

Towards an open registry of Earth observation instruments

David Montero¹, César Aybar², Miguel D. Mahecha^{1,3}, Luis Gómez-Chova²

¹ Institute for Earth System Science and Remote Sensing, Leipzig University,
04103 Leipzig, Germany - (david.montero, miguel.mahecha)@uni-leipzig.de

² Image Processing Laboratory (IPL), Universitat de València,
46980 Paterna (València), Spain - (cesar.aybar, luis.gomez-chova)@uv.es

³ Helmholtz Centre for Environmental Research (UFZ), 04318 Leipzig, Germany

Keywords: Earth Observation, Remote Sensing Instruments, Instrument Metadata, Open Registry, Open Science

Abstract

Earth observation (EO) is essential to understanding the Earth system, enabling the transformation of planetary properties into measurable variables that can be analysed, compared, and modelled. In recent decades, EO capabilities have grown rapidly, accompanied by an even faster expansion in the number and variety of available EO instruments. Today, EO includes instruments deployed on satellites, airborne platforms, and terrestrial or in-situ systems. However, despite this proliferation of instruments, users often lack a single, reliable source describing their existence and key characteristics. Although existing data catalogues have substantially improved dataset discovery, they primarily describe data products rather than providing persistent, curated metadata about the instruments that produced them. Here we present Awesome Earth Observation Instruments, an open, standardized, and community-oriented registry providing machine-readable metadata for EO instruments. The catalogue is hosted on GitHub and allows contributors to submit instrument metadata following a common schema. The schema combines a lightweight core with modular extensions covering spectral, geometric, and data access-related metadata, enabling both standardization and flexibility across diverse EO systems. All submissions undergo automated schema validation and human review. Because the schema is open, versioned, and extensible, the catalogue can continuously evolve as new instruments and metadata requirements emerge. This facilitates the discovery, interpretation, and analysis of EO data in light of instrument characteristics. We anticipate that the catalogue will serve as a community reference for EO instrument metadata while lowering barriers to discovery and integration. To support programmatic access and interoperability, we further envisage an API for integration within common EO analysis environments. The catalogue is openly available at <https://github.com/awesome-spectral-indices/awesome-earth-observation-instruments>.

1. Introduction

Earth observation (EO) through remote sensing is fundamental for understanding the Earth system because it provides repeated, spatially explicit measurements that transform planetary properties into variables that can be analysed, compared, and modelled (Running et al., 1994; Justice et al., 1998). As the demand for such observations has grown, EO capabilities have expanded rapidly over recent decades, accompanied by a substantial increase in the number and diversity of remote sensing instruments available to the community (Wulder et al., 2012). Contemporary EO includes instruments onboard satellite platforms, airborne sensors, and terrestrial or in-situ systems. Passive optical instruments measure reflected radiation using multispectral (Wulder et al., 2016; Drusch et al., 2012) or hyperspectral systems (Clark et al., 2024). These measurements commonly cover the ultraviolet, visible, near-infrared, and short-wave infrared regions, with some instruments extending into the thermal infrared. Active systems, such as synthetic aperture radar, provide illumination-independent observations and enable monitoring under clouds and variable atmospheric conditions (Torres et al., 2012). Airborne and uncrewed aerial vehicle (UAV) instruments can provide centimetre-scale spatial detail and flexible acquisition timing, while ground-based sensors offer direct in-situ reference measurements (Alvarez-Vanhard et al., 2021).

This diversity of platforms and sensing modalities creates a practical challenge: users often lack a single reliable source for identifying which instruments exist and for accessing their key

characteristics, including governance, mission status, spectral configuration, imaging geometry, spatio-temporal resolution, and data access points. Existing resources only partially address this need (Bugbee et al., 2021). Some instrument catalogues, such as the sensor information provided by the Index DataBase (Henrich et al., 2009), include useful entries but are limited in openness, completeness, and update frequency. The Committee on Earth Observation Satellites (CEOS) Database¹ provides one of the most comprehensive and standardised overviews of EO missions and instruments, but it does not systematically expose the level of instrument-specific detail required for many analytical workflows, such as spectral characteristics or data-access patterns. In parallel, data catalogues, including large data platforms and SpatioTemporal Asset Catalog (STAC)² catalogues, have substantially improved dataset discovery. However, they are primarily designed to describe data-product collections and associated assets, rather than to provide persistent, curated descriptions of the instruments that generated those data products.

This paper presents the first iteration of Awesome Earth Observation Instruments, an open, standardised, and community-oriented registry of EO instruments that provides machine-readable instrument metadata. The registry is designed to complement existing data catalogues and to support reproducible, automated geospatial workflows. It is implemented as a living catalogue hosted on GitHub, where contributors can add instruments and associated metadata according to a shared spe-

¹ <https://database.eohandbook.com>

² <https://stacspec.org>

cification. The specification is designed to be straightforward to read and implement while remaining strict where standardisation is essential. Contributions are reviewed by maintainers and validated against a YAML-based JSON Schema. The schema consists of a lightweight core and a set of modular extensions, enabling the catalogue to represent diverse EO systems while remaining extensible. In version 0.1.0, the registry includes 16 catalogue entries covering seven instrument families from widely used satellite EO systems: the Landsat archive (MSS, TM, ETM+, OLI, and TIRS), the Sentinel-2 MultiSpectral Instrument (MSI), and the hyperspectral EMIT instrument on board the International Space Station (ISS).

