



The RESEARCH DATA ALLIANCE

GEO BON Workgroup 8

WG: Brokering Governance
Wim Hugo – ICSU-WDS/ SAEON
/ GEO BON

research data sharing without barriers
rd-alliance.org

GEO BON

<http://www.earthobservations.org/>

<http://www.earthobservations.org/geobon.shtml>

The Group on Earth Observations Biodiversity Observation Network – GEO BON – coordinates activities relating to the Societal Benefit Area (SBA) on Biodiversity of the Global Earth Observation System of Systems (GEOSS). Some 100 governmental, inter-governmental and non-governmental organizations are collaborating through GEO BON to organize and improve terrestrial, freshwater and marine biodiversity observations globally and make their biodiversity data, information and forecasts more readily accessible to policymakers, managers, experts and other users. Moreover, GEO BON has been recognized by the Parties to the Convention on Biological Diversity.

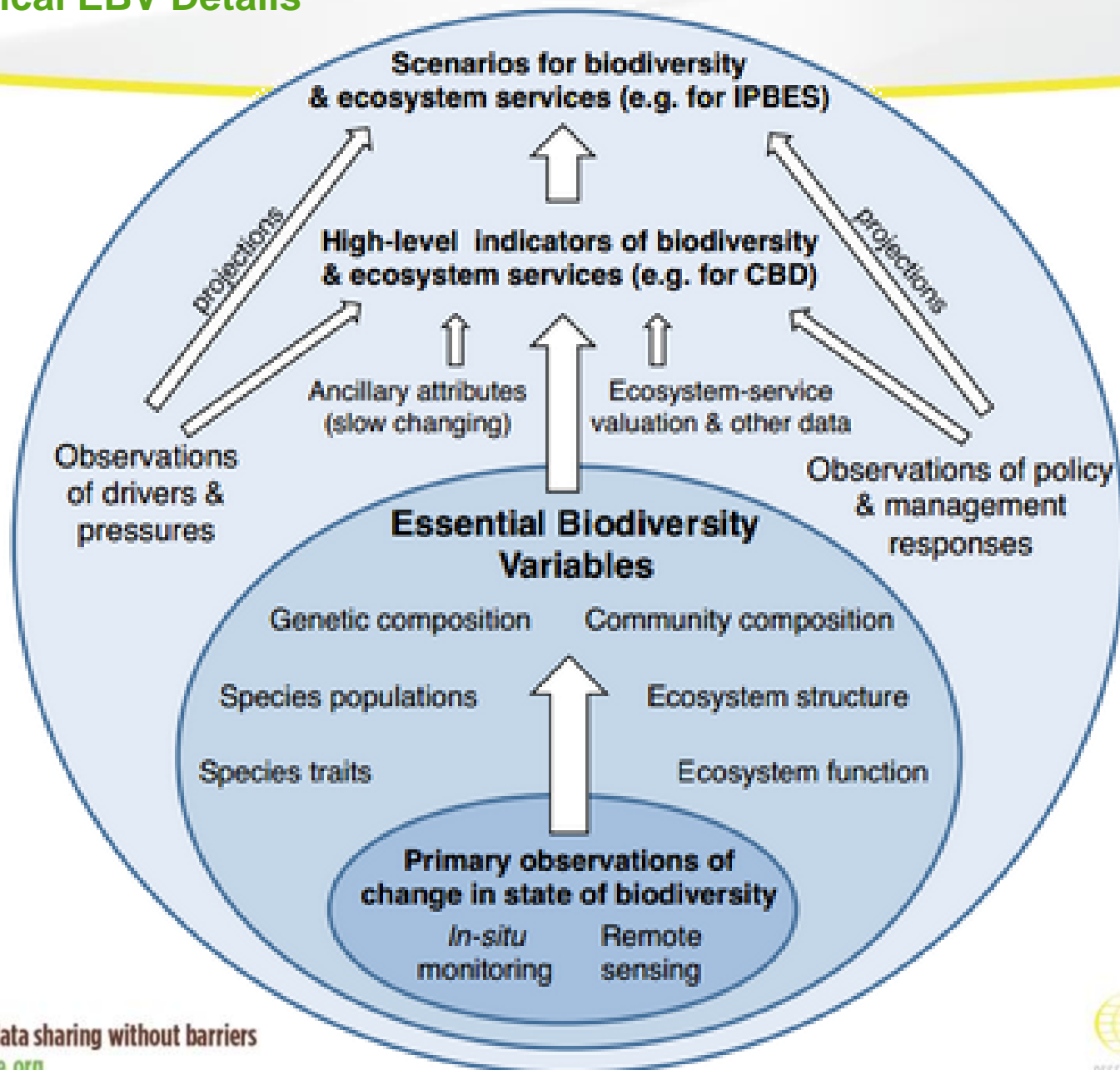
GEO BON has a Manifesto ...

