National Integrated Cyberinfrastructure System

A framework for the establishment and maintenance of a sustainable NICIS

Report of the International Committee

for the

Development of South Africa's

National Integrated Cyberinfrastructure System

Appointed by the

Department of Science and Technology

December 17, 2013

Executive Summary

Introduction

The impact that cyberinfrastructure has made and can make as a driver in research and development is now well recognised. This point was made in a very strong symbolic way as this report was finalised by the awards of the 2013 Nobel Prizes in Chemistry[1] (which recognised the impact that complex computational models can make in developing Chemistry) "*for the development of multiscale models for complex chemical systems*" and Physics *"for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider"*, in which experimental work and a massive worldwide cyberinfrastructure was required to complement the theoretical insight of the awardees. This "third mode of science" is likely to grow in importance. Cyberinfrastructure is critical to the competitive innovation capacity of industry and society for socio-economic benefit, as well as many research areas across all disciplines. e-Science, as a third-mode of science, relies on a strong cyberinfrastructure.

The framework for the establishment and maintenance of a sustainable National Integrated Cyberinfrastructure System (NICIS) aims to exploit the enormous synergies that can be derived from integrating national Cyberinfrastructure (CI) into a cluster of mutually supporting activities which leverage high-level financial and strategic planning, as well as oversight and management. This integration opportunity is important to competitive national science and technology strategies and implementations, to international scientific and engineering collaborations, and for socio-economic impact.

The South African national cyberinfrastructure is now about six years old. This report proposes significant changes to the organisation and mode of operation of the cyberinfrastructure. This is a reflection that the needs of new organisations in their set-up phase are different to that of organisations approaching maturity. Thus, the need for change is driven as much by the *successes* of the current structures as much as new challenges and improvements that should be made.

Benchmarks

The analysis of integrated CI systems in other countries provided the basis for developing benchmarks and identifying principles that would inform recommendations for an integrated framework for South African CI. The benchmarking exercise focused on the management structures for coordinating and overseeing CI programs, rather than the specific nature of those programs or the level of financial investment. There are general management principles for successful international frameworks that drive the guiding principles laid out by this Committee for the South African NICIS.

The Committee observed that centralised resources and capabilities function effectively, when the NICISs are well-run and responsive to the user community.

[1] Michael Levitt, one of the Chemistry Nobel Prize winners of 2013, was born in Pretoria in 1947.

Recommendation 1: *It is recommended that DST amalgamate the structure of separate organisations for each primary sector function into a single integrated CI organisation (NICIS) that encompasses the multiple sector functions (e.g. networking, compute, data).*

There is a natural tension between centralised (“Tier 1”) and distributed (“Tier 2”) resources and capabilities. While centralised Tier 1 facilities include the inherent advantages of economies of scale, distributed Tier 2 facilities create more diverse CI ecosystems.

Recommendation 2: *Given the realistic level of CI investments by DST, the Committee recommends that DST and its NICIS continue to focus investments in Tier 1 centralised facilities within each sector, with an expectation of collaboration with regional Tier 2 facilities.*

Evidenced-based needs analysis

There is already significant investment in CI in South Africa. The Centre for High Performance Computing (CHPC) is the largest component of the Country’s High-Performance Computing (HPC) infrastructure and plays an important leadership role. The Data Intensive Research Infrastructure for South Africa (DIRISA) is the newest CI initiative, born by the recognition that the need to manage very large data sets, goes beyond the physical storage requirements to issues such as curation, provenance, trust, digital preservation, and analysis techniques. The South African networking landscape has been revolutionised over the last five years by the sustained investments of the DST through South African National Research Network (SANReN) and the Universities through the Tertiary and Education Research Network of South Africa (TENET), coupled with effective use of new international links. However, a number of science projects continue to drive the need for cyberinfrastructure growth and expansion. Typically, these projects combine needs for high-performance computing, fast networking, and large data sets. Although, in the context of a fast research network, regional HPC facilities have only a small advantage over national facilities from an access perspective, there is space and need for shared and/or specialised facilities.

Recommendation 3: *It is important that researchers have access to a range of HPC facilities, including both national Tier 1 resources and institutional Tier 2 resources. The facilities need to be built around education and training needs, provision of specialised services, and efficient use of resources.*

Given the large investments made by DST in the CPHC, the National Research Foundation (NRF) currently does not fund proposals for HPC. The Committee believes that NRF and other related programmes should be open to projects with a CI component.

Recommendation 4: *The NRF National Equipment Programme (NEP), Department of Higher Education and Training (DHET) Infrastructure Programme and any other related national programmes should fund CI equipment (including Tier 2 computing and storage facilities) on an equal footing to other equipment requests*. *The Committee further recommends stronger coordination between the various funding initiatives.*

Proposed organisation of national cyberinfrastructure.

Based on the above, a minimal set of principles for an ideal National Integrated CI System would include:

- Joint planning and budgeting,
- Good governance: transparency, accountability, efficiency and effectiveness,
- Visibility of CI services,
- Sustainability, and
- Constructive stakeholder's engagement.

Other factors that have influenced the recommendations of the Committee include the need for innovation, the need for integrated services, international relations and cost efficiency. Responding to the challenges and opportunities of the digital age we propose a national integrated cyberinfrastructure, designed to become an effective national tool.

The vision is a NICIS that takes national leadership in the provision of a comprehensive cyberinfrastructure essential to 21st century advances for South Africa in research, education and innovation.

The mission of NICIS is to increase knowledge creation through provision of a national platform of essential CI.

The organisational structure of NICIS can be described at the high level of ownership and reporting structures, as well as at the level of the internal structure with detailed functions and responsibilities. The internal structure is largely independent of the high level ownership model.

The most powerful tool for DST to develop the national leadership for CI in the future is to converge NICIS in the long-term into an appropriate legal entity fully owned by the Department of Science and Technology (DST). The role of DST is to execute and further its professional role as owner of NICIS. DST decides the vision and overall aims of NICIS as well as the funding and budgetary aims. This model will optimally fulfill the basic principles enumerated above.

On a short to medium time scale, a viable alternative model is to organise NICIS as a "National Facility" under NRF. The astronomy sub-agency model may be appropriate. As NRF's mandate covers research infrastructure, CI could be naturally accommodated. The user base of NICIS overlaps with the user base of NRF. This scenario synergistically matches research priorities to CI priorities and harmonises the use of research and CI investments. Other models, like the current one, might be viable, but sub-optimal.

Recommendation 5: *The preferred organisation is a NICIS set up as an appropriate legal entity fully owned by DST. An alternative viable option is to consolidate NICIS into a "National Facility" or a sub-agency of NRF.*

Recommendation 6: *The Committee proposes an organisation that is horizontally integrated across CI elements.*

A NICIS Board will maintain an independent control function over the organisation on behalf of its owner, as well as develop a strategic plan. The Board appoints a CEO, who is responsible for the implementation of the strategy. Whatever the placement of NICIS, it is imperative that the Board has real, decision-making powers and the ability to hold the CEO to account. The activity of NICIS will be organised in Service Areas, including Networking, Computing, Data, Skills & Training, and Administration. Each Services Area will engage the user community through an appropriate *User Forum* and will be coordinated by a Director.

The integrated NICIS framework allows for a systematic and efficient approach to cross-cutting themes, including skills and human capital development. NICIS should run innovation projects that focus on developing new services, new capabilities and engaging with the commercial sector. Participation in international CI initiatives is important in ensuring high quality of the NICIS enterprise.

Funding models: Based upon international experience, NICIS will not be sustainable without continued significant support from DST. Two strong recommendations need to be made:

Recommendation 7: *Budgeting must be integrated across the NICIS. While ring-fenced funding for the system as a whole is a good idea, there needs to be joint planning between the different pillars of the NICIS. This planning should be aligned to the Medium-Term Expenditure Framework (MTEF).*

Recommendation 8: *There must be a high-level of accountability of the sectors to the stakeholders (DST, universities, etc.).*

In this respect, the different sectors have a responsibility to show that the money that is invested in them is well spent. This is important to justify the long-term funding that will be required for the sector, and for planning. Clear and transparent metrics to measure performance must be set. Responsiveness to user and other stakeholder concerns must be fundamentally transformed.

Other sources of funding are available and should not be neglected. As there is significant benefits of scale, DST should explore shared use of central NICIS facilities by other state entities. Commercial use of CI facilities should be encouraged but is expected to be limited as a source of financial sustainability.

Research funding. While there is need for innovation projects within the NICIS, the Committee does not see an extensive role for in-house research within the service areas of NICIS, as it would induce potential conflicts of interest with the research communities it is intended to serve. However, the Committee strongly supports CI-related research within the various research communities outside the scope of NICIS. Research funding processes need attention and NICIS as an entity should not serve as a funding agency.

Recommendation 9: *Much tighter integration between the allocation of resources from NICIS to researchers within existing research funding instruments is needed. NICIS should not have an extensive in-house research role. However, CI-related research should be strongly supported within the research communities outside the scope of NICIS.*

Large-scale access by researchers to the cyberinfrastructure should be made on a competitive basis.

Recommendation 10: *A common framework for the allocation of cyber-resources across all NICIS Services Areas (computing, networking, data) should be adopted, using merit-based scientific assessments.*

The Committee is aware that there are cost implications to the proposed recommendations that will have to be properly costed and evaluated.

Recommendation 11: *New funding is necessary for NICIS management, expanded Data Services (DIRISA) and a new Skills and Training Services area.*

Proposed model for each of the CI service areas

Recommendation 12: *The Committee recommends that the proposed Networking Services Area should be recognised as the National Research and Education Network* **(***NREN). It is important for DST to have significant involvement in the NREN both because it will be a primary funder and so will need to account for the spending of public funds and because it is in the position of having strategic oversight and can take into account all national interests. Other key parties such as the universities will also have a stake in the process and be able to hold the service area to account.*

TENET will continue to play a key strategic and operational role in the networking sphere. Although accountable to its shareholders, TENET's role in providing and defining network services will be important in the planning and operational processes.

Recommendation 13: *The Committee recommends that the CHPC in essentially its current form should take on the role of the Computing Services area, with some changes to its mandate. More formal processes to assess user satisfaction and needs must be established with requisite metrics for performance.*

Recommendation 14: *The Committee recommends that SA Grid be continued with modest additional support to increase robustness, and recommends that DST consult with stakeholders to determine whether to assign it to Networking Services or Compute Services.*

Recommendation 15: *The Committee strongly recommends that cloud computing be a key function of the Computing Services area.*

Recommendation 16: *Universities should be asked to respond to this report (either through the Higher Education South Africa (HESA) or the Deputy-Vice Chancellors (DVC) Research Forum) and draw up individual and collective strategies for provision of Tier 2 (and 3) facilities. Collaboration and coordination between DHET and DST are essential in prioritising mid-range facilities, connection to national leadership-class compute capabilities, human resource use, and benefit of procurement.*

Recommendation 17: *The Committee recommends that the NICIS Data Services Area be the leading organisation within South Africa to advocate for and implement data initiatives across the research community.*

The current DIRISA program should be transformed to a more vigorous expanded Data Services Area. By necessity, there will be both near-term and longer-term goals for this transformation. Long-term a substantial new investment would be required to achieve the ambitious goals in the expanded remit recommended by this Committee.

Recommendation 18: *The Committee recommends that NICIS works with the community to develop an ambitious proposal on data services to DST, and formulates convincing arguments to the Treasury for the return on that investment, predicated on economic competitiveness, human resource development, and industrial benefit.*

Recommendation 19: *NICIS should offer effective coordination of CI Skills and Training services within a sustainable framework. NICIS should maintain a coordinated plan to develop and grow the community of researchers benefiting from CI.*

The main targets of this new Services Area are

- CI professionals (support personnel), for developing operational knowledge,
- Researchers that use, or could benefit from using, CI services to enhance their research and or collaboration capacities.

Timeline for the implementation of the recommendations

In the following we suggest a near- and mid-term timeline for the implementation of the recommendations.

Near-term (<3 months)

- Decision where to host the NICIS (Recommendation 1, Recommendation 5).
- Establish an Implementation Group led by DST to implement the new organisation, including standing up Skills/Training Area.
- DST establish the NICIS Board (Recommendation 6, Recommendation 7).
- Increase funding to Data Services Area and bolster that program (Recommendation 11, Recommendation 17).
- Continue DST investments in Tier 1 facilities (Recommendation 2).
- Facilitate access to both Tier 1 and Tier 2 facilities (Recommendation 3, Recommendation 4).
- Draft job description and advertise for Chief Executive Officer (CEO) and Directors of Services areas (Recommendation 6).

Mid-term (3-6 months)

- When a hosting decision is made, implement the recommended NICIS structure and appoint the CEO and other positions necessary (Recommendation 6, Recommendation 12, Recommendation 13, Recommendation 14, Recommendation 15, Recommendation 17, Recommendation 19)
- Integrate allocation of resources (Recommendation 7, Recommendation 9, Recommendation 10, Recommendation 18, Recommendation 16).
- Engage user forums and establish metrics for success in sectors (Recommendation 8).

When DST decides, after consultation with key stakeholders, on the path it wishes to take, significant efforts will be required by individuals who are motivated to expeditiously implement the plan. There is a tension between starting as soon as possible, and ensuring that the right people are hired and setting up good governance procedures. Therefore it may be appropriate for DST to appoint on an interim basis a suitable implementation leader and team with the necessary technical expertise and leadership ability and time to dedicate to this task without contention with other demands.

Professor Francesco Petruccione (chair)

Dr Scott Hazelhurst

Dr Gudmund Høst

Professor Amanda Lourens

Dr Cynthia McIntyre

Dr Richard Moore

12 December 2013

Table of Contents

Executive Summary 2

Acknowledgements 12

List of Figures 13

List of Tables 14

1. Introduction 15

1.1 Background 15
1.2 Terms of Reference 16
1.2.1 Generic Terms of Reference 16
1.2.2 Specific Terms of Reference 16
1.3 The Constitution of the Committee and the Working Groups 17
1.3.1 Committee 17
1.3.2 SANReN and SAGrid Reference Working Group 17
1.3.3 CHPC Reference Working Group 18
1.3.4 DIRISA Reference Working Group 18
1.4 The Process 18

2. Benchmarks for National Integrating Cyberinfrastructure Frameworks 19

2.1 Objective and Overview of International Cyberinfrastructure Frameworks 19
2.2 Benchmark Observations and Conclusions 20

3. Evidence-based needs analysis 23

3.1 Current status of cyberinfrastructure 23
3.1.1 High-Performance Computing 23
3.1.2 SAGrid 24
3.1.3 The Data Intensive Research Infrastructure for South Africa Initiative 25
3.1.4 Cyber Network 26
3.2 Cyberinfrastructure drivers 27
3.2.1 Astronomy 27
3.2.2 Bioinformatics 28
3.2.3 High Energy Physics at CERN 29
3.2.4 Earth Observation 30
3.2.5 Humanities and Social Sciences 30
3.2.6 Virtualisation and cloud computing 30
3.2.6. Open Access Scholarly Publishing 30
3.2.7 Other areas 31
3.3 Needs for specialised and regional facilities 31
3.3.1 Recommendations 33
3.4 Cyberinfrastructure opportunities 34
3.4.1 Cyberinfrastructure and science 34
3.4.2 Cyberinfrastructure, the Humanities and Social Sciences 34
3.4.3 Cyberinfrastructure and Medical Informatics 35
3.4.4 Cyberinfrastructure and industry 36
3.4.5 Cyber-industry 37

4. Proposed model organisation of national cyberinfrastructure 39

4.1 NICIS Principles, Vision and Mission 39
4.1.1 Principles 39

4.1.2 Vision ... *39*
4.2 High level NICIS Organisation and Ownership ... 40
4.2.1 A Unit under the Meraka Institute at CSIR ... *40*
4.2.2 A separate unit under CSIR ... *42*
4.2.3 A National Facility under NRF ... *43*
4.2.4 A separate legal unit owned by DST ... *45*
4.2.5 Transitional considerations ... *47*
4.3 Governance, operation and management structures ... 48
4.3.1 Governance ... *48*
4.3.2 Management structure and operations ... *49*
4.4 Funding models ... 51
4.4.1 Overall funding of NICIS ... *51*
4.4.2 Other sources of funding ... *52*
4.4.3 Research funding ... *53*
4.4.4 Cost implications of our recommendations ... *54*
4.5 Stakeholders ... 55

5. Proposed model for each of the cyberinfrastructure service areas ... 58
5.1 Networking Services ... 58
5.1.1 Need for and feasibility of a single NREN ... *58*
5.1.2 Remit of the NREN ... *60*
5.1.3 Connectivity Policy ... *61*
5.1.4 Financing and strategic direction ... *62*
5.2 Computing Services ... 62
5.2.1 Requirements for Sustainable Computing Services ... *63*
5.2.2 Identification of key functions ... *64*
5.2.3 Establishment of an accredited national distributed computing system ... *67*
5.2.4 Framework for funding institutional mid-ranged facilities ... *68*
5.3 Data Services ... 69
5.3.1 Background and Context ... *69*
5.3.2 Remit, nature and function of Data Services ... *70*
5.4 Skills and Training Services ... 73
5.4.1 Background and motivation/Description ... *73*
5.4.2. Sustainable Skills and Training Services ... *73*
5.4.3. Identification of Key Functions ... *75*
5.5 Administrative Services ... 76

List of References ... 77

Glossary of Abbreviations ... 80

Glossary of Terms ... 83

Appendix A: List of People Interviewed ... 84

Appendix B: Committee Responses to SWG Inputs ... 86
B.1 High-Performance Computing Sector Working Group ... 86
B1.1. Response to the HPC SWG Document ... *86*
B1.2. Proposal for second node ... *87*
B.2 Networking Sector Working Group ... 88
B.3 Data Sector Working Group ... 89

Appendix C: International Benchmarking ... 91
C.1 European Union ... 91
C.1.1 EGI ... *91*
C.1.2 PRACE ... *92*

C.1.3 GEANT ... *93*
C.1.4 DANTE ... *93*
C.1.5 e-IRG ... *93*
C.1.6 NRENs ... *94*
C.1.7 TERENA (www.terena.eu) ... *94*
C.1.8 EIROforum ... *94*
C.2 Finland ... 95
C.3 Denmark ... 96
C.4 Netherlands ... 98
C.5 United Kingdom ... 100
C.6 United States ... 102
C.6.1 National Science Foundation ... *102*
C.6.2 Department of Energy ... *103*
C.6.3 US National Networking Programs ... *104*
C.7 Canada ... 105
C.8 Australia ... 106
C.9 Singapore ... 108
C.10 Taiwan ... 108

Acknowledgements

The report of the International Committee for the Development of South Africa's CyberInfrastructure System, commissioned by the Department of Science and Technology, is the result of contributions of a number of local and international experts and stakeholders. We thank everyone who contributed their time and for their valuable input.

The Committee would like to thank a number of Institutions and key stakeholders for their assistance, advice and input during the compilation of this report and duration of the project. In particular, we are very grateful to the DST, the CSIR, and the CHPC. The contributions and input from members of the Sector Working Groups and individual researchers are also acknowledged and the Committee would like to thank them for their support.

The Committee would also like to thank Dr Thomas Auf der Heyde, Dr Daniel Adams, Charles Mokonoto and Ntombi Ditlopo who continuously provided advice, general and administrative support to the Committee.

List of Figures

Figure 4.1 Current structure of the NICIS.

Figure 4.2 NICIS as a separate unit under CSIR.

Figure 4.3 NICIS as a National Facility under NRF.

Figure 4.4 NICIS as a separate legal unit owned by DST.

Figure 4.5 NICIS governance model.

Figure 4.6 The main stakeholders of NICIS.

List of Tables

Table 3.1 Utilisation of the Tsessebe cluster per discipline over the last year.

Table 3.2 Overview of the current status of the Sa Grid.

Table 4.1 Broad categorisation of stakeholders within the South African National System of Innovation.

1. Introduction

1.1 Background

It is now globally appreciated that enormous benefits can be derived from integrating national CI [2] into a cluster of mutually supporting activities which require high-level synergistic financial, strategic planning, oversight, and management. To determine the appropriate model for the optimal integration of the components of the South African CI, the DST has initiated this study and document, undertaken by this Committee and associated Working Groups.

The project outline for the development of a framework for a National Integrated Cyber-Infrastructure System specifies CI to comprise the technologies, software and skills which underpin e-science-driven and supported research. This document recognises CIs as an essential core component of Research Infrastructures. Research Infrastructures, in turn, are the bedrock on which much advanced education, research and innovation are pursued.

South Africa's *National Research and Development Strategy* (NRDS) (2002) identified ICT as a key technology mission to bring about improved quality of life and enhance the economic competitiveness of industry. The *Information Communication Technology (ICT) Research, Development and Innovation Strategy* was compiled in 2007 with the purpose of creating an enabling framework for the advancement of ICT R&D and Innovation, within the context of the National Research and Development Strategy. The vision being South Africa as an inclusive information society, where ICT-based innovation flourishes. The Strategy addresses human and research capacity needs, and focuses on ICT R&D in creating a robust innovation value chain.

In the report of the *Ministerial Review Committee on the Science, Technology and Innovation Landscape of South Africa* (2012), specific reference was also made to the need for the development of a CI roadmap for the country.

To realise these strategies' objectives, government committed itself to invest in high-end ICT based infrastructure or CI for the provision of high performance computing capability, broad bandwidth communication networks, and data storage and management systems.

Currently, the main components of the South African National Cyberinfrastructure are the CHPC, the SANReN, the Very Large Database and Management Systems (VLDB) and the SAGrid Initiative. It is also important to note that outside these, other parties own and manage diverse other components of the broader SA CI ecosystem.

[2] Like the physical infrastructure of roads, bridges, power grids, telephone lines, and water systems that support modern society, **cyberinfrastructure** refers to the distributed computer, information and communication technologies combined with the personnel and integrating components that provide a long-term platform to empower the modern scientific research endeavor (National Science Foundation, 2013)

1.2 Terms of Reference

1.2.1 Generic Terms of Reference

The generic Terms of Reference (ToR) for the advice to be offered to the Minister of Science and Technology extend to the following:

- the nature, governance and structure of a sustainable National Integrated Cyberinfrastructure System;
- the relationship between the different components of such a system;
- CI requirements not catered for within the current suite;
- the role of the state (government) relative to the users;
- roles of the stakeholders and players, including other Government Departments (e.g. DHET, Department of Communication (DoC), Department of Trade and Industry (DTI), Department of Health (DoH), NRF); and
- an appropriate financial sustainability model.

1.2.2 Specific Terms of Reference

The Specific Terms of Reference set out below for the key components of the NICIS list specific issues that the individual reports are required to consider, in order to advise the Minister accordingly.

South African National Research and Education Network (South African NREN)

- The need for and feasibility of a single NREN;
- The remit, nature and functions of NREN;
- The most appropriate operational, management, sustainability and governance model for a SA NREN;
- A suitable connectivity policy, that is, which institutions should qualify to be connected to the national research and education network;
- Added value service and human capital development of the participating community so as to maximise impact of NREN; and
- Different funding options, taking into account the nature of the subscribing institutions.

The South African Grid Initiative (SAGrid)

- Remit, nature and function of SAGrid;
- Sustainability of the Grid initiative; and
- Most appropriate operational, management, sustainability, structural and governance model.

Centre for High Performance Computing (CHPC)

- Remit, nature and function of CHPC;
- Sustainability of the CHPC;
- Most appropriate operational, management, sustainability, structural and governance model;
- A consolidated HPC Ecosystem integrating various sizes of systems for equitable access; and
- Industry policy to enable different enterprises to take advantage of HPC systems.

Data Intensive Research Initiative of South Africa (DIRISA)

- The need for and feasibility of a Data Intensive Research Initiative of South Africa;
- Remit, nature and function of a DIRISA;
- The most appropriate operational, management, sustainability and governance model; and
- The strategic role in national research data infrastructure provision, as well as infrastructure policies to support the Initiative.

1.3 The Constitution of the Committee and the Working Groups

1.3.1 Committee

The members of the Committee are as follows:

- Prof. Scott Hazelhurst (Wits), Associate Professor, School of Electrical & Information Engineering.
- Dr. Gudmund Høst (Nordic e-Infrastructure Collaboration), Director.
- Prof. Amanda Lourens (North-West University), Vice-Rector: Research and Planning.
- Dr. Cynthia McIntyre (Council of Competitiveness), Senior Vice President.
- Dr. Richard Moore (University of California San Diego), Deputy-Director of the San Diego Supercomputer Centre.
- Prof. Francesco Petruccione (University of KwaZulu-Natal), South African Research Chair for Quantum Information Processing and Communication (Chairperson).

1.3.2 SANReN and SAGrid Reference Working Group

The members of this Working Group are as follows:

- Dr Jeffrey Mabelebele (CEO: HESA)
- Dr Duncan Martin (TENET)

- Mr Attie Juyn (representative of ASAUDIT)
- Prof. Colin Wright (SANReN, Meraka) (Chairperson)

1.3.3 CHPC Reference Working Group

The members of this Working Group are as follows:

- Prof. Phuti Ngoepe (UL)
- Prof. Daya Reddy (UCT)
- Dr. Happy Sithole (CHPC, Meraka) (Chairperson)

1.3.4 DIRISA Reference Working Group

The members of this Working Group are as follows:

- Mr. Wim Hugo (SAEON)
- Dr Martie van Deventer (CSIR)
- Prof. Ekow Otoo (Wits)
- Dr Dale Peters (UKZN) (Chairperson)
- Dr Happy Sithole (CHPC, Meraka)
- Prof Colin Wright (SANReN, Meraka)

1.4 The Process

The Minister of Science and Technology, Mrs Naledi Pandor signed the invitations to serve on the Committee for the Development of South Africa's Cyberinfrastructure on the 11th of July 2012.

The Committee and the Working Groups were constituted at a first meeting in Cape Town (CHPC Offices) on 26 and 27 September 2012.

The Committee held regular (monthly or bi-weekly) videoconferences in between the Face-to-Face meetings.

A second meeting, including a User Community Workshop took place in Pretoria (DST) on 28-30 January 2013.

A third meeting took place in Durban on 24 and 25 June 2013. The Committee also met with the National Research Infrastructure Roadmap Committee in Cape Town on 28 June 2013.

Some members of the Committee took part in the 5th African Conference for Digital Scholarship and Curation, held at the University of KwaZulu-Natal in Durban from 26-28 June 2013. Dr Gudmund Host delivered the Opening Keynote address entitled "Some global, regional and national responses to the opportunities offered by research data infrastructures". Prof. Francesco Petruccione took part in the Panel Discussion on "South African Response to the opportunities offered research data infrastructures: DIRISA dependencies, activities, and recommendations" and gave feedback on the work towards a National Integrated Cyberinfrastructure System. Dr Richard Moore presented the invited paper "Scientists and Librarians working together: Results of Pilot Projects to Curate Digital Research Products".

A draft interim report was submitted to the DST prior to a presentation by F. Petruccione to the DST Executive Committee on 26 August 2013. Scott Hazelhurst, Amanda Lourens and Cynthia McIntyre attended the meeting as well.