The remainder of the paper is organised as follows. Section 2 presents the Awesome Earth Observation Instruments catalogue, including the core schema, extensions, repository structure, validation and catalogue-generation workflow, and release and distribution strategy. Section 3 presents a showcase in which an initial set of EO instruments was validated and released in the first catalogue version. Section 4 discusses the potential integration of the catalogue with existing resources, its relevance for artificial intelligence (AI) workflows, possible future extensions, programmatic access, and current limitations. Section 5 concludes the paper.

2. The Awesome Earth Observation Instruments Catalogue

The Awesome Earth Observation Instruments Catalogue (AEOI) is a curated, machine-readable, and community-driven registry of EO instruments. It is designed to describe instruments deployed on diverse platform types, including terrestrial systems, airborne and UAV platforms, and spaceborne missions. The catalogue is hosted on GitHub and distributed in JavaScript Object Notation (JSON) format through GitHub and Zenodo. Each registry item corresponds to an EO instrument and is contributed through GitHub according to a predefined schema. To balance human readability with machine interoperability, all input files, including the schema and its extensions, are written in YAML (Sections 2.1 and 2.2).

2.1 Schema

The registry schema defines how each instrument item must be organised and is implemented using JSON Schema draft 2020-12³. It is provided as a core YAML file that can be extended through references to additional schemas (Section 2.2). The core schema captures general instrument metadata, including identifying attributes such as name, acronym, type, and platform, among others (Table 1). These attributes are required for ingestion into the catalogue. Each instrument must also provide a unique identifier through the `id` attribute. We recommend combining the instrument acronym and platform identifier, for example `MSI_S2A` for the MultiSpectral Instrument on-board Sentinel-2A. The same identifier must be used as the filename of the corresponding YAML item.

2.2 Extensions

EO instruments differ substantially in their mission objectives, platform constraints, sensing principles, and available metadata. Including all possible properties in the core schema would make the catalogue unnecessarily complex and difficult

to maintain. Instead, following the general logic of the SpatioTemporal Asset Catalog (STAC) extension mechanism, AEOI uses a compact core schema complemented by optional extensions. The core schema includes an `extensions` property that can reference additional schemas. When an extension is used for a given instrument, the corresponding properties are validated according to the referenced schema.

The initial implementation includes four extensions: `spectral`, describing spectral characteristics such as bands and spectral response functions; `imaging`, describing optical and geometric properties; `ee`, describing Google Earth Engine (GEE) data access; and `planetary_computer`, describing Microsoft Planetary Computer data access.

2.2.1 Spectral: The spectral extension specifies the spectral characteristics of an instrument (Table 2). Although optional, it represents one of the most important metadata components for many EO applications and is therefore strongly recommended whenever the required information is available. The extension remains optional to accommodate instruments for which spectral characteristics are unavailable, variable, or not applicable, such as some RGB cameras, integrated UAV cameras, and flexible systems such as spectroradiometers. The extension can be added later if reliable spectral information becomes available.

The extension covers two main metadata components: spectral bands and spectral response functions (SRFs). For multispectral instruments, spectral bands can be defined either as an inline dictionary of band properties (the `bands` property in Table 2) or as a CSV file containing the same information in tabular form. For hyperspectral instruments, bands can alternatively be generated from a continuous spectral range (the `range` property in Table 2). For a given instrument, only one of these two alternatives must be provided. SRFs are represented as CSV files for simplicity and interoperability. In these files, the first column contains wavelength values and subsequent columns correspond to the identifiers of previously defined spectral bands; cell values represent relative response at each wavelength. Although CSV filenames are explicitly declared in the schema and therefore need not follow a fixed convention, we recommend using the instrument identifier followed by the suffix `_SRF` or `_BANDS`, depending on the file content (e.g. `ETM_L7_BANDS.csv` for the bands file of the Landsat 7 ETM+ instrument).

2.2.2 Imaging: The imaging extension specifies optical and geometric instrument characteristics (Table 3). It is optional and can be added whenever the relevant information is available; all properties in the extension are also optional. The extension includes properties commonly used for satellite instruments, such as `swath_width`, `across_fov`, and `along_fov`, as well as properties relevant to airborne and UAV systems, such as `hfov` and `vfov`. Unlike the spectral extension, the imaging extension does not require auxiliary files and is populated directly through numerical values.

2.2.3 Earth Engine: Google Earth Engine (GEE) is a geospatial platform provided by Google that combines cloud-computing resources with a large catalogue of raster and vector geospatial datasets, including many EO products (Gorelick et al., 2017). To support open data discovery and open-science workflows, the GEE extension provides the first data-access extension in AEOI (Table 4). Although the core schema includes a generic `data_links` property, specialised data-access extensions provide standardised entry points to datasets produced by a given instrument. The extension defines a required

