Response to action on AI, Survey outcome

Presented to CGMS-54 Working Group II session, agenda item 5

Coordination Group for Meteorological Satellites

CGMS

Background

Reference to CGMS-53 plenary discussions last year and response to action item 53.07:

“On AI/ML: CGMS Secretariat to survey CGMS contributing agencies. WGII and WGIV to propose initial standards to be presented at CGMS-54”

Coordination Group for Meteorological Satellites

CGMS

Executive summary of the WP

- Strong community interest in making EO data and services more AI-ready, but **maturity remains uneven across agencies**: Few agencies are already advancing with roadmaps, pilot services, Spatio-Temporal Asset Catalogues (STAC) adoption, cloud-native formats, and AI-oriented access approaches, while others remain in planning or early exploration phases

- Agencies consistently highlight that users need an **usable access for an efficient implementation into ML workflows** (efficient data filtering and extraction, chunking, machine-readable metadata, provenance, and practical tooling) remain central expectations

- There is a **demand for a shared AI/ML catalogue for EO application**, mainly to avoid duplication, accelerate uptake, improve benchmarking, and strengthen collaboration

- **Challenges** reported by users remain concentrated around data access speed, preprocessing burden, interoperability across sources, standards maturity, and platform or usage constraints

- There is a **clear interest in a community coordination on best "AI-ready" EO practices under a CGMS –WMO framework**, with respondents indicating that joint guidance, shared practices, and lightweight coordination mechanisms would be useful next steps

Coordination Group for Meteorological Satellites

CGMS

Key themes

I. AI-Ready EO Data Standards:
   - Metadata and catalogues
   - Data formats and cloud-native access
   - APIs, tools and user workflows
   - Coordination and next steps

II. AI/ML Applications for the Satellite Earth Observation Community:
   - Current use of AI/ML in EO
   - Need for a Centralised AI/ML Applications Catalogue
   - Governance and Sustainability

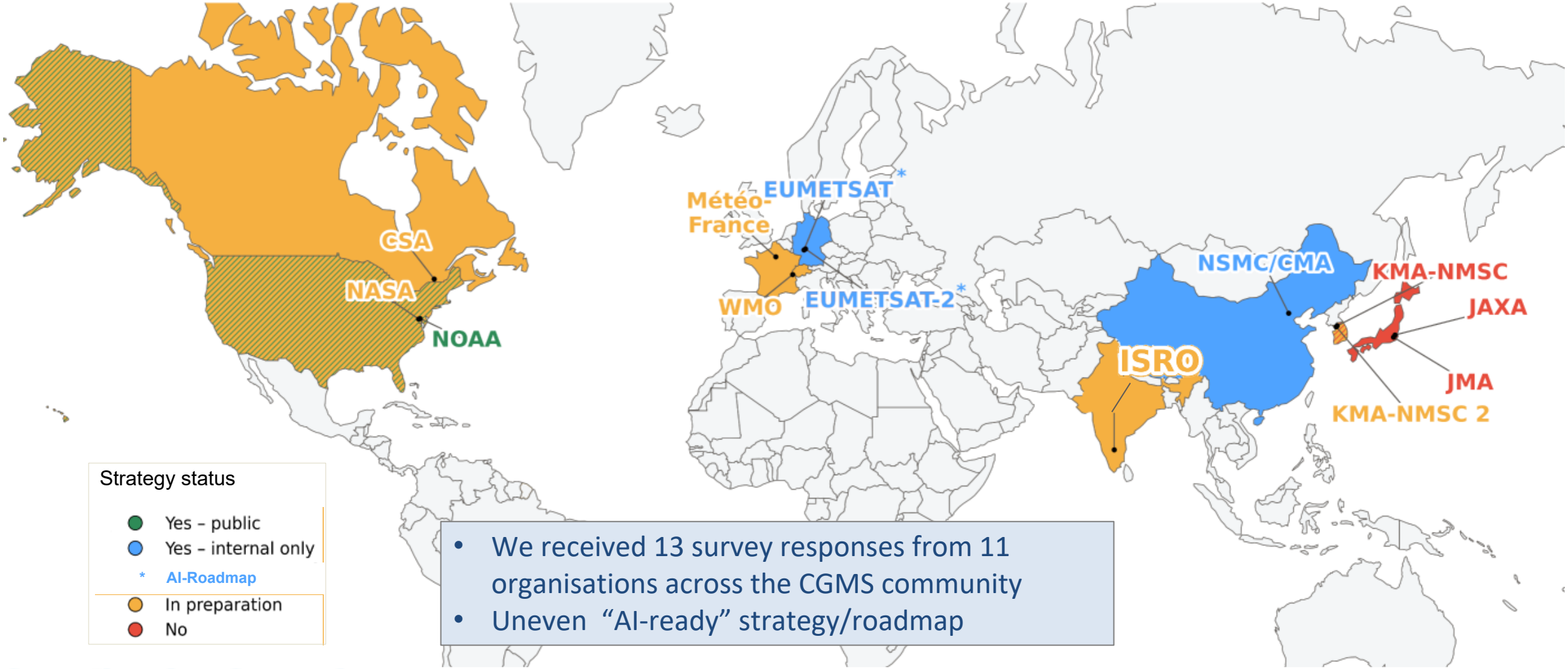
Coordination Group for Meteorological Satellites

CGMS

Introduction

- Earth observation data are currently managed differently across agencies. Organisations use different catalogues, metadata standards, data formats, and access methods to organise and provide their data.
- To support the broader use of AI, these data need to become more AI-ready. This requires a greater consistency across the community in cataloguing standards, metadata, access patterns, and interoperable formats.
- Data formats and machine learning operations (MLOps) are central to this transition. Formats determine how easily data can be integrated into ML workflows, while MLOps provides the operational processes and tools needed to support deployment, monitoring, maintenance, and reuse over time.

CGMS – WMO Survey on AI-Readiness



- We received 13 survey responses from 11 organisations across the CGMS community
- Uneven "AI-ready" strategy/roadmap

Coordination Group for Meteorological Satellites

CGMS

Definitions of Catalogues mentioned in the AI- Survey

- **Catalogue of AI-Ready EO datasets:** A structured, machine-readable inventory of Earth Observation datasets, accompanied by standardized metadata that describes elements such as contents, format, quality, uncertainty, clear provenance tracing as well as target applications, and make them directly usable as inputs for AI/ML workflows without requiring manual, custom preprocessing by end users.

