Focus Area Grant Application

Problem Identification

State main research question or problem statement

Information Management Systems (a.k.a. digital libraries, electronic libraries, networked information systems, etc.) are new to most developing countries, yet we in developing countries have the most to benefit from an unfettered sharing of electronic resources. In recent years, there has been a growing interest in South Africa (and other African countries) in the use of these systems for the setting up of institutional repositories, electronic thesis and dissertation collections, electronic journal management systems, citation management systems for funding agencies, etc. In order to support this emerging trend, there is a desperate need for appropriate technology – technology that can adapt to changing and different needs, resources and scales of operation that can be found across the various institutions and archives in Africa.

This project aims to investigate techniques, models and tools for the management of large quantities of data using scalable clusters of off-the-shelf computers. There has already been much work done to demonstrate that digital libraries can be built out of medium-sized components (see previous NRF-funded Flexible Digital Libraries project, and US-NSF-funded OCKHAM project) – this work seeks to look into how these medium-sized components can be distributed across computers in a dynamic local network so that the size of the data can scale with the number of machines, thus meeting the changing needs of a society where digital library technology is slowly gaining acceptance.

Rationale and Motivation

Provide background to the proposal together with a review of the literature and other relevant resources. State how the research is relevant to the selected NRF Focus Area (in the case of Focus Area Programmes and Thuthuka) or Research Niche Area (in the case of HBU and Technikon programmes)

Digital library systems are software systems that manage electronic resources and make them accessible to users. In recent years, much effort has gone into the software engineering of open source systems such as DSpace (Smith, et al., 2003), EPrints (OpCit, 2003) and Greenstone (Witten, et al., 2000), all of which are aimed primarily at ease-of-use and ease-of-installation while providing users with a wide range of services that operate on the data. Most of these systems have also begun to look into component architectures to support flexibility of service delivery, in keeping with the recurring theme that systems have to be developed in a component-wise fashion (Gladney, et al., 1994; DELOS, 2001).

Coupled with flexibility of service delivery, emerging systems need to support flexible collections of data and metadata. The Open Archives Initiative (OAI) had a major influence on the availability of data and metadata when its Protocol for Metadata Harvesting (PMH) (Lagoze, et al., 2002) was adopted by most digital archiving projects as a primary means for sharing metadata. One such digital archiving project is the Networked Digital Library of Theses and Dissertations (NDLTD) (Fox, 2005), which has set up a Union Catalog of all metadata from participating universities around the world (Suleman and Fox, 2003). This collection of data is then made accessible to various service providers who give users free access to portal-based search/browse services. This collection is a prime motivation for this proposal as the collection size has more than doubled each year for the last 4 years (7000 in 2001, 15000 in 2002, 40000 in 2003, 100000 in 2004, 185000 already in 2005). Current technology for management of such data as well as provision of services will not scale arbitrarily as the collection increases further in future. Given that NDLTD's Union Catalog contains only metadata from approximately 50 universities, and most sub-collections contain only recent data, there is much scope for expansion in the near future.

This problem is not unique to NDLTD. The Networked Computer Science Technical Reference Library (NCSTRL) (NCSTRL, 2003), replaced their legacy system with one that was designed from scratch to use components that communicated using the OAI-PMH (Anan, et al., 2002) – expansion in their system will encounter the same problems as NDLTD.

On the local front, there has been a sudden surge of interest in institutional repositories. In 2004, an Open Access Conference was hosted by SASLI and funded by OSI, to discuss this very important issue affecting the library, museum and academic communities in South Africa. Most institutions have accepted the advantages and are ready to embark on such open access projects – to this end, a workshop to train staff in setting up repositories is being organized in May 2005, hosted at CSIR in Pretoria. The applicant for this proposal has been an invited speaker at the former conference and is one of the organizers of the latter training workshop. Such activities will naturally lead to more open access electronic repositories being set up locally and the need for high quality services will emerge like it did elsewhere in the world, soon to be followed by services that can manage rapidly or irregularly increasing data collections.

