LSI-VC-16
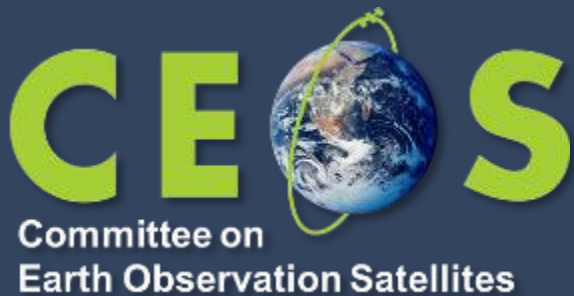
CEOS Interoperability

Handbook 2.0

Nitant Dube &
Tom Sohre,
WGISS

Agenda Item 7.1

LSI-VC-16, Canberra

23 - 25 September 2024

CEOS
Committee on
Earth Observation Satellites

CEOS Interoperability Handbook V 2.0

❖ Objective

- ▪ CEOS Interoperability Handbook Version 2.0 should provide <u>guidance to the organizations</u> for development of **Interoperable Data** and **Services** and help them in measuring their **maturity level**

Interoperability Handbook

Maturity Matrix

Approach

❖ Use the **interoperability framework** and develop a **handbook**, which will provide an overview of existing capabilities within CEOS, which could be used either by CEOS agencies or external users to develop approach for interoperability or to improve/develop interoperability for Earth Observation Data and Services

❖ We do not want to tell **How to Do**, we want to tell **What to Do**, with possible links to **How to Do**

Interoperability Framework : Factors

| Vocabulary (Semantics)	The (narrow) semantic aspect refers to the naming and meaning of data elements. It includes developing, harmonising, and maintaining vocabularies and schemata supporting provision, exchange, and analysis of data, and ensures that terms and data elements are understood in the same way by all communicating parties
Architecture	Architecture describes the organisational structure of concepts, processes, and assets, including data and workflows. It comprises the structural aspects of models and standards that govern the collection, storage, arrangement, integration, and use of data.
Interface (Accessibility)	Data exchange protocols, and application interfaces. These provide the means necessary to access and exchange data.
Quality	References are data and schemes that are used as benchmarks for (observational) data comparison or analysis. This could include instances such as geographic locations, product numbers, or official (authoritative) data and statistics.
Policy	Legal frameworks, policies and strategies regulating the relation between the different stakeholders.