- **AI/ML Applications Catalogue:** A curated registry of AI/ML models, algorithms, and processing workflows developed for EO applications, providing information on e.g. inputs, outputs, training data, performance metrics, and licensing to allow other agencies and users to discover, evaluate, reproduce, and potentially reuse or adapt these applications.

Link between catalogues: Note that the input data information used in the AI/ML Applications Catalogue should be shared as well within the Catalogue of AI-Ready datasets

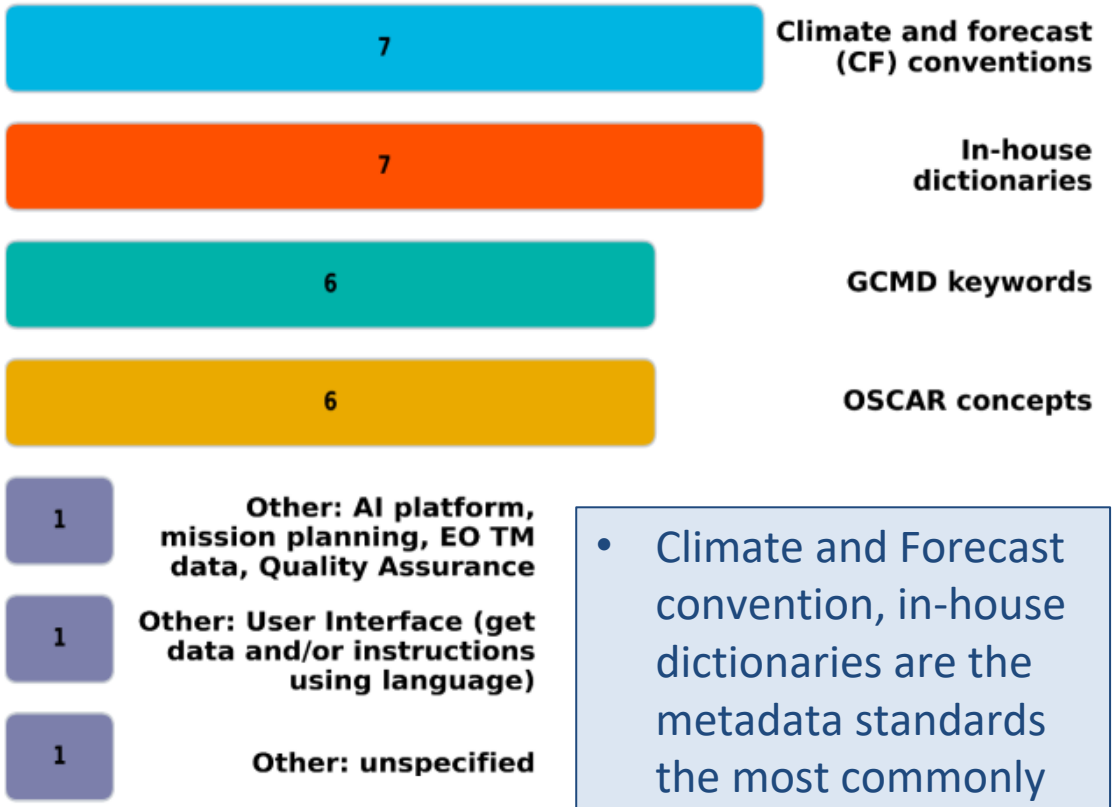
I. AI-Ready EO Data Standards - Metadata & Catalogues

Current use of data catalogue standard

Operationally 2	Pilot/prototype 3
Planned 3	Not planned 5

- Use of a common data catalogue standard (e.g. Spatio Temporal Asset Catalogue STAC) adoption remains uneven across the community: Only a limited number of agencies report operational use, while many are still in pilot, prototype, or planning phases.

Metadata standards currently used



- Climate and Forecast convention, in-house dictionaries are the metadata standards the most commonly used

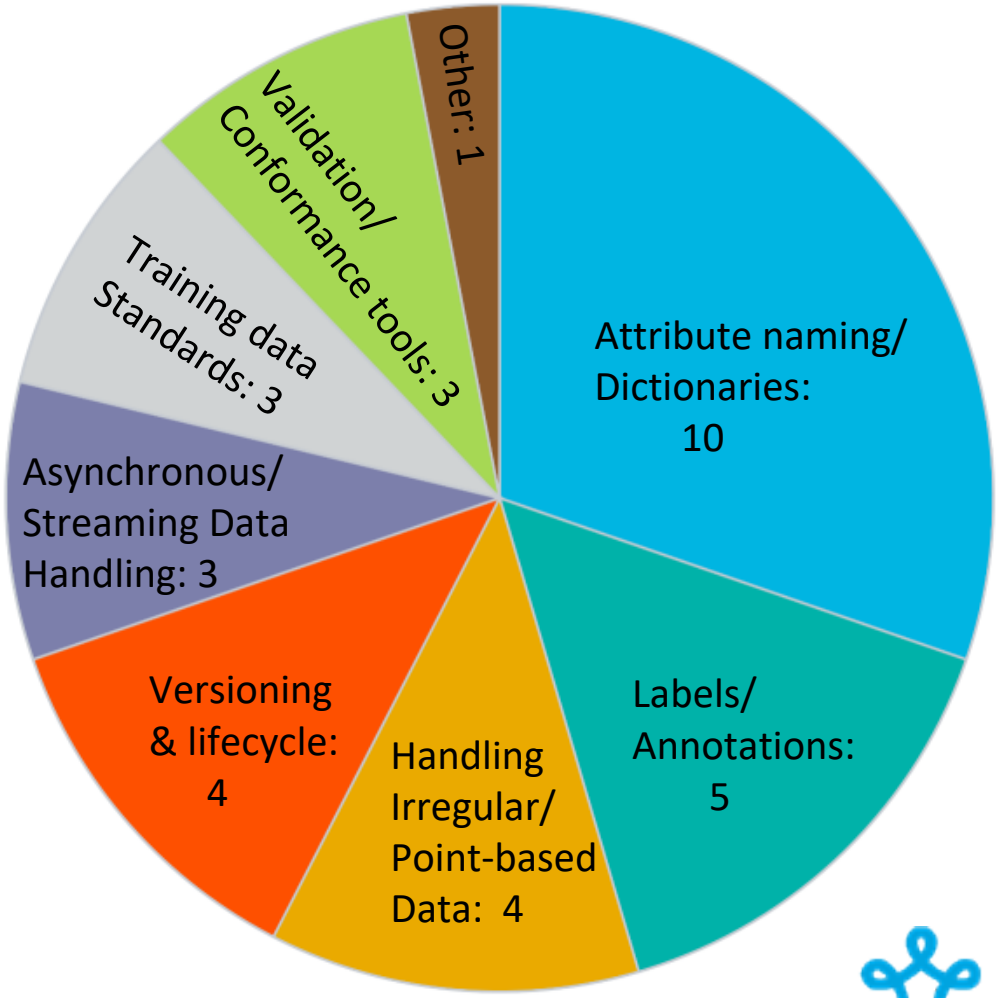
Coordination Group for Meteorological Satellites

