

## **A suitable technical platform for South Africa's Risk and Vulnerability Atlas**

- *Wim Hugo, SAEON*

The project team of the South African Risk and Vulnerability Atlas (SARVA) recently decided to approach the South African Environmental Observation Network ([SAEON](#)) for assistance in establishing an information technology platform for the Risk and Vulnerability Atlas.

The main reasons for this are:

- the shared benefits of collaborative development (most infrastructure of the kind required by the Atlas already exists, or is also required by parallel ventures, and are often funded by the same sponsors), and
- the reduced risk of sporadic development, the fact that support resources are often difficult to come by, and other sustainability issues that arise with a less collaborative approach.

The Atlas benefits from the fact that shared development programmes are already under way for SAEON itself, for the South African Earth Observation Strategy platform (SAEOS), and other smaller developments (the Nairobi Clearinghouse Mechanism, the World Data Centre for Biodiversity and Human Health, amongst others<sup>1</sup>).

It is already possible to respond very rapidly to new infrastructure requirements, and the process to be able to accommodate new requirements will probably be streamlined even more in future.

At the root of the infrastructure being built at present are high-level requirements placed on a shared platform for the provision of meta-data driven, interoperable systems for data management. The platform needs to make provision for

- support for a number of meta-data standards,
- support of a variety of data formats and data sources,
- a wide variety of stakeholder requirements and domains, and
- different levels of data use: from basic data provision and preservation to value addition and decision support.

Often overlooked is the requirement for minimum diligence in the creation of the infrastructure, because by definition, the shared infrastructure and all documentation pertaining to it are open source or non-restricted public goods. This imposes criteria in terms of the quality of design documentation and specifications, and the quality of acceptable coding/ development work that is produced.

---

<sup>1</sup> COASTAL MANAGEMENT INFORMATION SYSTEM; NATIONAL ANTARCTICA DATA CENTRE; AGULHAS-SOMALI LARGE MARINE ECOSYSTEM DATA CENTRE; AFRICAN COELACANTH ECOSYSTEMS PROGRAMME DATA CENTRE; MILLENNIUM ASSESSMENT SGA

Each of the supported initiatives has unique requirements, but in broad terms, the initiatives all share a basic requirement for meta-data and data management that is addressed in this document. It also shares requirements for data processing, presentation, and visualisation that can be shared.

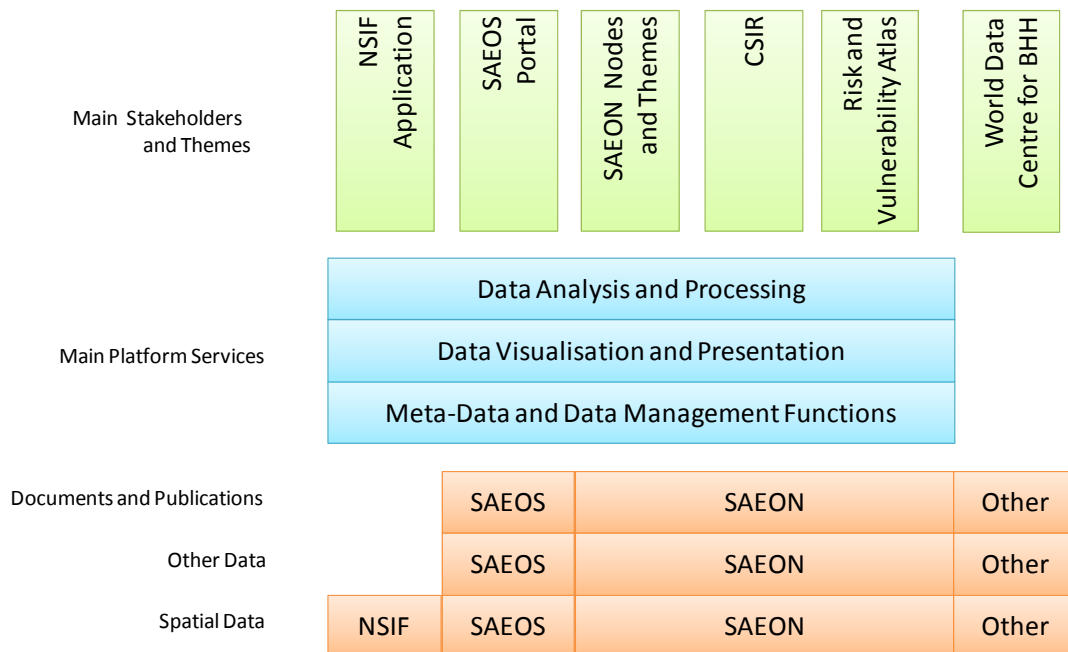
Furthermore, many of the meta-data repositories that these initiatives need access to (whether these are dynamically searched on demand or whether the meta-data repositories are replicated centrally) are the same. For example, repositories on biodiversity are required by SAEOS, SAEON, SARVA, WDC-BHH, and several other organisations. There is no need to duplicate the functionality and effort required to expose and integrate these repositories into portal environments where they can be better utilised.

