

Corso nazionale di formazione "Scienza Aperta ed INFN»
17-18 febbraio 2025

Dati Aperti (I)

Marcello Maggi
INFN Sez. di Bari

INDICE

- **I METADATI**
- **IL CICLO DEI DATI APERTI**
- **I DATI FAIR**
- **LE BARRIERE AI DATI APERTI**

La Politica dei Dati Aperti

Un Bellissimo Paradigma, ma:

Complessità

Risorse

Supporto

Infrastruttura



La Mitigazione della Complessità Attraverso L'Uso di Metadati

Gli Oggetti Digitali (DATI)



Documenti

Risultati: pubblicazioni, figure, tabelle

Dati strutturati = Datasets

Annotazioni = Datasets o Informazioni

Pagine web e siti

Codice di programmazione e di sistema

Introduzione ai Metadati

Glossario

- File
 - Format
 - Data
 - MetaData
- Dati sui Dati**

Exabytes

Octets

00000000:	2550	4446	2d31	2e33	0a25	c4e5	f2e5	eba7	%PDF-1.3.%.....
00000010:	f3a0	d0c4	c60a	3420	3020	6f62	6a0a	3c3c	...4 0 obj.<<
00000020:	202f	4c65	6e67	7468	2035	2030	2052	202f	/Length 5 0 R /
00000030:	4669	6c74	6572	202f	466c	6174	6544	6563	Filter /FlateDec
00000040:	6f64	6520	3e3e	0a73	7472	6561	6d0a	7801	ode >>.stream.x.
00000050:	bd5c	cd72	db48	92be	e329	701b	2ac2	82f1	.\.r.H...)p.*...
00000060:	4b82	73eb	71f7	4ecc	447b	7666	ad88	8d8d	K.s.q.N.D{vf...
00000070:	ed3d	d014	2d69	56a6	6451	f4b8	cf7b	f149	..=-iV.dQ...{.I
00000080:	ddaf	300f	d5ef	b35f	567e	89aa	0240	0ab4	..0..._V~...@..
00000090:	bd1b	8a10	5840	5556	fe57	6656	011f	d2bf	...X@UV.WfV...
000000a0:	a51f	d297	af76	45ba	dea5	b9fb	dbad	712b	...vE.....q+
000000b0:	cfca	5adb	f2a3	28d3	79b9	c896	65ba	7e9f	..Z...(y...e~.
000000c0:	fee1	223d	cfb3	a26c	cbf4	629d	ce17	ae1b	.. "=...l..b.....
000000d0:	2e4d	5167	4553	17e9	bc5a	2617	efd3	9717	.MQgES...Z&.....
000000e0:	1745	5aa4	17ef	d2ff	4c67	df5d	9ca5	e7cb	.EZ.....Lg.]....
000000f0:	3a9d	d9f5	07b6	dfc8	b5c1	fdae	c79f	e4ce	:.....
00000100:	3c9d	fdab	eb91	ccfe	c29e	29ef	7f2f	d72a	<.....)../.*
00000110:	9d19	84fe	fdff	61ff	bff2	fa23	afdf	9d25a....#...%
00000120:	0e83	576c	dbf8	d7d2	0ee0	d97c	82e9	0278	..Wl.....l...x

