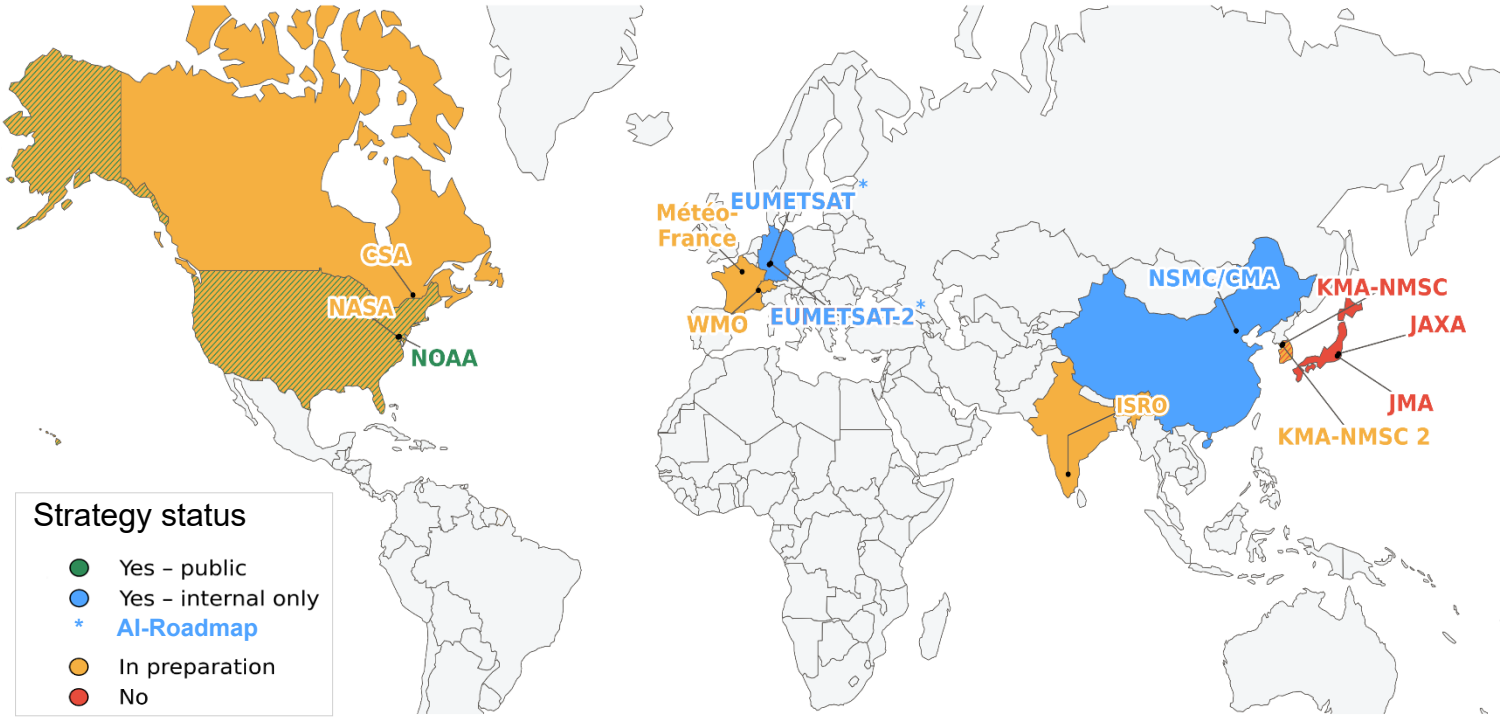
Annex of Response to action on AI, Survey outcome

Presented to CGMS-54 Working Group session

Prepared by CS Office (EUMETSAT)

