



OPENGOVWEEK  
19 - 23 maggio 2025

ITALIA  
opengov

# Il ruolo dei dati aperti nell'era dell'intelligenza artificiale

Giorgia Lodi

Prima tecnologa presso Istituto di Scienze e Tecnologie della  
Cognizione del CNR (CNR-ISTC)

[giorgia.lodi@cnr.it](mailto:giorgia.lodi@cnr.it)

# Nel 2023

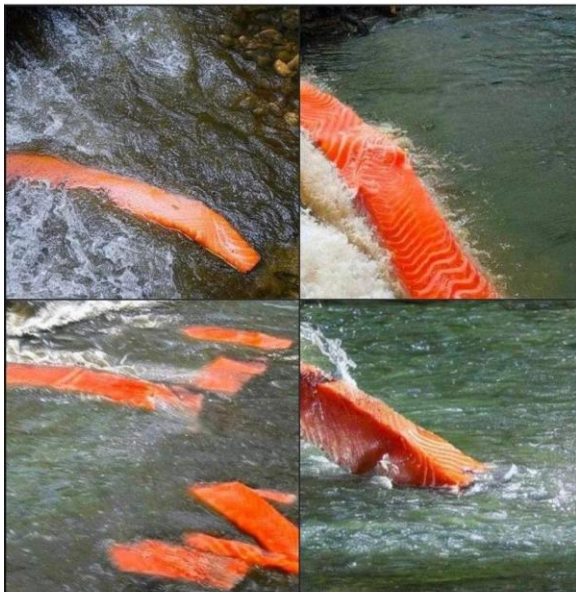
Avevamo ancora qualche problema con alcuni software di intelligenza artificiale

**JavaScript Developer** 852,977 followers  
4d • 🌐

When you ask an **#AI** to recreate « a salmon swimming down a river »

The end is near 😬

Thanks **Carlos la Orden Tovar** for sharing



👍❤️ 7,625

148 comments • 189 reposts

e per la specie *theragra calgogramma*?

## Nel 2025

Ho chiesto a  
ChatGPT in  
italiano...

Il codice FAO 3-alfa per la specie *Theragra chalcogramma* (nome corretto attuale: *Gadus chalcogrammus*, comunemente noto come merluzzo dell'Alaska o pollock dell'Alaska) è **PAC**.



# Nel 2025

e per la specie *theragra calcogramma*?

Ho chiesto a ChatGPT in italiano...

Il codice FAO 3-alfa per la specie *Theragra chalcogramma* (nome corretto attuale: *Gadus chalcogrammus*, comunemente noto come merluzzo dell'Alaska o pollock dell'Alaska) è **PAC**.



Gemini

2.5 Flash ▾

Upgrade

Ho chiesto la stessa cosa a Gemini...



mi dici qual è il codice FAO 3-alfa per la specie *Theragra chalcogramma*?



Il codice FAO 3-alfa per la specie *Theragra chalcogramma* (il merluzzo dell'Alaska) è **ALK**.



# Nel 2025

Ho chiesto a Gemini...

cloud you please provide the metadata DCAT-AP\_IT in RDF/Turtle for an hypothetical dataset on the water quality parameters?

✦ I can certainly do that! Here's the DCAT-AP\_IT metadata in RDF/Turtle for a hypothetical dataset on water quality parameters.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix adms: <http://www.w3.org/ns/adms#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .
@prefix locn: <http://www.w3.org/ns/locn#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix xkos: <http://eurovoc.europa.eu/schema/xkos#> .
@prefix schema: <http://schema.org/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

