

Development of a standard for data and metadata preservation (19165-2)

Content specifications for Earth observation data and derived digital products

Presented at “Standard in Action” seminar
TC 211 Plenary Meeting – Maribor, Slovenia

June 5, 2019

H. K. “Rama” Ramapriyan

Hampapuram.Ramapriyan@ssaihq.com

ANSI (United States)

Topics

- **Motivation**
- **Background (history – timeline)**
- **ISO 19165-1 – scope, content**
- **19165-1 → 19165-2**
- **ISO 19165-2 – scope, content**
- **Status**

Why Preserve?

- **Earth observational data, derived products and models are used to answer key questions such as**
 - **How is the global Earth system changing?**
 - **What are the sources of change in the Earth systems and what are their magnitudes and trends?**
 - **How will Earth system change in the future?**
 - **How can Earth system science improve mitigation of and adaptation to global change?**

Why Preserve?

- **Near term - important to provide easy access to the data and services commensurate with current information technology**
- **Long-term - when the focus of the research community shifts toward new missions and observations, it is essential to preserve the previous mission data and associated information**
- **Data are needed for a long time after the principals have moved on**

Examples of National and International Organizations Collecting/Managing Earth Observation Data



NASA Example – Earth Observing System Data and Information System (EOSDIS) - Extensive Data Collection

Started in the 1990s, EOSDIS today has distributed 11,000+ data types

- **Atmosphere**

- » Winds & Precipitation
- » Aerosols & Clouds
- » Temperature & Humidity
- » Solar radiation

- **Ocean**

- » Surface temperature
- » Surface wind fields & Heat flux
- » Surface topography
- » Ocean color

- **Cryosphere**

- » Sea/Land Ice & Snow Cover

- **Land**

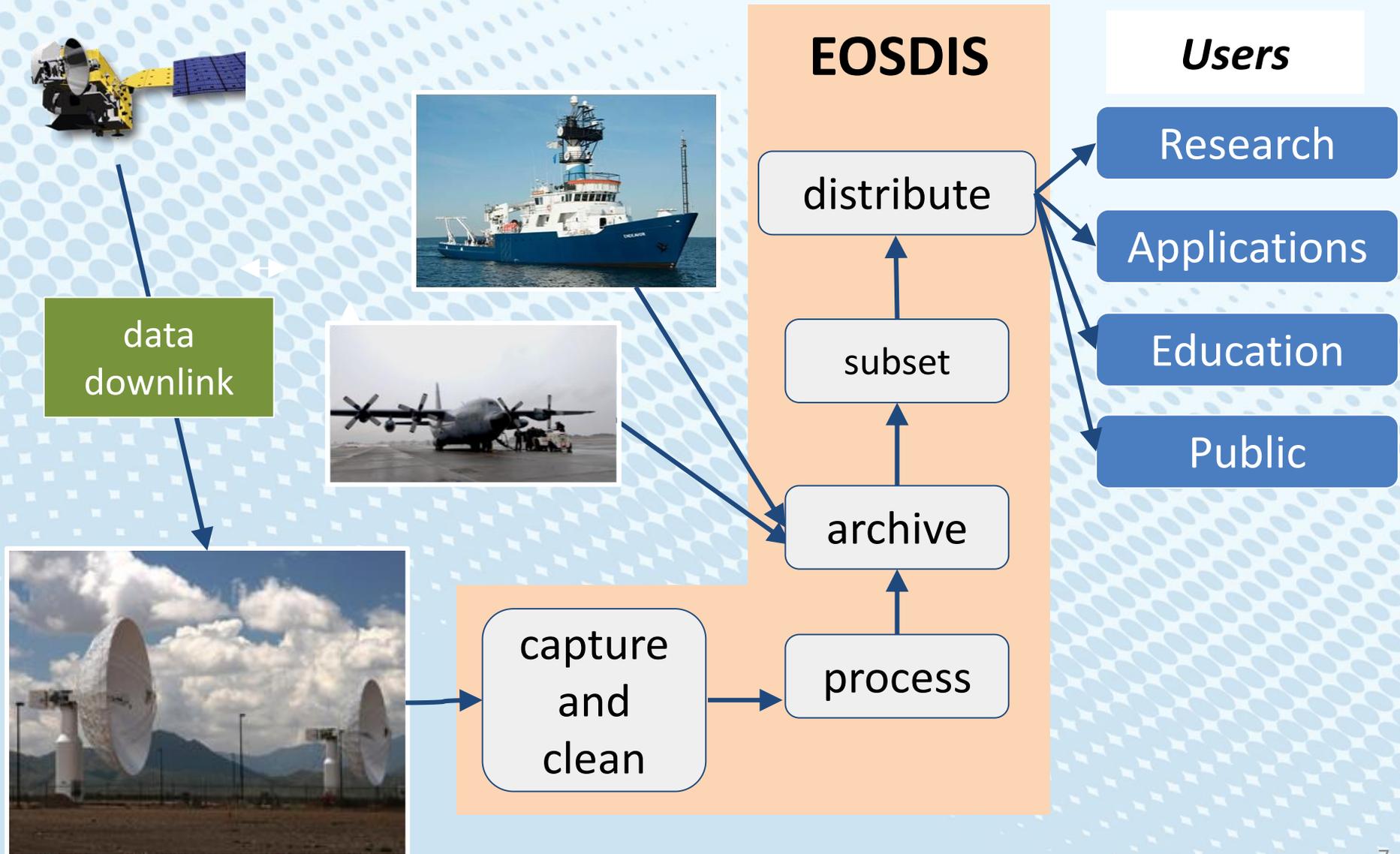
- » Cover & Usage
- » Soil Moisture
- » Topography & elevation
- » Temperature