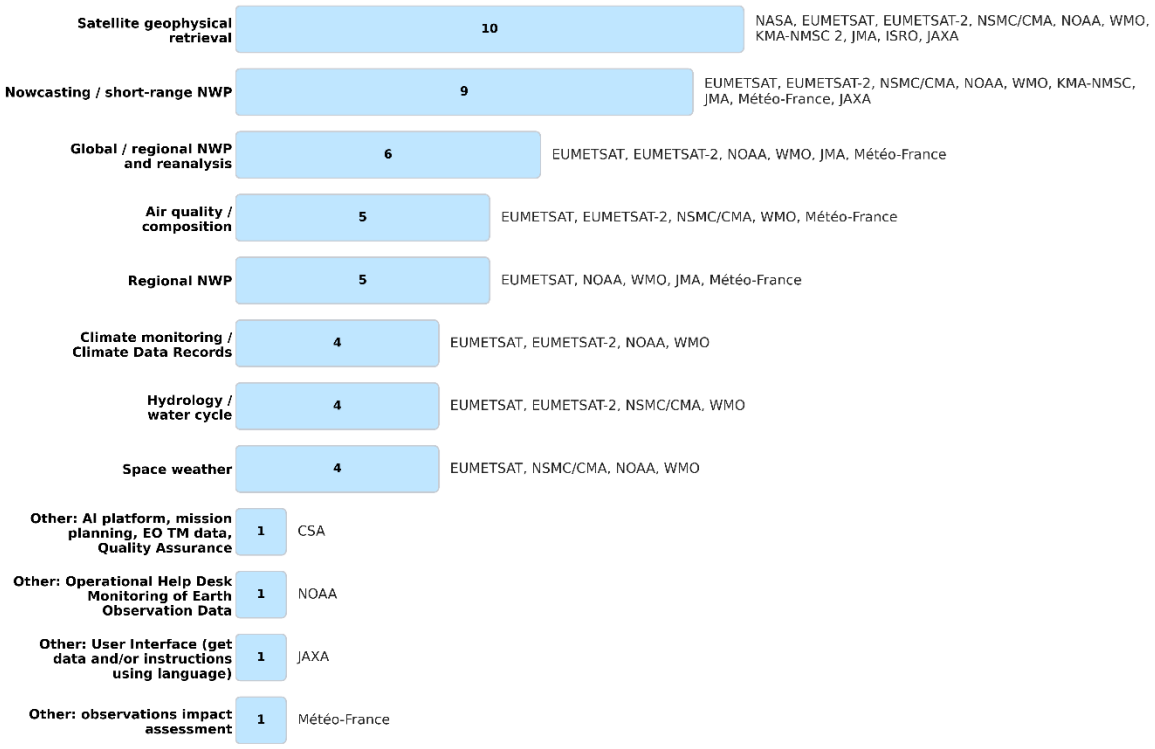
Part A

Q1.1 Do you have an explicit strategy/roadmap for “AI-ready” Earth observation data (metadata, formats, APIs)?



13 respondents across 11 agencies with uneven “AI-ready” strategy/roadmap

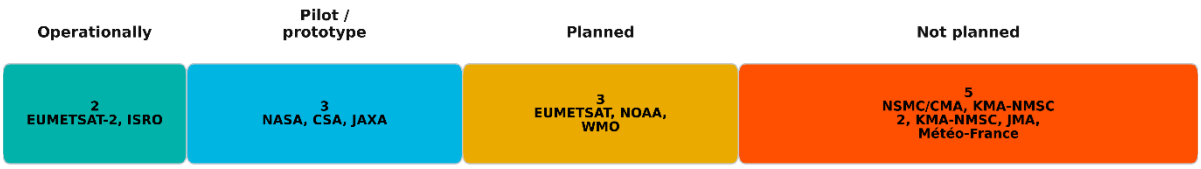
Q1.2 Which main AI/ML use cases are you targeting with your Earth observation data?



AI/ML use cases targets:

- Satellite geophysical retrieval
- Nowcasting/Short-range NWP
- Technical: AI platform, QA, User Interface, Operational Help Desk Monitoring for EO

Q2.1 Do you currently use SpatioTemporal Asset Catalogues (STAC) for Earth observation data discovery?



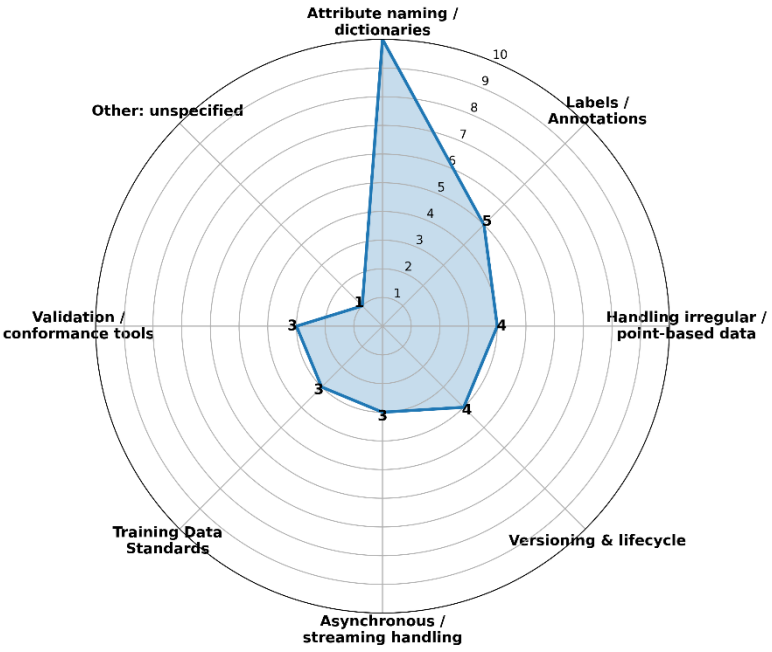
Spatio Temporal Asset Catalogues (STAC): Catalog standard for organizing, describing and easily discovering geospatial datasets to be used consistently across platforms and tools

STAC adoption is progressing: 5 respondents are currently operationally using STAC for EO discovery or at a pilot stage. 3 are planning using STAC and 5 are not planning it

Q2.2 Which semantic metadata standards do you use?

