

Digitisation as a Means to Document Shared Histories of the Liberation Struggle

Roger LAYTON¹, Marjorie JOBSON²

¹CEO, Roger Layton Associates (Pty) Ltd, Suite 230, Killarney Mall Office Towers, 60 Riviera Road, Killarney, Johannesburg, South Africa
Tel: +27 (82) 881-0380, Email: roger@rl.co.za

²National Director, hulumani Support Group, 2nd Floor Khotso House, 6 Marshall Street, Johannesburg, marje@hkulumani.net

Abstract:

In 2009, Roger Layton Associates was commissioned to create the National Policy on the Digitisation of Heritage for the Department of Arts and Culture. This culminated in a public participation workshop in Feb 2011, and has led to the creation of a best-practices document called the Digital Heritage Body of Knowledge (DHBOK), which is now in its second version.

Digitisation is seen as an inevitable evolutionary path in heritage documentation, and particularly for archival documentation. Once converted into digital form, all forms of heritage are maintained as digital archives and it is this archival nature which forms the basis of the large-scale digital repositories.

Archival repositories may be structured merely as collections of disconnected materials, but their value is greatly enhanced when the inherent and implicit connections between the archival materials is documented. Using the archives of Khulumani a pilot was established concerning the notion of shared histories and how this can be given substance using the power of modern repositories. Rather than seeing each oral history as an independent story, it is the massively-connected nature of these stories which provides a new, previously unseen, view of the collective story. Early work on this was presented at the 2012 Oral History Conference.

Our work has created a platform for the digitisation and documentation of the oral histories to enhance this shared value and as a tool that can help to create an eternal history, or an eternal heritage.

Keywords: digitisation, digital archives, digital repositories, eHeritage, digital heritage, oral history documentation, digitisation policy, Linked Data, Web 3.0, Semantic Web, RDF, liberation struggle, victims of conflict, biographies

1. Introduction

1.1 The Khulumani Archives

The Khulumani archives contain biographical information of more than 45,000 individuals who are victims of conflict and who have mostly not been recorded into the archives of the TRC and Freedom Park. For various reasons, these individuals were unable to participate in these programmes and yet their stories are equally as significant as those previously recorded.

Khulumani Support Group is the only national membership organisation of more than 65,000 victims and survivors of apartheid-era gross human rights violations. The organisation was formed in 1995 prior to the establishment of the TRC.

Khulumani's main focus remains that of transforming survivors of major apartheid trauma into community activists through processes of facilitating that they reclaim their political agency and contribute to community development initiatives for economic empowerment and for securing the accountability of local government in their local communities. The active citizenship demonstrated by empowered Khulumani members focuses on dealing with the ongoing reality that the post-1994 years have been marked by slow delivery on the promise of freedom with local government being beset by inadequate management and a slow roll-out of necessary services.

Khulumani's mission is summarised as - responsive to the needs of its members Khulumani exists to restore the dignity of all persons harmed by apartheid through their transformation from victims to active citizens. A major contributor to the restoration of the dignity of victims and survivors is the acknowledgement of what happened to individuals who sustained serious harm in the struggle for liberation.¹

Recording of biographical information may be carried out as individual life stories, concerning events that are specific to each individual. However an emerging approach is to develop "Linked Data" in which data is linked into larger information structures and archives, to enable a vision of interconnectedness of all data stored in all archives.

Some of the events identified in these archives are shared events, in which a number of individuals participated, and there are linkages between people and organisations that can be recorded to improve the quality of information retrieved from search engines, leading to a new form of information discovery that we have called the "connection engine".

The Linked Data movement is new and has its roots in modern Web technologies. Linked Data essentially does for individual data elements what hyperlinking has done for documents, and this movement is driven by the same organisations that created the World-Wide Web some 20 years ago.

Whereas traditional databases are limited to the handling of text-based data, the modern trends are to accommodate shared vocabularies as well as to include a range of non-textual data including scanned notebooks, digital images, as well as audio and video records.

This paper outlines how biographical information is being collected and stored into a modern semantic data structure, and how this is being included into a digital archive repository for the purposes of long-term preservation and improved access. This paper also outlines how we are managing the process of data collection of biographical information.

1.2 The Digitisation of Archival Materials

There have been major calls for digitisation of archival materials, with this being seen as a means to improve access to materials, as well as to ensure a long-term preservation of items of enduring value. Many archival materials are not stored under best-practice conditions, and are often readily inaccessible.

