A software architecture for an indigenous knowledge management system

Thomas Fogwill
CSIR Meraka Institute
Meiring Naude Rd
Pretoria, South Africa
tfogwill@csir.co.za

Ilse Viviers
CSIR Meraka Institute
Meiring Naude Rd
Pretoria, South Africa
iviviers@csir.co.za

Louis Engelbrecht
CSIR Meraka Institute
Meiring Naude Rd
Pretoria, South Africa
lengelbrecht@csir.co.za

Chris Krause
CSIR Meraka Institute
Meiring Naude Rd
Pretoria, South Africa
ckrause@meraka.org.za

Ronell Alberts
CSIR Meraka Institute
Meiring Naude Rd
Pretoria, South Africa
ralberts@csir.co.za

ABSTRACT

Indigenous knowledge (IK) is defined as the unique, traditional and local knowledge of people within a particular area. Typically, IK is stored in peoples' memories and passed down orally from generation to generation. However, with issues such as rapid urbanisation, commercialisation, marginalisation, westernisation and changes in technology, traditional IK systems are in decline.

IK holds significant value, both in terms of cultural heritage and identity, and in terms of its potential economic and environmental benefits. As such, there is a strong drive to capture and preserve IK worldwide.

The National Recordal System is an initiative undertaken by the South African Government's Department of Science and Technology to enable communities, guilds and other holders of IK to record their knowledge holdings for the purpose of preservation, future economic benefit and social good. To support the initiative, a National IK Management System (NIKMAS) is being developed. The system has a core Digital Knowledge Repository (DKR) component that stores IK items, together with a rich set of metadata fields and multimedia files describing the IK. The DKR is based on the Flexible Extensible Digital Object and Repository Architecture, and is enriched with semantic web technologies to support intelligent information retrieval and inference.

This paper presents a high level architectural overview of the NIKMAS software, using a number of architectural views to depict different aspects of the system. The views used are loosely based on Kruchten's "4+1" View Model of Software Architecture, and uses four views: use case view, logical view, data view and deployment view.

This paper is intended to capture and convey the significant architectural decisions which were made during development of the first version of the system. The architecture described here could be considered as a general starting point for similar systems.

Indigenous Knowledge Technology Conference 2011. Windhoek, Namibia.

ISBN 978-99945-72-37-3

General Terms

Design, Reliability, Security, Documentation

Keywords

Indigenous knowledge, indigenous knowledge management system, software architecture

1. INTRODUCTION

Indigenous knowledge (IK) is the local, traditional knowledge held by people of a particular area [11]. It often allows them to effectively utilise and preserve their environment [21]. IK is both cumulative and dynamic. Typically, IK is transferred orally between generations. However, as a result of socio-economic issues, IK is fragile and at risk of being lost [2].

IK is central to the cultural heritage and identity of indigenous people. It also offers significant potential for economic and environmental development and preservation [23]. Because of this value that resides in IK, a number of initiatives in various countries have attempted to implement software systems to capture and preserve IK. A prominent example is the Indian Traditional Knowledge Digital Library (TKDL) [12].

The National Recordal System (NRS) is another such initiative, undertaken by the South African Government's Department of Science and Technology (DST). It strives to capture, preserve and protect South African IK, using a National IK Management System (NIKMAS). The system has a core Digital Knowledge Repository (DKR) component that stores IK items, together with a rich set of metadata fields and multimedia files describing the IK. The DKR is based on the Flexible Extensible Digital Object and Repository Architecture (Fedora) digital library technology, and is enriched with semantic web technologies to support intelligent information retrieval and inference.

This paper presents a high level architectural overview of the NIKMAS software. It is intended to capture and convey the significant architectural decisions which were made during development of the first version of the system, and could serve as a guide for the architectures of similar systems.

The architecture of NIKMAS is presented here using a multi-view model loosely based on Kruchten's "4+1" View Model of Software Architecture [15]. Of Kruchten's views, the development and process views are not presented here. Instead, an additional view is presented to describe the data requirements and design. As such, the system is described here using four views: use case view, logical view, data view and deployment view.

The paper is arranged as follows: section 2 presents background, followed by section 3 which describes the NRS initiative in more detail. Section 4 describes the architectural framework, constraints and principles, and gives an overview of the technical platform. Sections 5, 6, 7 and 8 present the detailed use case, logical, data and deployment views, respectively. Section 9 concludes the paper.

2. BACKGROUND

IK is defined by Grenier [11] as the unique, traditional and local knowledge of people within a particular area. Quiroz [21] defines IK Systems as the knowledge and skills possessed by people in a particular geographic area, which they require to effectively use and preserve their environment. IK is cumulative, having developed over centuries, and dynamic, with new knowledge constantly being added. Typically, IK is stored in peoples' memories and passed down orally from generation to generation. However, with issues such as rapid urbanisation, commercialisation, marginalisation, westernisation and changes in technology, IK systems are in decline [2].

IK holds significant value, both in terms of cultural heritage and identity, and in terms of its potential economic and environmental benefits. The value of IK can be realised in many ways, such as through improved agricultural practices and food security, preservation of the natural environment, economic development through IK-based commercial offerings (products and services), drug discovery and biodiversity conservation [23].

A number of initiatives in various countries have attempted to implement software systems to capture and preserve IK. Some of these, such as the TKDL in India, focus on the capturing of previously recorded IK for the purposes of protection against biopiracy and foreign patents [13]. Similarly, a case study was conducted in Peru in order to evaluate the policy and legal context needed to protect traditional Peruvian knowledge from biopiracy [22]. Other initiatives, such as those in Australia and China, have focused primarily on preservation and conservation of traditional knowledge. A prominent South African example is the NRS, undertaken by the South African DST. The processes of the NRS are supported by the NIKMAS, a software system comprising a semantic digital library. The NRS and NIKMAS form the basis for the work presented in this paper, and are discussed in section 3.