CGMS

I. AI-Ready EO Data Standards - Metadata & Catalogues

- Catalogue and metadata standards that respondents would most like harmonised across agencies:

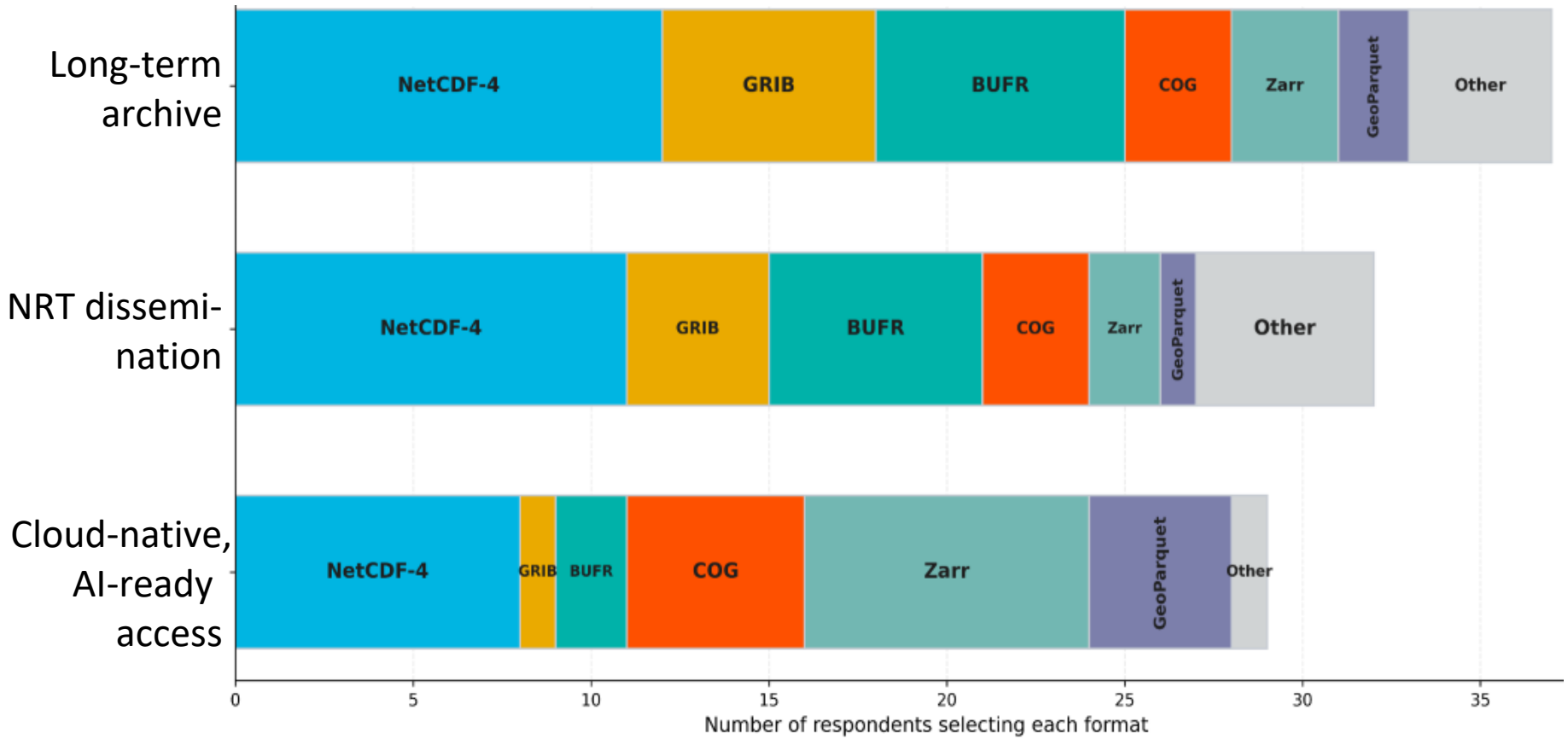
  1) Attribute naming/dictionaries: for using same names and definitions (for variables, parameters)
  2) Labels/annotations: additional information to describe, classify and interpret the data
  3) Versioning & lifecycle of collections/items to keep track of updates, changes and replacement
  4) Handling Irregular/ Point-based data: when data are not organised in regular grids