Digitisation of archives remains a costly exercise and it is essential that the limited funding available for such projects are well spent, and that the outputs of the digitisation projects is moved to digital repositories that provide for long-term digital preservation.

The output of digitisation is a set of digital objects, but when view by themselves they are limited in their usefulness. It is only when these objects are related and combined into stories that their true value becomes realised, and this is a value which can be achieved using modern technologies.

¹ Information about Khulumani Support Group is available on its website www.khulumani.net

In this paper we argue that this true value is inherent in the capability of the digital objects to be related to one another, rather than in terms of being viewed merely as a collection of unrelated items. We consider the shared histories which exist within the archives, and which can be brought into the open with the availability of the digital objects.

We also argue that the process of digitisation is a long and complex process, of which the actual capturing of the digital images, or the reformatting of the audio and video records, is only one part. We are far more concerned with what happens to these digital objects after they have been produced. These are sometimes provided on a DVD, since the institution itself does not have the capacity to store them on their own servers.

1.3 Recorded Narrative of the Liberation Struggle

Based on our work with Khulumani, including Zdena Mtetwa and Marjorie Jobson, it has become evident that there are a large number of recorded narratives of people involved in the liberation struggle and who did not, or could not, participate in the TRC, and yet who share the same historical timelines.

While the TRC created a ‘meta-narrative’ of the liberation struggle, the “story stock of the TRC” does not reflect the incredible richness and diversity of contributions to the birthing of democracy in South Africa, nor does it create space for the contestations that were part of the lived reality. There are many stories that were silenced and even suppressed by the TRC. The collation of narratives by the TRC was in fact, carefully managed. Ms Nomarussia Bonase, Khulumani’s Gauteng Provincial Coordinator, explains “*those who went inside to testify to the TRC were improperly prepared. They were told what to say to the TRC. Statement takers asked specific questions from a list; the “victim” would respond to those questions. In the list, rape and gender violence was not a question. If a woman said it happened to her, the interviewer often did not record it.*”

The ongoing documentation of stories remains crucial, not only to provide the bearers of these stories with acknowledgement but also to more accurately populate the historical record to prevent distortions that can and are being invoked for particular political purposes. The value of narratives is that they generate knowledge from the perspective of others’ lives and they help to prevent erasure while also refusing complicity in the production of domination.²

Many contributors to the historic liberation struggle in South Africa are at risk of being forgotten if they are not recorded somehow and somewhere within some formal and accessible structure. Khulumani have been recording these oral histories and have built up a large database, but this remains largely inaccessible due to limited resources, and yet holds many stories of South Africa’s struggle to democracy. In most cases these are the stories of the unsung heroes and heroines of the struggle, and it is for this reason that this is a useful starting point for exploring new methods for encoding of the oral histories, which may create, using the words of one of the Khulumani staffers “the facebook of the struggle”. This analogy to FaceBook is weak, since in the oral history we need to create modes of interconnectedness that are far beyond the limited structures available within the social networking systems, and in order to achieve this vision I have examined current and emerging methods for the encoding of biographical information as part of the search for the ideal approach.

1.4 The Motivation for this Work

Our motivation for this work has arisen from the observation that all histories are massively connected, and yet this level of connection is not directly evident from treating such histories as

² Presentation by Professor Louise Vincent of the Department of Political Science at Rhodes University on “A Narrative View of Institutional Culture”

disconnected personal life stories and as digital repositories. The true value of using a formal encoding process on the individual words and sentences of oral history transcriptions is that we can identify new connections, and can rebuild history in different ways, using a more powerful “connection engine”, which I foresee as the next generation of the current linear-structured “search engines”. My argument is that we cannot achieve this next generation of massive connections without a deliberate effort into fine-level encoding of the oral history materials we have at our disposal.

I have outlined my early approach to codification elsewhere [1, 2] in which I have introduced broad-brush techniques to encode oral history transcriptions, and this early work has since been extended to address more fine-grained approaches to encoding. My approach is to use modern methods of encoding of semantics within documents, as is becoming increasingly important within the next generation of the World-Wide Web. This is referred to as the Semantic Web, Web 3.0, or the Linked Data movement [3], and my approach is to apply this to biographical information obtained from oral history projects, no matter how large or small are the narratives recorded.