Digital libraries have enjoyed a lot of research and development since 1990. A digital library is a library in which collections are stored digitally and that is accessible via a computer [18]. Digital libraries have evolved to meet several needs, including those of overcoming physical space constraints, reducing costs, spanning physical and geographical constraints, 24 hour availability, support for multiple media and documents types, user empowerment and allowing faster, multi-access information retrieval [1]. Digital libraries also offer opportunities for improved preservation and conservation of materials. This, coupled with fast search technologies, makes digital library technology well suited to use for the storage of documents and media describing IK [23], as evidenced through its use in the TKDL [12] and Desa Informasi [24].

More recently, traditional digital libraries have been augmented with semantic web and web 2.0 technologies. These new digital libraries are called semantic digital libraries, and present opportunities for describing and classifying the items in the digital library more richly, using ontologies and user contributed classification (such as tagging). It also allows the specification of relationships between items, for example between a chapter and the book it appears in. This richer description allows for powerful searching and navigation. Furthermore, the use of these richer descriptive structures allows for better interoperability, flexibility and connectedness of libraries [16].

For the purposes of preserving and protecting IK, the use of semantic digital libraries, as implemented in NIKMAS, is novel. Semantic digital libraries offer an ideal platform for IK repositories, in that they provide the necessary services for storage and long term preservation of IK, but also provide the powerful search and discovery capabilities required to retrieve and infer pertinent information from the library.

A model often used to document software architectures is that of Kruchten [15], which describes a "4+1" view model of software architecture. The model defines a set of views for the architecture, each representing different aspects of the overall architecture. The views defined in the model are: the *logical view*, which describes the design; the *process view*, which deals with the dynamic processing aspects of the design, particularly where concurrent processing is required, and also covers some non-functional requirements such as performance and availability; the *physical view*, which describes the physical deployment environment for the system, including hardware; and the *development view*, which describes how the software is to be broken up into packages and organised during development. In addition to these four views, there is an additional view, which describes a set of *scenarios* (use cases) that describe the core functionality of the software and illustrate how the descriptions of the other views work together to ensure a working system.

3. THE NATIONAL RECORDAL SYSTEM

The NRS is an initiative undertaken by the DST to enable communities, guilds and other holders of IK to record their knowledge holdings for the purpose of preservation, future economic benefit and social good. It is driven by forward looking motivations, seeking to capture, preserve and protect IK, while simultaneously enabling processes to derive benefit from the IK. As such, it goes beyond most other IK database initiatives, which focus solely on preservation.

The NRS strives to be respectful of the rights of IK holders and indigenous communities. IK holders comprise individuals, IK practitioners, communities, guilds, traditional authorities and other entities that can be considered as possessing IK. A principle consideration of the initiative is the recording and annotation of IK in such a way that the IK can be legally protected as intellectual property belonging to the IK holder, thus helping to combat unjust exploitation (such as biopiracy). As such, a critical aspect of the NRS is the reforming of policy and legal regimes to better address the rights of IK holders, including the introduction of sui generis laws governing the individual and communal ownership of IK. These aspect of the NRS are presented in a separate paper.

The goals of the NRS extend beyond the defence and protection of IK. A second key aspect of the NRS involves the promotion of IK in such a way that it enables researchers, educators and commercial entities to interrogate the recorded IK, to identify IK with the potential for economic or social benefit. The NRS then brokers direct access between the researchers and owners of this IK for further research or commercial development. Throughout, the rights of the IK holders are protected, and measures are put in place to protect them and ensure that they benefit suitably. Examples of such measures include benefit sharing agreements with communities and IK holders, restriction of access to confidential IK information, and policy enforcement by requiring researchers to obtain bioprospepcting licences before accessing certain types of IK. In this way, the NRS acts as an honest broker between IK holders and interested third parties.

The third goal of the NRS is to combat the erosion of IK caused by social issues and changes in technological landscape. It preserves IK in a digital form that remains accessible to the original IK holders. It does so by recording and archiving audio/visual content that captures, to the extent possible, the oral, visual and performed aspects of IK.

The NRS gathers IK through two main processes, both of which focus heavily on community engagement. The processes are described in detail in related work, so only an overview is given here.

The first process involves cataloguing the IK that resides within communities and individuals. It identifies which communities and individuals hold IK, where they are geographically, and what IK they claim to possess. It also begins the process of establishing legal agreements with the communities and individuals to ensure that their rights are protected.

The followup process involves the responsible recording, analysis and annotation of detailed information about IK. During this process, aspects of the IK, such as the oral and performed components, are recorded as multimedia (videos, audio and photos). In addition, IK holders are interviewed about their knowledge. The multimedia content is stored for archival and preservation purposes. The interview data is used to annotate the multimedia content. It is also used to guide a process of analysis that translates the IK into an information structure that is suitably authoritative to enable the IK to be legally protected. This information structure also allows the IK to be better understood and navigated by the (western) scientific community. It could thus support, for instance, a prior art search tool for patent offices, or a sophisticated search tool for researchers.

Both processes are conducted using a combination of approaches. One approach focuses on the use of regional IK offices, or points of presence (PoPs), through which communities and individuals can capture their IK into NIKMAS for protection and preservation. These PoPs also act as regional points of access where IK holders and communities can access and manage their IK in the system.

A parallel approach uses the Infopreneur model of van Rensburg, Veldsman and Jenkins [25]. This model empowers local community members to become IK cataloguers and harvesters in their communities, and builds a network around them that will enable them to generate income and earn a livelihood, either directly on a commission basis, or indirectly by their offering of supplementary information services to the community. These “infopreneurs” act as points of contact within the community, allowing IK holders and practitioners to capture, access and manage their own IK.