# Dataset URI (replace with your actual dataset URI)
<http://data.arpalazio.it/catalogue/datasets/qualita-acque-fiume-[nomefiume]-[nomecomune]>
dct:title "Parametri Qualità delle Acque del Fiume [Nome Fiume] nel Comune di [Nome Comune]"
dct:description "Il dataset fornisce dati sui parametri di qualità delle acque del fiume [Nome Fiume] nel Comune di [Nome Comune]. I dati sono aggiornati al [data]."
dct:issued "2024-03-01"^^xsd:date ; # Date dataset was first published
dct:modified "2025-05-20"^^xsd:date ; # Last modification date, e.g., current date
dcat:theme <http://publications.europa.eu/resource/authority/data-theme/ENVI> ;
```

# Nel 2025

Ho chiesto a Gemini...

***Non sa cos'è il profilo nazionale di metadattazione DCAT-AP\_IT sebbene dica di riuscire a produrlo***

***(come ancora diverse PA italiane guardando la loro sezione amministrazione trasparente 😊)***

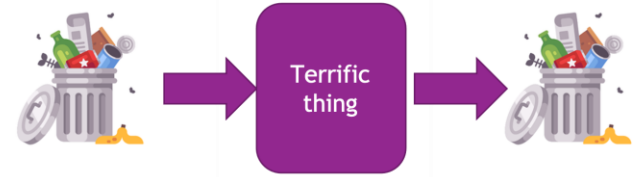
cloud you please provide the metadata DCAT-AP\_IT in RDF/Turtle for an hypothetical dataset on the water quality parameters?

✦ I can certainly do that! Here's the DCAT-AP\_IT metadata in RDF/Turtle for a hypothetical dataset on water quality parameters.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix adms: <http://www.w3.org/ns/adms#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .
@prefix locn: <http://www.w3.org/ns/locn#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix xkos: <http://eurovoc.europa.eu/schema/xkos#> .
@prefix schema: <http://schema.org/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

# Dataset URI (replace with your actual dataset URI)
<http://data.arpalazio.it/catalogue/datasets/qualita-acque-fiume-[nomefiume]-[nomecomune]>
dct:title "Parametri Qualità delle Acque del Fiume [Nome Fiume] nel Comune di [Nome Comune]"
dct:description "Il dataset fornisce dati sui parametri di qualità delle acque del fiume [Nome Fiume] nel Comune di [Nome Comune]. I dati sono aggiornati al [data]."
dct:issued "2024-03-01"^^xsd:date ; # Date dataset was first published
dct:modified "2025-05-20"^^xsd:date ; # Last modification date, e.g., current date
dcat:theme <http://publications.europa.eu/resource/authority/data-theme/ENVI> ;
```

# Software di intelligenza artificiale



Si **basano sui dati** - no dati no AI party!

Sono ancora delle **scatole oscure** (si sta lavorando sul cercare di spiegare il risultato)

Tanto più i dati in input sono di bassa qualità, tanto più il risultato rifletterà l'input e quindi sarà di bassa qualità (**GIGA law**)

Producono **allucinazioni** ma ci sono attualmente delle tecniche per mitigarle (**Retrieval Augmented Generation**) **anche** basate su **dati fortemente strutturati**

In Europa esiste una legge **AI Act** che li cataloga in base **al rischio**: più alto è il rischio più è necessario adottare accorgimenti nell'uso (trasparenza)

# Ripartiamo da qui

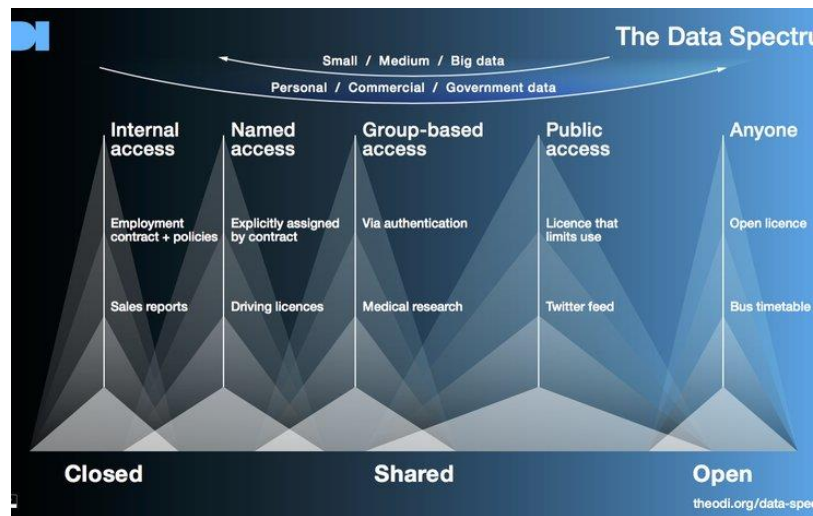
Nell'era dell'Intelligenza artificiale abbiamo bisogno di...