2. The “Elements” arising from Digitisation Projects

This paper follows on from papers presented by Roger Layton at the Oral History Conferences in 2010 Conference on “Digitising and Documenting the Oral History” [4] and in 2011 (with Zdena Mtetwa from Khulumani) on a “Linked Data Approach to the Biographical Documentation” [5]. The focus of the 2010 paper was on the “elements” arising from oral history digitisation and documentation projects and how these elements play a role in the context of the Living Heritage Policy [6], the National Policy on Digitisation [7], and the current technology trends that are providing impetus for large-scale, long-lived digital heritage repositories. One section of that 2010 paper concerned the codification of the oral history in order to provide a more detailed set of linkages to related items that are referenced within a single such element. This codification is concerned with metadata which is provided with digital objects, as well as more detailed information which concerns the content implicit in the digital objects, such as the subject matter of audio, images and video, and the full text information of archival records. The 2012 paper extended this work by exploring the role that Web 3.0 technologies will play in the future of online digital information, and in particular how these various digital archives and repositories, which may contain may hundreds of thousands of digital objects, can be interconnected.

This paper which we are presenting now moves on from this to explore how large-scale digital repositories can be used to extract shared histories. However, this can only be effective if the digital objects are stored, structured and marked up in the right manner. In this way, and only in this way, is it possible to create the vision of usable interfaces into large-scale archives. Some work is already underway by Google and others on creating timelines based stories using modern web interfaces, but this is merely the start of a new approach to the presentation of archival information.

The encoding of the oral history is a “hard problem” given that there is so much rich information and connections that can be gleaned from even small statements, and also given that each individual narrative contains references to a wide range of external concepts and things, each of which will be required to be encoded separately, and each of which should be commonly identified within well-known vocabularies in order to minimise the risk of duplication.

3. A Model for Linked Data Encoding for Biographical Information

3.1 Biographical Archives

Our focus on this paper is on the shared histories, and thus our focus is biased towards biographical information, in which the individual histories are documented in words, audio recordings and

pictures, and which then form the basis for the discovery of the shared history. The problem is that when these individual histories are documented they lack this shared perspective.

The “elements” of the oral history, for the purpose of this project, are the documents that have been captured such as; audio recordings, their transcripts, and their translations; digital images from scanned photographs; old newspaper cuttings; archival documents; film and video, and any other type of source materials.

When these “elements” are treated as the units of oral history information, they are indexed and made available as self-contained units in digital archives, and are mostly managed as parts of specific oral history collections. However the real value is to be found embedded deep within the content, and this is not directly available within a document-level indexing structure. When we look for specific information, such as the incidence of a particular type of event in a particular area within a given time frame, we will not discover this from the titles or the metadata associated with the document-level information. What is needed is to change our notion of the “unit of information” to be the internal parts of the documented historical recordings — from each chapter, paragraph, sentence and word, and also the subject matter contained within each image and video segment. In essence, we turn the current words and images into data. The motivation for this approach is that data can be readily queried, whereas documents as a whole and images cannot. Whereas there has been much emphasis on “full-text” searching in creating the modern search engines such as Google, this approach fails in cases in which there is considerable context in which the statements exist. The current search engines are good at finding web pages over the Internet but cannot find relationships between things, which is the essence of our approach to digitally recording the oral history. We need to build a new generation of information-based tools to accommodate these needs, which is the combination of a new encoding structure with a new form of data discovery using the “connection engine”. The work of encoding the oral history is posing challenges for our notions of recording and notating information and is pushing us to a new future of the information technologies.

Our approach makes an important assumption concerning the motivation for recording oral histories in the first place. We argue that such documentation is not to create an archive of independent histories, but to provide the basis for understanding the collective histories, and the massive connections that exist between these histories. We explore future, ideal forms of data representation that serve the basis for long-term management of oral histories as connected and linked data, rather than as independent stories.

In order to achieve this ideal data representation, we need to analyse our oral history recordings not as a narrative or as a story, but as a set of linkages to other things: to people, organisations, events, places, and concepts. This is a very large task when considered in the context of the totality of the oral history and in order to reduce the scale of this task for the purpose of a pilot implementation we have chosen to explore the smaller problem of the encoding of biographies, and have collaborated with Khulumani, who have extensive oral history resources available from years of recording narratives and stories of the unsung heroes of the liberation struggle.