In order to support these processes, the National IK Systems Office in the DST identified the need for NIKMAS to identify, collect, store and disseminate IK. NIKMAS development is an ongoing project, with some components fully developed, others as prototypes, and some that exist only in design. NIKMAS comprises a number of integrated components, each designed to directly support the goals and processes of the NRS. These components are described in section 6. The intended users of NIKMAS, as well as how they interact with the system, are described briefly in section 5.

4. ARCHITECTURAL OVERVIEW

4.1 Framework

The architecture of NIKMAS is presented here using a framework based on the “4+1” framework of Kruchten [15]. Of Kruchten’s views, the development and process views are not required for this system, and are omitted. Instead, an additional view is presented to describe the data requirements and design. As such, the system is described here using four views: use case view, logical view, data view and deployment view. These views are described below.

The *use case view* maps onto Kruchten’s *scenarios view*, and comprises the core, high-level use cases for system. The details are not presented in this paper, but were captured separately for the project in a detailed requirements specification. Section 5 describes this view.

The *logical view* of the system architecture represents the high-level system design. The important components and subsystems of the system are described, as well as the interactions between them. As this describes the main design of the system, it receives more coverage in this paper than the other views. Section 6 describes this view.

The *data view* is not one of the standard “4+1” views - it briefly describes the persistent data for the system, focusing on the data model for captured IK artefacts and the metadata associated with those artefacts. This view is important in that it describes how the captured IK is described. As such, it is described in some detail. Section 7 describes this view.

The *deployment view* maps onto Kruchten’s physical view, and describes the physical network and hardware configuration required for system deployment. Section 8 briefly describes this view.

4.2 Constraints

NIKMAS is constrained by a set of non-functional requirements, including those of performance and speed, availability, reliability, robustness, fault tolerance, capacity, scalability, security, auditability, legal compliance, maintainability and supportability. These are not presented here in detail.

An important additional architectural constraint is imposed by the requirement that NIKMAS must support distributed deployment, such that data can be captured at distributed PoPs. Thus, data replication and synchronisation is required to keep the central and distributed repositories consistent. This implies that the technology selected for the storage of digital IK artefacts and metadata should support federated search, metadata synchronisation based on open metadata exchange protocols, selective/filtered data replication and synchronisation of multimedia content, and distributed access control.

4.3 Principles

NIKMAS is designed and intended to be a robust, reliable and secure system. As such, technologies are chosen that are: based on open standards, to lower the dependence on any single technology vendor and to ensure longevity of the NIKMAS platform; scalable, to ensure that NIKMAS can scale to support the management of IK nationally; and robust, to ensure that the NIKMAS can operate reliably and that it does not permit the loss nor corruption of data.

In addition, NIKMAS must support complex security models and multi-tiered access control, to ensure that sensitive IK and information can be appropriately protected. It must support the storage of complex, multi-format multimedia content, together with complex and flexible metadata to allow the rich circumscription of IK items. The technology is required to support structured curatorship of captured IK by domain experts, to ensure that captured IK is accurate and of the highest quality, and must make that data available through next-generation search and information management architectures, to ensure that researchers can employ sophisticated query techniques to extract value from the captured IK.

The technology must be capable of multi-lingual interfaces and storing IK artefacts in multiple languages.

It is preferable for the software to be available under an open source licence, to promote the development of a South African high-tech support and development capability, and that it be developed in well-known programming languages, to ensure a continued supply of future talent for ongoing development and maintenance.

4.4 Technical Platform

NIKMAS is a server-side web application developed in Java using the JBoss Seam application framework. The core engine of the DKR is the Fedora Core Repository Service.

Functionality is exposed to users through the Hypertext Transport Protocol (HTTP) as Hypertext Markup Language (HTML) pages. Where required, functionality can also exposed to external systems and services using a variety of integration, data interchange and remote procedure call protocols, including Representational State Transfer (REST) or Simple Object Access Protocol (SOAP).

5. USE CASE VIEW

The use cases for NIKMAS are not discussed in this paper, but were specified elsewhere in a detailed requirements specification. The important high-level requirements of NIKMAS are: management of information about communities, including legal agreements signed with them; search for and access to community information; management of information about IK holders; capture, search and browsing of actual IK data in a secure digital knowledge library; access to other IK repositories.

The main users of the system are:

DST personnel: managing and reporting on captured IK, managing agreements with communities

IK cataloguers, recorders and capturers: capturing and managing information about communties and IK holders, recording IK entries

NIKMAS curators and subject specialists: managing, analysing and annotating IK entries in the repository

Community members and IK holders: viewing, managing and synthesizing their own IK in the repository

NIKMAS system administrators: maintaining the system, managing user accounts, etc.

Unregistered viewers of content: accessing and viewing open, public and promotional IK material

Registered viewers of content: searching and accessing restricted IK content (based on suitable acccess control), identifying and contacting IK holders, entering into agreements with communities and IK holders. Examples include: patent offices, scientists, bioprospectors, third party commercial entities.

6. LOGICAL VIEW

6.1 Overview

A high-level overview of the components and subsystems of NIKMAS is given in Figure 1. These components will be described in more detail in following sections, together with several other important aspects of the overall design.

6.2 Portal

The NIKMAS Portal is the main point of entry for users. It provides an aggregated view across a number of IK resources and websites. The NIKMAS Portal is responsible for handling all interaction between users and the system, and acts as the primary NIKMAS user interface. Users access the NIKMAS Portal over HTTP using their web browsers.

Currently, the NIKMAS portal consists of three implemented HTML user interfaces that provide access to the IK captured and stored in NIKMAS. These are: a simple browsing web interface that allows users to browse the IK within a specific theme; a search web interface that provides access to the standard NIKMAS search and the prototype intelligent (ontology-based) search facilities; a curator web interface that provides facilities for IK curators to capture, edit and manage the content provided by IK recorders. Each of these interfaces will be expanded in future versions.