The applicant also participated in discussions related to the SARIS project and the SARKIN proposal in 2004, aimed at improving research infrastructure in South Africa. Part of the proposal included the need for open access research repositories and central citation indices – if this is ever operationalised at South African institutions, there will be a need for high quality cross-institutional services based on a gradually increasing collection of data.

All of these current and future projects have a need for scalable information management systems that do not give up the flexibility of components or the interoperability afforded by standards such as the OAI-PMH. It is proposed that clusters of off-the-shelf machines can be used as the hardware platform for these generic digital library systems, but the software to scale as hardware is thrown at the problem still needs to be investigated, and is the subject of this proposal.

Past efforts in component models have demonstrated that they are indeed feasible for rapid and flexible deployment of information management systems. The Open Digital Library project (ODL) (Suleman and Fox, 2001; Suleman and Fox, 2002) has generalised the well-understood syntax and semantics of the OAI-PMH to support inter-component communication. This generalisation was then used as the basis for designing a suite of simple protocols to support search engines, category-based browsing, recommendation systems, annotation engines and other typical services expected by users of a digital library. The Flexible Digital Library project, currently underway, is looking into advanced interfaces for automatic management of components (Eyambe and Suleman, 2004), and lays the foundation for using a cluster environment for such components.

Building on these past and current projects, as well as known results from the distributed computing and data warehousing communities, cluster computing can be used to support projects where there is a need to start small and rapidly scale to large collections of data with changing service suites. This is very much in line with local attitudes to digitization, where millions of rands of resources cannot be committed to projects until some operational and practical proof-of-usefulness has been established.

Research into the scalable use of components to build digital libraries is of practical importance to local digital library efforts while simultaneously satisfying multiple themes of the research agenda defined by NRF in its ICT programme. Building distributed systems from components is directly related to the "software customisation and integration" theme and since components are distributed over the Internet, this work is also related to the "Internet and mobile application" theme. In terms of telecommunications and networking, componentised cluster-based digital libraries contribute to most of the identified research themes, if not all. Protocol design is a key part of developing components distributed over the Internet and such protocols must be evaluated further for their ability to scale arbitrarily to support both user-component and component-component interaction. Local conditions of poor network connectivity impact on such experimental protocol and system design to encourage minimalist approaches and robustness of algorithms. Lastly, this work will contribute to the areas of distributed systems and services since the component-based digital libraries under examination are network-distributed services. Investigations into aspects such as scalability and robustness of components generalise to other distributed systems, while services within the context of digital libraries are driven by user needs in a networked environment so include the full gamut of possibilities.

Finally, digital libraries as networked information systems fall within the research theme of "human-information interaction", especially the "use, storage, retrieval and sharing of information". The services provided by a digital library focus largely around the retrieval aspect while interoperability among systems is a direct realisation of the “sharing of information” research area. This sharing occurs at multiple levels within a component-based digital library: at the extremities where information is shared with external systems, and within the system where information is exchanged among independent components within nodes of the cluster.

From a developmental angle, digital libraries fundamentally change the landscape of access to information and therefore affect the quality of life for all. This work, by promoting simple models for scalable digital libraries, will make them more accessible to archivists working in varying resource conditions. This, in return, makes information more readily available to the ordinary student, teacher or researcher (within the constraints of basic Internet access), levelling the playing fields by removing access to information as a barrier to learning.

References

Anan, Hesham, Xiaoming Liu, Kurt Maly, Michael L. Nelson, Mohammad Zubair, James C French, Edward A. Fox and P. Shivakumar (2002), “Preservation and transition of NCSTRL using an OAI-based architecture”, in Proceedings of the Second ACM-IEEE Joint Conference on Digital Libraries, Portland, OR, USA, pp. 181-182.