Finally, several of the typical platform tools that are required to establish a domain or community-specific view on the meta-data and data repositories can be shared. These include the establishment and management of user communities and control to access and visibility of portal resources, collaboration environments, facilities for monitoring of portal resource usage and value, collaboration support, and many more.

**Technical discussion – conceptual architecture**

The diagram presents a summary of a meta-architecture on three levels: the communities of use, the portal functions, and the supporting meta-data and data repositories.

## Mandates and Architecture



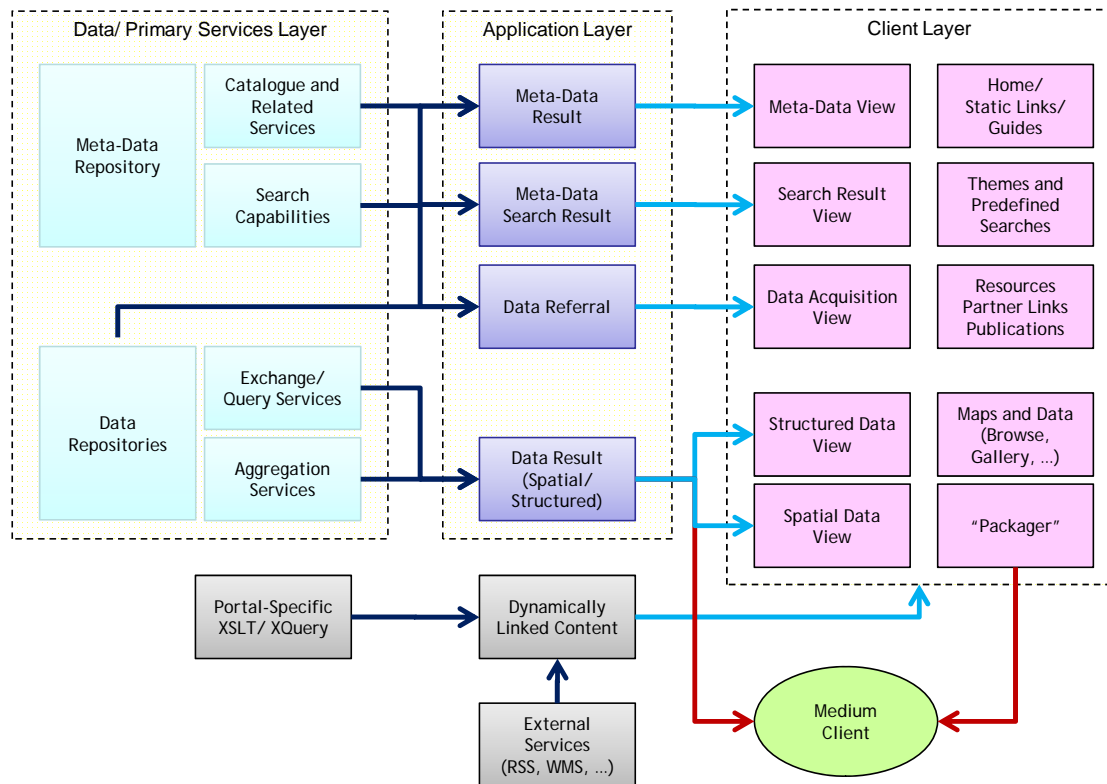
The notes and diagrams below provide a blueprint for a feasible architecture, based on the existing CoGIS platform, and making copious use of available open source components,

especially from the stakeholder community and the Global Earth Observation System of Systems (GEOSS).