These are not complete biographies as one would find in a “Who’s Who” but are mostly provided in the form of small snippets of individual narratives which highlight the specific role of individuals who participated in the liberation struggle, including the stories related by others concerning those who died or disappeared. Some of these narratives are small, and some are long and extensive — some are available with the original wording and in the original languages, whereas others are provided as translations, and in which the stories have been rewritten for the benefit of the reader. However, it is not the medium or the format or even the level of rewriting which is significant, but the content, and it is the finer details of this content which is of interest to me in the context of this project.

Attempts have been made to describe biographical information as a sequence of events that make up a person's lifetime, as in the BIO vocabulary [8], which is focussed on the relationships with other persons. However, for the oral history in general, and the biographies of those who participated in the liberation struggle in particular, this approach to encoding biographies is too limited since it omits the important elements of shared events and the ad hoc organisational structures that were formed spontaneously in response to the worsening political situation at the time. This BIO vocabulary also omits the contextual relationship with places and of the larger context in which individual biographies were a part.

Our approach has been to design a new encoding structure from the bottom up, by examining the nature of the materials available to me, and from this to create new structures that capture the essence of this interconnectedness.

3.2 Biographical Events

Events are an important element of biographical information, representing the dynamic elements of what happened to people ("I did this", "this was done to me", "I participated in this shared event"). The types of events that we see within these struggle narratives use verbs such as "teargassed", "picked up (by police)", "tortured", "arrested", "beaten (to a pulp)", "dies", "joined (the struggle)", "detained", "attacked", which all indicate a violent society which was both fighting with itself and with its oppressors.

The statements that describe events represent complex relationships that exist between people, organisations, places, outcomes, and also with other prior and consequential events.

As an example, consider the statement:

"I was badly beaten by the police here in Sikhulu Street when we were involved in a protest dance, the *toyi-toyi*. I was beaten on my back with an R1 rifle and spent three months in the local provincial hospital."

This shows a particular type of event (*beaten by police*), the place where this occurred (*Sikhulu Street*), and the reason given by the person concerned (*involved in a protest dance*). There is no date or time reference for this event. There is a reference to the *toyi-toyi* as a protest dance. The second sentence indicates the instrument used for the beating (*an R1 rifle*), and the body part affected (*back*). It also indicates the consequence of this event (*three months in hospital*).

Further in this short narrative the following statement is made

"When the weather is cold, I have severe pains in my body and I am unable to sleep. My back has never healed completely."

This shows the need to identify the long-term outcomes arising from an event, in this case the result of the beating on the person and how this impacted their life. This is a cause-effect relationship which is significant in identifying the narrative in perspective.

In order to formalise this encoding of the narrative it is necessary to break this into its constituent parts, identifying the various parts of each statement, what they refer to and how they link together. These individual parts in this story concern the following:

- Beating : a type of event
- *Toyi-toyi* : a type of event, sub-type of dance
- Police : an institution / organisation
- Provincial hospital : an organisation
- Back : part of body
- Three months : a duration

We also need to encode all of the referenced information (organisation, place, time, people) sufficiently to capture the connections and commonalities between the events and between individual people's histories.

From this small selection from a small narrative, a snippet of a single person's life, a large number of issues arise as to how to effectively encode this at a finer level, and this raises the question of the "grain-size" of the encoding.

3.3 Grain-Size

"Grain size" problems exist in situations in which we can develop finer and finer descriptions, leading to the question of how fine do our descriptions have to be to accomplish the outcomes we are seeking. The term "grain size" arises from the analysis of a sandy beach, which at one level is a single unit, the beach, and at another level can be seen as individual grains of sand. The sand granules themselves can be divided further into their molecular structure if we wish, but at some point we need to decide how fine a level do we need for our descriptions.

This issue forces us to make a key planning decision early in the encoding process concerning how much detail we encode, since encoding these narratives at a fine level will provide more detailed information but will be more costly, and encoding at a high-level will provide less connections but can be conducted quickly and at relatively low cost. We do not currently know the answer to this grain-wide question for the Khulumani archives and we hope to learn lessons from this project to help guide future projects.

Our recommendation, given the massive volume of un-encoded oral history documentation, is to start encoding at the broadest level and then to conduct future projects in which a finer level of encoding can be carried out on documents regarded as more significant than others.

There is always the possibility that future artificial intelligence programs will be able to read the archival documents and to automatically encode these, but this approach appears to be almost impossible given the nature of the oral history transcriptions, and is at least 50 years into the future.