Coordination Group for Meteorological Satellites

CGMS

I. AI-Ready EO Data Standards - Data formats and cloud-native access



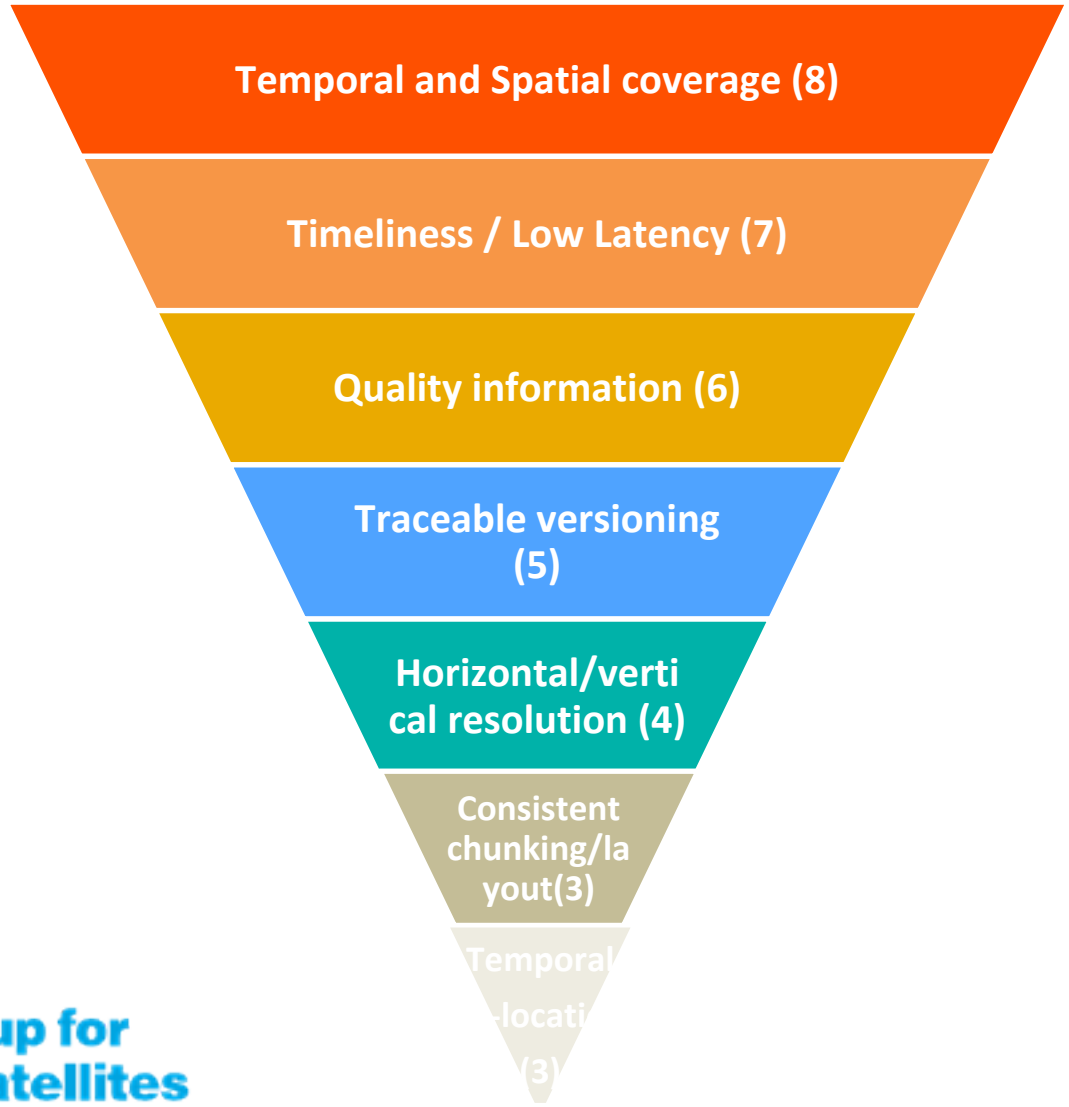
Adopted Formats by context

- Context matters:
-NetCDF-4, GRIB and BUFR are most commonly used for long-term archive and NRT dissemination.
-NetCDF-4 and Zarr are used for Cloud-native/AI-ready access
- Zarr adoption is emerging but not yet broadly used (7 respondents do not provide Zarr)

Coordination Group for Meteorological Satellites

CGMS

I. AI-Ready EO Data Standards - Data formats and cloud-native access



Critical data characteristics for Users in AI-workflows

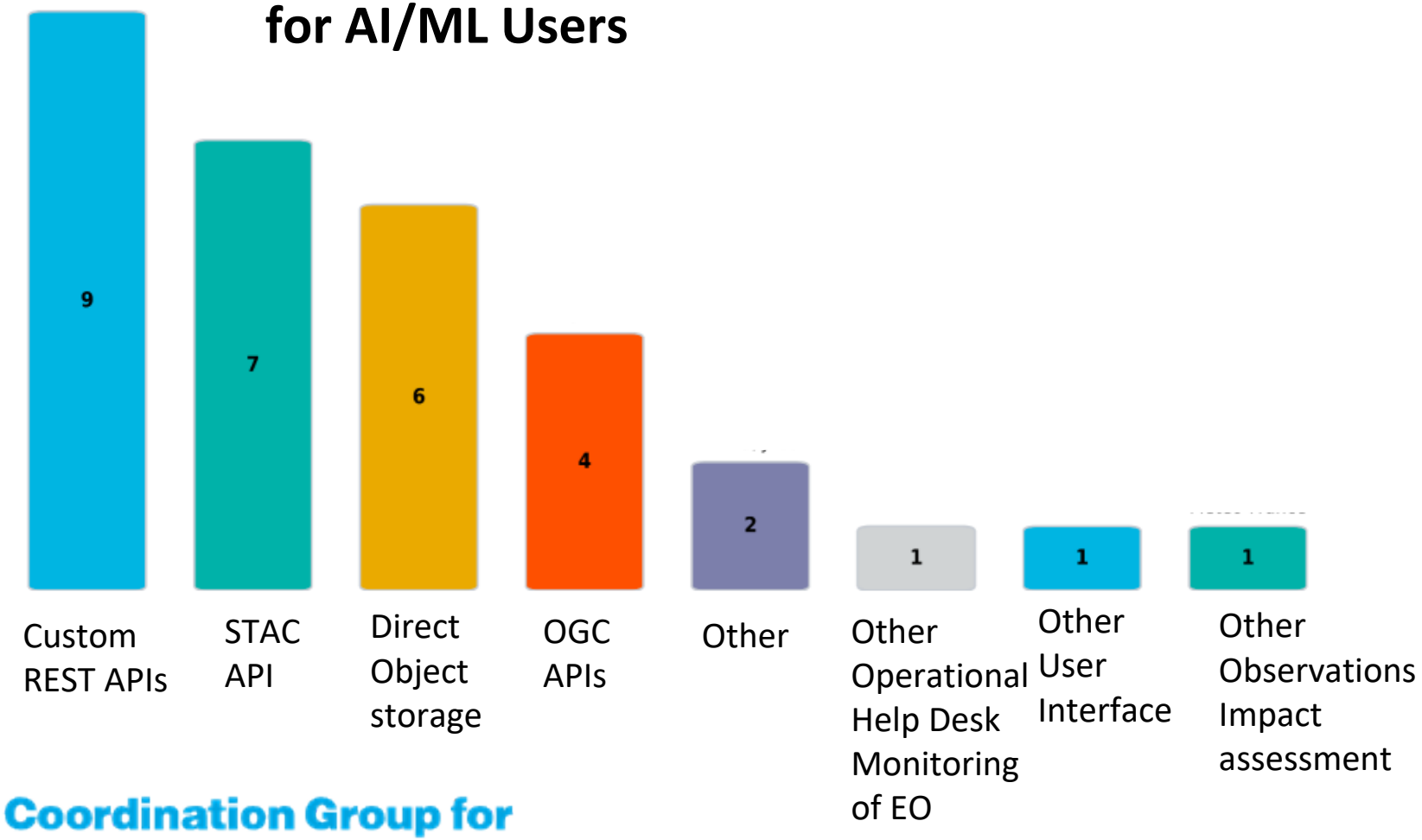
- Temporal & spatial coverage consistency as well as Timeliness/low latency and quality information are the most critical to users for AI-workflows