Semantic metadata commonly used: Climate and Forecast conventions as well as in-house dictionaries

Q2.3 Which aspects of STAC/metadata standardisation would you most like harmonised across agencies?



Attribute naming / dictionaries

NASA, EUMETSAT, EUMETSAT-2, CSA, NSMC/CMA, WMO, KMA-NMSC, ISRO, Météo-France, JAXA

Labels / Annotations

EUMETSAT, NOAA, KMA-NMSC, ISRO, JAXA

Handling irregular / point-based data

CSA, NSMC/CMA, WMO, Météo-France

Versioning & lifecycle

NASA, CSA, ISRO, Météo-France

Asynchronous / streaming handling

EUMETSAT-2, NOAA, KMA-NMSC

Training Data Standards

NSMC/CMA, NOAA, KMA-NMSC 2

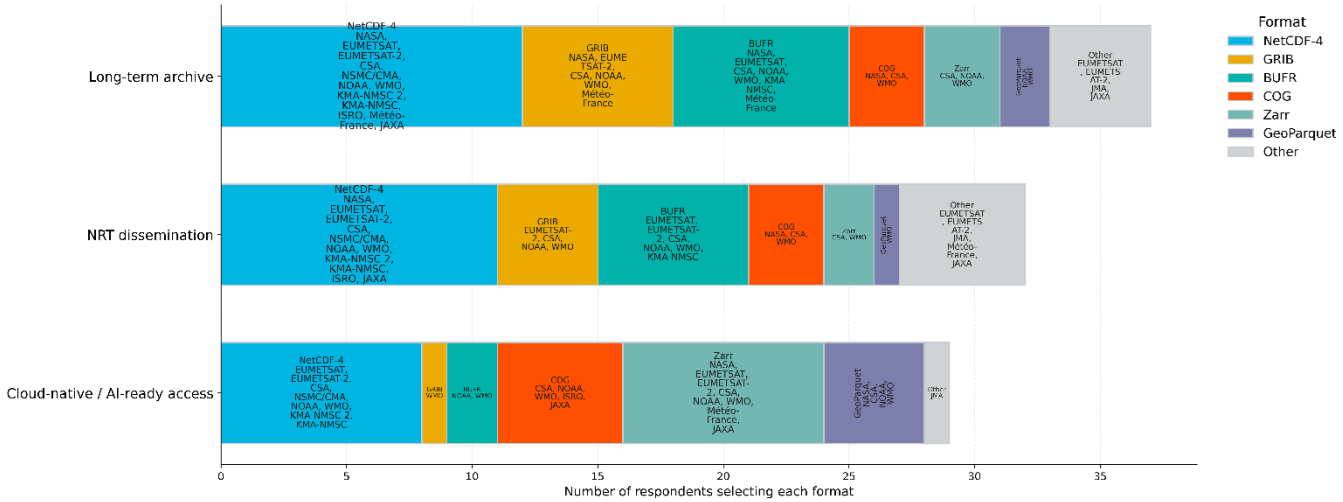
Validation / conformance tools

NASA, EUMETSAT-2, WMO

Other: unspecified

JMA

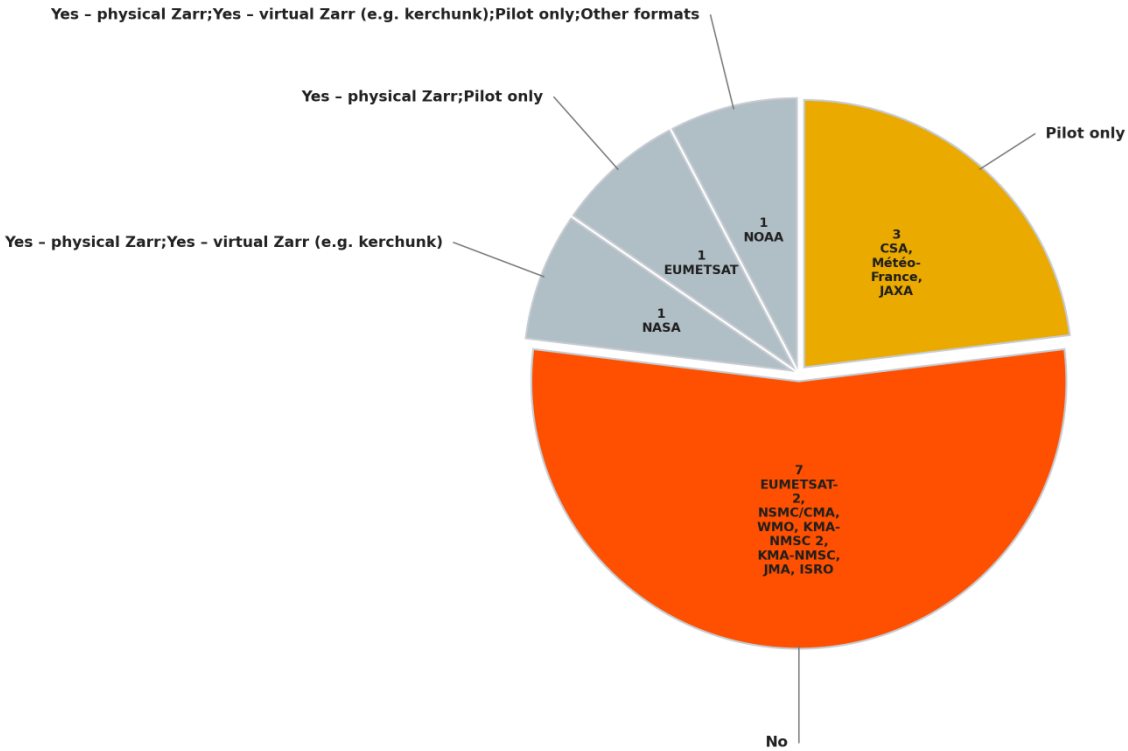
Q3.1 What formats do you use in each context?



