Towards 2011: An IT strategy for SAEON

- By Wim Hugo , SAEON IT consultant

SAEON recently completed a round of Information Technology planning on a broad front. The process followed and the results of this process are discussed briefly in this paper, concentrating on the Information Technology Strategy that resulted from the work.

SAEON is, by definition, a partly virtual organisation that relies on stakeholders, associated nodes, and collaboration with partner organisations to achieve its goals. Its mandate is wide-ranging, and its goals are ambitious. These factors add to the complexity of the environment in which SAEON must deliver information technology infrastructure to its stakeholder communities.

From an information technology perspective, SAEON thus presents an enormous challenge. These challenges range through the whole spectrum of information systems governance in an organisation, across all its functions, and at many different levels of detail. The partial list below provides some sense of the complexities involved:

(1) There is a growing emphasis world-wide on ***proper information systems governance*** in organisations, but the standard approaches and best practice require deep pockets and are designed for the multinational corporations and large government departments that dominate the information technology landscape. SAEON is not a large organisation, even in local terms. One of the major challenges, even ***before*** we embark on some form of governance, is to design a governance framework that suits the resources, size, and objectives of SAEON. We need a framework that is ***necessary and sufficient,*** but no more. See figure 1.

(2) Likewise, all properly managed information technology projects have some form of ***systems engineering and program oversight*** attached to it. Again, ***it is easy to do too much***, and most of the received wisdom is aimed at large, well-funded projects. We are challenged to define a systems engineering framework that is adequate, defensible, and effective at delivering information technology services and products.

(3) SAEON needs to provide or facilitate information technology products, services, and infrastructure to a wide ***variety of stakeholders and collaborators***, across many different platforms and computing environments, and using a mixture of in-house, outsourced, and 'negotiated' resources.

(4) The data, information, and knowledge that SAEON works with ranges across a many levels of detail, encompass a wide variety of formats, and derive from a variety of environmental systems and sciences.

(5) A high degree of ***interoperability needs to be maintained*** to foster collaboration and exchange with other systems, and in this regard the worldwide trend to standardization and specification-driven systems development must steer SAEON's own systems development efforts. This type of focus is imperative if the goals elaborated by Liz Gavin

and Craig Schwabe (See '"A view from the top" – how do we (get to) see our country?', elsewhere in this publication) are to be achieved.

A governance framework for information systems management

Our starting point, then, has to be the creation of a framework for governance and management of the information systems function within SAEON (and, because of the nature of the organisation, extending into stakeholder organisations at times). SAEON has adopted a governance model that unites business process engineering, systems engineering, and organisation development into a consolidated view.

