

01

# DATI CHE SI PARLANO: STANDARD E FAIR

Dal laboratorio a un ecosistema di open data

---

LUCIA LI DONNI

Sapienza Università di Roma  
(lidonnilucia@gmail.com)

---

“Oltre l’apertura: gli Open Data  
plasmano la società”  
INAPP-Roma, 10 dicembre 2025

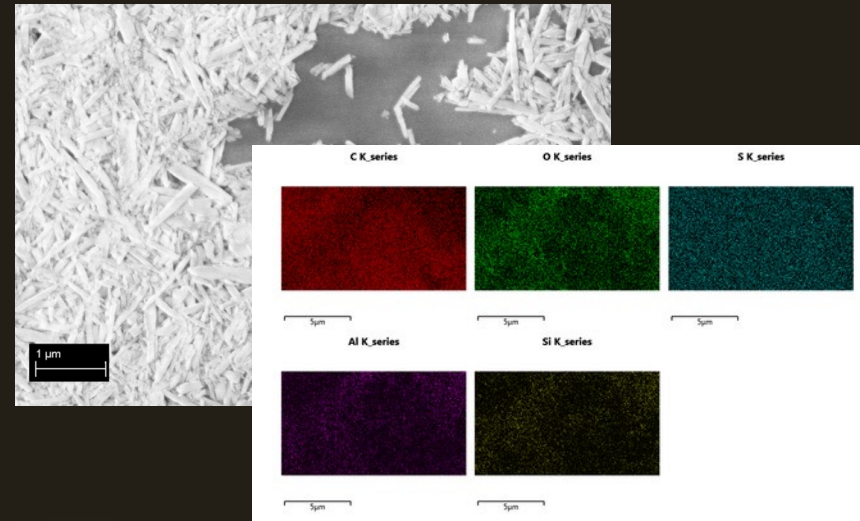
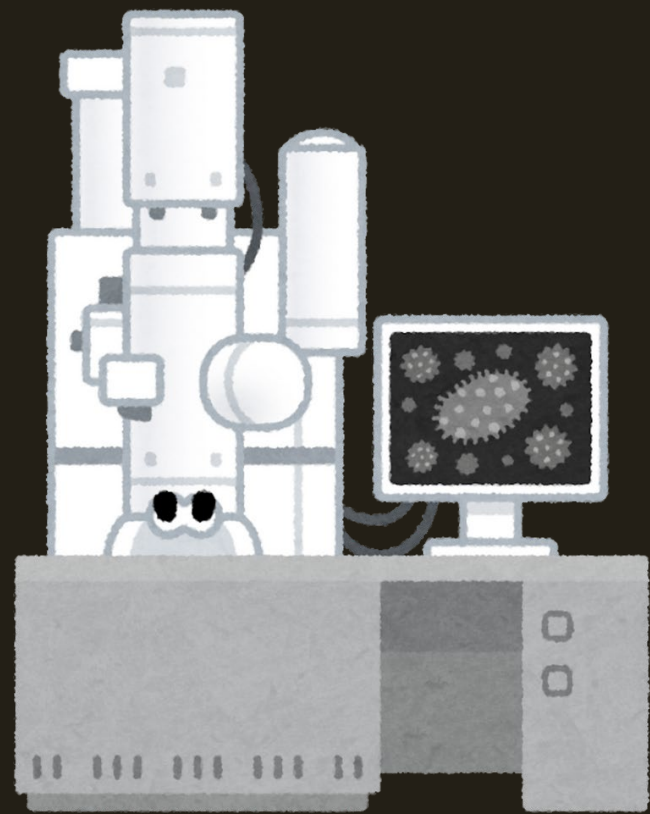
# Agenda