It is possible, desirable, and in the public interest to:

- Ensure that scientific data is described properly, preserved properly, and discoverable;
- Once discovered, its utility, quality, and scope can be understood, even if the data sets are large;
- Once understood; it can be accessed ***freely and openly***;
- Once accessed, it can be included into distributed processes, preferably automatically, and on large scales;
- Once processed, the knowledge gathered can be re-used.

... across multiple domains and dissemination channels.

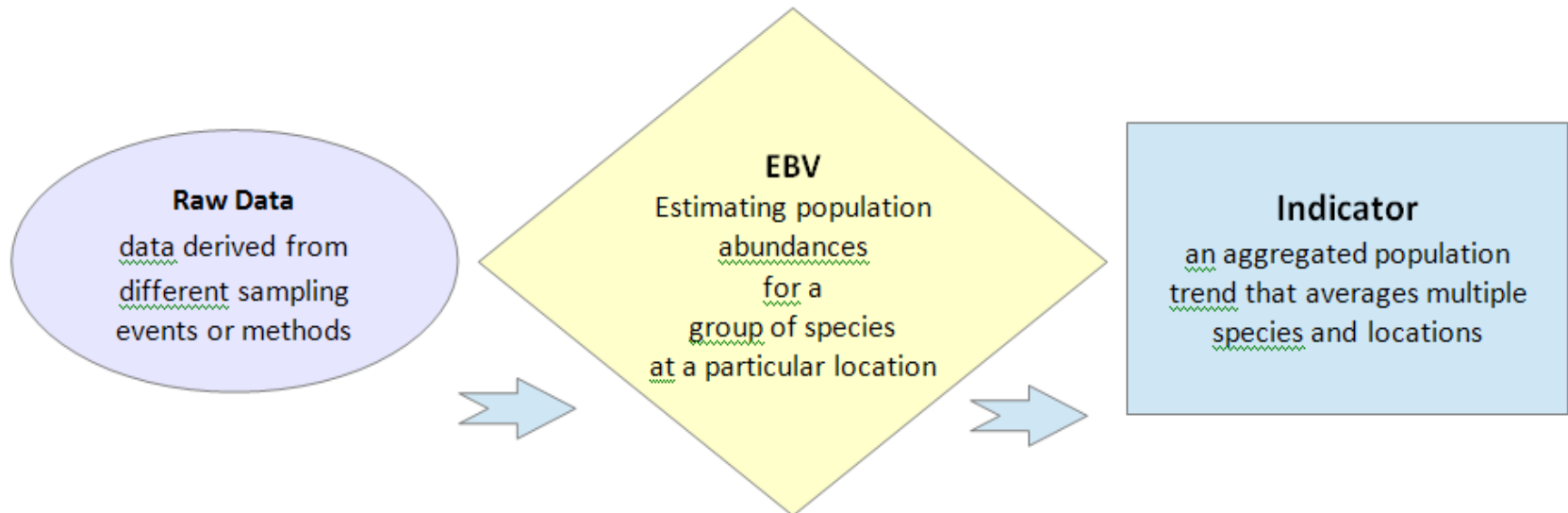
Typical EBV Details



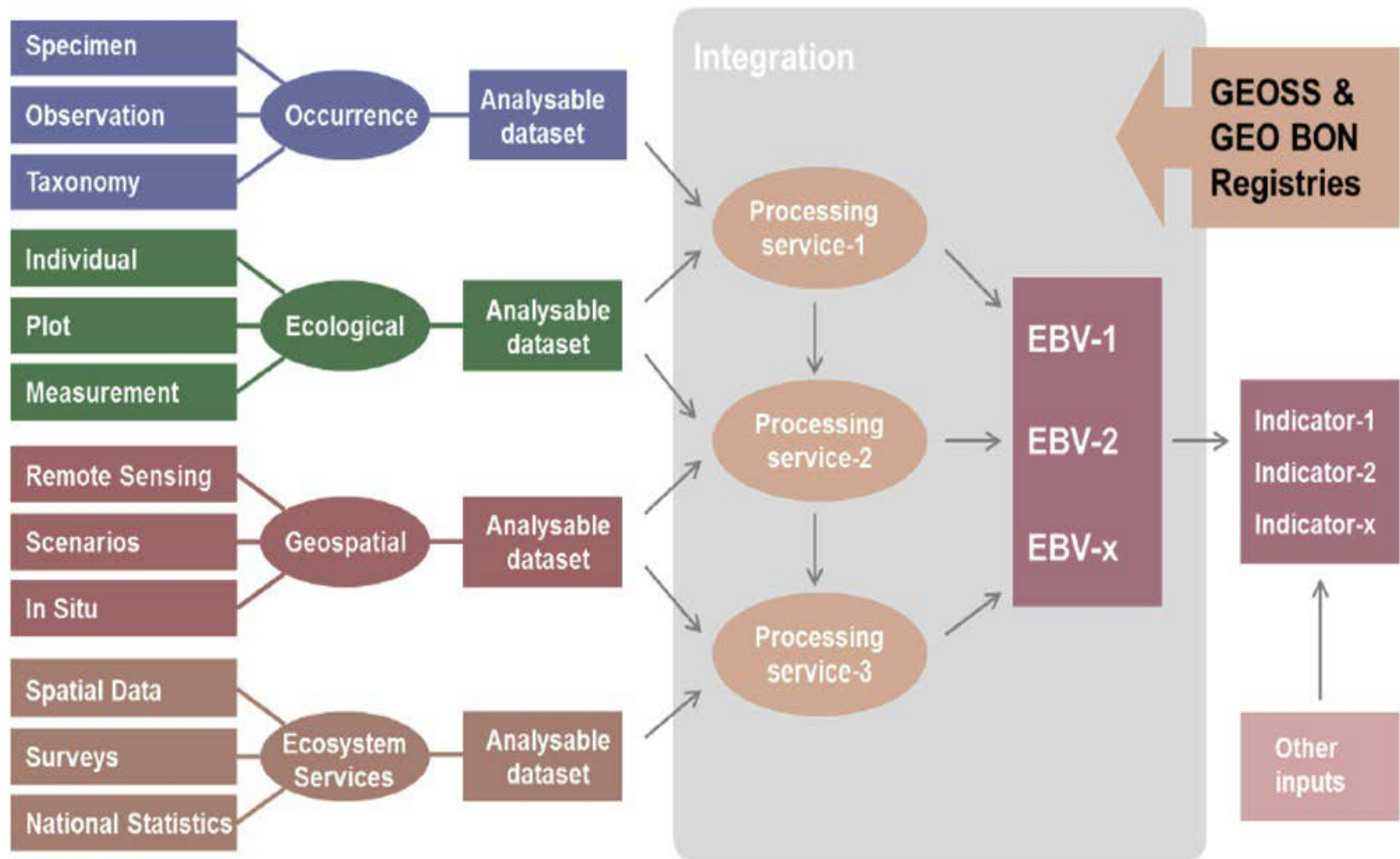
Typical EBV Details

EBV Class	EBV
Genetic Composition	Allelic Diversity for Selected Species
	Breed and Variety Diversity
Species Populations and Ranges	Abundances for a selected set of species
	Distributions for a representative set of species
Species traits	Phenology of selected functional groups
	Body Mass for Selected Species
Community Composition and Interaction	Overall taxonomic diversity for selected locations
	Species interactions
Ecosystem Extent and Structure	Ecosystem extent and fragmentation for a range of ecosystems
	Ecosystem structure
Ecosystem function and processes	Net primary productivity
	Nutrient retention