Overview

Handbook Chapter Layout
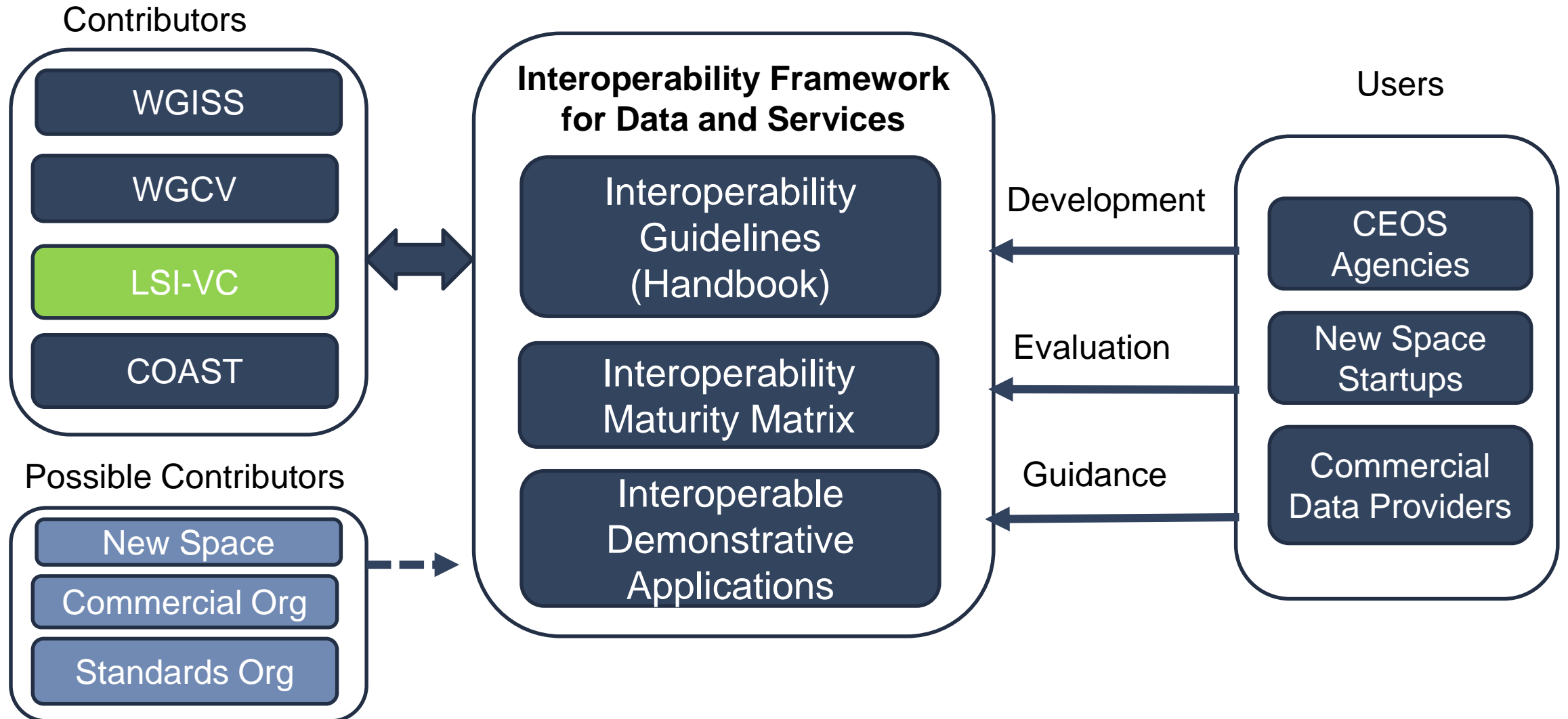
❖ Chapters

- Chapter-1 : Introduction
  - Requirements/Need, Scope, Audience
- Chapter-2: Interoperability Framework
  - Genesis of the Factors, Brief overview of the factors
- Chapter-3: Vocabulary (Semantics)
- Chapter-4: Architecture
  - Pre-Processing, Analysis Ready Data, Data Cubes, Cloud Compatibility; Data Stewardship; Data Dissemination
- Chapter-5: Interface (Accessibility)
  - **Discovery:** Metadata, STAC; **Access:** API, https; **Authentication**
- Chapter 6: Quality
  - Quality Evaluation and Tags, Calibration, Validation
- Chapter 7: Policy
  - Open Data Policy, Open Science, Open Source Science
- Chapter 8: Conclusion and Future Scope
- Annexures

Anticipated LSI-VC Contributions

```
Vocabulary ── Architecture ─┬─ Analysis Ready Data
                            ├─ Data Cubes
                            ├─ Data Archival
                            └─ Data Dissemination

Quality ─┬─ Calibration
         └─ Validation
```
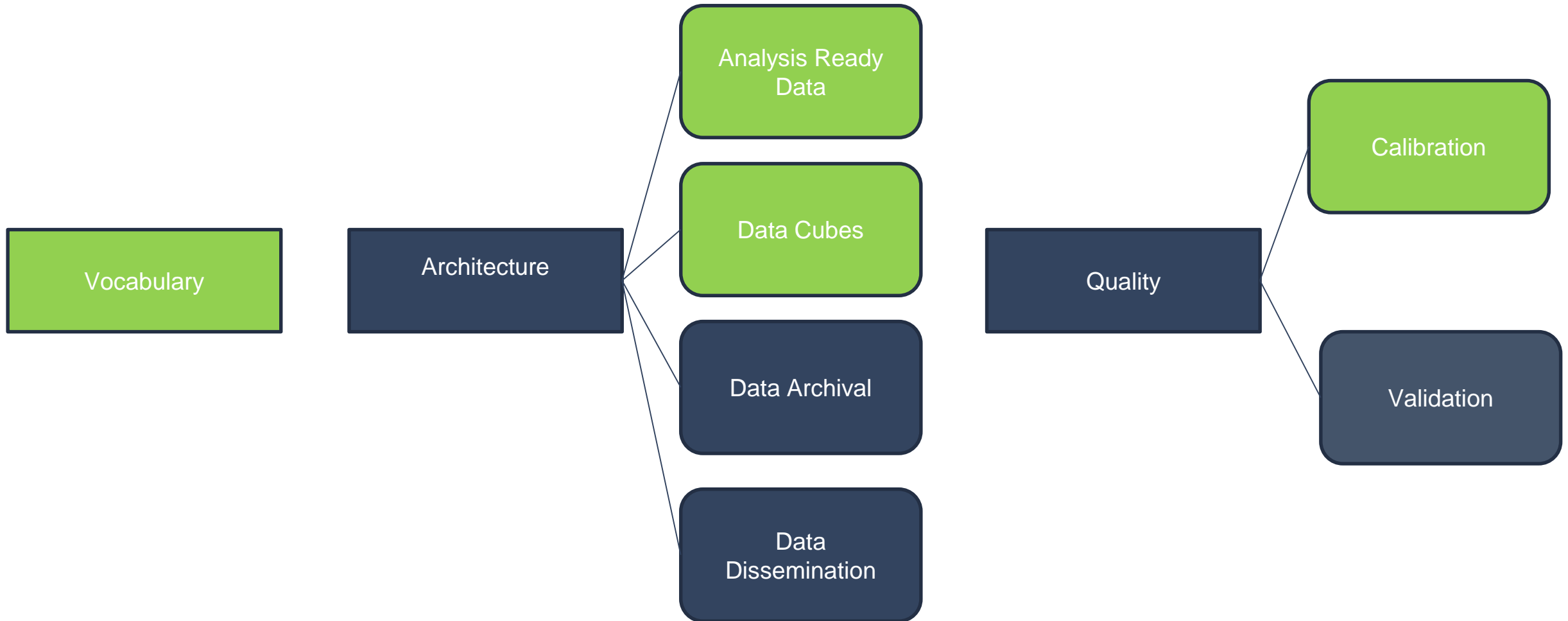
Vocabulary

❖ Recommendations are primarily based on work done and paper published by **Peter Strobl , Woolliams and Katrin Molch**

❖ Vocabulary also termed as semantics helps in providing a **consistent terminology for communication** between different stakeholders including humans and machines for acquisition, processing, archival, dissemination and analysis of Earth Observation data.