In this architecture, maximum use is made of existing CoGIS components.

1. The backbone of the architecture, as a set of ***primary services and data layer***, will be provided by CoGIS, specifically the following:
  - a. A meta-data repository, capable of managing all appropriate meta-data standards through:
    - i. Registration and maintenance of automated harvesters to replicate meta-data to the SAEOS repository;
    - ii. Validation of individual meta-data records against appropriate standards;
    - iii. Limited editing and maintenance of records (applicable if SAEOS is the primary source of the meta-data record);
    - iv. Provision of harvesting services to other collaborative portals.
  - b. Search capabilities against the meta-data repository, implemented as one or more services. These services must support at least the standards recommended by the Group on Earth Observations ([GEO](#)), so as to enable interoperability with other GEOSS components (R-004);
  - c. A data repository, capable of
    - i. storing spatial data in one of the following default map server environments:
      1. GeoServer,
      2. MapServer, or
      3. ArcGIS.
    - ii. Storing publications, images, reports, and structured data sets in a variety of formats;
    - iii. Allowing services to retrieve these items and to provide them to users as automated data acquisition links;
    - iv. Allowing map services to provide spatial data to the portal itself, and to other requesting systems or users.

## Architecture: Prototype



2. CoGIS will be extended to support a number of **back-end interoperability processes**, as follows:

a. EML, FGDC, and ISO 19115:

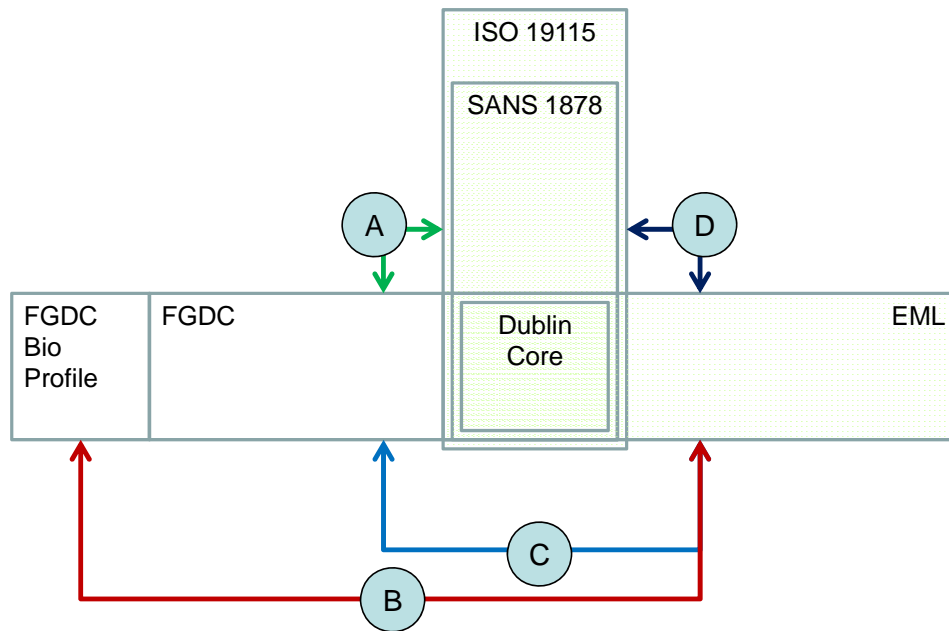
- i. translation between EML and the biological profile of the FGDC;
- ii. ability to align FGDC with the ISO 19115 spatial meta-data standard. Some work in this regard is under way at the USGS (O-031).
- iii. These translation artefacts (XSLT) will be extended and utilised to provide a seamless transition between meta-data clearinghouses utilising the standards applicable to their domain and/or jurisdiction. See figure below - "Standards Translations".
- iv. In doing so, US-based clearinghouses will be able to harvest meta-data unique to SAEOS and translate these to FGDC with biological profile extensions, and SAEOS will be able to harvest meta-data from US clearinghouses, and translate these to SANS 1878/ EML or ISO 19115/ EML, as required.