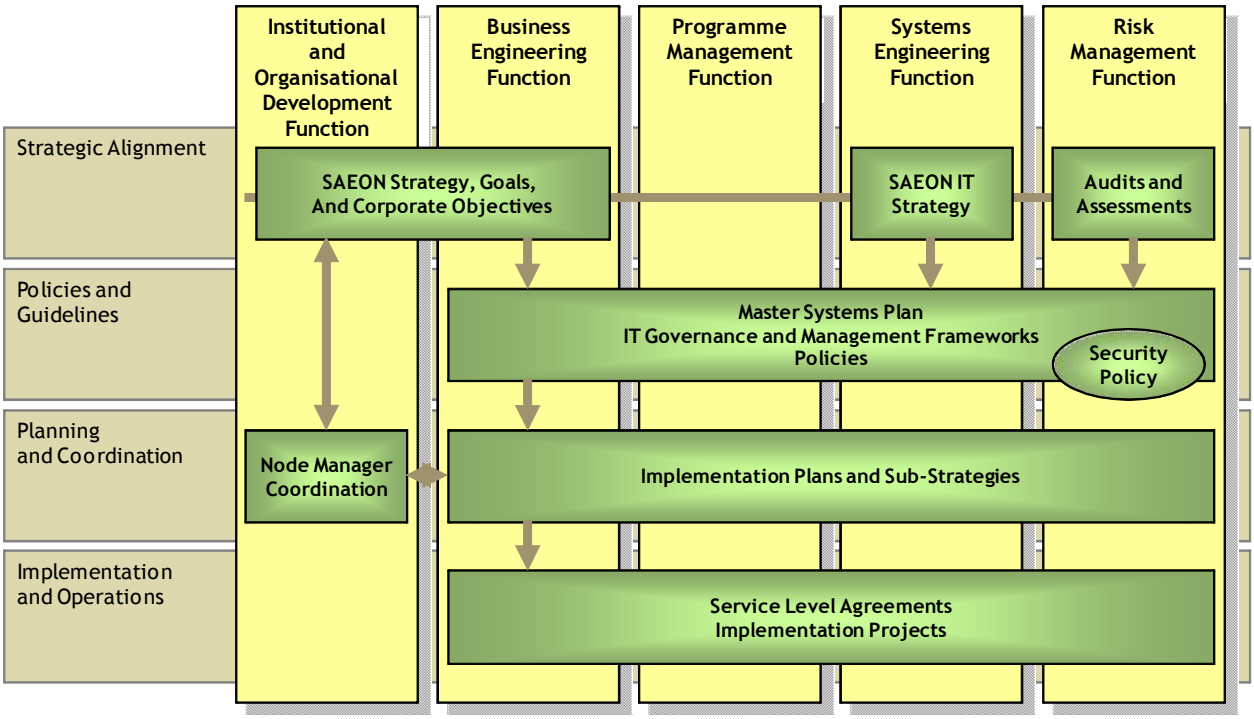

Figure 1: Governance Framework for Information Systems Management

The main objective of the framework is to provide and maintain systems in support of the business objectives of SAEON, taking environmental, organisational, and technical drivers and impacts into account. Strategies, architecture, policies, processes, and implemented systems need to be delivered within a governance and management framework ***that limits the technical, financial, and legal risks*** of SAEON to acceptable levels.

For the ***Systems Engineering Function*** specifically, we define four levels of detail and scope of activity, summarised in the table below:

Strategic Level	Strategic assessment of the concerns, drivers, and business imperatives impacting on systems and technical architecture, and alignment of these with systems through analysis, solution selection, and planning (**IT Strategy**). **High-Level Audits** to determine compliance with a number of aspects and to assess risks.
Governance Level	Definition of technical and domain architectures and implementation plans to address strategic goals. This forms one of three aspects of a **Master Systems Plan**: the others deal with business architecture and organisational development. Development of Governance Frameworks; **Policy** document(s) detailing systems engineering processes and methodology. **Templates and examples** of standardised artefacts (such as process definitions, enterprise models, class definitions, and so on). A **Quality Assurance** approach aimed at achieving the process(es). Cross-cutting abstract specifications ("**Perspectives**"). Publication of **Position Papers** on important aspects of abstract specifications, strategies, and perspectives.
Planning and Coordination Level	Development and dissemination of implementation plans and documented sub-strategies. Delivery of life cycle documentation (URS, SRS), based on research, engagement with stakeholders, vendors, specialists, and experience. Maintenance of a shared environment in which documents can be deposited and maintained under configuration control.
Execution Level Implementation and Operations	A systems engineering database, maintaining the interdependencies between system elements, versions, releases, and builds. A dictionary that maintains meaning and its interrelationships to data models and implementations, both for data elements and process elements. Provision of SLA/ contracting inputs. Technical (i.e. non-contractual) management of delivery against SLA's and development contracts.

Governance frameworks

One of the major internal goals of the governance and systems engineering function is to improve the standing of SAEON in respect of the 'Capability Maturity Model' (CMM) (1), which measures the extent to which the systems management and delivery capacity of the organisation has been stabilised.

The model proposes that the software engineering process can broadly be categorized into 5 stages of maturity – ranging from a 'chaotic' state to an 'optimised' one. While not perfect, it is clear that the model provides a framework for a measured process of goal-driven improvement – which lies at the heart of most quality assurance and improvement philosophies.

