Knowledge Management for Development Journal
Vol. 7, No. 3, December 2011, 291–307

Routledge
Taylor & Francis Group

Collecting structured indigenous knowledge for enhanced exploration: the National Recordal System (NRS), South Africa

Ronell Alberts*, Gugu Khalala and Onkgopotse Molefe

CSIR Meraka Institute, Pretoria, South Africa

Indigenous Knowledge (IK) is recognised as a valuable, often untapped resource. It is increasingly important to record un-captured IK for the purposes of protection, preservation and promotion. The potential of IK for the use of scientific research and development is known and can be financially beneficial to the holders of IK in communities when managed properly. The need to be able to effectively explore recorded IK for the purposes of scientific investigation and to prove prior knowledge is increasing. The exploration of IK recorded as free-form stories is found to be less effective for these purposes. In order to increase the effectiveness of exploration of recorded IK, an approach was defined to enrich the processes of IK collection by defining structured metadata to be collected in addition to the IK story itself. The National Recordal System (NRS) initiative in South Africa has been used as a case study for this approach. The approach, implementation, evaluation and results in terms of the effectiveness of the process, the effect of the introduction of structured metadata in the process and the resulting effectiveness of exploration are described in this paper.

1. Introduction

Indigenous Knowledge (IK) refers to

> the unique, traditional, local knowledge within and developed around the specific conditions, indigenous to a particular geographic area and stored in people's memories and activities and is expressed in stories, songs, folklore, proverbs, dances, myths, cultural values, beliefs, rituals, community laws, local language and agricultural practices. (Tella 2007, p. 185).

It can be understood as the local or traditional knowledge that is unique within a community or society, transferred through generations by sharing experiences and skills and by storytelling. IK is primarily implicit and holds the local and cultural practices of the people who have lived in a community for generations. This knowledge is usually transferred through stories and practice demonstrated from teachers to apprentices, parents to children and between neighbours (Okorafor 2010).

IK is important because: (1) IK provides problem solving strategies for local communities, especially for the poor; (2) it represents an important contribution to global development knowledge; (3) IK systems are at risk of becoming extinct; (4) IK is relevant for the development process; and (5) IK is an under-utilized resource in the development process (Rao 2006, Tella 2007).

*Corresponding author. Email: ralberts@csir.co.za

ISSN 1947-4199 print/ISSN 1871-6342 online

http://dx.doi.org/10.1080/19474199.2011.649009
http://www.tandfonline.com

IK is collected for various reasons such as to prevent the loss of IK as time passes (preservation), to protect IK from unfair exploitation and bio-piracy (protection) and to ensure the IK remains relevant and accessible to later generations (conservation and promotion) (Global Exchange 2001, Ruiz *et al*. 2004). In addition, IK is acknowledged as a valuable source of knowledge for scientific exploration and discovery (Agrawal 1995, Sillitoe 2009). To realise the potential of new scientific discovery based on IK, it is important to collect and store IK in databases where exploration of the knowledge is possible while the privacy of the knowledge is preserved (Ruiz *et al*. 2004, Sen 2005). Due to the fact that IK is used as the basis for scientific research and development, it is crucial to protect IK from bio-piracy and unfair exploitation by making it available for patent searches to prove prior knowledge (Rao 2006, Traditional knowledge digital library n.d.). Initiatives have been launched in various countries to record IK such as the Traditional Knowledge Digital Library from India (Rao 2006, Traditional knowledge digital library n.d.), the Biozulua database from Venezuela (Johnson 2002) and the China Traditional Chinese Medicine (TCM) Patent Database (Liu and Sun 2004, Feng *et al*. 2006).

Approaches to collect and record IK span a broad spectrum. These include recorders visiting people who hold IK and are willing to disclose it to record their stories on video, to field workers performing ethnographic studies to discover the IK and devise methods to record the IK in collaboration with the IK holders. Because IK is traditionally transferred orally in the format of stories, IK is most often recorded in video or audio format in the native language of the narrator (Telfer and Garde 2006, Bidwell *et al*. 2010).

A story is defined as ‘a narrative that can either be true of fictitious that is delivered in prose or verse and is designed to interest, amuse or instruct the hearer or reader’ (Dictionary.com n.d). The facts presented in a story depend on the memory and train of thought of the narrator and may be unintentionally incomplete or implied.

For the purposes of scientific exploration or searches to prove prior knowledge, the correctness and completeness of the information has a direct impact on the relevance and usefulness of the result. Due to the free-form nature of a story, exploration of the facts hidden in the story is not optimal. Specific, fact-based exploration questions cannot be effectively answered with a standard keyword search over the full text of IK stories. An example of such a query is:

> Give me all the plants where any of their aerial parts (i.e. leaves, twigs, flowers etc.) are used in medicine to treat respiratory diseases.

The objective of this paper is to report on an approach designed to enrich the collection of IK with structured metadata in order to enhance the completeness of the information and the potential for exploration. The approach has been implemented and evaluated in a case study. The approach, the case study and evaluation results are described in this paper. This work will be of interest to organisations or individuals collecting IK with the purpose to make the knowledge available for further exploration.

The paper is structured as follows: the research method is described in Section 2. This is followed by a description of the proposed approach in Section 3. The case study and how the approach was applied in the case study are described in Section 4. Sections 5 and 6 contain the detail of the evaluation and results. A discussion of the results is presented in Section 7 and the paper is concluded in Section 8.

