



OGC Metadata Summit 2026

Bolzano, 7–8 September 2026 · Hosted by EURAC Research

Summary of Sessions and Outcomes

Resources:

Recordings, presentations are available under [Events Public > 2026 > 09 Bozen](#).

Executive summary

- It is not always possible to precisely describe who consumes open data metadata. Usage evidence is thin across public bodies, research communities, and commercial providers.
- A layered, dual-audience model is emerging: a short discovery “business card” for humans and text-based search engines, plus a machine-actionable “full CV” with schemas, vocabularies, and validation rules for agents.
- There is a demand of guidance how to implement annotation type of data, namely quality descriptions with the flexibility, in additive processes, clearly correlated with data on different granularity levels
- AI systems benefit from explicit, typed, linked structure. Cryptic titles and free-text-only records lead to hallucinations. Providing schemas, vocabularies, and API-based access cuts inference cost and error rates. Yet a shared “AI profile” - a consistent set of properties across standards specifically chosen to help AI systems reach more consistent conclusions is needed.
- OGC Building Blocks were presented live and received well. They give small, testable, reusable specifications that can combine schema, vocabulary, examples, and SHACL validation, and can wrap legacy standards rather than replace them.
- The group asked OGC to define governance for Building Blocks: versioning, registers, formal review, and a path to good-practice status alongside INSPIRE and ISO.
- The room prefers less specification volume, more clarity: simpler entry points, better tooling, open access to standards, and multilingual guidance for national implementations.
- Convergent priorities for the handover to the Canadian summit: user evidence, dual-layer discovery to detail, machine-actionable validation, and shared governance of registers.

- Communities use different encodings (STAC, DCAT, custom schemas) and don't connect them back to what ISO already covers.
- metadata requirements are tool-specific rather than universal: even when the underlying data and its meaning are identical, different tools expect different sets of metadata elements, so there is no single way to describe a dataset once and have every tool consume it.

Context

The OGC Metadata Summit convened public-sector, research, and industry participants in Bolzano on 7 and 8 September 2026, hosted by Eurac Research. The event is one European convening in a wider OGC Testbed on Trusted Data and Systems and is paired with a companion summit in Canada. The program combined short plenary talks, live demos, and structured breakouts against a common set of guiding questions covering metadata value and related pain points, who consumes it, what works in current pipelines, how a flexible architecture could evolve, and how registers and sovereignty should be governed.

This document reports the outcomes of each session in the order they took place, using the guiding questions as the structure. It closes with cross-cutting themes, participant feedback, and the priorities to carry into the Canadian summit.

Day 1 · Monday 7 September

Opening plenary · Setting the scene: Metadata value, directions and challenges

Guiding questions. Why are we here? Who are the main data providers and users? What are the recurring pain points across communities?

Key contributions

- **Alexander Jacob (Vice Head of Institute for Earth Observation, Eurac Research):** opened the summit with a welcome and framed why Eurac chose to host this convergence event.
- **Jordi Escriu (Scientific Project Officer, JRC), Updates from Metadata-related EU Policies:** framed the European regulatory landscape and the complexity of INSPIRE. Metadata often ends up as a compliance task rather than a discovery tool without understanding the users' needs. Too many overlapping obligations, including parallel standards are not helping and are tricky in practice.
- **Maria Bellantone (Open Science Research Support Office, Eurac Research), Metadata: My FAIR Lady:** argued that a large share of research data goes unused for want of good metadata. A little metadata added correctly makes the whole record unusable. While efforts are taken to derive metadata from existing assets, also using AI, good practices are needed to limit bias, ethical and maintenance issues.
- **Benedikt Gräler (Managing Director, 52North), From Tedious to Automated:** introduced a multi-agent AI approach to reducing the metadata curation burden. He proposed an agent workflows that automate the generation and refinement of metadata to ensure datasets are AI-ready and reliable for autonomous GIS workflows.
- **Jean-Philippe Aurbout (Expert Spatial Data Management, European Environment Agency):** shared numbers from EEA holdings, their main discovery and consumption streams. He proposed a dual layer: a short business card for discovery and a full CV for machine use as they serve different touchpoints. It was also noticed that generalized data catalogs struggle to convert into website traffic and cryptic titles like "Copernicus N2K 2024" caused AI systems to hallucinate.