DELOS (2001) Digital Libraries: Future Directions for a European Research Programme, San Cassiano, Alta Badia, Italy, 13-15 June 2001. Available http://delos-noe.iei.pi.cnr.it/activities/researchforum/Brainstorming/brainstorming-report.pdf

Eyambe, L. and H. Suleman (2004). “A Digital Library Component Assembly Environment”, in G. Marsden, P. Kotzé and A. Adesina-Ojo (eds): Proceedings of SAICSIT 2004, Stellenbosch, 4-6 October 2004, pp. 15-22. ISBN: 1-58113-982-9

Fox, E. (2005) Networked Digital Library of Theses and Dissertations. Website http://www.ndltd.org

Gladney, H., Z. Ahmed, R. Ashany, N. J. Belkin, E. A. Fox and M. Zemankova (1994), “Digital Library: Gross Structure and Requirements”, Workshop on On-line Access to Digital Libraries, June 1994.

Lagoze, Carl, Herbert Van de Sompel, Michael Nelson and Simeon Warner (2002), The Open Archives Initiative Protocol for Metadata Harvesting – Version 2.0, Open Archives Initiative, June 2002. Available http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

NCSTRL (2003) Networked Computer Science Technical Reference Library. Website http://www.ncstrl.org

OAI (2002), Open Archives Initiative. Website http://www.openarchives.org

OCKHAM (2003), Open Communities for Digital Library Development. Website http://ockham.library.emory.edu/index.php

OpCit (2003), E-Prints. Website http://www.eprints.org/

Smith, MacKenzie, Mary Barton, Margret Branschofsky, Greg McClellan, Julie Harford Walker, Mick Bass, Dave Stuve and Robert Tansley (2003), “DSpace: An Open Source Dynamic Digital Repository” in D-Lib Magazine, Vol. 9, No. 1, January 2003. Available http://www.dlib.org/dlib/january03/smith/01smith.html

Suleman, Hussein, and Edward A. Fox (2001), “A Framework for Building Open Digital Libraries”, in D-Lib Magazine, Vol. 7, No. 12, December 2001. Available http://www.dlib.org/dlib/december01/suleman/12suleman.html

Suleman, H., and E. A. Fox (2002), “Designing Protocols in Support of Digital Library Componentization”, in Proceedings of 6th European Conference on Research

and Advanced Technology for Digital Libraries (ECDL2002), Rome, Italy, 16-18 September 2002, pp. 568-582.
Suleman, H. and E. A. Fox (2003), Leveraging OAI Harvesting to Build a Union Catalog, Library Hi-Tech, Vol. 21, No. 2, pp. 219-227, edited by Timothy W. Cole, Emerald Publishing.
Witten, I. H., R. J. McNab, S. J. Boddie and D. Bainbridge (2000), "Greenstone: A Comprehensive Open-Source Digital Library Software System", in Proceedings of Fifth ACM Conference of Digital Libraries, San Antonio, Texas, USA, 2-7 June 2000, pp. 113-121.

Research Aims

Provide details on the aims of the research project. These should focus on research-related aims and not capacity or human resource development aims.

This proposal outlines a course of research into the issues faced by digital library designers who wish to adopt component models for their systems while including support for arbitrary scalability. Specifically, the following questions will be investigated:
a) What network protocols and application interfaces are needed to embody the principle of simplicity while at the same time supporting the range of lower-level services required by components that operate in a cluster computing environment?
b) How are fault tolerance and load balancing included to address scalability and robustness of such systems?
c) How does the existing Web infrastructure support component-based systems, if we introduce arbitrary replication and migration for load-balancing? What are the needs of individual components and the base architecture (especially Web server technology) to improve on this situation?
In summary, this research is aimed at addressing the problems introduced by dividing what is conceptually a single fixed system into individual components to maintain flexibility of services but more importantly support arbitrary scaling for growing data collections and use patterns. The expected outcome of this work is a set of proven guidelines and experimental tools to move the digital library community closer to an ideal of simple, flexible, scalable and robust digital library architectures.