There are five levels to the model. The gist of each level can be summarised as follows:

- Level 1 (***Initial***): At maturity level 1, processes are usually ad hoc and the organisation usually does not provide a stable environment …
- Level 2 (***Repeatable***): Software development successes are repeatable. The organisation may use some basic project management to track cost and schedules…
- Level 3 (***Defined***): Processes are well characterized and understood, and are described in standards, procedures, tools, and methods …
- Level 4 (***Managed***): Using precise measurements, management can effectively control the software development effort …
- Level 5 (***Optimising***): Focuses on continually improving process performance through both incremental and innovative technological improvements …

By analogy, the CMM has been extended to serve as a management framework for all aspects of SAEON Information Systems Governance. Refer to the example taken from SAEON's meta-data management strategy below.

A process for strategic information systems planning

SAEON's Information Systems Strategy (2) methodology is based on the broad principles of two business tools: ***Strategic Planning Practice***, and on ***Systems Engineering Foundations***. The first determines our approach and the second the nature of our outputs.

Strategic planning compels us to understand (using any number of valid toolsets – see Figure 2) the current situation, as well as the environmental impacts that affect the status now and are likely to do so in future. Pitted against this are the objectives, needs, and requirements of the organisation: to some extent, the 'problem' for which a 'solution' must be found. This solution, while taking the environmental impacts into account, can rarely afford to be a complete ideal, nor can it afford to break completely with the past: resources and other constraints will typically not allow it. For this reason, we state an ideal solution and then, as a rule define, plans and measures that will move us from the current situation towards that ideal, while limiting the use of resources and the risk involved (Figure 3).

The output of this exercise is a technical one, leading to a set of design imperatives, policy level requirements, solution designs, and implementation plans.

Stakeholders	SAEON | DST/ NRF | Government Departments | Research Institutions | Suppliers | …
Concerns	Cost | Suitability to Task | Access | Flexibility | Specifications | Technology | …
Constraints	Principles | Standards | Policies | Resources | Guidelines / Best Practice | Status Quo …
Viewpoints	Functional | Information | Physical | Logical | Fail-Over | Operational | …
Tools	Knowledge | Patterns | Styles | Idioms | Methodology…
Perspectives	Security | Performance | Availability | Usability | Accessibility | Risk | Regulation | …

Figure 2: A Toolset for Information Systems Planning (3)

The toolset serves as a checklist, first and foremost, of the scope to be considered in planning for information systems on a strategic level.