Coordination Group for Meteorological Satellites

CGMS

I. AI-Ready EO Data Standards - APIs, tools & users workflows

Access or planned interfaces for AI/ML Users

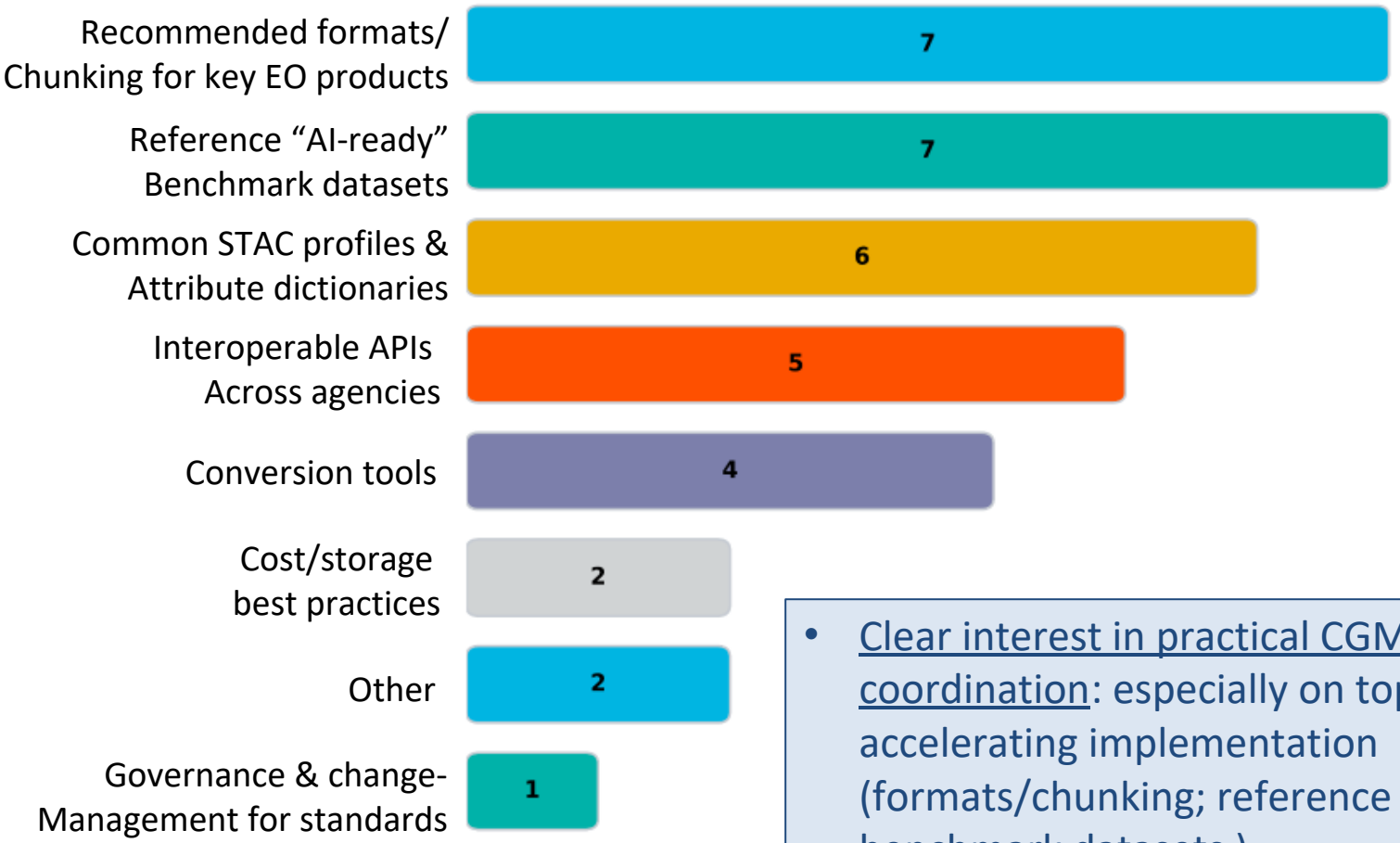


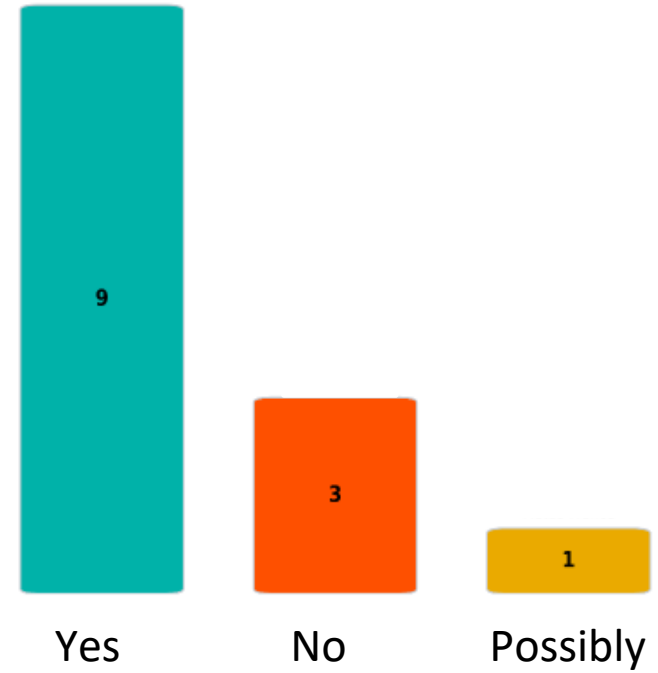
- Agencies are moving toward more AI-oriented access services, but implementation is still uneven across the community (Use of Custom REST APIs, STAC API, Direct object storage.., or other interface)
- For most respondents: agencies do not yet provide tools to help users re-chunk or reformat data for AI, but planned to do so
- Main pain points reported by users when using data for AI/ML:
  - Data access/transfer speed
  - Data format/ preprocessing burden
  - Standards catalogue immaturity

Coordination Group for Meteorological Satellites

CGMS

I. AI-Ready EO Data Standards - Coordination & next steps





Participation in task team on "AI-ready EO practices"