A provisional report was submitted to the Minister on 1 October 2013.

The Committee's fourth face-to-face meeting was held from 9–11 October at which the committee held a meeting with the Minister, senior officials from the DST and the chairs of the working groups. Based upon these meetings, the provisional report was revised and re-distributed to DST and the chairs of the working groups. Further written comments were received from DST, DIRISA and the Network Working group, which were taken into account in the final document.

2. Benchmarks for National Integrating Cyberinfrastructure Frameworks

2.1 Objective and Overview of International Cyberinfrastructure Frameworks

One element of the remit to the NICIS Committee is to provide benchmarks for integrating frameworks that have been developed for national research cyberinfrastructure (CI) in other countries, and to identify principles and lessons learned that would inform recommendations for a South African CI integrating framework. The focus of this benchmarking exercise is on the *management structures* for coordinating and overseeing CI programs, rather than the level of financial investments, the comparative size/capabilities of facilities, or the specific nature of those programs,

To support this benchmarking exercise, the Committee has reviewed published reports and presentations from many countries and leveraged personal knowledge of international CI programs and organisations. The scope includes efforts from the European Union (e.g. EGI, PRACE, GÉANT, TERENA and DANTE), the United Kingdom, Finland, Denmark, the Netherlands, the United States, Canada, Singapore, Taiwan, Korea and Australia. Limited space and time prevent a comprehensive review of all international CI programs, but the countries considered are diverse and represent a wide range of organisational models. The integrated observations and conclusions from the benchmarking exercise are summarised below; additional information about each of the international programs is included in Appendix C.

2.2 Benchmark Observations and Conclusions

The key observations and conclusions that emerge from this benchmarking exercise are:

- Many countries worldwide are investing substantially in CI to support scientific research. There is a broad recognition of the value of these investments to facilitate scientific and technological innovation and to drive economic development.
- The national economic impact of CI investments is recognised and widely supported, including the impact on the competitiveness of a country's industrial base.
- Some countries have a long history of significant CI investments for scientific research, while many others are making relatively new investments.
- Information technology infrastructure, including computation, data-intensive science, and networking, is pervasive across virtually all fields of science and much of industry.
- In addition to CI investments per se, there is increasing attention to the integrating frameworks for those investments in order to more effectively manage programs, deliver services to users, and realise potential synergies and efficiencies.
- There is no 'one-size-fits-all' or best approach when it comes to management models for integrating frameworks. Different models have been successfully tailored to a range of situations in various countries.
- The economic and educational benefits of pervasive internet access, at increasing bandwidths, are well recognised for science and society.
- Most CI programs focus on computing and networking, but data elements are now emerging at a rapid pace.
- Workforce development is crucial to educate students and train staff to develop the expertise necessary to deploy and operate CI.

For more than a decade, South Africa has recognised the importance of information technology as a key element of its national strategy for economic benefit (e.g. National Research and Development Strategy, 2002; Innovation Towards a Knowledge-Based Economy: Ten-Year Plan for South Africa", (2008-2018); Information and Communication Technology: Research and Development and Innovation Strategy, 2007; the ICT RDI Roadmap 2012). The DST has laid the groundwork for CI investments to support scientific research via CHPC, DIRISA and SANReN. The benchmarking exercise validates DST's decisions to invest in these programs and demonstrates that it is timely and crucial for South Africa to identify an integrating framework that is best tailored to the needs of South Africa's cyberinfrastructure going forward.

It is important to recognise that some of the well-known efforts in the European Union (EGI, PRACE), the United States (e.g. NSF's XSEDE, and DOE labs) and Canada (Compute Canada) represent frameworks for integrating *multiple centres with similar resources and overlapping functions*. The integration, coordination and governance functions for these frameworks are fundamentally different from those required for an integrating framework in South Africa, where there is only one organisation for each sector (computing, data and networking) and little overlap in the functions for each sector. Thus, while those international organisations are interesting, their experiences may be less relevant to South Africa's situation than the experiences of other countries where the focus is to integrate a small number of organisations across sectors.

Many different integrated framework models have been developed for countries more analogous to South Africa with a small number of (non-overlapping) facilities. While there is not a consensus that suggests one specific model that would be optimal for South Africa, there are general management principles for successful international frameworks that drive the guiding principles laid out by this Committee for the South African NICIS (see Section 4.1.1).

It is this Committee's observation that centralised resources and capabilities, when the organisations are well-run and responsive to the user community, can function quite effectively. For example, Finland's centralised cyberinfrastructure organisation (CSC) is an excellent example. The Finnish CSC model is a wholly government owned non-profit company. It caters for all CI elements within one organisation, managed as five focal service areas. This integrated organisational structure has enabled CSC to position itself as one of the strongest CI centers in Europe, which is quite remarkable for a country with a population of only 5 million. More details are given in Appendix C.2 and we will return to this model in Section 4.

This is relevant in the South African context in that the Committee recommends that DST amalgamate its current structure of separate organisations for each primary sector function into an integrated organisation encompassing all CI elements in a single management framework. There must be accountability and responsiveness of the national organisation to the user community. And there must be collaborations between the national organisation and the universities, particularly in the area of human skills and training.

Recommendation 1: *It is recommended that DST amalgamate the structure of separate organisations for each primary sector function into a single integrated CI organisation (NICIS) that encompasses the multiple sector functions (e.g. networking, compute, data).*

In many countries, there is a natural tension between centralised (e.g. "Tier 1") versus distributed ("Tier 2") resources and capabilities; for example in South Africa, a single national computing facility like CHPC and multiple university- or program-based computing facilities. The advantages of investing in centralised Tier 1 facilities include inherent advantages of larger-scale resources (e.g. highly-parallel capability computing), achieving the critical mass of technical expertise to effectively operate high-end cyberinfrastructure, operational economies of scale, and more streamlined management and overhead costs. On the other hand, distributed Tier 2 facilities create a more diverse ecosystem of resources and expertise, may offer more opportunities for staff development, and often have more political support due to an expanded stakeholder base.

Recommendation 2: *Given the realistic level of CI investments by DST, the Committee recommends that DST and its NICIS continue to focus investments in Tier 1 centralised facilities within each sector, with an expectation of collaboration with regional Tier 2 facilities.*

South Africa is committed to major international projects such as the Square Kilometers Array (SKA), and thus must aspire for its research CI facilities to not only serve South

African researchers well, but to be recognised as competitive on an international scale. Given differences in economic GDP and national budgets, this does not mean vying for the top spot in the Top500 computing list, but rather being internationally recognised for expertise and capabilities. For example, a success criterion could be that South Africa becomes a highly attractive partner for international collaborations, for example within the European Union (EU) Framework Programmes.

While the focus of this benchmarking exercise is on management models, several other questions were raised to this Committee for comparison to international benchmarks. For example, how does the level of South Africa financial investments or the scale of facilities achieved with those investments compare internationally? To provide an anecdotal comparison of cyberinfrastructure investments, the US has a gross domestic product (GDP) ~40X that of South Africa. The combined budgets for the US funding agencies supporting CI for non-military research (NSF Office of Cyberinfrastructure and the DOE ASCR Facilities division) is ~$500M/year, or ~25X that of DST's annual ~$20M/year investments in CHPC, SANReN and DIRISA. Thus, while the scale of South Africa's investments is much smaller, the investment is competitive as a fraction of national Gross Domestic Product (GDP).

Another benchmarking question regarded common practices in charging user fees versus providing resources free to researchers. It is much more common for the government to fund facilities directly with no cost passed on to researchers, than to construct a system where the costs of using facilities are borne by the researcher. While there are pros and cons of these alternatives, the former is simpler to implement and incentivises users to take advantage of what should be well-run, cost-advantaged national facilities rather than develop local solutions. Networking infrastructure can be an exception, with institutions (e.g. universities or centers) sharing in costs of national networks, but this is at an institutional level rather than at individual researcher level. Industry usage of government research facilities is often charged back to the industry users. Also, if scientific projects (e.g. SKA) want to leverage national facilities but have dedicated access and control of a portion of the resources for its specific purposes (rather than compete with other researchers and priorities), the government may decide to allocate costs to that project.

Finally, it is often challenging in a government budget process to plan for the large capital investments generally required to refresh CI systems every 3-5 years. Countries with many facilities can spread out their investments across facilities, but this is less applicable to South Africa with a small number of facilities. (This is one of the potential benefits of amalgamating the sectors into one central organisation.) While leasing equipment can level out the costs of these capital investments, this approach is not common internationally. Essentially, most countries prefer to find mechanisms to manage the uneven capital investments rather than pay the interest and administrative costs associated with leases.

3. Evidence-based needs analysis

This chapter briefly reviews the current state of CI in South Africa and then discusses the needs for advanced CI.

3.1 Current status of cyberinfrastructure

South Africa has already made significant investments in CI. This section describes the current state of CI in order to situate and justify the recommendations that the Committee makes. The section focuses on investments in the public sector; however, we are aware that the private sector — for example, telecommunications companies, large industry such as SASOL and ESKOM has significant CI capacity. Thus we primarily discuss the public sector, which is more coordinated and by its nature more open with respect to both current status and requirements.

3.1.1 High-Performance Computing

The CHPC is the largest component of the country's High-Performance Computing (HPC) infrastructure and plays an important leadership role. Established in 2007, its technology platforms have already gone through significant renewal. The major roles of the CPHC are to provide computational resources to the South African research community and to support human capital development through training and to a lesser extent funding of students in research projects. The CHPC also has industrial customers — although this is a relatively small component of the CHPC's work, it is an important one in supporting local industry, Most of the research that the CHPC supports is extra-mural, for example in universities. However, there is an important intra-mural research component through the CHPC's Advanced Computing Engineering Lab (ACE).

Overview of hardware

The iQudu cluster was the original major computing resource of the CHPC with 160 four core machines, each with 16GB of RAM, and 96 TB of storage, and a nominal peak rating of almost 3 teraflops. Although superseded by subsequent equipment, this cluster is still in service, and plays an important role as a Tier 2 grid facility for the LHC ALICE experiment.

IBM made a donation to the CHPC of a BlueGene/P machine in 2008. This is a 4096-core computer with a specialised interconnect, and 97 TB of external storage. Although a valuable addition to the CHPC's ecosystem, the difficulties of keeping the machine in service and promoting the use of it shows that hardware acquisition must be part of a careful strategy.

The CHPC's current flagship facility is the Tsessebe cluster, a heterogeneous cluster of Oracle and Dell equipment (mainly Intel Xeon CPUs) with a peak performance of 61 teraflops and over 7000 cores and 480 TB of external storage.

Some specialised facilities, a GPU cluster and a Convey FPGA-based machine, round up the CHPC offerings.

Utilisation

There has been a rapid growth in the number of users registered on the cluster. The utilisation of the Tsessebe cluster was reported to be 91% in April 2013. The figures presented in the working group's report show that the distribution of resources per discipline (as measured by numbers of jobs) over the last year was:

Discipline	Percentage (%)
Climatology	38
Chemistry/Biochemistry	20
Material Science	14
Physics	12
Engineering	11
Cosmology	4

Table 3.1 Utilisation of the Tsessebe cluster per discipline over the last year.

There has also been sustained activity in identifying bottlenecks to adoption. For example, there has been relatively small uptake in the use of the facilities by the bioinformatics community, which has significant HPC needs. In response, the CHPC and DST have established a steering committee to liaise with the bioinformatics user and appointed two members of staff with a specific mandate to serve their needs.

Human Capital Development

The CHPC has recognised from the start that people are the key critical success factor, and has identified a number of methods of improving skills in the country, including:

- funding of graduate students, through the flagship research programme;
- winter schools which give high-level training to graduate students and others;
- ad hoc training workshops, especially associated with the national HPC research conference;
- the student cluster competition inaugurated in 2012.

These training activities have been well attended and successful, and have shown the need for further investment in training

3.1.2 SAGrid

Grid computing provides mechanisms for sharing decentralised, autonomous computing facilities. For example, the current SA Grid setup, users at UCT may store files at UFS and execute jobs at UJ and the SAAO. This is accomplished through federated identity (user-certificates that promote mutual trust) and the use of grid-middleware — software that sits on top of the operating system and that allows sharing of storage and computational resources around the world.

The SA Grid is coordinated from inside SANReN/Meraka and utilises the hardware facilities of the CHPC and several universities. An overview of the current status of the SA Grid can be found in Table 3.2. It should be noted that the CHPC grid facilities are dedicated exclusively to grid activity, whereas the other major resources based at the universities are shared between local and grid users. SA Grid has been involved in several training initiatives both inside South Africa and across Africa. It has been actively involved in several EU programmes, with its current participation in the EU CHAIN-REDS project, which explores sharing resources across different cyberinfrastructures.

With regard to the location of SA Grid, currently it sits for historical reasons within SANReN. While an acceptable home in the short term, for grid computing to be successful it must have a more stable and integrated position in the national CI.

Site	Number of cores	Storage exposed to grid (TB)
CHPC	550	49
Free State	1308	3
UJ	236	7
UCT	16	
Wits	182	28
NWU	12	

Table 3.2 Overview of the current status of the SA Grid.

3.1.3 The Data Intensive Research Infrastructure for South Africa Initiative

The Data Intensive Research Infrastructure for South Africa (DIRISA) Initiative is the newest of the national CI initiatives. Shortly after the launch of the CHPC and SANReN, it was recognised that there would soon be a need to manage very large data sets. The technical requirements of efficiently and reliably storing these data sets are a huge challenge in themselves; however, the need for data-intensive research goes beyond the physical storage requirements to issues such as curation, provenance, trust, digital preservation, and analysis techniques. After a preliminary proposal was submitted to DST, DST agreed in principle to explore the concept and CHPC was charged with developing the concept further.

At a physical level, the current infrastructure comprises two petabyte storage systems in Cape Town and Pretoria, connected via SANReN. The first phase of work on DIRISA is to establish data repositories for some of the disciplines CHPC services, e.g., climate modelling, astronomy and earth observation.

The DIRISA Sector Working Group's reports and previous reports that led up to the current pilot have put up a convincing case for the need for such a data service, and it is one that the Committee are happy to endorse. However, more work is needed to flesh out the model and the budget. For one, the Committee is unclear about what size data repository is required, or what the costs of the higher-level services required to meet the DIRISA objectives will be. This is not intended as a criticism since an initial investment is needed in facilities before it can start to attract users, and to be able to build up some plans for large-scale usage.

3.1.4 Cyber Network

There has been sustained investment in the network by both the DST through SANReN and the universities through TENET. All participants in our discussions emphasised that the South African networking landscape has been revolutionised in the last 5 years, with remarkable investment in network infrastructure, coupled with effective use of new international links.

SANReN was established as a unit within Meraka starting from the 2006 financial year, and has been funded by DST with approximately 700 million Rands since then. Although it is a component within the CSIR, it has a dedicated mandate from DST, with the CSIR providing organisational support. Notable achievements have been the establishment of the 10Gbps backbone, fibre rings in four of the metropolitan areas, and a range of dedicated links to most universities and research institutions. Work is continuing on rolling out full connectivity as some institutions still have limited national connectivity.

TENET is a public benefit organisation controlled and funded by the universities and other research institutions and science councils. It is effectively the Universities' ISP and connects 160 campuses of 56 institutions across the country, and it acts as the operator of SANReN's network. TENET has primarily been responsible for the acquisition and paying for international bandwidth. Major components of the TENET's network are a 10 Gbps link to Amsterdam on the SEACOM cable system, a 5Gbps international link to London on West Africa Cable System (WACS), the SANReN 10 Gbps backbone, SANReN fibre rings in four metropolitan areas, as well as a number of other networks of a range of providers. TENET operates transit and peering links, including GÉANT.

Operating the network is expensive, and the 2012 budget was R112 million which is fully funded by TENET's members. As there is no direct subsidy from the state, TENET funds the operating costs of both the international links and the South African infrastructure including the operating costs of the SANReN links that it uses. However, there is a substantial indirect subsidy since these costs do not reflect the capital investment in the links.

Both TENET and SANReN see improved value added services, such as Eduroam and conferencing as part of their mandate.

These investments have resulted in substantially faster and more reliable networking. A full overview of network performance and utilisation is not available, but there is strong evidence to show growth in network performance and utilisation. Bandwidth usage is very variable through the year, and to some extent varies according to the time of day. TENET reported a 24-fold increase in peak incoming international traffic between October 2009 and October 2012, with peak daily international traffic exceeding 6Gb/s by the end of 2012 and 8Gb/s by the end of 2013. An examination of many institutions' network logs shows strong growth over the period, with their network links effectively being saturated during weekdays. This implies that additional investment is required.

Although the Committee has seen important strides forward, there are some concerns.

- Network bandwidth is still limited by international standards. For examples, the network bandwidth available for ALICE and ATLAS collaborations is approximately an order of magnitude less than for our European collaborators. This makes a substantial impact on the ability of scientists to participate in experiments. For example, at 10Mbps, transferring a 1TB data set will take over a week.

- As is pointed out in the network working group's reports, the configuration of the network for scientific computation takes planning and requires different measures and settings to those required for general services such as email and web browsing. Often, Networking Services at institutions are geared to support general services. In this context a 10 Mbps channel for a typical user probably provides a high quality service, but for moving large data sets, it is not.

- In some cases, some participants in our meetings noted that the connection of their campus to SANReN had not resulted in any improvement in network performance, because of local connectivity issues. Competing views about why this was the case were expressed, and it seems that different factors were at play in different places. However, the two leading factors were (i) lack of preparation on the campus; (ii) insufficient coordination of SANReN and university planning. As a case in point, a recent study by the H3Africa Bionet Infrastructure Working group on network connectivity showed marked differences in connectivity between South African sites. Effective bandwidth between the CHPC and three different urban universities showed an almost magnitude difference in available bandwidth. This cannot be explained by SANReN infrastructure and is either a sign of deliberate policies to throttle local networking connectivity (probably because networking policies are geared to dealing with commodity web browsing rather than big data) or dated or misconfigured local routers.

- At some institutions, network utilisation is significantly smaller than capacity. This in itself is not a reason for concern as the installation of infrastructure is a long-term investment. However it *may* be a marker of an issue to be addressed.

3.2 Cyberinfrastructure drivers

A number of science projects continue to drive the need for CI growth and expansion. These projects typically have a combination of needs for high-performance computing, fast networking and large data sets.

In this section we survey four of these areas as exemplars (it is not intended by choosing these four to suggest that they are the most important). Other examples such as the Applied Centre for Climate and Earth System Science (ACCESS), the Southern African Large telescope (SALT), Karoo Array Telescope (KAT7) and MeerKAT, electronic-Very Large Baseline Interferometry (e-VLBI) at Hartebeesthoek Radio Astronomy Observatory (HartRAO) and its extensions also illustrate the need for cyberinfrastructure renewal and investment.

3.2.1 Astronomy

South Africa's success in being the premier site of the SKA rests not only on our natural heritage and the skills of our scientists, but also on the belief that we have the capacity to support the SKA project computationally. The computational needs of the SKA by far exceed the demands of the other projects mentioned in this section.

It is estimated that each dish will produce 160Gb/s. Key design choices must still be made regarding how and where data will be processed and transmitted. Processing centrally will impose huge networking challenges, whereas a fully distributed framework imposes formidable infrastructure challenges of establishing large computing facilities in remote places. However this tension is resolved, the computational requirements will be in the petaflops range and the networking requirements in the terabit per second range at the minimum. The challenges for both infrastructure and skills are immense.

Work with the Meerkat and KAT-7 telescopes, which have acted as prototypes for the SKA, has shown that there is significant local capacity but the computational, storage and network demands are high. The success of the SKA will require huge advances in computing. This is an international effort and responsibility but South Africa should use the opportunities that are provided.

One of the opportunities of the SKA and related projects is that it gives South African scientists the opportunity to develop a significant local capacity in digital design. The large computational requirements of the SKA make novel architectures using technology such as FPGAs very attractive, because these architectures can support highly parallel applications at low cost (high flops per watt measure). The design of such architectures does not require anywhere near the same order of capital investment as VLSI design; it does instead require investment in people.

3.2.2 Bioinformatics

The South African bioinformatics community, as other bioinformatics communities is heavy users of computing infrastructure. Without high-performance computing, fast networks and large data storage repositories, it will be impossible for bioinformatics to meet national and international challenges.

The development of bioinformatics has been an important thrust of the DST since at least the 2001 publication of the national biotechnology strategy (Parker *et al*. 1999). The National Bioinformatics Network was funded by the DST from 2003-2009 to support capacity development, and subsequently the DST has continued to support bioinformatics through ring-fenced funding through the NRF, the CHPC and its other entities.

Modern bioinformatics is extremely computationally intensive. The rapid dropping of the cost of sequencing relies very heavily on high-performance computing. For example, the *de novo* sequencing of moderately complex genomes may require machines with memory of 1-2TB (to put into context, more than 100 times greater than a very well configured desktop machine). There are only a handful of such machines in the country. A number of projects have already established South African expertise and this need is likely to grow. Other examples include population structure analysis (a modest run of which may require a thousand CPU hours) and protein docking.

Storage requirements are high too. The proposed first phase of the South African Human Genome Programme is likely to generate 50 TB of raw data, before serious analysis done. Few sites in South Africa can accommodate this easily. There will also be a need to move large data sets internationally and between collaborating sites.

One of the stumbling blocks for the use of CHPC facilities was the complex and highly heterogeneous requirements for software by the bioinformatics community. The DST and the CHPC have taken the initiative to address this problem by hiring two specialists to support the bioinformatics community. A Bioinformatics Steering Committee, chaired by DST and with members from CHPC, Technology Innovation Agency (TIA), the universities and an international expert set the strategic goals and high level work plans. This is a model, which may be worth replicating in other areas, though we note that needs are domain specific.

There is a direct impact of the quality of the CI on the success of these bioinformatics projects, which tackle questions of fundamental science and are important for the broader biotech industry. The risks of not being successful are twofold. First, South Africa would remain merely a producer of raw material, as those who learn to exploit our biological heritage will be outside our borders. But a more severe risk is that opportunities for knowledge enhancement within problems of particular importance to South Africa would remain understudied.

3.2.3 High Energy Physics at CERN

South African scientists are actively involved in experiments at CERN, and formally collaborate on both ATLAS and ALICE experiments. The ability to participate effectively in these high-profile international experiments depend on significant computing and people infrastructure. It requires being able to move terabytes of data from the LHC to South African sites with significant computational resources, with high-level skills to support it. Data and processing for both depends on grid computing technology.

The CHPC provides the computational power to support ALICE. Currently 550 cores are dedicated to ALICE and are effectively 100% utilised.

The UJ and Wits nodes of the SA National Grid support ATLAS. In principle, 300 cores are available to ATLAS. However, these resources are also used by local users and other grid services so the effective availability for ATLAS is about 40% of this. Currently, the two sites provide 25 TB of storage dedicated to ATLAS. In 2012, over 20 core-years of computing was done for ATLAS. There are three problems that these two sites experience, the first two being particularly severe:

- Lack of technical support.
- Bandwidth constraints limiting the capacity to move large data sets from international sites.
- Lack of local storage.

In both cases, the resources directly support South African scientists who have experiments on ALICE and ATLAS. These resources are important for their research and for the high-level training of their students. These resources are also available to researchers from other countries and at any time there are likely to be dozens of jobs from around the world running on our infrastructure. This ability to support international science is not only important for our image, but is crucial for the long-term sustainability of our involvement in big science projects as we must be partners who can be relied upon to pull their weight.

3.2.4 Earth Observation

Earth observation projects are another class of projects, which require large data sets, fast networks and high performance computing. A good example is the Global Earth Observation System of Systems (GEOSS) which is a ten year plan to produce computing systems and tools that bring together different earth observation systems to support planning. The South African Earth Observation Network is a full partner of this initiative.

There are many different earth observation systems, producing a huge amount of data daily including daily images of the earth, monitor the weather, the state of the atmosphere and the oceans, seismic conditions, estimates of fish and bird populations and so on. The data storage requirements of GEOSS will be daunting in itself, but to be successful the sharing of such heterogeneous data will require sophisticated use of meta-data standards and methods of sharing and linking the different data types. Large amounts of data will need to be moved around the network, and the analysis of the data will impose very high computational resources.

3.2.5 Humanities and Social Sciences

There are relatively few examples in South Africa where the CI has directly been used to support research in the Humanities and Social Sciences. However, there have been some successful projects, which show the huge potential in the area. The fact that there have been relatively few projects is a challenge rather than showing a fundamental limitation. Some successful projects that show the potential in this area include digital libraries, digitising key heritage datasets (e.g., Bleek archives, Rock Art resources) and human language translation. The 2011 Charter for Humanities and Social Sciences also made several recommendations in strengthening the Humanities and Social Sciences in the country.

Although the Humanities and Social Sciences sector is currently not one of the major drivers of the cyberinfrastructure environment, it would be a mistake to ignore these sectors. Continued effort in engaging with scholars and cultural workers should be encouraged.

3.2.6 Virtualisation and cloud computing

Virtualisation and cloud computing are technologies that can be used for HPC, and we shall discuss them as such later in the report. However, their use reflects two important needs that some users have in HPC. Some users have very *bursty* HPC needs; they may go for months without significant needs followed by short bursts with high needs and with tight deadlines. Cloud computing may provide these users with cost-effective solutions. Some users may have operating system and software needs that do not allow them to make easy use of common facilities. By using virtualisation, users can be given root access to their machines and configure their systems as they see fit.

Finally, South Africa has niche expertise in this area, and we can see the driver of developing an indigenous software industry as being relevant here.

3.2.6. Open Access Scholarly Publishing

National and international initiatives to promote open access scholarly publishing are another area which a healthy CI can enable. These will dove-tail with the DIRISA. Already there have been exciting developments such as the *Scientific Electronic Library Online*[3]. These platforms are likely to grow in importance and provide researchers access to key research as well as providing platforms for research. Progress in this regard has been fast-tracked by the work done by ASSAf in promoting the visibility of all South African research articles through their Scholarly Publishing Unit. The South Africa Open Access Platform, SciELO has also been launched in July 2013 by the DST, which provides an excellent mechanism for researchers to publish their work. *SciELO South Africa was established in 2009 and certified in April 2013 as a fully operational collection indexed in the SciELO Network Global Portal .*

It is important to note that publications will not just be research papers, but also research data. This data will be of tremendous value to the scientific community as a whole. Examples of such data are demographic and survey data. South African-based projects such as the NIH and Wellcome Trust H3A projects, and the South African Human Genome Programme will generate huge data sets — the publication of the data that is generated is as important as the publication of scientific papers. Existing work in Digital Libraries needs to be encouraged.

3.2.7 Other areas

The list above is by no means comprehensive. For space reasons, the Committee only mentions a few more briefly:

- Although not uncontroversial, there is immense potential in e-learning for teaching and training at all levels (e.g., the development of Massively Open Online Courses or MOOCs) (Martin 2012; Waldrop 2013). These require significant CI.
- Much of the financial industry relies on financial mathematics, which in turn requires HPC.
- Animation, digital arts and gaming are areas of rapid growth. South Africa has some expertise and if we wish them to grow it is important that they be supported further.

