



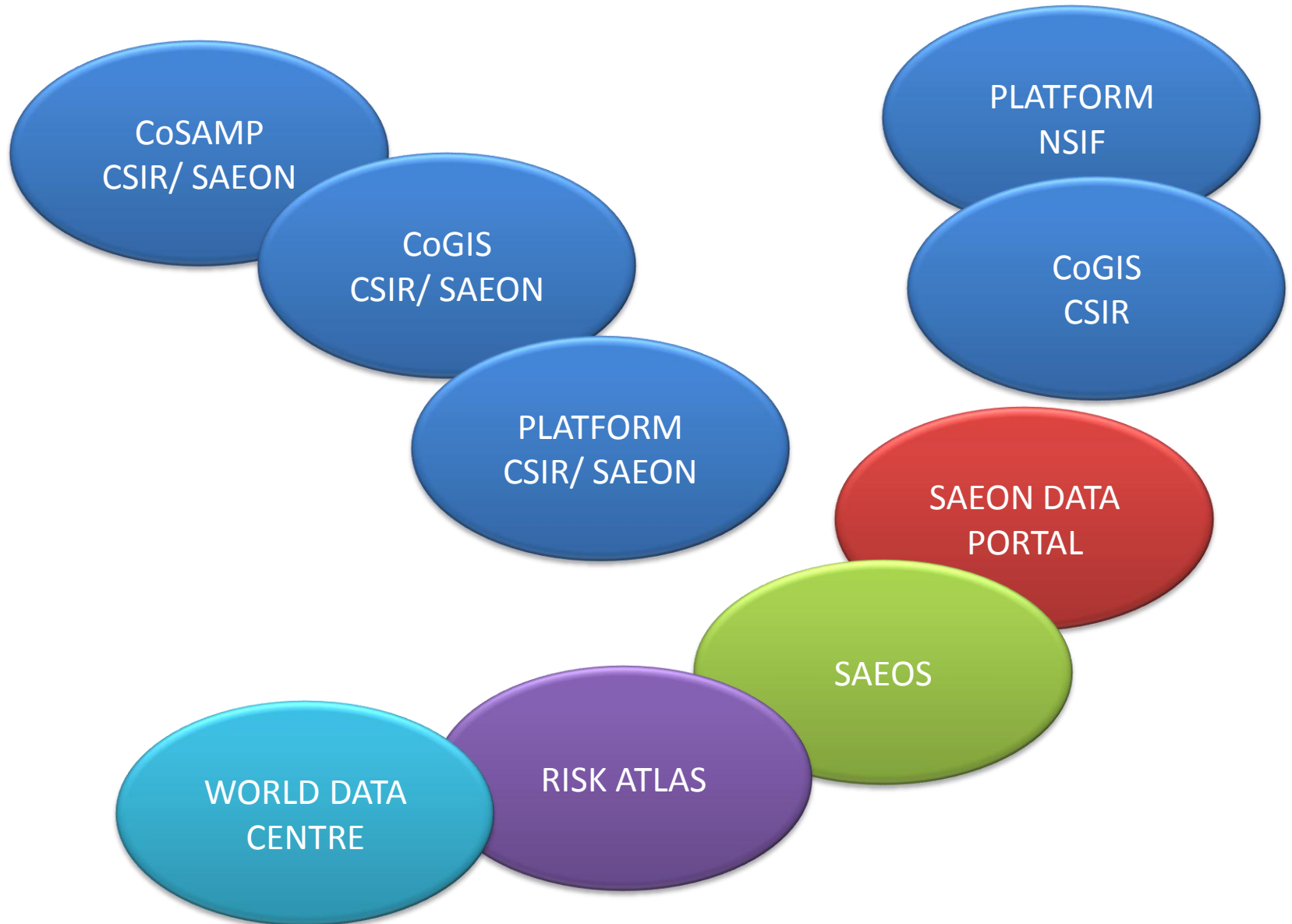
A South African Platform for Data Systems SAEON Data Portal and Sister Projects