In addition to these three user interfaces, development is underway on HTML pages for user registration and profile management, as well as an administrative and management web interface that allows NIKMAS administrators to monitor and administer the portal.

The NIKMAS Portal is also responsible for handling user login, authentication, authorisation and logout (via the Security Service), and for managing the user's login session.

The portal communicates with various back-end components and subsystems. All interactions with the back-end pass through the security service, which is responsible for enforcing the NRS access control policy and authorising user actions. The portal gains access to the knowledge items

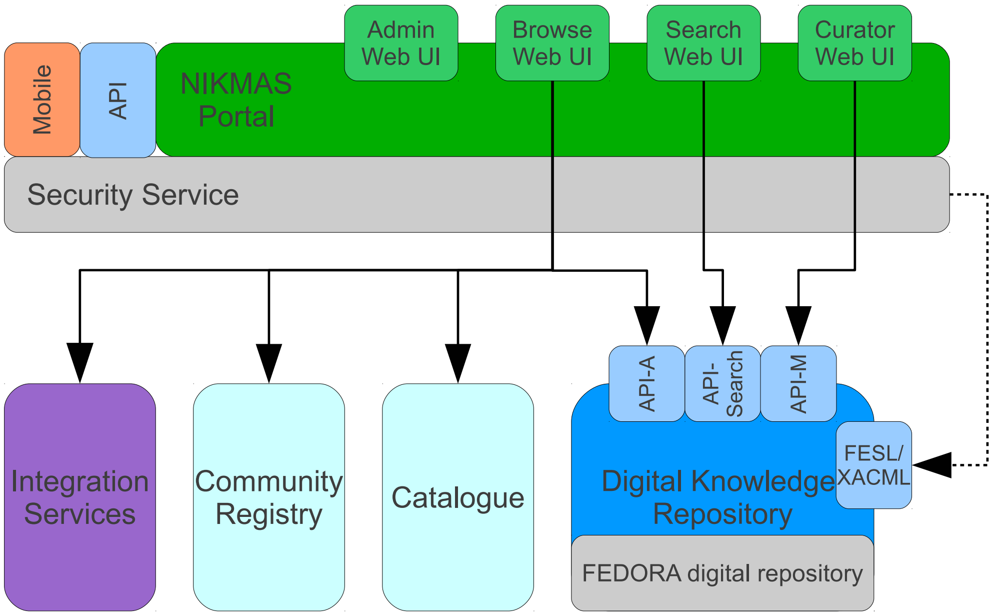

Figure 1: High level system overview

in the knowledge library and invokes required functionality by using one of several Application Programmer Interfaces (APIs) exposed by Fedora. These APIs include: access APIs (API-A and API-A-LITE) to browse and retrieve items; management APIs (API-M and API-M-LITE) used primarily by the curation functions to add, edit and delete items; basic search API (Search) to search for items based on their metadata; and Resource Index Search API (RISearch) used by the intelligent search module to access Fedora's resource index [20].

The portal accesses the catalogue and community register subsystems via direct database access and by invoking remote methods on those subsystems. The portal will also provide a web interface for authorised users to edit/manage items in the catalogue and registry.

In future, the portal will have the ability to display data from external resources using the Integration Services module; this access will be based on (remote) procedure calls against an Integration Services API.

6.3 Digital Knowledge Repository

The DKR is the core of the NIKMAS system. It is built on digital library technology to support the capturing, management, preservation and dissemination of IK content in digital formats. The DKR provides the capabilties to describe, store, index, search and retrieve the digital artefacts representing the IK gathered by IK recorders and captured by IK capturers. The NIKMAS DKR has been fully implemented and is built on the Fedora Core Repository service, which is extended using a custom data model (described in section 7), a custom ontology for the IK domains of interest, and a custom web front-end (described in section 6.2). The knowledge repository and the Fedora technology at its core are described in this section.

The DKR stores and maintains digital audio/visual recordings of the captured IK. While these audio/visual recordings may not be compatible with all real-life manifestations of IK, they are sufficient to capture oral, visual and some practice-based IK. The original recordings are archived and preserved in the DKR, and can later be accessed by authorised individuals (including the knowledge holders). In this way, the DKR acts as a long term preservation archive for IK. It maintains some correspondence with what Bowker calls memory "traces" (the informal and dynamic memory of people) [6, 7], by preserving IK in a near-original form and allowing it to be accessed naturally.

The DKR also stores annotations and metadata that describe or translate the IK into the formal information structures needed for legal protection and scientific interrogation. This information structure is presented in section 7. This information architecture imposes a data bias on the recorded IK, recontextualising it into the domains of intellectual property rights, western law and science. While this representation may be partially incompatible with, or incomplete with respect to, the original form of the IK, it does lend the IK legitimacy in those domains and allows it to be protected and form an integrated part of the NRS's future-looking processes. In this sense, the DKR forms a jussive archive [7] by imposing a formal information architecture on the IK and preferring certain access patterns (that do not necessarily correlate with the memory patterns of the IK holder). Although IK recontextualised this way may be poorly understood by indigenous communities initially, there is the opportunity for them to acquire new digital literacies as part of the NRS processes, which would empower them to derive more benefit from the system. Regardless, the audio/visual original recordings remain accessible to them, providing some degree of digital preservation of their knowledge and allowing them more easily to "reassmeble" their IK in terms and formats they are familiar with.