3.3 Needs for specialised and regional facilities

The HPC Working Group and several of the groups the Committee consulted raised the issue of mid-range HPC facilities. These would be HPC sites either based at the universities or some regional facilities. The question of funding and placing of regional and institutional facilities is within our terms of reference. The needs for such facilities will be discussed below, and return to a more comprehensive discussion of computing services in Section 5.3.

Referring to the terminology of the European PRACE infrastructure (see A.2.1), the regional and institutional facilities will be termed Tier 2 facilities, as compared to national (Tier 1) resources.

There cannot be a one-size-fits-all model for HPC because the needs are very different. It would not be efficient to try to centralise all HPC in South Africa under one roof. There is clearly a need for a variety of computing facilities from Tier 1 systems such as those hosted by CHPC to Tier 2 systems consisting of departmental clusters of up to a dozen computers.

[3] http://www.scielo.org.za

There is a very strong argument for *centralising* the Tier 1 resources, and that is efficiency. By sharing infrastructure such as backup power, air-conditioning, and system administration costs, overall costs can be brought down and service improved. It is easier to optimise allocation of computer cycles to users in a single large cluster than two more modest clusters. It is easier to monitor performance and plan upgrades. This argument is a strong one, and for this reason we urge that the idea of distributing the Tier 1 resources across the country is one that would need very strong motivation. And, while the Committee argues that further Tier 2 facilities be established, they need to be carefully planned in form, function and funding model. In particular, the Committee believes that limited DST funds will be better spent for the national Tier 1 resources while Tier 2 resources should more naturally be subject to institutional strategies.

The following are commonly used arguments for availability of Tier 2 HPC facilities:

- Training needs alone justify the necessity for clusters and other computing facilities of different sizes. A large, central production site such as the main cluster of the CHPC is not a good learning environment for beginners, as seen from the perspectives of both the HPC site (it is wasteful to have poorly tuned software or training software running on the premier hardware) and the trainee. While learning to write HPC code or to tune new code, it is useful to have equipment to learn on and where mistakes can be made and analysed. The challenge of training and skills will be dealt with in Section 5.4.

- Some computational problems have relatively modest computational requirements. A modest size cluster with low cost interconnects may be a much more suitable environment than a large HPC resource. A very large production HPC environment may also be a poor environment for the development and early testing of novel HPC code, where low latency and dedicated access to the computing facilities may be very useful.

- Although, there is a significant economy of scale in building one very large cluster than several smaller ones, the challenge of system administration increases as the range of disciplines and applications running on the cluster increases. Interaction with customers may become more impersonal, and the system configuration for one set of applications may not be the best for another set.

- In the future, there may be a need for system architectures that are not directly compatible with the current Tier 1 systems. For example, currently South Africa does not have a large-scale Hadoop-style[4] facility and it would be difficult to integrate such a facility into the current cluster. However, if such a facility was to be established , it may still be more cost-effective to physically house it at the current Tier 1 facility than to build a new one.

- There may be operational reasons why additional Tier 1 facilities should be built. For example, if the CHPC continues to grow at some point it will outgrow its current facilities. What will happen then will depend on the relative costs of establishing and

[4] The Apache Hadoop project is open-source software for reliable, scalable, distributed computing. For more information see http://hadoop.apache.org.

operating a completely new facility compared to establishing and running an additional facility elsewhere.

- Regional (Tier 2) facilities should be established so that DST can be seen to be spreading resources nationally rather than concentrating in one or two places. While this argument is understandable, it is one that can lead to waste of resources so this is the weakest of all the arguments in favour of decentralisation and the Committee urges that it should be treated with considerable rigour. The success of Tier 1 facilities depends on a high-speed network. Without a fast network, an HPC facility cannot be effectively used by researchers in the same city; with a fast network it can be used by people all round the country. It is far better to concentrate time and effort on building a fast Internet and a Tier 1 facility that works well than to fragment resources.

3.3.1 Recommendations

For the reasons above, it is important that researchers have access to a range of HPC facilities. These need to be built around education and training needs, provision of specialised services, and efficient use of resources.

Recommendation 3: *It is important that researchers have access to a range of HPC facilities, including both national Tier 1 resources and institutional Tier 2 resources. The facilities need to be built around education and training needs, provision of specialised services, and efficient use of resources.*

The building block of any system is the Tier 2 facilities at institutional level, which will be especially important for training of new scientists. The funding of the Tier 2 facilities is primarily the responsibility of the institution concerned. However, some national funding, through existing research-funding instruments such as the National Equipment Programme, to act as an incentive for universities to work together may be sensible for increasing overall capacity in the system.

Innovative ideas of how to strengthen these facilities have been presented by the Director of the CHPC in a March 2013 proposal. Essentially the proposal is that a donation of significant cluster by the Texas Advanced Computing Center be used by deploying parts of the system to individual universities. Whether this specific proposal materialises, broadly, we welcome the thinking behind it and are particularly encouraged by its attention to the need to develop human capacity. However, it is important that the receiving institution show capacity to use the equipment, otherwise equipment and time will be wasted. This does not mean that the university concerned should have existing expertise in HPC, but should demonstrate a plan and willingness to acquire such skills (both scientific and system administration) and show concrete commitment to allocation of its own resources to support this programme.

Although, in the context of a fast research network, *regional* facilities at the Tier 2 level have only a small advantage over *national* (Tier 1) facilities from an access perspective, there is space for shared and/or specialised facilities either co-located with Tier 1 facilities at the CHPC in Cape Town or in other facilities around the country. These facilities can be provided either by top-down or bottom-up initiatives. In summary we cannot endorse the concept of *regional* facilities *per se*, but do support the need for different types of facilities.

One area where regional resources may be useful is by developing pockets of expertise in different parts of the country. The CHPC has already described a model where some of the systems administrative staff are based in places other than at the CHPC in Cape Town. This is an interesting idea to be explored. However, it would be important for these people to be embedded with users so that they can help with training and use of equipment; this is likely to be a pre-condition for successful implementation. The Committee believes that such pockets should most naturally be coordinated with relevant educational institutions.

In our consultations, the Committee was told that the NRF National Equipment Programme effectively does not fund proposals for high-end computing. Given the large investment that the DST has made in HPC it is understandable that the NRF review committee responds like this. However, the Committee believes that this programme should be open to funding proposals with an CI component and evaluation committees should be guided accordingly. Certainly, any proposals must demonstrate a clear need for a new facility that cannot be met by the CHPC and must be critically and rigorously evaluated. However, it would be counter-productive to have a blanket exclusion of HPC requests. The NEP requires co-funding of equipment and so supporting HPC acts as an incentive to bring new money into the system. Similar remarks can be made about funding from DHET.

There are no doubts some challenges due to the existing demands on the NEP and the relatively short life-cycle of HPC. However, this must be balanced against the damage done by not funding Tier 2 facilities. Alternative sources of funding, such as the DHET, should be considered. The Committee emphasises that Tier 2 facilities should be funded through an open and competitive process with significant co-investment from host institutions.

Recommendation 4*: The NRF NEP, DHET Infrastructure Programme and any other related national programmes should fund CI equipment (including Tier 2 computing and storage facilities) on an equal footing to other equipment requests*. *The Committee further recommends stronger coordination between the various funding initiatives.*

3.4 Cyberinfrastructure opportunities

3.4.1 Cyberinfrastructure and science

The great boost that CI can provide for science has been described above. There are few areas in modern science which cannot take advantage of a research CI. At a minimum an effective national research network is required, but from evolutionary sciences to biology, to chemistry, physics, astronomy and engineering, high performance computing is essential.

3.4.2 Cyberinfrastructure, the Humanities and Social Sciences

Although these sectors are perhaps the least developed areas in terms of application of HPC and high-end CI, there is huge potential and some good track record.

The Committee points out the following areas as ones with huge potential:

- Preservation and utilisation of key heritage information (good examples are Rock Art, key historical records);
- Storage and analysis of demographic data;

- Economics and econometrics;
- Computational tools to assist historical research;
- Human Language Translation and Linguistics Research;
- Visualisation and modelling;
- Text mining.

3.4.3 Cyberinfrastructure and Medical Informatics

Medical Informatics is a very important discipline, which requires CI support. Medical Informatics encompasses methods for storage, transmission and analysis of biomedical data (including health records, images, patient data, genomic records, …) (Denny 2012). Many international projects have shown the huge contribution that medical informatics can provide in supporting a cost-effective medical service as well as high-end research in medicine. The effective large-scale use of health and medical data such as electronic health and medical records has been shown to have significant benefits in areas ranging from clinical research to hospital and financial management (Jensen et al 2012; Koyuncugil and Ozgulbas 2012). There are risks and difficulties of technical, social and ethical nature, but there is very strong evidence for the contributions a strong medical informatics programme can make. The recent announcement of a global alliance of 69 institutions to promote data sharing of both sequence and clinical data underlies the importance and urgency of this project The proposed National Health Insurance programme makes this area particularly pertinent.

Medical informatics is challenging because the large, interconnected data sets, the complex algorithms required to analyse them, and the social demands placed on the management of such data (ethical issues such as the confidentiality and limits to how the data may be shared) and the very serious consequences if things go wrong.

Medical informatics requires a solid HPC infrastructure and the development of skills in the area. South Africa has some expertise already and a huge need — making this an exciting and necessary opportunity for development.

3.4.4 Cyberinfrastructure and industry

HPC has a major impact on industry and commerce and is an established and indispensable tool in many industrial and societal sectors. The benefits of HPC to industry go beyond a positive return on investment. Improvements in healthcare, the development of efficient transportation systems, the quest for renewable and clean energy sources, and support to decision making through faster real-time simulations on dynamical data streams in emergency response are examples of how HPC can have a major societal impact.

Many areas of South African industry already use HPC and whose need for HPC is likely to increase. Some obvious examples of sectors

- the financial sector;
- mining and chemical companies;
- biotechnology and pharmaceuticals;
- transportation
- telecommunications;
- electrical supply and demand (e.g., ESKOM);
- aeronautical and marine design;
- creative sector (e.g., animation);
- heritage.

There are very good examples of success stories in these areas in South Africa already, so our own experience as well as international case studies have shown that HPC is critical to many sectors.

A healthy national CI for industry will support industry broadly, by providing skills and technologies that can be used by industry. There are few industrial organisations that can afford a type of facility that the CHPC can provide; however, by being a shared facility CHPC can be made available to a range of industrial partners both large and small.

Planning and involvement of industry is important. The following extracts from the *A Strategic Vision for UK e-Infrastructure* (Tildesley, 2012) are particularly pertinent:

> *Firms cannot interact easily with a fragmented national infrastructure or without a clear forward commitment to a strategy. Any move from a local, in-house provision of computing to the use of a shared infrastructure will only be achieved if a clear plan is developed and adhered to. (This problem applies equally to academe.)*
>
> *There is no single coordinating body for e-infrastructure in the UK that can represent all relevant stakeholders, both providers and users, in the public and private sectors. This has resulted in the lack of a coherent strategic plan for e-infrastructure. Unless action is taken to coordinate the UK provision the socio-economic benefits from e-enabled science and innovation will stall.*

There is no coordinated plan for skill development. A steady stream of software engineers and developers is needed. In particular we need experts in mathematical and discipline-based skills to be trained additionally in software engineering for e-science. These types of developers are in short supply.

Develop a single coordinating body to drive closer cooperation and enable effective industrial access, while insuring that UK academe has access to leading edge capability. This coordinating body should develop a vision for UK e-infrastructure that will increase productivity and efficiency for all of the partners.

Pilot studies suggest that for every £1 of public spending in the industrial exploitation of e-infrastructure, £10 of net GVA will be generated within two years rising to £25 after five.

3.4.5 Cyber-industry

The application of HPC in industry has been described in the previous section. However, there is also an opportunity to grow a local cyber-industry. This will help build up a pipeline of local expertise, reduce imports and reliance on foreign expertise and even create opportunities for export. The Department of Trade and Industry already has programmes in place to support the local software industry so should be a partner in developing further strategies to support HPC.

Of course, one care needs to be taken — direct investment in high-end VLSI design and fabrication is unlikely to be competitive given the huge startup and operating costs, but there are opportunities in software, services and hardware design. There are some high-profile South African successes such as Thawte and the Amazon EC2 development group based in Cape Town. More examples can be found.

For example, in the cloud-computing environment, some of the large SA telecoms provide very significant Infrastructure as a Service (IaaS) to South African industry. At the other end there are small start-ups that provide sophisticated Software as a Service (SaaS) to a range of companies. These services need to be encouraged and there are opportunities in working with these industrial partners:

- They may be able to provide services to the scientific community in a cost-effective way either at commercial rates or as part of collaborations with the university.
- These companies need employees with the same high-level skills that the universities need for HPC. This is an additional motivation for skills development — it is not just for the research community and universities but for industry more broadly. And, while in the short-term these companies may be in competition for the same skills as those that the universities are looking for, in the long-run this demand is healthy because it makes a career in scientific computing attractive when it is clear that there is a healthy private sector demand.
- There is potential in development of niche areas in HPC software design with an international reach.
- Part of the mission of developing the national research CI should be to strengthen private sector companies.

There is some potential tension in supporting companies that have a need for HPC with developing a healthy South African HPC sector. If CHPC provides computing resources at significantly below market cost, the growth of the local computing sector may be stifled. This issue have been addressed elsewhere (see Section 5.2).

4. Proposed model organisation of national cyberinfrastructure

4.1 NICIS Principles, Vision and Mission

4.1.1 Principles

Based on our reviews of various national CI organisations and the current state of practice for the South African sector, the Committee has converged on a minimal set of main principles for an ideal national integrated CI system. These principles are:

1. Joint planning and budgeting: While investments are needed across the full spectrum of CI elements, a consolidated national strategy should be followed.
2. Good governance: transparency, accountability, efficiency and effectiveness.
3. Visibility of CI services: It will not be sufficient to offer a wide range of CI services if these are distributed over several providers and a myriad of web locations.
4. Sustainability: The technological volatility of CI must be countered with a smooth and solid funding stream. For the same reason, and due to frequent paradigm shifts, it is essential to maintain and nurture the development of the skills of the human capital base, including support personnel as well as current and future users of CI.
5. Constructive stakeholder engagement: There is a wide range of stakeholders that can affect or are affected by the development and operations of the national CI enterprise. Through systematic stakeholder engagement NICIS can better identify new opportunities manage risk and build reputation, thereby increasing value creation for both partners and society at large.

Other factors that have influenced the recommendations of the Committee include the need for innovation, the need for integrated services, international relations and cost-efficiency.

4.1.2 Vision

The DST's Ten-Year Innovation Plan aims to drive South Africa's transformation towards a knowledge-based economy, in which the production and dissemination of knowledge leads to economic benefits and enriches all fields of human endeavour (DST, 2007). The proposed model for a NICIS aims to support this plan. Responding to the challenges and opportunities that the digital age is posing for South Africa, the Committee proposes a national integrated cyberinfrastructure organisation, designed to become an effective national tool for the DST.

The vision is a NICIS that takes national leadership in the provision of a comprehensive cyberinfrastructure essential to 21st century advances for South Africa in research, education and innovation.

4.1.3 Mission

NICIS will establish a CI that is human-centred, world-class, supportive of broadened participation, sustainable, and stable but extensible. Investments should be user-driven, recognise the uniqueness of SA, provide for inclusive strategic planning, enable SA international competitiveness in research and industry, promote partnerships and integration with investments made by others in all sectors, both national and international.

The mission of NICIS is to increase knowledge creation through provision of a national platform of essential cyberinfrastructure.

4.2 High level NICIS Organisation and Ownership

The organisational structure for NICIS will be described at two levels: (i) the high level ownership and reporting structures, and (ii) the internal structure of NICIS with detailed functions and responsibilities. In the following, various alternatives for high-level organisation will be discussed, while the internal structure will be described in Section 4.3 and details for the various CI elements will be described in Section 5. The internal structure will largely be independent of the high level ownership models to be described in the following.

The Committee considered the suitability of other departments other than DST as a suitable home for the cyberinfrastructure. In the Committee's view, DST remains the best home for NICIS, taking into account DST's mission, the relationship to other DST-related entities and organisational history. There is a strong alignment between the vision of DST and the vision of the committee for NICIS. Strong links to other departments such as the Department of Higher Education and Training, the Department of Health and the Department of Communications will be important to build to meet the objectives of NICIS; these departments have a strong interest in the national cyberinfrastructure and will be important stakeholders and partners. In the longer run, as NICIS grows and proves itself, other organisational structures and relationship — perhaps as envisaged under the proposal national system of innovation — may be appropriate.

Recommendation 5: *The preferred organisational model is to set up NICIS as an appropriate legal entity fully owned by DST. An alternative viable option is to consolidate NICIS into a "National Facility" or a sub-agency of NRF.*

4.2.1 A Unit under the Meraka Institute at CSIR

Description

The Meraka Institute is an operating unit of the CSIR focused on research and development in Information and communication technology (ICT). Meraka covers the following areas:

- Earth observation science and information technology
- Human language technologies and knowledge technologies
- Networks and media
- Integrative systems, platforms and technologies
- Cyberinfrastructure

As compared to the current status quo, CHPC and SANReN are implemented as units (projects) under Meraka. In large, this alternative would be a continuation and strengthening of the current organisational model.

Pros: The main advantage would be continuity, providing a secure framework for current personnel.

Cons: One disadvantage that the Committee sees with this model is a lack of visibility on the national level. Furthermore, the main focus of CSIR is research and development; and not ICT service delivery. The Committee has not seen strong evidence of interaction between current CI elements and other units under Meraka. Thus, few synergies have been documented, although CHPC participates on some of the cross-disciplinary teams that CSIR has for various projects. The general impression of the Committee is that researchers at Meraka and CSIR are not driving or inspiring CI service development.

The customer base of Meraka and CSIR appear to be different from customers of CI. Furthermore, the customers of Meraka and CSIR do not appear to be dominating users of CI.

During interviews with stakeholders, the Committee has heard comments that the current organisation is experienced as unnecessarily bureaucratic, resulting in lengthy procurement and budgeting processes.

Assessment:

Since CSIR is not generally a service provider, it may be difficult under this organisational model to promote an ambitious national CI vision as part of a Meraka vision. Furthermore, the reporting line between NICIS and DST would be unnecessary complex, since it would go partly through the Meraka Executive Director and CSIR Executive; and partly through a project reporting line. As will be described in Section 5.3 we will recommend an authoritative governing board for NICIS. This role will be difficult to maintain if NICIS is to be embedded into an existing organisation, and it is likely that such a NICIS board in practise would end up having mainly an advisory function.

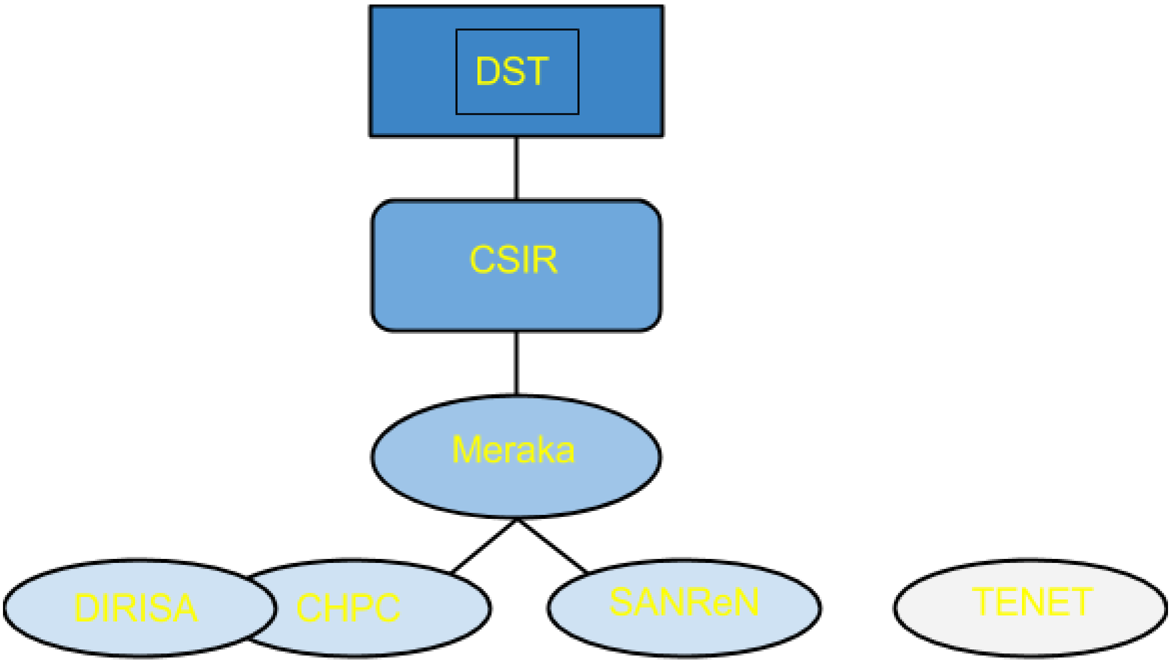


Figure 4.1 Current structure of the NICIS.

4.2.2 A separate unit under CSIR

Description

The CSIR is one of the leading scientific and technology research, development and implementation organisations in Africa. The CSIR undertakes directed and multidisciplinary research, technological innovation as well as industrial and scientific development. This takes place in domains such as biosciences; the built environment; defence, peace, safety and security; materials science and manufacturing; and natural resources and the environment. CSIR is organised accordingly into 9 research areas, one of which is research and development in information and communication technology (the Meraka Institute). Each research area is directed by a CSIR Executive Director[5].

A NICIS organisation could be established as a separate "research area" under the CSIR with a separate Executive Director reporting to the CSIR Executive. As compared to the present status quo, this organisational model would move CI out of the Meraka Institute.

Pros: This model would give a separate budget for CI within CSIR.The visibility would be increased, particularly within the CSIR organisation.

A further advantage would be that there would be one reporting level less between NICIS and DST, as compared to the present situation. The focus on CI would be strengthened within CSIR, and also nationally.

[5] http://www.csir.co.za

Cons: Many of the disadvantages given in the previous “Meraka model” still hold. There is a possibility of tension between such a CI area and other CSIR research areas. In particular this tension could emerge since the main focus of CSIR is research and not ICT service delivery. Also, the synergies are not apparent, since there seems to be little cross-fertilisation currently between CHPC/SANReN and other groups at CSIR. There may also be tension between the NICIS board and the CSIR board.

Assessment

The added value of this alternative is not obvious to the Committee. The Committee does not see significant differences between the current status quo and the present alternative.

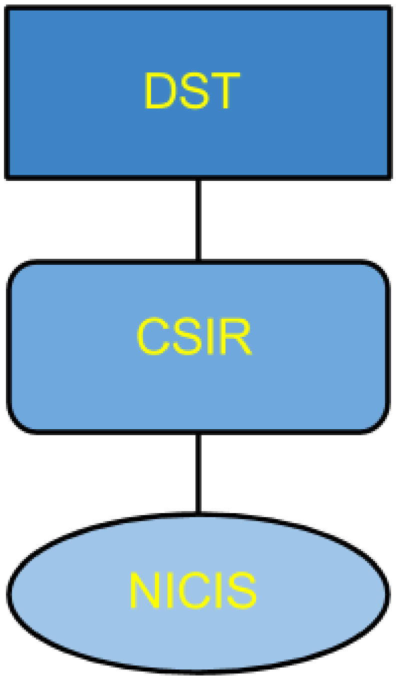


Figure 4.2 NICIS as a separate unit under CSIR.

4.2.3 A National Facility under NRF

Description

The NRF is an independent government agency which promotes and supports research in all fields of knowledge. It also conducts research and provides access to National Research Facilities. The NRF provides services to the research community and aims to uphold excellence in all its investments in knowledge, people and infrastructure.[6] Unlike other Science Councils whose role is research performance, the NRF primarily fulfils an agency role and funding from the NRF is largely directed towards academic research[7]. The main source of income for NRF is DST (R1,019 M in 2011) and a parliamentary grant (R762 M in 2010), with R66 M from other government organisations.

[6] http://www.nrf.ac.za
[7] http://en.wikipedia.org/wiki/National_Research_Foundation_of_South_Africa

A NICIS could be organised as a National "Research" Facility under NRF.[8] As a national facility, NICIS would report to the NRF Executive Management through the Vice-President for Research Infrastructure and National Research Facilities. Strategic priorities would be set by the Board of the NRF.[9]

Pros: The mandate of NRF covers infrastructure, so CI will fit naturally into this mandate. An NRF national facility would increase the visibility of CI and place CI alongside other National Research Facilities having a unique position in South African knowledge production. The user base of NICIS corresponds in large to the user base of NRF.

Furthermore, having NICIS within NRF gives the possibility to match national research priorities with CI priorities and the allocation of CI resources to projects. This is important in order to ensure appropriate return on research investments. Having NICIS organised as a facility under NRF would give one reporting level less than the current organisation under the Meraka Institute.

Cons: A NRF National CI Facility would have to "compete" internally within NRF with research funding. Such a NICIS would also have to "compete" with funding and attention to other national facilities. The reporting line to DST will still be fairly long, through the NRF executive management to the NRF Board to DST.

Assessment

The Committee finds this a viable model. However, one concern is that NICIS would have to compete with disciplinary research and research initiatives within NRF. The unique cross-cutting and generic nature of CI could be given too little attention in this context. DST would also have less opportunity to use NICIS as a strategic tool for implementing its priorities. Instead, the NRF will be given the opportunity to use NICIS as a tool for realising and supporting important national research priorities. However, fluctuating research budgets and science trends may affect the long-term sustainability of CI. Experiences from organising CI under the national research funding organisation other countries are variable.

---

[8] *Ref NRF Annual Report 2010-11: According to the White Paper on Science and Technology (1996), an institution qualifies as a National Research Facility when it has:*

- [8] *A unique position in South African knowledge production;*
- [8] *Core technologies, research methods or data pools/collections that live up to international standards;*
- [8] *Goals that are aligned with the overall objectives of the NSI;*
- [8] *Critical mass of equipment, skills and users;*

[8]*Potential for networking and attracting international collaborators; and l Opportunities for human resource development, with special efforts to involve researchers from formerly disadvantaged communities. An additional characteristic of a National Research Facility is that it provides a platform for science advancement and for bringing science and society together. The facilities offer unique education and training opportunities for students and learners. As science awareness platforms, the National Research Facilities aim to promote a culture of science and to stimulate young people to follow careers in science and technology.*

[9] NRF Directors' Advisory Committees operate iThemba LABS and SAEON. "The process for establishing committees for the other National Research Facilities will be completed during the first half of the 2011/12 financial year". Ref.: NRF Annual report 2010-11.]

As a sub-alternative, still within NRF, NICIS could be organised as a sub-agency with a separate board. The Committee has been informed that a similar discussion is on-going for astronomy. Again, this might be a viable model provided that the high-level governance structure can be set up in a transparent manner that ensures buy-in from the relevant stakeholders. The committee also has reservations regarding the role that the NICIS Board would have if NICIS is to be hosted by NRF, and it is likely that the NICIS Board would end up being an advisory body to NRF. The Board that will be described in Section 5.3 is intended to have a stronger authority to drive the development of national CI. Thus, for an NRF-hosted model the committee recommends a serious exploration of the possible high-level governance structures.