³ <https://json-schema.org/draft/2020-12>

Property	Required	Type	Description
id	✓	string	Unique alphanumeric identifier of the instrument.
name	✓	string	Full instrument name.
acronym	✓	string	Short instrument acronym.
type	✓	string	Sensing modality; one of: <code>multispectral</code> , <code>hyperspectral</code> , <code>radar</code> , <code>rgb</code> , <code>lidar</code> , <code>other</code> .
platform_type	✓	string	Platform class; one of: <code>satellite</code> , <code>airborne</code> , <code>uav</code> , <code>terrestrial</code> .
platform	✓	array	List of platform names (e.g., satellite mission, UAV model, aircraft).
operator	✓	array	List of operating organizations; predefined (e.g., ESA, NASA, JAXA, NOAA, DLR, ISRO, Copernicus) or custom strings for non-predefined organizations.
start_date	✓	string	Start of operation (YYYY-MM-DD).
status	✓	string	Lifecycle state; one of: <code>operational</code> , <code>retired</code> , <code>experimental</code> , <code>planned</code> .
availability	✓	string	Data access level; one of: <code>public</code> or <code>private</code> .
end_date		string	End of operation (YYYY-MM-DD), if applicable.
notes		string	Free-form additional information.
references	✓	array	List of reference URLs (papers, documentation, web pages).
data_links		array	URLs providing access to data products.
extensions		object	Optional structured extensions: <code>spectral</code> (band/SRF info; see Table 2), <code>imaging</code> (FOV, IFOV, optics; see Table 3), <code>ee</code> (Google Earth Engine dataset metadata; see Table 4), <code>planetary_computer</code> (Planetary Computer dataset metadata; see Table 5).

Table 1. Schema definition for instrument metadata.

Property	Type	Description
bands	oneOf	Either (i) CSV filename, or (ii) dictionary of bands keyed by ID (e.g., B1). Each band requires <code>center_wavelength</code> (nm) and <code>bandwidth</code> (nm), and may include <code>band_description</code> , <code>common_name</code> (e.g., <code>blue</code> , <code>red</code> , <code>nir</code> , <code>swir</code>), <code>gsd</code> (m), and <code>snr</code> .
range	object	Continuous spectral coverage for sensors without discrete bands: <code>min</code> and <code>max</code> wavelength (nm, required), plus either <code>sampling</code> (nm) or <code>total_bands</code> .
spectral_response_function	string	Filename of the spectral response function (SRF) CSV file.

Table 2. Metadata fields for the Spectral Extension (either `bands` or `range` must be provided).

primary data-access point for the instrument or platform, for which we recommend bottom-of-atmosphere (BOA) data when available. It also supports optional raw, BOA, and top-of-atmosphere (TOA) data-access points. Each property is an object containing a link to GEE documentation through `docs` and the corresponding collection identifier in the GEE catalogue. This structure enables users and software tools to retrieve data-access entry points programmatically in a consistent manner.

2.2.4 Planetary Computer: Microsoft Planetary Computer also provides access to an extensive EO data catalogue, but organises its collections through STAC. Because it is a widely used platform for EO data discovery and retrieval, we implemented a dedicated Planetary Computer extension following the same general data-access pattern used for GEE (Table 5). The extension adds a required `stac_endpoint` property, fixed to the Planetary Computer STAC API endpoint. The remaining properties are equivalent to those in the GEE extension: documentation links should point to the corresponding Planetary Computer dataset documentation, and collection values should match the appropriate STAC collection identifiers.

2.3 Repository Structure

The repository is organised to facilitate community contributions to the catalogue (Figure 1). The root repository is divided into four main folders: `catalogue`, `docs`, `schema`, and `src`.

- **catalogue:** This folder contains the generated catalogue

JSON file created after audit and validation. Its content is generated automatically.

- **docs:** This folder contains files used to generate the README. It can be extended in the future to support website deployment.
- **schema:** This folder contains the catalogue core schema and its extensions, stored in the `core` and `extensions` subfolders, respectively. The `core` folder contains only the core YAML schema (Section 2.1). The core schema can be modified by maintainers or through community pull requests after maintainer review, for example when adding new core properties or registering new extensions. The `extensions` folder contains the YAML schemas used to extend the core schema. New extension schemas can be contributed through pull requests and must be referenced from the core schema.
- **src:** This folder contains all input files required to generate the catalogue and is organised into four subfolders:
 - **instruments:** This folder contains the YAML files corresponding to each instrument added to the catalogue. These files must comply with the core schema and with any extensions they use. Users can contribute new instruments through pull requests.
 - **bands:** This folder contains CSV files describing spectral bands associated with specific instruments.

Property	Type	Description
swath_width	number	Swath width (km).
across_fov	number	Across-track Field of View (FOV) (degrees). Relevant for satellite platforms.
along_fov	number	Along-track Field of View (FOV) (degrees). Relevant for satellite platforms.
across_ifov	number	Across-track Instantaneous Field of View (IFOV) (microradians). Relevant for satellite platforms.
along_ifov	number	Along-track Instantaneous Field of View (IFOV) (microradians). Relevant for satellite platforms.
hfov	number	Horizontal Field of View (HFOV) (degrees). Relevant for UAV or similar platforms.
vfov	number	Vertical Field of View (VFOV) (degrees). Relevant for UAV or similar platforms.
entrance_pupil	number	Diameter of the entrance pupil (mm).
focal_length	number	Focal length (mm).
fnumber	number	F-number.
gsd	number	Ground Sampling Distance (m). If GSD varies by band, it can be defined per band in the spectral bands CSV file.