SAEON Summit

5-6 October 2010

Wim Hugo - SAEON

Portal Evolution



Partners/ Funding

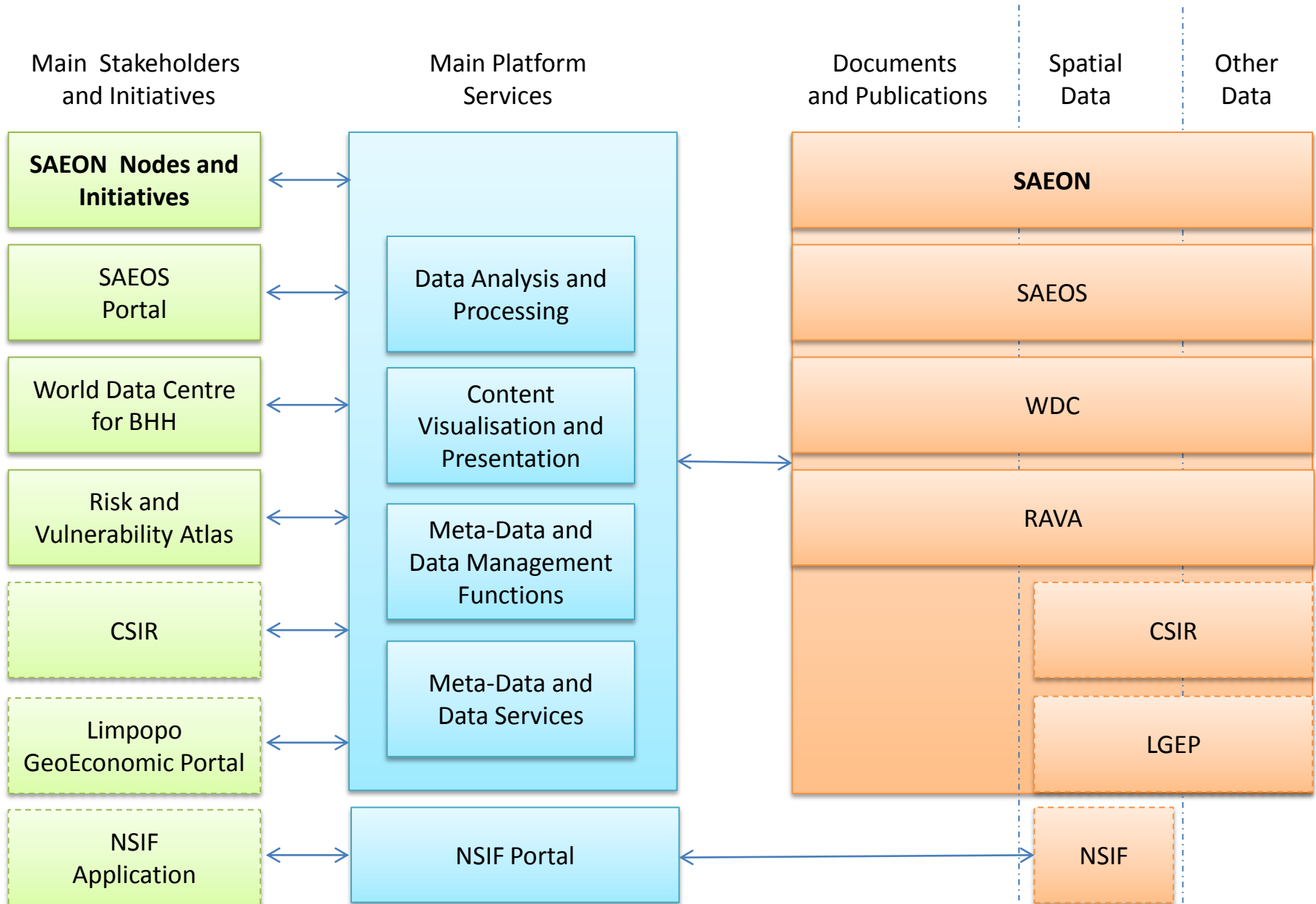


science
& technology

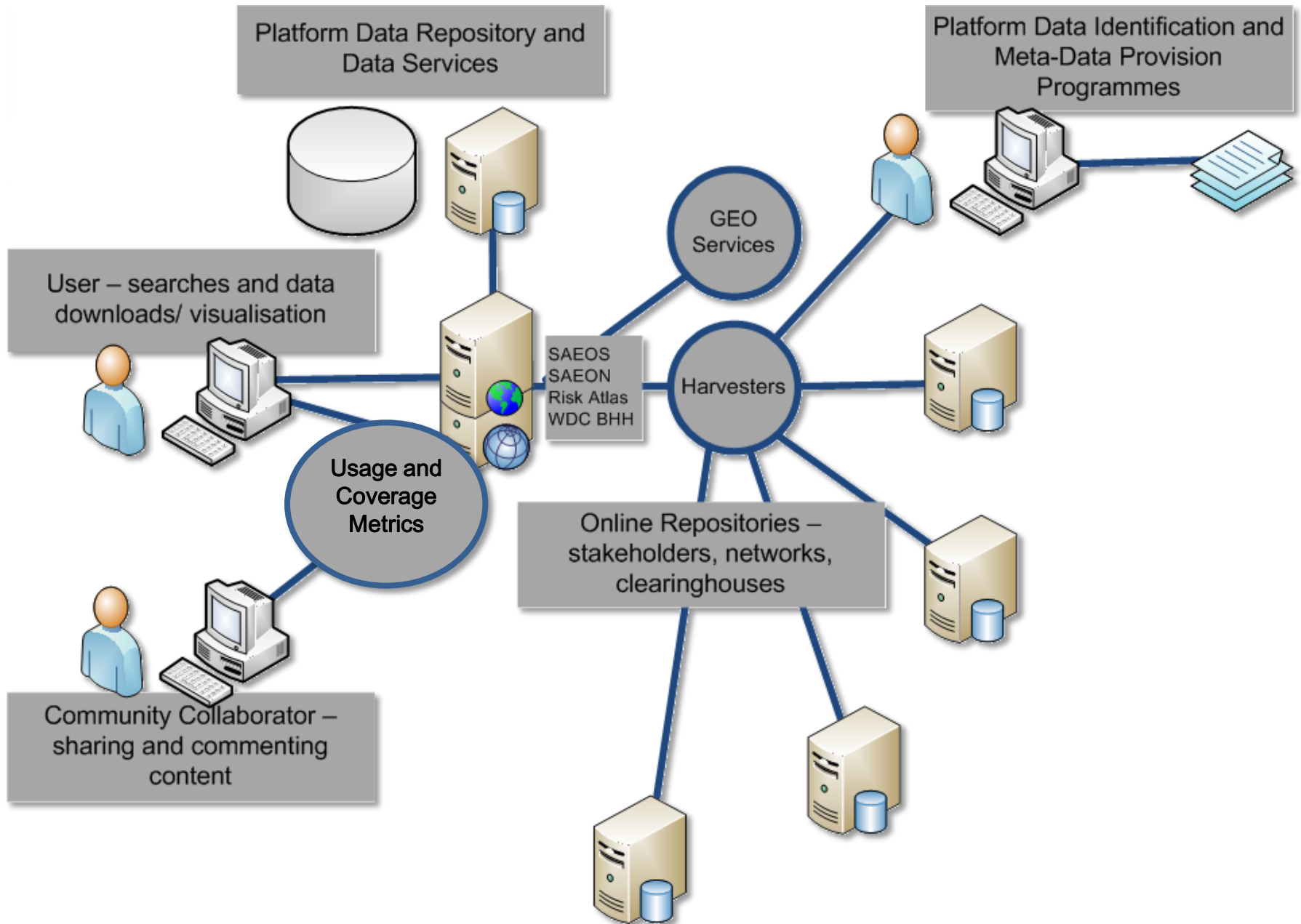
Department:
Science and Technology
REPUBLIC OF SOUTH AFRICA