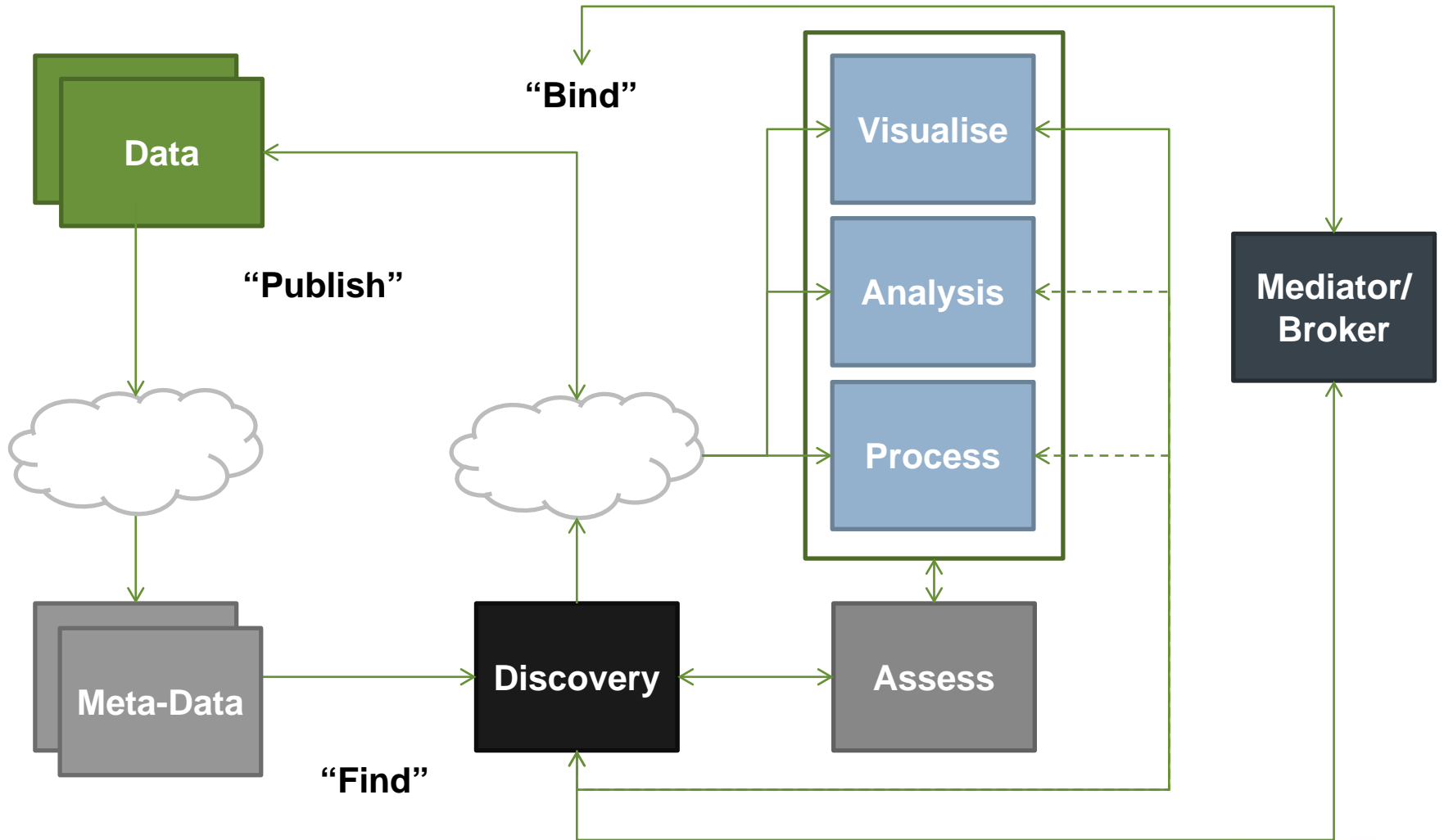
Example: Simplified Objective



Generic Use Case



Main Components

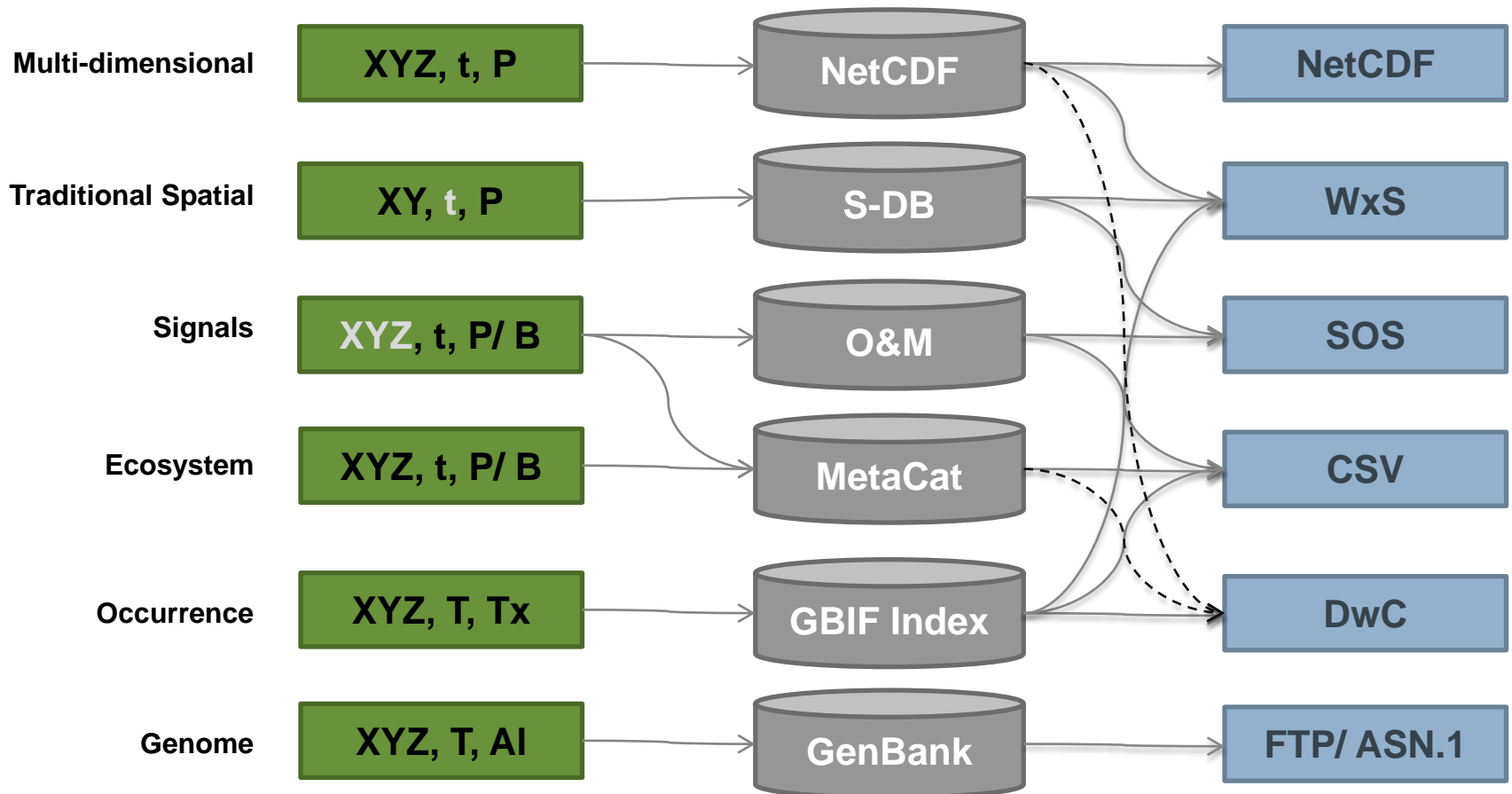


Generic Dimensions of Data

- Spatial Coverage
 - XYZ
- Temporal Coverage: T
- Topic or Semantic/ Ontological Coverage
 - P: Phenomenon
 - mostly physical, chemical, or other contextual data
 - B: Biological
 - Tx: Species and Taxonomy (with some extensions)
 - Al: Allele/ Genome/ Phylogenetic
- Each unique combination of these, supported by a vocabularies/ ontology is a generic data family