2. Method

An approach to enrich the IK collection process through the introduction of structured metadata was developed and applied in a real-world problem area in the form of a case study (Walsham 1995). The applied approach was evaluated to ascertain the applicability and effectiveness of the approach to enable appropriate improvement.

The evaluation process employed techniques of observations, interviews and document reviews of the developed questionnaires being used for IK collection instruments. Interviews were used as a source of primary data by the researchers to seek data directly from a participant or interviewee by asking questions. For the purpose of this paper, semi-structured interviews were done to get more insight on the process that the IK recorders followed and to understand the challenges and issues that they encountered during the IK collection process. Participant observation was used to assess and learn how the defined processes and instruments were applied in the field. During participant observation the observer took part in the situation being studied rather than merely observing as in the case of a non-participant observer (UCEL 2004, Cooper 2004).

The potential of the enhanced effectiveness of exploration of IK collected with the aid of structured metadata has been assessed in the following manner. A number of structured exploratory queries were defined based on the articulated needs of researchers interested to explore IK in the selected domains of interest. A subset of the IK collected during the piloting phase was captured into the National Indigenous Knowledge Management System (NIKMAS) system. The defined queries were launched using two techniques and the results were compared. The two querying techniques are: (1) launching the queries as keyword searches over the free-form information of the recorded story only; and; (2) launching structured queries over the information collected using the questionnaires and captured as metadata describing the detail of the stories.

3. The approach

For this study, an approach was developed to enrich the IK collection process through the introduction of structured metadata. The approach consists of a sequence of activities that should be followed. The activities that comprise the approach are briefly described in the following list:

(1) *Define conceptual model:* The first step is to define a conceptual model of the domain of interest that will be covered by the IK collection effort (also referred to as the theme). The conceptual model is developed in collaboration with experts in the selected domains. The conceptual model describes the elements of interest in the domain and the relationships between them in a vocabulary familiar to the domain experts.

(2) *Metadata definition:* The conceptual model developed in step 1 is used to define a metadata structure that will be used to describe the collected IK in terms of the structured elements of interest in the specific domain. This metadata definition will be used for the design and implementation of the system to capture the collected IK and will be employed during exploration.

(3) *Questionnaire design:* The metadata definition developed in step 2 is used to design questionnaires to be used during the IK collection process. The questionnaires will be used by the person(s) conducting the IK collection interview

(referred to as IK recorders) to prompt the IK holder about certain aspects that he/she did not mention in the story in order to increase the completeness of the collected IK. The questionnaire also enables the IK recorder to classify certain elements in the disclosed information to increase the accuracy of the collected knowledge.

(4) *Process definition:* The process to be followed during IK collection is developed based on the context of the environment where the IK will be collected and the selected domain of interest.
(5) *Piloting:* The processes and questionnaires defined in steps 3 and 4 are implemented at selected IK collection sites. During the piloting phase the defined processes are followed and the questionnaires are used during the IK collection process.
(6) *Evaluation and refinement:* The effectiveness of the processes and questionnaires implemented during the piloting phase are evaluated and refined accordingly. This step will be repeated until the effectiveness of the defined processes and questionnaires is at a satisfactory level.
(7) *Implementation:* The refined processes and questionnaires are implemented wider than the initial pilot sites.

4. Case study

The approach described in Section 3 has been developed and implemented as part of a project undertaken by the National Indigenous Knowledge Systems Office (NIKSO) in the Department of Science and Technology, South Africa (DST). The design aspects of the implementation of the approach, as described in Section 3 points 1 to 4, has been implemented as part of the design of the NIKMAS system. The piloting, evaluation and refinement of the instruments in communities, as described in Section 3 points 5 and 6, have been performed in a pilot site in Vhembe, Limpopo. The project in this case study is still in the refinement phase, therefore the implementation activity of the approach, as mentioned in Section 3 point 7, has not been performed yet. The NRS initiative is described in more detail in Section 4.1. The implementation of the approach in the initiative is described in Section 4.2 to 4.6.

4.1. The NRS initiative

The case study described in this paper is based on the National Recordal System (NRS). The NRS initiative was raised in the Indigenous Knowledge Systems (IKS) policy that was adopted by the South African government in 2004 (Department of Science and Technology 2004). This policy provided a framework for the recognition, affirmation, development, promotion and protection of IKS in South Africa. The aim of the NRS initiative is to record, document, preserve, protect, and promote IK and to secure the legal rights of IK holders where appropriate. The National Indigenous Knowledge Management System (NIKMAS) is being developed to support NIKSO in realising the aim of the NRS initiative. The NIKMAS consists of a number of components including, amongst others, a number of paper based questionnaires being used as instruments during IK collection, a stand-alone cataloguing tool that is used off-line in communities, and a web-based repository based on digital library technology. The paper-based questionnaires and stand-alone cataloguing tool has been deployed in pilot sites since 2009. The first version of the web-based

NIKMAS repository has been completed in April 2011, but due to the sensitive nature of the collected information on the system, it is not yet available to the general public.