The Fedora Core Repository service at the core of the

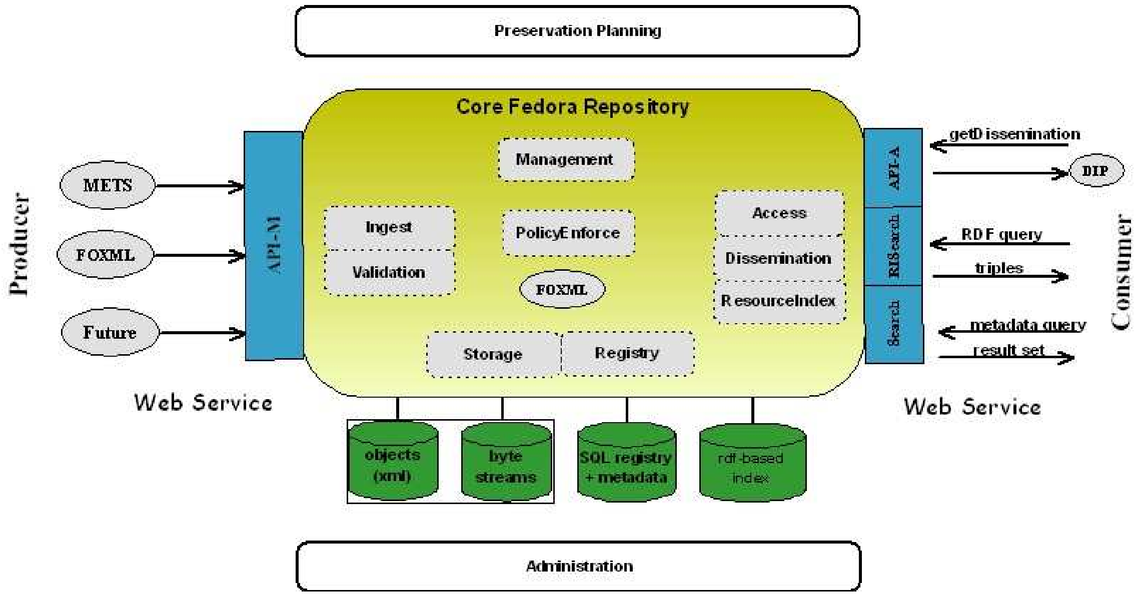

Figure 2: Fedora Core Repository Service (sourced from DuraSpace wiki [10])

DKR is depicted in Figure 2. Its architecture comprises a number of important elements which are described in the following paragraphs. Some of the components map directly onto components of the Open Archival Information System (OAIS) reference architecture [3].

A web service interface for *producers* of digital objects, called API-M, allows the management of objects in the repository. API-M supports the Metadata Encoding and Transmission Standard [8], as well as Fedora's own Fedora Object eXtensible Markup Language (FOXML) format.

Ingest and *validation* services handle the submission of objects by producers (in a supported format), validating the contents, and preparing them for storage and management by the repository.

Access and *dissemination* provide a collection of web service interfaces (API-A, RISearch, and (Basic)Search) for *consumers*. API-A supports access to objects in the repository in the package format of OAIS. RISearch allows searching over the resource index, using Resource Description Framework (RDF) queries and returns RDF triples. (Basic)Search allows metadata queries on Dublin Core (DC) metadata and FOXML and returns object results sets.

Storage and *registry* provides functions for storing, managing and retrieving objects in the repository. In Fedora, objects are stored as FOXML. Both the objects and their associated data (including multimedia artefacts) are stored. In addition, a registry of the objects and core metadata is stored in a database and an index is kept of the RDF triples representing the objects and their relationships. The data model for IK objects is described in more detail in section 7.

The *resource index* is an index of the RDF triples representing the objects in the repository, the relationships between datastreams within objects and the relationships between objects.

Policy enforce provides services for enforcing security policies in the repository.

As the primary responsibility of the knowledge repository is to store and manage the IK content in a way that ensures the durability and preservation of the data, Fedora provides a pluggable low-level storage subsystem that interacts with a persistent store [14].

It is also responsible for providing secure interfaces that allow the curatorship and management of IK stored in the repository. To accomplish this, the repository comprises two security and access control components for user authentication and authorisation. The older of these is based on the eXtensible Access Control Markup Language (XACML) [19] and uses a policy enforcement module. XACML is a standard XML-based language used to encode access control policies. The Fedora Security Layer (FeSL) is being developed as a replacement for the XACML module. It improves and extends the XACML-based module.

The repository also provides an implementation of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), for replication and mirroring of metadata across distributed repositories.

6.4 Security Service

The NIKMAS Security Service handles all security-related interactions. It provides a callable service which other components can use to authenticate users and to authorise them for specific actions. Authorisation of user actions is facilitated using an encrypted, token-based authorisation system.

The Security Service is tightly integrated with FeSL, and provides an API that allows other modules to create, manager, authenticate and authorise users. In addition, it acts as a security broker between the NIKMAS Portal components and the back-end subsystems, including the Fedora Digital Repository and the Community Register. As such, it is responsible for ensuring that only authorised security tokens are issued, and thus that no unauthorised access to the back-end subsystems is permitted.

The Security Service also exposes facilities for administrators to manage/add/edit/suspend/delete user accounts. Currently, a minimal implementation of the security service exists, but it will be further developed in ongoing development efforts.

6.5 Catalogue

The catalogue subsystem maintains a database of IK holders, together with data such as their contact details, location, the traditional authority under which they fall, and a description of their specialised IK. It provides other modules with access to this catalogued IK data, and exposes this data to authorised members of the public via the NIKMAS Portal.

It also provides software support for the NRS cataloguing process, through a standalone application allowing users in the field to catalogue and record information about IK holders and their knowledge claims. This application is based on the InTouch system developed to support the Infopreneur model of van Rensburg, Veldsman and Jenkins [25].