Continuous or Near-Continuous: Uppercase
Discrete or dispersed: Lowercase

Some Generic Data Standards and Interoperability Requirements



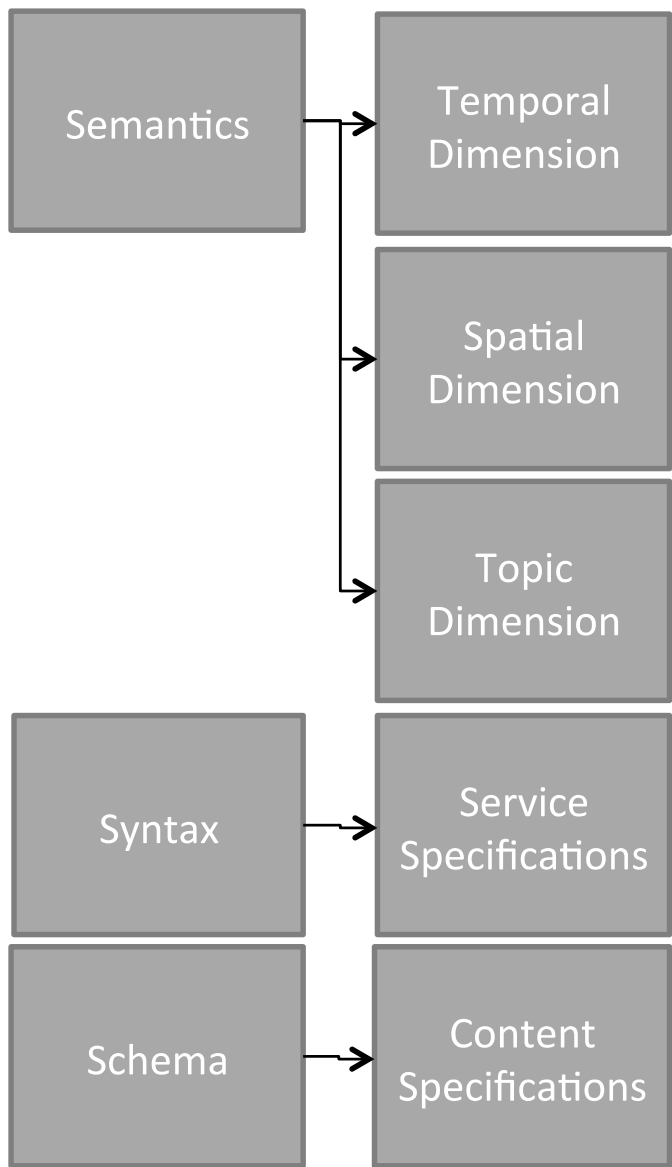
Status: Working Demonstrator

- Extending functionality as and when we have opportunity within existing projects. No dedicated funding.
- SAEON is building a loosely coupled open prototype
- EU BON is building a closely coupled operational system
- Supported by ongoing efforts in GBIF, DataOne, and other stakeholders