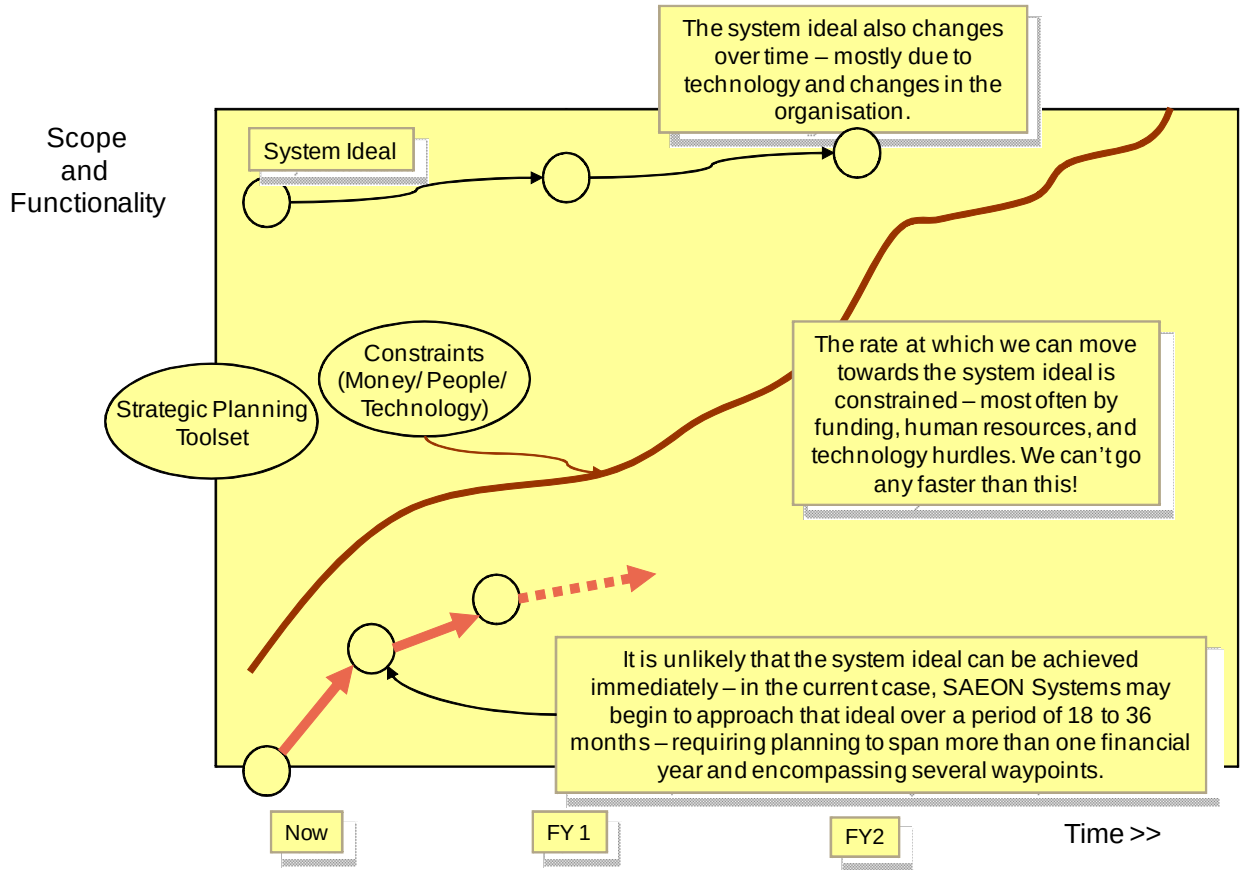

Figure 3: A Bird's-Eye View of the Strategic Planning Process

Resourcing options

One of the significant challenges faced by SAEON involves the procurement of human resources in the Information Technology field. There are several reasons for this, amongst others that the organisation is not large enough to sustain a critical mass of in-house resources, coupled to the fact that the organisation's focus and objectives does not always align with the career goals of IT professionals. For this reason, some time was spent on analyzing and selecting feasible options for provision of information technology resources, and making recommendations in respect of the functions that can best be supported.

The following generic procurement options are available:

1. In-house services – salaried personnel (includes "Selfsourcing" – the idea of supporting natural sciences professionals to contribute to the information technology resource pool with assistance from experts).
2. In-house services – contracted personnel. This is usually the least attractive option in the long run, unless the requirement is of a temporary nature.
3. In-house services – advisor-supported. Advisory services are distinguished from contracting by the specialist or expert nature of the service. It can be used with great effect in a mentoring arrangement.
4. "Insourcing" (independent service organisation, but bound to the client). These arrangements recognize that in-house personnel may not have enough work or scope in a typical organisation's IT environment – and allows a semi-independent existence that can also do work for third parties.
5. Outsourcing (independent organisation). This is a very common and familiar arrangement – but needs proper oversight to provide consistent value.
6. "Crowdsourcing" (small contracts to a large, but indirect supplier base). These are very effective in some circumstances, and the contracts can be paid or voluntary. Facilities to accelerate and manage interactions (4), (5), (6) are increasingly available and sophisticated.