Introduzione ai Metadati

Esterni

```
-rw-r--r-- 1 mmaggi staff 3521 Dec 17 23:05 MIB.json
-rw-r--r-- 1 mmaggi staff 4583 Nov 29 12:57 WHO.json
-rw-r--r-- 1 mmaggi staff 19635854 Nov 29 11:37 affiliations_ror.txt
-rw-r--r--@ 1 mmaggi staff 306708145 Dec 17 23:06 input.json
-rw-r--r-- 1 mmaggi staff 1555 Nov 29 11:50 jsontoyaml.py
-rw-r--r-- 1 mmaggi staff 568 Dec 2 15:00 jsontoyaml_mapping.json
-rw-r--r-- 1 mmaggi staff 2075 Nov 29 13:01 jsontoyaml_v1.py
-rw-r--r-- 1 mmaggi staff 2213 Nov 29 13:11 jsontoyaml_v2.py
-rw-r--r-- 1 mmaggi staff 2837 Nov 29 13:19 jsontoyaml_v3.py
-rw-r--r--@ 1 mmaggi staff 22030525 Dec 17 23:07 output.yaml
-rw-r--r-- 1 mmaggi staff 306708174 Nov 29 11:33 prova.json
-rw-r--r-- 1 mmaggi staff 127 Dec 11 11:10 schemaINV.yaml
-rw-r--r-- 1 mmaggi staff 398 Dec 11 11:10 schemaROR.json
-rw-r--r-- 1 mmaggi staff 2367 Nov 29 11:37 short_prev.yaml
-rw-r--r-- 1 mmaggi staff 23308 Nov 29 11:35 small.json
-rw-r--r--@ 1 mmaggi staff 34634627 Oct 31 19:04 v1.55-2024-10-31-ror-data.csv
-rw-r--r--@ 1 mmaggi staff 306708145 Oct 31 19:03 v1.55-2024-10-31-ror-data.json
-rw-r--r--@ 1 mmaggi staff 40265842 Oct 31 19:03 v1.55-2024-10-31-ror-data_schema_v2.csv
-rw-r--r--@ 1 mmaggi staff 225420385 Oct 31 19:03 v1.55-2024-10-31-ror-data_schema_v2.json
-rw-r--r-- 1 mmaggi staff 2072 Nov 29 12:06 xcheck.py
```

Interni



Entrée véranda.jpg

JPEG image - 325 KB

Information

[Show Less](#)

Created	Monday, 22 January 2024 at 12:52
Modified	Today, 12:38
Last opened	--
Content created	Monday, 29 April 2019 at 20:20
Dimensions	1000x750
Resolution	72x72
Colour space	RGB
Colour profile	sRGB IEC61966-2.1
Device make	samsung
Device model	SM-G930F
Aperture value	1,53
Exposure time	1/25
Focal length	4,2 mm
ISO speed	250
Flash	No
F number	f/1,7
Exposure program	Normal
Metering mode	Centre-weighted average
White balance	Auto
Content Creator	G930FXXU3ERL3

Gli Standard nei Metadati

Grande Varietà

Dublin Core 1995 aveva 15 elementi

Dublin Core 2008 ha 55 elementi + estensioni

ISO15836

CrossRef

DataCite

Tanti!! Articoli scientifici

Immagini Satellitari

ISO19115

.... E tanti altri...

Traduttori di Standard cercasi

Gli Standard nei Metadati

Vocabolari

Un esempio: Dublin Core 1995 per i documenti

#	Elemento	Descrizione	#	Elemento	Descrizione	#	Elemento	Descrizione
1	Titolo	Il nome dato alla risorsa	6	Contributore	Un'entità che ha contribuito alla creazione della risorsa	11	Fonte	Una risorsa dalla quale la risorsa attuale è derivata
2	Creatore	Il responsabile della creazione della risorsa	7	Data	Un punto temporale associato alla risorsa	12	Lingua	La lingua della risorsa (es. IT, EN, FR, ...)
3	Soggetto	L'argomento della risorsa	8	Tipo	La natura o il genere del contenuto della risorsa	13	Relazione	Una relazione con altre risorse
4	Descrizione	Una sintesi del contenuto della risorsa	9	Formato	Il formato fisico o digitale della risorsa	14	Copertura	L'estensione o la copertura spaziale e temporale della risorsa
5	Editore	Il responsabile della pubblicazione della risorsa	10	Identificatore	Un identificatore univoco per la risorsa (DOI, URI, ...)	15	Diritti	Informazioni sui diritti e sulle licenze associate alla risorsa

L' ARBITRARIETÀ È UNA BARRIERA

L'uso dei Metadati



**Gli Standard nei metadati
Rendono **intelligibili** la
tipologia dei dati in uso e
permettono ai **sistemi** di
operare automaticamente**

