# L'edizione pilota del «Master in Data Management and Curation (MDMC)»

Mariarita de Luca
Area Science Park

*presentato al Secondo convegno nazionale – Laboratori nazionali di Frascati dell'INFN, 27-28 novembre 2024*

DOI: 10.15161/oar.it/xxxx

# Summary

- Why MDMC?
- Supporting Institutions and projects
- Schedule and details of MDMC

# WHY MDMC?

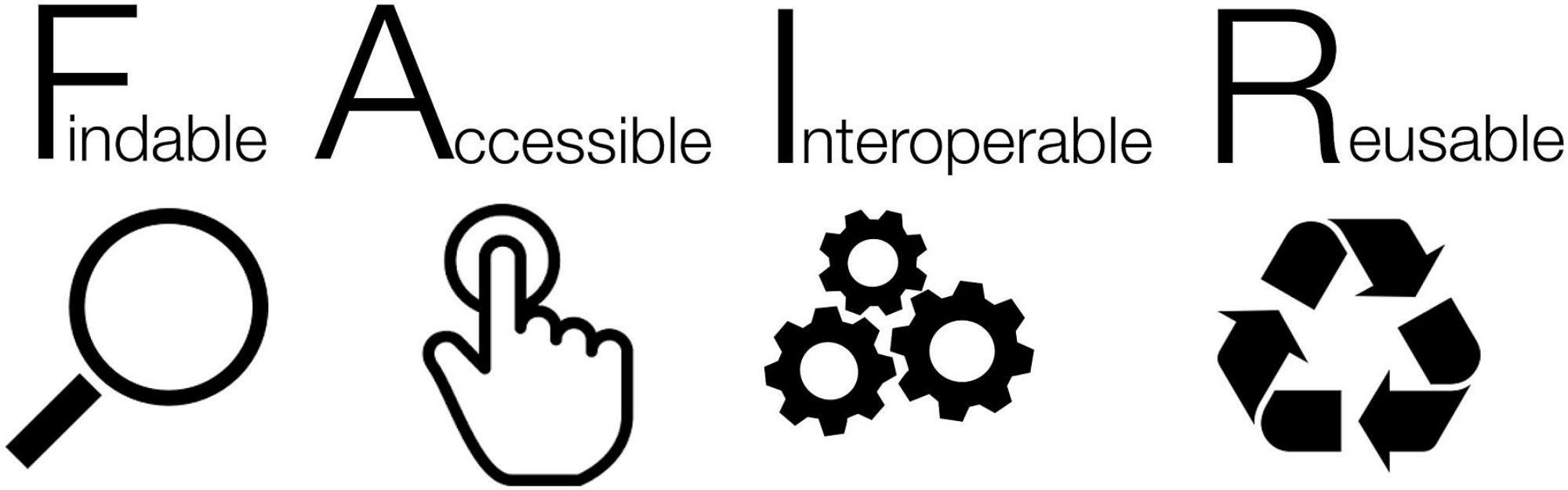
# IMPORTANCE OF DATA MANAGEMENT IN SCIENCE


- "*Research cannot flourish if data are not preserved and made accessible. All concerned must act accordingly*".

- "*Data management should be woven into every course in science, as one of the foundations of knowledge*"

'Editorial: Data's Shameful Neglect'
(10 September 2009) in Nature 461, p. 145, doi:10.1038/461145a.