Each of these options is to a lesser or greater degree suitable for the main functions associated with information technology management, governance, and provision. The following is a recommendation on the immediate, medium, and long-term sources of IT-related functions:

Aspect/ Element	**Immediate Term**	**6-18 months**	**18-36 months**
Strategic Alignment	Advisory	Advisory	Advisory
IT Governance/ Audit	Advisory	Advisory	Mentored
Planning and Coordination	Advisory	Mentored Insourcing	Mentored Insourcing

Implementation/ Life Cycle			
- Project Management	Advisory Insourcing	Insourcing Contractors	Insourcing Contractors
- BRS and URS	Advisory	Advisory CrowdSourcing	Advisory CrowdSourcing
- Test and Acceptance	Advisory	Advisory	Advisory
- Systems Development	SLA	SLA Self-sourcing	SLA Self-sourcing CrowdSourcing
- Hardware and Infrastructure	SLA	SLA	SLA
Content Provision	Salaried Knowledge Workers	Salaried Knowledge Workers CrowdSourcing	Salaried Knowledge Workers CrowdSourcing

SAEON's information systems strategy

The SAEON information systems strategy can be summarised as three ***alignments***:

(1) An alignment ***between the objectives of the organisation***, its stakeholders, and partner organisations ***and the systems available*** to them. Systems ultimately exist not for their own sake, but to support these objectives.

(2) An alignment with the ***expectations***, ***requirements***, ***and preferences of the user community***.

(3) An alignment with the ***standards, specifications, technology trends, and best practice*** that exist in the wider user community and technology space.

The information systems strategy identifies a series of programs that will achieve this alignment. These are summarised below:

Program	Discussion
Establishment of Processes	Processes for Governance, Systems Engineering, Procurement, Issue Resolution and Management, and
Communication Program	This includes a variety of publications, and capacity to engage with

	stakeholders and partner organisations. The program is supported by position papers, policies, guidelines, and published reports. Position papers, policies and guidelines have been completed and published: (1) Data policy (2) Meta-Data Management (3) Data Management at Nodes (4) Portal Re-use To be considered in addition for the short term: (1) Source Code Sharing (2) Knowledge Sharing
Risk Management Program	The scope includes (1) Disaster recovery, backup, and fail-over. (2) Supplier Diversification (3) Platform Diversification (4) Licensing Agreements (5) Management of Stakeholder Divergence
Reference Implementations and Capacity Building Workgroups	The following reference implementations are required: (1) EML (MetaCat and Morpho) (2) Data Products (GAP-3 or similar) (3) Project DB/ Meta-Data (Self-Sourcing) (4) Public Face (Self-Sourcing/ CoGIS/ Google Earth) Capacity building will be strengthened at node level with the establishment of 'Implementation Workgroups' – essentially combining the SAEON Systems Engineering role with knowledge workers and 'self-sourcers' at node locations who are interested in the systems aspects of their work. One group is foreseen for each reference implementation.
Collaboration and Sharing Program	Establishment of the collateral (documentation, agreements, source code packages and projects) to allow a number of typical collaboration arrangements. These will probably be categorized as : (1) Aimed at close cooperation with other data providers and meta-data repositories. (2) Aimed at very close cooperation with primary stakeholders, SAEON user community, and other data centers. (3) Data-set specific restrictions on use for commercial ends. (4) Data-set specific restrictions on use for private ends. Direct engagement with stakeholders, specifically SANBI, CSIR, GBIF, NCEAS, SAEOS, and others.
Systems Provision Program	Establishment, over time Full systems life cycle management of a number of systems required to serve SAEON's objectives. These categories are: (1) Collaboration and Sharing Systems (CoGIS)

	(2) Public Face Systems (3) Operational Systems: Data and Meta-Data Management, Dissemination, and Gathering Systems These systems need to be established within the architecture framework defined for SAEON (See Figure 3) and adhere to applicable international, national, and internal standards and specifications.

Guideline architecture for SAEON systems

One of the important outputs from the Strategic Planning process is an affirmation and refinement of the logical architecture that SAEON needs to establish in the next 18 to 36 months.