- **Human Dimensions**

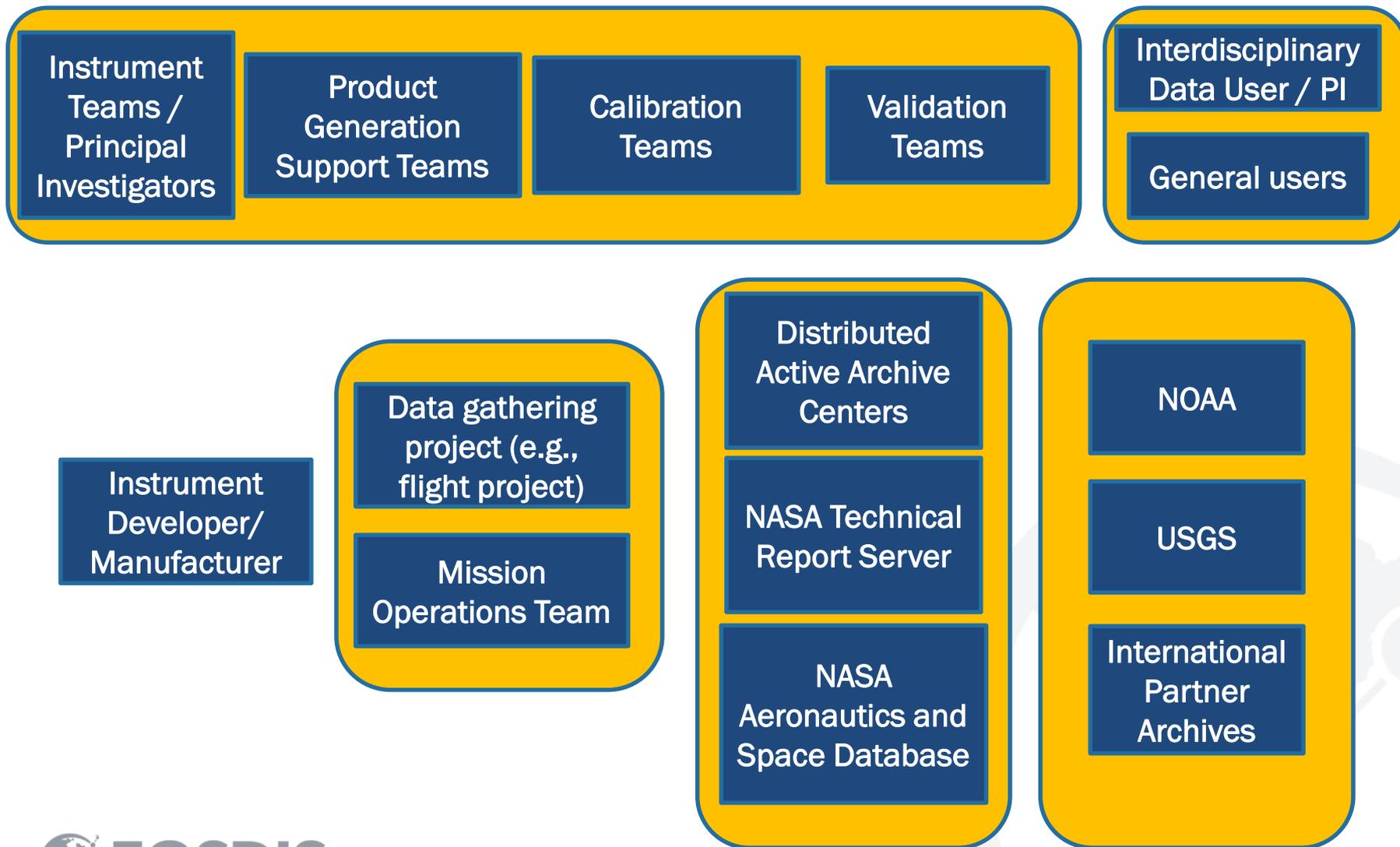
- » Population & Land Use
- » Human & Environmental Health