-NetCDF-4, GRIB and BUFR are most commonly used for long-term archive and NRT dissemination.

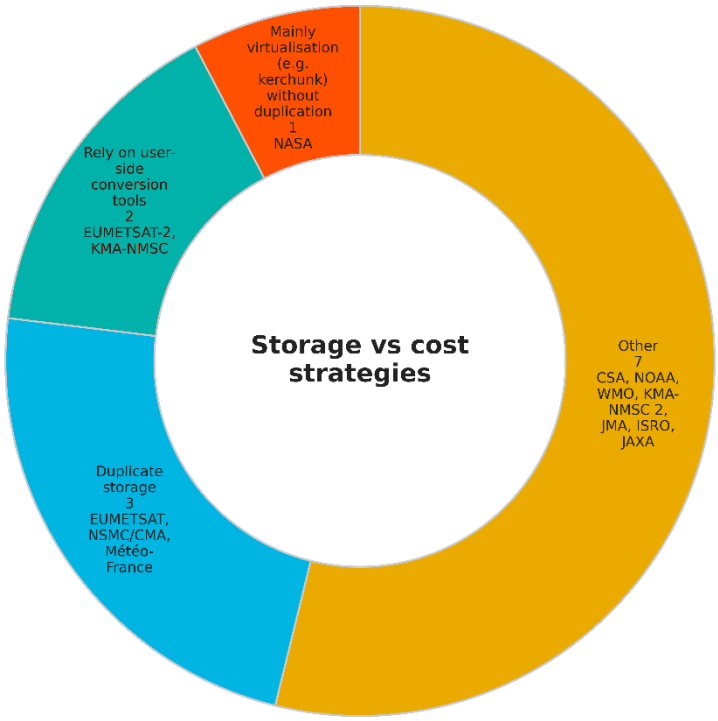
-NetCDF-4 and Zarr are used for Cloud-native/AI-ready access

Q3.2 Do you already provide Zarr datasets?



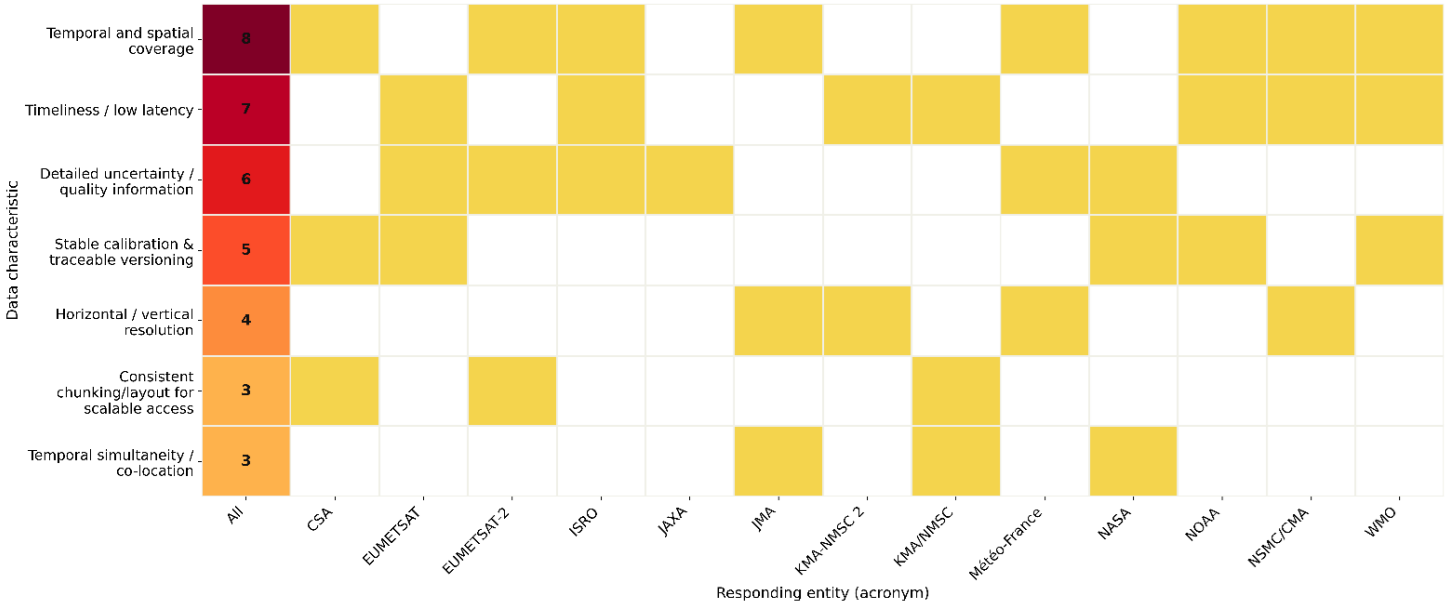
Zarr adoption is emerging but not yet broadly used (7 respondents do not provide Zarr)