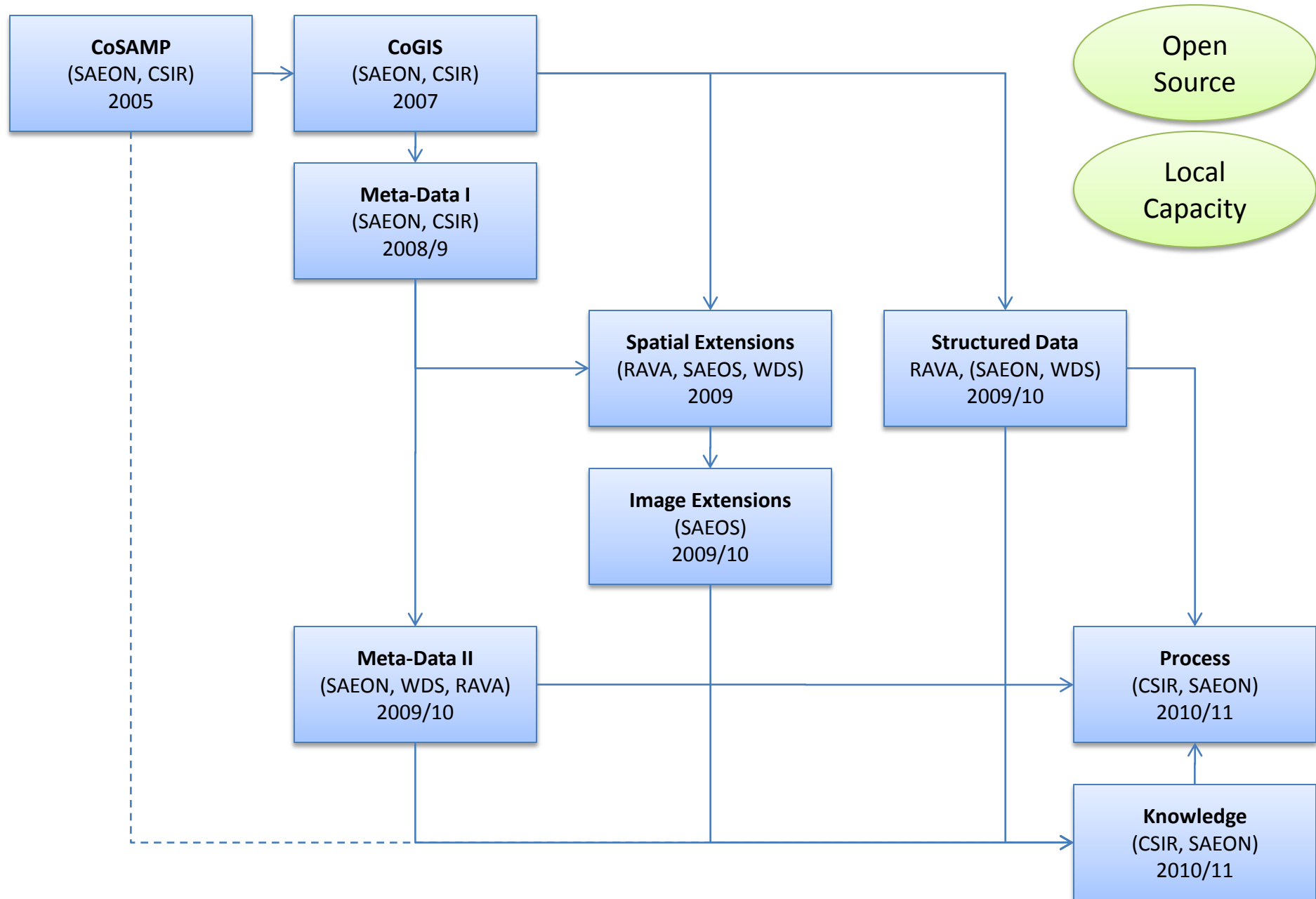
Portals and Platform



Main Use Cases



Aligned Development/ Road Map



Current Programmes

Strategy, Policies and
Guidelines/ Funding

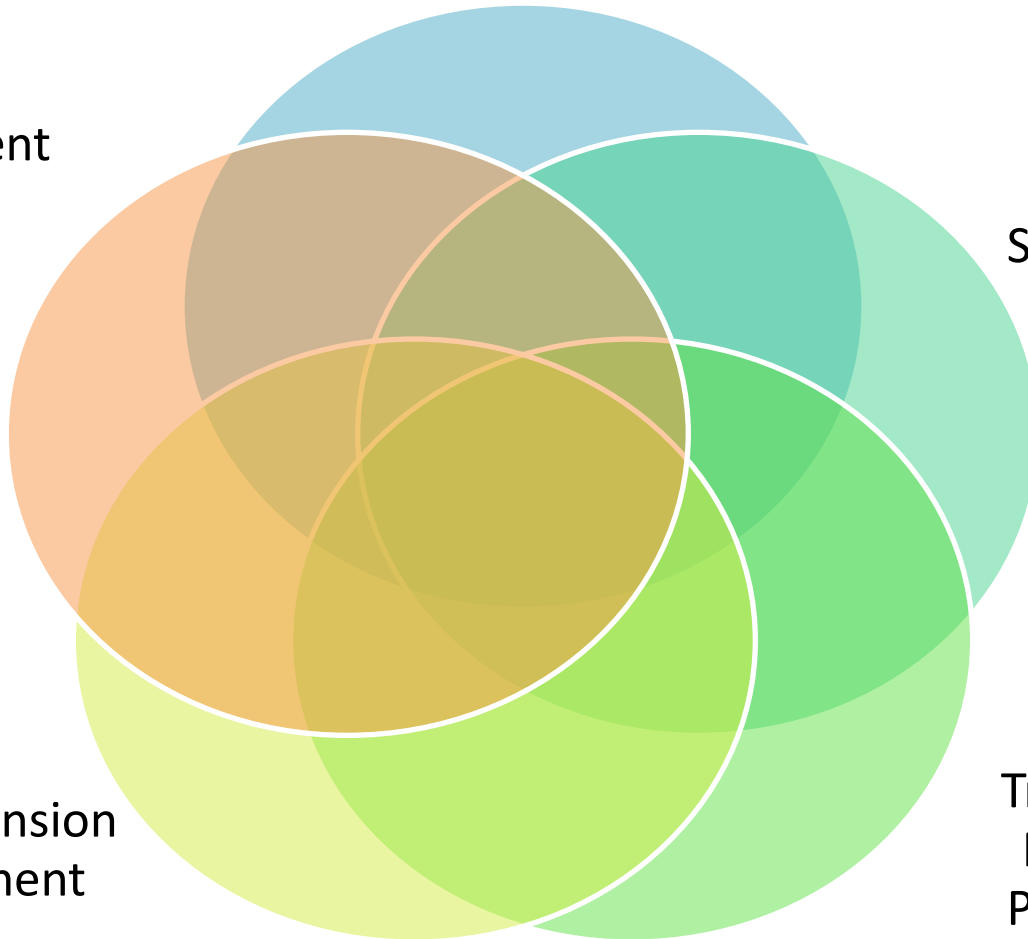
Networking and
Strategic Alignment

- NBII/ USGS
- CHPC
- GEOSS
- LiquidPub
- SA EOL

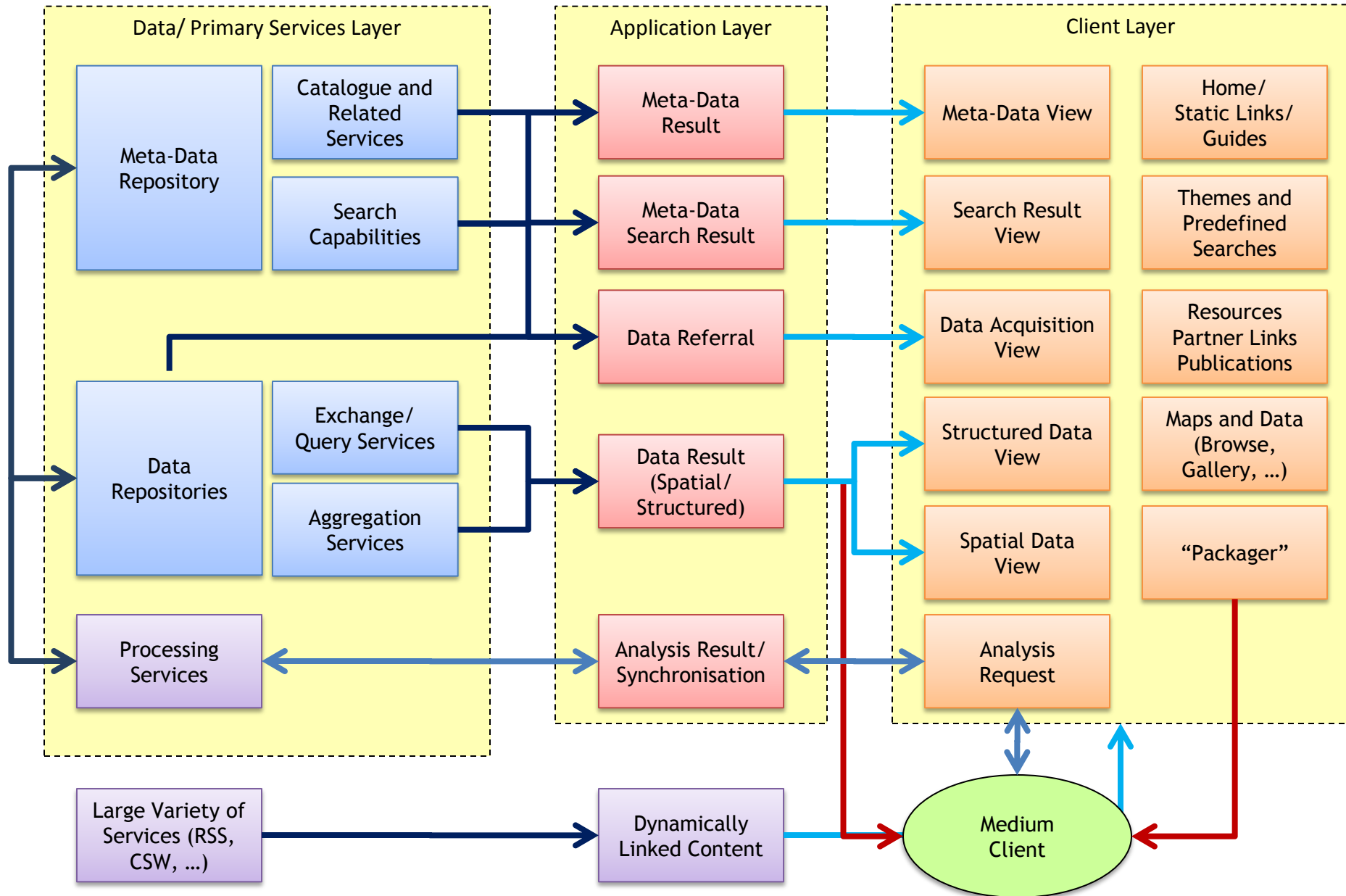
Stakeholder Meta-