❖ Recommendations

  o Thesaurus should be the preferred way of implementing vocabulary in an organization. Use **ISO 25964-1, 2011** for developing Thesaurus and **ISO 25964-2, 2013** to ensure Interoperability with other vocabularies

  o **CEOS Common Online Dictionary** can be used as a reference for Vocabulary implementation

  o Use existing vocabulary and enrich it by involving a wider community with a balance of **centralized and distributed control.**

  o The terms used in vocabulary should be Consistent and divided into classes such as **Base, Core, Controversial and High Impact**

Vocabulary

❖ **Recommendations** cont..

o **Active web links** to the standard centralized definition of terms should be used as against the documentary standards providing their own list of references.

o Vocabulary should have **version control and change management** at the individual term level and historical and alternative definitions should be linked to current term definition.

o The terms used in the vocabulary should be Interrelated with **clear relationships between other terms** (parent, sibling, child) avoiding, in particular, circular (child becomes parent) relations. Overlaps between terms that are supposed to delineate more generic concepts (siblings) should be avoided or minimized

o The vocabulary should have capability to record **public comment and discussion** and should provide options for adding new terms.

Analysis Ready Data (ARD)
(For discussion)

❖ Analysis Ready Data (ARD) are starting point for interoperability in analysis and hence Data Providers are encouraged to develop ARD.

- ▪ **Recommendations**
  - o **CEOS ARD Framework** should be used as a starting point for development of Analysis Ready Data
  - o CEOS **Product Family Specifications (PFS)** should be used for development of ARD products. In case if a new ARD is to be developed, use PFS template and submit to CEOS for approval
  - o CEOS ARD compliance of the product requires two level of assessments, first is self assessment (**CEOS ARD Self Assessment Guide**) and second is peer review by CEOS Experts
  - o ....

Data Cubes
(For discussion)

- ❖ Data Cubes provide capability to pack a collection of data and provide capability for fast access and analysis
  - ▪ Recommendations
    - ○ The CEOS supported **Open Data Cube** can be taken as a reference for Data cube implementation
    - ○ Data cubes should support **spatial and/or temporal dimensions** and capability for publishing available variables/properties as metadata.
    - ○ Data cubes should **abstract the underlying data storage architecture** to support hybrid data cubes and interoperability among different types of data cubes.
    - ○ Data cube may contain raw sensor data, analysis ready data or decision ready information. **Analysis Reay Data should be preferred** to avoid pre-processing overheads and fast analysis
    - ○ Data cubes should be able **to publish the supported pre-built analysis functions**
    - ○ Data Cubes should provide **REST API and OGC web services** based access for easy integration of end user applications

Calibration and Validation

❖ **Calibration** helps in quantitatively defining a sensor response to known and controlled signal inputs. **Validation** is the process of assessing the quality of the data

❖ Recommendations

   o CEOS CAL/VAL portal can be used as a reference site for insuring interoperability for calibration and validation

   o CEOS/WGCV developed Data Quality Assurance Strategy should be followed for ensuring interoperability

   o CEOS endorsed Cal/Val sites should be used for satellite cross-comparison

Interoperability Demonstrator

E PnP (Earth Observation Plug and Play)

❖ Development of Plug and Play modules to demonstrate interoperability best practices
  ○ Discovery and access of Analysis Ready Data

❖ Possible collaboration with CEOS Analytics Lab (CAL) & **LSI-VC**

CEOS Common Online Dictionary

❖ Develop CEOS thesaurus for access and management of CEOS controlled vocabularies and to demonstrate Vocabulary interoperability best practices

❖ Possible collaboration with **WGCV** and **LSI-VC**

How to Contribute

❖ Join Interoperability mailing list (send request to libby@symbioscomms.com)

❖ Check our development at https://github.com/ceos-org/interoperability-handbook (Most recent updates will not be available)

❖ The Draft Document is available for Edit/Feedback on google drive (link can be obtained for contribution from nitant@sac.isro.gov.in or tsohre@usgs.gov)

Other WGISS Initiatives

❖ **Technology Exploration Interest Group (ikehata.yousuke@jaxa.jp from JAXA)**

- White Paper on AI/ML
  - With increasing use of AI/ML techniques for Earth Observation data, a need was felt to study the capabilities available across CEOS organizations and share AI/ML use-cases and best practices being followed.
  - During Discussions on AI/ML, concept of AI/ML ready data came out
  - A Google survey was circulated, with an aim of capturing what people mean and understand by AI/ML ready data and try to understand if there is any standardization on AI/ML ready data
  - **Possible LSI-VC help/contribution for AI/ML ready data**

Thanks

nitant@sac.isro.gov.in
tsohre@usgs.gov