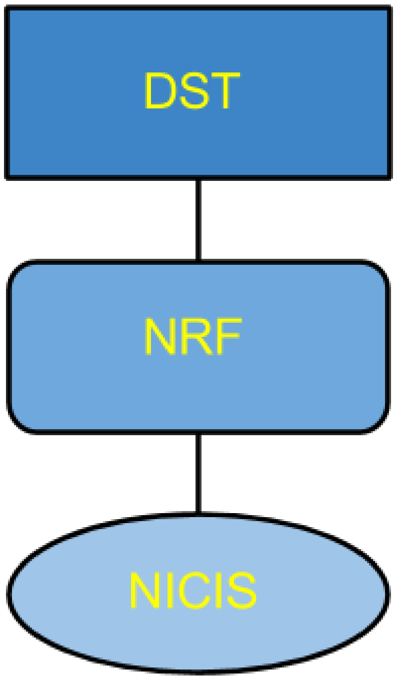


Figure 4.3 NICIS as a National Facility under NRF.

4.2.4 A separate legal unit owned by DST

Description

Under this model, NICIS will be organised as an appropriate legal entity owned by DST. Our main inspiration for this is CSC – The Finnish IT Centre for Science, which is Finland's national cyberinfrastructure organisation. CSC has proved highly successful in providing CI services to Finnish scientists. CSC is also a major player in international CI projects, which is remarkable for a country with population 5 million. The governance structure of NICIS at the highest level should be matched by the appropriate form of company structure, to be decided by DST. If, for example, NICIS is organised as a government-owned limited liability company, DST would be the shareholder and execute through the "Annual General Meeting of Shareholders", while the NICIS Board to be described below would constitute the "Board of Directors". The structure proposed below is tentative and its details will have to be reviewed in dialogue with DST officials.

The role of DST is to execute and further develop its professional role as owner of NICIS. DST will decide the vision and overall aims of NICIS and decide on the funding and budgetary aims for NICIS (for example not-for-profit). DST will adapt the funding for NICIS to the vision and current situation of NICIS.

DST will approve of the annual report and financial statement from the Board. DST appoints the members of the NICIS Board through some agreed nomination process. The composition of the NICIS Board will be characterised by competence, capacity and diversity; based on the specific need of NICIS high-level stakeholders. DST decides on the remuneration of the Board.

Pros

Referring to the main principles set out in the beginning of this chapter, it is seen that:

- The model will enable joint planning and budgeting according to a consolidated national strategy.
- The independent role of the organisation will provide for good governance with minimal conflicts of interest. The high-level organisational structure is transparent with short reporting lines, paving the ground for an accountable, efficient and effective organisation.
- The model will give high visibility of the CI services through a one-stop-shop concept .
- The model will be highly sustainable due to the close link to funding at the government level and the establishment as a separate legal entity. This will again provide a stable hub for skills and human capital development which is essential for South Africa and of particular relevance for the new fields of data-driven research.
- The model will enable constructive stakeholder engagement at the strategy level. This can effectively and dynamically be managed by DST through selection and replacement of representatives to the NICIS Board.

Cons

The separate organisation will come at some additional administrative costs for separate legal, HR and communication services, etc., though these should be weighed up against existing cost charges. Furthermore, the establishing of the organisation would require a lengthy legislative process and an intermediate organisational model might be required.

Assessment

The Committee finds that this model will give the most powerful tool for DST to develop the national leadership for CI in the future. The model resembles the Finnish CSC organisation, which has proved highly successful in the European CI ecosystem. CSC efficiently provides a broad range of high quality tools for academia, it is a strong international player, and is characterised by close ties with the government level.

Thus, the Committee believes that the unique enabling role of CI will be best served by a separate organisation directly under DST[10]. This will ensure the proper level of attention and

[10] The Committee is aware of the Ministerial NSI Report (2012), where it is recommended to constitute a National Advisory Panel on Cyberinfrastructure, reporting to a proposed National Council for Research and Innovation. The Committee notes that there is a potential overlap in the mandates

ownership from DST in the long term and discharge the potential of CI as a key enabler for modern science, innovation and human capital development across all sectors of society.

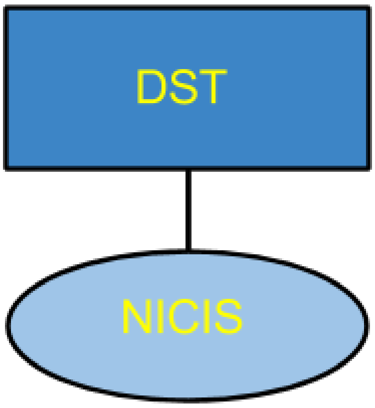


Figure 4.4 NICIS as a separate legal unit owned by DST.

4.2.5 Transitional considerations

During the committee's consultations with DST, implementation issues and possible intermediate solutions have been discussed. We will briefly comment on this in the following, noting that implementation of NICIS is not within the remit of the committee.

The establishment of a new legal entity is a lengthy process, and there might in this case be necessary to consider a transitional set-up. It is possible to look at the NRF-hosted model as an intermediate step towards a separate legal entity. The committee's assessment is, again, that this is a viable alternative to the ideal model.

Another interim solution could be to establish NICIS as a "fifth" programme under DST. Such a program may be regarded as a division of the DST, with a programme officer (interim NICIS CEO) reporting to the Deputy Director General of DST. Such a programme could be set up with a Steering Committee and it could be spun off as a separate legal entity at a later stage.

A major concern of the committee is that intermediate solutions oftentimes grow into semi-permanent solutions. Thus, we are reluctant to recommend a particular intermediate set-up, and rather emphasise the need to decide on a sustainable model that can firmly position South Africa to benefit from the technological advancements to be expected in the 21st century.

When DST decides, after consultation with key stakeholders, on the path it wishes to take, significant efforts will be required by individuals who are motivated to expeditiously

---

of the NICIS Board and a National Advisory Panel on Cyberinfrastructure which should be resolved if the latter panel should be created.

10

implement the plan. There is a tension between starting as soon as possible, and ensuring that the right people are hired and setting up good governance procedures. Therefore it may be appropriate for DST to appoint on an interim basis a suitable implementation leader and team with the necessary technical expertise and leadership ability and time to dedicate to this task without contention with other demands.

4.3 Governance, operation and management structures

Irrespective of high level organisation or ownership, the Committee proposes an organisation that is horizontally integrated across CI elements. Vertical integration, from national systems to institutional CI is not treated here in detail as it crosses several ministerial borders and will have to be resolved through agreements at the ministry level.

Recommendation 6: *We propose an organisation that is horizontally integrated across CI elements.*

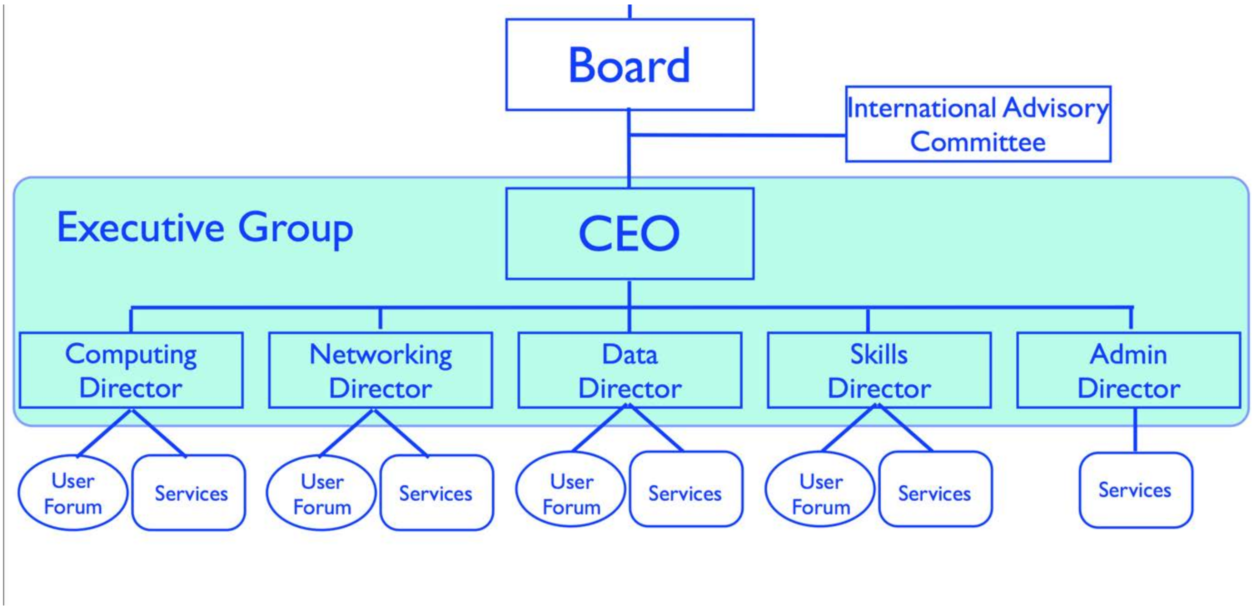


Figure 4.5 NICIS Governance model.

4.3.1 Governance

NICIS Board

The NICIS Board will maintain an independent control function over the organisation on behalf of its owners. The NICIS Board prepares and maintains a strategic plan for NICIS. The Board supervises the implementation of the strategic plan and oversees that the organisation is managed accordingly. The Board will prepare and maintain a plan for its own work that includes board-level competence building. The NICIS Board decides on the salary paid to the key managers, including the CEO and on the salary policy of NICIS employees.

The NICIS Board

- Shapes and monitors the company's framework for accountability, control and risk,
- Monitors key stakeholder relationships associated with generating value,
- Approves the budget, the financial statements and the report of the Executive Group,
- Controls and monitors the current management,
- Appoints and dismisses the CEO of the company.

The Board should be appointed from representatives of the main stakeholder groups, cf. Section 4.5. For effectiveness, it is recommended that the number of Board members is limited to approximately 10 persons. The Committee emphasises that whichever organisational model is adopted (i.e., independent agency, NRF sub-agency, Meraka sub-agency) the NICIS Board must be a *decision-making* board and not an advisory board. Problems of accountability and layers of governance will be exacerbated if this Board is seen to be a talking shop.

4.3.2 Management structure and operations

NICIS CEO

The Board appoints the CEO for the company. The CEO's terms of office are defined in a written contract approved by the Board. The CEO is responsible for the implementation of the strategy. S/he is responsible for the general execution of all NICIS activities and for the successful daily operation of NICIS. The CEO is also responsible for the exploitation of new service areas and collaboration opportunities. The CEO will interact with the NICIS Board and engage with stakeholders.

Services Areas

The activity of NICIS will be organised in Service Areas. The Service Areas will at the start-out include Networking, Computing, Data, Skills & Training and Administration. The Service Areas may change over time and they are described in detail in Section 5 below. Here, a *service* denotes everything from core services such as provision of raw bandwidth, compute cycles and data storage; to value-added services such as advanced user support, data curation, cloud services or training activities. The Service Areas largely follow the current elements of the South African CI, with the addition of a new area for Skills & Training and a common Administrative Service. The Skills and Training Service Area will coordinate the cross-cutting theme of human capital development in strong interaction with the higher education sector. It will be implemented in a highly distributed fashion, where the bulk of activities will be based at the HEI, while NICIS provides the glue to ensure a coherent approach to training and education in cyberinfrastructure.

User Engagement

Each Service Area will engage the user community through an appropriate User Forum. The User Forum will be composed of leading edge users and representatives from major user communities that can provide guidance for the development and deployment of CI services. It is important that the range of user representatives also includes possible future users and users that might be critical to the current service portfolio. The User Forum will advise on plans and progress, and identify issues pertinent to the success of the Service Area. The

User Forum will be chaired by a user representative. The Chairs of the User Fora meet annually with the NICIS Board.

Cross-Cutting Themes

The integrated NICIS framework allows for a systematic and efficient approach to cross-cutting themes. Such cross-cutting themes include skills and human capital development, where the integrated approach enables a coherent take on competence building for advanced use of ICT. It will also provide a better environment for building career paths for CI personnel.

The development of integrated CI is another area where an integrated framework will provide an optimal platform for development and deployment of integrated services such as the staging of large data sets for HPC systems, bandwidth-on-demand applications for virtual organisations; and integrating national and local services with commercial cloud services.

An integrated national CI with a long-term vision will provide a stable one-stop shop for the broad portfolio of advanced ICT services for users in academe and industry. This will be essential in enabling these users to benefit fully from the opportunities for value creation offered by digital science. Access to a "guaranteed" national CI will reduce the cost of application for industry; enable scalability as opposed to home-grown solutions, give access to the national skills pool and reduce risk by introducing a shift from capital expenses to operational expenses. These advantages will be particularly relevant for SMEs, who are the most vulnerable to risks of relying on in-house infrastructure, software and expertise.

Innovation project

The main focus of NICIS is to deliver high quality services to the users. In order to keep up with the technological development and deliver leading edge services, NICIS should also run an innovation project which focuses on developing new services, new capabilities, engaging with the commercial sector[11]. This should be organised in a flexible manner, pragmatically cutting across e-Infrastructure elements, involving dynamic consortia with participation from relevant actors inside and outside the organisation. The innovation project should have a separate funding stream organised in 3-4 year project cycles. Such an innovation project could for example embrace themes that support the present DIRISA activities, as proposed by the DIRISA SWG. Examples include innovative technology developments in global research infrastructures; the alignment of national and institutional policy on the provision of efficient and cost effective technology for research data infrastructures; best practice in data management and Open Science mechanisms of knowledge sharing. Inspiration for setting up a national innovation project can be found from the Netherlands' GigaPort program, see A.2.4. The Committee draws a distinction between innovation projects and *research*. Although, the innovation projects require staff that is capable of high-level research, the goals of research and innovation are different; this topic is explored further in section 4.4.4.

---

[11] A similar innovation project within the networking area is run by SURFnet, the Dutch NREN. The GigaPort3 program aims to take the SURFnet network infrastructure to a higher level through piloting new services. More information is available online: http://www.surfnet.nl/en/Innovatieprogramma's/gigaport3/.

International activity

Science is a global enterprise. In order for South-African researchers to be able to collaborate globally the NICIS must be interoperable with the important cyberinfrastructures outside SA. Participation in international *fora* is essential for such relevance. Participation in international CI initiatives is also important in ensuring high quality of the NICIS enterprise. It is recommended that NICIS aims to be an attractive partner for collaboration in international and bilateral CI initiatives. Participation in international initiatives, such as the framework programmes of the European Union, will be an important quality mechanism for NICIS. Furthermore, a measure of success for NICIS would be to become the preferred partner for SKA in delivering CI services. Other quality mechanisms should include regular reviews by the International Advisory Committee.

Directors

Each Service Area will be coordinated by a Director. Each Director is responsible for the development and delivery of services within his service area. This responsibility includes proposing and executing technical projects for development, deployment and operation of services.

Each Director will engage with users and other stakeholders, perform requirements gathering, and exploit future opportunities for innovative projects within his service area. Users will be engaged in the technical development and they will be involved in requirements gathering. Within each Service Area each Director coordinates user engagement through appropriate user forums. The Director is responsible for stimulating the organisation of new user communities.

Executive Group

The CEO chairs the Executive Group. The Executive Group is composed of the Directors, the Administration Director and the CEO. The Executive Group has the primary responsibility for management of NICIS. The Executive Group implements the strategy, engage with stakeholders, exploit and review opportunities, discuss and develop activity plans, monitor quality and progress of projects. The Executive Group will have frequent (i.e. weekly) meetings to monitor progress, prepare quarterly and annual reports, to resolve occurring issues and to plan the implementation of future activities.

4.4 Funding models

4.4.1 Overall funding of NICIS

Cyberinfrastructure — both equipment and people — is extremely expensive, and it is often difficult to quantify direct payoff of the investment. DST has already contributed substantial funding to the CI, both HPC and networking.

Based upon international experience, our view is that the NICIS will not be sustainable without continued significant support from DST, so that funding of NICIS will continue to consume a significant portion of the national science budget. The Committee suggests some measures below to ameliorate this load, but strongly urge DST to remain committed to the NICIS.

Money must be spent effectively. There are two strong recommendations that the Committee proposes:

Recommendation 7: *Budgeting must be integrated across the NICIS. While ring-fenced funding for the system as a whole is a good idea, there needs to be joint planning between the different pillars of the NICIS. This planning should be aligned to the medium-term expenditure framework (MTEF).*

Recommendation 8: *There must be a high-level of accountability of the sectors to the stakeholders (DST, universities, etc.).*

In this respect, the different sectors have a responsibility to show that the money that is invested in them is well spent. This is important to justify the long-term funding that will be required for the sector, and for planning. Clear and transparent metrics to measure performance must be set. With some exceptions, the Committee believes that current performance metrics and reporting are insufficient, which has complicated the writing of this report.

One of DST's challenges in funding the current networking, compute and data programs is the unevenness of the expenditures to create and then refresh every 3-5 years the substantial capital equipment required for these CI elements. By creating NICIS as a single program that encompasses multiple elements within a single budget planning process, there is an opportunity to synchronize the phases of capital refresh across various components (e.g. networking, computing and data) and thus level out the spikes in capital expenditures. Another mechanism to smooth out capital expenditures is to lease rather than purchase equipment. Our benchmarking analysis (Sec 2.2) indicates that most countries/agencies avoid leasing and find a way to purchase equipment rather than incur the additional interest costs and complexities of leasing equipment. Finally, the NICIS multi-year budget plan could request level funding for capital refresh, even if the planned capital expenditures are not level - e.g. save funds some years for planned higher expenditures in later years. One way or another, it is important that NICIS and DST work with the Treasury to predictably provide the funds necessary for the capital refresh of equipment on the 3-5 year timescales typical of CI equipment.

4.4.2 Other sources of funding

Other state entities. As there are significant benefits of scale, DST could explore shared use of central NICIS facilities by other state entities (e.g., the South African Weather Services). Even if shared computing systems are not feasible, co-location of equipment and use of other infrastructure can be explored. For example, the cost of backup power and air-conditioning is a very large part of the cost of any system, even if amortised over a ten-year period. Sharing would allow costs to be brought down and managed.

Commercial income. As discussed elsewhere in the report, limited commercial use of HPC facilities and other computing services should be encouraged. However, it is unlikely that this can take place at the level that a significant amount of income could be brought in without impacting on the key priorities of the NICIS. Moreover, this usage has the potential of discouraging the development of an indigenous high-tech IT sector. Thus, while limited industrial use of computing services is to be welcomed this must be managed.

Similar remarks apply to the other service areas of the NICIS, though in the case of networking services, this is complicated by licensing and/or contractual agreements that limit the type of traffic that can be carried across the SANREN.

User fees. Currently, the only significant user fees are those paid by TENET for international bandwidth. This is a very significant amount of money that comes into the system and without which the national research and education network would not be sustainable. All parties appear to agree that this will continue. These user fees are useful not only because they bring in significant resources, but act as a point of accountability for the universities in their usage of the facilities and for the (present day) SANReN as a provider.

The idea of user fees for other service areas of the NICIS is in principle attractive. However, the Committee cannot see practical methods for implementing this that will bring in more money than the cost of implementing. In the case of the network, there are only a small number of large users of the networking services. In the case of computing and data services, there are many, relatively small users. The flow of funding is more complex, and the nature of the relationship between the customers and service providers is different. Thus, the Committee does not recommend user fees as a way of bringing in funds to the system for the other service areas of NICIS. However, the Committee emphasises the importance of proper metrics for performance so that objective criteria for evaluating both the users (in terms of their resource usage) and the service providers (in terms of their effectiveness and efficiency).

4.4.3 Research funding

A component of the existing DST-funding is channeled to research, either within SANReN and the CHPC or to the universities.

As discussed in Section 4.3.2, the Committee sees a role for innovation projects within the NICIS, to be carried out by dynamic consortia involving NICIS employees and external partners. Such innovation activities could explore better ways in which services can be provided, or to explore new technologies, or to provide better support to users. Note that the goal of this innovation activity is fundamentally different to the type of research that universities and research councils might conduct.

The Committee does not see an extensive role for in-house research within the service areas of the NICIS. It would place the NICIS in a conflict of interest position, and may have the effect of weakening existing research programmes within universities. Clearly, there are activities that lie in a grey-area between research and innovation project, but nevertheless an in principle distinction can be made.

Currently, the CHPC provides significant funding for researchers, either through the allocation of processor cycles or through direct funding of projects (students, postdocs, travel). The first component of this is non-controversial: this is exactly what the CHPC should be doing. The second component is less universally accepted.

The virtue of funding research through the CHPC now (and NICIS in the future) is that it gives ring-fenced funding to research in this area and a way for DST to influence the direction of national research strategies. However, separating funding for HPC-related research from other funding instruments might cause lack of cohesion and send the wrong message. For example, how would a sociologist wishing to conduct an extensive demographic study and computational analysis apply for funding. Only through NICIS? Only through the NRF? Two applications? What if one was funded and the other not? Second, high-quality, independent review of applications is essential. The NRF already has a system in place. Will NICIS duplicate a review system? What are the costs of this?

Recommendation 9: *Much tighter integration between allocations of resources from NICIS to researchers within existing research funding instruments is needed. NICIS should not have an extensive in-house research role. However, CI-related research should be strongly supported within the research communities outside the scope of NICIS.*

As discussed in Section 4.2.3, this may be an argument for NICIS being placed under the NRF umbrella.

Allocation of networking and data resources. Just as there is a need for scientists to get access to HPC resources, there is a need for access for some projects to bandwidth above what an individual university could provide (here we think not so much of SKA, which is *sui generis*, but of experiments like ATLAS). Our understanding is that currently this is handled in an *ad hoc* way, but as the NICIS matures, there will be a need to regularise the allocation process by using rigorous scientific assessment within a clear, transparent system. The same principles will apply to the allocation of data services to users.

Recommendation 10: *A common framework for the allocation of cyber-resources across all NICIS service areas (computing, networking, data) should be adopted, using merit-based scientific assessment.*

4.4.4 Cost implications of our recommendations

The Committee is aware that there are cost implications to our proposals.

The director of NICIS would be a new post and that person would require at least a PA to support them, and our preferred model would require HR functions being developed inside NICIS. These new costs may be offset by the freeing of some of the resources that are currently paid to Meraka as overhead[12].

As discussed below in Sections 5.3 and 5.4, additional funding is required to expand the current DIRISA program and to establish a new Skills and Training effort. The latter will require a modest level of extra funding, though much of its initial efforts will be to coordinate efforts already funded under existing components of the national CI.

Recommendation 11: *New funding is necessary for NICIS management, expanded Data Services (DIRISA) and a new Skills and Training Services area.*

The proposals made by the Committee will need to be properly costed and evaluated. The Committee is of the view that the new structures and improved governance model will lead to a more effective and accountable CI.

[12] Presumably, some of the charges for renting space would remain.

4.5 Stakeholders

Referring to the main principles set out in Section 4.1.1, NICIS will focus on constructive and systematic stakeholder engagement. Through stakeholder engagement NICIS can better identify new opportunities manage risk and build reputation; thereby increasing value creation for both partners and society at large. A broad range of stakeholders within the South African National System of Innovation that are affected by the NICIS have been identified and listed in Table 4.1.

All stakeholders contribute, directly or indirectly, to NICIS's capacity for value creation. Stakeholders can be potential beneficiaries or risk bearers of NICIS activity. By maintaining and improving relationships with stakeholder groups, NICIS will obtain a better collaborative environment for its operations. This may lead to innovation, as well as development and deployment of relevant services that address stakeholder needs. A strong focus on stakeholder engagement may also lead to long-term cost reduction and increased value creation for NICIS.

NICIS therefore needs to engage with a broad range of stakeholders. The ways of engagement will range from direct representation in the Board or User Fora to regularised meetings, workshops and social media. The stakeholders may be grouped as depicted in the graph below and the stakeholders that add value to NICIS in several ways, as indicated in Table 4.2.

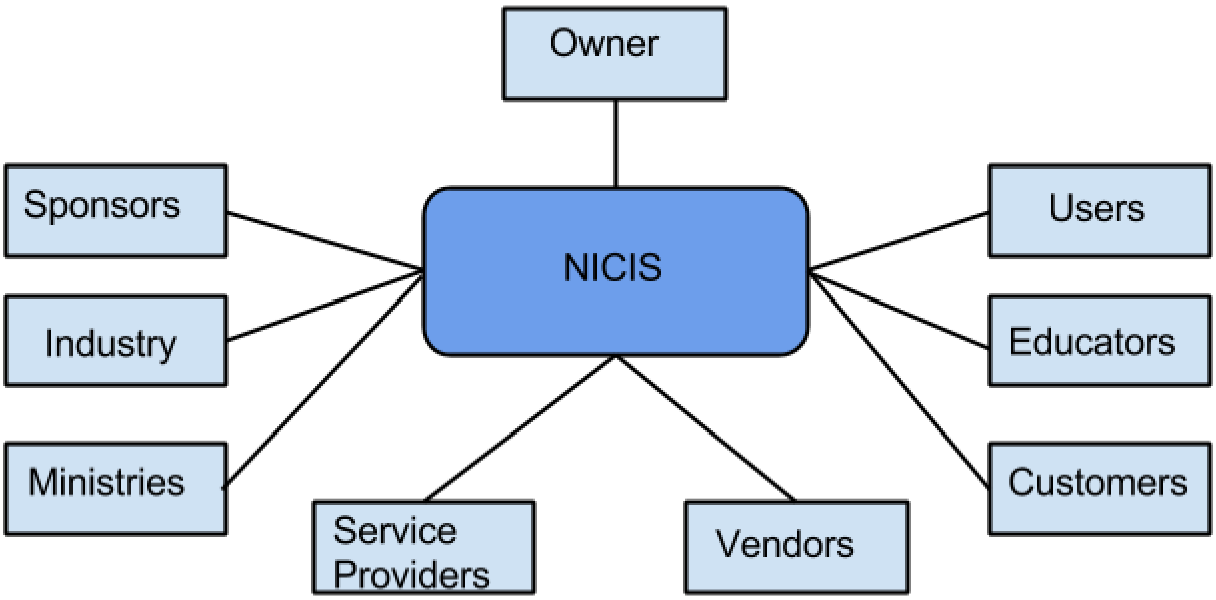


Figure 4.6. The main stakeholders of NICIS.