Table 3. Metadata fields for the Imaging Extension.

Property	Required	Type	Description
primary	✓	object	Primary dataset; requires docs (documentation URL) and collection (Earth Engine collection ID).
raw		object	Optional raw (unprocessed) dataset; same structure as primary (docs, collection).
boa		object	Optional bottom-of-atmosphere (BOA) dataset; same structure as primary (docs, collection).
toa		object	Optional top-of-atmosphere (TOA) dataset; same structure as primary (docs, collection).

Table 4. Metadata fields for the Earth Engine Extension.

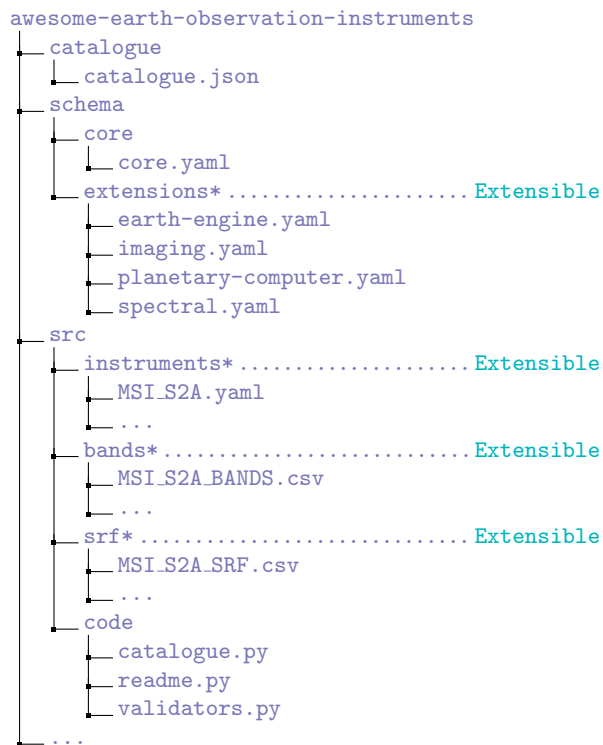


Figure 1. Repository structure. Only the folders catalogue, schema, and src are shown here. Folders marked with an asterisk indicate locations where the community can contribute new files.

- **srf**: This folder contains CSV files describing the spectral response function (SRF) associated with specific instruments. If the spectral extension is used and the `spectral_response_function` property points to a CSV file, that file must be stored in this folder.
- **code**: This folder contains the Python code used to validate catalogue items (`validators.py`), generate the catalogue (`catalogue.py`), and generate the repository README.md file (`readme.py`).

2.4 Instruments Validation and Catalogue Generation

Catalogue generation requires a complete validation process for each instrument provided as a YAML input file. The validation workflow has two phases: a human audit performed by repository maintainers and automatic validation of the input data. Instruments are included in the generated catalogue only after both phases are completed successfully.

2.4.1 Human Audit The human audit is performed by a maintainer of the GitHub repository, following a procedure similar to that used in the Awesome Spectral Indices (ASI) catalogue. During this phase, the maintainer checks whether the information provided in the instrument YAML file is supported by the sources listed in the `references` property. A revision is requested if the references do not point to authoritative or trustworthy sources, such as the organisation responsible for the instrument or its operation. A revision is also requested when the referenced sources are reliable but the submitted metadata do not match the information reported in them.

2.4.2 Automatic Validation After the human audit, the YAML file is automatically validated using a JSON Schema draft 2020-12 validator. This step verifies that the file and its property values comply with the core schema and with any declared extensions. If validation fails, the file must be corrected

If the spectral extension is used and the `bands` property points to a CSV file, that file must be stored in this folder.

Property	Required	Type	Description
stac_endpoint	✓	string	STAC API endpoint (fixed: https://planetarycomputer.microsoft.com/api/stac/v1).
primary	✓	object	Primary dataset; requires docs (documentation URL) and collection (STAC collection ID).
raw		object	Optional raw (unprocessed) dataset; same structure as primary (docs, collection).
boa		object	Optional bottom-of-atmosphere (BOA) dataset; same structure as primary (docs, collection).
toa		object	Optional top-of-atmosphere (TOA) dataset; same structure as primary (docs, collection).

Table 5. Metadata fields for the Planetary Computer Extension.

before the instrument can be included in the catalogue. Maintainers may correct errors when sufficient information is available in the submitted file and references; otherwise, a revision is requested from the contributor. If the YAML file uses the spectral extension and provides CSV files for either the `bands` or `spectral_response_function` properties, these files are also validated automatically. For `bands` files, validation fails if required columns are missing or if unsupported columns are present. For SRF files, validation fails if the wavelength column or the columns corresponding to band identifiers defined in `bands` are missing. As in the YAML validation step, maintainers may correct issues when sufficient information is available, otherwise a revision is requested.

2.4.3 Catalogue Generation After all instruments pass validation, the catalogue is generated as a single JSON file using `catalogue.py`. During generation, each YAML file is read and validated with helper functions from `validators.py`. If CSV files are provided for the `bands` or `spectral_response_function` properties, they are converted into dictionary objects and nested under the corresponding property. If bands are specified through a spectral range, band properties such as centre wavelength and bandwidth are generated automatically from the minimum wavelength, maximum wavelength, and total number of bands.

