AI-Augmented Metadata
A JRC study

82nd INSPIRE MIG-T Meeting, 27 June 2025

Margherita DI LEO, Ilyas TIOUASSIOUINE, Jordi ESCRIU, Hagar LOWENTHAL
European Commission, Joint Research Centre (JRC)

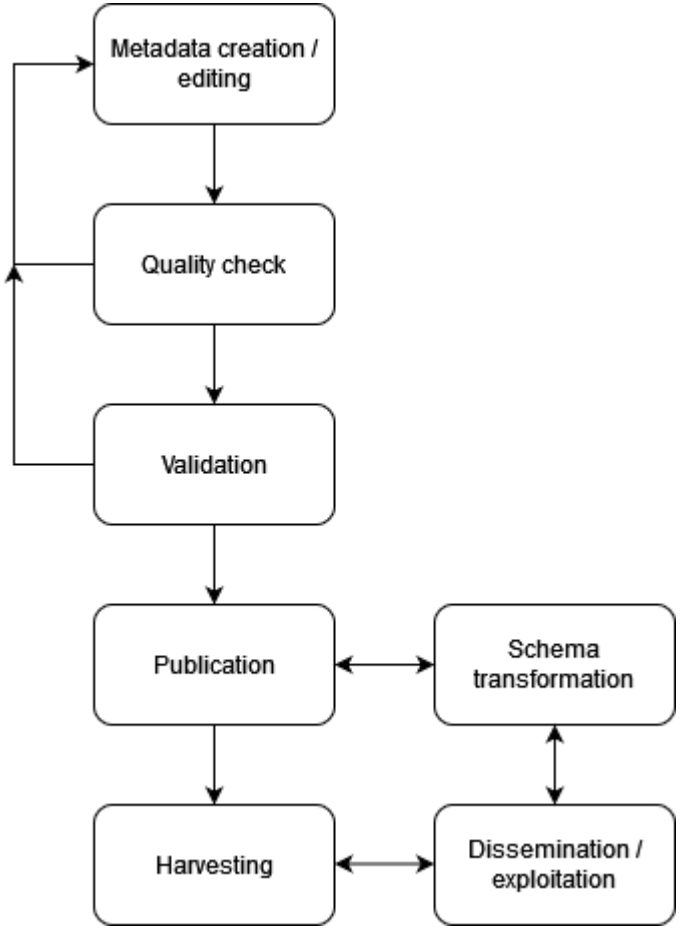
AI in Every Field: Enhancing Productivity and Efficiency

Leveraging AI for Tedious Tasks

- AI is increasingly used in various domains.
- AI can automate repetitive and mundane tasks, freeing humans for more strategic work.
- Metadata management: a tedious task. Crucial for data discoverability, accessibility, and usability.
- The metadata lifecycle involves multiple stages, each with its own challenges.