Table 4.1: Broad categorisation of Stakeholders within NICIS

Stakeholder Group	Stakeholders
Owners	Government (DST, DTI, DHET)
Sponsors	NRF, TIA
Employees	Staff members of Government, Industry and Universities

Users	Knowledge creation through products, services and processes
	• Research Communities
	• Research Infrastructures
	• Universities and Universities of Technology
	• Science Councils
	• Agencies:
	    • National Research Foundation (NRF)
	    • Technology Innovation Agency (TIA)
	    • South African National Space Agency
	    • Medical Research Council (MRC)

Supply chain	On-campus resources (for example: Technology Transfer Offices, Research Offices and IT Departments)
Joint venture partners and alliances	• ASAUDIT • Higher Education South Africa (HESA) • Human science Research Council (HSRC) • Council for Scientific and Industrial Research (CSIR) • Business (private sector and state-owned) • Industry
Educators	• University academic staff members
Local communities	• Universities (including Comprehensive Universities) • Universities of Technology • FET Colleges
Ministries	• Department of Science and Technology (DST) • Department of Higher Education and Training (DHET) • Department of Trade and Industry (DTI) • Department of Minerals (DoM) • Department of Energy (DoE) • Department of Agriculture (DoA) • Department of Health (DoH) • Department of Water Affairs and Forestry (DWAF) • Economic Development Department (EDD)
Commercial suppliers and	• Business and Industry

service providers	
Special interests groups	• National Science and Technology Forum (NSTF) • National Advisory Council on Innovation (NACI) • HESA and HESA (RISG) • DVC Research Forum • Southern African Research and Innovation Management Association (SARIMA) • Association of SA University Directors of Information technology (ASAUDIT)

5. Proposed model for each of the cyberinfrastructure service areas

5.1 Networking Services

First-class networking services are fundamental for the national e-infrastructure both strategically and operationally. The provision of inexpensive, high-quality bandwidth to the sector is a necessary building block for the entire sector. This need extends to the provision of higher-level networking services and high levels of innovation within the sector.

International experience has shown that state involvement in networking services is key. As important as private sector involvement is, commercial networking services are not aimed at providing cost effective solutions that meet research and high end industrial needs. The introduction of innovation is often highly risky for commercial enterprises, and there is sometimes little incentive for the private sector to respond directly to national priorities, some of which may require significant initial investment.

This need for state strategic involvement is particularly relevant in South Africa, which is still dealing with dysfunctionalities inherited from apartheid, geographical disparities, market failures in networking service provision post-1994, and which as an economy which requires strategic investment.

5.1.1 Need for and feasibility of a single NREN

A specific term of reference of this Committee is to investigate the need for and feasibility of a single NREN — a point emphasised several times during our discussions with sectors.

Our starting point in making recommendations is based on the Committee's general principles that joint planning in e-infrastructure is desirable and that appropriate, active involvement of all key stakeholders is important. This point is broadly supported by the recommendation of the NREN Sector Working Group.

Currently, there are two key organisations providing Networking Services in the sector — SANReN, funded by the DST, and TENET, funded by the universities. One concern is how the overlapping mandates of these organisations affects the sector and the Committee can identify two issues:

- How can strategic decisions and planning take place across the two organisations?
- Which organisation has the status of being "the" NREN? The question of who represents South Africa internationally has clear reputational and coordination impact. The same applies to a lesser extent nationally and questions of organisational and personal importance may complicate both strategic and operational functioning.

Role of TENET: TENET was established by the universities to provide cost-effective networking. It acts as a provider of networking services to the universities, but in relation to other players in the e-infrastructure community, it acts as a consumer of services, directly representing the user community. A narrow view would be to see TENET primarily as a buyers' club — by banding together members get better pricing than they could get individually. This view would see TENET as an extension of PURCO (a consortium established by the universities to reducing procurement costs generally). However, TENET has had to play a much more important role. Cost-effective networking is absolutely mission critical and market failures and strategic gaps posed significant risks. Without the establishing of TENET in 2000, there could have been very serious adverse impacts on the sector.

TENET has been very successful in its mission. TENET is directly accountable to the universities, and the universities are still mindful of a history when they were put at risk. Thus, the continued existence of TENET as a major player was stressed to the Committee by the university sector. The Committee is sympathetic to this experience. We must learn from history; but it is important that we plan for 2030 based on what happens now rather than 1995.

It is important to note that TENET's budget derives from contractual arrangements with the universities, out of the universities' budgets. Much of the university's needs are commodity needs that increasingly could easily be met without strategic involvement by an increasingly mature and competitive private sector. For example, provision of raw bandwidth to support email and commodity web browsing may not merit particular strategic involvement or state investment. An important driver of networking strategy must be actual needs of the universities. Each university has the autonomy to decide what proportion of its budget should be spent on networking, and this is desirable as one part of the planning mechanism.

For these reasons, TENET will continue to play an important separate role, though this may change in nature.

Role of SANReN. SANReN plays a strategic role within the sector. By using the very significant financial investment from DST, SANReN has been able to provide infrastructure, services, coordination and planning that (1) has reduced costs and improved services dramatically; (2) helped alleviate networking connectivity problems of institutions that are geographically remote. All parties who provided information to us recognised that SANReN's contributions have revolutionised what can be done.

Coordination between TENET and SANReN. On the whole, coordination between SANReN and TENET appears healthy at both an organisational and personal level. This was stressed to us repeatedly by representatives of both organisations, though different views were expressed by other individuals. As good as these links are, there are clear challenges.

- National planning is compromised. All parties need to be involved in planning, although this does not preclude individual universities or the sector as a whole from independent activities.
- Both TENET and SANReN have played part of the role of an NREN and should be congratulated for what they have done. But it is not healthy for both SANReN and TENET to be claiming to be the NREN (as can be found on their web sites).
- There have been examples where failures of coordination caused unnecessary conflict. On the other hand, it is recognised that organisational re-structuring might just move conflict from one interface to another.

Recommendation 12: *The Committee recommends that the Networking Services Area proposed above should be recognised as the NREN. It is important for DST to have significant involvement in the NREN both because it will be a primary funder and so will need to account for the spending of public funds. Moreover DST is in the position of having strategic oversight taking into account all national interests. Other key parties such as the universities will also have a stake in the process and be able to hold the service area to account.*

The proposed model of the Networking Services Area as the NREN, collaborating with TENET as the Service Provider of the Universities, could easily be generalised to accommodate the need of facilitating the access to the NREN backbone to FET Colleges, Schools, and Hospitals. These groupings of users could form their own "Internet Service Provider" and cooperate with NREN on a similar basis as TENET does at the moment.

5.1.2 Remit of the NREN

The main responsibilities of the NREN will be:

- The development of long-term networking services plans;
- Recommending policies and priorities for funding to the DST;
- Development and coordination of short and medium-term plans, policies and budgeting framework;
- Cyber-security of the CI is a critical function. The development of a CSIRT, appropriate processes and skills across the systems are crucial. Regular audits must be undertaken.
- Determination of value-added services and how they will be provided and funded.
- Management and responsibility of the work of the current SANReN;
- Coordination with TENET and major users of networking facilities;
- Coordination of international representation.

This new structure will primarily be responsible for coordinating international representation. However, this does *not* mean that it would have the monopoly on international representation. In many cases, it will be appropriate for other bodies — for example, TENET, HESA or a discipline-specific organisation such as the South African Institute of Physics — to represent South Africa in international forums. It will be incumbent on the NREN to involve these parties, as attempts to monopolise representation will be technically and socially undesirable and lead to breakdown in cohesion of the sector. However, the NREN would be a well publicised single point of call. This proposal is not meant to downplay the importance of TENET; indeed, without TENET's skills and active contribution, an NREN is impossible to visualise. TENET brings strategic and operational expertise, and, as agent of the universities and science councils, brings considerable financial resources and the needs of the key users into the system.

The roles of the Networking Services Area and its components as well as TENET should be redefined. Although the Committee can make some suggestions, it is probably more appropriate for those directly involved to get together and make joint recommendations. The current model where TENET acts as the operator of the SANReN network appears sensible. However, as the network landscape changes, this may change.

There needs to be clarity in the long-term funding model for the operating costs of the research network. DST must also make a decision about its long-term involvement in the network space. DST must continue to play a long-term role in strategic investments and decisions, but the operation of a commodity ISP is a different issue. DST's role will therefore evolve as the network infrastructure matures.

5.1.3 Connectivity Policy

The overall policies and activities of the NREN should be aimed at improving national levels of competitivity in both the public and private sector.

Primary criteria for access will be:

- need for access to these facilities — where more cost-effective alternatives exist for the need of the organisation, these should be strongly considered.
- mission of the organisation — broadly organisations within the state-funded sector (e.g., universities, museums, research councils) which have a public benefit or mission.

The fact that an organisation is not a profit-making institution should be an important factor, but is in itself neither a necessary nor sufficient condition for access. Before discussing the principles, it is important to note that there are significant contractual limitations on commercial services that the current SANReN can provide that are independent of the argument below.

Role of the private sector The innovation and improvements in infrastructure within the NREN should act as models for the private sector. By using the successes (and failures) of the NREN as learning experiences, risks in innovation should be reduced. A direct outcome of state investment will be improved skills and that should have a knock-on effect. The Committee does not see wide-scale use by the private sector as being desirable but there should be exceptions where there are public-private partnerships, and in well-defined areas where it can be justified. The appropriate cost-recovery model should be applied — in rare cases it may be desirable to provide full access without cost; in some cases recovery of marginal costs will be acceptable; in other cases full cost recovery will be required. This is a highly policy-laden question that is beyond our remit. The factors to consider are

- impact of access (either because of direct public benefit, or because of the impact that the innovation can make);
- possible inappropriate subsidy of private interest;
- possible stifling of local networking industry (provision of services to the private sector at preferential rates can stifle innovation by local networking companies who cannot compete in the short-term)

Local networking resources. One of the conditions for access to the national network must be ability to use it effectively. Although there was general agreement that the impact of SANReN and TENET has been enormous across the sector, we heard from more than one institution that access to the SANReN network has had no effect on the users because of configuration or lack of appropriate local equipment — although it does not appear to the Committee that lack of money was the primary problem and above all, it appears that a lack of skills at the institutions concerned is the fundamental problem. The NREN will have a key responsibility in overcoming these skills problems and working with each institution. The Committee will address this through the establishment of a Skills Service Area, described in Section 5.4. We recognise that all institutions do not start at the same level and there needs to be national input. However, each institution must display a real commitment to overcoming these problems. To this end, a set of externally verifiable metrics should be adopted that describe the performance that real users at each site experience so that it can be determined whether there is an acceptable return on investment.

5.1.4 Financing and strategic direction

The state has and should continue to be the primary funder for infrastructure that has been required to meet (1) high-performance networking needs; (2) strategic projects such as Meerkat; (3) address access for institutions outside of the metropolitan areas.

For operational costs, users should pay at least marginal cost recovery that gives appropriate incentive for good planning and usage. The current model that TENET uses for international bandwidth delivery is a good one for international usage. However, its applicability to local networking is less clear since there isn’t the same direct link between usage and costs. Although TENET is the operator of the physical infrastructure and pays running costs, including that of the staff required to run the network, there is still indirect subsidy by the DST as maintenance costs of the network are not covered by the universities.

The role of the state is likely to change as the networking infrastructure matures, and commercial offerings improve. The Committee envisages that planning would happen through a 3-5 year rolling plan drawn up by the networking services area after consultation with DST and TENET (the major funders) and other interest groups. The plan would address (a) strategic opportunities; (b) capital investments; (c) maintenance; (d) running costs; and (e) performance metrics and auditing.

5.2 Computing Services

South Africa has invested in Tier 1 computational assets for national and international research competitiveness in addressing the most complex science and engineering problems. This investment has been recognised internationally in the Supercomputer Top 500 rankings. The Computing Services area must continue to provide the facilities that are:

- compute-intensive, requiring the exploitation of massively parallel computation involving a very large number of processing elements;
- communication-intensive, requiring the high-speed transfer of data among processing elements; and
- data-intensive, involving the high-speed manipulation of very large quantities of data.

This Tier 1 capability must be used to advance national strategic goals of economic and societal security in addition to its utilisation by the academic community. Economic security derives from greater industrial innovation, productivity and job creation. Consequently, appropriate industrial access to national Tier 1 computing services must be a priority in order for the South African industrial base to be competitive in the global marketplace. Additionally, societal goals pertaining to quality of health, agricultural productivity, mitigation of natural disasters, anticipating climate change and others can be addressed through the use of Tier 1 resources by the appropriate government department and agencies.

Advanced computing services (HPC, grid, etc.) play a key role in addressing a number of national science and engineering research priorities (see Section 3.2) as well as socio-economic issues as they have in other countries. The continued, integrated development of HPC is crucial.

The Committee **strongly recommends** the following approach to maintaining South Africa's leadership-class computing systems and services for internationally competitive research activities.

5.2.1 Requirements for Sustainable Computing Services

The national computing services are currently provided through the Centre for High Performance Computing (CHPC), which would form the core of the Computing Services area. The CHPC is a nationally administered and shared HPC resource that offers cost savings in procurement and efficient allocation of high-end compute resources. Sustainability of the computing services rests in its ability to understand and meet user requirements and creating the business conditions for effective delivery to users. Meeting user requirements undergirds sustainability as outcomes such as a competitive research community, increased multidisciplinary collaborations that exploit computing services assets, and expanding adoption of advanced computing services are most likely to persuade DST and Treasury to continue significant funding of DST.

Useful metrics that can be used for planning and assessment of the new service area will include effective business processes, level of appropriate technical staffing (system administration, domain experts and computation experts), mix and range of equipment, utilisation of hardware and staff, research output and increase of user community. Harder to measure, but very important will be socio-economic impact.

The CHPC has grown its user base from 15 in 2007 to approximately 600 in 2013 and further growth can be expected. Good work has been done by the CHPC. It is likely that the flagship cluster of the CHPC will continue to be the most important single piece of equipment in the CI, and for the CI to be competitive and sustainable this must continue to be supported and upgraded. However, the CI eco-system will require other types of equipment and facilities.

Recommendation 13: *The Committee recommends that the CHPC in essentially its current form should take on the role of the Computing Services area, with some changes to its mandate as described below. More formal processes to assess user satisfaction and needs must be established with requisite metrics for performance.*

5.2.2 Identification of key functions

A suite of advanced computing services is needed to optimise impact of science and engineering research and education. This suite includes access to user support services on the national level coupled to regional compute assets, cloud computing, and advanced user support that enables productivity on High Performance Computing platforms.

User Desk Support, Advanced User Support and Consulting. Providing professional technical support for the effective use of national CI resources is very important. This activity includes (a) ensuring that researchers can access resources from their own location, (b) facilitating multidisciplinary research, (c) supporting data analysis, possibly including statistical analysis, (d) providing scientific application expertise, and (e) supporting existing and emerging technologies. The most basic service needed by the widest user community is an easily accessible centralised help desk with the objective to become a 24 hour function. However, advanced user support and consulting services are needed too. This is a real challenge because this often requires both advanced computing and domain expertise. Assistance with porting, parallelising and optimising code will be very important, especially for the the high-end applications on the high-end equipment. The use of tools such as advanced debuggers, concurrency checkers and performance tuners is essential and the provision and training in these tools will be an important function.

High Performance Computing Services – Providing national computational resources dependably as a production service. Providing this service with the right level of technology capabilities requires understanding the user community. CHPC has appropriately kept compute resources current and identified opportunities to deploy new types of resources (GPGPUs, accelerators, etc.).

Grid Computing Services — The SA Grid involves several institutions across the country. South Africa has high profile commitments to the ALICE and ATLAS projects which presently rely on the South African Grid (SA Grid). Two distinct visions of SA Grid were presented to the Committee:

1. SA Grid as a federation of autonomous computing units, with CHPC as the largest partner and some facilitation conducted nationally (this is the view of the NREN SWG).

2. SA Grid as essentially a subcomponent of the work that CHPC does with minimal inputs (this is the view of the HPC SWG).

These are distinct because in the former, while the CHPC provides very significant computing facilities and plays a leadership and coordination role, there is also substantial, autonomous computing available at the universities.

The Committee generally supports the first perspective as put forward by the NREN SWG. The advantage of this view is that it provides a mechanism for small and mid-range HPC facilities to be shared nationally, while allowing universities to control their facilities. It also provides access to international resources, both in services and training. However, perhaps the most important role is that it provides a way of coordinating and training technical staff at the universities and other research facilities. This can be a valuable tool in integrating across different tiers, and so is an important role of the SA Grid. SA Grid also provides mechanisms for linking with and between international cyberinfrastructures.

Whichever vision is most appropriate, the Committee observes that SA Grid currently appears to be fragile and over-reliant on one or two (dedicated) people. For SA Grid to be sustainable and organisationally stable, it needs modest additional technical resources to provide strategic, stable operational functioning and ongoing training. Also, there must be increased support from the research institutes that make up the current SA Grid; the current commitments are too *ad hoc* and informal for future growth. There is also scope for grid research projects that would make the grid more usable to people outside the physics community.

There is a reasonable debate whether SA Grid should fall under the Compute Services or Networking Services area of the NICIS. There are technical linkages to both areas; while the end deliverable is computing resources; the grid relies heavily on networking, security, authentication and grid certificates often aligned with networking services. Functionally the Committee sees a more natural synergy with the Computing Services area and hence it is included in this section. However, the networking sector has a keen interest in SA Grid and the key individuals most closely engaged with SA Grid are aligned with the networking sector; this commitment and expertise should not lost in the interest of optimising a paper organisational chart. (In either case, training could be coordinated through the Skills & Training Services Area.)

Recommendation 14: *The Committee recommends that SA Grid be continued with modest additional support to increase robustness, and recommends that DST consult with stakeholders to determine whether to assign it to Networking Services or Compute Services.*

Visualisation Support – An important component of data analysis is visualisation. As the volume of data produced in research continues to grow at rapid rates, using visualisation to analyse that data continues to grow in importance. Visualisation resources range from workstation tools to dedicated visualisation clusters with graphic accelerators to specialised installations that support three-dimensional immersive graphics at extremely high resolutions.

Software Development – The development of software tools, libraries, and techniques to improve the usability of national cyberinfrastructure resources is based on mission and subject to funding. This typically involves research and development efforts that focus on the latest, often leading-edge, CI resources in order to ensure optimal utilisation by researchers. In-house software development can be a mission critical service

Data Storage Services Providing data storage services for national users are critical to the current CHPC's mission both in support of computing per se and its VLDB role. The exploding volume of data being produced by scientific instruments, distributed data sensors and computer simulations make this a growing challenge — a challenge of the current CHPC as it develops data storage strategies and implementations, that will impact how datasets are created, formats that are used, metadata solutions, methods for tracking provenance, and, in some cases, long-term curation. The Committee recommends that CHPC continue its storage role in support of its computational users, particularly for high-performance short-term storage, while the long-term persistent storage (e.g. VLDB) and related data services be transitioned to the new Data Services Area (see Section 5.3). (As mentioned below, CHPC is a logical organisation to continue to operate the VLDB, but this would be under the oversight and funding of the Data Services area.)

Cybersecurity National and regional computing networks must be designed to support cyber-security while also supporting the performance and robustness needed by CI. The goal is to safeguard science and technology information of national computing facilities and resources by:

- establishment and operation of security infrastructure
- research and development of security technology
- sharing of information and supporting services pertaining to cybersecurity.

Key activities in this area are event monitoring (anomalous events), analysis of latest trends for information security, researching security and monitoring technologies, developing detection patterns for monitoring, planning security projects, sharing information with regional and university computing resources, and consulting for risk management. On a periodic basis, national and regional computing assets should undergo external security audits.

Cloud Computing Services. Cloud Computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of advanced computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned, configured and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics. These five essential characteristics are on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. The global research community is exploring cloud computing as a cost-effective and elastic environment for productivity.

As discussed in section 3.2.6, virtualisation and cloud computing have become a key computing technology (Armbrust et al., 2010). The technology was not originally designed for high performance computing; a typical application might be a highly redundant web service. However, flexibility, elasticity and scalability are very attractive in many scientific applications and many case studies have been conducted — a sample of recent studies includes (Berriman et al 2013; Juve et al 2013; Nagasaki et al. 2013; Saini et al 2012).

In summary, these case studies show that despite technical and usability limitations, there are many scientific computations that can benefit from cloud computing. At least for the foreseeable future, cloud computing cannot support top end tightly-coupled applications due to communication and virtualisation overhead, and so the traditional role of a supercomputing centre must continue to be supported. However, there are many applications that can benefit in a cost-effective way from cloud computing, and the flexible environments that they provide.

We point to two important international activities. In the United States, the Department of Energy has undertaken an extensive study of the use of cloud computing for scientific computing, examining the use of private and public clouds (Yelick et al 2011). This study pointed out many of the problems of cloud computing. However, it concludes that cloud computing has a very important role to play in the repertoire of computing technologies that the Department of Energy provides and makes very useful recommendations, many of which are likely to be of direct relevance to South Africa. These emphasise the benefits of flexibility and availability, balanced with the challenges of cloud computing.

Australia's National Computational Infrastructure has launched a federated Australian Research Cloud (http://www.nectar.org.au/research-cloud, Kennedy 2013). This cloud supports data storage, collaboration and high performance computing.

Although we engaged with the working groups on the question of cloud computing, none of them addressed cloud computing in their visions. However, we believe that cloud computing is an important way of making HPC available more widely in the research community. This will not supplant the traditional HPC model, but will complement these services.

Interaction with the commercial sector is important too. There is a significant local niche industry in cloud computing and virtual hosting. Improving local skills is important. Second, public, commercial clouds have already been used by South African researchers and there may be certain applications which can be met in a cost-effective way by the private sector.

In summary, we see the medium-term mandate of the Computing Services Area with respect to cloud computing as the investigation and provision of the development of a private clouds for research, provision of training in virtualisation and cloud computing for universities, investigation of the role the private sector partners in cloud computing.

Recommendation 15: *The Committee strongly recommends that cloud computing be a key function of the Compute Services area.*

Education and Training – Providing various levels of education and training is critical for user productivity and adoption. This is extremely important for research outcomes and essential in developing a workforce equipped to compete globally. Education and training involves helping researchers and students learn computational and data analysis skills, including programming, parallel programming, use of computational resources (local and remote), numerical analysis, algorithm development, debugging, and performance optimisation. The Committee recommends that these services in the future should be coordinated under the new Skills and Training Services Area.

5.2.3 Establishment of an accredited national distributed computing system

South Africa science and engineering research and higher education communities require ready access to needed advanced computational and data-intensive compute capabilities.

The Department of Science and Technology seeks to position and support the entire spectrum of DST-funded research communities across the range of advanced computing technologies, hardware, and software. It also aims to promote a more complementary, comprehensive, and balanced portfolio of advanced computing infrastructure and programs for research and education to support multidisciplinary computational and data-enabled science and engineering that support the entire scientific, engineering, and education community.

Development of advanced computing capabilities appropriate for a range of research needs for the broader science and engineering research community are required. These are facilities that all researchers can use and support staff who would be trained and available for consultation, as well as strategic investments in domain-specific advanced computing systems.

A carefully planned ecosystem of high-end computing services (Tier 1), where there is a coordinated growth of institutional-based Tier 2 compute resources, and mid-range Tier 2 computational facilities that are linked with the national high-end compute resources will well serve the research and education communities. The importance of this ecosystem allows for a reduction in the effort to migrate applications to high end compute and thus accelerate the realisation of capability computing.

5.2.4 Framework for funding institutional mid-ranged facilities

It *may* be appropriate for national mid-range facilities to be established nationally (e.g. at the CHPC) and this would be funded through the national, integrated planning. The Committee emphasises its very strong recommendation that in such cases these facilities be co-hosted with existing Tier 1 facilities to capitalise on existing investment in staff, security, backup, power and air-conditioning.

Institutional mid-range facilities (either owned by a single institution or a consortium) are to be encouraged. Our preference would be for these to be planned bottom-up because they reflect real interest and need, as shown by investment of the host institution. Some incentive, especially where the equipment is shared, would be beneficial. Existing instruments such as the NRF's National Equipment Programme or infrastructure grants from the DHET may be the appropriate mechanisms. Collaboration and coordination between DHET and DST are essential in prioritising mid-range facilities, connection to national leadership-class compute capabilities, human resource use, and benefit of procurement.

Currently most South African researchers do not have access to these mid-range facilities. A few institutions have mainly departmental based machines acquired through the grants of Principal Investigators (PIs), which in most cases are widely accessed by the other departments. Recently, some institutions, such as NWU, UFS, UNISA, UP etc... have invested in the mid-range facility managed by the local IT division and shared amongst departments. This model of institutional facilities is more sustainable as it reduces the dependency of HPC systems on the drive of an individual researcher and the passion of current students, which create problems as students graduate or professors move from the institution. In this model, the skills and support are budgeted for in the institution IT infrastructure and thus provides for continuity and promotes equal access by researchers.

The HPC and Skills & Training Services Areas should play a crucial role of advocacy and hands-on training. Capacity development is crucial as is engagement with computing departments throughout the country.

Universities need to feed into the planning process.

Recommendation 16: *Universities should be asked to respond to this report (either through the HESA or the DVC Research forum) and draw up individual and collective strategies for provision of Tier 2 (and 3) facilities. Collaboration and coordination between DHET and DST are essential in prioritising mid-range facilities, connection to national leadership-class compute capabilities, human resource use, and benefit of procurement.*

5.3 Data Services

The emergence of data, from search engines to social networks to the scientific enterprise, is well recognised as a crucial element of modern information technology.There is an explosion in the volume of scientific data produced, from computational simulations, from large-scale instruments such as CERN's LHC or the SKA, and from vast numbers of distributed sensors and instruments such as gene sequencers. But "big data" is not just about storing petabytes of data. Data movement, access and management are critical. New training and skills are demanded in data mining and analysis to use data effectively. And data sharing, with associated requirements for metadata, provenance, access policies, search/discovery and curation, are increasingly important to funding agencies wanting to facilitate new scientific discovery and to maximise their investments. These trends will require cultural shifts for researchers, consistent policies at the national level with appropriate incentives, new tools and capabilities for sharing data effectively, and training and curricula for 'data scientists'. These trends also call for a more integrated approach to the organisational aspect, where all CI elements can be developed within a similar context. Numerous reports have been published highlighting the importance of data in research and identifying the paths forward for funding agencies, universities and researchers (e.g. Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2010; Hey, Tansley and Krist, 2009). Science funding organisations in many countries are developing policies for the management, preservation and sharing of data produced by researchers funded by them — e.g. US NSF Data Management Plans, NIH Data Sharing Plans, and broader OSTP policies (Holdren, 2013). It is vital that South Africa stay at the forefront with not only infrastructure for storing and accessing data, but also with aggressive policy development regarding data use and sharing, and a broad set of expertise,tools and training for supporting data science.

5.3.1 Background and Context

DST was ahead of its time in the international community when it recognised the importance of data as a third element of its cyberinfrastructure investments and created the Very Large Data Base (VLDB) program in ~2011. Later the charter of the program was expanded beyond the VLDB storage facility and renamed the Data Intensive Research Initiative of South Africa (DIRISA), with additional efforts in eScience support, metadata, data curation, publication, and digital repositories.

DST launched the VLDB program by providing ~R17.5M additional funding to the CHPC to procure, deploy and operate a 2 PB storage facility, with replication between CHPC's Cape Town facility and a datacenter in Pretoria (1 PB each location), and accessible at 10 Gbps via SANReN. This facility went into production in 2012 and is utilised as a digital repository by a number of researchers. Currently ~60% of VLDB's usage is by research communities storing simulation and observational data in bioinformatics, astronomy (Meerkat), climate modelling (SAEON) and heart disease, ~40% of usage is by CHPC computational users.