The final JSON catalogue contains four top-level properties: 1) `name`, the catalogue name, Awesome Earth Observation Instruments; 2) `version`, the catalogue version; 3) `link`, the URL of the JSON file in the GitHub repository; and 4) `instruments`, a key-value object in which each key is an instrument identifier and each value contains the validated metadata for that instrument.

2.5 Release and Distribution

The catalogue is distributed through GitHub releases, with each release version matching the catalogue version. The repository is also linked to Zenodo, enabling archival releases with a dedicated DOI for each version in addition to a general DOI for the repository.

3. Showcase: Pilot Instruments

To evaluate the catalogue implementation and the behaviour of the specified schema, we ingested 16 YAML files corresponding to seven instrument families (MSS, TM, ETM+, OLI, TIRS, MSI, and EMIT) across multiple platforms. Separate catalogue entries are needed when an instrument family has platform-specific configurations or operational histories. For example, the MSI instruments onboard Sentinel-2A, Sentinel-2B, and Sentinel-2C have platform-specific characteristics and operational dates, and are therefore represented as separate items. All submitted files complied with the specified schema.

These instruments were selected because of their relevance to existing resources such as Awesome Spectral Indices (ASI) and the Transparent Access to Cloud-Optimized datasets (TACO) specification. They also cover a broad range of characteristics useful for testing the schema, including multispectral and hyperspectral systems, retired and operational missions, and datasets available through Earth Engine or Planetary Computer. This initial set demonstrated the functionality of both the core schema and the four implemented extensions: spectral, imaging, Earth Engine, and Planetary Computer.

Given the importance of the spectral extension, we ingested CSV files describing the band specifications for all multispectral instruments and, when available, corresponding SRF CSV files. For EMIT, the hyperspectral test instrument, band information was generated from a spectral range defined by minimum wavelength, maximum wavelength, and number of bands. In both cases, ingestion was successful and the catalogue was correctly validated and generated.

All pilot instruments were successfully catalogued in AEOI and are summarised in Figure 2. To further demonstrate the functionality of the generated catalogue, we used the stored spectral metadata to plot the available SRFs in the optical and thermal regions (Figures 3 and 4, respectively). These examples illustrate that the catalogue can store and expose instrument-level spectral information in a form that is directly useful for EO workflows and analyses. This initial set corresponds to version v0.1.0 of the catalogue; later versions may include additional instruments.

4. Discussion

This section discusses the potential integration of AEOI with existing open resources, its relevance for AI workflows, possible future extensions, programmatic access, and the limitations inherent to maintaining a free, open, and community-driven catalogue.

4.1 Potential Integration with Existing Resources

4.1.1 Awesome Spectral Indices (ASI): Awesome Spectral Indices (ASI) is an ecosystem that provides an open catalogue of spectral indices together with interfaces for Python, Julia, and GEE (Montero et al., 2023). The ASI catalogue could be enriched by linking each spectral index to 1) the instrument for which the index was originally developed, when applicable, and 2) the instruments in AEOI for which the index can be calculated. Such links would help researchers identify the instrumental context in which a spectral index was defined and determine which other instruments can support its computation.