Data Acquisition

Platform Extension
and Refinement

Training and
Marketing
Programme



Technology Platform Architecture: 2009-11



State of Meta-Data

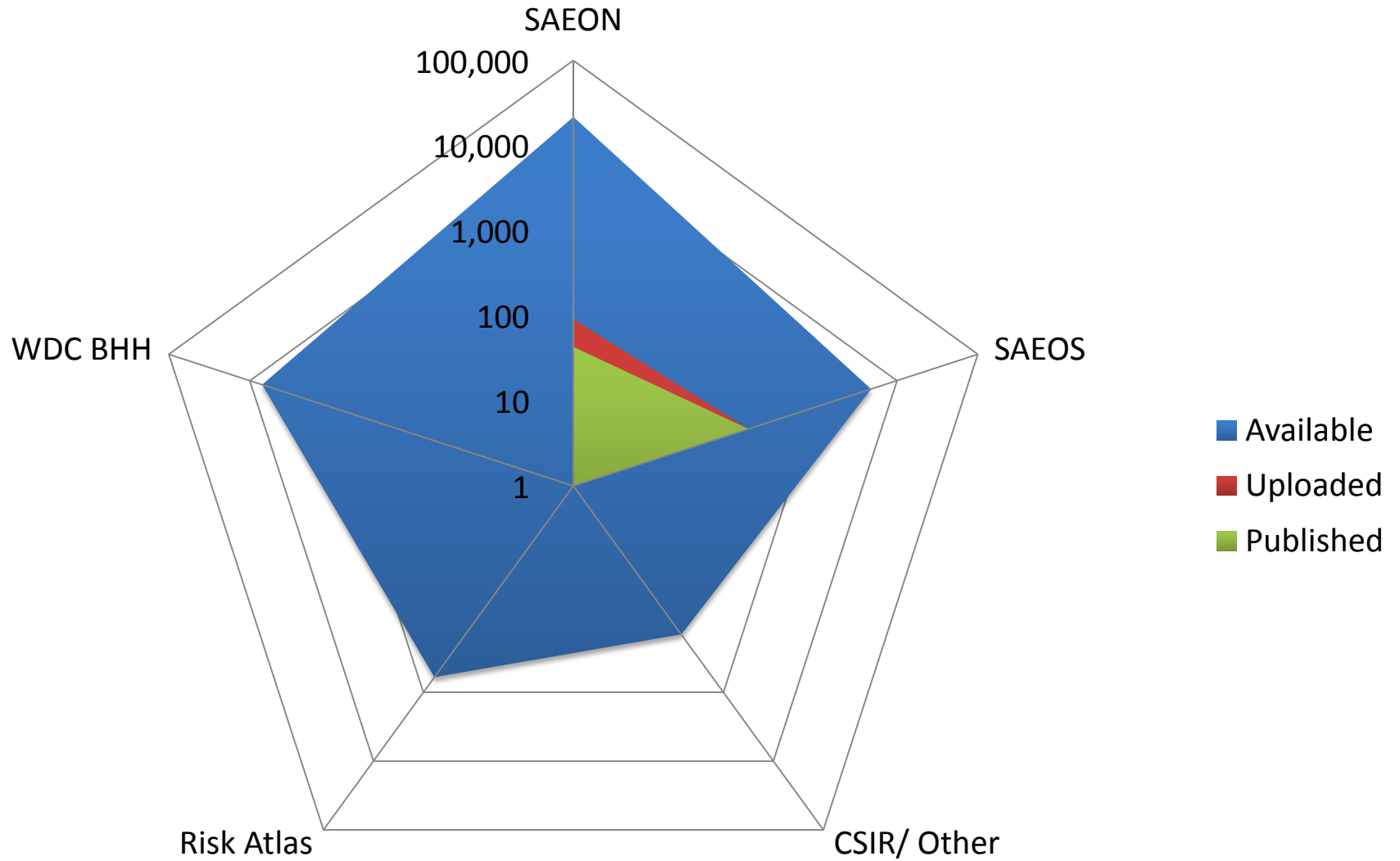
Project/ Initiative/ Custodian or Source			Format	Total	Uploaded	Published
SAEON	Nodes	General	EML	80	80	43
	Nodes	Catch Data - Historical	EML	6250	0	0
	Nodes	EKZNW	EML		0	0
	Nodes	Estuaries Database	ISO 19115	260	0	0
	Nodes	MRSU	ISO 19115		0	0
	Nodes	OceansAfrica	ISO 19115		0	0
	Nodes	SHBML	EML	319	0	0
	Nodes	SANBI-OMPA	EML	34	0	0
	Nodes	SADCO	EML	9715	0	0
	NCCH	ACEP	ISO 19115	210	0	0
		FAO	ISO 19115		0	0
		SANHO	ISO 19115	51	0	0
	PLATFORM	ADU	EML	1337	0	0
		NSIF	ISO/ FGDC	291	0	0
		SABIF/ GBIF	EML	3120	0	0
		SANPARKS	EML		11	0
		VEGMAP	ISO 19115	17	0	0