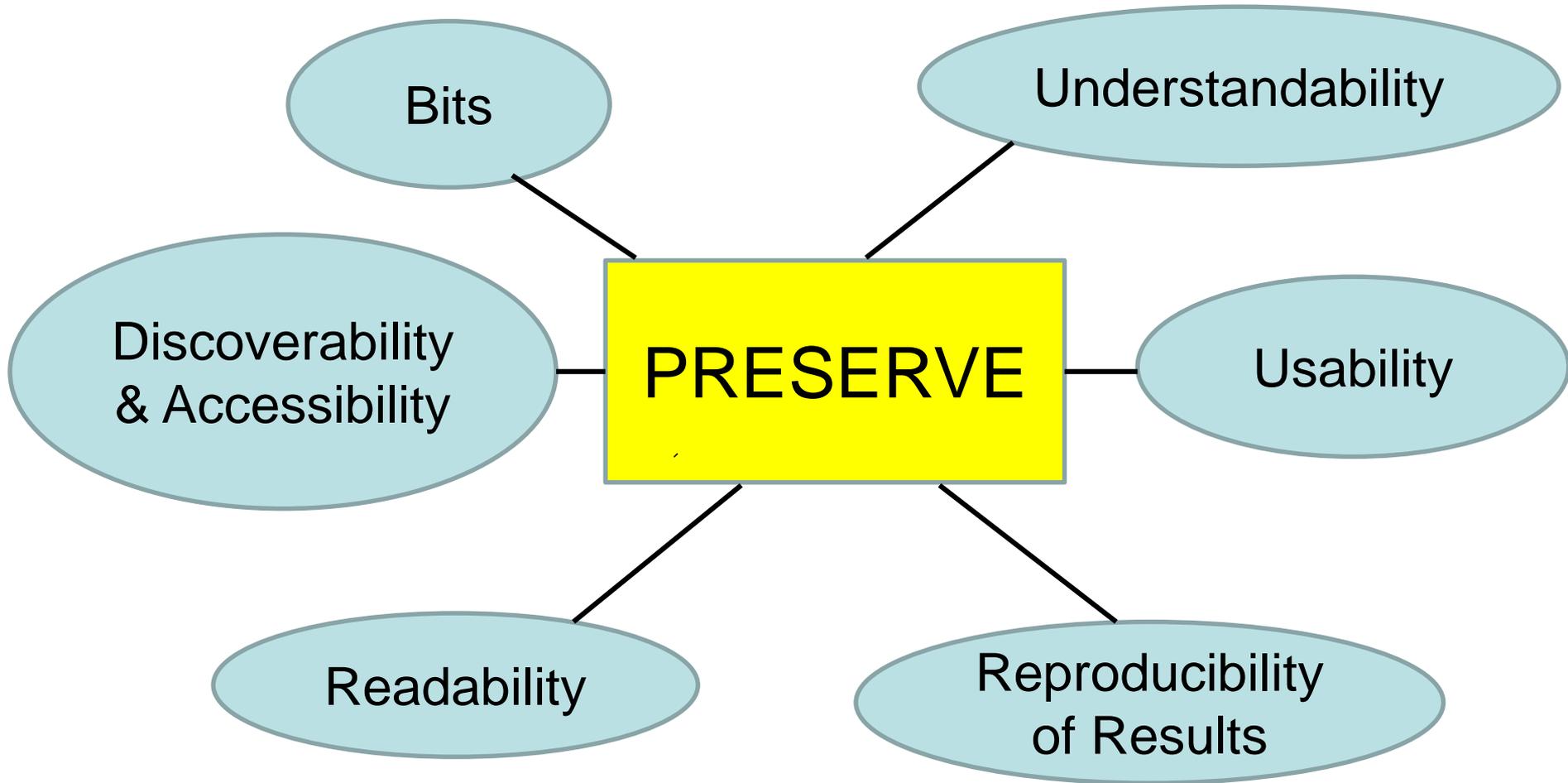
Earth Observing System Data and Information System (EOSDIS)