The catalogue supports the distributed deployment architecture of NIKMAS through a hierarchical arrangement of nodes and custom-developed data replication services. Cataloguers operating in communities can capture their information locally, and the system efficiently handles the replication of that data via a PoP to the central catalogue database. Cataloguers do not need to be online during cataloguing - the system opportunistically replicates whenever connectivity becomes available.

The NIKMAS catalogue is fully implemented. Presently, based on user feedback, it is being revamped to better serve the needs of the user community.

6.6 Community Register

The Community Register is a subsystem that complements the catalogue system and tracks registered communities within NRS. It incorporates a component that stores electronic copies of any legal or contractual agreements that have been signed with community representatives.

The Community Register provides other modules with access to the community and agreement data. It also exposes some of its data to authorised users via the NIKMAS Portal, by providing community contact and agreement information linked to IK data viewed on the portal.

This component is partially implemented and is currrently being expanded. It could be further expanded in future to incorporate social networking and media features that will allow communities to interact and discuss their IK online.

6.7 API

This component comprises a limited web service API (using REST and SOAP) that exposes some of the NIKMAS features to external systems, for purposes of integration.

This component is also used to retrieve non-Fedora data from the central NIKMAS node by distributed PoP nodes for the purposes of data replication and synchronisation (see section 6.10 for detail).

The NIKMAS API is not yet implemented.

6.8 Mobile Communications

The mobile component provides integration into the Mobi4D platform [5] to offer users and IK holders access to their data stored in NIKMAS, via their mobile phones using a combination of Short Message Service (SMS) and Unstructured Supplementary Services Data (USSD) [4]. The current implementation is a proof of concept - future iterations of NIKMAS will include more sophisticated applications of mobile technology.

6.9 Integration Services

The integration services of NIKMAS provide a framework and functions for integration from NIKMAS into external IK resources, websites and databases. It is not discussed further here.

The NIKMAS integration services are not yet implemented.

6.10 Distributed PoP Node

As illustrated in section 8, NIKMAS has a distributed deployment architecture. There is a central NIKMAS node, with a number of distributed PoP nodes. PoP nodes contain only a subset of the full set of components and services available at the central node.

Synchronisation and replication between the central and distributed nodes is required. Metadata is synchronised in full using OAI-PMH. Synchronisation of the actual IK content and multimedia artefacts require a custom synchronisation tool.

In order to respect the confidentiality of information, IK content and media (as opposed to basic metadata) is not synchronised down to other PoP nodes, unless the content is marked as being viewable by the general public. Any IK artefacts or data that are marked as private are not synchronised/replicated downward, and will reside only on the central NIKMAS server and at the PoP where it was captured. In such cases, only the basic metadata associated with the artefact will be synchronised to other PoP nodes.

Metadata search will be active on all nodes, as will metadata browsing and retrieval and browsing of all local and publicly viewable content.

Some security-related information such as usernames, role and group assignments and permissions will have to be replicated to all PoP nodes as well.

Interaction between the central NIKMAS node and the PoP node occurs along three streams, as depicted in Figure 3. The synchronisation can be initiated by either node using the Synchronisation API. The actual synchronisation and data exchange happen in 3 distinct protocols, namely: OAI-PMH for core metadata synchronisation, with a filter to appropriately exclude items that will not be synchronised; a custom content synchronisation protocol for partial synchronisation of IK content; a custom security information synchronisation protocol for partial synchronisation of security data.

The NIKMAS PoP synchronisation is currently being developed.

6.11 Intelligent Search

The NIKMAS architecture makes provision for a sophisticated search and information retrieval facility based on description logics and using techniques from knowledge representation and data integration. It provides a way of performing inference and information retrieval over the IK captured in the NIKMAS system using a formal ontology, and using the rich semantics of ontologies to obtain search results that would not have been possible with only keyword-based searching.

The Intelligent Search component can be accessed via a web interface on the portal. It uses its own ontology and interacts with the Fedora Repository using the RISearch API.

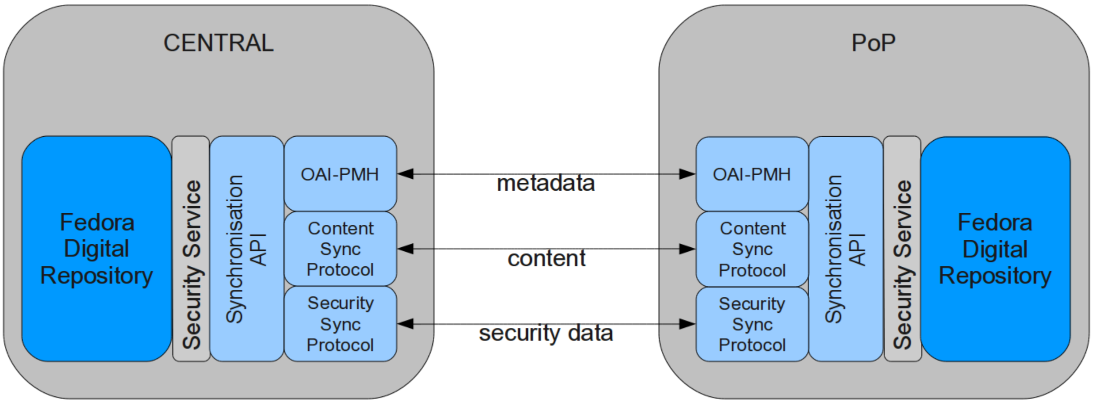

Figure 3: PoP synchronisation

The intelligent search prototype is currently being developed, and will be the topic of a future paper.

7. DATA VIEW

The Fedora digital object model [9] describes the way in which objects in the digital repository are logically stored.