- Clear interest in practical CGMS-WMO coordination: especially on topics for accelerating implementation (formats/chunking; reference "AI-ready" benchmark datasets)
- 9 positive answers to participate in a small task team on "AI-ready EO practices"

Coordination Group for Meteorological Satellites

CGMS

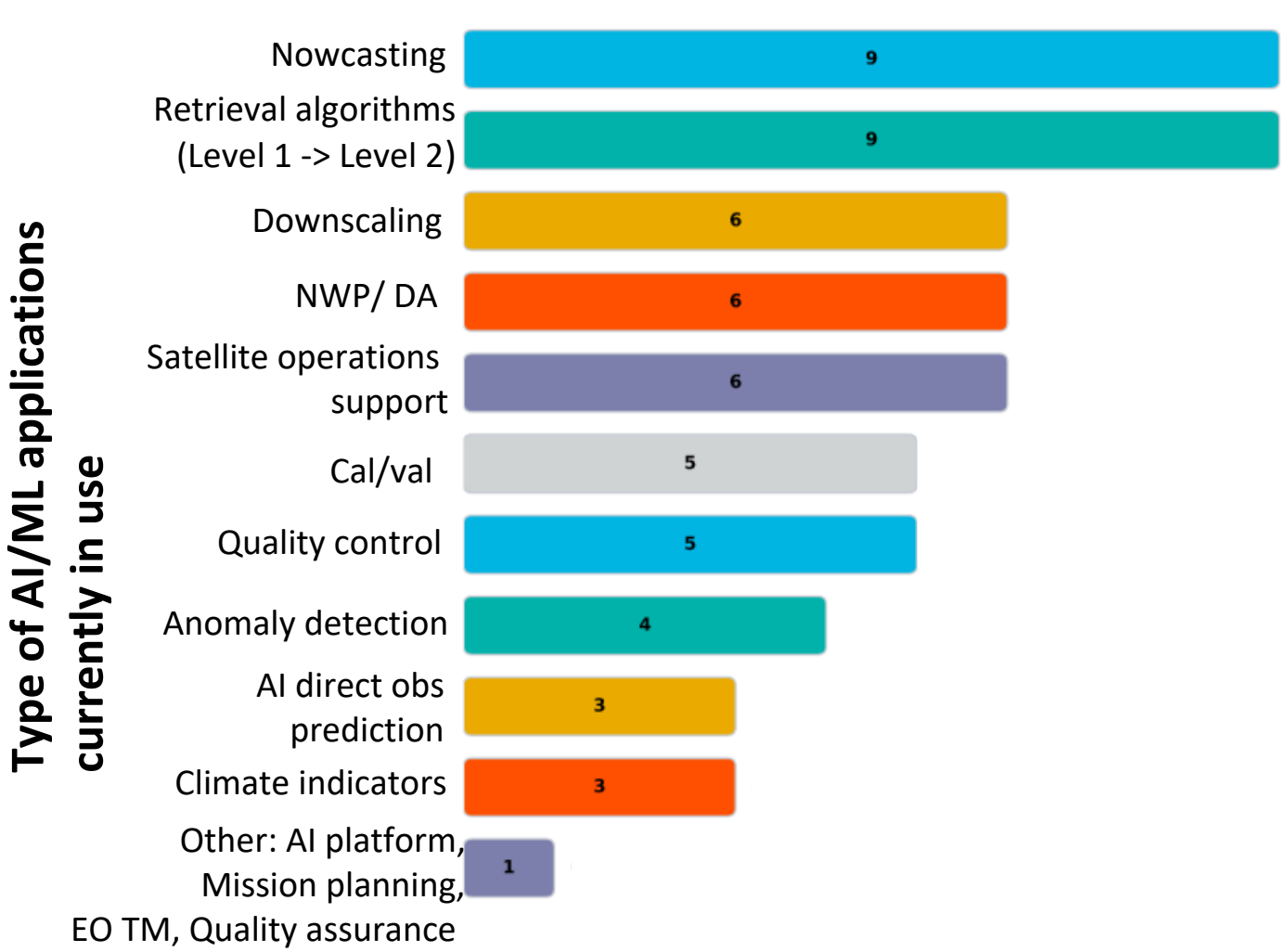
Concluding remarks Part I.

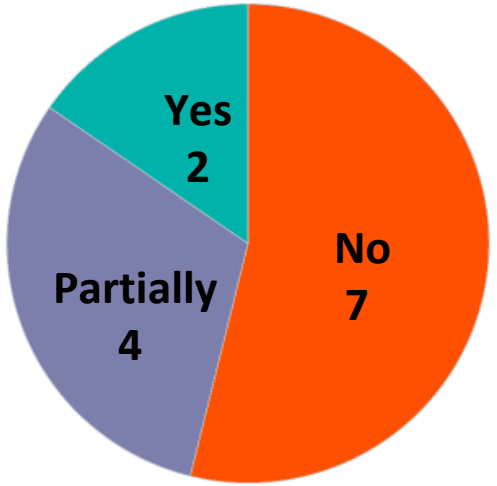
- Survey responses indicate a clear community interest in making Earth observation data more AI-ready, although implementation maturity remains uneven across organisations
- The strongest common needs relate to metadata and catalogue standards, improved discoverability, more usable access patterns, and data structures that reduce preprocessing effort for implementation into machine learning workflows
- The results suggest that agencies are moving in a common direction, but have not yet converged on shared practices for cataloguing, standards, formats, and AI-oriented data access
- These findings set the context for the second part, which focuses on resources for AI/ML applications, including code, scripts, tools, platforms, workflows, guidance, and standards, and how these could be better shared and reused through a common reference resource.