4.2. Conceptual model definition

The first step when collecting IK is to recognise and identify the different kinds of IK and choose or define the area of focus (Tella 2007). The domains of interest for the first phase of the NRS initiative were identified as African Traditional Medicine (ATM) and Food Security (FS). A conceptual model was developed for the selected two themes through collaboration with stakeholders in the domains and scientists interested in research into the domains. The stakeholders and scientists originated from state departments, higher education institutes and science councils. The collaboration was based on the analysis of the aspects of the domain that were deemed important to be addressed in the information to be gathered in order to support useful interrogation, decision support and scientific discovery. The conceptual model was developed graphically using Object Role Modelling (ORM) (Halpin and Morgan 2008) and describes the elements of interest in the domains of ATM and FS and the relationships between them. A subset of the conceptual model constructed for the themes of ATM and FS is depicted in Figure 1.

The conceptual model depicted in Figure 1 shows a subset of the elements of interest in the themes of ATM and FS. It is important to note that the two themes overlap significantly

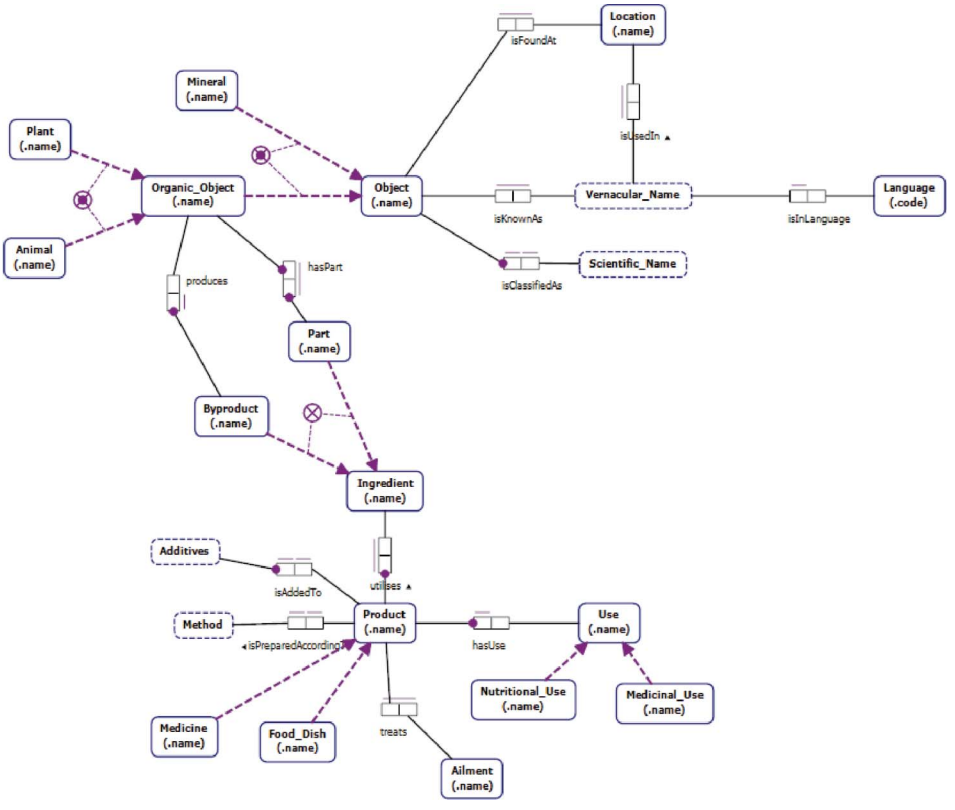

Figure 1. Sample conceptual model for the ATM and FS themes.

and share many elements such as Ingredients that are Parts or By-products of Organic objects that in turn can be either Plants or Animals. The distinction between these two themes is in the type of Product that can either be a Medicine (in the case of ATM) or a Food dish (in the case of FS). Provision also had to be made for functional food, which is food that also has a medicinal use. It will not be the case that additional themes, to be added later, will overlap so extensively with the existing themes. This implies that the conceptual model may be extended significantly with each additional theme.

4.3. Metadata definition

Metadata is defined as data describing an information resource (Hillmann 2005). In the context of the proposed approach for IK collection, the concepts that were defined in the conceptual model in Section 4.2 were used to define the metadata with which the collected IK stories will be described and classified. Dublin Core (Weibel and Koch 2000) was selected as the metadata standard to be used for the metadata definition for the IK stories. This selection was because (1) Dublin Core is used by most metadata based systems for interoperability and data exchange; (2) it is not prescriptive and will therefore not impact on the definition of metadata for IK; and (3) it allows for qualification to support more specialised interrogation. The metadata for the IK stories required for standard interrogation and data exchange were defined according to the Dublin Core standard.

Concepts specific to the domains of ATM and FS were defined as thematic metadata that are supplementary to the Dublin Core metadata. The thematic metadata will be used for specialised, theme based information classification and interrogation.

The use of controlled vocabularies during the population of the metadata increases the quality of captured metadata and improves the effectiveness of data interrogation through searching and browsing. Controlled vocabularies were defined as far as possible during the metadata definition phase of the case study. International standards were used for classification in the defined metadata where appropriate such as the International Patent Classification (IPC) codes of the World Intellectual Property Organisation (WIPO) (WIPO 1971) and the International Classification of Diseases codes of the World Health Organisation (WHO) (WHO 1994). Care was taken to prevent that the defined controlled vocabulary restrict or influence the IK collected.

This defined metadata structure was used in the design and implementation of the NIKMAS system that is used to capture the collected IK and will eventually make the recorded IK available for dissemination and exploration. The metadata defined for the case study is briefly described in the next sub sections.