3.4 Common Vocabularies

There is a significant level of commonality found in the contents of individual oral history elements, including transcribed narratives, photographs and videos. These elements will also include future forms of historical documentation, such as blogs, emails, online discussions and web sites, but for the present these are excluded.

This commonality must be encoded using commonly available vocabularies, and the call for a common set of national vocabularies has been made previously [7, 9] with some prior work in creating these having been performed in 2006-7 [9].

Creating vocabularies is a complex and time-consuming activity, and it is important to start correctly so that the vocabularies are useful and as widely-known and accepted as possible. Once vocabularies are in place, they not only provide a fixed-point for encoding, but they also impose a single way of encoding, which itself may be a risk. To mitigate this risk I recommend that all encoded source information be retained in its original state for re-encoding in the future, and also that the vocabularies are also flexible and evolving, allowing for redefinitions, the inclusion of new words, terms, and named things, and also the ability to represent different viewpoint formally and explicitly within the structure of the vocabulary.

As an example of these different viewpoints, it is common that experts hold different positions, and disagree on interpretations of specific events and situations, and when such differences are analysed, they can be reduced, in some cases, to different views of how the evidence should be structured, and into which semantic box a particular piece of evidence should fit. This does not belittle the work of the experts, but it provides a means to formalise and clarify multiple viewpoints and is a move towards the "Glass Bead Game" representation that I cited in my paper in 2010 [1].

The common approach to vocabularies has been to structure them into one of two types: thesauri, representing abstract terms, and authority files, representing named items.

Authority files are concerned with individualised items that we have already named, and this includes People, Organisations, Places and Named Events – essentially the concrete nouns. Thesauri are concerned with ways in which we use categories and classifications to make sense of our world, to identify groups of items using these categories – which represent abstract nouns, adjectives, verbs, and adverbs from our language structures.

I have produced extensive prior work on the nature of these vocabularies and authority files [8, 9] and it is not possible to provide details here, and consequently I will rather outline only some high-level issues as they apply to encoding of oral history biographical narratives.

3.5 People and Organisations

The first authority files I have examined are those concerning individuals (Persons) and how individuals organise themselves into groups for various purposes, formal and informal (Organisations).

Persons exist by name, and ideally each person should be recorded only once within this vocabulary. However, in practice this is difficult, since the same person is referred to in different ways, and often by different names, in the context of the oral history narratives. In some cases the name is their full name, in others it may be partial (such as Mr X), and in other cases, nicknames are used. In most cases there is no identification information on which to base the linkage such as an ID Number. There also a number of people who share the same name.

Organisations may represent large-scale national organisations (the Police), or informal local organisations (CRADORA – the Cradock Residents Association, CRADOYA – Cradock Youth Association).

Some of these organisations are widely known and widely used as names and abbreviations (e.g. COSAS), whereas others are local and relatively unknown outside of their geographical areas.

In other cases, groupings represent geographical communities (such as CRADORA), tribe names (Zulu, Xhosa), political parties (ANC, Inkatha Freedom Party / Inkatha / IFP).

There are also common names used for various groups that are significant due their histories being highlighted within the context of the struggle (PEBCO 3, Cradock Four).

There are many challenges in identifying and recording these names into a formal database, due to many of these having different meanings to different people, differences in spelling, and changes to the names of the organisation including name changes that occur over time.

3.6 Places

The Places authority file includes the names of towns, villages, suburbs, streets, and buildings. For example, from our work on the Cradock narratives some of the place names that appear are:

- Cradock, Langa Township, Tembisa
- Sikhulu Street
- Provincial Hospital, Livingstone Hospital, Walton Hospital
- Matthew Goniwe High School
- Pollsmoor Prison

When we read these names in a narrative, we create a mental image of the place as we read the narrative. We also need to record the older names used, as well as common, colloquial names and to include the type of place.

An international Thesaurus of Geographical Names (TGN) is maintained by the Getty Foundation, but this does not accommodate the level of detail required for our current work, although it serves an important purpose for more commonly used names.

3.7 Named Events

This is an authority file concerning the names of specific events which are well-known and used frequently. This will include the names of wars, battles, sieges, etc. as well as the names of large-scale events which cannot be accurately placed on a time line. However, each and every named event does have some mapping to a time line, even if this is inaccurate.

