Belmont Forum Scoping Workshop

Thoughts on Important Focus Areas for a Smaller (Southern) Funder

28-29 November 2016

Wim Hugo (NRF/ SAEON, ICSU-WDS)

Andrew Kaniki, Jonathan Diederiks, Peter Shisani, Lazarus Matizirofa (NRF)

DISTRIBUTED QUERY/ SUBSET/ DOWNLOAD/ PROCESSING

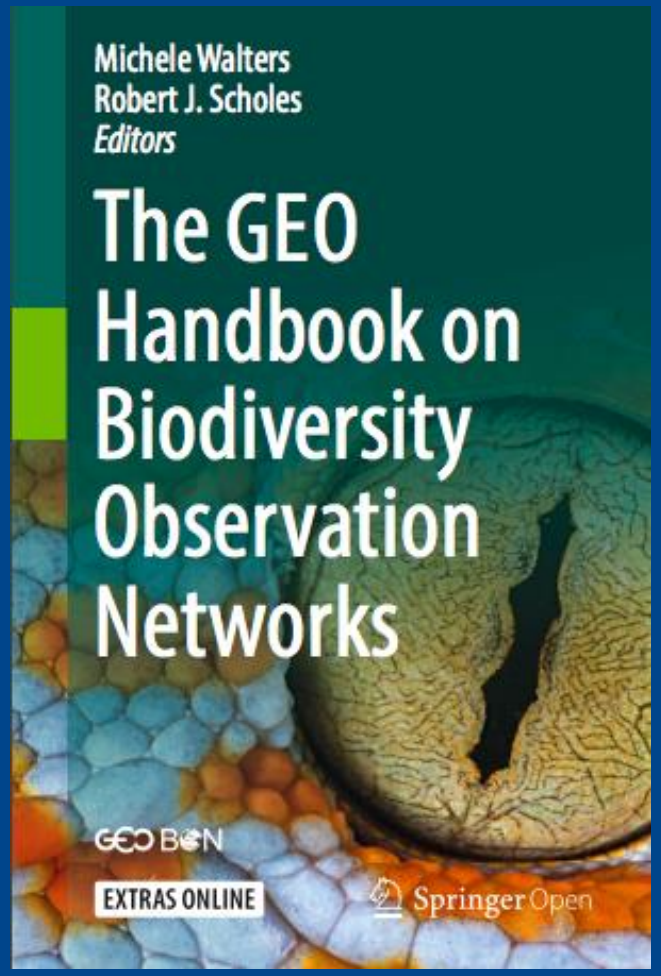
This technology is important for the typical distributed infrastructure in South Africa (and by extension regional or global initiatives).

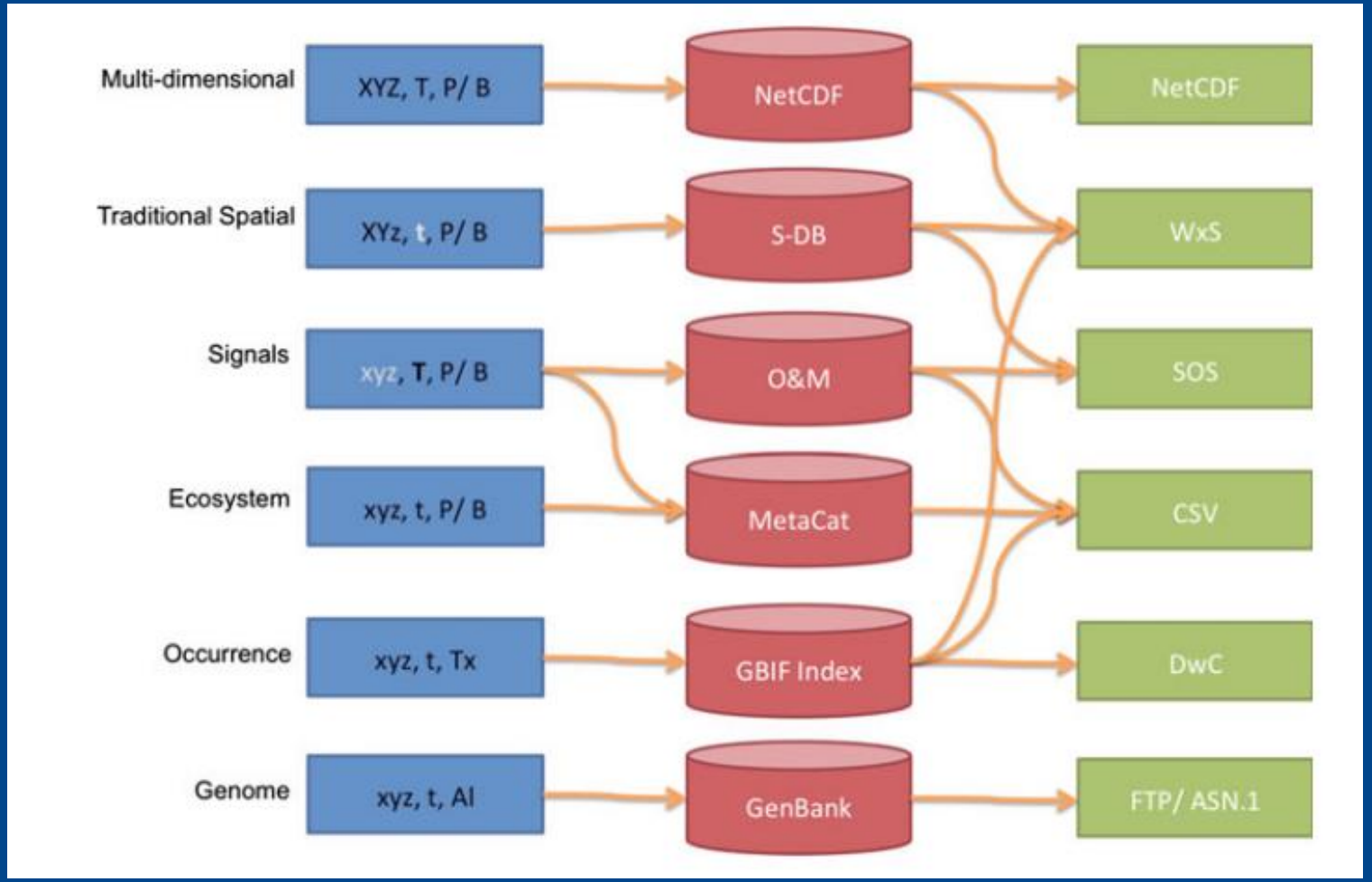
Protocols, standards, and implementations are required for those families of data where true distributed query, subset, and download services are not yet available.