4.3.1. Dublin Core elements

The Qualified Dublin Core standard allows for a selection of standard elements that are used by any metadata based system that uses Dublin Core. The Standard Dublin Core elements defined for the IK stories are:

(1) DC:Title – The name of the story.
(2) DC:Subject – The subject of the story, including a list of keywords.
(3) DC:Description – A description of the story.
(4) DC:Type – The type of story.
(5) DC:Source – The source of the story (i.e. the person who told the story).
(6) DC:Relation – Relations to any other stories in the repository.

(7) DC:Coverage – The geographical or temporal coverage of the story.
(8) DC:Creator – The person who created the story in the repository.
(9) DC:Publisher – The person who made the story available in the appropriate access level.
(10) DC:Contributor – Any other persons who contributed to the story.
(11) DC:Rights – The access level of the story in terms of who can access it and who has the rights of the story (which will be the IK holder or the community of the IK holder).
(12) DC:Date – Any dates related to the story including the date the story was collected.
(13) DC:Format – The format of the entry (e.g. text, video, audio etc).
(14) DC:Identifier – An internal unique identifier of the story.
(15) DC:Language – The language the story is captured in.

In addition to the standard Dublin Core elements, qualified elements were defined to further qualify the standard elements for more specific search and interrogation processes. In the interest of space all the qualified Dublin Core elements will not be listed, but for example, the Standard Dublin Core element Coverage is qualified further into Place for a description of the location; Spatial for the GPS location; Province for the name of the province name etc. The naming convention for Qualified Dublin core is defined as follows:

DC:<Standard element name>:<Qualified element name>

E.g. the Qualified Dublin Core elements for the GPS location are defined as:

DC:Coverage:Spacial:Longitude
DC:Coverage:Spacial:Latitude

4.3.2. Thematic metadata elements

The thematic metadata is used to capture the theme specific metadata as defined in the conceptual model. The thematic metadata was defined for the themes of ATM and FS and contain elements specifically of interest of the selected themes. Some example elements of the thematic metadata for African Tradition Medicine (ATM) are:

- Name – The name of the medicine.
- Ailment.Description – A description of the ailment the medicine can be used for.
- Ailment.Classification – The classification of the ailment according to the ICD.
- Preparation.Method – A description of the method to prepare the medicine.
- Preparation.Additives – A description of additives used during the preparation of the medicine.
- Preparation.Rituals – Description of the rituals during the preparation method.
- Preparation.Ingredient.Item.Type – The type of item the ingredient is sourced from (i.e. plant, animal or mineral).

Some example elements of the thematic metadata for Food Security (FS) include:

- Name – The name of the food.
- Consumption – A description of how or when the food is consumed.

- Preparation.Ingredient.Item.Type – The type of item the ingredient is sourced from (i.e. plant, animal or mineral).
- Item.VernacularName – The vernacular name of the plant, animal or mineral.
- Item.Acquiring.Method – An indication if the item is cultivated or collected in the wild.
- Item.Habitat – A description of the habitat of the plant, animal or mineral.
- Item.Cultivation – A description of how the plant is cultivated.
- Item.Harvesting – A description how the plant is harvested.

4.4. Questionnaire design

The metadata design described in Section 4.3 was used to develop questionnaires for the selected domains of ATM and FS. The aim of the questionnaires is to aid the IK recorder during the IK collection interview to ask specific questions in order to collect the IK story to the appropriate level of detail as defined in the conceptual model. The questionnaires include instructions to the IK recorders and example leading questions that will aid the IK recorder and the source to explore the IK story in the required level of detail. For example, to explore the shelf life of a prepared medicine the questionnaire will include the following leading questions:

(1) How do you store the medicine you prepared?
(2) How long can you store the medicine after preparation?
(3) Where do you store the medicine after preparation?
(4) Why do you store the medicine like this?

Elements defined during the metadata definition that are too specific, and will not be productive during the interview such as the classifications of diseases are not included in the questionnaires. Instead, descriptive questions are asked that will allow classification at a later stage.

The questionnaires provide instructions for certain media to be recorded by the IK recorders such as the picture of the final product, a picture of the plant used, a picture of the parts of the plant used, a video of the preparation method, etc. The questionnaires also had to provide sufficient space where the answers to the questions can be recorded for later classification and capturing into the NIKMAS system. Figure 2 depicts an example of a page from the questionnaire for the collection of IK in the ATM theme.

The following challenges were experienced during the development of the questionnaires:

(1) The sequence of the questions is extremely important to follow the logical flow of the information and story from the IK holder.
(2) Since the number of ingredients used for a specific food dish or medicine is unknown, additional pages had to be made available to record the detail of the ingredients and IK items used. These additional pages had to be added to the questionnaire manually as needed.
(3) The questions on the questionnaires had to be rephrased to ensure that IK recorders and IK holders understand the intended meaning.
(4) Very large controlled vocabularies such as IPC and IDC codes were too complex to include in the questionnaire.

4. Plant/Animal/Inorganic Material Detail (Please fill in a form for each plant used and attach all the forms in one (1) report. Refer to step 2 of the instructions).		
Indigenous items used	4.1.1. What do you use to prepare the medicine? ☐ Plant ☐ Animal ☐ Inorganic material	
Ingredients used	4.1.2. What parts of the item do you use to prepare this medicine? ☐ Parts (e.g. Flower, leaves) ☐ By-products (e.g. Gum)	
Vernacular Name	4.1.3. What do you call the plant / animal/Inorganic material? (Vernacular name)	
	4.1.4. In what language is this name?	
	4.1.5. Where is this name used?	