Areas of Collaboration: GEO-BON Workgroup 8

- Updates to GEO BON Handbook WIKI pages on standards, software, and best practices
- Identify/ Develop Content Standards and Vocabularies for EBVs and Data Families
 - Including name services for
 - Taxonomy
 - Traits
 - Location
 - Time
 - Habitats
 - Species Interaction
 - ...

For Each Data Family...



	Data Services	Meta-Data Services	Visualise / Explore	Analyse	Process
Semantics - Temporal Dimension	O	M	M	M	M
Semantics - Spatial Dimension	O	M	M	M	M
Semantics - Topic Dimension	O	O	O	M	M
Syntax - Service Specifications	M	M	M	M	M
Schema - Content Specifications	O	M	O	M	M

Typical Guidance

For Each EBV ...

EBV 22	<i>Ecosystem Composition by Functional Type</i>
Class	<i>Ecosystem Structure</i>

Measurement and Scalability	Functional types can be directly inferred from morphology (in situ) or from remote sensing.
------------------------------------	--

Feasibility	Implicitly part of current ecosystem maps. Some models (e.g. DGVMs, marine ecosystem models) are based on functional groups.
--------------------	--

Preferred Data Family	Remotely Sensed Data: Multidimensional Data In-Situ Data: Indeterminate Schema
Alternatives	Spatial Data: Traditional Vector/ Raster Data

Meta-Data Preference	Remotely Sensed Data: Convention to be decided In-Situ Data: EML	>	Software	NetCDF can support a convention
Alternatives	Spatial Data: ISO 19115	>		GeoNetwork GeoNode INSPIRE

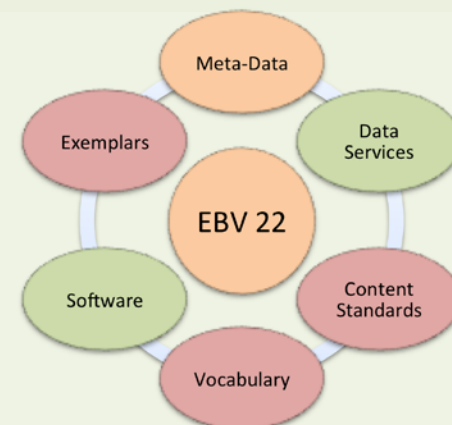
Data Services Preference	Remotely Sensed Data: netCDF/ HDF 5 In-Situ Data: EML downloads as CSV	>	Software	NetCDF can support a convention
Alternatives	Spatial Data: OGC Web Coverage Service	>		GeoNetwork GeoNode INSPIRE

Data Content Standards	None	>	Resources	None
-------------------------------	------	---	------------------	------

Existing Vocabularies	None	>	Resources	None
------------------------------	------	---	------------------	------

Required Vocabularies	Functional Type			
------------------------------	-----------------	--	--	--

Exemplars	None	>	Resources	None
------------------	------	---	------------------	------



Translations Required	NetCDF -> WMS NetCDF -> SOS CSV -> SOS CVS - WMS/ WFS
------------------------------	--



research data sharing without barriers
rd-alliance.org

... described properly, preserved properly, and discoverable

- *Meta-data standards implied.*
- *Harvesters, brokers, and meta-data interoperability implied.*
- *Persistent identifiers implied.*
- *Protocols and standards for data exchange/ uploads implied.*
- *Preservation standards and formats implied.*
- *Tools and approaches to make searches more efficient (vocabularies, ontologies, dealing with massive meta-data collections, ...).*
- *Sustainable, accredited data centers and long-term archives are implied – depositor SLA and contract.*

How long is the ‘Long Term’?

Who funds this?

Distributed or Centralised Infrastructure?

... its utility, quality, and scope can be understood ...

Implies:

- *Visualisations, Collations, Data Exploration Tools,*
- *Utility metrics ('Like' ..),*
- *feedback on quality, quality metrics and standards,*
- *viewing search results in relation to reference spatial, temporal, and ontological/ taxonomic coverages,*
- *ability to dynamically extract 'thumbnail' views of large datasets, ...*

'Big' Data: Different protocol – not HTTP but maybe RPC?

... accessed freely and openly ...

Implies:

- *Standardised services, licenses and policies,*
- *Standardised, generic conditions and exceptions to free and open access,*
- *Simplified, effective distribution channels, even if costs are involved, ...*
- *Equal opportunity to discover and apply.*

... included into distributed processes ...