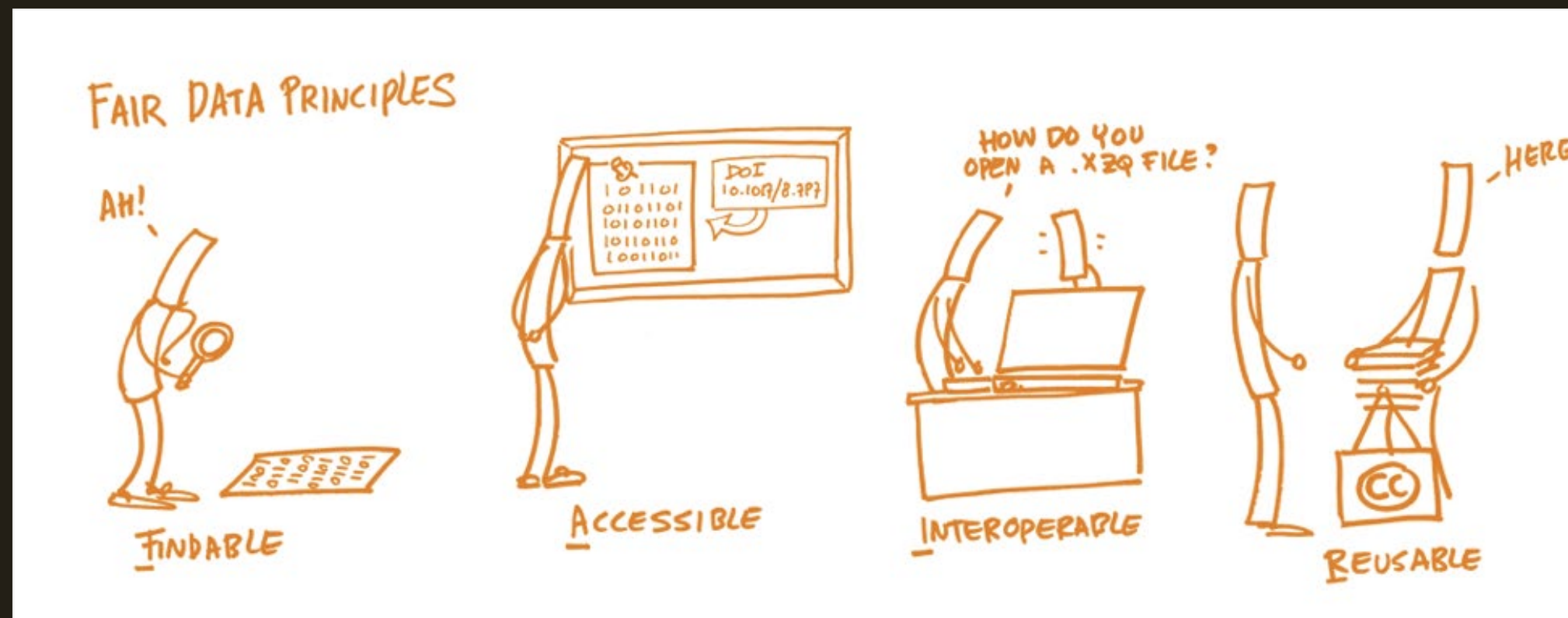
1. Qual è il problema nella **gestione dei dati** ?
2. Cosa significa progettare in ottica di **FAIRby-design**
3. **Cosa cambia** concretamente nel lavoro quotidiano
4. **Sfide** a frontate

03

# Dati preziosi, ma fragili



# Cosa significa FAIR-by-design



<https://book.fosteropenscience.eu/>

- NON È SOLO “METTERE ONLINE” UN FILE: SERVE CONTESTO E STRUTTURA
- **BY-DESIGN** SIGNIFICA PENSARE FAIR FIN DA QUANDO NASCONO I DATI IN LABORATORIO

# Un laboratorio reale, con limiti reali



- Laboratorio **SEM**, CNR-IMM di Catania
- Strumentazione **legacy**, **PC isolato**, software **proprietario**
- Immagini salvate in **locale**, informazioni su campioni ed esperimenti **sparse** tra Excel, quaderni, email e la **testa del ricercatore**

- **Pipeline FAIR**
- Dati e metadati **strutturati** secondo **standard** internazionali
- Pubblicazione in repository dedicati **Open**

# Workflow Outline



Pianificare  
“Come possiamo  
essere FAIR”



- Immagini TIFF
- Metadati esperimento
- Informazioni strumento
- Operatore

- DOI
- FAIR dataset

# Ciclo di vita del dato

---

1. **SCATTO** : il microscopio genera immagini e parametri tecnici
2. **ESTRAZIONE AUTOMATICA** : il sistema recupera dai file le informazioni “nascoste”, **metadati** dello strumento
3. **ARRICCHIMENTO** : il ricercatore aggiunge i dati sul **campione** , **esperimento** e **progetto**
4. **STRUTTURAZIONE** : tutto viene organizzato in una scheda dati **standard** , linguaggio **NeXus**
5. **CONDIVISIONE E RIUSO** : il file è pronto per repositories e ambienti di **open data**



# Un linguaggio comune per far parlare i dati

*NeXus*

- **Struttura standard comune**
- **Racconta un esperimento leggibile da persone e macchine**
- **“Scheda standard” specifica per la microscopia: NXem**

In questo modo i dati SEM possono entrare in ecosistemi di open data, invece di restare forme artigianali

# Per visualizzare

```
Tag: CZ_SEM, Type: <class 'dict'>, Value: {'': (0, 0, 0, 7.352941e-08, 5000.0, 6, 3000.0, 2.32, 2e-07, 0.003023319, 1, 7.352941e-08, 5000.0, 6, 3000.0, 2.32, 2e-07, 0.003023319, 2, 7.352941e-08, 5000.0, 6, 3000.0, 2.32, 2e-07, 0.003023319, 3, 7.352941e-08, 5000.0, 6, 3000.0, 2.32, 2e-07, 0.003023319, 0)}
```

```
"CZ_SEM": {  
  "Pixel Size": 7.352941e-08,  
  "Magnification": 5000.0,  
  "Output Device Index": 6,  
  "EHT (Electron High Tension)": 3000.0,  
  "Filament Current": 2.32,  
  "Probe Current": 2e-07,  
  "Working Distance": 0.003023319,  
  "Max Scan Speed": 1  
},
```



