Author: Wim Hugo
Review: Johan Pauw
02 August 2016
Version: Draft

Content Deposit, Publication, and Discovery Resources

SAEON has established an ***Open Data Platform (ODP)*** for publication, discovery, dissemination, and preservation of Earth and Environmental Data with funding from NRF and DST. This platform hosts several portals[1] and gateways[2], including SARVA, The South African Earth Observation System of Systems (SAEOSS), the BioEnergy Atlas, and SAEON's own data portal. It also serves as a platform for hosting the South African Spatial Data Infrastructure (SASDI), and has been used for internationally funded exploratory work to establish Africa-wide prototypes for data management in the domains of biodiversity, human health, and socio-economic sciences.

The ODP operates on a principle of mutual benefit, and by design is capable of providing access to metadata and data or digital content across all portals and gateways, as well as allowing improvements and extensions funded by a specific initiative to be available to other potential users at low or no cost, depending on their requirements.

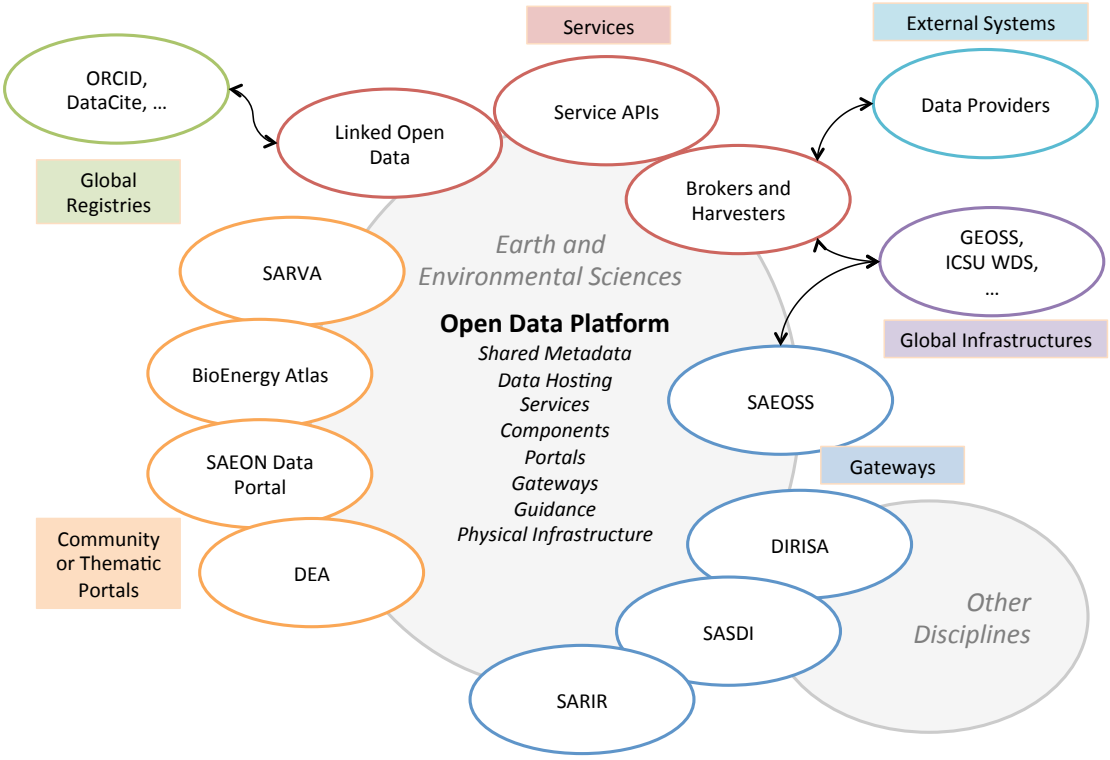

Figure 1: National and Global Research Data Infrastructure Context

[1] Portal: providing theme or community-specific access to resources
[2] Gateway: providing generalised user, service, and system interfaces to resources

The ODP allows any number of ***harvesters*** (capable of brokering several mainstream metadata standards and service protocols) to be configured for any portal that it supports, and as such can automatically synchronise metadata collections from as many contributors as needed, including SANSA Earth Observation Data Centre (EODC). With the operationalization of SASDI, this portfolio will grow to include most government departments.[3] Over time, several research institutions and contributors have been added to the portfolio as and when required.