Machine-actionable metadata

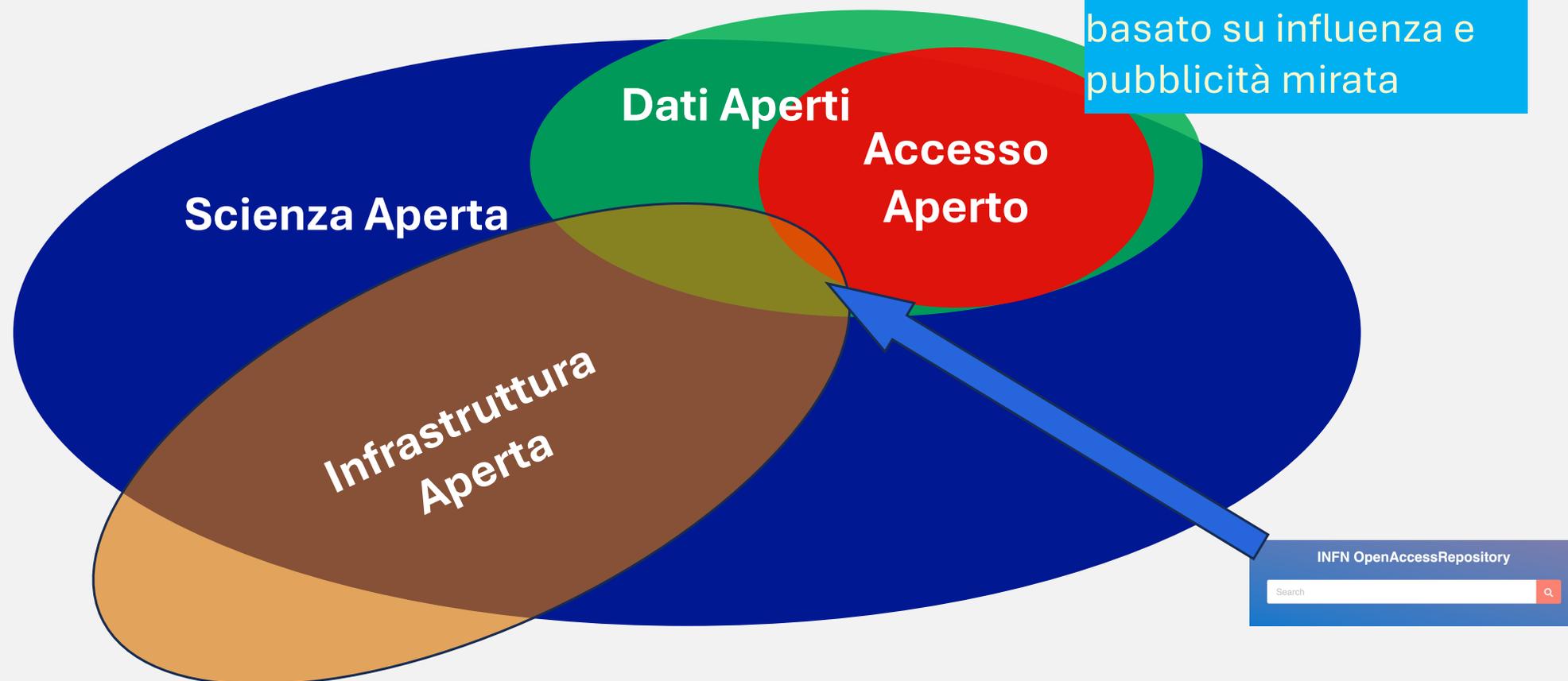


**I Dati Aperti
Nel Contesto Della
Scienza Aperta :
Il Cittadino che fa Scienza**

Dati Aperti nella Scienza Aperta

Governo Aperto
(PA – AGID)

Economia
Modello Business
basato su influenza e
pubblicità mirata



INFN OpenAccessRepository

Dati Aperti nella Scienza Aperta



Dati Aperti nella Scienza Aperta



**Come Riusciamo a Incapsulare
La Conoscenza nei Metadati
Ed Anche nei Dati
Per Continuare ad Usare i Dati?
I Principi FAIR**

I Principi FAIR

L'origine

*There is an **urgent** need to improve the **infrastructure** supporting the **reuse** of scholarly data.*

*A diverse set of stakeholders —representing academia, industry, funding agencies, and scholarly publishers— have come together to design and jointly endorse a concise and measurable **set of principles** that we refer to as the **FAIR Data Principles***

Accessible



Indicizzazione nei repository

I metadati devono essere esposti con protocollo OAI-PMH o API semantiche.