Workplan – Research Activities

This section includes the workplan which should focus on research activities and their associated milestones. Description should also include information on timeframes and responsibilities, how students will be involved, availability of specialized equipment, infrastructure and resources and other relevant information.

Student support is a primary aim of this project. As such, the project will be conducted as a series of related sub-projects, with collaboration on the interfaces and external dependencies.

These can be enumerated as follows:

Design of experimental interfaces and reference components based on prior work and current trends for mobile component migration, replication and management. This work is currently being addressed by one MSc student, under supervision of the principle investigator. It is envisaged that a second student will work on additional aspects of the project related specifically to individual information retrieval services. Time-frame: 2005 to 2007

Design and implementation of a universal Web server infrastructure for Web-based component farms is a separate sub-project that will be addressed by a single MSc student, under supervision of the principle investigator. Time-frame: 2006 to 2007.

It is intended that student assistants will help with the building of prototypes, as part of their research training in (pre-MSc) Honours degrees.

Some equipment for this work is already available and is being acquired during the course of 2005. Additional machines will be needed to bolster the computing cluster in 2006 and beyond.

Workplan – Research Methods

This section includes three areas that should be addressed as fully as possible viz, **methodology, methods and techniques**, including data collection and analysis, and other relevant information.

The methodology for this work has two identifiable aspects: those elements applicable to all of the sub-questions that will be addressed; and elements that relate to each individual major issue. Thus, these will be treated as separate sections:

1. Overall research design
1.1. Data collection
As an enabling mechanism for the other aspects of this project, data collections need to be established in the form of local archives. These will be obtained by working with researchers to archive their publications in subject repositories recommended by their individual communities e.g., by setting up local nodes for the NCSTRL project, which already has more than 150 participants worldwide, notably excluding South African universities. In addition, institutional repositories for documents such as electronic theses and dissertations are becoming popular (South Africa has at least three members in the international Networked Digital Library of Theses and Dissertations (NDLTD) (http://www.ndltd.org) and it is proposed that these will be buttressed by additional components to make them interoperable with our and other efforts. In all of these cases, the work involves consultation with staff and the use of standard or emerging tools in the digital library community – no new software development will be necessary. The collections built over time will serve a two-fold purpose: the data can be used by experimental systems and the researchers – who would have had exposure to digital libraries – can provide a base group for evaluation

of systems. Essentially, this nurtures a tradition of digital library use and understanding, which is lacking in South Africa and must be developed to support research and development efforts.

As a second front to data gathering, the OAI Protocol for Metadata Harvesting, which is pivotal to the component model, can be used to obtain data from remote sources. The OAI maintains a registry of archives currently sharing their collections of metadata and/or data freely with external entities. This includes notable organisations such as the US Library of Congress and collections such as OCLC's WorldCat – a collection of 4 million pointers to dissertations. Such data, or subsets thereof, can readily be obtained as and when necessary, over the Internet, in order to supplement and complement local collections and to expose local audiences to international resources and vice versa.

1.2. Specialized equipment and infrastructure

As one of the aims of this project, the techniques evolved must be applicable to a wide range of systems, from large institutional, national and international repositories to small collections of important resources. However, recognising that very large projects have the personnel and finances to develop custom solutions, this project is aimed primarily at defining the architecture of digital libraries that operate with tight constraints on resources. Thus, there is no unconventional equipment required. Instead, medium- to low-end PC servers using open source software are becoming the norm in the digital library community and these are all that is needed. Multiple component farm machines are needed to test inter-component interaction over a network. As much of the work in this project is designed to be conducted by postgraduate students working independently, commodity workstations are required for each of those researchers, who begin working on the project in 2006.

Some funding has already been obtained from UCT and NRF for the primary server to host experiments with components and an initial component farm – some of these will require upgrades and additional nodes. Also needed for all servers and workstations is a local network infrastructure and a realistically fast external Internet connection. While the local department provides some external connectivity, there is a need for internal gigabit connections and supporting equipment (e.g., routers/switches) within the component farm. External research collaboration and data collection from outside sources may require a stable and broadband connection.