Pain points named in the room

- Providers cannot say who consumes their data. Portal analytics rarely reach the record level.
- Free-text metadata reads well for humans but confuses agents.
- Regulation stacks on regulation. Providers meet the letter of INSPIRE but lose the intent.
- Metadata is treated as a box-tick at publication rather than a living record.
- Fear of misuse and job-security concerns reduce willingness to describe data well.

Why we need metadata and who consumes it – workshop

Guiding questions. For each dataset, who is the consumer, why do they need it, how do they access it (machine to machine, AI, portal), what query or prompt would they use, and what evidence do we

have? Which categories are involved: Discovery, Spatial, Temporal, Trust, Fitness, Access, or AI and machine use?

What the groups reported

- Almost every group admitted the same gap: real consumer evidence is thin. Providers infer users rather than measure them.
- Concrete machine-to-machine cases were named. WMO station metadata is consumed by ECMWF models. Marine vocabularies drive dataset descriptions across observatories.
- AI consumers prefer strongly-typed, nested structures with linked vocabularies. Human consumers prefer flat, readable summaries. This tension recurred across tables.
- Trust, license, and access metadata came up as often as discovery fields. Consumers want to know if they use the data before inspecting its content.
- Disincentives to good metadata were named openly: fear of misuse, loss of control, and no reward for the extra work.

Current developments and success stories

What works

- **Stefania Morrone (Senior Data Architect, EPSILON IT), Metadata Experiences while Supporting the EEA:** tested AI readiness across formats. Agents hallucinate against plain HTML pages, perform better against APIs, and best against GeoJSON and JSON-LD. Only an agent with download and analysis rights produced a usable answer end to end.
- **Cayetano Benavent (Director of Engineering, CARTO), AI-ready Metadata in Practice, Lessons from Building Portolan:** presented the Portolan specification. It is STAC-based, cloud-native, and intentionally designed for a dual audience with a “howling fish” approach for people and agents. The specification is public on GitHub.
- **Bartolomeo Ventura and Andrea Vianello (Senior Technicians, Eurac Research), EDP-Portal, A FAIR and Collaborative Platform:** combines GeoNetwork, STAC, openEO, and Jupyter, with F-UJI FAIR scores integrated into the catalogue.
- Homogeneous, well-scoped catalogues work. Problems appear when catalogues mix standards and domains.
- South Tyrol reported a controlled pipeline where all producers follow the same standard. Multilingual publication in Italian and German still creates friction.

What does not

- Harvesting duplicates rather than federates. Copies drift from source and updates are lost.
- Harmonization to the lowest common denominator loses domain-specific detail.
- Links between metadata and the actual asset break too often. Agents cannot act on records they cannot resolve.

- Commercial imagery providers publish thin metadata. Users pay and then hunt in forums for missing details.

Implementation patterns and pipelines - workshop

Guiding questions. Build a wall of real pipelines. Where does metadata originate, how is it managed and validated, and where does it break?

Pipeline observations

- Metadata should be born with the data. When the data expert is not in the loop at the start, downstream records inherit misinterpretations.
- A metadata record has its own lifecycle. Records should be reviewed and updated at each processing stage, not only at publication.
- Standards work well for high-value operational products. Long-tail research and custom data have weaker curations and rely on informal practices.
- AI-driven pipelines need shared identifiers, a common metadata layer, and a model and data registry across steps. Bad prompts and missing context cause hallucinations.
- Provenance and licence interoperability were flagged repeatedly. Licences must be machine-readable and comparable, not only present.

Final plenary talks · Toward a managed ecosystem

- **Niels Hoffman (Architect, Geonovum), From Profile Proliferation to a Managed Ecosystem:** introduced the OGC Building Blocks framework as the way out of profile proliferation. The framework was the anchor for the Day 2.
- **Núria Julià (UAB, CREAM), Geospatial Metadata Manager:** presented the Geospatial Metadata Manager tool and the workflow it supports practitioners maintaining catalogues.

Day 2 · Tuesday 8 September

Opening plenary · Report back on what works and what breaks

Guiding questions. Where in the metadata lifecycle do we succeed, and where do the joins fail? Which layers of the architecture are most exposed?