Implies:

- *Persistence of mash-ups, derived works, and mediations,*
- *Web context documents,*
- *Web processing services,*
- *Standards and guidelines for grid computing,*
- *Ability to construct decision support models, indicators, and standardized, interoperable final products, ...*

***What moves? Data, Processes, or Both?
Concept of a 'Distributed Indicator Standard'***

... due recognition is afforded to the creators ...

Implies:

- *Data publication and citation,*
- *Data and service citation indices,*
- *Linking to scholarly articles, ...*

... the knowledge gathered can be re-used ...

Implies:

- *Defining and storing templates and examples of finished work, processes, mash-ups, ...*
- *Liberalising Meta-Data and building formal knowledge networks, ...*

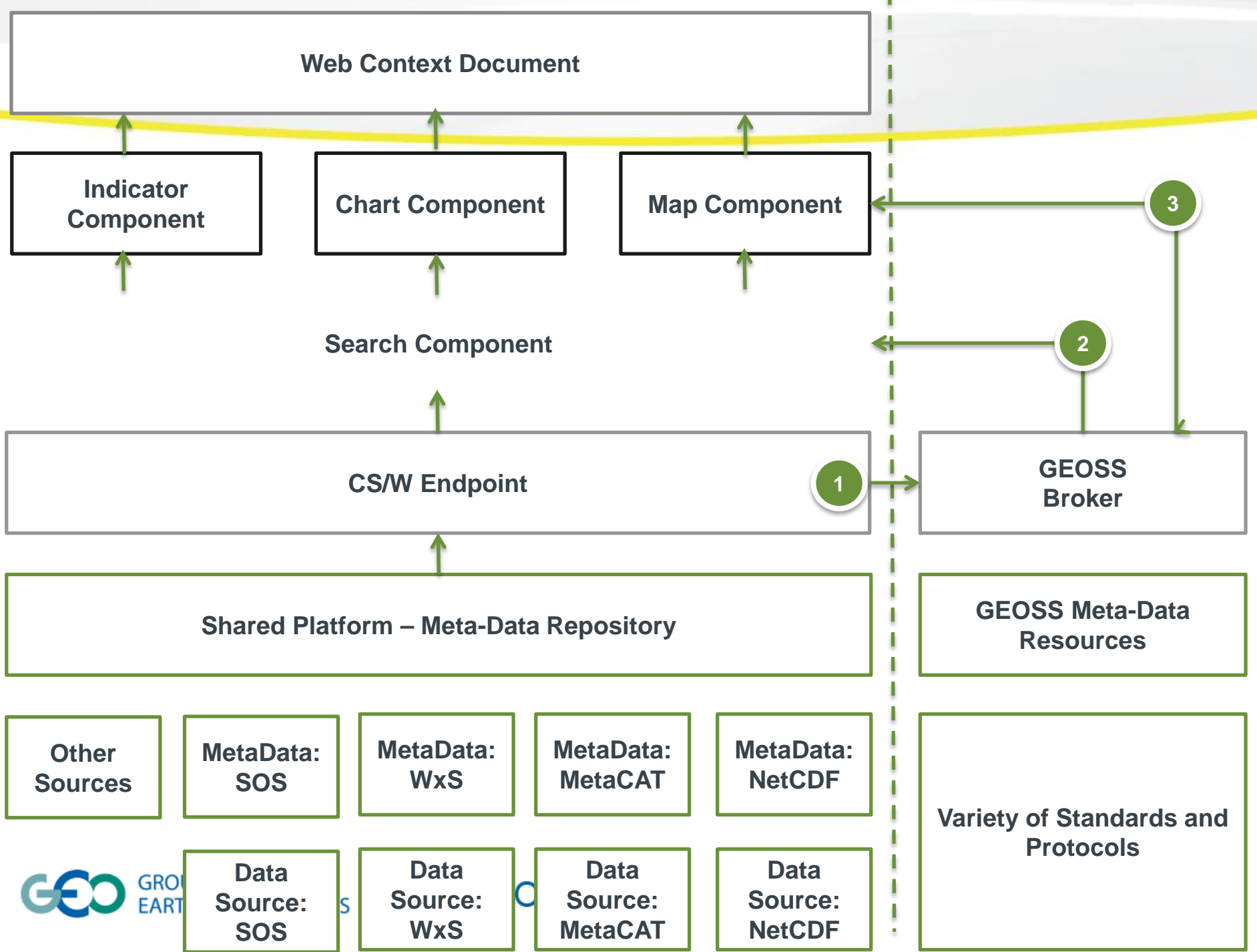
***ICSU-WDS Working Group on Knowledge Networks
(seeking a home in an RDA Collaboration)
Collaboration with RDA on Trusted Digital Repositories***

... against a backdrop of ...