Q3.3 How do you manage storage vs cost when introducing cloud-native formats?



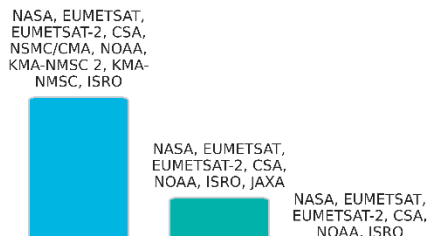
Storage vs cost: No single model strategy is emerging, agencies use rather diverse strategies (see other. Ex: Kerchunk virtualisation for deep archive)

Q3.4 For AI workflows, which data characteristics are most critical to your users?



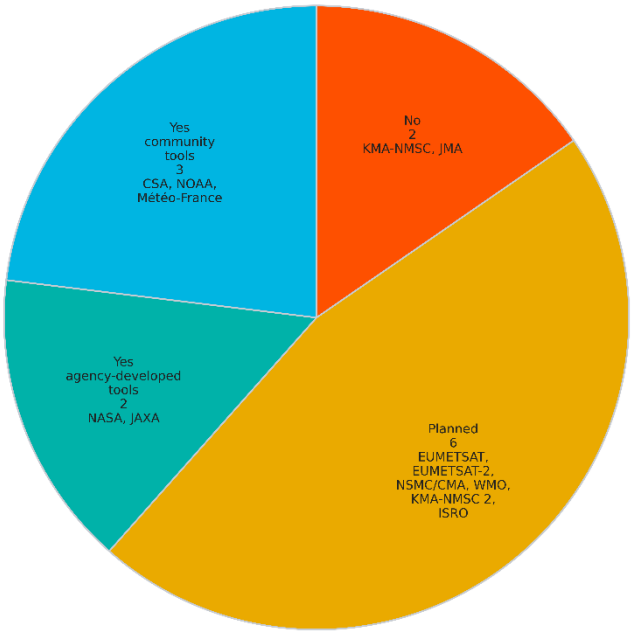
Temporal & spatial coverage as well as Timeliness/low latency and quality information are the most critical to users for AI-workflows

Q4.1 What access interfaces do you provide / plan 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)

Q4.2 Do you provide tools to help users re-chunk or reformat data for AI?



For most respondents: agencies do not yet provide tools to help users re-chunk or reformat data for AI, but planned to do so

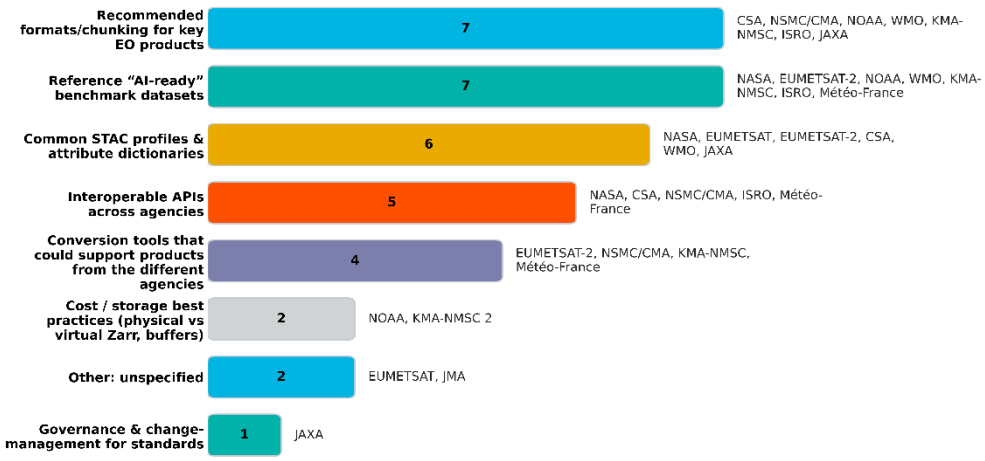
Q4.3 What are the main pain points that your users report when using your data for AI/ML?