- event\_1
  - end\_time
  - identifier\_sample
  - image\_1
    - image\_2d **NX**
    - instrument
    - optical\_system**
      - magnification
      - probe\_current
      - working\_distance
    - image\_2
    - image\_3
    - image\_4
    - start\_time
  - instrument
- user**
  - address
  - affiliation
  - email
  - name
- sample**
  - atom\_types
  - identifier\_sample
  - is\_simulation
  - name
  - physical\_form
  - preparation\_date
  - start\_time

10

# Cosa cambia nella pratica di laboratorio

PRIMA	DOPO
<ul style="list-style-type: none"><li>• Nomi di file spesso <b>inconsistenti</b></li><li>• Metadati tecnici <b>copiati a mano</b> in report o quaderni</li><li>• Informazioni su campioni ed esperimenti <b>sparse</b> tra Excel, Word, “testa”</li><li>• Ogni operatore ha il <b>suo modo</b> di documentare</li></ul>	<ul style="list-style-type: none"><li>• Nomi di file legati a un <b>identificatore</b> di esperimento/campione</li><li>• Metadati tecnici <b>estratti automaticamente</b> dalle immagini</li><li>• Dati di campioni ed esperimenti raccolti in un <b>modulo unico e strutturato</b></li><li>• File finale NeXus generato in automatico: il laboratorio <b>compila moduli e non deve conoscere lo standard</b></li></ul>

# Gli ostacoli incontrati

11



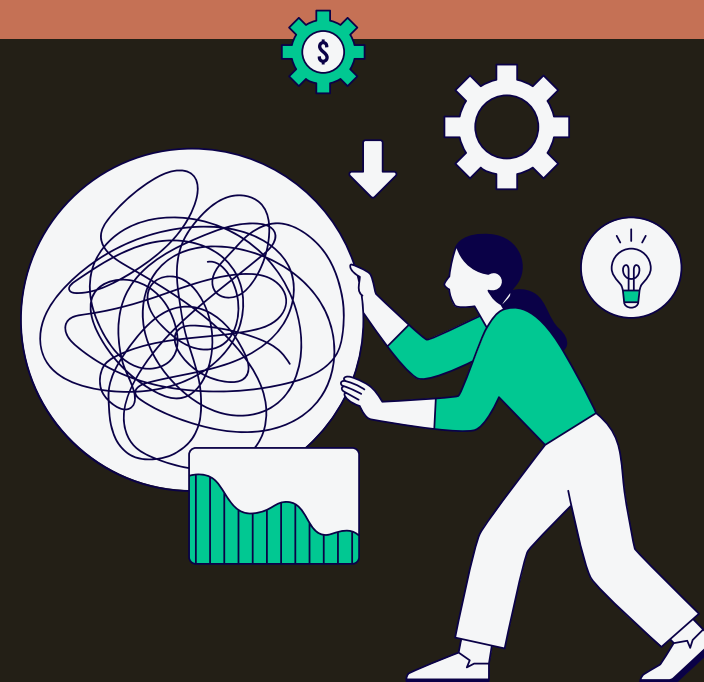
ADOZIONE CULTURALE

COMPATIBILITÀ  
LEGACY



SOSTENIBILITÀ

INCENTIVI  
ISTITUZIONALI



# Dal laboratorio al bene comune digitale



## TRASPARENZA E FIDUCIA NELLA RICERCA

Dati ben **documentati** e **standardizzati** sono più facili da verificare, combinare e riusare



## COLLABORAZIONE

STANDARD APERTI E STRUMENTI RIUSABILI PERMETTONO ANCHE A LABORATORI PICCOLI O PERIFERICI CONTRIBUIRE ALLO STESSO ECOSISTEMA DI DATI



## DIALOGO

DATI STRUTTURATI SU MATERIALI, AMBIENTE, SALUTE ETC. POSSONO IN FUTURO DIALOGARE CON ALTRI DATASET E CONTRIBUIRE A DECISIONI PUBBLICHE PIÙ INFORMATE.

13

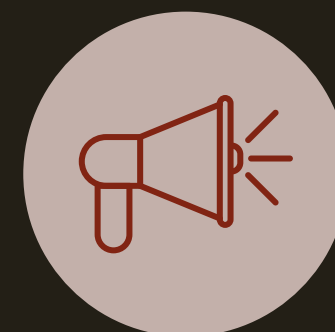
# Tre idee da portare a casa



IL MODO IN CUI NASCONO I DATI  
DECIDE GIÀ QUANTO POTRANNO  
ESSERE RIUSATI



GLI STANDAR SONO  
UN'INFRASTRUTTURA INVISIBILE  
CHE PERMETTE AI DATI DI  
PARLARSÌ OLTRE IL SINGOLO  
PROGETTO



IL FAIR-BY-DESIGN PUÒ  
FUNZIONARE ANCHE IN CONTESTI  
NORMALI, RIDUCENDO IL CARICO  
SULLE PERSONE E AUMENTANDO IL  
VALORE DEI DATI

14

# Grazie per l'attenzione!

---

Ci sono domande?

Contatti : [lidonnilucia@gmail.com](mailto:lidonnilucia@gmail.com)