Each object in the Fedora repository has a Persistent Identifier, which is a string that uniquely identifies that object. It also contains a set of system-defined properties for tracking and managing the object in the repository. Finally, it contains a set of datastreams - each datastream represents a content item, such as a document or media file. Objects also contain a number of system-managed datastreams that hold the DC metadata, audit information, security policy information and information about the object's relationship to other objects [17].

Each datastream consists of a collection of fields, namely: datastream identifier to uniquely identify the datastream within the digital object; descriptive label for the datastream; type of the datastream; datastream state (active, inactive, or deleted); date and time the datastream was created; date and time the datastream was modified; indicator whether the repository should maintain version information about the datastream; optional format identifier, using emerging schemes like PRONOM or the Global Digital Format Registry; list of alternate identifiers for the datastream (possibly those used by other systems - examples could be Handles or Digital Object Identifiers); integrity checksum to ensure that the data is not corrupt; actual content (e.g. document, digital image, video, metadata record); and control group indicator, describing the approach used by the datastream to represent its content [20].

Each piece of IK managed by NIKMAS is categorised into an IK theme. Each theme has its own schema of thematic IK metadata that is used to describe the knowledge elements in that theme. In the NIKMAS data model, each IK element is decomposed into a set of interrelated components and each of these components is stored in the repository as a separate object. These objects are described using standard DC metadata, enriched DC metadata using extended DC and theme-specific descriptive data describing the IK content of the object.

Consider, as an example, the data model shared by the African Traditional Medicine (ATM) and Indigenous Food (IF) themes, depicted in Figure 4. In this schema, the main object is of type *IK Item* (with subtypes *IK Traditional Medicine* and *IK Indigenous Food*), which describes the actual IK item in question (remedy or food), including thematic information such as its method of preparation, storage, administration or consumption. Each of these *IK Item* objects can have zero or more media objects attached to it, for example a photo of the dish, a video of the method of preparation, or an audio recording of the IK story.

Each *IK Item* object can also be linked to a number of *IK Ingredient* objects, which describe the ingredients used in the preparation of the item. *IK Ingredient* objects also carry specific thematic information about the ingredient, such as how it is collected, what part of the plant or animal is used, etc. As with *IK Item* objects, *IK Ingredients* can be linked to multiple media objects that can be used to describe the ingredient.

Each *IK Ingredient* object is linked to a single object of type *IK Plant*, *IK Animal* or *IK Inorganic Material*. These objects carry specific thematic information about the plant, animal or mineral from which the ingredient is collected, such as its habitat or cultivation information. Again, these objects can be linked to multiple media objects.

Each of these objects, regardless of type, follow the generic Fedora object schema and have their own standard DC metadata. They also carry additional, extended DC metadata in a special datastream called DC_EXT, as well as thematic data in a special datastream called THEMATIC_METADATA in the figure. In practice, a naming convention is used to name the metadata datastream, based on the type of the object. For example, the datastream containing thematic data for an *IK Ingredient* is IK_ING.

A fuller treatment of this IK object data model will be presented in a future paper.

Besides the IK content stored using the Fedora Digital Object Model, NIKMAS has other data that needs to be persisted, including user and security data, community register data, and catalogue data. However, the data models for these are simple, so they are not further discussed here.

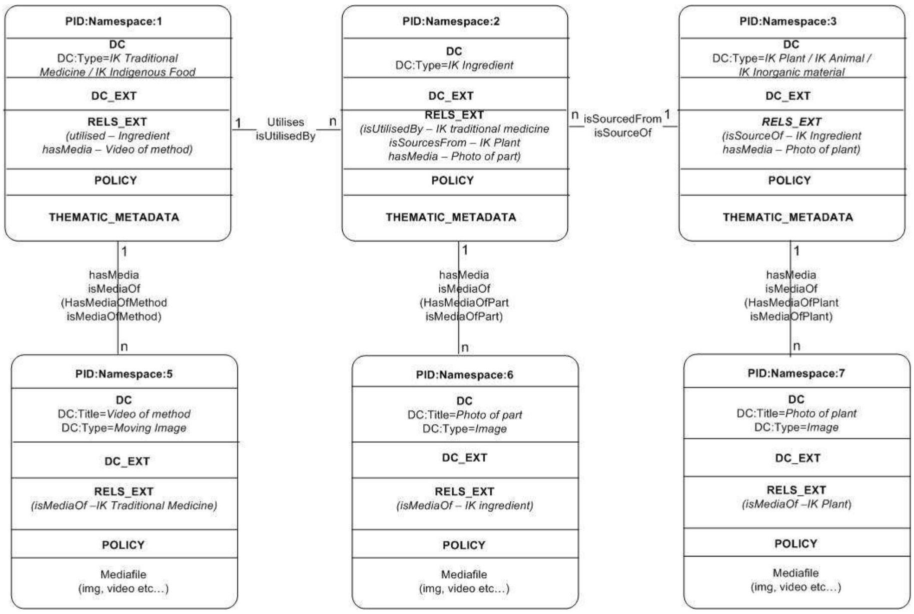

Figure 4: NIKMAS data model for ATM and IF

8. DEPLOYMENT VIEW

NIKMAS is designed to be deployed as a distributed system, with a central server hosting the entire set of services and components, and with distributed PoPs that host a trimmed down version of the system. The PoPs act as local access points for communities to upload and access their IK. IK data is replicated between the central and distributed nodes, but in a way that prevents secret and confidential data from being made available to non-authorised users.

The following network requirements must be met: high-speed, IP-based connection between PoP and central nodes for synchronisation of metadata and content; preferably persistent connection between PoP and central nodes, otherwise the synchronisation service must be configured to run whenever the link is established; and local area, IP-based network at the PoP to allow the upload, searching, viewing, curatorship and editing of IK content.