Access speed / latency	4	0	1	0	0	1	1	0	0	1	0	0	0	0
Preprocessing / data preparation	2	0	1	1	0	0	0	0	0	0	0	0	0	0
Formats / Zarr conversion	4	0	0	1	1	0	1	0	0	0	0	0	1	0
Temporal-spatial alignment / gaps	2	0	1	0	0	1	0	0	0	0	0	0	0	0
Cloud-native access / chunking / STAC	2	0	0	0	1	0	1	0	0	0	0	0	0	0
Cost / egress / storage constraints	2	0	0	0	1	0	1	0	0	0	0	0	0	0

Main pain points reported by users when using data for AI/ML:

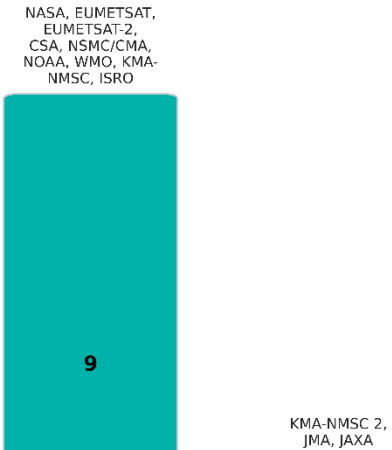
- Data access/transfer speed
- Data format/ preprocessing burden

Q5.1 On which topics would you most welcome CGMS–WMO coordination/common guidance? (select up to 3)



Clear interest in practical CGMS-WMO coordination: especially on topics for accelerating implementation (formats/chunking; reference "AI-ready" benchmark datasets)

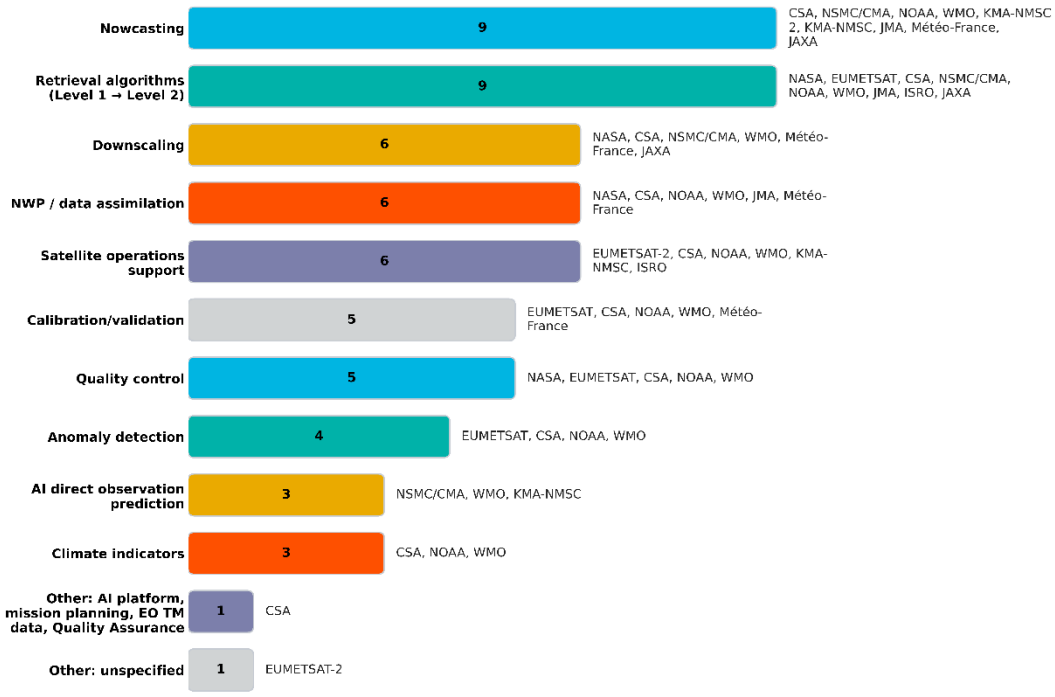
Q5.2 Would your agency participate in a small task team or workshop on "AI-ready EO practices" (STAC, Zarr, APIs, tools)?



9 positive answers to participate in a small task team on ”AI-ready EO practices”

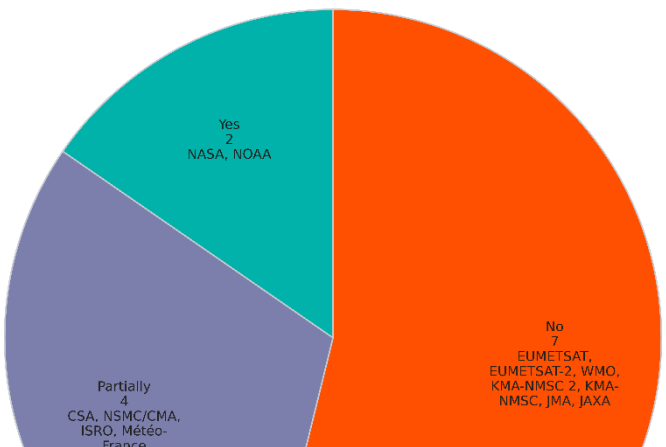
Part B

Q1.1 Which types of AI/ML applications do you currently develop or use?