1.3. Component Framework Design and Systems Integration

Primary responsibility for further developing the component framework to adhere to current Internet and digital library standards will rest with the principal investigator. This includes the dynamic process of re-design, ongoing maintenance and testing of the ODL/FDL framework, in collaboration with local and international collaborators, in response to emerging requests for changes. A component suite will be maintained to support all experiments by students and collaborators associated with the project. Interest in the ODL/FDL framework from external parties has spurred discussion that may lead to the standardisation of inter-component interaction in the long term. The principal investigator will participate in and initiate such activity as and when an appropriate body of research is available to substantiate a need for standardisation.

In addition, while the research question has been divided into sub-questions aimed to define and encourage student participation in the larger project, integration among all of these parts is essential to prove the viability of the overall approach. This integration will be led by the principal investigator, to oversee the various parts and maintain a consistent and workable model as time progresses.

Dissemination is usually considered to be a separate activity post-experimentation, but in the context of networked digital libraries, it is a primary need since collaborators frequently provide support for experiments with remote systems. The principal investigator will encourage dissemination and actively seek collaborators to help with experimental validation of the approach taken by this project.

2. Specific issues

2.1. Migration, Replication, Scalability of Component-based Systems

There are many specific concerns related to the management of remote components, including registration mechanisms for components to support remote configuration, load balancing and possibly a form of process migration. All of these support the move to a location-independent swarm-based system architecture. This is in keeping with a growing interest in the use of grid and peer-to-peer architectures for digital library systems, requirements that place very specific constraints on the design and interfaces of components. Also, they require a particular execution environment that supports dynamic mobile components, one that is rarely available to practitioners in archiving communities. This part of the project will look into bridging the gap between scientific community best practices for cluster computing, agent-based architectures that have evolved from AI and the processing and storage needs of users and managers of digital information. The work will concentrate on addressing issues related to specific services, such as information retrieval, that can be parallelized/replicated/migrated and open architecture issues that relate, such as security and authentication in peer-based systems.

This work will be carried out in conjunction with one Master’s student in 2005-2006, possibly in conjunction with a group of Honours students. A second student will be recruited in 2006, to work on specific aspects of the problem area. Substantial responsibility for the design and evaluation will be delegated to the Master’s students, under supervision of the principal investigator, with Honours students focusing mainly on the implementation of tools.

2.2. Universal Web Application Containers

Many modern Web servers are optimised for static documents and handle one or more types of back-end technology. This works well in single-user single-technology situations (such as many commercial websites), but creates many problems when a single machine hosts many websites with different technical requirements. Theoretically, it should be possible for a Web server to support PHP, Perl, Java and other languages without integrating them into the core system and while retaining the user/group context switching performed by tools such as suexec and CGIWrap. The SpeedyCGI tool demonstrates that this idea is feasible in the context of a single language - Perl. This project will begin by investigating the feasability of a language-independent secure, dynamic and efficient environment for Web-based components and how past work in this area has touched on various aspects of a universal solution. A prototype will be developed, or adapted from prior work, to incorporate support for multiple languages and platforms. Then the packaging technologies of Web Services will be considered in conjunction with the new universal platform. This research will concentrate on how Web Service components need to be structured to support rapid and portable deployment on a language-agnostic platform. Key issues will include deployment mechanisms, security models for controlled access to individual suites of components, labelling of service endpoints, component configuration and local resource allocation.

The anticipated end-result is a system which allows a non-privileged user community to easily install and make accessible software components that are in essence Web Services, without having to deal with a myriad of different technologies and without having to hardwire hooks into the Web server and similar system-level resources.

This project is aimed at a single Masters student and will build on prior work done by Honours students (for an Honours project) and other efforts related to information management components. The results of this project can have a major impact on wider adoption of components for Web Services and the Service-Component paradigm of computing, especially in component farms where scalability is a prime concern. The project will run over the period 2006-2007.