Disponibilità dei metadata

A lungo termine
In caso di dato rimosso
Sempre e comunque

Interoperability

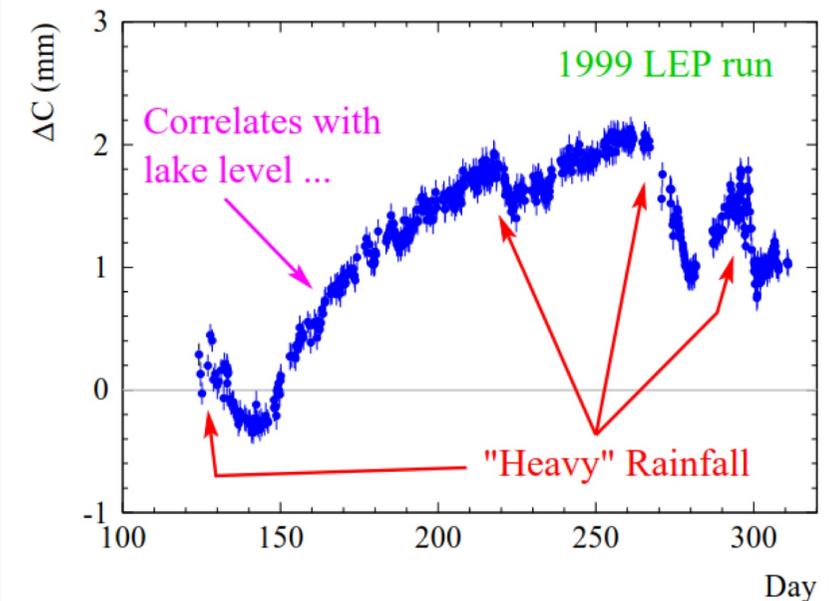
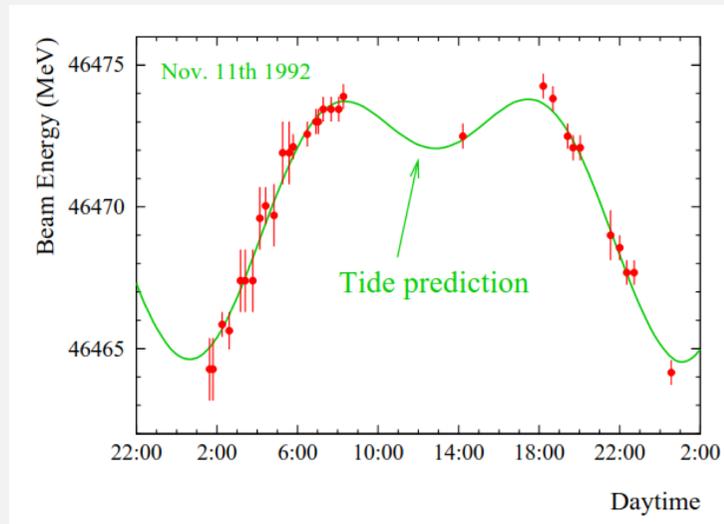
Uso di linguaggi standardizzati

(Meta)dati, accessibili, condivisi e applicabili per la rappresentazione della conoscenza.

Vocabolari conformi ai principi FAIR.

Collegamenti a (meta)dati correlati

Riferimenti qualificati



Reuse

Phys. Lett. B 856 (2024) 138957

Contents lists available at [ScienceDirect](#)

 **ELSEVIER**

Physics Letters B

journal homepage: www.elsevier.com/locate/physletb



Letter

Long-range near-side correlation in e^+e^- collisions at 183-209 GeV with ALEPH archived data

Yu-Chen Chen^a, Yi Chen^a, Anthony Badea^b, Austin Baty^c, Gian Michele Innocenti^a, Marcello Maggi^d, Christopher McGinn^a, Michael Peters^a, Tzu-An Sheng^a, Jesse Thaler^a, Yen-Jie Lee^{a,  ,*}

^a Massachusetts Institute of Technology, Cambridge, MA, USA
^b University of Chicago, Chicago, IL, USA
^c University of Illinois Chicago, IL, USA
^d INFN Sezione di Bari, Bari, Italy



Metadati ricchi e dettagliati

Devono descrivere chiaramente contesto e metodologia

La provenienza dei dati

Fonti, modifiche e versioni precedenti

Licenze d'uso.