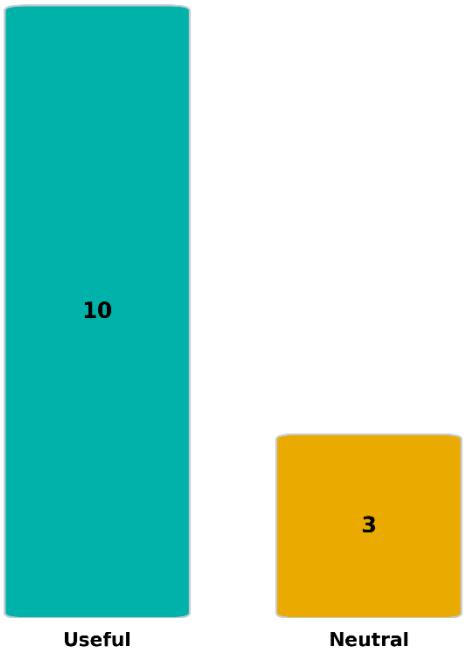
Broad spread of AI/ML application types (mostly for nowcasting or retrieval algorithms (L1 -> L2)

Q1.2 Do you have an internal catalogue or repository of AI applications?



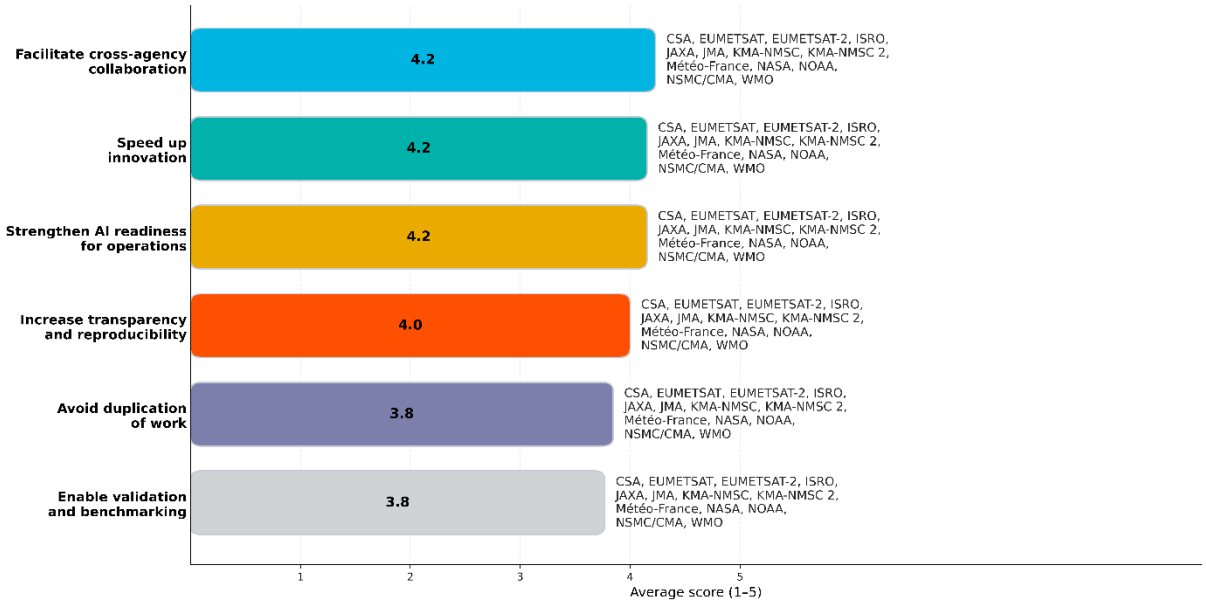
Internal Cataloguing of AI/ML application is uneven: 7 respondents do not have an internal AI-catalogue. 4 partially hold one. 2 possess one.

Q2.1 Would a shared catalogue of AI/ML applications for EO be useful to your organisation?