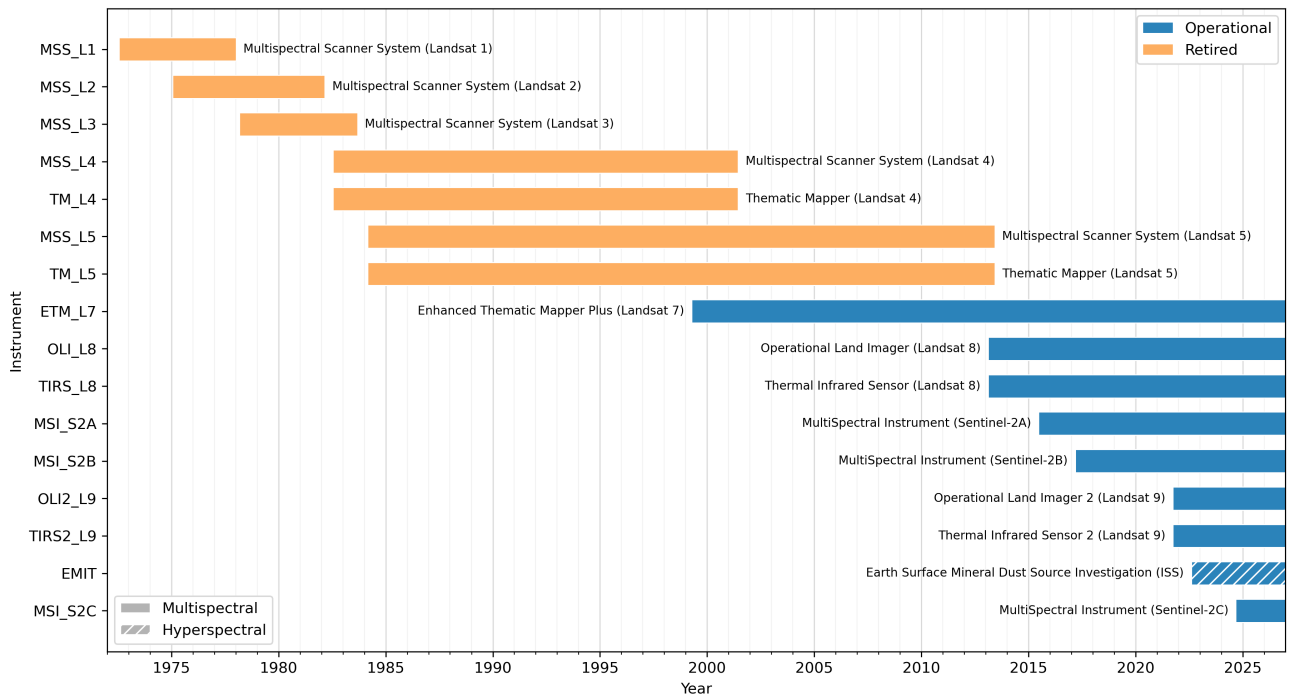


Figure 2. Overview of the Earth observation (EO) instruments included in the pilot release of the catalogue. The metadata used to generate this figure were programmatically extracted from the core schema properties ingested for each instrument.

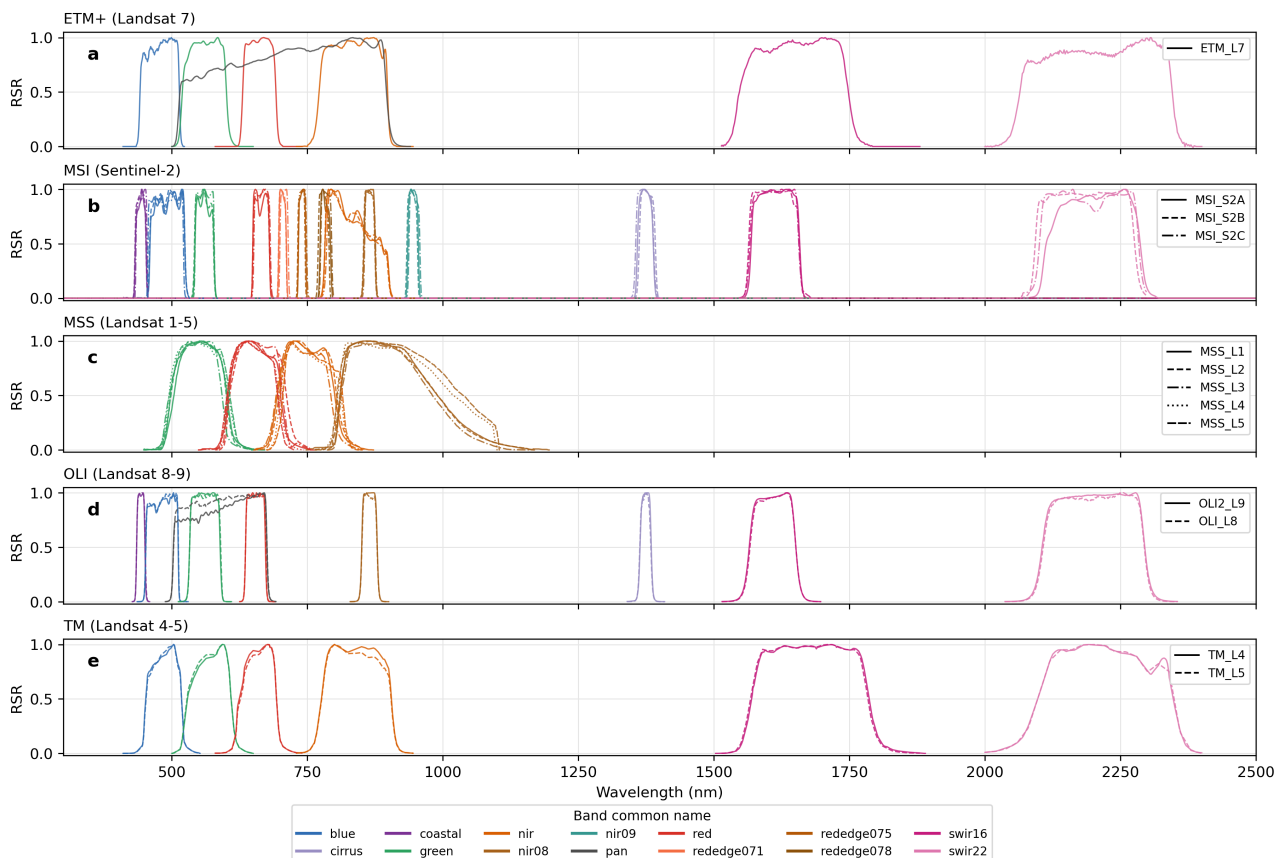


Figure 3. Spectral response functions (SRFs) of the optical bands represented in the pilot release of the catalogue. Curves show the relative spectral response (RSR) of each band for the multispectral instruments included in the pilot release. Band groups follow the common_name attribute used by the STAC Electro-Optical Extension when available; for MSS instruments, this categorisation was inferred from the band names and spectral configuration.