As discussed above, storing data is but one tangible element of the infrastructure, tools and expertise necessary for the 'data pillar' of a national CI. The expansion of the charter from the VLDB storage infrastructure to the broader DIRISA program was an excellent idea, but additional funding was not provided to support this expanded charter. The current DIRISA efforts beyond VLDB represent a "coalition of the willing," a forum of interested experts represented to this Committee by members of the Data Sector Working Group. Furthermore there appears to be relatively weak coupling between CHPC's operation of the VLDB storage facility and this forum of data experts. These factors result in a confused identity for DIRISA - is it a storage facility, a forum of experts or both? And if it is both, how do these two components work together and how is the funding utilised?Just as an example, CHPC staff are responsible for the allocation and use of the storage facility, with relatively little interaction between that process and the forum of experts involved in data policies and sharing.

5.3.2 Remit, nature and function of Data Services

Recommendation 17: *The Committee recommends that the NICIS Data Services Area be the leading organisation within South Africa to advocate for and implement data initiatives across the research community.*

The remit for Data Services would include:

- provide a robust foundation to enable data science in South Africa in the 21st century;
- develop an integrated set of data services across the data lifecycle, from storage facilities to data management to metadata tools/processes to curation and preservation practices;
- operate national storage facilities (possibly out-sourced to an organization such as CHPC);
- develop and implement a well-defined open allocations process for storage and other resources that supports the national research community;
- work with funding agencies to formulate and implement data policies regarding sharing and access to research data;
- coordinate a national dialog regarding scientific journal licensing processes, a potential national online publishing facility, and open access policies;
- in coordination with the Skills Service Area (see Section 5.4) develop training for the human expertise necessary for data scientists, including data management, sharing and analysis; and
- actively participate in international forums related to Data Services to promote South African activities and gain knowledge from other international efforts.

The current DIRISA program should be transformed to the more vigorous, expanded Data Services Area proposed in this report. By necessity, there will be both near-term and longer-term goals for this transformation. Long-term a substantial new investment would be required to achieve the ambitious goals in the expanded remit recommended by this Committee.

Recommendation 18: *The Committee recommends that NICIS works with the community to develop an ambitious proposal on data services to DST, and formulates convincing arguments to the Treasury for the return on that investment, predicated on economic competitiveness, human resource development, and industrial benefit.*

The Committee notes that SKA/CHPC have recently put forth a “Big Data Africa” proposal to DST to establish South Africa as a leader in Big Data by 2020. This document proposes a collaborative consortium of universities, industry and government programs (including SKA and CHPC) in South Africa and beyond to other African countries, to coordinate activities, advance research, develop necessary expertise, and spin off commercial ventures in Big Data. While this proposal should be vetted by the data services community in South Africa, it can perhaps serve as a starting point for the ‘ambitious proposal’ in data services recommended by this Committee, particularly with more balance paid to data sharing policies, data management, metadata tools and processes, curation and preservation. Also, as a matter of principle, the Committee recommends that elements that focus on research not be captive within NICIS activities but rather be presented as openly-competed solicitations to the community.

Prior to this ambitious proposal to significantly expand the scope of data services, it is important near-term to clarify the identity of the current DIRISA program and better integrate the efforts of the storage infrastructure and data services elements of the current organisation. NICIS Data Services should define, design and then provide the entire range of data services across the data lifecycle, with storage infrastructure being just one of many elements of these services. Other elements include data management tools and practices, metadata standards and tools and best practices, search and discovery tools, data curation, and long-term preservation practices. It is interesting to look at comparators in other countries. For example there are two separate data programs in Australia – the Research Data Storage Infrastructure (RDSI), similar to DIRISA/VLDB, and the Australian National Data Service, which is analogous to the non-infrastructure program elements proposed here. And Infrastructure is but one of five working groups established by Research Data Canada, with the others being Policy, Standards/Interoperability/Interdisciplinary Liaison, Education and Training, and International Liaison. These examples illustrate the breadth required for an effective, balanced data services program.

Furthermore there must be a better balance between *funding* for the storage infrastructure *per se* and the development of broader data services. There must be a reasonable near-term increase in funding from DST to fund the organisation beyond the current “coalition of the willing”. While cyberinfrastructure is readily measured in Teraflops and Petabytes, scientific impact requires investments in people.

Particularly as data programs are relatively new and yet demanding substantial attention and investments, it is crucial for the NICIS Data Services area to participate in international forums such as the Research Data Alliance. NICIS Data Services should facilitate South Africa’s role in international data programs and policy development, including metadata standards, exchange methods, sharing policies, etc. The SKA program is indeed a major international driver for data infrastructure and management, and it is crucial for South Africa to play a leadership role in SKA that can be leveraged to broader engagement.

One step in the transformation is that the funding that presently goes to CHPC for operating the DIRISA/VLDB storage should go to the Data Services Area and that NICIS should subcontract the storage hardware facility component necessary to support their planned data services. (The current CHPC appears to be a logical organisation to acquire and operate such a facility, with the requisite facilities, expertise and staff knowledge for the infrastructure.)

Furthermore, the Data Services Area should draw on input from the research community to develop an allocation procedure for national storage facilities (presently the VLDB, but also to include any future DST-funded repositories), including policies for retention of data and criteria for storage allocations (e.g. scientific impact of data retention, whether and how data are shared, metadata available for search and discovery). It is expected that allocations would involve a proposal process, open to individual university researchers in any field, with peer review by a panel of scientists representing a variety of domains, analogous to procedures for CHPC's computing resources. By opening access to all researchers and receiving their proposals, the Data Services Area can better gauge the demand for storage and plan for it, even if that demand cannot all be accommodated near-term. There may be economies of scale by leveraging a data storage infrastructure for use by large-scale programs such as Meerkat or SKA; however, those costs should be borne by the individual programs, particularly as large programs have requirements that cannot be subject to a competitive allocation process for individual researchers.

One issue is whether NICIS should focus on a small number of large-scale, centralised storage facilities or plan for a larger number of distributed facilities (e.g. at the university or program level), which may be federated in some fashion. As discussed in Section 2, it is generally best that DST focus its investments in centralised (Tier 1) facilities; there are economies of scale with operating centralised storage facilities, particularly to ensure that appropriate data reliability and management procedures are in place. However, universities are likely to establish (Tier 2) facilities and policies for retention of crucial research data, in part because the incentives for curation and retention of data may be more effective at the local university level than the national level. And one of the principles espoused by the Data Sector Working Group is that data is best curated at its source, by its owner. The autonomous nature of universities and research programs suggests that institutional storage facilities will always be a reality, so a centralised physical infrastructure which can virtually federate distributed repositories will allow more data sources to be accessed through a national repository, as well as provide more training opportunities for people at universities and labs around the country.

An important function of NICIS is to work with DST and other science funding agencies within South Africa to develop policies for data sharing and access, with appropriate incentives to researchers. This is receiving substantial attention in Europe, the US and other countries. For example, the US NSF and NIH funding agencies have required some form of Data Management Plans for some time. Recently the US Office of Science Technology and Policy has promulgated a much-expanded effort in data sharing across all federal funding agencies (Holdren, 2013); this mandate is now being translated to specific researcher guidelines across various agencies.

Another element of the Data Services portfolio is to coordinate a national dialog regarding journal licensing, a potential national online publishing facility, and open access policies. There is a rapid transformation in the journal/publishing industry that has profound implications for how researchers publish their work and access the publications of others, and how universities/researchers pay the ever-increasing costs of journals. NICIS can facilitate the national dialog.

5.4 Skills and Training Services

5.4.1 Background and motivation/Description

As pointed out in the *Cyberinfrastructure Vision for the 21st Century Discovery* (NSF, 2007), CI is enabling powerful and unprecedented opportunities. These opportunities include new modes of collaboration, modelling and visualisation of complex scientific and engineering concepts, as well as creation and discovery of scientific and educational resources for use in a variety of settings. These changes will support a new level of technical competence in the science and engineering workforce and in our citizenry at large. However, the complexity, novelty and changing nature of cyberinfrastructure means that there is a high risk of under-utilisation, or non-optimal exploitation without adequate investment in skills development through education and training (e-IRG, 2008).

The enabling role of human factor development was also pointed out to the Committee by the Sector Working Group on data, emphasising the need to build a cohort of data professionals to support research infrastructure development.

The shortage of IT skills is a global phenomenon. In UK there were 3,420 A-Level students in computing in 2011/12, as compared to 12,529 in 1998 (Next Gen Skills, 2013). Also in UK, a report from the National Audit Office estimates that it will take 20 years to close the skills gap within IT security (NAO, 2013). Furthermore, McKinsey Global Institute (2011) predicts that there will be a shortage of 190,000 data scientists in the US in 2019. In recognition of the particular national challenges of human capital development in South Africa, the Committee identifies skills development as a key element in an integrated strategy to ensure optimal societal return on CI investments.

5.4.2. Sustainable Skills and Training Services

Recommendation 19: *NICIS should offer effective coordination of CI Skills and Training services within a sustainable framework. NICIS should maintain a coordinated plan to develop and grow the community of researchers benefiting from CI.*

The main targets of these services are

- Cyberinfrastructure professionals (support personnel), for developing their operational knowledge of CI.
- Researchers that use, or could benefit from using, CI services to enhance their research or collaboration capacities.

The services should also have a link to the “next generation” of students in computer science that could become future CI experts and students in other disciplines that could become future users.

The services within Skills and Training will be delivered in collaboration with the higher education institutions (HEIs), as HEIs have a natural responsibility for curriculum development and student education. In line with the Finnish CSC model, the role of NICIS will be to co-design and co-organise training courses together with the relevant stakeholders (e.g., the admins of the NREN, coordinators of doctoral programmes). This collaborative approach will enable the organisation and coordination of a larger set of events due to the inkind resources provided by the end user institutions. Furthermore, it will ensure relevance and buy-in from the stakeholders.

Together with the stakeholders, NICIS should co-develop a portfolio of training modules in topics such as computational science, numerical algorithms, grid-computing, parallel programming, cloud computing, data-centric computing, e-science, computer animation and computer graphics. As suggested in *A strategic vision for UK e-infrastucture* (Tildesley, 2012), elements of this training should be included in established doctoral training programmes where escience would sit alongside disciplines such as physics, biochemistry, archaeology, social sciences.

NICIS will offer streamlined support, coordination and practical arrangement services for courses and events, as well as a pre-planned course calendar balanced to the user needs. Such a support infrastructure will facilitate seasonal events (summer schools, etc.) as well as *ad hoc* events within a national NICIS brand.

Education in the general concepts of computational thinking and data-driven discovery for the broad mass of students falls outside the responsibility of NICIS, as does the education of a critical mass of experts who can drive the future innovation of CI services. Still, it is important that NICIS have good communication with the HEI sector in this capacity and also that NICIS has the leverage to provide stimulus to institutions that want to take a frontrunner role in these areas. A successful NICIS needs to relate to the entire value-chain of knowledge creation. NICIS skills and training services should make use of and further stimulate the existing CI elements and competences in South Africa, and link to existing programmes where possible. An infrastructure for remote education in CI should be maintained and training material should be standardised and made openly available (e-IRG Report, 2008).

As suggested in *Cyberinfrastructure Vision for the 21st Century Discovery* (NSF, 2007), on-going attention must be paid to the education of the professionals who will support, deploy, develop, and design current and emerging CI. This is particularly pressing within the area of data-driven scientific inquiry, where there are vast needs for digital data management or data curation professionals. This point is strongly supported by the DIRISA SWG, which suggests a professional oversight is needed of curriculum development in the area of data management in higher education institutions.

The Committee supports the proposal in (NSF, 2007) of developing new careers involving the development of hybrid degree programs that marry the study of library science with a scientific discipline. Similarly, the power that visualisation and other presentation technologies bring to the interpretation of data may call for specialised career opportunities that pair the graphic arts with a science or engineering discipline. These points find support in "*A Strategic Vision for UK e-Infrastructure*" (Tildesley, 2012), which also recognises the need to put higher value on the scientific programmer and data scientist/curator/librarian.

Topics that need to be covered include computational science, numerical algorithms grid computing, parallel programming, cloud computing, data-centric computing and analysis, e-science, computer animation and computer graphics. Furthermore, in Section 5.1.3. The Committee has identified a need for a more focused effort on building networking skills at the institutional level. This is necessary for the institutions to be able to integrate and benefit from new networking capabilities as these emerge. The area of data management needs particular attention in view of the data deluge that is experienced across all sectors of society. According to the DIRISA SWG, important topics within data management include maintaining the essential research data infrastructure; data science including analysis and visualisation; curation and long term preservation; and auditing. As recommended by the HPC SWG, training programs to accelerate the research community competence in the quickly changing multi- and many-core computer architecture is essential to enable efficient utilisation of parallelism in new generations of computer systems.

Providing domain specific workshops in partnership with the research community to advance efficient HPC utilisation by researchers;

5.4.3. Identification of Key Functions

The key functions that should be covered by NICIS Skills and Training Services are described below. Many of these are parallel activities of the US XSEDE initiative[13], which runs an extensive training, education and outreach program.

- ***Training:*** NICIS will coordinate and co-develop courses together with the stakeholders. NICIS will maintain a course calendar at the national level for CI subjects and organise the events. NICIS will run road shows and ensure presence at conferences. NICIS will facilitate on-line training opportunities and webinars. NICIS will also support the development and maintenance of standardised training material. accelerate the Research community competence in ever changing multi- and many-core computer architecture and to enable efficient utilisation of parallelism within chipset technology. Training modules should be offered on "commercial basis" to industry and the public sector in order to stimulate workplace skills development.
- ***Campus Champions.*** The concept of campus champions have been successfully implemented by XSEDE, and a similar program is currently run by EGI in Europe. A NICIS Campus Champions program will support campus representatives as liaisons to the NICIS services and local sources of knowledge on CI concepts with the aim of empowering local researchers, educators and students to advance scientific discovery.
- **Curriculum and Educator Programs.** NICIS will support the community in their efforts to (further) develop educational activities and formal programs within CI concepts. NICIS will support the creation of model competencies of CI skills in order to build the future South African workforce. Activities could include promoting postgraduate cohort training in topics such as data management through Postgraduate Studies Units in higher education institutions. As an example recommended by the HPC SWG, a formalised curriculum in HPC is important for development and sustainability of skills, and can be introduced at multiple levels, from under-graduate to post-graduate levels.
- ***Student Engagement.*** This NICIS program will prepare a larger and more diverse group of students to be future researchers and educators. The program should include training, internships, fellowships, mentoring. Students are to be recruited nationally, with focus on underrepresented groups and underrepresented institutions. The aim is to provide students with real-world research and development experience

[13] https://www.xsede.org/

that can encourage them to pursue a future career or advanced degree in digital science or services.

The skills development will need collaboration across a number of disciplines and faculties at the HEI, building on the best practises that exist within disciplines and institutions. There must be formal engagement with key-stakeholders (e.g. computer science and electrical engineering departments, and academic organisations such as the South African Institute of Computer Scientists and Information Technologists).

Furthermore, it is natural to involve the Department of Communications' e-Skills Institute in the development of activities at the national scale. Libraries and library schools need to be involved in the development of data management and curation skills development activities.

5.5 Administrative Services

The Administrative Services is headed by the NICIS CEO and assisted by the administration director. In the case that NICIS is placed under an existing host organisation, then the NICIS office is located at the hosting organisation premises and the administrative tasks are generally performed by hosting organisation employees. The administration director will coordinate all NICIS office activities and supervise the personnel contributing to the execution of administrative tasks. The administration director will also follow up day-to-day activities and act on issues arising from work in the technical areas. The administration director is responsible for financial affairs, meeting arrangements, maintaining archives, legal matters, for internal and external communications. The administration director will assist the CEO in preparing agenda and meeting documents for the NICIS Board.

The administrative director is responsible for the development and follow-up of a communications plan. Internal communications include arrangements for face-to-face meetings, teleconferences and providing appropriate communication and collaboration tools. This includes providing mechanisms that are needed for a distributed NICIS organisation to build and maintain a high team spirit and workforce morale, as well as mechanisms for escalating issues and resolving conflicts. A central NICIS calendar will be maintained for all regularly scheduled meetings to avoid conflicting appointments. These meetings include Executive Group, NICIS (administrative) office meeting, other technical services meetings, NICIS Board meetings. There will be an annual NICIS all hands meeting to foster interaction between NICIS participants, stakeholders and users. External communication includes communicating the NICIS concept and its successes, and establishing strong brand identification targeted to the entire range of external stakeholders. In addition, all Executive Group members will deliver presentations and participate in external events as essential elements of their job descriptions.

List of References

Armbrust, M. et al. 2010 A View of Cloud Computing. *Communications of the ACM* 53(4), pp. 50-58

Berriman et al., 1983. The Application of Cloud Computing to Scientific Workflows: A Study of Cost and Performance. *Proceedings of the Royal Society* A, 371, pp. 1–14.

Blue Ribbon Task Force on Sustainable Digital Preservation and Access, February 2010. *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information*. [online] Available at: <http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf> Date of access: 26 September 2013

CSIR and Department of Science and Technology, Republic of South Africa, 1 February 2012. *ICT Roadmap. Towards Digital Advantage: Roadmapping South Africa's ICT Future.* Version 2. [online] Available at: <http://www.meraka.org.za/ictroadmap/sites/default/files/Roadmap%20Report%20DEV4.pdf> Date of access: 26 September 2013

Denny, J. 2012. Mining electronic health records in the genomics era. *PLoS Computational Biology* 8(12), e1002823.

Department of Higher Education and Training. *The Charter for the Humanities and Social Sciences.* 30 June 2011. [online] Available at: <http://www.info.gov.za/view/DownloadFileAction?id=150166f> Date of access: 26 September 2013

Department of Science and Technology, 2007. *Information and Communication Technology. Research & Development and Innovation Strategy.*[online] Available at: <http://www.dst.gov.za/images/ICT_RDI_Strat_2007.pdf> Date of access: 25 September 2013

Department of Science and Technology, 2008. *Innovation Towards a Knowledge-Based Economy. Ten-Year Plan for South Africa (2008-2018)*. [online] Available at: <http://unpan1.un.org/intradoc/groups/public/documents/CPSI/UNPAN027810.pdf> Date of access: 25 September 2013

Department of Science and Technology, 2012. *Ministerial Review Committee on the Science and Technology and Innovation Landscape in South Africa*. [online] Available at: <http://www.dst.gov.za/images/FINAL_MINISTERIAL_REPORT_MAY_16__4_.pdf> Date of access: 25 September 2013

Global Corporate Governance Forum, 2009. *Stakeholder Engagement and the Board: Integrating Best Governance Practices*. [online] Available at: <http://www.ifc.org/wps/wcm/connect/19017b8048a7e667a667e76060ad5911/FINAL%2BFocus8_5.pdf?MOD=AJPERES> Date of access: 25 September 2013

Hey, T., Tansley, S. and Tolle, K. (Eds.), 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research. [online] Available at:

<http://research.microsoft.com/en-us/collaboration/fourthparadigm/> Date of access: 25 September 2013

Holdren, J.P, *Memorandum for the Heads of Executive Departments and Agencies.* Executive Office of the President, Office of Science and Technology Policy, 22 February 2013. [online] Available at: <http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf> Date of access: 25 September 2013

Jensen, P. et al. 2012. Mining electronic health records: toward better research application and clinical care. *Nature Reviews Genetics* **13**(6), pp. 395-405.

Juve, G et al. 2013. Comparing FutureGrid, Amazon EC2, and Open
Science Grid for Scientific Workflows. *Computing in Science and Engineering*. [online] Available at: <http://doi.ieeecomputersociety.org/10.1109/MCSE.2013.44> Date of access: 26 September 2013

Kennedy, S. 2013. Supercomputing heads for the cloud. *The Australian*, July 23. [online] Available at:
<http://www.theaustralian.com.au/australian-it/cloud-computing/supercomputing-heads-for-the-cloud/story-fn8lu7wm-1226683406191> Date of access: 25 September 2013

Koyuncugil, A. and Ozgulbas, N. 2012. Early warning system for financially distressed hospital via data mining application. *Journal of Medical Systems* **36**(4), pp. 2271-2287.

Martin, F.G. 2012. Will Massive Open Online Courses Change How We Teach? *Communications of the ACM*, 55(8), pp. 26-28

McKinsey Global Institute, 2011. *Big data: The next frontier for innovation, competition, and productivity*. Report by James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers. [online] Available at: <http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation> Date of access: 25 September 2013

Nagasaki, H. et al. 2013. DDBJ Read Annotation Pipeline: A Cloud Computing-Based Pipeline for High-Throughput Analysis of Next-Generation Sequencing Data. DNA Research 20(4), May,pp 384-390. doi: 10.1093/dnares/dst017

National Audit Office (NAO), 2013. *The UK cyber security strategy: Landscape review*. ISBN: 9780102981285. [online] Available at: <http://www.nao.org.uk/report/the-uk-cyber-security-strategy-landscape-review/> Date of access: 25 September 2013

National Planning Commission; Department: The Presidency, Republic of South Africa, 2012. *Our Future - Make It Work. National Development Plan 2030.* [online] Available at: <http://www.npconline.co.za/pebble.asp?relid=25> Date of access: 25 September 2013

National Science Foundation. Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. January 15, 2003 .[online] Available at: <http://www.nsf.gov/cise/sci/reports/atkins.pdf> Date of access: 26 September 2013

National Science Foundation (NSF), 2007. *Cyberinfrastructure Vision for 21st Century Discovery*, NSF 07-28. [online] Available at: <http://www.nsf.gov/pubs/2007/nsf0728/index.jsp> Date of access: 25 September 2013

NextGen Skills. *Low take up Computing across Country and London hurts high-tech growth.* Press Release 29 January 2013. [online] Available at: <http://www.nextgenskills.com/computing-a-level-entries-fall-again-worries-for-hi-tech-talent-in-england-and-london/> Date of access: 25 September 2013

Parker, I., *et al*. 2001. *A National Biotechnology Strategy for South Africa*. Department of Arts, Culture Science and Technology. [online] Avaialble at: <http://www.esastap.org.za/download/sa_biotechstrat_jun2001.pdf>Date of access: 25 September 2013

Saini, S. et al. 2012. An Application-based Performance Evaluation of NASA's Nebula Cloud Computing Platform. In *Proc. Fourteenth IEEE International Conference on High Performance Computing and Communications*, pp. 336-343. doi:10.1109/HPCC.2012.52

The e-IRG Education and Training Task Force Report, 2008. [online] Available at: <http://www.e-irg.eu/images/stories/publ/task_force_reports/ettf_long_report_final_july08.pdf> Date of access: 25 September 2013

The Government of the Republic of South Africa, 2002. *National Research and Development Strategy*. [online] Available at: <http://www.dst.gov.za/index.php/resource-center/strategies-and-reports/174-national-research-a-development-strategy-2002> Date of access: 25 September 2013

Tildesley, D. 2012. *A strategic vision for UK e-infrastructure: a roadmap for the development and use of advanced computing, data and networks* (2012). Report commissioned by David Willetts, Minister for Universities and Science. [online] Available at: <https://www.gov.uk/government/publications/e-infrastructure-strategy-roadmap-for-development-of-advanced-computing-data-and-networks> Date of access: 25 September 2013

Waldrop, M.M. 2013. Massive Open Online Courses, aka MOOCs, Transform Higher Education and Science. *Nature* **495**, 14 March, pp. 160-163.

Yelick, et al. 2011. The Magellan Report on Cloud Computing for Science. US Department of Energy Office of Advanced Scientific Computing, December. [online] Available at: <http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Magellan_Final_Report.pdf> Date of access: 25 September 2013

Glossary of Abbreviations

ACE	Advanced Computing Engineering lab (CHPC)
ACCESS	Applied Centre for Climate And Earth System Science
ALICE	A Large Hadron Collider Experiment
ASAUDIT	Association of South African University Directors of Information Technology
ASSAf	The Academy of Science for South Africa
ATLAS	One of the very large experiments at CERN's LHC (A Toroidal LHC Apparatus)
CERN	European Organisation for Nuclear Research
CHPC	Centre for High Performance Computing
CI	Cyberinfrastructure
CSIR	Council for Scientific and Industrial Research
CSC	Finnish IT Centre for Science
DHET	Department of Higher Education and Training
DIRISA	Data Intensive Research Initiative of South Africa
DST	Department of Science and Technology
DTI	Department of Trade and Industry
DoA	Department of Agriculture
DoE	Department of Energy
DoH	Department of Health
DoM	Department of Minerals
DVC Forum	Deputy Vice-Chancellor Forum
DWAF	Department of Water Affairs
EDD	Economic Development Department
Eduroam	Education Roaming
e-VLBI	electronic Very Large Baseline Interferometry
FPGA	Field-Programmable Gate Array
GDP	Gross Domestic Product

GÉANT	Pan-European Research and Education Network
GEOSS	Global Earth Observation System of Systems
GRDI	Global Research Data Infrastructure
GVA	Gross Value Added
HartRAO	Hartebeesthoek Radio Astronomy Observatory
HEI	Higher Education Institutions
HESA	Higher Education South Africa
HPC	High Performance Computing
HSRC	Human Science Research Council
IaaS	Infrastructure as a Service
ICT	Information and Communications Technology
ISP	Internet Service Provider
KAT	Karoo Array Telescope
LHC	Large Hadron Collider
MeerKAT	MeerKAT array
Meraka	Meraka Institute
MRC	Medical Research Council
NACI	National Advisory Council on Innovation
NEP	National Equipment Programme
NICIS	National Integrated CyberInfrastructure System
NIH	National Institutes of Health (US)
NRF	National Research Foundation
NSTF	National Science and Technology Forum
NWU	North-West University
OEM	Original Equipment Manufacturer
PRACE	Partnership for Advanced Computing in Europe
RDA	Research Data Alliance
RDI	Research, Development, and Innovation

SaaS	Software as a Service
SAEON	South African Environmental Observation Network
SKA	Square Kilometer Array
SALT	Southern African Large Telescope
SANReN	South African National Research Network
SANBI	South African National Bioinformatics Institute
SANSA	South African National Space Agency
SARIMA	Southern African Research and Innovation Management Association
SCAP	Scholarly Communication in Africa Programme
SEACOM	not an acronym
SME	Small and Medium-sized Enterprises
SWG	Sector Working Group
TENET	the Tertiary Education and Research Network of South Africa
TCO	Total Cost of Ownership
TIA	Technology Innovation Agency
UCT	University of Cape Town
UL	University of Limpopo
UJ	University of Johannesburg
UFS	University of the Free State
VLSI	Very Large Scale Integration
VRE	Virtual Research environment
WACS	West Africa Cable System
WITS	University of the Witwatesrand
WOS	Web Object Storage

Glossary of Terms

Tier 1 denotes a high performance facility at the top of an HPC provisioning pyramid (National Facility).

Tier 2 and 3 denote high performance facilities in the middle and at the bottom of an HPC provisioning pyramid (Regional and University Centres).