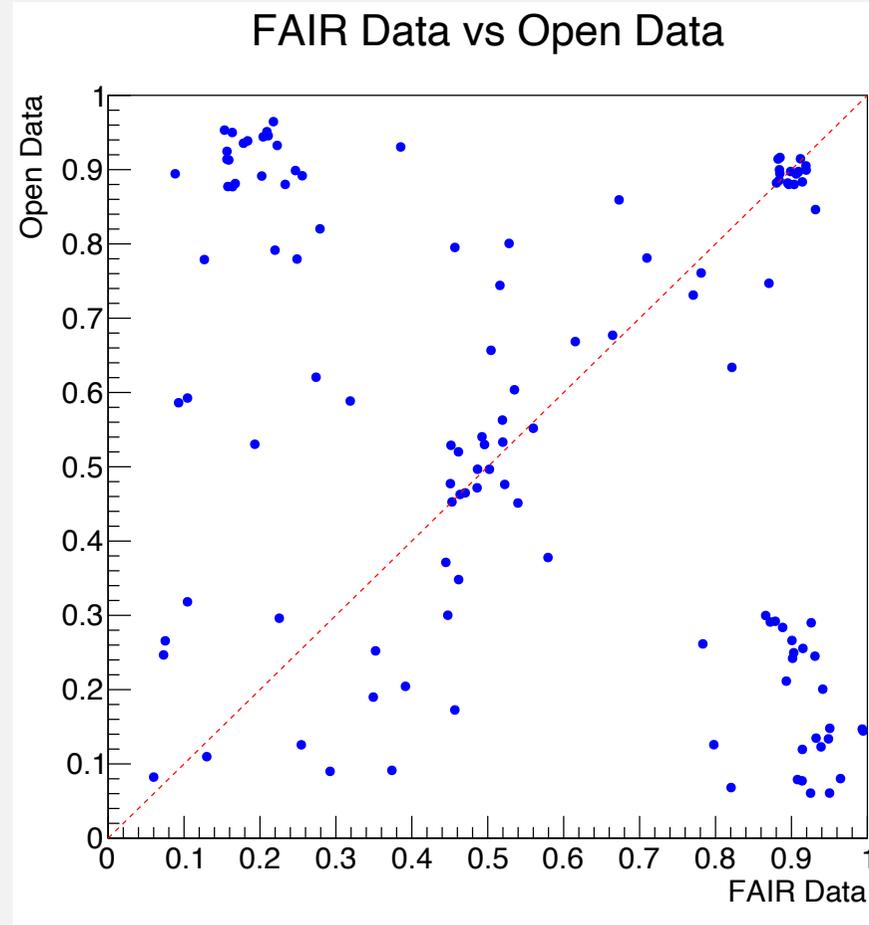
preferibilmente **aperte**: (CC0, CC-BY, ODC-By, ODbL)

Aderenza a standard disciplinari (FITS, EDM4HEP, MIAME)

Dati FAIR e Dati Aperti

Documenti
Amministrativi

Documenti
Governativi



Dati perfettamente
riutilizzabili

Dati Astro sotto
Embargo

Barriere all'Apertura Dei Dati in ambito scientifico (e non)