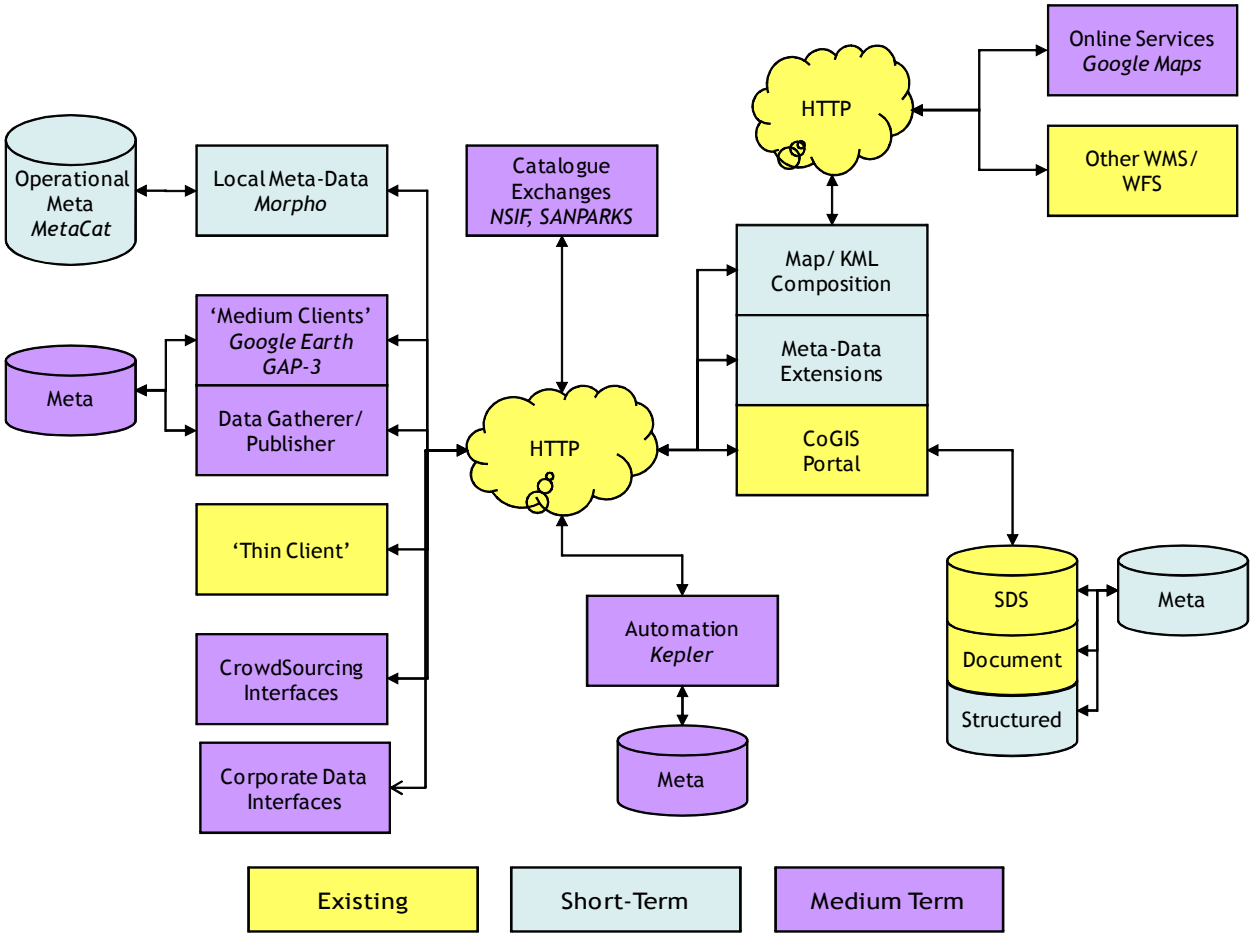

Figure 3: Guideline Architecture for SAEON Systems

The systems and subsystems identified in this architecture have the following roles:

