concept being captured in many different ways (see Table 3 - bat-eared fox). For any kind of machine or computer program

each of these different spellings is a different concept. A website built upon such a list of keywords would not be able to provide paths to distinct people, places or concepts.

The challenge was then to consolidate these different spellings into single entries, so that when each of the clips were uploaded onto the exhibition site, they would point to a unique list of concepts or entries. This required the use of a digital tool called OpenRefine, which specialises in cleaning up “messy” data. In the example in Table 3 below, OpenRefine was able to identify that the five different spellings of “bat-eared fox” all refer to the same thing and consolidate the entries by combining the columns with the tapes being referenced as one line-item. This would result in the keyword of “bat-eared fox” being tied to distinct media items, enabling the user to journey through all of the clips which have it mentioned.

Keyword	Tape	Tape	Tape[...]
bat eared fox	1998_05-12		
bat-eared fox	1998_01_04	1997_07-13	2001_04-06
bat-eared foxes	1998_03-05		
bat-earred fox	1998_07-01	2000_02-07	1998_05-15
beard	1997_02-03		
beat-earred fox	2001_02-04		
beatig	2001_06-03		

Table 3: An example of data clean-up using OpenRefine to identify and group like entries for consolidation

This process of data clean-up in relation to the keywords found in the transcript files moved towards establishing a controlled vocabulary which could be used in future as further transcripts are added. This controlled process ensures the continual building of the collection by connecting to the existing dataset and forming immediate links. In other words, when a new tape is added, the keyword referring to a certain individual from that tape e.g. “Una”, would be tied to the existing vocabulary and therefore link the new media clip and its associated transcript file, to the existing data containing the same entry.

5.3 Multilingual Support

It was imperative to capture the multilingual and cultural fluidity of the community in the repository. This gives the option of linking data objects and is crucial to the creation of the digital collection and to the objectives of the project. The linked data capabilities of Omeka-S allow for singular concepts to be mapped out to different versions or expressions, allowing for multilingual support. For example, a search for a word (e.g. a place name) in any of the languages within the collection would link back to the same concept, object, or person that it refers to irrespective of the language used in the search

5.4 Navigating the Website

With respect to organising the media and the content on the collections’ feature website, the site needed a number of different entry points. Omeka-S lends itself to multiple presentations, due, in large part, to the nature of the database of links that are formed as the items are uploaded. This means that complex webpages populated by diverse elements can be setup through a simple query, as opposed to any hard coding or laborious design and arrangement work. For example, by setting up a simple template and a query to list all elements which reference a specific individual, the Omeka-S software can generate a page that presents all of the various media elements, together with any captured metadata, as well as certain custom texts. It therefore becomes easy to navigate through the collection by spending time on a page dedicated to each individual community member and accessing the transcripts, photos, and videos in which they are present. This way of navigating can be repeated with any of the conceptual elements within the collection (places, cultural ceremonies, plants, etc), thanks to the established keyword vocabulary which allows for multiple expressions to tie to a singular item, i.e. a many to one relationship. This flexibility allows for example, a cultural ceremony to be explored from many different personal narratives, or a particular place to be referred to by more than one speaker, or a specific plant to have multiple uses.

Furthermore, the website can also be explored chronologically through the journeys of the researchers who built the original collection, looking through the media in sequential order. With a view towards a kind of meta curation, the website also has custom pages that draw on a few key items to deliver a more pointed story. An example of this builds upon the keywords that had been lifted from the transcript sheets, consolidated and then used as the keywords in the upload of each of the metadata items. Building upon the description in the previous paragraph, it would be possible to allow visitors to the website to search for specific individuals talking about specific cultural ceremonies, or to listen and view all those who talk about a specific physical location. These custom pages can then be built by the collection team and scaffold upon such queries from the collection to highlight specific issues or themes. These custom pages can continue to be built as more individuals engage with the collection and express their interests or concerns. Such individuals could be researchers or members of the community or general public. These custom pages can be composed of text that is written specifically for that page, and then placed alongside some of the media items, which may be related to people, places, cultural ceremonies or plants. A culmination of custom designed pages based on the user engagement with the content of the collection results in a unique and continuously evolving resource.