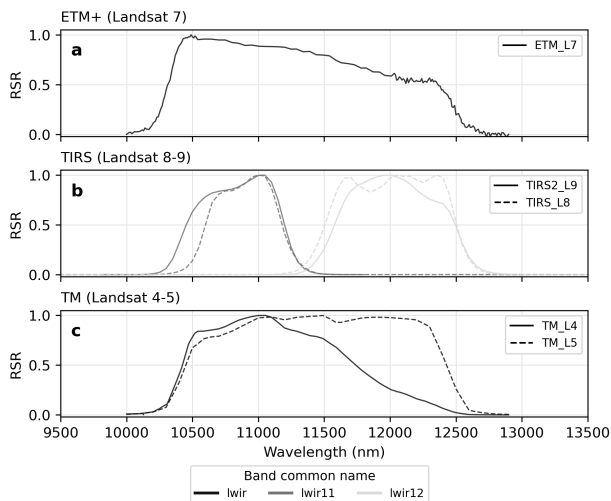


Figure 4. Spectral response functions (SRFs) of the thermal bands represented in the pilot release of the catalogue. Curves show the relative spectral response (RSR) of each thermal band for the multispectral instruments included in the pilot release.

Band groups follow the `common_name` attribute used by the STAC Electro-Optical Extension when available.

4.1.2 Transparent Access to Cloud-Optimized datasets

(TACO) Specification: The TACO specification⁴ is a FAIR-compliant, cloud-native specification that defines a formal and scalable format for packaging and sharing AI-ready EO datasets (Aybar et al., 2026; Pellicer-Valero et al., 2026). Datasets created with TACO, such as CloudSEN12+⁵ and MethaneSET⁶, could benefit from links to the instruments registered in AEOI. Such links would provide a clearer description of the instruments used to acquire the data in a dataset, helping users identify characteristics relevant to AI model development and potentially use this information directly as model inputs (Section 4.2).

4.1.3 SpatioTemporal Assets Catalog (STAC): STAC is one of the most widely used specifications for describing and cataloguing EO data. STAC already provides a comprehensive set of attributes for describing datasets and a well-established extension mechanism for representing additional metadata. An AEOI-related STAC extension could link catalogue instrument properties directly to STAC collections, helping users understand the instrument characteristics associated with specific datasets in a manner analogous to the proposed integration with TACO.

4.2 Artificial Intelligence

The metadata gathered in AEOI are relevant for AI modelling with EO data. Instrument characteristics can inform the design of model architectures, preprocessing strategies, and downstream task definitions. Moreover, because AEOI provides these characteristics in a standardised machine-readable form, they can be used directly by metadata-aware AI models. Such models could incorporate instrument metadata as prior information or as explicit encodings, enabling more robust sensor-aware or sensor-agnostic approaches (e.g., Francis et al., 2022).

⁴ <https://tacofoundation.github.io>

⁵ <https://huggingface.co/datasets/tacofoundation/cloudsen12>

⁶ <https://huggingface.co/datasets/tacofoundation/methanaset>

In this setting, a single model could potentially operate across datasets from different instruments by conditioning predictions on the relevant instrument metadata.

4.3 Further Extensions

One of the main advantages of the proposed schema is its extensibility. New extensions can be introduced to address additional instrument properties, data-access mechanisms, or domain-specific requirements. The current implementation has a strong focus on optical instruments, but the same approach can support extensions for radar and LiDAR systems. Such extensions could describe properties relevant to current radar missions such as Sentinel-1 and ESA Biomass, retired radar missions such as TerraSAR-X and ALOS-PALSAR, spaceborne LiDAR missions such as GEDI, and terrestrial laser-scanning systems.

Additional data-access extensions could also be developed. The existing Earth Engine and Planetary Computer extensions share a common structure that could be adapted to other access points, such as EarthData. In particular, the Planetary Computer extension includes a STAC endpoint property, making it a useful template for other STAC-based catalogues while preserving a standardised representation.

A further extension could represent cross-links to external resources. Because AEOI is not the only resource that records EO instrument characteristics, such an extension could connect AEOI items to equivalent or related entries in other catalogues. For example, the Index DataBase (IDB), which catalogues spectral indices, also includes sensor information. AEOI instruments could also be linked to relevant data-source catalogues such as EarthData.

4.4 Programmatic Access

The catalogue was designed to be machine-readable from the outset. This model has proven useful in the ASI ecosystem, and AEOI extends this idea by providing the complete catalogue as a JSON file. This format makes the registry accessible from most programming languages and facilitates the development of APIs for common EO data-processing environments, including Python, R, Julia, and GEE interfaces in JavaScript or Python.

4.5 Limitations and Future Work

AEOI was designed to make community contributions straightforward: missing instruments can be added through simple YAML files. This design, however, also implies a maintenance burden. Thousands of EO instruments have been developed, and keeping a registry complete and up to date requires sustained effort. Version v0.1.0 includes 16 items, and this number is expected to increase in future releases. Community engagement will therefore be essential. A similar pattern occurred in Awesome Spectral Indices, which grew from approximately 60 spectral indices in version v0.1.0 to more than 270 in version v0.10.0 through contributions from maintainers and the remote sensing community. We expect AEOI to grow in a similar manner, beginning with widely used instruments and gradually expanding as users request and contribute the instruments needed for their workflows. Future item creation could be accelerated with large language models that retrieve information, draft input files, and iterate against validators. However,

such workflows should remain under human supervision, particularly for the audit phase. AI systems could instead support maintainers by preparing draft pull requests for new instruments. As the catalogue grows, an initial Python API is planned to provide programmatic access in one of the most widely used languages for EO data operations.