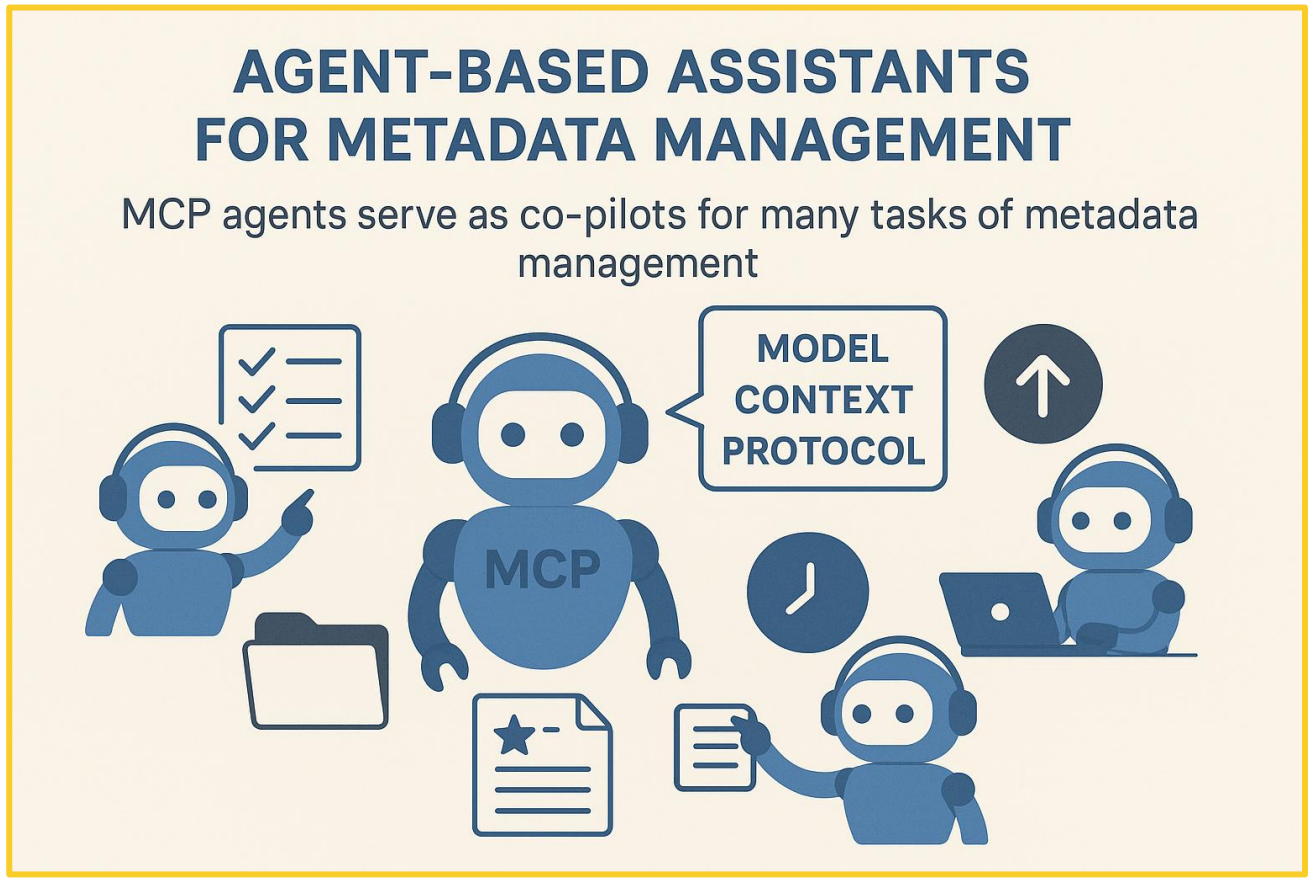
Metadata Lifecycle Analysis
Identifying Areas for AI-Driven Improvement

- The metadata lifecycle involves stages like creation, validation, publication, harvesting, refinement, and dissemination.
- Each stage has its own set of challenges, such as quality control, consistency, standard compliance, discoverability, and interoperability.
- AI can be applied to various stages to improve efficiency, accuracy, and scalability.