State of Meta-Data (2)

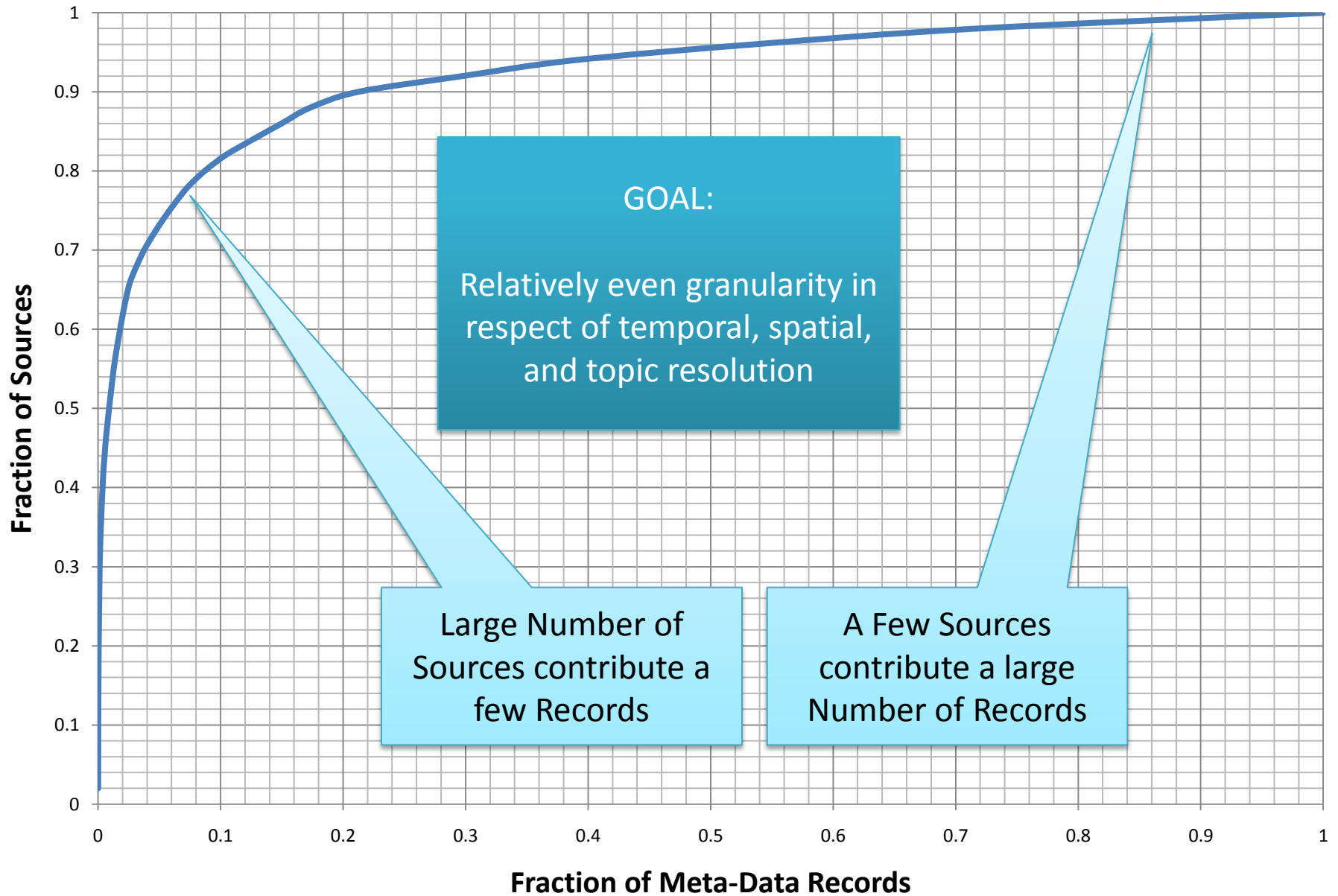
Project/ Initiative/ Custodian or Source			Format	Total	Uploaded	Published
SAEOS	PLATFORM	AFIS	ISO 19115	1	1	1
		AGIS	ISO 19115	90	0	0
		ARC	ISO 19115		0	0
		DWA	ISO 19115		0	0
		Council for GeoScience	FGDC/ ISO	800	40	40
		ICSW	ISO 19115		0	0
		NDMC	ISO 19115		0	0
		SAC - SPOT5	ISO 19115	2918	100	100
		SAC -CBERS	ISO 19115	180	0	0
		SAC - SCCA	ISO 19115	800	0	0
		SAWS	ISO 19115		0	0
		WAMIS	ISO 19115		5	5
Risk Atlas	PLATFORM	BGIS	ISO 19115	111	1	1
		Case Studies	EML	4	4	4
		CSAG	ISO 19115	466	166	2
		Weather-CSIR	Dublin Core		0	0
		TIP	ISO 19115	22	4	4