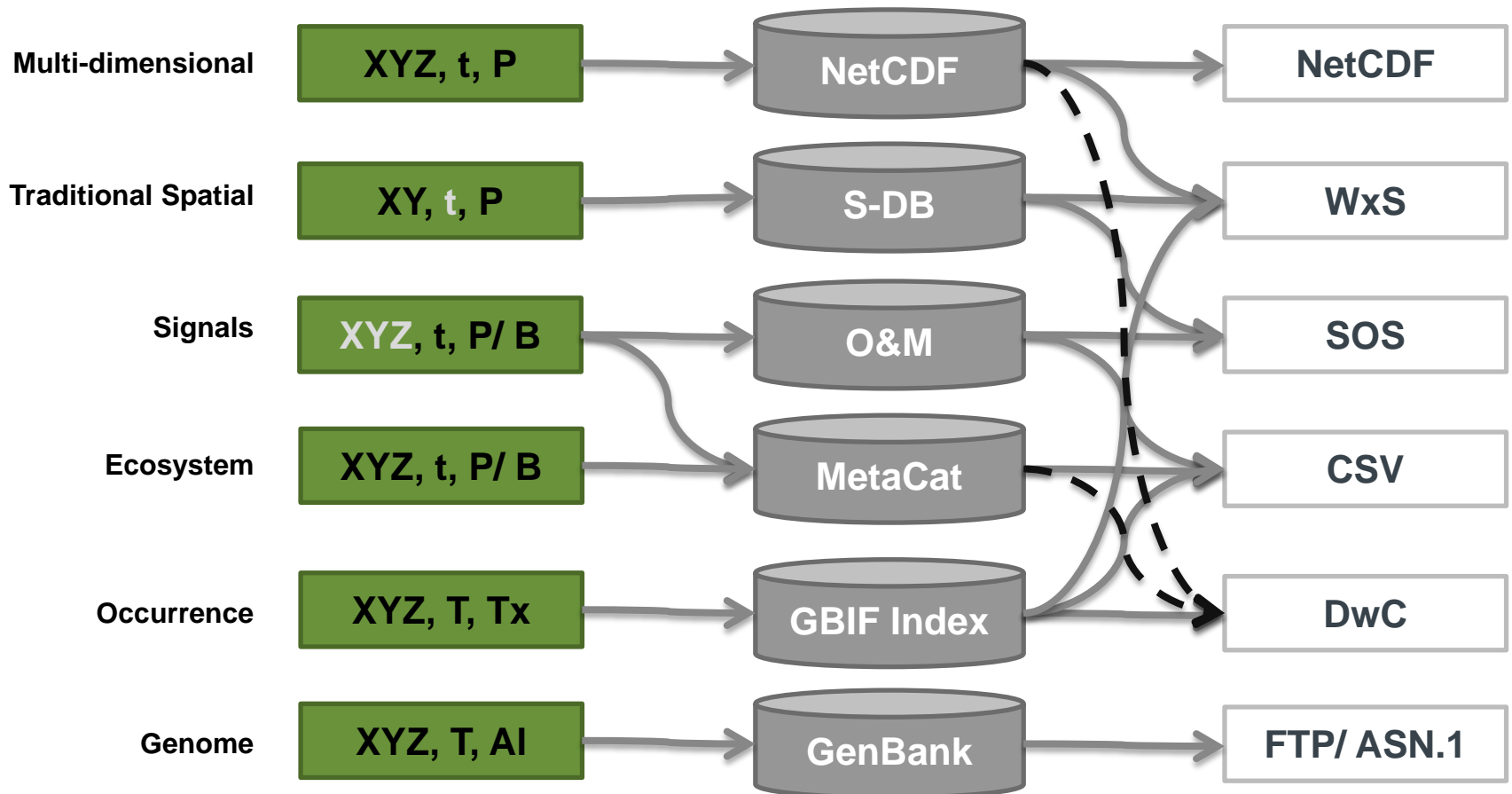
- *The push to extend formal meta-data with linked open data;*
- *The increased availability of crowd-sourced and citizen contributions;*
- *A proliferation of devices and sensors;*
- *And the construction of knowledge networks.*

Building Infrastructure

- Is NOT a Research Task – it is an Engineering Task
- We can realise large parts of the GEO BON infrastructure already
- Issues are not so much technological as institutional
- Our first principle should be to engage and amend existing infrastructure components
- Infrastructure cannot be funded through projects or through voluntary contributions.



Some Generic Data Standards and Interoperability Requirements



Use Case to be achieved