SAEON now operates significant physical infrastructure in its own right (up to 75 TB of online storage, split between operational, test, and fail-over/ disaster recovery facilities), and the ODP allows rapid deployment of new portals and gateways at relatively low cost.

SAEOSS serves as a gateway to GEOSS (GEO[4] System of Systems) through the GEOSS Broker, exposing locally produced research outputs to a global user base, and in principle affording South African researchers access to globally available data sets.

The components for linking specific quality assured data sets to the ICSU World Data System[5] (WDS) are also in place, and once other aspects of sustainability and governance have been addressed, portals within the ODP can be accredited as members of the WDS. This accreditation serves as recognition by peers that the data platform is properly managed, serves quality assured data, and will be available for the foreseeable future.

Finally, the technical and licensing aspects of issuing data sets with Digital Object Identifiers[6] (DOIs) via DataCite have also been addressed. This allows data sets to be published internationally and for data sets to be cited reliably in scholarly publications. Data sets thus become formal scientific outputs that attract a citation index.

SAEON is also in collaboration with Meraka (since May 2015) to operationalize DIRISA. This collaboration will likely result in additional benefits to the user community, including

1. Additional storage and fail-over capacity to become available in a private cloud;
2. Increased access to deposited data and digital content from universities.

SAEON ODP infrastructure will initially serve some of the functions envisaged for DIRISA.

---

[3] SASDI Act, http://www.sasdi.gov.za/About/SDI%20Act.aspx

[4] Group on Earth Observations - http://www.earthobservations.org/index.php

[5] http://www.icsu-wds.org/

[6] http://datacite.org/

The discussion that follows highlights options for participation from a legal and technical perspective. We make mention at times of expected improvements and new releases: please consult the ODP Roadmap for more details.

Options for Participation

Options for participation depends largely on the factors detailed below.

1. **Funded Projects**: Examples include the Risk and Vulnerability Atlas (SARVA), and the BioEnergy Atlas. Funding allows construction and maintenance of a dedicated portal, support and maintenance, specific enhancements and improvements, active content management by domain experts, and interaction with the community through workshops and forums.
2. **Funded Infrastructure (Gateways)**: Examples include DIRISA, SAEOSS and SASDI – these are dedicated gateways to a specific user community, but funding is directed towards maintenance and support of hard and soft infrastructure, while content is largely managed and provided directly by the end users and automated harvesters. It is expected that many end user organisations will interact with the infrastructure through services.
3. **Public (Free) Services:** SAEON provides free services in the ODP, generally in three cases:
    a. **Mandate Alignment:** SAEON has some resources to assist data providers and depositors that wish to publish and archive environmental observation data. The constraining factors include human resources and non-standard metadata and data formats.
    b. **Funded Project Alignment:** These include SAEON collaboration projects and projects as described in (1). If data is broadly useful to one of these projects, SAEON will assist with its publication and description on a best effort basis, and can host data if required by the provider.
    c. **Small deposits and individual researchers:** SAEON can and will accommodate such research outputs in the near future.
    d. **Limits to Free Service:** SAEON supports free services on a best effort basis and within the limits of available human, infrastructure, and financial resources.

Each participant (depositor or depositing institution) enters into a data agreement with SAEON, which defines roles and responsibilities, service levels, license and access provisions, and cost implications - if any.

For a comprehensive review of the typical agreements that SAEON enters into, please have a look at our Data Agreement Template.

Typical Contributions by a Participant

The table below provides an overview of the typical contributions from a participant:

	Content	Funded Participation	Non-Funded Participation

	Portal or Gateway Presence	A dedicated portal or gateway is constructed and the participant determines layout, style, and structure.	A simple gateway or home page is provided by SAEON.
	Metadata	SAEON can assist with standardisation of metadata and implementation of a metadata management system that can be harvested automatically.	Metadata must be provided as a standardized service or in a standard metadata template.
	Static Web Page Content	SAEON can assist with establishment of a knowledge repository for the storage of static content and assist with its automated rendering.	Static content is provided and maintained by the end user within a web-based framework provided by SAEON.
	Data	Data set formats and services are agreed by SAEON and the depositor, and implemented by SAEON if it is not currently supported.	The depositor needs to select one of the data formats and service options already supported by SAEON.
	Other Digital Content	Standard digital content is supported, and specialized content can be accommodated as required.	Only standard file objects and selected media (video, audio) are supported. A specialised object ("Stub") can be used for limited knowledge management tasks.