Project/ Initiative/ Custodian or Source		Format	Total	Uploaded	Published	
World Data Centre	PLATFORM	CHELYS	GeoRSS	18	0	0
		FAO	ISO 19115	5	0	0
		In-Depth	ISO 19115	2	0	0
		MARA	ISO 19115	584	0	0
		PANGAEA	Proprietary	5000	0	0
	NBII-USGS	LIVE	FGDC	356	0	0
		ESA	EML	2	0	0
		NCEAS	EML	51	0	0
		NRS	EML	6	0	0
		GBIF	FGDC	766	0	0
		GBIP	FGDC	1	0	0
		HBMP	FGDC	1	0	0
		NBII	FGDC	11	0	0
		MA	FGDC	71	0	0
		NCDDC	FGDC	1	0	0
		NCEAS	FGDC	47	0	0
		NOS	FGDC	14	0	0
		NRPI	FGDC	58	0	0
		NRS	FGDC	6	0	0
		NSMN	FGDC	2	0	0
		OBFS	FGDC	1	0	0
		PASDA	FGDC	33	0	0
		UAF	FGDC	2	0	0
		VNRGC	FGDC	1	0	0
		WDC	FGDC	5	0	0
CIESIN	FGDC	5	0	0		

Meta-Data Sources



Meta-Data Resolution Problem

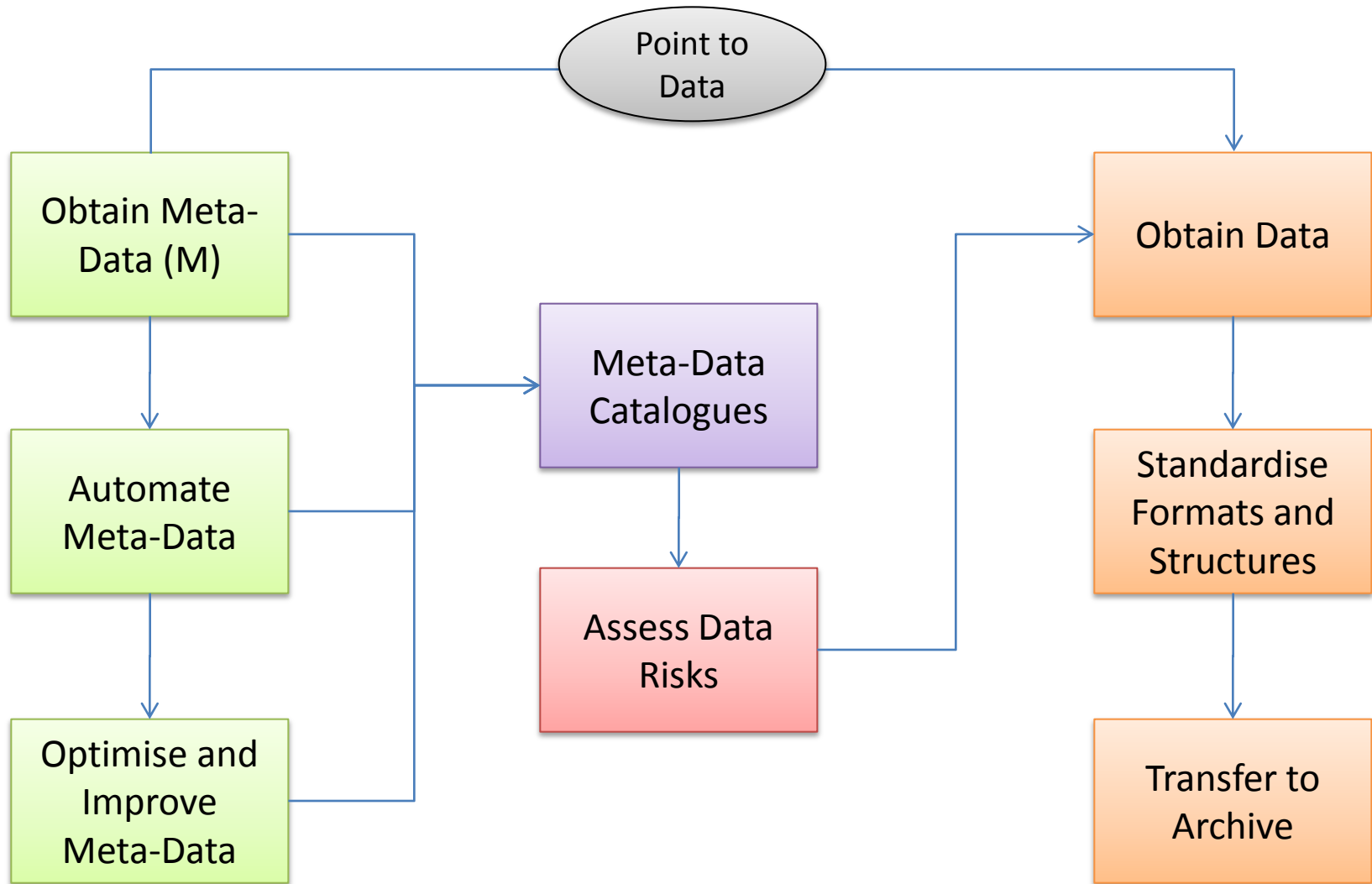


Some Perspectives ...

... or opinions:

- Availability of ***historical data*** and ***large anthropogenic/ natural impacts and resulting anomalies*** determine study areas.
- Variability determines required sampling frequency.
- Meta-data populates a multidimensional space containing at least
 - Spatial coverage,
 - Temporal coverage,
 - Topic coverage (natural and anthropogenic drivers, responses)***This space can be pre-defined from the perspective of the science plan.***
- Some meta-data is better than nothing at all.

Meta-Data and Data Strategy



New Developments ... CHPC

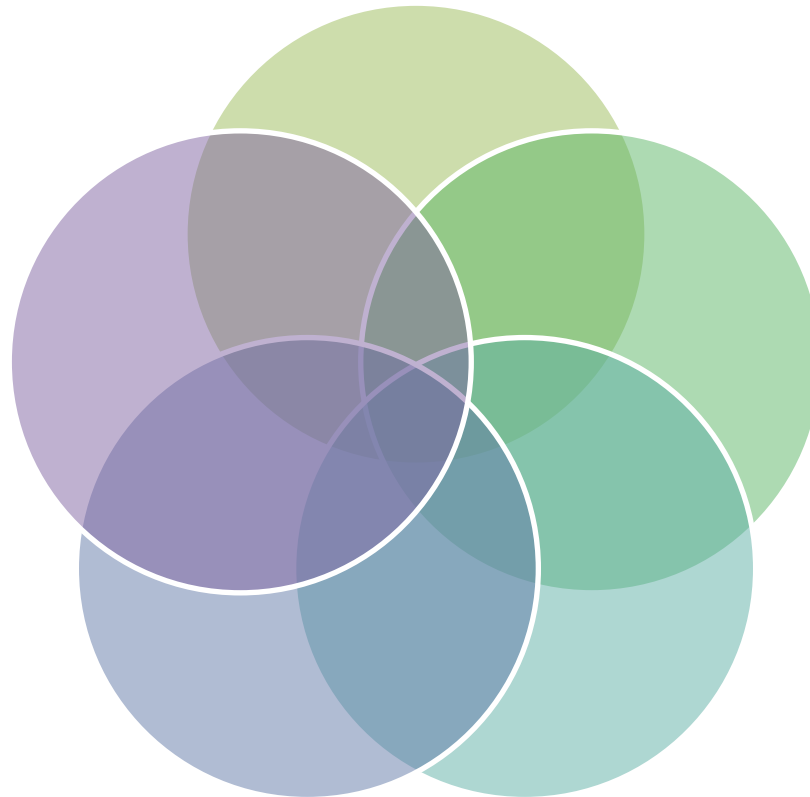
Long-term, managed preservation of Environmental Data	SAEON will require an archiving facility that can preserve data for the long term (beyond 20 years from the present).