Figure 2. Example questionnaire snippet for ATM.

4.5. IK collection process definition

The processes to be followed in order to collect the IK in the communities have been defined. This included pre collection activities such as engaging the management structure of a specific community, setting up a steering committee in the community responsible for coordinating the IK collection and appointing individuals from the communities as IK recorders. The detail of these processes is outside the scope of this paper.
The defined process for IK collection comprises of two sub-processes:

(1) Cataloguing process: This process includes the identification of potential IK holders in the community, where they reside and what IK they claim to have. The detail of the cataloguing process is outside the scope of this paper.
(2) Detail IK collection process: This is the actual process of visiting the IK holder, conducting the IK collection interview and recording the information through video, photographs and the appropriate questionnaire.

The defined detailed IK collection process consists of the following steps:

(1) *Identify IK holder to visit.* The IK holders to be visited are identified from the information captured through the cataloguing process. The catalogue information is evaluated by appointed members of the community and the individuals to be visited for detailed IK collection are identified.
(2) *Assign IK recorder.* The IK recorder that will visit the IK holder is selected from the IK recorders appointed in the community.
(3) *Make visit appointment*. The IK recorder makes an appointment with the IK holder for the IK collection visit.
(4) *Perform IK collection interview.* The IK recorder visits the IK holder. The IK recorder must firstly ensure that the IK holder is fully informed of his/her rights, comfortable with the process and equipment and has signed the required agreements. During the IK collection interview the IK holder is invited to tell his/her

story (which is captured on video or audio). The story is supplemented with questions asked by the IK recorder based on the defined questionnaires. The video footage, photographs and GPS locations are recorded as required.

(5) *Verify the recorded detail with the IK holder.* The photos, videos and text recorded during the IK collection interview are verified with the IK holder after the interview to ensure that the detail is a true reflection of the disclosed knowledge.

(6) *Verify recorded detail with IK owner.* In certain instances the IK is owned by the community. In this case a specific person in the community is appointed to tell the story and provide the detail during the IK collection interview. The detail collected during the interview is then verified firstly with the person who told the story and secondly with the community itself in order to verify the accuracy and completeness.

(7) *Send information and media for capturing in NIKMAS.* The completed questionnaires with the associated media (photographs and videos) are sent to the nearest Indigenous Knowledge Systems Document Center (IKSDC) for classification and capturing into the NIKMAS system.

The capturing and classification of the recorded information and media into the NIKMAS system is part of the curation process. This process includes tasks such as classification of the information into the controlled vocabularies as defined in the metadata definition, language translation where necessary, clean-up of the recorded media and definition of additional metadata such as abstracts etc. The detail of this process is outside the scope of this paper.

4.6. Piloting

The developed questionnaires and defined processes were piloted in a community in the Vhembe district in Limpopo. This pilot site was selected because of the following: firstly, there was an existing, active network of trained recorders from the community collecting information on small enterprises that could easily be extended to include IK; and secondly, the area is rural with a rich heritage of knowledge in the areas of traditional medicine and traditional food. This made the establishment of the pilot site cost and time effective and it may be postulated that the results from the pilot could be applicable to other rural areas.

The cataloguing process was conducted in the Vhembe district in 2010 and during this time 385 IK holders were catalogued. The piloting of the detailed IK collection process started in February 2011. Over a period of four months 22 IK collection interviews were conducted using the defined questionnaires. The first versions of the questionnaires were used for three months in the field. The questionnaires were refined based on the feedback from the IK recorders and the quality of the completed questionnaires received. The second versions of the questionnaires were rolled out for use.

5. Evaluation

For the evaluation of the impact of the defined processes and the developed questionnaires on the IK collection process, two researchers from the CSIR visited the IK recorders in the Vhembe district in May 2011. During the evaluation visit the researchers observed a number of IK collection interviews and conducted a semi-structured interview with a group of IK recorders to discuss their experiences on following the defined IK collection processes and using the questionnaires. The IK holders were members of the communities in the

district who were trained in video recording, writing skills, socioeconomic transformation, the IK collection processes and the IK collection questionnaires to be used.

The observations involved observing the process of how IK recorders conduct the IK collection interview, accompanying the IK holders to collect the ingredients for traditional medicine in the wild and recording the preparation of the medicines at the home of the IK holder.

The semi-structured interview was conducted with six IK recorders who were responsible for IK collection in the Makhado community in the Vhembe district. The aim of the interview was to explore how the defined processes were applied in the field and how effective the defined questionnaires were in conducting the IK collection interview and recording the structured information from the IK holder's story. During the interview a document review was conducted on the questionnaires to assess the usefulness and comprehensibility.

The effectiveness of exploration of the IK collected in the structured format as defined through this approach was evaluated by capturing 10 of the IK entries, collected during the piloting process using the defined questionnaires, into the first version of the NIKMAS system. A number of queries, defined by the scientists interested in exploring the IK captured in the system, were launched as structured queries over the metadata as well as keyword queries over the free-format story text.

6. Results

The results of the evaluation were analysed, collated and grouped into feedback on the defined IK collection process, feedback on the use of the developed questionnaires and feedback on the effectiveness of the exploration of the captured IK. The evaluation results are described in the following sub sections.