The hardware configuration for the central node should comprise a high-end server with sufficient RAM and storage (preferably network attached) to accommodate a large repository of IK items, including multimedia files. Redundancy, backup and disaster recovery should be in place to ensure continuity and data preservation and integrity. Additionally, given the potentially sensitive nature of IK items, adequate physical and network security is required.

The PoP node requires a mid-range server with sufficient RAM and permanent storage to accommodate the IK content of that PoP, and optionally the replicated public IK from the central server. Backup and disaster recovery must be in place to ensure that private IK items are safe, and adequate physical and network security is required to prevent unauthorised access.

9. CONCLUSION

This paper presented a high level architectural overview of the NIKMAS software, using a number of architectural views, based on the "4+1" model, to depict different aspects of the system. The architecture of NIKMAS was presented using four views: use case view, logical view, data view and deployment view. Particular attention was given to the logical view, which encapsulates the high level design of the software, and is most likely to be of general interest to other initiatives.

The paper conveyed the significant architectural decisions which were made during development of the first version of the system. As such, the architecture presented here could serve as a starting point or reference for similar projects.

Future work will involve further development of the system, and research and development into the semantic aspects of the system and on improved intelligent information access using the ontology. Future papers will describe the intelligent search technology and the detailed data model.

10. ACKNOWLEDGEMENTS

The authors would like to acknowledge the hard work and contributions of the entire NRS team.

11. REFERENCES

[1] W. Arms. *Digital libraries*. The MIT Press, 2001.

[2] F. Berkes, C. Folke, and M. Gadgil. *Traditional ecological knowledge, biodiversity, resilience and sustainability*. Beijer International Institute of Ecological Economics, 1995.

[3] B. Book. Reference model for an open archival information system (oais), 2002.

[4] A. Botha, I. Makitla, F. Ford, T. Fogwill, D. Seetharam, C. Abouchabki, J. Tolmay, and O. Oguneye. Mobile phone in africa: providing services to the masses. In *CSIR 3rd biennial conference 2010, Science Real and Relevant, CSIR International Convention Center, Pretoria, South Africa*. CSIR, 2010.

[5] A. Botha, I. Makitla, J. Tolmay, M. Ford, D. Seetharam, L. Butgereit, O. Ogunleye, and

C. Abouchabki. Mobile4d platform. In P. Cunningham and M. Cunningham, editors, *IST-Africa 2010 Conference Proceedings, IIMC International Information Management Corporation, 2010*, 2010.

[6] G. Bowker. *Memory practices in the sciences*. MIT Press Cambridge, MA, 2005.

[7] G. Bowker. The archive. *Communication and Critical/Cultural Studies*, 7(2):212–214, 2010.

[8] L. Cantara. Mets: The metadata encoding and transmission standard. *Cataloging & classification quarterly*, 40(3):237–253, 2005.

[9] D. Davies. Fedora digital object model. `https://wiki.duraspace.org/display/FCR30/Fedora+Digital+Object+Model`, Dec. 2009. [Online; accessed 28-May-2011].

[10] D. Davies. Service framework. `https://wiki.duraspace.org/display/FCR30/Service+Framework`, Aug. 2010. [Online; accessed 28-May-2011].

[11] L. Grenier. *Working with indigenous knowledge: A guide for researchers*. IDRC (International Development Research Centre), 1998.

[12] V. Gupta. Traditional knowledge digital library. In *Sub-Regional experts Meeting in Asia on Intangible cultural heritage, Bangkok, Thailand*, pages 13–16, 2005.

[13] M. Hirwade. Protecting traditional knowledge digitally: a case study of tkdl, 2010. In M. Hirwade and A. Chikate, editors, *National Workshop on Digitization Initiatives & Applications in Indian Context, DNC, Nagpur*, volume 3. Principal, DNC, Nagpur, 2010.

[14] S. Hitchcock, D. Tarrant, A. Brown, B. OâĂŹSteen, N. Jefferies, and L. Carr. Towards smart storage for repository preservation services. *International Journal of Digital Curation*, 5(1), 2010.

[15] P. Kruchten. Architectural blueprints - the "4+1" view model of software architecture. *IEEE software*, 12(6):42–50, 1995.

[16] S. Kruk and B. McDaniel. *Semantic digital libraries*. Springer Verlag, 2009.

[17] C. Lagoze, S. Payette, E. Shin, and C. Wilper. Fedora: an architecture for complex objects and their relationships. *International Journal on Digital Libraries*, 6(2):124–138, 2006.

[18] D. Levy and C. Marshall. Going digital: a look at assumptions underlying digital libraries. *Communications of the ACM*, 38(4):77–84, 1995.

[19] T. Moses et al. Oasis extensible access control markup language (xacml) version 2.0. *OASIS Standard*, 1, 2005.

[20] S. Payette and C. Lagoze. Flexible and extensible digital object and repository architecture (fedora). *Research and Advanced Technology for Digital Libraries*, pages 517–517, 2009.

[21] C. Quiroz. Biodiversity, indigenous knowledge, gender and intellectual property rights. *Indigenous knowledge and development monitor*, 2(3):12–15, 1994.

[22] M. Ruiz, I. Lapena, and S. Clark. Protection of traditional knowledge in peru: A comparative perspective, the. *Wash. U. Global Stud. L. Rev.*, 3:755, 2004.

[23] B. Sen. Indigenous knowledge for development: Bringing research and practice together. *The International Information & Library Review*, 37(4):375–382, 2005.

[24] T. Tjiek et al. Desa informasi: The role of digital libraries in the preservation and dissemination of indigenous knowledge. *The International Information & Library Review*, 38(3):123–131, 2006.

[25] J. van Rensburg, A. Veldsman, and M. Jenkins. From technologists to social enterprise developers: Our journey as ict for development practitioners in southern africa. *Information Technology for Development*, 14(1):76–89, 2008.