Improved Indexing and Searching Capabilities	SAEON will, in time, replace its current folder-based meta-data repository with an object-relational database capable of supporting improved indexing and searching algorithms, using controlled vocabularies and eventually semantic ontologies.
Client-side Map Image processing	Possibilities exist for leveraging client-side web map service images to develop statistical cluster analysis between layers obtained from distributed sources.
Improved visualisation of large databases	CHPC will be doing work in this regard, and SAEON will be able to utilise such services to apply to large databases (for example NetCDF data sources).
Standards-based structured data aggregation	Opportunity exists to test the aggregation of physically disparate, but logically related databases.
Networked Databases	Concepts of network databases allow re-utilisation of data in distributed locations to fit many semantic models. Opportunity exists to evaluate and test.
Processing of physically distributed data.	Several applications require the processing of spatial data that is physically distributed. SAEON can develop prototypes as reference implementations for the standardised implementation of Web Processing Services.
Augmented reality and semantic interoperability	User experience of portals can be improved by implementation of “augmented reality” principles.

Potential Programmes

Data Processing and
Preparation

Citizen Science/
Live Capture/
E-Research

Meta-Data
Harvesting and
Indexing R&D



Exploratory
Modelling

Long-Term Data
Preservation

We need your help ...

Free
T-Shirt...

...with your
first meta-data
submission

Fights Invader
Bush ...



Braai Hard