6.1. IK collection processes

The detailed IK collection process as defined in Section 4.5 was mostly followed as defined. Based on the observations and the interviews conducted, the following lessons were learned regarding the IK collection process:

(1) Continuous interactions and follow-up between the IK recorders and IK holders are needed to build a trust relationship before the IK collection visits.
(2) A well-structured steering committee, appointed within the community, is needed to interact with the community members before IK collection processes commence. In the absence of a steering committee, the IK holders are hesitant to disclose their IK.
(3) Most IK holders require a letter of recommendation from a village chief before they can disclose any IK. Without a letter of recommendation from the chief, some IK holders refused to disclose their IK.
(4) IK holders may refuse to disclose their IK at the time of the IK collection visit. This may result in wasted time in preparation and unnecessary travelling.
(5) Communities and IK holders can be spread over a large geographical area. Travelling arrangements are often difficult to manage.
(6) An IK collection interview usually needs the involvement of three IK recorders. One to conduct the interview, one to take the photographs and one to capture the video recording.

(7) An IK collection interview can be extremely time-consuming (sometimes more than three days). This may result in cancelled appointments, incomplete information or IK holders losing interest in the process. It should be investigated how the interview process can be made more effective without losing the richness of the knowledge.
(8) The collection of the ingredients for a specific medicine or food dish may require a lot of time and cover large distances. This is because it may involve going to the forest or into the mountains or having to travel to another village to collect an ingredient. Time and transport must be planned accordingly.
(9) In some cases it is necessary to buy additional ingredients required to demonstrate the preparation of a product such as a food dish. It might be necessary to provide IK recorders with a small budget for this purpose.
(10) The IK recorders feel it is necessary to present the IK holder with a small gift of appreciation for the time they spent during the IK collection interview as it is part of the culture and almost expected. The implication of giving gifts in this context needs to be investigated further.
(11) The IK recorders may need additional training in interviewing skills to effectively conduct the IK collection interview.
(12) The IK recorders may need additional training on video recording and photography to increase the quality of the media captured during the interviews.

6.2. Questionnaires

The IK holders found the questionnaires helpful during the IK collection interview as a tool to lead the conversation and a prompt to explore the IK in more detail where necessary. However, a number of issues were raised that need to be addressed in the following versions.

The issues are:

(1) The questionnaires are in English, but the IK collection interview is conducted in the native language of the IK holder. This implies that the IK recorder must read the questions in English, ask it in the native language, listen to the answer in the native language, translate the answer into English and capture it in English on the questionnaire. Due to this elaborate process information may be missed or misrepresented in the completed questionnaire. It should be investigated to provide the questionnaire in the native languages allowing the complete interview to be conducted and recorded on in the same language.
(2) The IK recorders and IK holders found it very difficult to use the classifications defined for certain information items such as the part of a plant or the classification of a disease. This is because of technical terms being used that are not in the general vocabulary of the community. It was decided to remove complicated classifications from the questionnaires and ask for sufficiently detailed answers instead that will inform the classification at a later stage.
(3) Elements of interest that cannot be supplied by an IK holder (e.g. the scientific name of a plant) must not be asked in the questionnaire.
(4) The process of capturing the answers of open-ended questions on the paper based questionnaire during the interview may result in a misinterpretation of the answer by the IK recorder. Possible interventions to improve the effectiveness of this process must be investigated.

(5) Keeping the completed questionnaire and the associated recorded media together is a difficult task and needs very careful management. This process must be addressed and mechanisms must be implemented that will make this more reliable.
(6) The process of capturing the answers to the questions on a paper-based questionnaire and sending this for capturing at another location is not secure since the possibility exists of losing a completed questionnaire in the process. An investigation into a technology based questionnaire that can be captured directly in the field during the interviews must be conducted.

6.3. Effective exploration

Due to the fact that the exploration features of the NIKMAS system are still under development, the authors can only report on initial experimentation. The initial exploration of the IK captured in the NIKMAS system in the structured format, as defined by the metadata definition and collected using the defined questionnaires, proved to be very promising. Due to the structured nature of the information, the browsing of the information can be presented in certain areas of interest. For instance the stories can be grouped either by plant used or by disease treated (in the case of ATM). This would not be possible if the stories were merely collected as free-form stories.

For the evaluation of the querying potential of the structured information, a number of exploration questions were submitted on the captured information. Structured queries were submitted over the defined metadata while the same phrases have been submitted as keyword searches over the free-form text of the related IK story. Examples of the queries include:

- Give me the plants that are used as medicine treating diseases of the respiratory system
- Give me the food dishes prepared from peanuts and beef
- Give me the medicines that are prepared from the leaves of the plant commonly known as the Asparagus tree

In the case where a structured query was launched over the metadata of the stories, the results were very accurate and returned the expected stories. However, when the query phrase was launched as a keyword search over the free-form IK story itself, the results were less accurate and depended on the following:

- The phrases used by the source when telling the story must correspond to the phrases used in the keyword search for the system to return the result.
- If the source used a different, vernacular name for the plant in the story, the system will not return the story as a result.
- The stories are recorded in different local languages. The result will therefore depend on the language of the query and that of the stories. Only the matching stories in the same language of the query will be returned.
- Since controlled vocabularies are not used in the freeform stories, the same aspect can be described in different ways making querying inaccurate.
- International codes such as IPC and IDC cannot be used in queries over the free-form stories.