Organizations holding relevant content during project life cycle – NASA Example



Preservation Implies...

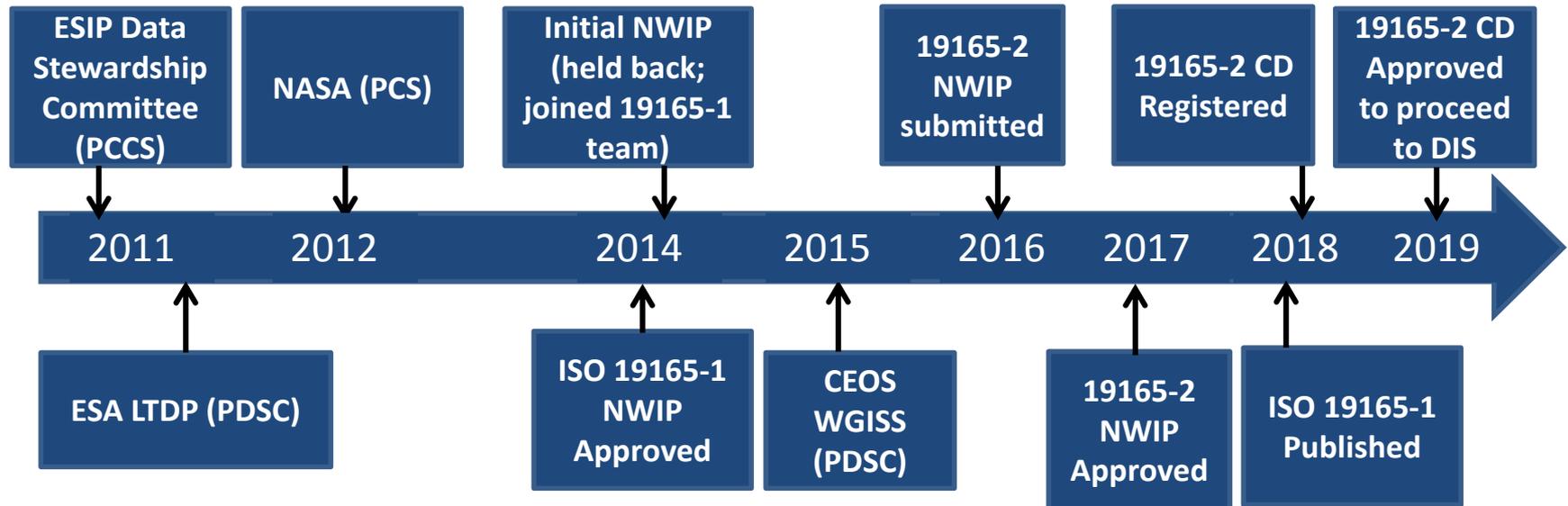


... for future generations to be able to reuse data and derived products

Background

- Importance of preservation of data and metadata for reusability, understandability and ideally reproducibility has been recognized by a large number of organizations
- Focused attention since 2011 to Earth observation data and metadata preservation

Time-Line



CD	Committee Draft	NASA	National Aeronautics and Space Administration
CEOS	Committee on Earth Observation Satellites	NWIP	New Work Item Proposal
DIS	Draft International Standard	PCCS	Provenance and Context Content Specification
ESA	European Space Agency	PCS	Preservation Content Specification
ESIP	Earth Science Information Partners	PDSC	Preserved Data Set Content
LTDP	Long-Term Data Preservation Program	WGISS	Working Group on Information Systems and Services

ISO 19165-1

- **Title – *Geographic information -- Preservation of digital data and metadata -- Part 1: Fundamentals***
- **Summary of scope**
 - **Defines requirements for long-term preservation of digital geospatial data**
 - **Data also include metadata, representation information, provenance, context and any other content items that capture the knowledge that are necessary to fully understand and reuse the archived data**
 - **Standard also refers to characteristics of data formats useful for the purpose of archiving.**

ISO 19165-1: Table of Contents (1 of 2)

- **Foreword, Introduction, Scope, Conformance, Normative references**
- **Terms and definitions, Symbols and abbreviated terms**
- **Preservation**
 - **Prioritization**
 - **Structure**
 - **Data Format**
 - **Data Structures**
 - **Software and Algorithms**
 - **Properties of geospatial data**
 - **Gold copy**
 - **Rights/licensing**
 - **Time**

ISO 19165-1: Table of Contents (2 of 2)

- **Geospatial information model**
 - Overview
 - Designated community
 - Metadata
 - 11 subsections, e.g. preservation metadata classes, format specifications, value & strategy, fixity
 - Open Packaging Convention (OPC)
- **Annexes**
 - Abstract test suite
 - Data dictionary
 - Case-specific archival concept
 - Functional requirements for a preservation archive
- **Bibliography**

ISO 19165-1 → 19165-2

- Section 7.3.1 of ISO 19165-1 states:
 - specific content items needed to preserve the full provenance and context of the data and associated metadata depend on the needs of the designated user community and types of datasets (e.g., maps, remotely sensed data from satellites and airborne instruments, physical samples). Follow-up parts to this standard may be developed detailing content items appropriate to individual disciplines.
- 19165-2 addresses these for Earth observation data

ISO 19165-2

- **Title - *Geographic information -- Preservation of digital data and metadata -- Part 2: Content specifications for Earth observation data and derived digital products***
- **Summary of Scope**
 - Provide details about content describing the provenance and context specific to data from missions that observe the Earth using spaceborne, airborne or in situ instruments.
- **Includes what needs to be preserved and when (i.e., starting at what stage of an activity)**