Potential Impact on HR Development

Outline the impact this proposal has on human resource development in terms of:

- Extent and appropriateness of research training (e.g. opportunities for training and experience provided to students)
- Nature and significance of contribution of co-investigators and research associates
- Nature and significance of inter-institutional collaboration
- Appropriateness of Human Resource Capacity Development for a specific science domain/discipline/area of application/employment sector or for the development of scarce skills.

Issues relating to redress and equity should be specifically addressed in the next section.

This proposal is specifically designed to encourage the training of students in the principles and techniques involved in building modern networked digital libraries. This is a critical skill needed in the IT community to develop and enhance future infrastructure for information sharing. While postgraduate students will benefit directly, their experiments will involve scores of researchers and fellow students who will gain a better understanding and appreciation for digital libraries by exposure to the technology. There are currently 5 MSc students working on related projects in digital library systems at UCT, three of whom are funded through NRF, one of whom has external funding and a fifth who is not eligible. One of these students has just finished her studies and submitted her dissertation for examination, while the others are continuing in 2005. As this project continues to investigate emerging problems, 2 additional MSc students will be recruited and trained in 2006/2007. In addition, Honours and 3rd year students will assist where possible, as a way of encouraging them to pursue/continue research-led degrees in future.

This project also aims at promoting, directly and indirectly, collaboration among researchers through the medium of digital libraries. Such collaborations enhance the quality of individual research works, while contributing to a sense of community among researchers.

Potential Impact on Redress and Equity

Outline the impact of this proposal on redress and equity (in terms of race and gender) in relation to the issues listed under “Potential Impact on HR Development”.

Digital libraries, by their very nature, address imbalances in the information arena by making information more accessible to regular users. This is especially relevant in the area of journal publications. These are traditionally expensive to obtain and therefore only easily available to universities that subscribe at high costs to themselves, the state and their students/academics. One of the main aims of digital library research is to make it possible for academics to archive, review and publish work without the need for expensive intermediaries such as publishing houses. Even if such works are already published, simple digital libraries can be used to make the publications available locally, as is allowed by many publishers already (e.g., ACM, Springer). This project aims to make the technology accessible to researchers so that they, in turn, may make the products of their research accessible to the larger research community.

Component-based scalable systems are important to support research into online information systems because they provide the mechanism for highly specialised projects to be incorporated into larger systems. This allows researchers in information systems at smaller universities to easily contribute to digital library (and related) research without the need to first develop their own basic tools and frameworks.

Eventually, digital libraries are the mechanism to get high quality information about Africa out to the rest of the world, and in the reverse direction, for us to obtain high quality electronic resources. Further, this will be possible without having to develop massive software systems if our models, and the models adopted by the rest of the world, are based on simplicity and flexibility.

Potential Outcomes

Indicate the relevance of the proposed research in relation to the following:

- Expected national and/or international acclaim for the research and contribution of research outputs to building the knowledge base
- Exploitability of outputs e.g. applicability to community development, improved products, processes, services in SA, region and/or continent
- Expected effects of research results on user sectors

Appropriateness of the knowledge dissemination strategy in view of the above

This project has many potential outcomes for the research community in digital libraries and the archiving and library community.

The research community in digital libraries traditionally sees service provision as being an atomic operation, but by demonstrating that scalability can be achieved on clusters of component servers, it is hoped that there will be a shift in thinking about services to incorporate notions of well-defined interfaces to support aspects such as migration, replication and remote management in general.

This project has implications for the library and archiving community in terms of cost savings and the ability to gradually adopt new technology. Using information management systems that generalise to clusters, it will be possible for archivists to start every project in proof-of-concept mode and scale upon (expected) success or with the arrival of funding or larger data sets. This eliminates the need for large proposals and injection of funds with unknown returns.