Lastly, ontology-based querying tools can be used to explore the captured information in even further detail, based on the semantics of the information rather than the words being

used. For example, the query asking for plants which aerial parts are used in medicine to treat respiratory diseases as mentioned in Section 1, will return a story describing using the leaves of a plant to treat coughing as well as a story describing using the fruit of another plant to treat asthma. This is not possible with standard keyword searches.

7. Discussion

Constructing the conceptual model proved a very useful exercise for stakeholders to consider and discuss what elements should be addressed in the theme. This enabled the stakeholders to create a shared view of the theme. The fact that the two selected themes (ATM and FS) overlap considerably made the task easier, but this cannot be assumed for other themes to be added. One criticism of the application of this activity in the NRS is that only high-level stakeholders from government departments and higher education institutes were included. It will be imperative to include practitioners in the field in this activity in the future.

The metadata definition activity is advantageous since the formal definition of the metadata using an international standard ensures structured exploration and future interoperability with other metadata based systems using Dublin Core. Care must be taken, however, to not overanalyse the theme and make the metadata so elaborate that the capturing of the IK is too complex.

The introduction of the questionnaires into the IK collection process proved a useful tool for IK recorders. However, there are two challenges that need to be addressed: firstly, the issue of language needs careful consideration. The questionnaires must be translated into the local language to enable the interview to be performed in the native language of the source. The translation of the information can either be done by the IK recorder after the collection interview, or later by professional translation services. The pros and cons of both scenarios should be investigated and the best solution should be implemented. Secondly, the fact that the IK recorder must capture the answers to the questions in the questionnaire during the interview must be addressed since it breaks the flow of the interview and results in inaccurate capturing of information. An approach where the recorder merely records the answers to the questions in the questionnaire on video or audio during the interview and transcribes it later is being developed and will be tested in the next six months. Care should be taken not to overburden the IK recorder with too many administrative tasks or to create an unnecessary bottleneck that may result in a very time-consuming IK collection process.

The defined IK collection process was received well, but it must be sufficiently flexible to be adjusted according to the context in other communities. In particular, the process should also be piloted and evaluated in urban areas.

The pilot conducted in the Vhembe district, Limpopo was very small. It provided the researchers with valuable insights into the shortcomings of the proposed approach. The approach should be adjusted accordingly and applied in other pilot sites to get a more stable view of the effectiveness of the approach.

The results in this study were based on a small pilot, and the technology to be used (the NIKMAS) is still under development. As the approach is rolled out in other areas and communities, the results will inform the shape of the approach and technology in future.

The experimentation on the efficiency of exploring the information collected using the approach was very preliminary for this study. Although the results are very promising currently, a more extensive evaluation will have to be performed on this aspect as the technology matures.

8. Conclusion

The defined approach to enrich the IK collection processes with the introduction of structured metadata to be collected in addition to the IK stories has been described. This approach entails tasks of (1) defining a conceptual model for the domain(s) of interest; (2) defining the metadata to be collected; (3) designing structured questionnaires to be used during the IK collection interview; (4) defining the IK collection processes; (5) piloting the defined processes and questionnaires and refining them based on evaluation results; and (6) implementing the refined processes and questionnaires.

This approach has been applied successfully on the NRS initiative as a case study. For the case study the defined approach was applied in the context of the NRS initiative focusing on two domains of interest namely African Traditional Medicine (ATM) and Food Security (FS).

A context model was developed for the selected themes through collaboration with domain experts and relevant scientists. The conceptual model informed the definition of the metadata to be collected for the two domains, based on the Dublin Core standard. The defined metadata will be used for structured exploration of the collected IK.

The metadata definition informed the development of theme specific questionnaires to be used during the IK collection interviews to ensure the completeness of the collected information. An IK collection process was developed to be followed by IK recorders. The processes and questionnaires were piloted in the Vhembe district in the Limpopo province.

The pilot study revealed that the proposed approach is successful in effectively collecting IK in a structured way. However, the following issues need to be addressed:

- Practitioners in the themes of interest must be involved during the definition of the conceptual model.
- Defined metadata must be kept as simple as possible.
- IK collection interviews must be conducted in the local language of the source.
- The issue of translation of the collected metadata must be addressed without creating bottlenecks in the process.
- Technology must be applied to enable an IK collection interview without interruption for the purposes of recording.
- The processes must be adjustable to cater for the reality in the communities.
- This approach must be tested in more pilots to obtain stable results.

The effectiveness of exploration of the collected IK based on structured metadata was evaluated on a subset of collected IK. The preliminary results show great potential for (1) context-based browsing of collected IK; (2) structured querying of collected IK; and; (3) semantic-based querying of IK retrieving results based on the meaning of the information. This method of exploration has been shown to be possible and much more effective than mere keyword searches over the full text of free-form IK stories.

Further research will entail addressing the lessons learned and identified issues through refining the processes and questionnaires; implementing the approach in more IK collection sites; investigating the development of technology interventions to more effectively support the IK collection process; and; completing the exploration feature in the NIKMAS system to further evaluate the effectiveness of exploration of IK collected through structured metadata.

Acknowledgements

The authors wish to thank NIKSO DST who is the owner of the NRS initiative and primary sponsor of the project for their inputs, direction and ongoing support. In addition, a special thanks to the IK recorders in Vhembe for their ongoing efforts in cataloguing and detailed IK collection in their communities and their willingness to contribute to this study.