*Principle 1: "a **strong data infrastructure**" as "the foundation for building an open, trustworthy data ecosystem on a global scale and that this can help address our most pressing challenges"*

*Principle 2: Strong data infrastructure includes data across the spectrum, from open to shared to closed.*

**But the best possible foundation is open data,** supported and sustained **as data infrastructure**

*ODI - <https://theodi.org/about-the-odi/our-vision/>*





# Ripartiamo da qui

Open data cosa vuol dire veramente?

Accessibile: da macchine, in formato aperto e **corredato di metadati**

Disponibile: **disaggregato**, secondo i termini di una **licenza aperta** che ne consenta il riutilizzo, anche per **finalità commerciali**

Gratuito: gratuito o a costi marginali sostenuti per la riproduzione, messa a disposizione e divulgazione

# Ripartiamo da qui

Open data cosa vuol dire veramente?

Accessibile: da macchine, in formato aperto e **corredato di metadati**

Disponibile: **disaggregato**, secondo i termini di una **licenza aperta** che ne consenta il riutilizzo, anche per **finalità commerciali**

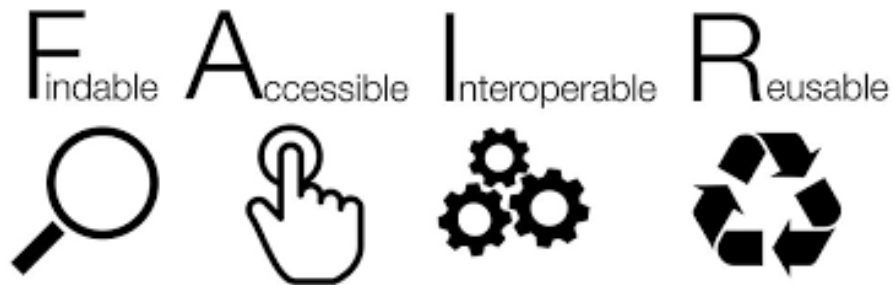
Gratuito: gratuito o a costi marginali sostenuti per la riproduzione, messa a disposizione e divulgazione

Findable  
Accessible  
Interoperabile  
Reusable

## In altri termini

I dati aperti sono TUTTI quei dati che DOVREBBERO seguire i principi FAIR, NON SOLO i dati aperti della ricerca scientifica

(ed è soprattutto vero nell'era dell'AI)



# Il ruolo degli open data nell'AI?

Possono essere quell'infrastruttura di base per l'allenamento dei software di AI

«What **AI-ready data** means in practice depends on many factors, but **established practices like FAIR-ness** (findable, accessible, interoperable, reusable) and **linked data** are a strong starting point»