Content Supported by SAEON

The Open Data Platform supports the following standard and specialized content types:

1. **Digital Objects**: Any digital object can be uploaded and described in metadata, including popular formats such as PDF, images, video, zipped archives, presentations, and documents. *Status: mature release available.*
2. **Spatial Data**: Data can be hosted by the depositor (in one of a number of standard formats or services), or by SAEON in a geospatial server. Data can be downloaded, visualized, or compiled into Atlases and composite maps. *Status: mature release available.*
3. **Tabular Data**: Text and CSV files are supported for visualization and download. Hosted either by the depositor or uploaded to platform servers. Can be piped to charting applications if the schema is reasonably simple. *Status: beta version available and can be made operational on demand.*
4. **MetaCAT Data:** Widely used in long-term environmental observation, can be uploaded to ODP servers. Data can be downloaded, visualized in tables, or visualized in charts. *Status: mature release available.*

5. **Multidimensional Data**: Specialized interfaces were developed for WAMIS data and for 2D map-based visualization of NetCDF. *Status: beta versions of these are available for operationalization. Operational implementation will follow in 2016.*
6. **Sensor Data:** Sensor observation services standards are supported in end user interfaces, and an operational time series database is available – currently used by SAEON for internal data streams. *Status: SensorWeb services – beta version. Time series database – operational within SAEON, with extension for visualization, query, and download planned for Q3 2016.*

Search and Discovery Options

The ODP provides the following options in respect of search and discovery. There are two main sets of functionality: services and components that search through the metadata available to the ODP, and services that search the contents of a specific Portal or Gateway.

Metadata Searches

1. **GUI**: A search interface that supports free text, spatial, temporal, and selected vocabulary-based search actions. Results can be viewed as summary metadata, detailed metadata, map representations, and filterable lists. *Status: mature implementation available, new release end August 2016.*
2. **Predefined Queries:** Complex, composite search criteria can be saved and invoked by name. *Status: mature implementation available, new release end August 2016.*
3. **Embedding**: The search interface can be embedded into any other web server or web application. The search interface can be manipulated by way of query string parameters under control of the end user. *Status*: *mature implementation available, new release end August 2016.*
4. **API Services**: API services can be used to invoke a search, with query parameters. There are two search protocols, with a third being planned:
    a. **CS/W**: A CS/W endpoint (Catalogue Services for the Web) is available and can be invoked through HTTP or via a web front end. Output is Dublin Core XML with extensions. *Status*: *Operational beta version, with a new release planned to offer DataCite XML schema as an option.*
    b. **Proprietary:** A RESTful web interface (GetRecords) that can be appended to any portal offered by the ODP, and producing output as HTML, GeoRSS, of Dublin Core XML. *Status: Operational beta version, no enhancements planned.*
    c. **OAI-PMH**: *Status: Planned.*

Portal Content Searches

It is also possible to retrieve the contents of a portal or gateway, in one of three ways:

1. **Search Facility**: Each portal or gateway exposes a site search facility – either as a 'QuickSearch' option or as an advanced option that can also search elements other than free text. *Status: mature implementation available.*
2. **Content Service**: Any content within a portal or gateway can be queried and returned as JSON for further use. *Status: mature implementation available*
3. **RSS Push Service**: The content of any folder can be published as an RSS feed, which allows notifications of changes to the content of the folder to be pushed to a mail client, an RSS reader, or any other RSS-aware client. If the contents of the folder are geo-tagged images or objects, the service can also be used as GeoRSS. *Status: beta version available. Future plans – also allow RSS 2.0 enclosures, enabling the syndication of video and audio feeds.*

Metadata Upload and Synchronisation Options

Metadata can be provided to the ODP in a variety of ways:

1. **Through Harvesters**: this is the preferred method and can be automated to maintain synchronization between metadata collections hosted by the provider and the aggregate maintained by SAEON. See the table below for options in respect of metadata schema and harvesting protocols. *Status: Operational, see table for planned extensions.*
2. **Through a Push (RESTful) Service**: Generally speaking, an XML file compliant with any of the supported metadata standards can be uploaded through a web service to one of the repositories offered by the ODP. *Status: Operational, no changes foreseen.*
3. **Creation from Template**: users can create and edit metadata in the ODP from user-configured templates. This method is only workable for a small number of records or incremental additions. *Status: Operational, improvements to template configuration are planned.*
4. **Upload a File**: Users can upload individual XML files. This process is best assisted by SAEON personnel, since it may require prior validation. *Status: Operational, no changes foreseen.*
5. **Upload Wizard**: A web application that allows simultaneous upload of data, metadata, publication, and styling of data for visualization. *Status: beta version available, with finalization by end September 2016.*