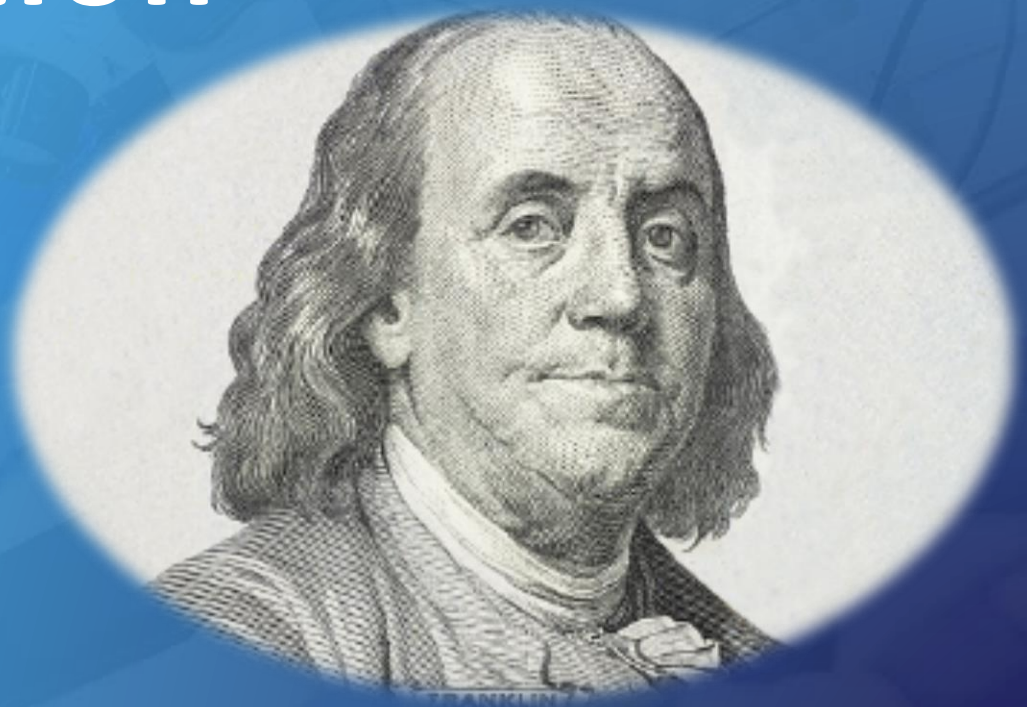
# FAIR PRINCIPLES






Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

“If you think education is expensive, Try ignorance”



Benjamin Franklin

“If you think ~~education~~ data management is expensive, Try ~~ignorance~~ ignore it”

Stefano Cozzini

# COST OF NOT HAVING FAIR DATA

| AREA | INDICATORS | COST (Million EUR per year) |
|---|---|---|
| Impact on research activities | 1. Time spent | 4500 |
| | 2. Cost of storage | 5300 |
| | 3. License costs | 630 |
| Impact on opportunities for further research | 4. Research retraction | 4,4 |
| | 5. Double funding | 25 |
| | 6. Cross-fertilization | N.A. |
| Impact on innovation | 7. Potential economic growth (as % of GDP) | N.A. |
| | | 10459,4 |

European Commission, Directorate-General for Research and Innovation, *Cost-benefit analysis for FAIR research data – Cost of not having FAIR research data*, Publications Office, 2018, https://data.europa.eu/doi/10.2777/02999

# ROLES AND RESPONSABILITIES IN RESEARCH DATA MANAGEMENT

Research Data management responsibilities relies first on PI or researcher who created or collected the data.

They extend than to other people involved in the research process that play a role in ensuring quality data stewardship.

It is fundamental that the roles and responsibilities of data management are clearly defined and assigned, rather than assumed.

AREA SCIENCE PARK

# EOSC SYMPOSIUM 2024

**Scientific keynote:**

- FAIR, what else? (Claudia Draxl)
- FAIR for AI and AI for FAIR (Julia Lane)