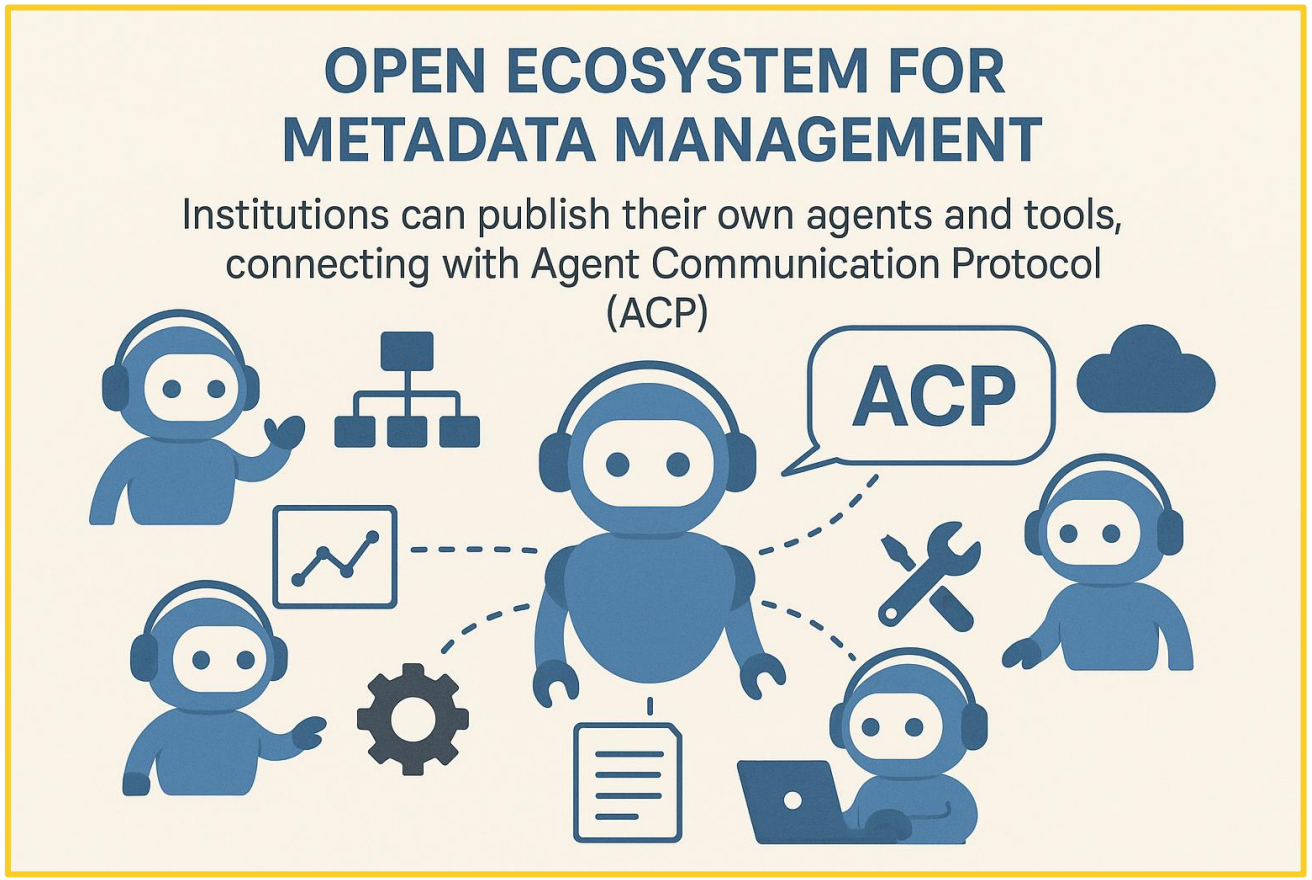
Proposed approach

Modular Infrastructure with AI Agents



- Proposed architecture: modular, cooperative, adaptive.
- Based on Model Context Protocol (MCP) – open source, scalable, proposed standard.
- Specialised AI agents can assist (co-pilot) in tasks like metadata creation, service metadata, validation, transformation.
- Modular infrastructure allows for flexibility, scalability, and easier maintenance.