1. **ArcCatalog Adapter:** This is a requirement expressed by several stakeholders. At present, SAEON can harvest and partially map ArcInfo metadata to DataCite and SANS 1878 schema. *Status: Operational, improvements to error reporting and extent of mapping is planned for late 2016.*
2. **IPT:** This adapter links species observation data to the GBIF registry – SAEON will do internal development work in the 2016 financial year to implement the option to pass species observations to GBIF.

	CS/W	OAI-PMH	HTTP	FTP	IPT
Dublin Core		Yes	Yes	Yes	
ISO 19115 …	Yes		Yes	Yes	
FGDC	Yes		Yes	Yes	
SANS 1878	Yes		Yes	Yes	
EML		Yes	Yes	Yes	
DataCite		In Process	Yes	Yes	
ArcCatalog	Yes				
DDI		In Process	In Process	In Process	
Darwin Core					Scheduled

Data or Content Visualisation and Presentation Options

The ODP offers a variety of options for visualization of standard data services. These include:

1. **Spatial Data Formats**: combine data services from OGC WxS (vector and raster), KML, GeoRSS, and Mapping Services (Google, OpenStreetMap). *Status: Operational version available, improvements planned for release in September 2016.*
2. **Tabular Data**: Brokering and mediation is available for charting (graphing) and tabling of tabular data obtained from any of the following sources:
    a. **EML 'EcoGrid"**: CSV data can be tabled or graphed. *Status: Beta version.*
    b. **SensorWeb**: Time Series data from SensorWeb Services. *Status: Beta version.*
    c. **Specialised Interfaces**: SAEON has developed specialised interfaces for the WAMIS Service and for its own Observations Database. *Status: New release planned for September 2016*.
3. **Saving Context**: SAEON has adapted the OGC Web Map Context specification to allow persistence of mashed-up views that contain data services from a variety of standard servers. *Status: Operational version available, improvements planned for release in September 2016.*
4. **Special Composites:** These components allow end users to request template copies of popular composite visualisations (for example an Atlas composed of distributed data sources) and populate them using discovery services. *Status: Atlas is operational, charts and indicators planned for September 2016*.

Data Publication and Citation

The ODP supports data publication and citation via a DataCite License, and this service is currently free to SAEON collaborators and stakeholders.

In short, the service allows depositors to

1. Indicate and select data sets for publication;
2. Request Digital Object Identifiers for such objects (DOIs);
3. Obtain feedback and statistics on citation of their deposited data;
4. Query DataCite services for additional information.

Once a DOI has been allocated, SAEON has a responsibility to ensure continued access to the original data set, and, as such, will require additional information and commitment from depositors.

In summary, the following options exist for use and integration of the Open Data Platform:

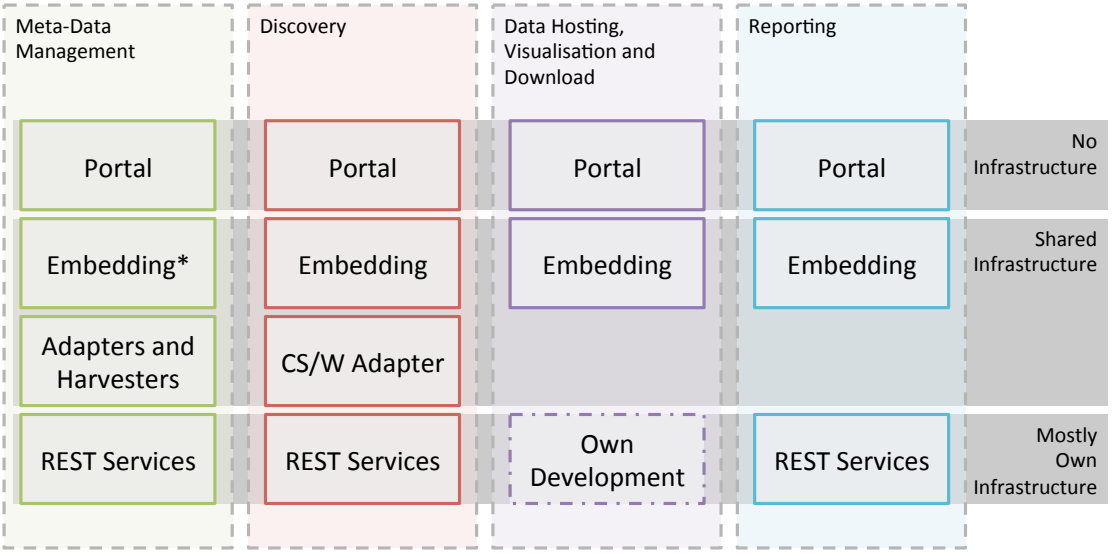

Figure 2: Options for use and integration of the Open Data Platform

Roadmap for Data Infrastructure

This section provides a high-level summary of the roadmap for Data Infrastructure development in the next 3 years, with a more detailed discussion of the expected progress to be made through SAEON development and collaborative projects.

A Short Description of Our Vision and our Mandate

SAEON has been active in the development of data management systems, largely funded by NRF and DST, since 2008. This effort has resulted in a collection of services, tools, and components that is referred to as the 'Open Data Platform (ODP)'. SAEON's vision is to make the ODP available to other government funded programmes to assist with appropriate aspects of data stewardship, leverage such funding as may be available for systems development so that a collective resource is enhanced and extended, and support publication and the long-term preservation of data sets that are indicated by its mandate.

Within a broader context, this resource will be managed as a major component of an Earth and Environmental Data Centre that serves the community, is aligned with and contributes to DIRISA as appropriate, is linked to global initiatives such as GEOSS and ICSU-WDS, and generally promotes two concepts: standards-driven interoperability, and free and open access.

While SAEON has access to funded baseline resources (people, hardware and software infrastructure, and management), it has limited capacity to serve its stakeholder community free of charge. Since SAEON does not make a financial profit and is focused on creating economies of scale for efficiency of delivery, it will likely remain as a reliable and cost-effective partner for local institutions in respect of data management, publication, and preservation.

Stakeholders

The ODP has two sets of stakeholders: direct stakeholders and collaborators, and indirect stakeholders that benefit from the platform in general terms. We discuss only the direct stakeholders here, together with the main collaboration focus.

	Stakeholder	Internal SAEON Focus	Collaboration with Stakeholder
	NRF	SAEON will develop a set of guidance documents for data management, DOI implementation, and an institutional repository that can	NRF will do the same as part of its **Open Access Statement** implementation. NRF wants to apply soft and hard infrastructure as applicable to other NRF Facilities

		be accredited.	for institutional repositories. Collaborate with DIRISA to build Research Management Infrastructure integrated with RIMS.
	DIRISA Project	In addition to the above, SAEON will be developing the ODP as a trusted digital repository that is integrated with research cloud-based storage and preservation infrastructure.	DIRISA serves as additional hardware and software back-up for the ODP, and ODP software and services are used as appropriate by DIRISA.
	SANSA and EODC	SAEON has created an interface to the EODC metadatabase, funded by DST.	The EODC/ SAEOSS integration has to be operationalized and linked into the GEOSS Broker architecture.
	GEOSS and AfriGEOSS	SAEON already exposes all public domain, standardized spatial data sets to the GEOSS Broker and the GEO Core Data collection.	The SARVA and BioEnergy Atlas portals are available for integration into African and Global initiatives such as IRENA and the BioEnergy Atlas for Africa.
	DEA	Comprehensive set of collaboration projects has been agreed and some are under way (SANBI, SASDI contribution).	Meraka Institute has developed Carbon Atlas data sets for DEA, and SAEON has published this as part of SARVA for public availability. Future work will incorporate the bulk of publicly available data sets available from DEA.
	SANBI	SAEON has or is busy developing a suite of products that overlaps with SANBI's needs. SAEON and SANBI are in the process of concluding a collaboration MOU and this will cover assistance to SANBI with data publication and component development.	SANBI has use for existing and planned SAEON infrastructure, and may be able to contribute to development and refinement or funding.
	SALGA	SAEON is developing an engagement programme for SARVA and BEA that will involve accreditation of service providers (consultants).	SALGA needs to publish municipal metrics as data sets that are accessible (probably through SARVA), and wants to participate in the engagement programme.
	DEA Oceans and Coasts	SAEON has already developed several of the data publication and dissemination channels required by DEA Oceans and Coasts, and have previously converted SADCO metadata to published ISO schemas. Will in future include infrastructure for	DEA Oceans and Coasts is in process of implementing data description, publication, and dissemination infrastructure in collaboration with SAEON. DEA can assist with OPeNDAP and THREDDS implementation.