Resistenza Culturale

Concorrenza

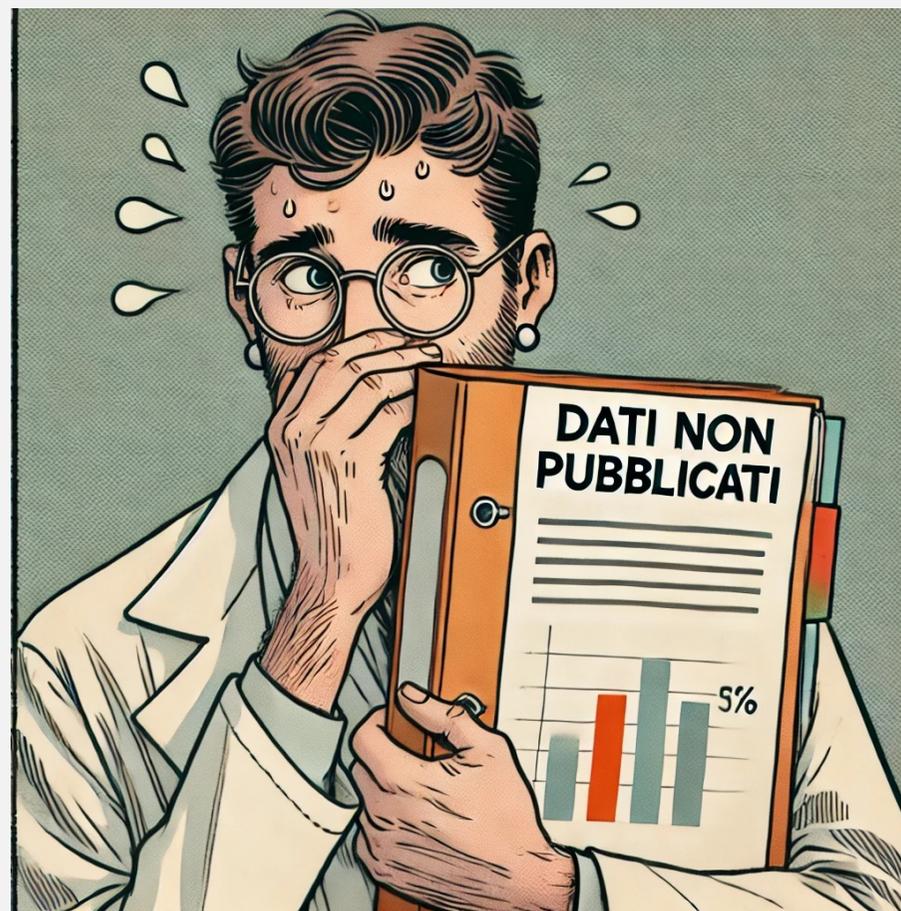
Gruppi con maggiori risorse

Reticenza

Gruppi con migliori risultati

Competenza

Complessità del dato per essere descritto correttamente



Nonostante l'Embargo

Incentivi?