Open Ecosystem for Metadata Management:
Sharing and collaboration

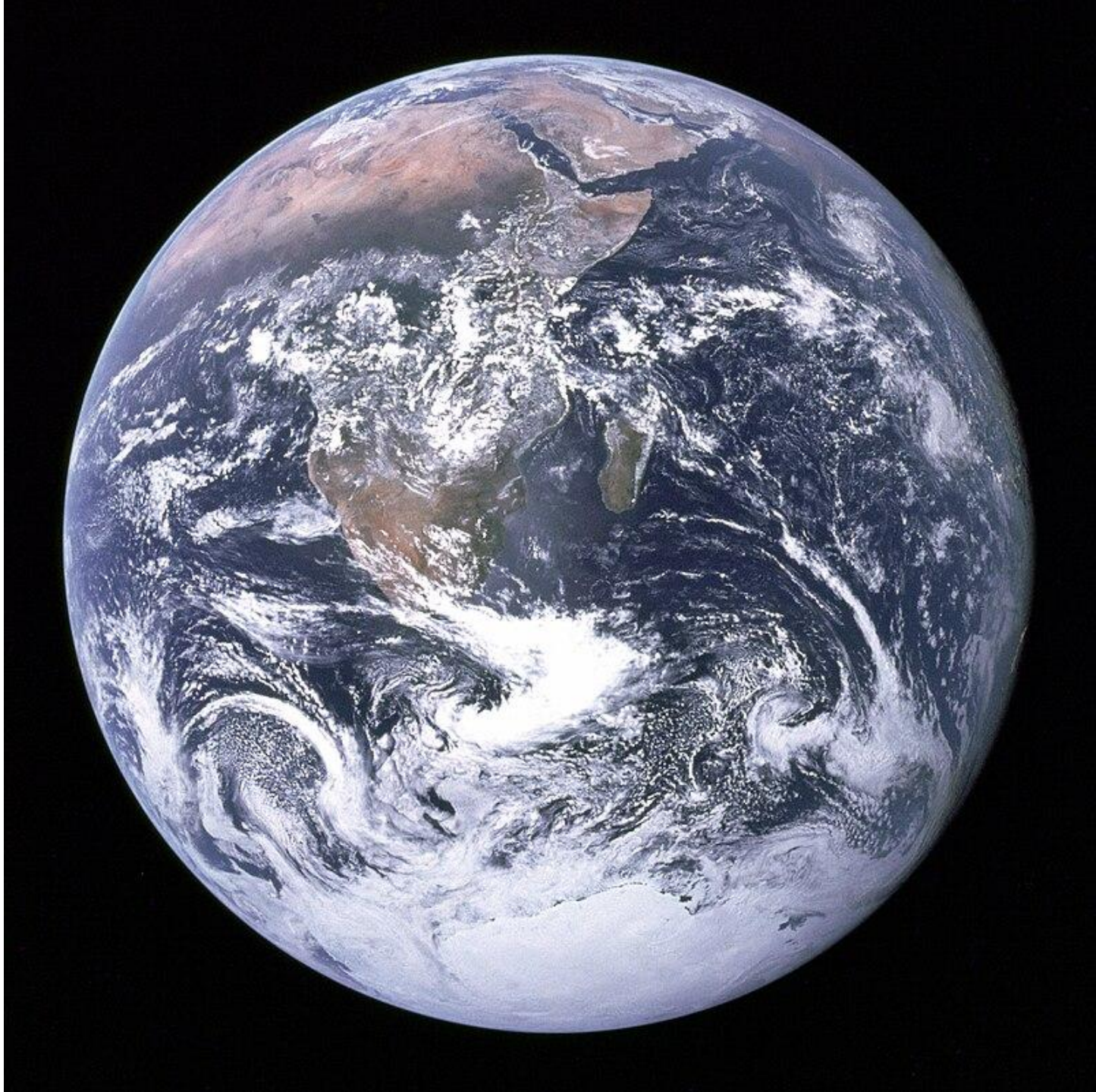


- Institutions can publish their own agents and tools, connecting with ACP (Agent Communication Protocol), promoting sharing and collaboration.
- The ACP enables agents to communicate and interact with each other, regardless whether they are built with MCP or not.
- Benefits of an open ecosystem:
  - Reuse, interoperability and cooperation among institutions.
  - Faster development and deployment of new agents and tools.
  - Improved management through shared knowledge and expertise.
  - Improved metadata quality.

Prototype

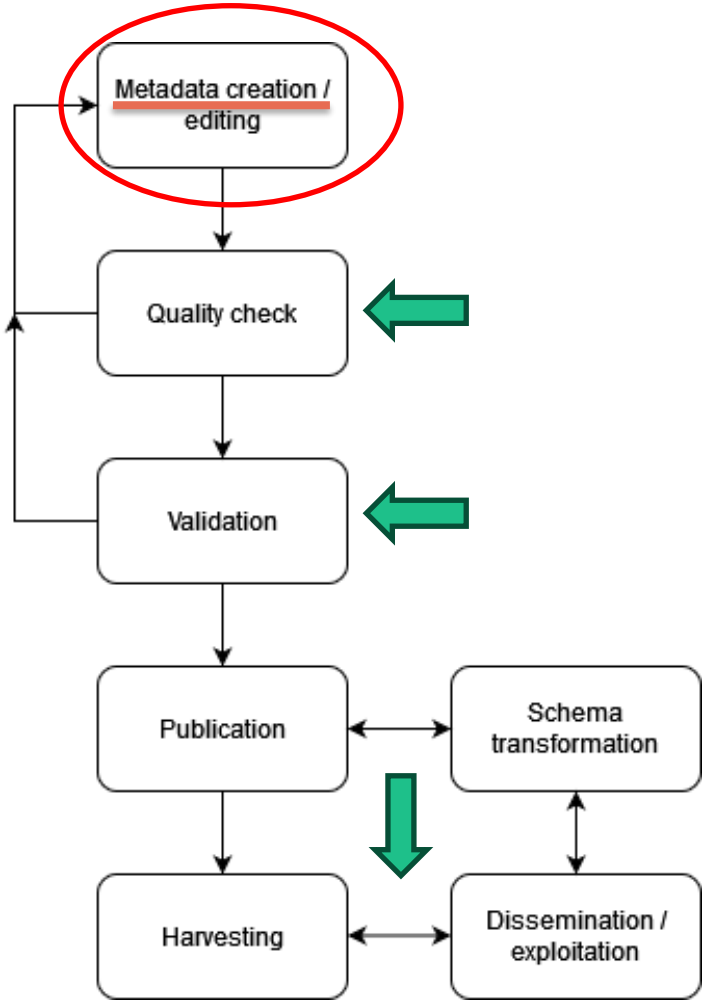
AI-powered metadata editor

- Geospatial metadata
- INSPIRE and GeoDCAT-AP schemas



Prototype

AI-powered metadata editor



Information-rich fields

→ Description

Vocabulary:	DCMI Metadata Terms
URI	http://purl.org/dc/terms/description
Label	Description
Definition	An account of the resource.