Coordination Group for Meteorological Satellites

CGMS

II. AI/ML Applications for the Satellite EO Community – Current use of AI/ML in EO





Internal AI-catalogue for EO application availability

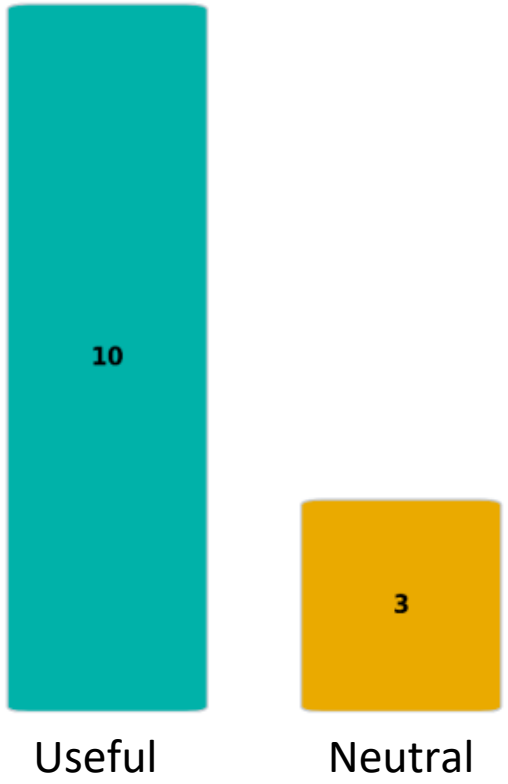
- Broad spread of AI/ML application types (mostly for nowcasting or retrieval algorithms (L1 -> L2)
- Internal Cataloguing of AI/ML application is uneven: 7 respondents do not have an internal AI-catalogue. 4 partially hold one. 2 possess one.

Coordination Group for Meteorological Satellites

CGMS

II. AI/ML Applications for the Satellite EO Community – Need for a Centralised AI/ML Applications Catalogue

Shared catalogue of AI/ML applications



- Shared AI/ML Applications Catalogue definition:
  Common reference resource supporting the effective use of AI/ML by CGMS contributing agencies and meteorological data users. The catalogue shall provide guidance and standards for preparing, using, and converting meteorological EO data into suitable form for machine learning applications.

- Main benefits of using a shared catalogue: FAIR principles (Findability, Accessibility, Interoperability and Reusability)
- A shared AI/ML catalogue is seen as useful by a clear majority (10 out of 13 respondents)

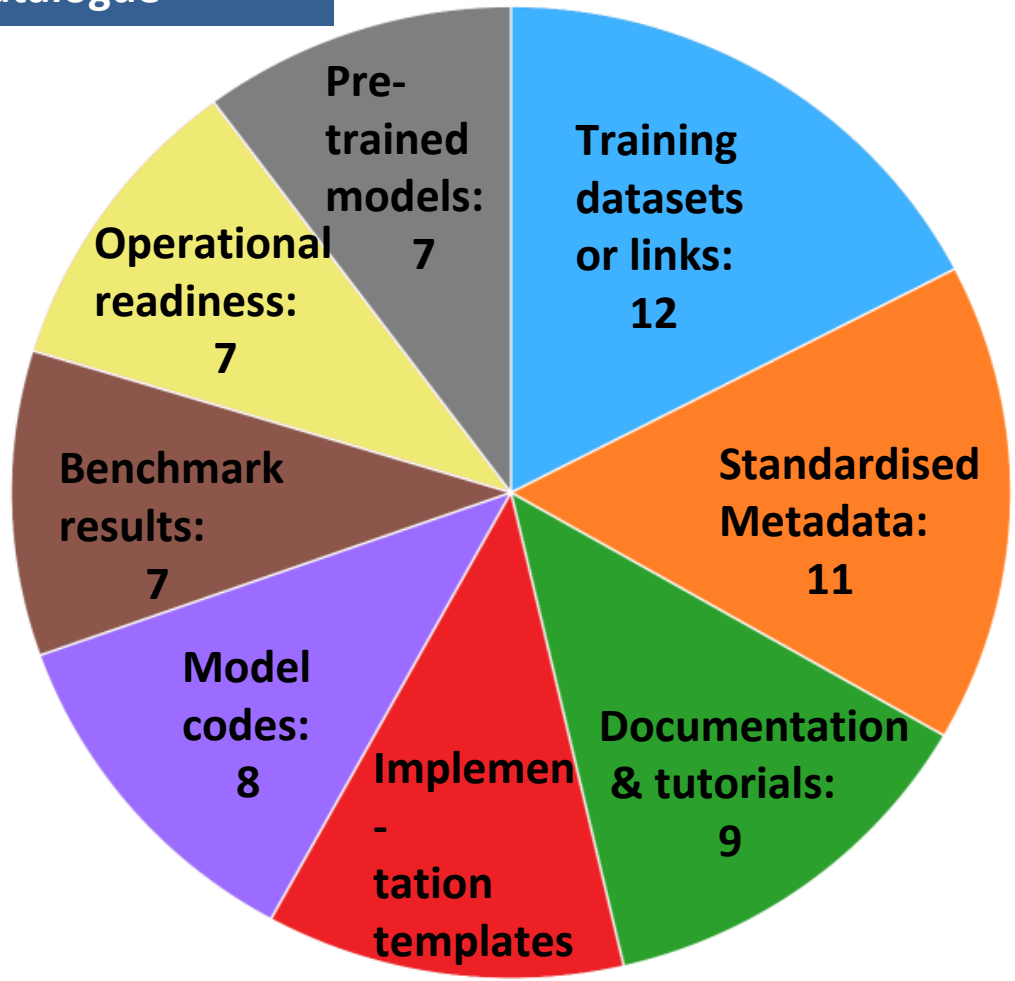
Coordination Group for Meteorological Satellites

CGMS

II. AI/ML Applications for the Satellite EO Community – Need for a Centralised AI/ML Applications Catalogue

- Respondents would like the catalogue to be practical and implementation-oriented (training datasets or links; standardised metadata).
- Strong demand for an adoption of support materials (documentation and tutorials)

=> Agencies wish for a shared AI/ML Applications catalogue that supports **collaboration, reuse, and implementation.**