Project Team

Member Body	Expert	e-mail address
Australia (SA)	Margie Smith	margie.smith@ga.gov.au
	Bob Wilson	bob.wilson@bom.gov.au
	Kate Roberts	Kate.roberts@bom.gov.au
Canada (SCC)	C. Douglas O'Brien	douglasobrien@idontech.ca
Denmark (DS)	Jan Hjelmager	jnh@sdfе.dk
France (AFNOR)	Clément Drouadaine	Clement.Drouadaine@ign.fr
Japan (JISC)	Koki Iwao	iwao.koki@aist.go.jp
South Africa (SABS)	Serena Coetzee	serenacoetzee@gmail.com
South Africa (SABS)	Thandi Msibe	MsibeTP@eskom.co.za
Spain (UNE)	Antonio RODRIGUEZ	afrodriguez@fomento.es
United Kingdom (BSI)	James Reid	james.reid@ed.ac.uk
United States (ANSI)	Hampapuram Ramapriyan	Hampapuram.Ramapriyan@ssaihq.com

ISO 19165-2: Table of Contents (1 of 2)

- **Foreword, Introduction, Scope, Normative references**
- **Terms and Definitions, Symbols and abbreviated terms, Conformance**
- **Mission Stages**
 - **Mission Concept**
 - **Mission Definition**
 - **Mission Implementation**
 - **Mission Operation**
 - **Post Mission**

ISO 19165-2: Table of Contents (2 of 2)

- **Preservation Content (one subsection for each stage; each subsection has rationale and list of recommended content items)**
 - Mission Concept
 - Mission Definition
 - Mission Implementation
 - Mission Operation
 - Post Mission
- **Annexes**
 - Abstract test suite
 - Stages and Phases
 - XML Representation for ISO 19165-2

Columns in Content Item Tables

- **ID** – a short identifier for the content item
- **“Need for”** – indicates which type of missions the specified item is needed (satellite – SAT, aircraft – AIR, field campaign – FLD, all types – ALL)
- **Type** – indicates whether the item is a document, data record, software, etc.
- **Identification** – a phrase identifying the content item (longer than ID)
- **Description** – a brief description of the content item
- **Examples of Types of Quality Information** – information to be contained in the content item for the item to be considered to be of high quality.

Sample table (Mission Concept Stage)

ID	Need for	Type	Identification	Description	Examples of Types of Quality Information
MC_1.0	ALL	Doc	Preservation metadata	List of items preserved at this stage to satisfy content requirements specified below. (Note that this item will be updated at each of the subsequent stages.)	Identification of items by title and persistent ID. Access method (e.g., link). Indication of access restrictions if any as well as whether and when the restrictions will be removed.
MC_1.1	ALL	Doc	Scientific/ Applications Scenario, data producer and User Communities	Defines scientific/applications scenario and expected goals. Lists Principal Investigator, designated user communities and third party actors.	Required uncertainty bounds for services and applications, lifetime, data availability, data accuracy, ... Names of key science team leads and product team members, ...
MC_1.2	SAT	Doc	Mission Requirements Document	Defines scientific/applications mission and sensor requirements, processing methods, qualification methods. Includes instrument specifications (e.g., frequencies, ...), data products to be produced, and operations concepts.	Calibration plan and quality assessment plan for the mission. Uncertainty requirements for instrument product (e.g. radiometric/geometric uncertainty bounds, coverage, revisit times, etc.). Justification for the design decisions (e.g. band selection).
MC_1.3	ALL	Doc	Mission Operation Plan	Defines the plan for how the mission will be conducted.	Initial operations concept.
MC_1.4	ALL	Doc	Mission Cost and Schedule	Defines planned cost and schedule for the mission.	Initial estimated costs and schedule timelines

Status

- **Initial Document Based on ESA, CEOS/WGISS and NASA documents**
- **CD approved to proceed to DIS in February 2019**
 - 47 comments were received from 5 countries
- **Comments have been addressed in revised version**
 - Editing Committee met on June 3, 2019
 - Editing is complete and expect to submit DIS within a few weeks