Aspect	Discussion
CoGIS Portal	The CoGIS Portal comprises of several sub-system components: (1) A content management and access/ security layer (2) Client-side extensions for the management of meta-data, integrated searches, discovery and related functions. (3) Client-side and mid-tier components for the composition of maps from distributed sources, and service interfaces for these. (4) Service interfaces to external data sources, such as OGC-compliant services, Google/ Yahoo/ Microsoft Map sources, and similar. (5) A spatial data store, its associated services, and client-side mapping to visualize the available data sets. (6) A document repository for storage of data sets, documents, meta-data sets, and other information objects. (7) A structured data set repository for standards-based storage and dissemination of data. (8) A structured meta-data store (for example MetaCat. GeoNetwork or an extension of these).
Crowdsourcing Interfaces	These interfaces would provide (1) generic tools for the informal contribution of data to the CoGIS Portal environment (2) specific tools for the micro-contracting of data cleanup and meta-data extraction services.
Thin Client	Specifically aimed at high-quality, CoGIS-integrated web mapping interface. Required by meta-data, spatial data, and custom home pages.
Stand-Alone Data Gatherer	A packaged data gathering application with spatially enabled capabilities, for the gathering of structured, spatially referenced data in the field. This sub-system implies a number of components: (1) A client application or component for the dissemination and presentation of data; (2) A filtering component that can be used to 'author' the extraction of structured data and its publication/ export. (3) A map composition component that can be utilized to package maps to collate with the structured data obtained in (2). (4) A component for the synchronization/ upload of data gathered in the field.
Medium Clients	Web-enabled, stand-alone applications in the style of Google Maps, aimed at extended analysis and display capabilities for spatial and structured data. These are used primarily as data dissemination vehicles.
Medium Client Meta-Data	Both stand-alone data gatherers and medium client implementations can be built on the same basic architecture and components, and will require local meta-data repositories for its

	configuration and data descriptions.
Operational Meta-Data: MetaCat	One or more operational meta-data repositories will be required. At present, we would make use of the SANPARKS node, and establish an additional node at the Egagasini site for use by Cape Town and Grahamstown nodes.
Local Meta-Data Management (Morpho)	Client installations can be made in as many locations as needed.
Catalogue Exchanges	These components (possibly on both sides of a transaction) will be required to upload one or more meta-data records from Meta-Cat to the formal CoGIS repository, and to transfer catalogue information between repositories.
Automation (Kepler, but not limited to it)	Implementation of automated processes (especially in respect of modeling/ data processing). Kepler is a domain-standard specification, bit we need to align with mainstream business process automation standards where appropriate.
Corporate Data Interfaces	Mechanisms are needed for the definition, configuration, and population of data structures for corporate management. This is a SAEON-only requirement – examples include project databases, etc.

Conclusion

SAEON has concluded a comprehensive review of the impacts and drivers that shape its Information Systems Strategy, and have identified a series of interrelated programs aimed at alignment with its stakeholders, user community, domain(s) of operation, and corporate objectives. Programs are already underway to create the guideline architecture through extension to the CoGIS Portal environment and the establishment of a number of supporting systems.

Governance measures, processes, capacity building programs and other supporting initiatives are in place, and SAEON looks forward to facilitating increased cooperation and sharing of knowledge, information, and data with a growing number of collaborators.

Glossary

Acronym	**Label**	**Description**
CoGIS	Collaborative Geographic Information System	A content management, meta-data management and spatial data infrastructure aimed at providing the primary interface for SAEON stakeholders to publish and share data.
CoSAMP	Collaborative Spatial Analysis and Modelling Platform	A forerunner of CoGIS that provided a wider, knowledge management-directed scope.