Grid computing is the collection of computer resources from multiple locations to reach a common goal. The **grid** can be thought of as a distributed system with non-interactive workloads that involve a large number of files. (http://en.wikipedia.org/wiki/Grid_computing)

Cloud computing relies on sharing of resources to achieve coherence and economies of scale similar to a utility (like the electricity grid) over a network (http://en.wikipedia.org/wiki/Grid_computing).

Cyberinfrastructure refers to research environments that support advanced data acquisition, data storage, data management, data integration, data mining, data visualisation,and other computing and information processing services distributed over the internet beyond the scope of a single institution. In scientific usage, cyberinfrastructure is a technological and sociological solution to the problem of efficiently connecting laboratories, data, computers, and people with the goal of enabling derivation of novel scientific theories and knowledge (http://en.wikipedia.org/wiki/E-infrastructure).

Mid-range facility refers to a cyber-facility that provide resources that are limited in scope.

Appendix A: List of People Interviewed

The Committee interviewed the following individuals:

- Mr Laurens Cloete (Executive Director: Meraka, CSIR)
- Dr Jeffrey Mabelebele (Acting CEO: HESA)
- Mr Jeremy Main (IT Manager: SKA)
- Dr Romilla Maharaj (Executive Director: NRF)
- Dr Werner Janse van Rensburg (Sasol)
- Dr Albert van Eck (UFS, HPC)
- Dr Bruce Becker (SAGrid)
- Mr Leonhard Staphorst (SANReN)
- Mr Duncan Greaves (TENET)
- Prof. Peter Clayton (RU/DVC Research Forum)
- Mr Kabelo Bokala (ASAUDIT)

The User Community Workshop on 29 January 2013 was attended by

- Dr Dale Peters (UKZN)
- Mr Wim Hugo (SAEON)
- Dr Lucia Lotter (HSRC)
- Dr Bruce Becker (CSIR)
- Dr Albert van Eck (UFS, HPC)
- Prof Marius Potgieter (NWU, Physics)
- Prof Gert Kruger (UKZN, Chemistry)
- Prof Enrico Lombardi (Unisa, Physics)
- Mr Jeremy Main (SKA)
- Dr Daisy Selematsela (NRF)
- Mr Duncan Greaves (TENET)
- Mr Leonhard Staphorst (SANReN)
- Dr Happy Sithole (CHPC)
- Prof Colin Wright (SANReN)

- Dr Werner Janse van Rensburg (SASOL)
- Prof Peter Clayton (DVC Forum/RU)

Responses were received from

- Prof Kevin Naidoo (UCT)
- Prof Alan Christoffels (SANBI)
- Prof Mike Inngs (UCT)
- Dr Bernie Fanaroff (SKA)

Appendix B: Committee Responses to SWG Inputs

B.1 High-Performance Computing Sector Working Group

B1.1. Response to the HPC SWG Document

This section addresses the report of the HPC SWG. Many of the ideas and recommendations have been incorporated into the main body of this report. In this section, we comment on some of the issues where the committee may have had a different perspective, or where we need to integrate and harmonise the proposals with those of other SWGs.

Page 2: Research Chairs: The idea of South African Research Chairs Initiative (SARChI) chairs in HPC has significant merit and needs discussion. We are aware that at least one university put forward a proposal for such a chair in the last round of SARChI applications. Whether a chair should be ring-fenced in the area could be discussed with the NRF.

Page 2: The need for a formalised HPC curriculum has been motivated. The CHPC has played and should continue to play a role both as an advocate and provider of training. However, for this project to succeed, there should be full engagement with universities and academic societies (e.g., SAIEE and SAICSIT). University academic IT department produce graduates that serve a range of needs. Some universities may decide that they have other priorities besides HPC. But the HPC and S&T service areas can play a powerful role in coordinating and helping develop HPC.

Page 3: The HPC SWG recommends that a comprehensive status report of HPC systems at universities and in industry should be carried out. We did ask participants in our meetings for this information, and we endorse the proposal.

Page 6: We endorse the recommendation that while Top 500 membership is a useful benchmark, it should not be an end in itself.

Page 6-9: This presents a hardware technology roadmap. Largely we agree with this, though point out the dangers of extrapolating to 2025 based on only a few years history, and that the choice of 10% as a growth rate may be too high or low. More seriously, the Committee feels that the hardware technology roadmap must go hand-in-hand with a software and services roadmap. Different sectors have very different needs at a system, software and hardware level, and the Computing Services Area must have the flexibility to provide service to all these needs. These differences are likely to become more marked as the HPC usage expands across the disciplines.

Page 9: The funding model is one the Committee agrees with. However, as with other parts of the HPC SWG, the metrics of performance must be much clearer and transparent.

Page 10: Industrial User Forum: The idea of an industrial user forum is an excellent idea and the Committee endorses the idea and the bulk of the planned activity. The idea of industry access to CHPC equipment is good, and we endorse these subject to the concerns we express in this report (section 4.4.3): the development of an indigenous high-tech IT industry may be stifled unless the costing and support model is correct. The Computing Services Area must also see a role in supporting nascent the nascent HPC industry.

Page 12: There are interesting ideas on funding of research and the role of the NRF. The HPC SWG has pointed out the difficulty that the CHPC has in evaluating research proposals.

B1.2. Proposal for second node

The proposal for the establishment of a new major facility based at the CSIR was only sent to us after the body of our report was submitted to DST.

The motivation of the proposal is primarily operational. The Committee is fully aware of the limits of the current physical facility. The question of expansion and/or moving of the facilities is one that will be justified by capacity constraints and the costing. These operational issues are not within the remit of the Committee nor does it have access to the data required to assess the merits of the proposal. The Committee therefore addresses only process and strategic issues.

First, a major operational change like this will have significant strategic implications. In section X, the Committee presented its views on the possibility of a multiple Tier 1 sites. The report points out that there are significant benefits of scale in having one larger facility than two smaller facilities as it allows resource allocation to be done more efficiently and larger problems to be tackled. Moreover, the very significant overhead costs such as physical security, power security, air-conditioning and so on can be more effectively amortised.

Secondly, the argument in the proposal that the physical proximity of the new facility to northern universities has some merit. Much more important, however, is that there is very fast network access to the facilities and that the facilities are powerful and effectively managed. Moreover, physical proximity is important for access to people and expertise; except for very specialist purposes (e.g., rare access to the Convey) physical proximity to the equipment is only marginally useful. As was pointed out in early submissions of the HPC SWG, it is not necessary for all CHPC skills to be co-located with the equipment and we have already been told of successful CHPC activities in this regard.

Third, the establishment of this new major facility is likely to lead to a downgrading of the status of the current site, with implications for current relationships and staffing. The Committee has no view one way or the other on this issue, but points out that this should be factored into decision-making.

Fourth, the costs are likely to be significant which will have implications on strategic options available. The proposal must be fully costed, and particularly alternative proposals must be fully investigated in the light of strategic objectives and the disaster mitigation and recovery policy of the CHPC. What are the alternative facilities available in Cape Town? What quantity and type of staffing is required to support a backup facility, and to what extent do they have to located at the facility? There are large commercial facilities based in different parts of the country that could easily host the proposed storage facility. Could space be leased? A proposal of this magnitude can only be assessed when the alternatives are fully explored, particularly when the complex costing implications are explicit. Costing of this proposal and alternatives must be provided.

B.2 Networking Sector Working Group

Governance: (ref NREN SWG v17 08/12/13 10:27:23 PM, Section II.A., item iv.) The SWG refers to the Ministerial NSI Report (2012), where it is recommended to constitute a National Advisory Panel on Cyberinfrastructure, reporting to a proposed National Council for Research and Innovation. The SWG suggests that it would be appropriate to populate a substructure of this advisory panel to provide strategic governance to the NREN. The Committee believes that the combination of our proposed NREN "User" Forum and the NICIS Board will be well-suited structurally of accommodating for stakeholder engagement, both in the context of high-level integrative CI strategy (Board) and in the context of specific NREN challenges (NREN Forum).

Establishment of a single NREN.

The Committee believes that its proposals are consistent with the NREN SWG in principle.

Key functions. The Committee's views on core functions are largely the same as that of the SWG. The SWG also lists several value-added services. The Committee agrees with most of the points made and sees these as valuable and as being within the remit of the NREN service area. The only open issue appears to be the funding of maintenance costs; this should be explicitly covered in the SANReN/TENET contract.

On the question of grid computing, the Committee agrees with the views of the SWG on a vision of grid computing. However, the Committee questions whether the Computing Services Area would not be a better home. The Network SWG makes a logical distinction between the coordination of the grid and the underlying computational resources (which may be managed by the Computing Services Area or universities) which explains why their view has merit.

The SWG makes no recommendations regarding cloud computing.

The other value-added services are sensible and should be encouraged. The management of and operations of these services is not fully elaborated upon. Whether these services are practically provided by TENET or directly by the service area, the Committee believes that the planning of these should be done under the aegis of the service area and clear lines of accountability and metrics for performance are implemented. A common help/user desk would be provided so that user's can request help or report problems at one place and avoid situations where users are referred from pillar to post. It must be recognised by everyone that a risk in keeping two organisations and not having a full merger that both operational and strategic accountability can be compromised.

The SWG makes a strong case for a much broader mandate for the services area — for example providing network services to schools and hospitals. The SANReN experience and track-record has been very positive so this is a tempting vision. However, the risks of significant expansion to providing core services must be carefully managed.

B.3 Data Sector Working Group

In general, the Committee concurs with the report provided by the Sector Working Group. The discussion below identifies those areas where there may be differences in the recommendations or priorities.

- The SWG report makes little mention of the current VLDB data repository (operated by CHPC), and, with respect to storage infrastructure, emphasises the need for DIRISA to federate distributed data repositories. In some respects, this illustrates the DIRISA/VLDB identity crisis discussed in Section 5.3 of the Committee's report, as well as the natural tension between centralised and distributed facilities discussed in Section 2. The Committee's report (Sec 5.3) acknowledges the current weak coupling between the VLDB infrastructure and DIRISA's 'coalition of the willing' working in broader data services, but recommends that the Data Services area integrate and leverage the VLDB investment as a large centralised data repository, while also enabling virtual federation of distributed repositories.

- There are significant differences in the recommendations regarding governance of the integrating framework above the Data segment itself – e.g. Figure 2 of the SWG report and and the discussion in Section 5.1 of that report. The SWG report has a number of potential organisations and programs participating in a virtual NICIS structure, with CHPC, NREN and DIRISA being only three initial organisations, with direct oversight of all programs by an Implementation Oversight Committee, and strategic guidance provided by a higher-level Cyberinfrastructure Board. In any of the permutations of the governance model recommended by the Committee, the governance is more structured and hierarchical with a formal organisation, led by a CEO with direct reports for each CI element (four Services Areas: Computing, Data, Networking, Skills). The SWG's propose strong user representation in the Implementation Oversight Committee, while the Committee's proposed model has a User Forum for each Services Area to guide the deployment of services. The Committee finds that the proposed model would be preferable to the SWG's suggestion due to a more hierarchical and transparent structure, leading to more accountability in the system. In particular, the Committee believes that it is preferable to have a CEO accountable to the board rather than a (chair of a) committee.

- The SWG report states that "Similar to the existing components of the national CI, CHPC and SANReN, DIRISA requires annual ring–fenced funding, budgeted in a three year business plan, and tabled as an agreed line item in the National Budget." The Committee recognises the need for sustained, multi-year funding for operating any infrastructure, particularly the data element of CI, where the long-term security of data is paramount to programs, data owners and data users. At the same time, the Committee recommends a consolidation of planning and budget across all service areas of the NICIS, without 'ring-fencing' at the individual component. (Ring-fencing around the *entire* NICIS budget may be appropriate.)

The SWG report mentions the Australian model which has two separate programs, RDSI and ANDS respectively, for data storage infrastructure and the higher-level data services beyond the infrastructure. It appears that the SWG report envisions a Data element similar to

ANDS, while the Committee specifically recommends a Data Services Area that addresses and integrates both of the functions represented by RDSI and ANDS.

Appendix C: International Benchmarking

One element of the remit to the NICIS Committee is to provide benchmarks for integrating frameworks that have been developed for national research CIs in other countries, and to identify principles and lessons learned that would inform recommendations for an integrating framework for South African cyberinfrastructure. The focus of this benchmarking exercise is on the *management structures* for coordinating and overseeing CI programs, rather than the level of financial investments, the comparative size/capabilities of facilities, or the specific nature of those programs.

To support this benchmarking exercise, the Committee has reviewed published reports/presentations from many countries and relied on personal knowledge of some international CI programs. The international programs considered include the European Union's EGI, PRACE, GEANT, Terena and Dante programs, the United Kingdom, Finland, Denmark, the Netherlands, the United States' NSF and DOE programs, Canada, Singapore, Taiwan, Korea and Australia. Limited time prevents a comprehensive review of all countries' CI programs, but the countries considered are diverse and represent a range of organisational models. The integrated observations and conclusions from the benchmarking exercise are summarised in Section 2 of the report. This appendix provides more extensive information about each of the international programs considered.

C.1 European Union

C.1.1 EGI

EGI.eu is a not-for-profit foundation established under Dutch law to coordinate and manage the European Grid Infrastructure (EGI) federation on behalf of its participants: National Grid Initiatives (NGIs) and European International Research Organisations (EIROs).

The services provided by EGI.eu to the wider EGI community aims to

- Oversee the operations of EGI to guarantee the integration of resources from providers around Europe into a seamless and secure e-infrastructure.
- Coordinate the support provided to EGI's user communities.
- Work with technology providers to source high-quality and innovative software solutions to answer users' requirements.
- Represent the EGI federation in the wider Distributed Computing Infrastructures (DCI) community through coordination and participation in collaborative projects.
- Coordinate the external services provided by partners in the community.
- Steer the evolution of EGI's policy and strategy development.
- Organise EGI's flagship events and publicise community's news and achievements.

Governance structure

EGI.eu is governed by the EGI Council (members), which is responsible for defining the strategic direction of the EGI federation. The Council acts as the senior decision-making and supervisory authority of EGI.eu. The Council participants (table) are the National Grid Initiatives (NGIs) and European Intergovernmental Research Organisations (EIROs).

The Council delegates oversight of the day-to-day running of EGI.eu to the Executive Board currently with seven members. The Executive Board devolves financial and organisational responsibility to the Director, who is supported by a staff of 22 people based at the EGI.eu headquarters in Amsterdam. EGI.eu's work is supported by the equivalent of another 22 full-time workers spread across many organisations around Europe.

EGI.eu participates in several EC FP7-funded projects, the main focus is the coordination of the EGI-InSPIRE project, which expires in 2014.

C.1.2 PRACE

The mission of PRACE (Partnership for Advanced Computing in Europe) is to enable high impact scientific discovery and engineering research and development across all disciplines to enhance European competitiveness for the benefit of society. PRACE seeks to realise this mission by offering world class computing and data management resources and services through a peer review process.

PRACE also seeks to strengthen the European users of HPC in industry through various initiatives. PRACE has a strong interest in improving energy efficiency of computing systems and reducing their environmental impact.

PRACE Research Infrastructure (RI)

PRACE is established as an international not-for-profit association (aisbl) with seat in Brussels. It has 25 member countries (as of October 2012) whose representative organisations create a pan-European supercomputing infrastructure, providing access to computing and data management resources and services for large-scale scientific and engineering applications at the highest performance level. The PRACE resources consists of very high-end systems (Tier 0) that only the principal partner countries can afford. These Tier 0 resources are complemented with national systems (Tier 1) which are offered by the general partner countries of PRACE.

The computer systems and their operations accessible through PRACE are provided by four PRACE members (BSC representing Spain, CINECA representing Italy, GCS representing Germany and GENCI representing France) who committed a total funding of €400 million for the initial PRACE systems and operations. In pace with the needs of the scientific communities and technical developments, systems deployed by PRACE are continuously updated and upgraded to be at the apex of HPC technology.

The PRACE project partners receive EC funding under the PRACE Preparatory and Implementation Phase Projects (PRACE-1IP, 2010-2012, RI-261557, PRACE-2IP, 2011-2013, RI-283493, PRACE-3IP, 2012-2014, RI-312763) for a total of €67 million complemented by the consortium budget of over €43 million.

PRACE HPC Access

PRACE systems are available to scientists and researchers from academia and industry from around the world through 3 forms of access:

- Preparatory Access is intended for resource use required to prepare proposals for Project Access. Applications for Preparatory Access are accepted at any time, with a cut-off date every 3 months.
- Project Access is intended for individual researchers and research groups including multi-national research groups and has one-year duration.
- Multi-year Access is available to major European projects or infrastructures that can benefit from PRACE resources and for which Project Access is not appropriate. Multi-year and Project Access are subject to the PRACE peer review process. Leading scientists evaluate the proposals submitted in response to the bi-annual calls.

Other activities:

PRACE has an extensive education and training effort for effective use of the RI through seasonal schools, workshops and scientific and industrial seminars throughout Europe. PRACE also undertakes software and hardware technology initiatives with the goal of preparing for changes in technologies used in the Research Infrastructure and provide the proper tools, education and training for the user communities to adapt to those changes. One goal of these initiatives is to reduce the life-time cost of systems and their operations, in particular the energy consumption of systems and the environmental impact.

C.1.3 GEANT

The GÉANT backbone comprises a hybrid network architecture delivering a service to NRENs' users with choice of connectivity options spanning IP (v4 and v6) and point-to-point circuits. This has involved the installation of 50,000 km of network infrastructure, including 12,000 km of optical fibre across Europe (2011). Over 80% of users across the GÉANT service area are able to access to point-to-point circuits at up to 10 Gbps dedicated capacity.

GÉANT is the pan-European communications infrastructure serving Europe's research and education community. The current network and associated programme of activities, known as GN3+ is co-funded by the European Commission and the NRENs, with total EC funding of €93 million and €88 million from NRENs over four years from April 2009.

C.1.4 DANTE

DANTE: (standing for 'Delivery of Advanced Network Technology to Europe') is a limited liability company and a not-for-profit organisation that plans, builds and operates advanced networks for research and education. It is owned by fifteen European NRENs and works in partnership with them and in cooperation with the European Commission.

C.1.5 e-IRG

The e-Infrastructure Reflection Group was founded to define and recommend best practices for the pan-European electronic infrastructure efforts. Its mission is to pave the way towards a general-purpose European e-Infrastructure. The e-IRG consists of official government delegates from all the EU countries. The e-IRG produces white papers, roadmaps and recommendations, and analyses the future foundations of the European Knowledge Society.

C.1.6 NRENs

National Research and Education Networks are national institutions in the EU member states and associated states responsible for providing research and education networking services within their territories. 32 NRENs representing 36 countries are full members of the NREN PC and 4 are special members.

Further information on NRENs' capabilities and activities can be found in the annual publication TERENA Compendium of National Research and Education Networks In Europe, www.terena.eu/compendium.

NREN PC: The governance body of the GÉANT backbone consisting of senior representatives of participating NRENs, DANTE and TERENA as well as 4 observers.

C.1.7 TERENA (www.terena.eu)

Trans-European Research and Education Networking Association offers a forum to collaborate, innovate and share knowledge in order to foster the development of internet technology, infrastructure and services to be used by the research and education community.

The GÉANT Expert Group in its report Knowledge without Borders (2011) proposed a new structure for European networking towards 2020 addressing three core functions:

- community building, high-level strategy and coordination;
- connectivity and services provision; and
- innovation.

This is in line with what is also promoted by e-IRG for European e-Infrastructure at large.

C.1.8 EIROforum

EIROforum is a partnership between eight of Europe's largest inter-governmental scientific research organisations that are responsible for infrastructures and laboratories: CERN, EFDA-JET, EMBL, ESA, ESO, ESRF, European XFEL and ILL.Its mission is to combine the resources, facilities and expertise of its member organisations to support European science in reaching its full potential.

A working group of the European EIROforum recently proposed to create a common e-Infrastructure platform for the future that builds on the experience of the last decade and is flexible enough to adapt to technological and service innovations. Such a platform must provide the underlying layers of common services, but must be adaptable to the very different and evolving needs of the research communities. The overall motivation is to address the challenges for the diverse, emerging "long tail of science" conducted by researchers that do not have access to significant in-house computing resources and skills. Again, this can be seen as an argument for an integrated CI system.

Sources (C.1)

http://www.e-irg.eu/images/stories/e_irg_whitepaper_and_comments_2011.zip

http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/geg-report.pdf

http://www.e-irg.eu/images/stories/e-irgs_reaction_geg_a5.pdf

e-IRG Roadmap December 7, 2012 http://www.e-irg.eu/publications/roadmap.html

http://go.egi.eu/EGI2020

"A Vision for a European e-Infrastructure for the 21st Century", EIROforum IT Working Group, 21th May, 2013.
http://indico.cern.ch/getFile.py/access?contribId=0&resId=0&materialId=0&confId=212498

http://www.eiroforum.org

C.2 Finland

CSC - IT Center for Science Ltd. is a wholly government owned company operating on a nonprofit principle. It is administered by the Ministry of Education, Science and Culture. In the Ministry, CSC belongs to the Department for Education and Science Policy.

CSC has existed since 1971 and is the only e-Infrastructure organisation for research in Finland. It has about 250 employees.

CSC's mission statement is "CSC, as part of the Finnish national research structure, develops and offers high quality information technology services". The CSC operational goals are:

- to improve conditions for research and product development in universities, polytechnics, research institutions and commercial life
- to comply with the information strategy of the Ministry of Education, Science and Culture by providing national services that would be impracticable to establish at the university level
- to promote collaboration between universities and polytechnics, research institutions and companies that utilise information technology for science
- to provide internationally competitive supercomputing and data communication services
- to serve as a pioneer and information provider in the latest information technologies for science.

CSC's statutory administrative bodies consist of the General Meeting of Shareholders, the Board of Directors and the Managing Director. The operative management of the company is under the responsibility of the Managing Director together with the Management Group. Senior scientific advisers support the Managing Director's and Management Group's work.

CSC has divided operations into five focal service areas:

- Data services for science and culture

Solutions for data storage, management and analysis needs

- Funet Networking Services Fast backbone network for academia – Versatile Networking Services
- Computing services
  High-performance computing and IT consulting services
- Application services
  Support, development and infrastructure for computational science and engineering.
- Information management services
  Information technology services for the needs of academia, research institutes, state administration and companies.

CSC's services are meant primarily for the use of the national research community. According to the contract between CSC and the Finnish Ministry of Education, Science and Culture the services for university and polytechnic researchers are mainly free of charge. Services for research institutes and companies are commercial use and the invoicing is based on organisation's service agreement with CSC.

Sources (C.2):

http://www.csc.fi/english

http://www.csc.fi/english/csc/publications/brochures/services

http://www.csc.fi/english/csc/publications/reports/csc_annual_report_2011

C.3 Denmark

Reorganisation and strengthening of Danish e-Science was planned as part of a research infrastructure roadmap process in 2010-2011. The Roadmap committee proposed that:

> *"A new, consolidated national organisation will be established that will possess the requisite resources and competences for supporting Denmark as an e-Science nation. The organisation will develop new funding instruments for responding to the challenges surrounding high performance computing facilities, network connections, provision of advice/guidance and training to new users, storage of scientific data and cloud computing."*

Scientific and technological advances entail that investment in computer technologies are relatively short-lived, which makes it a challenge to continually provide researchers with access to the latest tools and services they require in order to pursue research at the highest international level. In order to meet these challenges, and make efficient use of public-sector resources, there is a need for new funding instruments, on-going strategic assessment and increased joint national coordination and governance. A new national organisation for eScience will be conditional on the involvement of researchers from many different branches of science, as well as decision-makers at the individual research institutions and at central government level.

For a number of years, the Danish national e-Infrastructure has been based primarily on two organisations: the Danish Research Network, which hosts network connections between Danish research institutions (and via NORDUnet connections to institutions abroad), and the

Danish Center for Scientific Computing (DCSC) which supplies Danish researchers with computing power and large data repositories. In the interests of ensuring Danish e-Science of favourable development financially, technically and scientifically, efforts will be made to bring the Danish Research Network and DCSC under a joint board. The preliminary plan is for the new organisation to accomplish the following:

- contribute to formulation and implementation of a consolidated Danish e-Science strategy
- contribute to coherence and synergy between the research institutions' activities and to more explicit division of responsibilities between national and local activities
- perform those tasks currently performed by the Danish Research Network and DCSC
- ensure cost-effective solutions in line with advances in the field
- support training and advisory activities in the field, including establishing and operating a national centre of excellence for e-Science
- handle Danish membership of international cooperation, including NORDUnet, Nordic Data Grid Facility (NDGF), ESS-DMSC and PRACE.

Consortium partners
The Danish Agency for Science, Technology and Innovation appointed a committee in spring 2011 composed of key stakeholders with a view to preparing the statutes for the new organisation.

Funding
The present National Budget accounts for the Danish Research Network and DCSC are expected to be continued until a new organisation can come into effect. In addition to these existing National Budget allocations, provided that a satisfactory basis can be created for establishing the new organisation, the project will be eligible for additional funding of DKK 50m for launch of the new activities under the organisation. The DKK 50m should in principle cover a three-year running-in period. The details of funding allocations will be agreed in connection with preparation of the new statutes.

Current status
DeIC - Danish e-Infrastructure Cooperation was formed on April 19 2012 as a merger between Forskningsnettet (the Danish Research Network) and Danish Center for Scientific Computing (DCSC). DeIC is established under the Ministry of Research, Innovation and Higher Education by Act 70 of April 19 2012. Organisationally, DeIC belongs to the Danish Agency for Science, Technology and Innovation and is not an independent legal entity. The purpose of DeIC is to support Denmark as an e-Science nation through delivery of einfrastructures (computing, storage and network) to research and research based teaching. DeIC is a virtual organisation, which means that all employees are employed in other organisations, currently at the Danish Technical University, at the Danish Agency for Culture, at UNI-C and to some extend at Aarhus University, Aalborg University, University of Southern Denmark and University of Copenhagen.

Sources (C.3):
www.forskningsnettet.dk
http://en.fi.dk/publications/2011/danish-roadmap-for-research-infrastructure-2011

C.4 Netherlands

The Dutch government's key economic sectors policy focuses on collaboration between the commercial sector, knowledge institutions, and government. ICT links the various key sectors, and is seen as an important basis for innovation. SURF sees its role in this context.

In 2012, SURF received funding from the Dutch government to create a national ICT research infrastructure. In the years ahead, SURF's focus will be on constructing that infrastructure, which will consist of networks, supercomputers, grids, and data. This will be carried out by SURFnet, SARA and eScience Center.

Being able to apply the knowledge generated by outstanding research is recognised by SURF as an important factor in reinforcing its position on the knowledge front. SURF aims to contribute to the transfer of knowledge to the commercial sector. This takes place as part of a close relationship with the Netherlands eScience Center and will in the future also involve collaboration with SARA.

Organisational structure of SURF

Parent organisation SURF
Within the parent organisation, SURF, the emphasis is on administrative coordination and initiating ICT-driven innovation in higher education and research. SURF promotes knowledge sharing via theme-specific networks, professional development meetings, and the provision of publications, project results, and good practices. SURF finances innovative ICT projects within funding programmes. Each of the SURF operating divisions has its own focus in which it collaborates with national and international partners.