Valutazione Accademica

Non è una pubblicazione

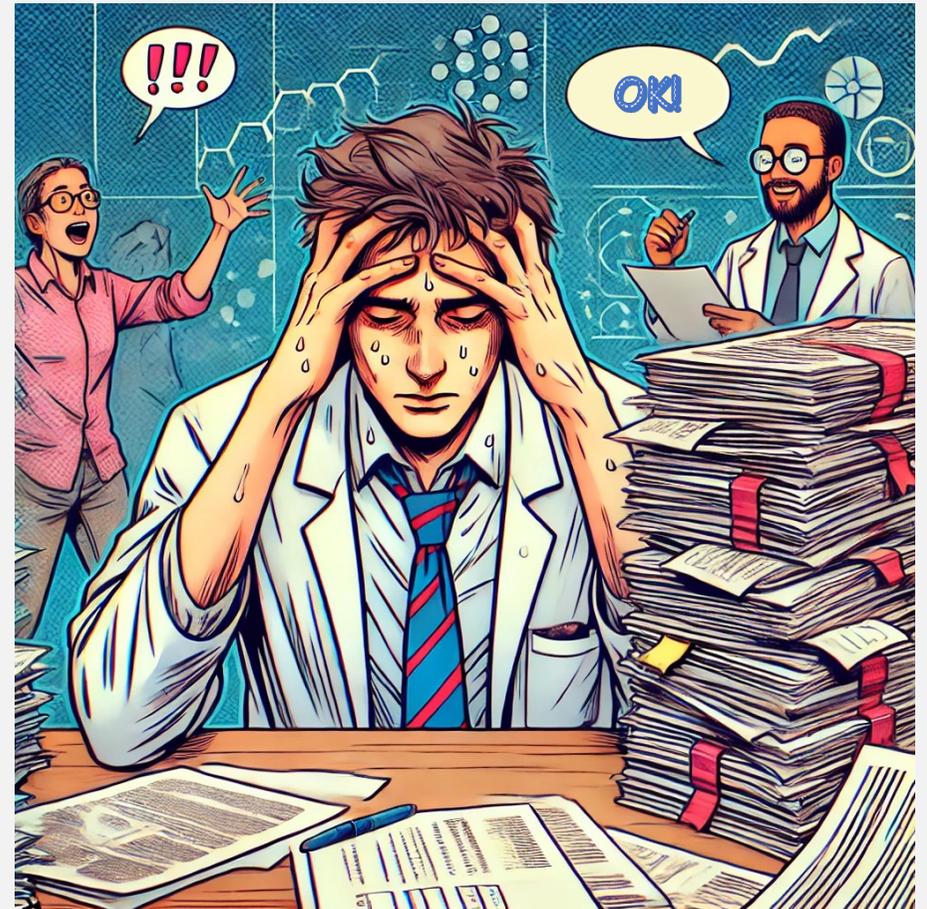
Valorizzazione

Una responsabilità ma senza visibilità

Fondi

Risorse aggiuntive

Personale



Attività Impegnativa

Vincoli

Norme

GDPR protezione dati personali

Proprietà

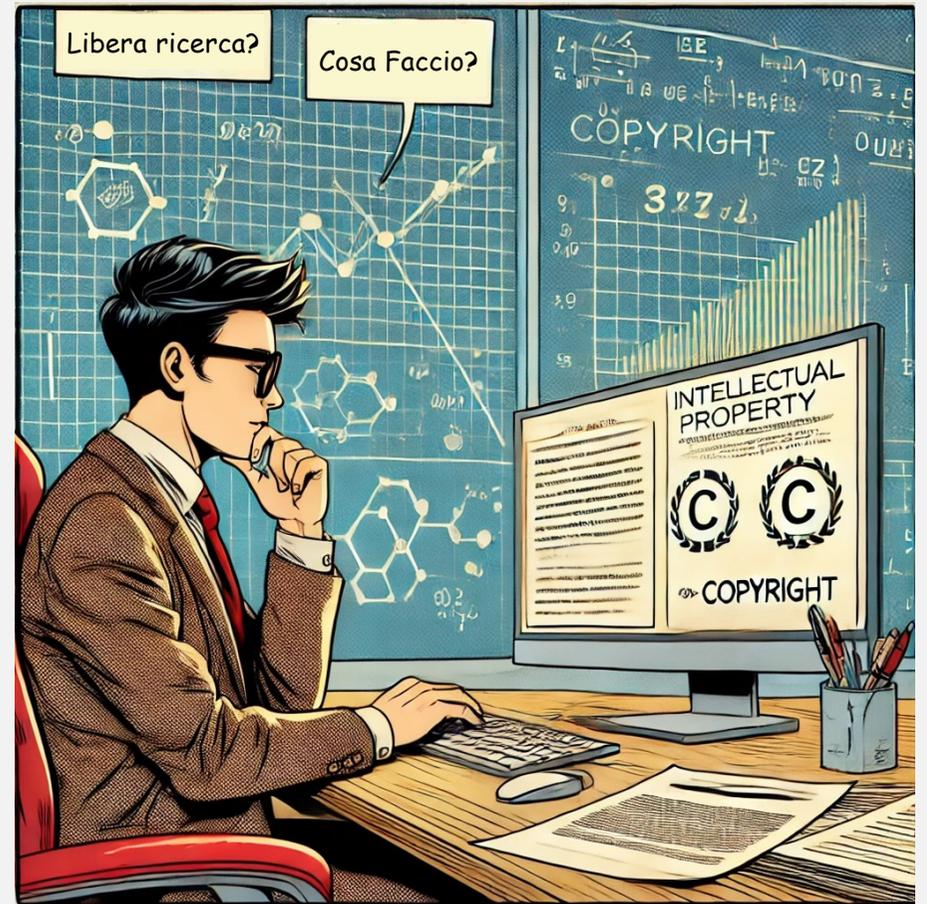
Segreto industriale

Brevetti

Restrizioni

Biomedicina

Sicurezza Nazionale



Attività Impegnativa

Dunque:

**Dati Aperti sono un Pilastro della
Scienza Aperta**

**I Principi FAIR aiutano
i Dati Aperti ad
Essere Riutilizzabili**

**Servono Politiche
per Abbattere
Le Barriere**



Grazie!

Contatti

- marcello.maggi@ba.infn.it
- Gruppo di lavoro Open Science dell'INFN: openscience@lists.infn.it