Convergent findings across groups

- Well-defined profiles work when they cover both the container and the content, including shared vocabularies for band names and units.
- Documentation works. Compliance with stated standard works. Problems concentrate on boundaries: mappings between standards, aggregation across domains, and legacy conversions such as INSPIRE to DCAT.
- Persistent identifiers were flagged as a shared prerequisite. Without them, links break as records evolves.
- Data spaces and Destination Earth are moving forward but are not yet fully operational, which creates friction in EU tenders that assume they are.
- South Tyrol added a specific national point: multilingual publication requires care, and Italian to German/English conversions lose information.

Plenary · A flexible metadata architecture

- **Bente Lilja Bye (BLB):** presented Clear-EO event-driven workflows for floods, drought, agriculture, and air quality. She raised the question of metadata that triggers actions and provided the cases where OGC standards can describe event identity, time, area, phenomena, severity, and next allowed action so that a notification from one digital twin can safely trigger another. GEO related work with Data Management Principles, Knowledge Hub with open Language models were presented as potential contributor and multiplier.
- **Martin Klopfer (OGC):** framed the direction of further discussion with observations: there is no single ring to rule all metadata. Cities and sectors cannot be forced into one universal schema. Instead, the proposal is to work on grouping requirements into small, testable Building Blocks that combine schema, vocabulary, examples, and validation rules.
- **On AI readiness:** the key is explicit, declared meaning. If an agent can follow a link to an authoritative definition, it does not need to be inferred. Every inference has a token and error cost.

Live demo · OGC Building Blocks

Presenters: Niels Hoffman (Geonovum) and Piotr Zaborowski (OGC), with contributions from Martin Klopfer and the summary by Alex Jacobs.

What was shown

- A Building Block is a GitHub folder with a JSON definition, optional schema, examples, and a JSON-LD context. Dependencies point to other Building Blocks.

- The Dutch base register for buildings was demonstrated as an OGC API Features Building Block, with attributes linked to the Cadaster vocabulary. Semantics enables the linked definitions and computational consistency and reusability to allow to reuse existing abstract models.
- The Dutch DCAT-AP profile was shown declaring its lineage on DCAT-AP and adding Dutch SHACL constraints. Validation runs against parent and local rules combined.
- Cross-domain patterns and transformer plug-ins were shown, including RML and XSLT engines that make transformations machine-actionable.
- The presenters walked through the several of the large portfolio of existing Building Blocks that exist at different maturity levels, from experimental in the OGC incubator repository to compliant and comprehensive with the suggested potential of readiness to be governed in the main OGC register.

Questions raised in the room

- Matthias Morh (online) asked how Building Blocks are referenced from data. Answer: through compliance declarations, similar to STAC extensions.
- Jeff Yutzler asked how long-running efforts like INSPIRE and SDI can align with Building Blocks. Answer: this is exactly what the OGC Testbed on Trusted Data and Systems asks organizations to shape.
- Karel Charvat (Plan4All) offered a provocation. Standards should evolve to become AI-friendly rather than expect AI to fit legacy structures. He argued that unstructured text alongside metadata can help language models understand meaning better in many cases.
- A participant asked whether governance of Building Blocks needs to be defined. Answer from OGC staff: yes, the discussion started in Helsinki, continues in Cape Town, and is part of the wider OGC architecture reforms.

Afternoon plenary · Live-demo reflections and metadata generator

Speakers reflected that “Building Blocks” as a name creates confusion with other communities, but the substance is clear once shown. The reflection produced an open request for a longer deep dive at a future event.

- **Jordi Escriu (JRC):** demoed an AI-assisted, schema-agnostic metadata generator, published under open licence in Code Europa. It splits attributes into free-text (Type A), data-intrinsic (Type B), and institutional (Type C), uses retrieval-augmented generation, and keeps the human in the loop for validation.
- **Niels Hoffman (Geonovum):** argued that metadata tools should point to shared Building Block registers instead of embedding schemas privately, so the ecosystem grows around agreed schemas.
- **Piotr Zaborowski (OGC):** presented a digital building permit example. SHACL rules from Building Blocks validated completeness of data in legal submission workflows across different municipalities. Mixing different building models (IFC, OGC) with city geospatial rules is complicated, but encapsulation allows to create libraries of machine executable rules.