In terms of artefacts, this project will contribute a suite of tools that may be used to design digital libraries or that may be adapted to similar projects in distributed systems and/or Web Services. Software packages produced as part of the work also may be used in research environments to make local resources accessible to local and/or international audiences. Finally, the collections developed during this project will serve as testbeds for future digital library experimentation locally and abroad, where they will encourage future international efforts to acknowledge and cater for the needs of countries such as South Africa.

Ultimately, by advancing the science of building digital libraries, this project hopes to make information more accessible to users. In South African society, where the cost of information is abnormally high, any move towards acceptance of digital libraries makes it easier for people to access information that would normally be beyond their reach (philosophically and physically). A typical example of this is the Networked Digital Library of Theses and Dissertations (NDLTD), which is making it possible for students and researchers to obtain electronic copies of theses and dissertations online. This was previously only possible through inter-library loan. Now, anybody anywhere in the world can find and get access to a thesis if it is part of a member digital library at a participating institution. NDLTD achieves this by using some components developed as part of the ODL development, and is therefore a suitable representative for the breadth of opportunities made possible by the adoption of a simple, flexible digital library model. Scalability will add another dimension to the capabilities and services afforded to the NDLTD community.

Progress to date: Summary

A statement of process progress

For new applications: Provide an informative summary of the relevant work that the applicant(s) has undertaken preceding this application.
For ongoing applications: Provide a summary of progress since commencement of the project

The principle investigator has actively developed and supports a framework for building digital library systems as collections of inter-connected components. These are used in various projects, including NDLTD, and have formed the basis for a number of experiments at UCT and beyond. Experiments to validate the architecture were conducted with various communities and case studies. The results of initial work has been published in conferences, journals and magazines (as listed in the appropriate sections).

During 2005, an MSc student (Muammar Omar) began work on migration and replication of components in a small component farm. This work is continuing and will end in 2006.

Some collaboration has been established with the new Centre for High Performance Computing (CHPC), to be established by UCT/CSIR. Digital libraries were listed as one of the application areas at a CHPC workshop held in late 2004.
From a data collection perspective, a local archive of research publications has been established and is in a prime position to serve as a future testbed for experimental services. Also, working with the ETD Africa initiative, the principle investigator is assisting universities in South Africa and Africa in general to set up electronic thesis and dissertation projects – these will serve not only the needs of the local communities but also jettison the African community into the digital library age and provide case studies and testbeds for current and future research. To this end, two workshops were held recently: one for South African universities in September 2003 and one in Addis Ababa for Ethiopian universities in February 2004.
The NRF-funded Flexible Digital Libraries project runs from 2004-2005. Four MSc are investigating various aspects of building systems from components – a number of initial results have informed the need to move towards clusters with remote management of components.

Progress to date:

Research outputs with relevant data – findings or results as well as publications and papers submitted to journals, prepared for conferences etc. other achievements or outputs from the research.

For new applications: Provide an informative summary of the relevant work that the applicant(s) has undertaken preceding this application.
For ongoing applications: Provide a list of the relevant research outputs associated with this project with relevant dates

The projects denoted "Open Digital Libraries" – which defined basic architecture and core interfaces - and "Flexible Digital Libraries" – which investigated higher level integration of components – together form the core underpinning for building generic information management systems. The former was funded by the US-NSF and Mellon foundation and was the subject of the applicant's PhD. The latter was funded by the NRF in 2004-2005 and both have resulted in various publications and technical reports as listed in the relevant section of the proposal. Some of these publications have a direct bearing on the concepts of scalability and remote management and will therefore be built upon in future work.

Progress to date:

Report on Progress of students involved - completion record/ status quo report for M and D students supported (if in first year of degree, for example "has progressed to second of two year degree; if in final year, submits thesis, exam, degree awarded etc.

For ongoing applications: Provide a report on progress of students involved.

There is 1 MSc student currently working on this project.

Co-Investigator Outputs

Please enter the best peer-reviewed research outputs of Co-investigators (no more than 10 in total)

(not applicable)