Nel creare dati aperti, avere in mente le potenzialità per i software di AI

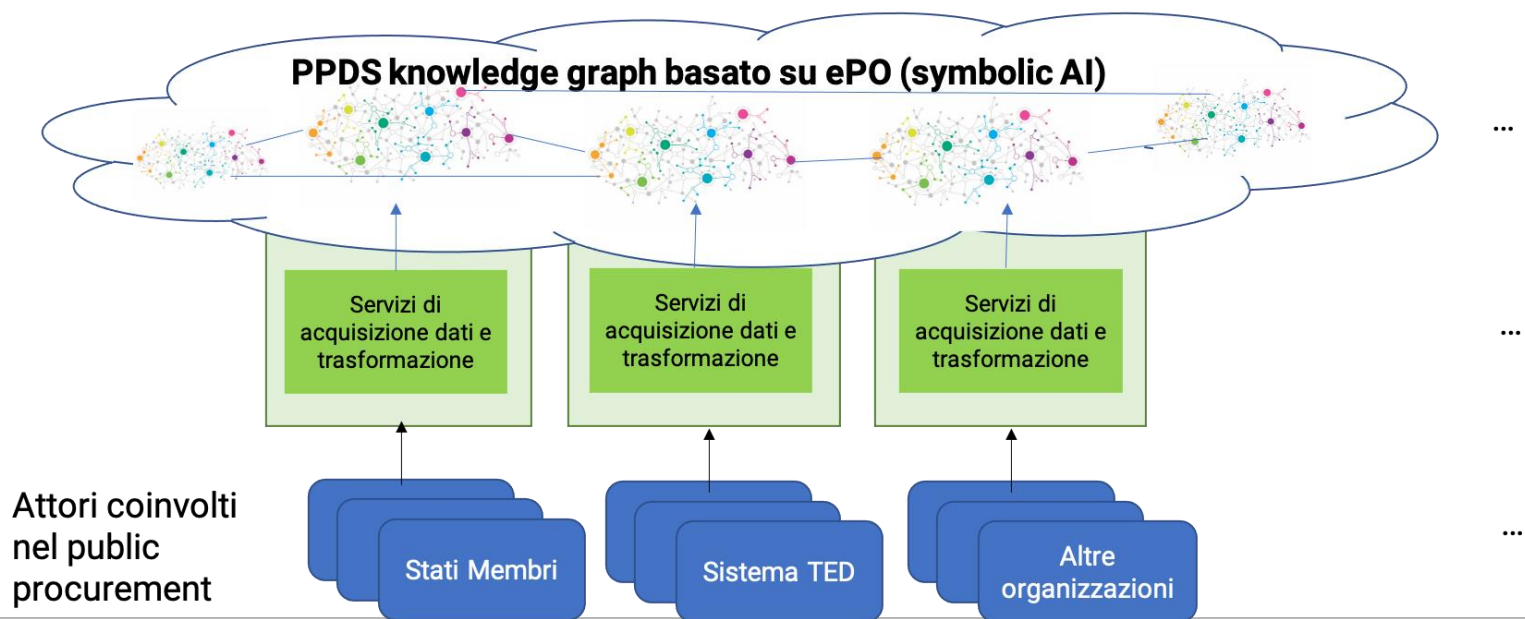
«The government **must guarantee that there are simply no 'gaps to be filled' by secondary sources and hallucinations.** To do so, they should ensure that they are the **first and most important data provider for foundational models** in all subject matters related to governmental affairs»

---

ODI - [https://theodi.cdn.ngo/media/documents/UK\\_government\\_as\\_a\\_data\\_provider\\_for\\_AI.pdf](https://theodi.cdn.ngo/media/documents/UK_government_as_a_data_provider_for_AI.pdf)

# Data space europeo sul public procurement

DATA DISTRIBUTION / DATA ANALYTICS ANCHE SULLA BASE DI AI



# Principali messaggi conclusivi

Nel 2025, con il potenziale dell'Intelligenza artificiale

1. La pubblica amministrazione può avere un ruolo cruciale nella costruzione dell'infrastruttura dei dati per l'AI (**data as a service**)
2. Meglio puntare sulla **qualità dei dati** aperti più che sulla quantità (e molti dati anche sulla trasparenza non lo sono)
3. Non si può più lavorare con i dati di ogni natura come si lavorava anche solo 5 anni fa (**cultura del dato**)
4. Necessario **revisare i processi di gestione dei dati** anche in ottica di apertura per alimentare l'equità di accesso a quei dati utilizzati poi nell'allenamento di software di intelligenza artificiale



**OPENGOVWEEK**  
**19 - 23 maggio 2025**

@opengovitaly | #OpenGovWeek | [open.gov.it](https://open.gov.it)

**ITALIA**  
●●●●● OPENGOV