Evening workshop · What do you need from metadata standards

Guiding questions. What should we stop, keep, and try? Which standard elements are strong, which are weak, and where are the gaps in the stack?

Table findings, condensed

- **Stop.** Reading manual-heavy specifications end to end. Overlong standards. Ambiguity in units and measurement. Manual conversions.
- **Keep.** Versioning, template examples, good editors, openness, OGC as a governance body, and validation of data and metadata records.
- **Try.** Break big profiles like GeoDCAT-AP into composable Building Blocks. Add statistical vocabularies so consumers know how a mean or an average is computed. Attach uncertainty. Describe which analytics are meaningful on a dataset. Define a discovery process for registers.

What is well standardised

- Time, location, title, keywords, description, contact, publisher, citation, temporal and geographic footprint, file type, and calibration.

What is missing or weak

- Compliance documentation across data chains and processes.
- Practical data-quality documentation, even where the standard has room for it.
- Space-data interoperability, notably for space weather.
- Processing, provenance, and lifecycle description.
- Governance for standards evolution and versioning.
- Implementation examples, best practice, and low-barrier documentation.
- Auxiliary files, images, and supporting documents. Referencing them cleanly is hard.
- Partial-metadata authorisation. Some metadata should not be fully open.
- Access to standards themselves. ISO documents behind paywalls slow adoption. National-language versions are missing.

Discussion

- A research organisation can host its own Building Block register that reuses national and international definitions as a lightweight interoperability hub.
- Governance should exist at three levels: international standards of bodies, national bodies, and domain- or organisation-specific registers.
- Federated architecture is more agile but more fragile. Trust between register maintainers matters as much as the technology.
- Converting STAC into ISO or DCAT is possible but loses item- and asset-level detail. Transformers in Building Blocks are the natural place to hold such mappings.

Cross-cutting themes summary

Theme	What we learned
User evidence	Providers routinely lack analytics on who reaches their records. Both human and machine consumers must be measured before we can prioritise categories.
Dual audience	A short discovery record for humans and a linked, typed, validated record for machines. Portland spec and Building Blocks converge on this pattern.
AI readiness	Explicit meaning, linked vocabularies, typed schemas, and API-based access reduce hallucination and cost. Cryptic titles fail.
Governance	Building Blocks need a formal registration, versioning, and review process alongside INSPIRE and ISO. OGC will decide the shape in the following TCs.
Sovereignty and federation	Registers can be federated across international, national, and domain layers. The Federation depends on trust and stable references.
Simplification and access	Standards should be free, translated, and paired with tools that hide complexity. Better editors and lower barriers to entry drive uptake.
Provenance and trust	Provenance, licenses, and quality must be first-class and machine-comparable, not free-text notes.
Quality information models	Few groups have raised the issues around support of quality descriptions of data/datasets in flexible, sometimes dynamic/version-able annotations
Dynamics	Current systems already require live metadata, interactive query mechanisms, up-to-date information; M2M consumption already shows new representations needed, but none claimed it is optimal for close future systems. Demand of dynamic but well-grounded metadata systems.

Priorities and handover to the Canadian summit

Working consensus from the two days points to five priorities the Bolzano group asked to carry into the Canadian summit and the wider OGC Testbed on Trusted Data and Systems.

- Build a shared evidence base on metadata consumers. Instrument portals and APIs, publish the numbers, and use them to prioritise categories.
- Adopt a dual-layer approach for discovery to detail. A short business card for humans plus a linked, typed, validated technical layer for machines.
- Advance Building Blocks with clear governance. Define versioning, registration, and a good-practice path that connects to INSPIRE and ISO rather than replacing them.

- Make standards machine-actionable and cheaper to use. Attach schemas, SHACL rules, transformers, and examples. Keep the AI-readiness cost of inference in mind.
- Lower the barrier to entry. Open standards, translations, better editors, and short guidance that a non-specialist can follow.

Acknowledgements: The summit was organized thanks to the support of EURAC Research, event host. Organization was co-sponsored by the Horizon Europe Research and Innovation Programme funding under Grant Agreements No. 101156488 (SeaDOTs), 101137787(FOCAL), 101181408(BioClima)