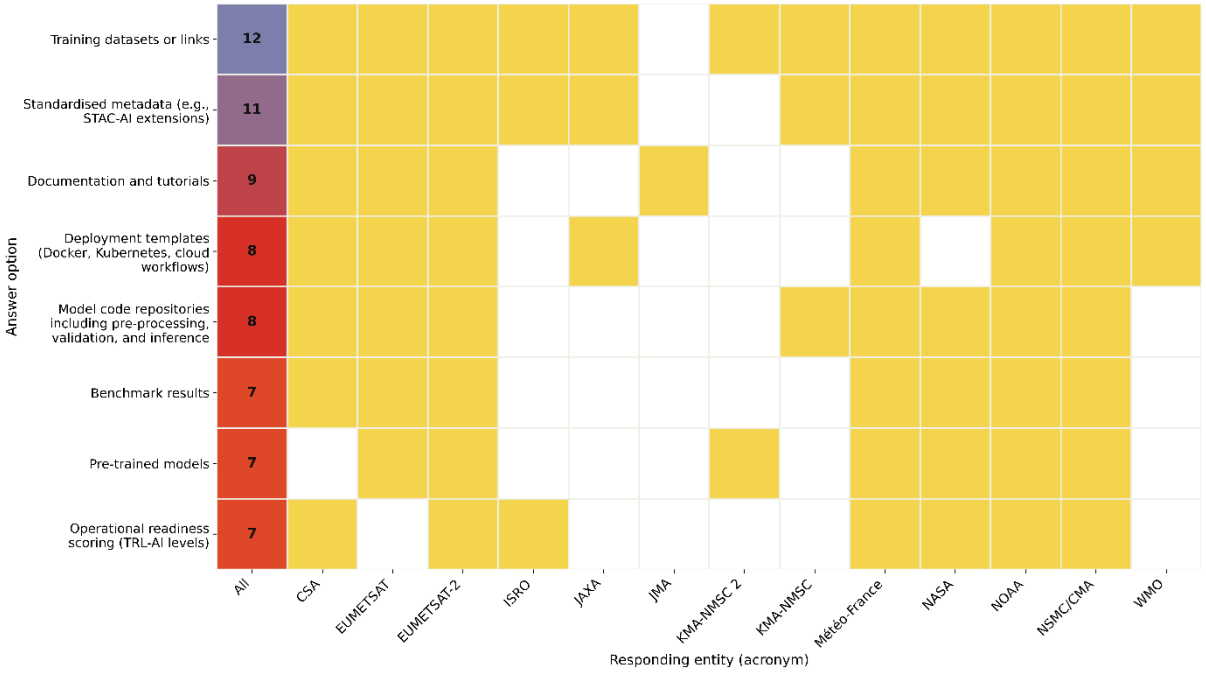
A shared AI/ML catalogue is seen as useful by a clear majority

Q2.2 Perceived benefits of a shared catalogue



Main benefits to expect for: cross-agency collaboration, followed closely by speeding up innovation and strengthening AI readiness for operations

Q2.3 What elements should the catalogue contain?

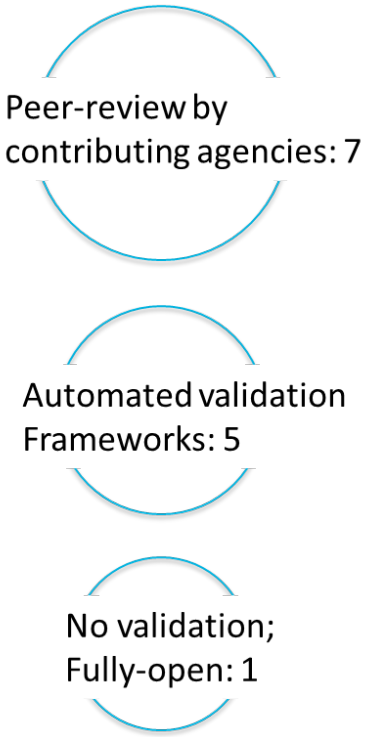


Respondents would like the catalogue to be practical and implementation-oriented (training datasets or links; standardised metadata). Strong demand for adoption support materials (documentation and tutorials)

Q3.1 Who should host and maintain the catalogue?

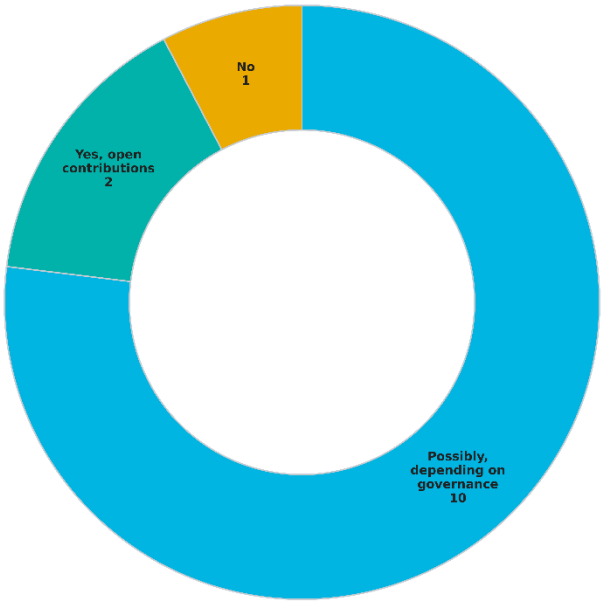
Preference for a shared governance model with suggestion that WMO/CGMS should host and maintain the catalogue

Q3.2 Who should validate contributions?



For validation, respondents favour a controlled contribution model, with peer-review by contributing agencies preferred over fully open submission without oversight

Q3.3 Would your organisation contribute models or tools?



For actual contribution: agencies are broadly positive, but depends on capacity, mandate, or readiness