This is more complicated than just “moving computation to data” – large data sets may not be in the same location and some complex analyses require more than one data set as an input.

DATA FAMILIES AND DIMENSIONALITY

- Spatial Coverage: XYZ
- Temporal Coverage:
  - T (continuous or near-continuous);
  - t (discrete)
- Thematic or Semantic/Ontological Coverage





SEMANTIC INTEROPERABILITY

Syntax –Schema - Semantics

Develop brokering services for important semantic name services (such as taxonomy) and contribute the research outputs to global platforms (such as Global Names Architecture (GNA))

Develop a common referencing system for different data families to link virtually via 3-dimensional space, time, and thematic referencing – essentially virtual array database technology. Also requires 3-D on-the-fly re-projection. Ideas include teaching mainstream spatial technologies how to serve virtual array data.

http://link.springer.com/book/10.1007%2F978-3-319-27288-7

OPEN SOURCE COLLABORATION

South African e-infrastructure funding is best applied as a partner in larger open source e-infrastructure projects.

Generic examples include OPENAIRE (D-NET) and DataVerse, but could extend to problem- or domain-specific infrastructures (e.g. Globe)

Requires collaborative, long-term infrastructure funding (not state of the art, not research).

BROKERING FRAMEWORK

Develop a framework for brokering components aligned with important open source software offerings and regional or global e-infrastructures. Contribute to RDA efforts.

Example: G-CAT (basis of GEOSS Broker), Pangaea, DataOne and EarthCube

CAPACITY BUILDING

Funding for development of courseware and support platforms for capacity development aligned with international good practice and widely used software platforms.

Include data management and principles into Honours-level courses

OPEN SCIENCE

Open access licenses and open data policies are a prerequisite for data sharing, but this does not always incentivize researchers and conditions can only be imposed on grant-funded research.

Three or four open science licenses required in addition to Creative Commons to deal with legitimate constraints.

Mechanisms (incentives, automated discovery, …) are required to cast the net wider.

AUTOMATED SCIENTIFIC WORKFLOW

Publication of growing data volumes will increasingly be constrained because of availability of scientists. Steps in the scientific workflow can be automated to boost capacity - especially adduction steps (hunches and pattern recognition).

GLOBALLY AGREED FRAMEWORKS FOR DECISION SUPPORT

Bridging the gap between scientific evidence and policy or management action requires a pre-defined framework for each major policy theme. Research is needed to define generic mechanisms for such frameworks and to develop frameworks for specific cases.

Example: CMIP/ IPCC

Example: ARCTIC/ Pan-ARCTIC

EXAMPLE FRAMEWORK

Integrated environmental, engineering, social, economic, and population health data systems:

Risk and Vulnerability Assessment in South and Southern Africa

Rogers, A & Hugo, W 2016

In preparation