SURF Platforms
The innovation activities of the parent organisation SURF are structured within three platforms:

- Platform Board ICT & Education (focus area Innovation in Education)
- Platform Board ICT & Research (focus areas Scholarly Communication and Digital Rights)
- Platform Board ICT & Business Management (focus areas Organising with ICT,Security and privacy, Cloud computing and Green ICT)

SURFnet

SURFnet ensures that researchers, instructors, and students can work together simply and effectively with the aid of ICT. It enables users to fully exploit the opportunities offered by ICT by promoting, developing, and operating a trusted, connecting ICT infrastructure. In doing so, SURFnet focuses on two specific areas:

- network infrastructure: a hybrid fixed-wireless network as the basis for all collaboration, providing efficient, unlimited data transport;
- collaboration infrastructure: a pioneering collaboration environment that seamlessly connects systems, services, tools, and people.

In addition to provision of networking services, SURFnet runs the innovation project GigaPort3. GigaPort3 is the successor to the GigaPort Next Generation Network project, again succeeding the first GigaPort project. The aims are to take the SURFnet network infrastructure to a higher level, to make possible the next generation of networks, and finally to support dynamic services with a state-of-the-art and scalable hybrid network. GigaPort3 builds on the implementation of SURFnet6, the first national hybrid optical and packet-switching infrastructure in the world. By utilising new technology in and on top of the existing SURFnet6 network, GigaPort3 creates the new SURFnet7 network.

The budget for the GigaPort3 project (which runs from 2009 to 2013) is EUR 37 million. SURFnet is carrying out the project under the supervision of SURF and in cooperation with suppliers, users, and research partners.

SURFmarket

SURFmarket is the ICT marketplace for higher education and research and it facilitates the use of ICT. SURFmarket negotiates with ICT providers on behalf of the institutions that are affiliated to SURF. This enables the institutions to choose from a range of software, digital content, services, and hardware. Large-scale purchasing (using a digitized process) makes SURFmarket.nl an inexpensive and efficient licensing platform for higher educational institutions and university libraries, and for university museums and medical centres. SURFspot.nl is the webshop where students and staff can purchase software and other ICT products for home use at affordable prices.

SURFshare

SURFshare provides an organisational setting for SURF innovation projects that now operate independently and ICT collaboration initiatives in higher education and research. It focuses on the provision of specific services that are not available on the market. The innovation projects and initiatives concerned require management but are not yet on a large enough scale, or do not yet have the right conditions or opportunities, to be efficiently self-managing. In 2011, two former SURF initiatives, Studiekeuze 123 and Studielink became part of SURFshare.

SURF's Scientific Technical Council (WTR) is made up of recognised authorities. Its task is to offer independent advice to SURF, institutions, and government bodies regarding their ICTpolicy and activities.

The Project Monitoring Committee (PMC) monitors the progress of SURF projects.

SARA

In order to combine the national ICT facilities for research, it has been decided that SARA should be incorporated into SURF. SARA will be integrated as an independent division within the SURF organisation, and will continue to concern itself with the infrastructure and services in the field of high-performance computing and visualisation. SARA provides largescale computing capacity, an integrated ICT infrastructure, and support for researchers.

It also focuses on the cloud and associated developments, advanced and large-scale data storage, and the basic facilities for eScience.

Netherlands eScience Center
The Netherlands eScience Centre is a joint initiative of SURF and NWO. Its aim is to support and reinforce multidisciplinary and data-intensive scientific research in the Netherlands and to encourage the innovative use of ICT in research.

Sources (C.4):
http://www.surf.nl/en/
http://www.surfnet.nl/
http://esciencecenter.com/
https://www.sara.nl/
http://www.surf.nl/en/publicaties/Pages/SURF'sStrategicPlan2011-2014.aspx

C.5 United Kingdom

UK e-Infrastructure Leadership Council (ELC)

The ELC is responsible for developing a strategy to provide a world class e-infrastructure and High Performance Computing (HPC) capability for the UK. It works in partnership with stakeholders across the academic community, industry, government and society. ELC advises government on all aspects of e-infrastructure including networks, data stores, computers, software and skills.

The council was established in March 2012 following recommendations from the report ‘A strategic vision for UK e-infrastructure’.

The 23 members of the ELC come from the academic community, industry, the research and funding councils, government departments, and the charitable sector.

UK e-Infrastructures Academic User Forum

The goal of the UK e-Infrastructure Academic User Community Forum is to maintain and grow the community of UK researchers who use computers of any shape or size in their research, regardless of discipline or domain. The community was central to the development of the Strategy for the UK Research Computing Ecosystem (Oct 2011), which largely influenced the Department for Business, Innovation, and Skills (BIS) publication A Strategic Vision for UK e-Infrastructure. The Forum holds meetings to give the e-infrastructure community regular updates on developments in the UK e-Infrastructure space and within ELC, to provide a forum for discussions about the community's future, and to allow the e-infrastructure community the opportunity to influence Government and help define e-infrastructure policy in the UK.

Project Directors’ Group

This Group coordinates activities between the national projects that are part of the National E-Infrastructure. Interaction with BIS is conducted via the RCUK e-infrastructure Group, a coordinating group for e-Infrastructures across the research councils.

The Group undertakes projects to prototype the technologies and systems that will allow the integration of the compute, data and networking services of the National E-Infrastructure Eco-system. This is to achieve the aim of creating a one-stop-shop for access to resources for academics and non-academic users/customers.

Organisation of e-Infrastructure Services:

National Service – operated by Edinburgh Parallel Computing Centre.

National Networking - JANETUK

Specialist Systems – GRIDPP, DiRAC, STFC Hartree Centre, STFC Scientific Computing Department, Sanger (Wellcome), CRUK, Data-Archive (ESRC), NERC services such as JASMIN and CEMS services, National Oceanographic Service, BBSRC and MRC.

Regional HPC services

The main Central Research Computing Service of Higher Education Institutes (HEIs) Research Community Organisations, such as the Computational Collaborative Projects.

The support landscape in the UK for software used in research is characterised by a split into national facilities (e.g. STFC software engineering group, GRIDPP/EGI Training Portfolio, the Hartree Centre, HECToR CSE service, and the Software Sustainability Institute); institutional support (often through research computing services or IT services); and large scale scientific consortium (e.g. the Collaborative Computational Projects (CCPs)).

There is also large portfolio of software development projects for computational research funded by the Research Councils (e.g. the recent EPSRC Software Infrastructure and HPC Software Development calls). The EPSRC software as infrastructure initiative and the STFC Hartree Centre are both examples of a long term commitment to software development and support.

The National Service

The National Service provides a general purpose 800Tflop/s system for use by the EPSRC, NERC and BBSRC research communities. Other research areas and user groups can purchase time on this system. This service supports 2300 of our most advanced users. In addition a 9PB Research Data File system is also available for use by National Service and other UK HPC users, regardless of which system has been used.

National Networking

Janet is the UK's National Research and Education Network (NREN) that interconnects all of the Universities and research council facilities which contribute to the e-Infrastructure environment, and has well provisioned peering points where traffic is exchanged with commercial networks within the UK, and with both commercial and NREN networks within Europe and beyond. Janet is predominantly funded by government, and connection services are provided to eligible organisations on the basis of need within the budgets available.

Janet is owned by Jisc, a “company limited by guarantee and a charity”, which is owned by stakeholder organisations from the national higher education sector. More precisely, Jisc owns a subsidiary company, Jisc Collections and Janet Ltd, of which Janet is a part.

Specialist systems

The Specialist Systems are offered by GRIDPP, DiRAC, STFC Hartree Centre, STFC Scientific Computing Department, Sanger (Wellcome), CRUK, Data-Archive (ESRC), NERC services such as JASMIN and CEMS services, National Oceanographic Service, BBSRC and MRC. A total of 53 specialist systems have been deployed.

The main Central Research Computing Service of Higher Education Institutes (HEIs) and Regional HPC services (HEIs). 35 institutes have deployed a total of 49 systems. Of these, 5 are regional centers resulting from a call by EPSRC in 2012.

Sources (C.5)

“A Strategic Vision for UK e-Infrastructure: A roadmap for the development and use of national advanced computing, networks and data”, http://www.bis.gov.uk/assets/BISCore/science/docs/S/12-517-strategic-vision-for-uk-e-infrastructure.pdf.

C.6 United States

Several agencies within the United States government make major CI investments, including the National Science Foundation (NSF), the Department of Energy (DOE), the Department of Defense and the intelligence community. Here we describe NSF and DOE programs that support scientific research at universities and national laboratories.

C.6.1 National Science Foundation

The nexus for NSF’s foundation-wide CI investments is the Division for Advanced Cyberinfrastructure (ACI), which “coordinates and supports the acquisition, development, and provision of state-of-the-art CI resources, tools, and services essential to the conduct of 21st century science and engineering research and education.” ACI-funded resources and programs include high-performance computers, storage facilities and data repositories, high-speed networks, software and middleware tools, as well as extensive education/outreach/training. All resources and programs are individually competed under separate solicitations, with proposal evaluations by peer review.

NSF/ACI funds a number of individually competed “Service Provider” (SP) programs to multiple institutions to operate specific resources, primarily HPC resources. One of NSF/ACI’s programs, called XSEDE (eXtreme Science and Engineering Discovery Environment) and led by the University of Illinois with many national partners, provides integration across the SP programs. The XSEDE program focuses on providing common services to users – e.g., there is a single proposal process for requesting and allocating time across all SP resources, a central help desk, central user documentation, common accounts/accounting, and coordinated security processes. In addition, XSEDE funds and operates a private high-speed network that links NSF’s SP resources. Thus XSEDE in many respects represents NSF’s integrating framework.

The following are observations regarding how the NSF/ACI framework compares and contrasts to the situation in the Republic of South Africa (RSA).

- XSEDE is designed to integrate *multiple providers of similar services,* primarily compute providers, while RSA currently invests in essentially one provider in each of the three relatively distinct areas of compute, data, and networking. The multiplicity of providers creates a clear need for the common services/interfaces to the users that drives much of XSEDE's charter. In contrast, for example if RSA funds only CHPS as an HPC center, it need not develop common documentation/support/ allocations/accounting across multiple centers, but rather just operate CHPC effectively.

- Individual SPs remain accountable to NSF/ACI and not to XSEDE. While NSF mandates that SPs will adhere to certain common policies and procedures established by XSEDE, the XSEDE-SP interactions are based on communication and joint dialog, not management lines of authority.

- NSF conducts open competitions for the service providers, as well as the integrating framework organisation, so organisational structures must be fluid and adaptable. While the RSA government maintains accountability over its service organisations, it does not currently plan on open competitions. The RSA management structure can be simpler, because the participating organisations are known, fixed and small in number.

- While ACI funds DataNet and a growing set of programs doing work in digital repositories/preservation programs, these are not currently included in the scope of programs integrated by XSEDE. This should not necessarily be considered a model or lesson learned by NSF; in fact it is likely that NSF is moving from its legacy focus on computing to be broader in its scope, while RSA is doing this from the outset.

C.6.2 Department of Energy

Two offices within the Department of Energy, the Office of Science and the National Nuclear Security Administration (NNSA), operate large CI programs. In the context of integrating frameworks, these two programs have similar approaches to each other, which are significantly different from NSF's approach. These differences include:

- DOE invests in national labs and wants stability in those labs. There is accountability of the labs to the funding agency, but the expectation is for continuity of the labs and their roles. In contrast, NSF formally competes all its CI investments as well as the integrating framework organisation. NSF has recently moved towards longer awards designed to promote organisational stability, with a system award that has an option for a single renewal (system refresh) upon NSF approval but without competition, and NSF's XSEDE award is for five years, with an option to renew without competition.

- While NSF has a separate large award (XSEDE) to provide an integration framework across its resource investments, DOE does not have a separate program for this integrating function. The integration is provided by:

- Active involvement of DOE program officers in managing across different labs and providing the guidance necessary for integration,
- An expectation by the government of collaboration across all levels of the participating labs, with people and funding committed from the lab budgets to support this collaboration,
- Sponsoring programs (e.g. SciDAC and many of the exascale research programs) which are open to competitive proposals from all labs, with the results expected to be shared across the labs.

The DOE approach illustrates that with active involvement of government staff, collaboration can be established that facilitates integration across the multiple investments. There certainly is a competitive spirit amongst the labs, but a healthy spirit of competition exists in the context of shared high-level goals.

C.6.3 US National Networking Programs

There are interesting parallels between national networking programs in the US and SANReN/TENET in RSA. Internet2 is analogous to TENET in that it is funded by educational and research institutions to provide shared network infrastructure to the higher education community. Internet2 is a national, private non-profit organisation with its own Board of Directors, and is funded by its university members as well as by additional contracted networking services that leverage its extensive infrastructure. ESNet, funded by DOE/Office of Science, "provides services to more than 40 DOE research sites, including the entire National Laboratory system, its supercomputing facilities, and its major scientific instruments. ESnet also connects to 140 research and commercial networks, permitting DOE-funded scientists to productively collaborate with partners around the world." Similarly NSF/ACI funds the XSEDE program to operate a private network available to XSEDE users that currently connects XSEDE Service Providers at 10 Gbps or higher.

ESNet and Internet2 are separately managed and funded but have a history of collaboration, both in terms of developing technology and tools, as well as operating infrastructure. Interestingly, XSEDE has recently contracted with Internet2 which, along with regional network providers, provides the actual networking infrastructure for XSEDE's private network. So while the charters, management and funding are reasonably distinct for Internet2, ESNet and XSEDE, there is a significant level of collaboration between these organisations and there are mutually beneficial contractual relationships. However, even though all these networks are national and support researchers, they each target a reasonably well-defined user community and none would assert that they were *the* national US research network.

<u>Sources (C.6):</u>

http://www.nsf.gov/od/oci/about.jsp

http://science.energy.gov/ascr/

http://www.internet2.edu/

http://www.es.net

C.7 Canada

Canada makes substantial investments in CI for the research community at the national level by the Canada Foundation for Innovation (CFI, comparable to the US NSF) at the regional (provincial) level, and by universities. CFI funds an independent non-profit organisation, Compute Canada, to provide an integrating framework for its CI investments. Compute Canada has its own Board of Directors, Advisory Groups, and a modest-size management/administrative staff. Compute Canada then works with four “regional grids”, which align with Canadian provinces (Ottawa, Quebec, and partnerships of eastern and western provinces), as well as university partners that ultimately house/operate the physical resources. All CFI investments flow through and are under the direction of Compute Canada, however provinces and individual universities provide matching funding for resources at the universities/regional grids; thus there is shared responsibility for procurements and operations, with specific decision-making and management processes being evolved. (CFI invests ~$50M (Canadian), with provinces and universities investing an additional ~$75M.) The resources focus on compute, visualisation and storage resources, not networking or data services beyond storage itself. There is a central Canada-wide proposal and allocation process for access to nationally-funded resources, similar to that in the US and EU.

The Canada’s Advanced Research and Innovation network (CANARIE) was established in 1993 as a non-profit institution, funded largely by the national government (CFI). “CANARIE’s roots are in advanced networking, and CANARIE continues to evolve the national ultra-high-speed backbone network that enables data-intensive, leading-edge research and big science across Canada and around the world. One million researchers, scientists and students at over 1,100 Canadian institutions, including universities, colleges, research institutes, hospitals, and government laboratories have access to the CANARIE Network.” CANARIE works closely with Optical Regional Advanced Networks (ORANs), which focus on networking reaching directly into educational institutions, to form a national network. The ORANs are funded primarily by provincial governments and educational institutions (subscription fees).

Research Data Canada is “a collaborative effort to address the challenges and issues surrounding the access and preservation of data arising from Canadian research. This multi-disciplinary group of universities, institutes, libraries, granting agencies, and individual researchers has a shared recognition of the pressing need to deal with Canadian data management issues from a national perspective.” The Steering Committee is led by members of the National Research Council and has diverse stakeholder representation. The group’s efforts appear to be focused around Working Groups in Policy, Education and Training, Infrastructure, International Liaison, and Standards, Interoperability, and Interdisciplinary Liaison.

<u>Sources (C.7)</u>

Compute Canada https://computecanada.ca/index.php/en/

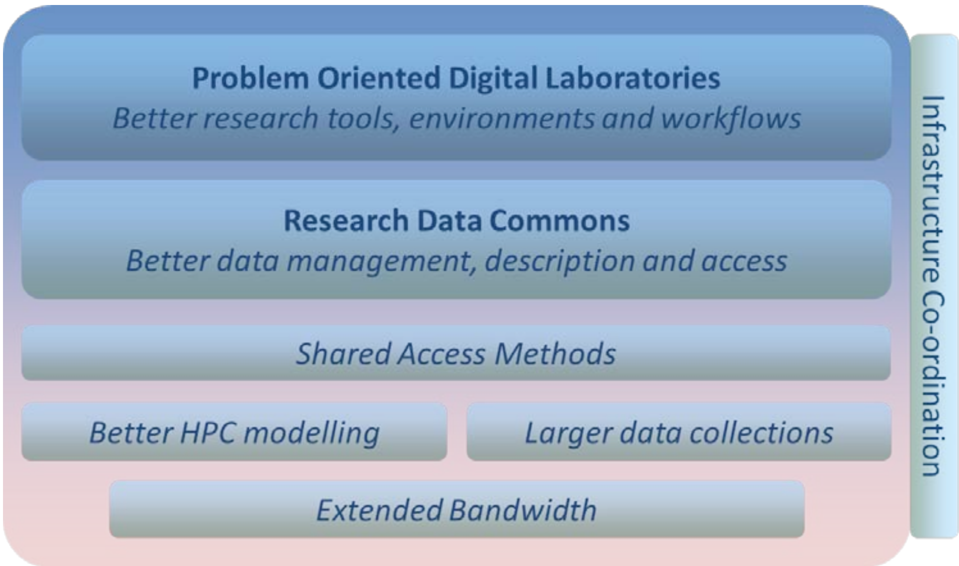
CANARIE http://www.canarie.ca/en/about/aboutus

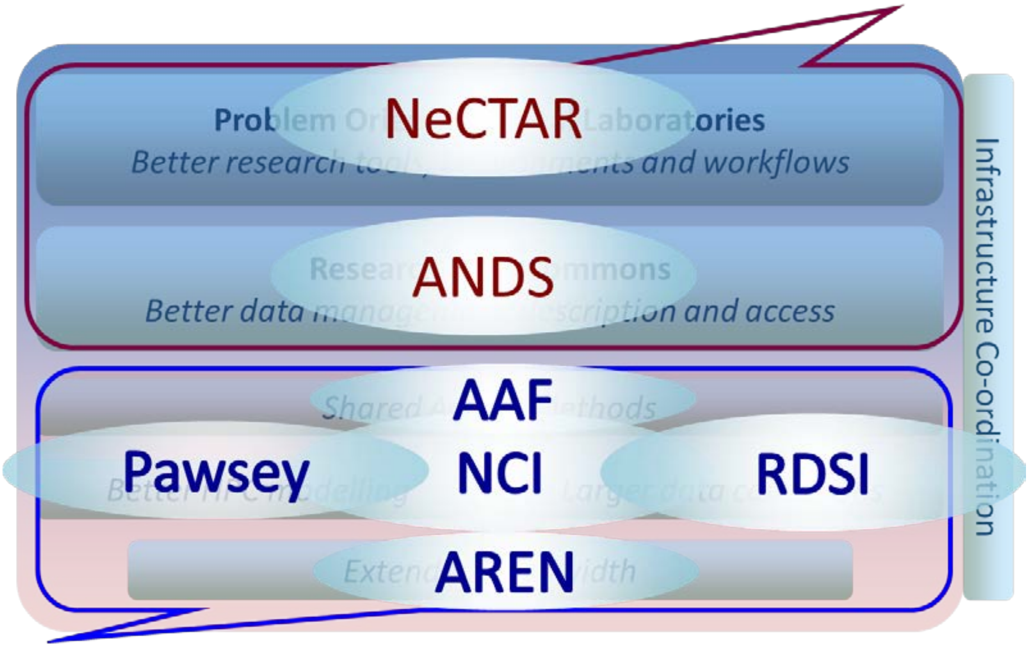
Research Data Canada http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/eng/

C.8 Australia

National CI is supported through the Department of Industry, Innovation Science, Research and Tertiary Education (DIISRTE). In 2004, the federal government committed to a 7-year program through 2011 (with total funding of ~Australian$542M) based on the National Collaborative Research Infrastructure Strategy (NCRIS). The goal was to "support major research infrastructure to encourage collaboration between the research sector, industry and government in Australia to conduct world-class research. NCRIS is designed to provide Australia's research sector with on-going access to high-quality, operational research infrastructure facilities. This will ensure that Australian research continues to be competitive and rank highly on an international scale." These initiatives have been extended for 2013-2014 as a program titled "Collaborative Research Infrastructure Strategy". In addition, the "Super Science Initiative was announced in May 2009 and will contribute $1.1 billion to priority areas of Australian research until 2013. $901 million of this money is for research infrastructure that will make a lasting contribution to Australian science." Major investment areas under Super Science initiative include Space and Astronomy, Marine and Climate Change, and Future Industries (Biotech, Nanotech and ICT).

The NCRIS investment strategy is a layered approach as shown in the figure below, with software and data management tools building on top of infrastructure facilities at the bottom (from Francis, 2012). Various programs address different elements of the strategy, as illustrated in the second figure. Some of these infrastructure programs have counterparts in South Africa's DST repertoire - e.g. the Australian Research and Education Network (AREN), the National Computational Infrastructure (NCI) and the SKA-affiliated Pawsey computer center, and the Research Data Storage Infrastructure (RDSI), similar to DIRISA/VLDB. Interestingly the Australian National Data Service (ANDS) is analogous to the data 'services' side of DIRISA as discussed in Section 5.3. The Australian Access Federation (AAF) program provides mechanisms for trusted access, and the National eResearch Collaboration Tools and Resources (NeCTAR) program has four program areas: Virtual Laboratories, Research Cloud, eResearch Tools and National Servers Program. Awards are made at a range of levels to implementing organisations for various program elements.





In terms of integrating frameworks across various elements, the management and coordination occurs within the government funding agencies at the program level - e.g. NCRIS, Super Science or CRIS. The Australian e-research and innovation council (AeRIC) provide an overseeing guide to these programs, reporting through the National Research Infrastructure Council (NRIC). There is a higher level advisory committee, the Australian Research Committee (ARCom), which provides integrated and strategic advice on investment across the science, research and innovation system, including in the areas of human capital, infrastructure and collaborative activities. This committee has very broad representation from many sectors of government, universities and industry, and is chartered to take a "whole-of-government" perspective on research priorities and investments.

While not directly related to research CI, the 2011 report *"National Digital Economy Strategy: Leveraging the National Broadband Network to Drive Australia's Digital Productivity,"* outlines a national strategy for pervasive broadband networking as a critical means to create positive economic benefits and improve the quality of life.

Sources (C.8):

http://www.innovation.gov.au/science/ResearchInfrastructure/Pages/default.aspx

http://eresearchau.files.wordpress.com/2012/11/03_rhys_francis.pdf)

"Australian eResearch Infrastructure: Advancing Discovery in an Information Rich World," Rhys Francis.http://eresearchau.files.wordpress.com/2012/11/03_rhys_francis.pdf

Australian Research and Education Network (AREN, http://www.innovation.gov.au/science/ResearchInfrastructure/FundedProjects/Pages/AREN.aspx)

National Computational Infrastructure (NCI, http://nf.nci.org.au/)

Research Data Storage Infrastructure (RDSI, http://rdsi.uq.edu.au/)

Australian National Data Service (ANDS, http://www.ands.org.au/)

The Australian Access Federation (AAF, http://www.aaf.edu.au/)

National eResearch Collaboration Tools and Resources (NeCTAR, https://nectar.org.au/)

"National Digital Economy Strategy: Leveraging the National Broadband Network to Drive Australia's Digital Productivity," http://www.nbn.gov.au/files/2011/05/National_Digital_Economy_Strategy.pdf.

"Reflections on Australia's e-research infrastructure in an international context," Andy Pitman, http://www.ccrc.unsw.edu.au/news/news/2012-03-27_reflectionsoneresearch.html.

C.9 Singapore

Although its focus is more oriented to the general public and industry, as opposed to supporting scientific research, the 2010 report *"Realising the iN2015 Vision: Singapore, an Intelligent Nation, a Global City, Powered by Infocomm"* is an important illustration of a country's commitment to infocomm (networking and IT services) and its critical economic role. The national strategy, to be implemented via the Infocomm Development Authority (IDA), is "to establish an ultra-high speed, pervasive, intelligent and trusted infocomm infrastructure; to develop a globally competitive infocomm industry; to develop an infocomm-savvy workforce and globally competitive infocomm manpower; and to spearhead the transformation of key economic sectors, government and society through more sophisticated and innovative use of infocomm." A large focus is on pervasive high-speed network penetration throughout society, but the effort also includes access to shared computing resources (e.g., grids and cloud infrastructures), spurring innovation in industry, and training an IT-savvy workforce.

C.10 Taiwan

The Taiwan National Science Council and Ministry of Education established the National Center for High-Performance Computing (NCHC) with the mission to be "Taiwan's premier HPC resource provider by supporting local academia and industry with cutting-edge hardware and software resources, advanced R&D and application development, and professional training". The NCHC is Taiwan's primary facility for HPC-related resources including large-scale computational science and engineering, cluster and grid computing, middleware development, visualisation and virtual reality, data storage, networking, and HPC-related training. The NCHC is responsible for the operation of the national 20 Gbit/s Taiwan Advanced Research and Education Network (TWAREN) R&D network. TWAREN provides high-speed networking to Taiwan's national labs and universities, as well as coordinates and develops networking technologies.

The NCHC is one of eleven national-level research laboratories under Taiwan's National Applied Research Laboratories (NARLabs). NARLabs is a non-profit independent institute, with the main missions: to establish R&D platforms; to support academic research; to promote frontier science and technology; and to foster high-tech manpower. Funding for NCHC and the other NARLabs is primarily from the Taiwanese government.

Sources (C.10):

NCHC http://www.nchc.org.tw/en/

NARLabs http://www.narl.org.tw/en/

C.11 Korea

The Korea Institute of Science and Technology Information (KISTI) is a government-funded research institute and is the primary national organisation for research CI. KISTI has a broader mission than some research CI organisations, including “(a) contribution to the promotion of national science, technology and industry, (b) comprehensive collection, analysis and management of data for science, technology and industry, (c) research on technology, policies and the standardisation of information management, and (d) development and management of infrastructure for R&D on science and technology.”. Some of these roles are clearly driven by industry needs and economic competitiveness. As part of its R&D infrastructure role, KISTI operates a supercomputing center with significant resource capabilities as well as research initiatives, and a high-performance research network (KREONet) that builds on the capabilities of 16 regional networks.

Sources (C.11):

http://en.kisti.re.kr/mission-vision/

http://en.kisti.re.kr/information-analysis/