COTS	Commercial, Off the Shelf Software	Commercial software for which proprietary licenses are purchased prior to use and for which source code would generally not be available.
GBIF	Global Biodiversity Information Facility	An international organisation that focuses on making scientific data on biodiversity available via the Internet using web services.
EcoGRID		EcoGrid is a next generation internet architecture for data storage, sharing, access, and analysis. It combines the features of a Data Grid for storage of ecological data and a Compute Grid for analysis and modeling services.
EML	Ecological Markup Language	An internationally supported XML standard for ecological meta-data.
FOSS	Free and Open Source Software	Software licenses that allow code re-use, usually governed by one of several public-domain license agreements. Sometimes (but not always) free.
GeoNetwork		Open source, free software for the management of ISO-compliant spatial meta-data records.
GEOSS	Global Earth Observation System of Systems	"GEOSS seeks to connect the producers of environmental data and decision-support tools with the end users of these products, with the aim of enhancing the relevance of Earth observations to global issues."
Kepler		Is an open source software tool that allows scientists to design scientific workflows and execute them efficiently using emerging Grid-based approaches to distributed computation
Morpho		An open source data management software tool for ecologists. It provides a way for ecologists to share data by defining a common structure (an XML file) to document their data (i.e. create metadata) so that other ecologists and/or software programs can correctly interpret the data.
NCEAS	National Centre for Environmental Analysis and Synthesis	
NSIF	National Spatial Information Framework	Responsible for establishment and maintenance of South Africa's national spatial information framework – primarily by providing a spatial meta-data management function/ repository as mandated by law.
OpenDAP	Open-source Project for a Network Data Access Protocol	A data transport architecture and protocol widely used by earth scientists. The protocol is based on HTTP and the current specification is OpenDAP 2.0 draft. OpenDAP includes standards for encapsulating structured data, annotating the data with attributes and adding semantics that describe the data.
Sparrow		Sparrow aims at combining algorithms and techniques from logic-based knowledge representation and databases into a single, open-source toolkit. Sparrow will also include new approaches to handle semantic information in scientific applications.
THREDDS	Thematic Realtime Environmental Distributed Data Services	Is developing middleware to bridge the gap between data providers and data users. The goal is to simplify the discovery and use of scientific data and to allow scientific publications and educational materials to reference scientific data.
URS-I URS-II	User Requirements Specification versions (I) and (II)	User requirements published for CoSAMP (Aug 2005) and CoGIS (Sep 2007) respectively.

References

(1) The Capability Maturity Model was developed by the Software Engineering Institute (SEI) at Carnegie Mellon University in Pittsburgh - http://www.sei.cmu.edu/cmmi/.

(2) Hugo, W.J. "SAEON: Information Technology Strategy – towards 2011", Internal Publication, June 2008. (Available on request from SAEON).

(3) Rozanski, N and Woods, E: "Software Systems Architecture", Addison-Wesley, April 2005.

(4) "Mechanical Turk" – a service provided by Amazon (http://www.mturk.com/mturk/welcome)

(5) Guru.com – a service to connect freelance service providers and clients (www.guru.com).

(6) Elance.com – a service to connect freelance service providers and clients (www.elance.com).

Annexure: Example of a governance framework for meta-data management

Meta-Data Framework						
		Initial Phase	**Repeatable**	**Defined**	**Managed**	**Optimised**
		Level 1	**Level 2**	**Level 3**	**Level 4**	**Level 5**
ST	**Standards Implementation**	Alignment is arbitrary and ad-hoc. Awareness exists of applicable standards, but implementation is patchy	A set of meta-data standards are defined and reference implementation software identified to maintain meta-data sets with	A sub-strategy is available and the processes required have been documented/ implemented.	Extent of compliance can be measured	Overlap between standards are addressed by software
DE	**Depth of Implementation**	No implementation exists.	Minimum requirements as stipulated by law are implemented and mandatory/ obligatory parts are provided	Sufficient definition exists to profile a standard and to decide which aspects will be optional in future.	Extent of compliance can be measured	Optional meta-data elements are implemented and maintained by practitioners.
SO	**Software Implementation**	No software is in use or no consensus is achieved.	A set of meta-data standards are defined and reference implementation software identified to maintain meta-data sets with.	Portal implementations allow multiple software outputs to be linked to a single data set.	Portal implementations allow searches across multiple meta-data sets and implements interfaces to all important catalogue/ harvesting functions.	Overlap between standards are addressed by software

IN	**Interoperability**	No interoperability	Standardised software allow standards-based exchanges on an ad-hoc basis with external applications.	Meta-data interoperability requirements are defined and translated into a set of portal functions.	Tasks such as harvesting from participating portals are routine and can be managed by using portal functionality.	Data and Processes are automated
SE	**Semantic Web**	No ontology implementations or implementations are ad-hoc.	Node managers agree on a standardised implementation of ontologies.	Implementation in the portal is defined and allows multiple, domain-specific ontologies to be registered and used.		OWL-type semantic translation is available in systems

Key	Current Status	Realistic Goal