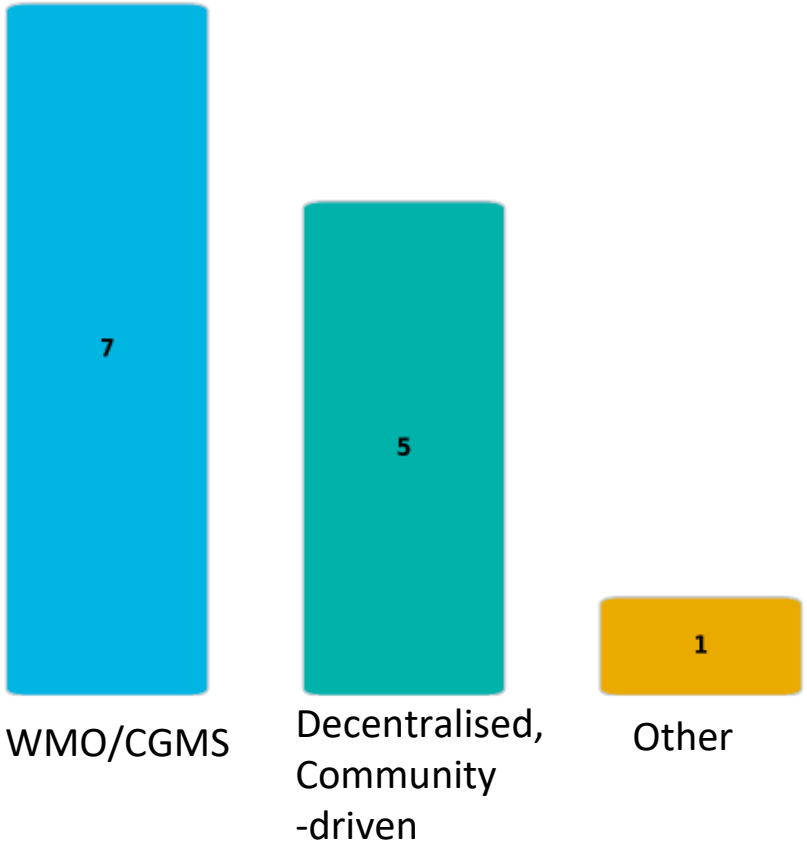
Catalogue elements

Coordination Group for Meteorological Satellites

CGMS

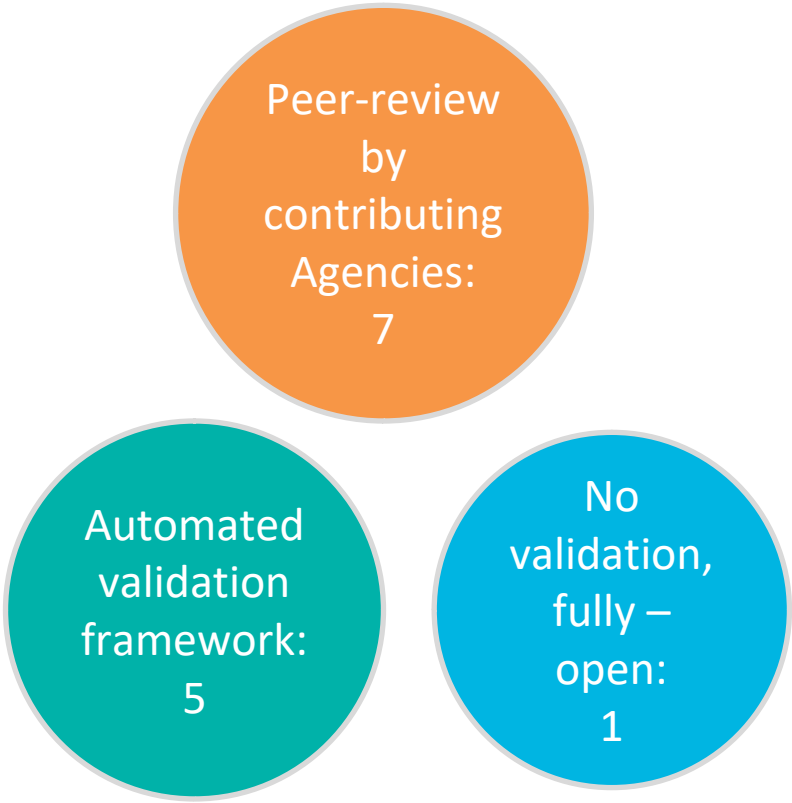
II. AI/ML Applications for the Satellite EO Community – Governance & Sustainability

Catalogue maintain by:



- Preference for a shared governance model with suggestion that WMO/CGMS should host and maintain the catalogue
- Respondents favour a controlled contribution model, with peer-review by contributing agencies preferred
- For actual contribution: agencies are broadly positive, but depends on capacity, mandate, or readiness

Contributions from:



Coordination Group for Meteorological Satellites

CGMS

Concluding remarks Part II.

- A shared AI/ML Applications Catalogue is important for CGMS because, as AI/ML moves from research into operations, it can help avoid the development of siloed modes within agencies. This Catalogue would ensure to make existing applications and expertise more visible, reduce duplication of effort across the CGMS community and enhance coordination across agencies.
- The AI/ML community already uses repositories and platform-based cataloguing functions through tools such as GitHub, Hugging Face, and other MLOps environments. However, these mainly indicate where code or models are hosted. A CGMS AI/ML Applications Catalogue would add context, such as data and sensors used, the intended application, demonstrated performance, and the conditions under which the application can be used reliably in operations.
- For CGMS, the question is therefore whether a shared metadata catalogue is needed: not to replace existing technical platforms, but to describe AI/ML applications in a common, discoverable, and reusable way across agencies.
- Such a catalogue would require a clear investment and governance model: the community would need to decide who defines the metadata framework, who implements and maintains the catalogue, and how contributions are reviewed over time.

Coordination Group for Meteorological Satellites

CGMS

CGMS – WMO Survey on AI-Readiness

Conclusions

- The survey shows a community that is ready to **move from experimentation toward coordinated implementation of AI-ready EO data practices**
- Across agencies, the strongest common needs are: **better discoverability, more consistent metadata, usable cloud-native access, and practical support for ML workflows**
- There is **clear interest in a shared AI/ML application catalogue**, but only if the catalogue becomes an operational resource, not just a static inventory. It should provide: structured, quality-controlled, and supported by documentation, standards, code, and examples
- Agencies favour a shared AI/ML application catalogue managed under a joint CGMS–WMO governance model, with peer-reviewed contributions to ensure quality, consistency, and community relevance before broader dissemination.

Coordination Group for Meteorological Satellites

CGMS

Discussion points:

- CGMS members to take note of the survey result
- The CGMS Secretariat recommends establishing annual CGMS–WMO workshops on AI-readiness for EO data, starting in 2027 and potentially coordinated with relevant WMO AI groups. These workshops could define priority topics for community action, including for example a discussion and adoption of a minimal set of common metadata attributes, building on existing work by organisations such as NASA, NOAA, OGC, and ESA. As a follow-up, CGMS could develop a small pilot activity around the priorities identified during the workshops, moving beyond coordination with existing initiatives towards practical implementation.
- The CGMS Secretariat recommends to promote practical standards and good practices to improve interoperability across agencies and help reduce discrepancies between the community in AI-ready maturity
- **<u>Action</u>: CGMS Head of Delegation and Agencies are invited to discuss the above points and provide guidance on the proposed way forward.**

Coordination Group for Meteorological Satellites

CGMS