Examples in usage within our current work include the following:

- Liberation Struggle / the struggle
- State of Emergency
- Langa Massacre
- School boycott
- Sharpeville massacre

These names events has a parent-child relationship in which a larger event (such as “the struggle”) contains within it a range of other events, which themselves may overlap in time, and which may also have their own child events. I have found that this parent-child relationship exists with a number of types of vocabulary, and that there are few purely linear structures of names or knowledge.

3.8 Event Types

Another thesaurus is needed to describe the types of event that happen to people and organisations, and this is represented by the verbs used within the sentences. We find the words used repeat themselves in many contexts, and have much the same meaning, but the true meaning is often not defined anywhere and is implied. Some of the common event types in our sample set of narratives include:

- detained
- arrested
- beaten
- shot
- killed
- tortured
- compensated

Each of these has a range of parameters which define their situation, which correspond to the subjects and objects used within sentences of which these event types are the verbs.

3.9 Roles Types

The final thesaurus I outline for the purpose of this paper is that of the relationships or roles that exist between entities, and particularly between people and organisations.

For example

- father/son / other family relationship
- member
- founder / co-founder
- chairman

These few examples of the various vocabularies have barely scraped the surface of this complex area for encoding information. Much work is needed to ensure that such a common set exists for the benefit of all, since without such a common set, it may prove impossible to allow for queries that span repositories and collections. The worst-case scenario is that many repositories will be built that are not interoperable, and the development and widespread usage of common

vocabularies, and associated standards for representation of dates and times, will be sufficient to mitigate this eventuality.

4 Example of Encoding

For the purpose of this paper, I am illustrating the encoding processes we are using for a single narrative, and how this relates to other vocabularies.

4.1 Raw Content

This information has been used in its raw and original form with real names used. This is because in order to tell these stories they cannot be about anonymous people, but the real people who wish their stories to be told – it is they who are the “unsung heroes” that we identified in the title of this paper.

NOMBULELO MBANJANA, 11 Kwintshi Street, Cradock

I was not directly affected by torture but my uncle Mr Madoda Jacobs was an activist and a leader. Madoda Jacobs was an activist since 1977. He was arrested and beaten to a pulp on several occasions up until 1990. He was detained under Section 28 of the Internal Security Act and spent time in Pollsmoor Prison in Cape Town in 1984 together with Matthew Goniwe who was brutally assassinated. He spent nine months in Pollsmoor prison and was then a listed person who could not be quoted. He was detained in 1985 until the lifting of the state of emergency in 1986 at which time he became the chairperson of COSAS. COSAS led the school boycotts from 1984 to 1987 which resulted in 200,000 school children participating. Mr Jacobs was a member of the Cradock Youth Congress which was affiliated to the United Democratic Front. He could not complete his studies and my family is still bitter that he died in 1984, unemployed and uneducated, having spent most of his student years behind bars. Other people went to school and obtained qualifications that enabled them to enjoy the fruits of freedom. But he was a forgotten hero, neglected by those in power. Nowadays few people are committed to the people as he was. I am proud of him. Long live the spirit of Madoda Jacobs. Long Live!

4.2 Analysis of the Content

Here we read a narrative about a person whom the author claims is a “forgotten hero”. Within this narrative we see a large number of “things” mentioned, which represent the key elements from which this story is weaved. My recommendation is to identify and encode these nouns as the first step in the process of encoding the entire narrative.

In this first level of encoding I identify the key nouns and encode them using the XML structures we have developed. I have only selected the first half of the above narrative, and you will see that these include the XML Tags of Person, Organisation, Law, Place, and NamedEvent. In each case the “Code” attribute is linked back to the vocabulary entry for these nouns. One example of each of these markups is highlights in **bold**.

I was not directly affected by torture but my uncle **<Person Code='madoba.jacobs'>Mr Madoda Jacobs</Person>** was an activist and a leader.