		multidimensional data.	
	Eskom	SAEON has collaborated with Eskom in the recent past by making data on BioEnergy potentials and feasibility available.	Continued collaboration as and when required on dissemination of Eskom's publicly available data, and on providing support and access in respect of BioEnergy Atlas data.
	DRDLR	SAEON is using ODP as a basis for implementation of the South African Spatial Data Infrastructure.	SASDI/ DRDLR requires many of the functions offered by the ODP and is funding extensions and improvements. It will facilitate access to government-held spatial data.

Soft Infrastructure

	Aspect	Description	Actions Required
	Guidance on Data Management	Advice and guidance on the elements of proper management of research data throughout its life cycle. Includes sub-elements of deposit and publication, standards, and preservation.	SAEON will develop a framework document and a draft guideline. NRF/ DIRISA working group to refine in a 6-month period and publish. Integration with grant application process – NRF.
	Policies and Licenses	Guidance on elements of data policy and supporting Open Access licenses.	SAEON will develop a framework document and a draft guideline. NRF/ DIRISA working group to refine in a 6-month period and publish.
	Guidance on Permanent Identifiers	Advice and guidance on permanent identifiers in the web, and the SAEON/ NRF DOI implementation.	SAEON will develop a framework document and a draft guideline. NRF/ DIRISA working group to refine in a 6-month period and publish.
	Guidance on Metadata and Data Standards	SAEON has existing guidance documentation that needs to be extended and aligned with the requirements of DOIs and DataCite. Detailed guidance exists for Biodiversity Data and General Environmental/ Earth Observation	Existing SAEON documentation to be generalized and used as a blueprint for non E&EO scientific domains. This can be done by a joint DIRISA/ NRF working group, driven by DIRISA.
	Trusted Digital Repositories	SAEON, via its involvement with WDS, is party to a WDS/ DSA effort in RDA to develop a broad guideline for assessment and accreditation of Trusted Digital Repositories. This will also guide its own application to accredit	DIRISA, with inputs from the SAEON effort and NRF, to develop an accreditation framework for institutional repositories. NRF to use this as a guide to individual researchers in respect of deposit and preservation. DSA and WDS,

		the ODP as a WDS member.	within an RDA working group, are supportive partners.
	Capacity Building	SAEON has no national mandate or immediate plans to build capacity in data management, except internally.	DIRISA, NRF, and selected HEIs have indicated that they will pursue development of honours-level courses in data management. These courses can be informed by the guidance developed collectively (see above).
	Domain Reference Groups	SAEON needs to establish a reference group and governance input in respect of the ODP for E&EO data.	This can serve as a test case for similar reference groups.
	RDA Alignment and Coordination	SAEON is active in several RDA working groups.	DIRISA is placed to coordinate local contributions to RDA and to integrate soft infrastructure efforts into a coherent effort. Possibility of creating "SA-RDA" as a initiative promoted by NRF/ SAEON, DIRISA, and HEIs (UCT, UJ,).
	GEO BON Advisory	SAEON has contributed to development of a framework for standards and specifications to support the GEO BON Essential Biodiversity Variables.	Publication of WIKI-style, community-maintained service and standard advisory pages with collaboration of DIRISA, GBIF, DataOne, EU-BON and iDIV.

Software and Services

The software and services offered by SAEON in the ODP are described in detail in the first section of this document. The following portfolio of improvements and new development is foreseen:

	Aspect	Description	Partners *Potential Partners*	Timeline *Funding*
	Brokering Framework	SAEON has developed a pro-forma brokering framework for registration and filtering of ODP components that it is in process of deploying. This framework is of interest to RDA Brokering Interest Group.	WDS-RDA PANGAEA *OpenAire* *DataCite*	12 months
	Depositor/ Custodian Management	SAEON is busy developing services for the remote registration and editing of custodian and depositor entries, allowing its own systems and those of others to create such entries without access to a portal or gateway, and to embed it into workflow.	*DIRISA*	3 months *SAEON*