→ AI-ready documentation

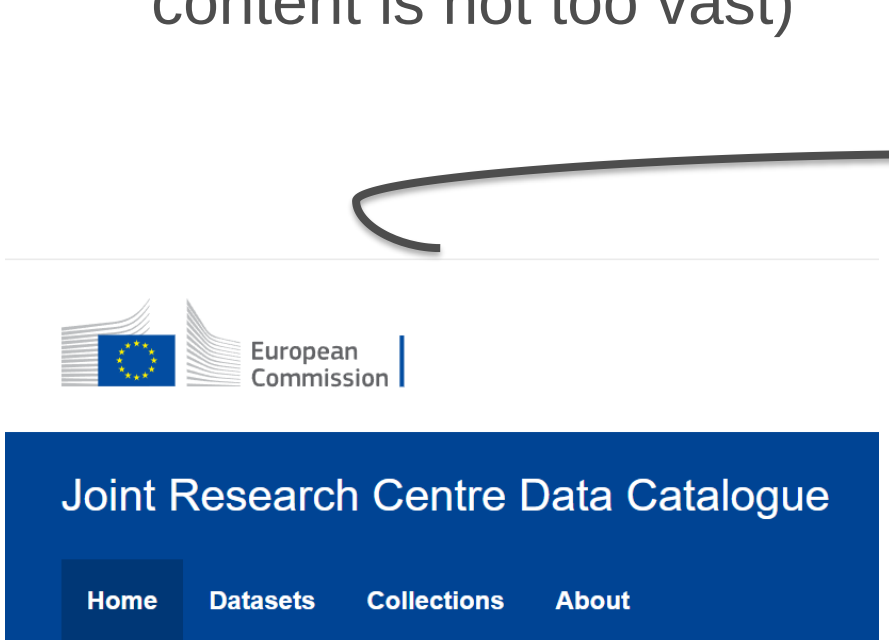
is useful for human producers

Prototype

AI-powered metadata editor

Assumption:

Natural language processing tools (NLP) with attention mechanisms (typically transformers) do work well for synthesising textual content (if the content is not too vast)



European Commission

Prototype

AI-powered metadata editor

Example:

Structuring content of GeoDCAT-AP free-text metadata fields

This dataset provides [*broad topic of dataset content*].

It depicts [*topic or theme*] for [*geographical scope*] from [*beginning temporal scope*] to [*ending temporal scope*], with a resolution of [*resolution, technical details*].

The data have been collected using [*methodology*] and sourced from [*data sources*].

The data are available in [*data format*] and have [*dataset characteristics*].

Potential applications include [*intended use or application*].

Related publications can be found in [*related publications or studies*].

[*availability*] and [*access*]. [*contact information*]. [*update Frequency*]. NOTE: [*limitations or notes*].

Prototype

AI-powered metadata editor

Example:

Structuring content of GeoDCAT-AP free-text metadata fields

This dataset provides [*broad topic of dataset content*].

It depicts [*topic or theme*] for [*geographical scope*] from [*beginning temporal scope*] to [*ending temporal scope*], with a resolution of [*resolution, technical details*].

The data have been collected using [*methodology*] and sourced from [*data sources*].

The data are available in [*data format*] and have [*dataset characteristics*].

Potential applications include [*intended use or application*].

Related publications can be found in [*related publications or studies*].

[*availability*] and [*access*]. [*contact information*]. [*update Frequency*]. NOTE: [*limitations or notes*].

Prototype

AI-powered metadata editor

Early lessons learned

- Good generative results can be obtained with SLMs and RAG-like architecture even if the documentation is not optimal. They need though to be supervised by humans.
- It is not possible to produce the full range of metadata out of only data: to produce satisfactory metadata with AI we also need “meta”-documentation. Meta-documentation should probably be written by humans.

Key takeaways

- Modular design: agents specialise on small, well – defined tasks
  - Aid to implementation of the data-service linking simplification; geospatial-HVD tagging good practice; metadata encoding changes; finetuning of code lists; language-related transformations; …
- Co-pilot tool for assisting metadata managers, not substituting them.
- Early stage of the research, we welcome your experiences.

Thank you

Ilyas.TIOUASSIOUINE@ec.europa.eu

Margherita.DI-LEO@ext.ec.europa.eu