5. Conclusions

In this paper, we introduced Awesome Earth Observation Instruments, an open-source, machine-readable, and community-driven catalogue of EO instruments. The catalogue is based on a compact JSON Schema that describes general instrument properties and can be extended through additional schemas for spectral, optical, geometric, and data-access metadata, including Earth Engine and Planetary Computer access points. The catalogue is designed to be extensible by the community while maintaining a rigorous validation workflow that combines human audit and automatic schema validation before ingestion. Community members can contribute missing instruments or propose new extensions. We foresee AEOI becoming a useful integration point for existing resources such as Awesome Spectral Indices and TACO, and a practical reference for EO data users who require instrument characteristics and metadata for reproducible analysis workflows.

Code and Data Availability

The Awesome Earth Observation Instruments (AEOI) catalogue and the code used to generate Figures 2-4 are publicly available on GitHub^{7,8}.

Declaration of AI Use

We used ChatGPT for language editing and Codex for programming assistance. All content and code were reviewed and edited as needed, and we take full responsibility for the published work.

Acknowledgments

All authors acknowledge support by the Climate Change AI Innovation Grants program, hosted by Climate Change AI with the support of Quadrature Climate Foundation, the Global Methane Hub, Google DeepMind, and the Canada Hub of Future Earth. D.M. and M.D.M. acknowledge support from the European Space Agency — ESA (‘cube.ng — No Gap Earth System Data Cubes’, CCN1 of ARCEME — Adaptation and Resilience to Climate Extremes and Multi-hazard Events, ESA Contract No. 4000144482/24/I-KE). L.G.C. and C.A. acknowledge support from the Spanish Ministry of Science, Innovation and Universities (grant PID2023-148485OB-C21/C22 funded by MCIU/AEI/10.13039/501100011033 ERDF, EU).

References

Alvarez-Vanhard, E., Corpetti, T., Houet, T., 2021. UAV & satellite synergies for optical remote sensing applications: A literature review. *Science of Remote Sensing*, 3, 100019. dx.doi.org/10.1016/j.srs.2021.100019.

⁷ <https://github.com/awesome-spectral-indices/awesome-earth-observation-instruments>

⁸ <https://github.com/davemlz/srf-aeoi>

Aybar, C., Contreras, J., Ma, C. et al., 2026. The missing piece: Standardising for AI-ready earth observation datasets. *Terra-Bytes - ICML 2025 workshop*.

Bugbee, K., le Roux, J., Sisco, A. et al., 2021. Improving Discovery and Use of NASA’s Earth Observation Data Through Metadata Quality Assessments. *Data Science Journal*, 20. dx.doi.org/10.5334/dsj-2021-017.

Clark, R. N., Swayze, G. A., Livo, K. E. et al., 2024. Imaging Spectroscopy: Earth and Planetary Remote Sensing with the PSI Tetracorder and Expert Systems from Rovers to EMIT and Beyond. *The Planetary Science Journal*, 5(12), 276. dx.doi.org/10.3847/PSJ/ad6c3a.

Drusch, M., Del Bello, U., Carlier, S. et al., 2012. Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120, 25–36. dx.doi.org/10.1016/j.rse.2011.11.026.

Francis, A. M., Mrziglod, J., Sidiropoulos, P. et al., 2022. SEN-SEL: A Deep Learning Module for Creating Sensor Independent Cloud Masks. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–21. dx.doi.org/10.1109/TGRS.2021.3128280.

Gorelick, N., Hancher, M., Dixon, M. et al., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. dx.doi.org/10.1016/j.rse.2017.06.031.

Henrich, V., Jung, A., Götze, C. et al., 2009. Development of an online indices database: Motivation, concept and implementation. *Proceedings of the 6th EARSeL Imaging Spectroscopy SIG Workshop: Innovative Tool for Scientific and Commercial Environment Applications*, Tel Aviv, Israel.

Justice, C., Vermote, E., Townshend, J. et al., 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, 36(4), 1228–1249. dx.doi.org/10.1109/36.701075.

Montero, D., Aybar, C., Mahecha, M. D. et al., 2023. A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research. *Scientific Data*, 10(1). dx.doi.org/10.1038/s41597-023-02096-0.

Pellicer-Valero, O. J., Aybar, C., Czerkawski, M. et al., 2026. TACO: Operationalizing AI-Ready EO datasets. Copernicus.

Running, S. W., Justice, C. O., Salomonson, V. et al., 1994. Terrestrial remote sensing science and algorithms planned for EOS/MODIS. *International Journal of Remote Sensing*, 15(17), 3587–3620. dx.doi.org/10.1080/01431169408954346.

Torres, R., Snoeij, P., Geudtner, D. et al., 2012. GMES Sentinel-1 mission. *Remote Sensing of Environment*, 120, 9–24. dx.doi.org/10.1016/j.rse.2011.05.028.

Wulder, M. A., Masek, J. G., Cohen, W. B. et al., 2012. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, 122, 2–10. dx.doi.org/10.1016/j.rse.2012.01.010.

Wulder, M. A., White, J. C., Loveland, T. R. et al., 2016. The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment*, 185, 271–283. dx.doi.org/10.1016/j.rse.2015.11.032.