	Data Publication and Citation	SAEON has recently developed the interface to automate registration of DOIs through DataCite, based on collections of existing or harvested metadata. This capability has to be extended to provide a seamless, wizard-driven interface for users ranging from single depositors through to custodians managing multiple thousands of records. Cosmetic updates planned.	*DIRISA*	6 months *SAEON* *DIRISA*
	Collection Management and Reporting	Individual depositors and custodians all require management reporting and collection management services, some of which already exist. These have to be packaged so that it can be accessed as embeddable pages/ web components, or as services within external systems. Cosmetic updates planned.	SAEON *DIRISA*	3 months *SAEON* *DIRISA*
	Data Management Plans and RIMS Integration	SAEON will be developing software and services to support depositors in management of data. This includes selection of standards and specifications, software, a trusted repository, and publication of data as proof to the funder that requirements have been met.	SAEON *DIRISA* *NRF*	18 months *SAEON* *DIRISA*
	Data Visualisation and Assessment	The existing signature-recognising broker will be improved to include additional standardized data sources and services, and extended to allow persistence of special mash-ups (such as 'indicators'. Cosmetic updates planned.	SAEON *SALGA* *Meraka* *DEA*	SAEON BEA SARVA
	Data Preservation and Curation	SAEON has, to date, not implemented formal data preservation and curation software. iRODS is under consideration.	*DIRISA*	24 months *Not funded*
	Model Web Automation	Linking of data sources and web processing services that are standardized or can be mediated to produce automated new data sources or services.	GEO BON WG8 GBIF iDIV (Leipzig) Macquarie Univ.	6 months *Voluntary*
	"Gateway on Demand"	Leveraging the ability already available within the ODP to create a Science Gateway or Collaboration Portal within a short (although	*DIRISA* *SANREN* *CHPC*	18 months *Not funded*

		manual) space of time, SAEON would like to develop the automation of this function with transparent access to computing and data storage facilities within the South African Research Infrastructure cloud.		
	Composite (Mash-Up) Applications	The current spatial data publication application (Atlas) has to be extended to allow analysis and drill-down, and to allow more sophisticated configuration – including automated configuration. Cosmetic updates planned. Equivalent components are planned for Charts (time series data), multidimensional and model data, and for decision support frameworks such as SARVA and BEA.	*SAEON* *DEA*	6 months *SAEON* *SARVA* *BEA*
	Improved Discovery	Current discovery mechanisms are limited in terms of large metadata collections. Implementation is needed to improve speed and handling of large return sets.	PANGAEA ElasticSearch/ Lucene *SANSA*	6 months *SAEON* *SARVA*

Specific Operational Systems and Data Products

	Aspect	Description	Partners *Potential Partners*	Timeline *Funding*
	Observations and Measurements Database	SAEON has developed a database for the storage and dissemination of time series data. This has to be extended to allow improved visualization, querying, and automated maintenance of metadata records and DOIs. In addition, data needs to be offered as standardized services (SOS, netCDF).	*DEA* *UWC* *Meraka*	6 months *SAEON*
	MetaCAT	MetaCAT serves as a metadata and data repository for structured, but arbitrary, ecological observation data. It currently supports only user-driven input. Extensions required include service implementation to allow machine interaction, and cosmetic updates.	*NCEAS*	12 months *SAEON*
	GeoServer/ GeoNode	SAEON currently operates 8 GeoServer instances. These have to		12 months *SAEON*

		be migrated to load-balanced servers, and if feasible, updated to GeoNode.		
	OPeNDAP/ THREDDS Server	SAEON and its stakeholders will increasingly require multidimensional data storage.	*DEA O&C*	6 months *SAEON*
	SAEON SPECIES	A species-annotation and occurrence database, which can be in time extended to define ecosystem structure, species interaction, traits, etc. Can also be extended to automate annotation and to process video.	*LifeClef/ CIRAD* *SANBI* *SAIAB*	12 months *SAEON*
	MEDIA	SAEON is investigating the deployment of cloud-based (Dropbox-like) software as a supplement to research content management systems. Will make provision for deposit and management of large collections of images, video, audio, and data logger files in a distributed environment.	*DIRISA*	3 months SAEON

Data Sharing Agreement

General Data Sharing Agreement

Conditions of Archiving

1. The SAEON ODP offers the facility to store datasets in a secure environment that protects the intellectual property rights of the Data Providers. All Data Providers are required to fill out and sign the Data Provider Agreement. Signed Agreements can be returned to the relevant SAEON Node.