b. Darwin Core needs a crosswalk to EML, in both directions. This will enable the unification of searches against SABIF/ GBIF – type species occurrence

repositories with searches against all other meta-data *in a single search operation*.

3. An **application layer** is required, to marshal services and data in support of the following:

### Architecture: Standards Translations ('Crosswalks')



A: Translation between FGDC and ISO19115 / SANS 1878 (Spatial Meta-Data)  
 B: Translation between FGDC Biological Profile and EML  
 C: Translation between FGDC and EML 'Spatial Coverage Element'  
 D: Translation between ISO 19115/ SANS 1878 and EML 'Spatial Coverage Element'

- a. A consolidated set of information pertaining to a single meta-data record;
- b. A result set of meta-data records, resulting from a search against the clearinghouse repository, either by a user or a system;
- c. A data referral, as one of the following options:
  - i. A local data set/ content item (hosted in the data repository);
  - ii. A pointer to a remote URL, as provided by the associated meta-data;
  - iii. A page containing contact information whereby the data resource or content item can be obtained, should a URL not be available.
- d. A data result, which can be a spatial data set offered as a Web Map Service, or a structured data set offered for download, or a content item offered for viewing/ download.

- e. In addition, the application layer handles aspects such as visibility of content to users and communities, users and their communities, the publication life cycle of the content referenced in the portal, and other security/ access/ and privilege related tasks. Note that harvested meta-data records, with minor exceptions, will all be considered as 'published', especially in prototyping.
4. **Dynamically linked content** or **portlets** can be marshalled by the client or by the application layer, typically by requesting RSS feeds, or invoking services, through HTTP. Examples of this behaviour include:
- a. Google Earth can be invoked as a backdrop for layers in a map viewer, even though the application layer need not be aware of its existence as a map service through meta-data.
  - b. RSS feeds can be invoked to provide links to external resources (images, publications) on the basis of a common identifier (such as LSID), without the application layer requiring any additional information about the service.
5. A **client layer** is required – this presents
- a. A combination ('mash-up') of the following:
    - i. Data streams obtained from the main function of the portal (meta-data, content, or spatial data), usually as XML to be transformed.
    - ii. Portlets or external linked content as appropriate for the view,  
typically for standard pages containing
      - 1. a single meta-data record, either as text or a map, or both;
      - 2. multiple meta-data records, either as text or a map;
      - 3. a content instance (map, portal item, download page);
      - 4. a composite map of several spatial data content instances, all of which are map services.
  - b. Typical static or less dynamic pages, representing the following generic groups:
    - i. Home pages, background to the portal, legal documents (disclaimers, privacy statements, conditions of use), guidelines and user support, and similar.
    - ii. "Theme" pages, which can be thought of as home pages devoted to a special partnership or theme, and makes substantive use of predefined map views, predefined searches, and predefined dynamic links.
    - iii. Sets of resources, publications, and links to partner and contributor sites.

6. A *medium client*<sup>2</sup> may be of some use, but this forms part of our ambitious goal. If implemented, it can benefit from packaging assistance (i.e. marshalling and shaping data sets for download and client-side analysis). In our task list for the project, a 'simple aggregator' is defined, capable of some client-side manipulation of spatial data representations, using technology such as SVG/ JavaScript or JavaFX, together with Ajax.

---

<sup>2</sup> In the CoSAMP/ CoGIS environment, a medium client is taken to be a feature-rich mapping and data analysis environment that enhances thin-client mapping tools through client-side plug-ins or extensions.