<Person Code='madoba.jacobs'>Madoda Jacobs</Person> was an activist since 1977. He was arrested and beaten to a pulp on several occasions up until 1990. He was detained under **<Law Code='ISA' Section='28'>Section 28 of the Internal Security Act</Law>** and spent time in **<Place Code='pollsmoor'>Pollsmoor Prison</Place>** in Cape Town in 1984 together with **<Person Code='matthew.goniwe'>Matthew Goniwe</Person>** who was brutally assassinated.

```
He spent nine months in <Place Code='pollsmoor'>Pollsmoor prison</Place>and
was then a listed person who could not be quoted. He was detained in 1985
until the lifting of the <NamedEvent Code='stateofemergency'>state of
emergency</NamedEvent> in 1986 at which time he became the
chairperson of <Organisation Code='cosas'>COSAS</Organisation>.
```

Next we markup the relationships that exist, such as that he was the chairperson of COSAS, and that this started around 1986. Another relationship is that he was at Pollsmoor with Matthew Goniwe.

```
He was detained in 1985 until the lifting of the <NamedEvent
Code='stateofemergency'>state of emergency</NamedEvent> in 1986 at which
time he became the <Relation Type="chairperson" Person="madoda.jacobs"
Organisations="COSAS" DateStart="1985" DateEnd="?"> chairperson</Relation>
of <Organisation Code='cosas'>COSAS</Organisation>.
```

Finally, it is then possible to encode the specific events that occurred in the life of Madoda Jacobs, including the roles he played, his arrest, beating, detention, and the fact he was a listed person under the State of Emergency.

```
He was <Event Type="detention" DateStart="1985" DateEnd="1986"
DateEndEvent="stateofemergency">detained</Event> in 1985 until the lifting
of the <NamedEvent Code='statsofemergency'>state of emergency</NamedEvent>
in 1986 at which time he became the <Relation Type="chairperson"
Person="madoda.jacobs" Organisations="COSAS" DateStart="1985"
DateEnd="?">chairperson</Relation of <Organisation
Code='cosas'>COSAS</Organisation>.
```

Only limited examples have been provided of the above due to lack of space. However, these indicate the nature of our markup process for the encoding. It is important to note that this markup does not impact what the users will see when they access this. Each of these marked up areas embeds a small part of the text, and will allow for a link to the item mentioned if this points to another entity (such as a Place or Person).

The next step is what happens to these oral history files once they have been encoded in this way.

5. Discussion

This paper is too short a space to outline our goals in building this “connection engine” repository structure for the oral history, and much of our work is engaged in research into new ways of recording and encoding this information.

This work falls within our programme called “ETHER” (Eternal Heritage) which has the vision of heritage information systems which can live forever, through continual attention to reformatting and migration of the content and upgrading the content as new technologies become available. ETHER also includes the development of new standards for museum and library collection management in the digital age, and the development of collection management systems that are more accessible to smaller and less-wealthy institutions.

Our work on the Khulumani materials is self-funded at present, although we are looking to complete a full pilot to use as the basis for attracting funding. This will enable new ways of accessing the vast stores of oral history materials, in which they will appear to the users as items in a massive connected universe for which even a 3D visualization is insufficient to capture the complexity of the relationships. However, such visualization is a useful tool in seeing how these connections are structured, and one future project we have considered is the usage of such virtual reality tools to help us understand and grasp the complexity of this data by seeing it within a 3D or 4D landscape rather than using the limited interface of the 2D 15” computer screen.

For now, our work continues on simple data sets, using these as the pilots to help us work on our prototype for the Connection Engine.

We would like to use a range of different types of oral history, beyond the biographical accounts, and would welcome the opportunity to work with others as part of our ongoing research and development of this Connection Engine.

References

- [1] Layton, R. Codification of Intangible Heritage for the South African Heritage Resources Information System (SAHRIS). SAMA National Conference 2006. 13 June 2006, Champagne Sports Resort, Drakensberg, Kwazulu-Natal.
- [2] Workshop conducted on Oral History Codification with the Women's March Oral History Project. 2007.
- [3] Linked Data. World Wide Web Consortium. Retrieved from <http://www.w3.org/standards/semanticweb/data> on 2 October 2011.
- [4] Layton, R. Digitising and Documenting the Oral History. National Oral History Conference, 2010, White River.
- [5] Layton, R., Mtetwa, Z. A "Linked Data" Approach to Biographical Documentation. National Oral History Conference, 2010, Mafikeng.
- [6] Department of Arts and Culture. Living Heritage Policy Draft for Public Review. December 2009.
- [7] Department of Arts and Culture. Draft National Policy on Digitisation. 2011. Available from www.dac.gov.za.
- [8] BIO: A vocabulary for biographical information. Retrieved from <http://vocab.org/bio/0.1/> on 2 October 2011.
- [9] Unpublished report on Data Coding Standard for Heritage, Part D of the South African Heritage Resources Information System Project, Phase 1, SAHRA, 2007.
- [10] The ETHER Initiative. www.ether.co.za