2. The Data Provider is free to archive his/her/its Dataset(s) with the SAEON Data Portal, without charge. SAEON agrees to archive these data provided they meet the conditions in (3), (4), and (5) below.

   OR: The Data Provider can host the data in non-SAEON infrastructure of choice, provided access is aligned to the license conditions, and conditions (3) and (4) are adhered to.

   OR: The Data Provider can host the data in SAEON infrastructure on the basis of a service level agreement, for which a cost of service fee applies.

Metadata

3. All Datasets that are archived in the SAEON Data Portal must be accompanied by Metadata. All Metadata will be publicly accessible, to allow potential users to locate Datasets of interest to them. Metadata must be maintained using the service stipulated by (15) below.

Annual review

4. The Data Provider will review the status of the public availability of his/her/its dataset in the SAEON Data Portal on an annual basis. This pertains to datasets that have been archived and are not yet publicly available.

Regular updates

5. The Data Provider will notify the SAEON Data Portal Gatekeeper if he/she/it has any updates to add to a particular Dataset archived in the SAEON Data Portal. Such updates include changes of contact details; changes in or additions to metadata information; changes in or additions to the dataset.

Service Levels and Responsibilities of SAEON

6. SAEON does not make any express or implied commitment in terms of service levels in support of the data deposit.

   OR: SAEON will adhere to the service level agreement entered into with the data provider for a paid service portfolio.

7. SAEON will apply due diligence in respect of backups of the data and metadata supplied to it and will keep a copy of both in a separate physical location to the SAEON servers at all times. Backup frequency is at least once a week.

   OR: SAEON will develop a backup and failover strategy with the data provider as a part of a paid service portfolio.

8. SAEON will endeavour to provide high availability for its servers and services associated with metadata and data deposit, but does not make any express warranty in respect of this.

   OR: SAEON will agree an availablity and uptime with the data provider as a part of a paid service portfolio.

Intellectual Property Rights

9. SAEON allows access by default to data sets by application of a Creative Commons License (see https://creativecommons.org/choose/).

10. The standard license in use allows sharing, as long as subsequent users can share alike, and allows commercial application. The terms of this license can be found here: http://creativecommons.org/licenses/by-sa/4.0/

11. This license protects the rights of IP owners across multiple jurisdictions.

12. Data Providers should indicate the following:

    a. If the Creative Commons License is acceptable or
       i. Whether use is limited to non-commercial applications;
       ii. Whether a publication embargo is applicable;
       iii. Whether the data refers to species that are endangered;
       iv. Whether the data is private to an individual;
       v. Whether there are ethics considerations attached to the data;
       vi. Whether end users should be identifiable.
    b. If the Creative Commons License is not applicable
       i. provide an end user license (terms and conditions) instead;
       ii. Disallow access to the data.

Digital Object Identifiers and Data Publication

13. Data Providers should indicate whether a Digital Object Identifier should be issued for the data set. This will normally be done by default for all publicly available data under the Creative Commons License, and allows citation of data in journals and scholarly articles.

Metadata Harvesting

14. The Data Provider is aware of the fact that SAEON will offer applicable metadata records to global networks, including but not limited to DataCite, the ICSU World Data System, GEOSS, and ILTER. Metadata is automatically made available to South African resources such as SASDI, SAEOSS, The Risk and Vulnerability Atlas, and the BioEnergy Atlas. Any institution can harvest metadata from SAEON.

Specific Services

15. This agreement makes provision for the following specific services:
    a. SAEON will create a custodian area for the Data Provider, for the purpose of management and harvesting of metadata.
    b. SAEON will provide a GeoServer instance to the data Provider for the purpose of publication of spatial data, and for the purpose of maintaining metadata.
    c. Metadata will be harvested from GeoServer by the SAEON Data Portal for purposes as in (14) above.

OR: A paid service portfolio can stipulate any or all of the following requirements, in addition to the above, as appended to the Data Agreement:

    a. SAEON will create a custom portal or gateway for the service.
    b. SAEON will provide operational database access and support for the purpose of publication of a variety of supported data formats.
    c. SAEON will define specific development tasks based on user requirements in respect of gateway and portal functionality.

General Services

16. SAEON offers additional web services and embeddable components to the Data Provider on a best effort basis.