Notes on contributors

Ronell Alberts is a senior researcher at the CSIR Meraka Institute in South Africa. She has obtained a Masters degree in Computer Science through the European Masters Program in Computational Logic. Her research fields of interest include software engineering; business and systems analysis; complex databases; knowledge representation and reasoning; and applying semantic web technologies in information systems.

Gugu Khalala is a business analyst and a researcher for the Meraka Institute, at the Council for Scientific and Industrial Research (CSIR) in Pretoria, SA. Gugu specializes in analysis but also enjoys interacting with users and doing fieldwork. She also writes research papers and her article titled 'Demystifying the shrinking pipeline of women in ICT education and careers: a South African case study' was published in the *Journal for New Generation Sciences.* She has worked on various research projects and her latest project that she is working on is the National Recordal System.

Onkgopotse Molefe is a researcher at the CSIR Meraka Institute. She is pursuing a career path in research, business processes and business analysis. She holds an MTECH degree from University of South Africa (UNISA) in Information Technology acquired in 2009 representing a completed dissertation focusing on Open Source Software (OSS) and Process Reference Models (PRMs), which happens to be her areas of interests in research. She has produced five publications in the areas of OSS and PRMs. She also holds a BTECH degree in Information Technology acquired in 2006 where she specialized in Software Development at Tshwane University of Technology (TUT).

References

Agrawal, A., 1995. Dismantling the divide between indigenous and scientific knowledge. *Development and Change*, 26, 413–439.

Bidwell, N.J., Reitmaier, T., Marsden, G., and Hansen, S., 2010. Designing with mobile digital storytelling in rural Africa. CHI 2010: Storytelling.

Cooper, J., Lewis, R., and Urquhart, C., 2004. Using participant or non-participant observation to explain information behaviour. *Information Research*, 9 (4), paper 184. Available from: http://InformationR.net/ir/9-4/paper184.html

Department of Science and Technology, 2004. *Indigenous knowledge systems policy* [online]. Available from: http://www.wipo.int/tk/en/databases/creative_heritage/policy/link0007.html [Accessed 30 May 2011].

Dictionary.com, n.d. Story [online]. *Dictionary.com.* Available from: http://dictionary.reference.com/browse/story [Accessed 30 May 2011].

Feng, Y., Wu, Z., Zhou, X., Zhou, Z., and Fan, W., 2006. Knowledge discovery in traditional Chinese medicine: state of the art and perspectives. *Artificial Intelligence in Medicine*, 38, 219–236.

Global Exchange, 2001. Biopiracy a new threat to indigenous rights and culture in Mexico. *Global Exchange* [online]. Available from: http://www.globalexchange.org/countries/americas/mexico/biopiracy.pdf [Accessed 30 May 2011].

Halpin, T. and Morgan, T., 2008. Information modeling and relational databases. 2nd ed. Morgan Kaufman.

Hillmann, D.I., 2005 (November). Using Dublin Core. Available from: http://dublincore.org/documents/usageguide/ [Accessed 30 May 2011].

Johnson, O., 2002. Venezuelan project establishes indigenous plant database. *British Medical Journal*, 325(7357), 183.

Liu, Y. and Sun, Y., 2004. China traditional Chinese medicine (TCM) patent database. *World Patent Information*, 26, 91–96.

Okorafor, C.N., 2010. Challenges confronting libraries in documentation and communication of indigenous knowledge in Nigeria. *The International Information & Library Review*, 42, 8–13.

Rao, S.S., 2006. Indigenous knowledge organization: an Indian scenario. *International Journal of Information Management*, 26, 224–233.

Ruiz, M., Lapeña, I., and Clark, S.E., 2004. The protection of traditional knowledge in Peru: a comparative perspective. *Washington University Global Studies Law Review*, 3, 755–797.

Sen, B., 2005. Indigenous knowledge for development: bringing research and practice together. *The International Information & Library Review*, 37, 375–382.

Sillitoe, M., 2009. Future of indigenous knowledge research in development. *Futures*, 41, 13–23.

Telfer, W.R. and Garde, M.J., 2006. Indigenous knowledge of rock kangaroo ecology in Western Arnhem land, Australia. *Human Ecology*, 24, 379–406.

Tella, R.D., 2007. Towards promotion and dissemination of indigenous knowledge: a case of NIRD. *The International Information and Library Review*, 39, 185–193.

Traditional knowledge digital library, n.d. About TKDL [online]. Available from: http://www.tkdl.res.in/tkdl/langdefault/common/Abouttkdl.asp?GL=Eng [Accessed 30 May 2011].

UCEL, 2004. Universities collaboration in eLearning glossary of key terms [online]. Available from: http://www.ucel.ac.uk/rlos/pnpobservation/2.html [Accessed 30 May 2011].

Walsham, G., 1995. Interpretive case studies in research: nature and method. *European Journal of Information Systems*, 4(2), 74–81.

Weibel, S.L. and Koch, T., 2000. The Dublin Core metadata initiative. *D -Lib Magazine*, 6(12), 1082.

WHO, 1994. International classification of diseases (ICD) [online]. Available from: http://www.who.int/classifications/icd/en/index.html [Accessed 30 May 2011].

WIPO, 1971. International patent classification (IPC) [online]. Available from:http://www.wipo.int//classifications/ipc/en/ [Accessed 30 May 2011].