6. The Importance of Collaboration

Researchers from diverse fields were instrumental in collating all the relevant data required for a successful land claim through a cultural heritage audit. Such fields included: visual anthropology, law, linguistics, cartography, conservation and social development (as exemplified in Figure 6). Such an endeavour would not have been possible without the collaboration of #Khomani San community members, researchers from diverse fields, charity organisations, NGOs and Universities alike. Such unity in diversity of expertise is testament to a potentially new methodology that embraces a decolonial ethos for one of South Africa's most unique narratives.



Figure 6: Cartographer Bill Kemp, working alongside linguist Levi Namaseb and former leader of the #Khomani San, Dawid Kruiper to create detailed maps of significant places in Khoekhoe, Afrikaans and English

7. Summary

This paper highlights the social, historical and linguistic context of the #Khomani San | Hugh Brody collection with a particular emphasis on the digitisation process applied by the University of Cape Town's Digital Library Services to deliver an accessible end-product.

Until the final URL is available please contact Michal Singer, Special Collections, UCT Libraries, michal.singer@uct.ac.za for current access details to the collection or either of the authors of this paper.

8. Acknowledgements

This digital collection was made possible due to the input provided by the #Khomani San community, Hugh Brody, Open Channels, the South African San Institute, African Tongue, University of Fraser and the University of Cape Town Libraries.

9. Bibliographical References

- Adamo, J. (2011). Omeka and the NLM Digital Repository: A New way of Creating HMD-Curated Websites. National Library of Medicine. (Online) Available: https://www.nlm.nih.gov/about/training/associate/associate_projects/AdamoOmekaReport2011.pdf
- Crawhall, N. (2001). *Written in the Sand: Auditing and managing cultural resources with displaced indigenous peoples: A South African case study*. Mowbray: SASI.
- Gruber, E. Building Omeka Exhibits with Fedora Repository Content. Published December 15, 2010. (Online) Available: <https://scholarslab.lib.virginia.edu/blog/building-omeka-exhibits-with-fedora-repository-content/>
- Güldemann, T. (2008). Greenberg's "case" for Khoisan: the morphological evidence. In D. Ibrizimow (Ed.), *Sprache und Geschichte in Afrika*, 19 (pp. 123-153). Köln: Rüdiger Köppe
- Hardesty, J.L. (2014). Exhibiting library collections online: Omeka in context. (Online) Available: https://scholarworks.iu.edu/dspace/bitstream/handle/2022/17627/HardestyJulietL_ExhibitingLibraryCollectionsOnlineOmekaInContext.pdf;sequence=1
- Heine, B and Honken, H (2010). The Kx'a Family: A new Khoisan genealogy. *Journal of Asian and African Studies*, 79: 5-36.
- Jing, J. (2015). Digital asset management (dam) systems used in Libraries. Queen's University Library. (Online) Available: <https://www.slideshare.net/happyrainb/digital-asset-management-dam-systems-used>
- Jing, J. (2016). The workflows for the ingest of digital objects into a repository/digital library. (Online) Available: <https://www.slideshare.net/happyrainb/the-workflows-for-the-ingest-of-digital-objects-into-a-repositorydigital-library>
- Jones (2019). Contemporary Khoesan languages of South Africa. *Critical Arts*. DOI: 10.1080/02560046.2019.1688849
- Kandybowicz, J., and Torrence, H., (2017). Africa's endangered languages: Documentary and theoretical approaches. New York: Oxford University Press. Omeka-S. (Online) Available: <https://omeka.org/s/> Penn Libraries Guides: Omeka. (Online) Available: <https://guides.library.upenn.edu/omeka>
- Sands, B., Miller, A., Burgman, J., Namaseb, L., Collins, C., Exter, M., (2006). 1400 item N|uu Dictionary. Manuscript.
- Statistics South Africa (2018). General Household Survey 2018. Statistical Release, P03182018, Available at: <http://www.statssa.gov.za/publications/P0318/P03182018.pdf>

- du Plessis, M. (2018). *Kora: A lost Khoisan language of the early Cape and the Gariep*. Pretoria: Unisa Press.
- Voßen, R. (2013). *The Khoesan Languages. Routledge Language Family Series*. London: Routledge.
- Witzlack-Makarevich, Al. (2006). *Aspects of Information Structure in Richtersveld Nama*. MA thesis, Institut für Linguistik der Universität Leipzig.