**Take away message:**

We need more data engineers and data curators

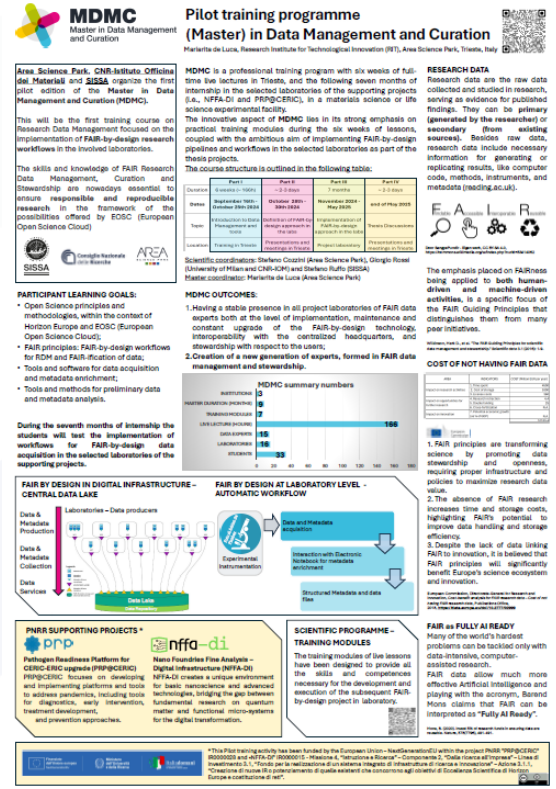

Picture from «The Turing Way» 10.5281/zenodo.3332807

**FAIR (2016)**

**Findable Accessible Interoperable and Reusable**

**FAIR (2024)**

**Fully AI Ready**




To build the EOSC nodes federation,
Interoperability is the key

# SUPPORTING INSTITUTIONS AND PROJECTS

MDMC
Master in Data Management and Curation

# PILOT TRAINING COURSE MASTER IN DATA MANAGEMENT AND CURATION

Area Science Park, CNR-Istituto Officina dei Materiali and SISSA organize the first pilot edition of the **Master in Data Management and Curation (MDMC).**

The skills and knowledge of **FAIR** Research Data Management, Curation and Stewardship are nowadays essential to ensure responsible and reproducible research.

The structure and program of the master are entirely innovative on both **national and international** levels, with a specific focus on the implementation of **FAIR-by-design processes** in the involved laboratories.

**Scientific commitee:**
Prof. Giorgio Rossi *(University of Milan and CNR-IOM)*
Dr. Stefano Cozzini *(Area Science Park)*
Prof. Stefano Ruffo and prof. Eugenio Piasini *(SISSA)*

**MDMC Coordinator:**
Mariarita de Luca *(Area Science Park)*

# PNRR SUPPORTING PROJECTS*

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

## MATERIALS SCIENCE PATH

### NANO FOUNDRIES FINE ANALYSIS – DIGITAL INFRASTRUCTURE (NFFA-DI)

NFFA-DI creates a unique environment for basic nanoscience and advanced technologies, bridging the gap between fundamental research on quantum matter and functional micro- systems for the digital transformation.

## LIFE SCIENCE PATH

### PATHOGEN READINESS PLATFORM FOR CERIC-ERIC UPGRADE (PRP@CERIC)

PRP@CERIC focuses on developing and implementing platforms and tools to address pandemics, including tools for diagnostics, early intervention, treatment development, and prevention approaches.

AREA SCIENCE PARK

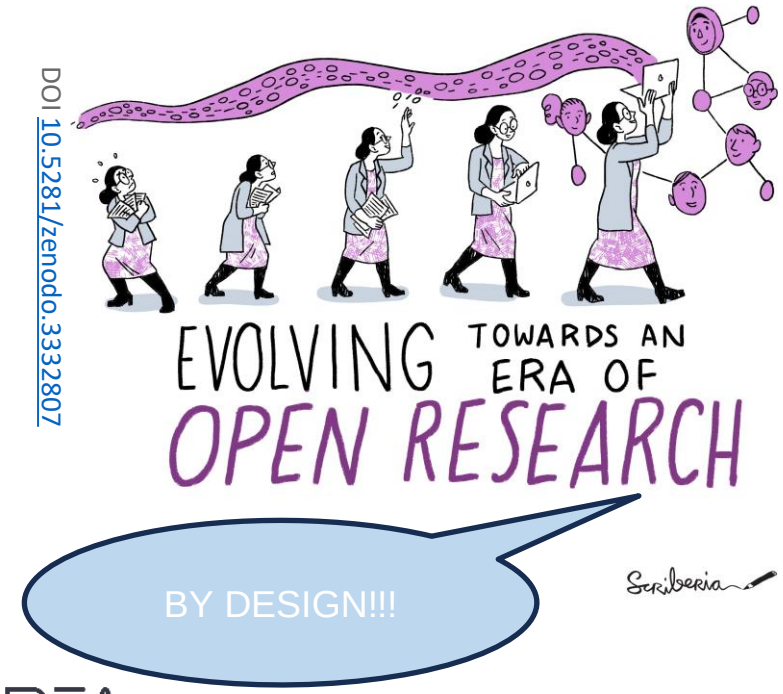
* Funded by the European Union through the National Recovery and Resilience Plan (NRRP), part of Next Generation EU, as part of Mission 4 “Education and Research”, Component 2 “From Research to Business”, Investment Line 3.1 “Fund for the creation of an integrated system of research and innovation infrastructures”.

# SCHEDULE AND DETAILS OF MDMC

# FAIR-BY-DESIGN

The next generation of researchers should be trained and supported to do transparent and reproducible science from day one — that is, to be “open science natives”. Open By Design — Open By Design at Stanford (dsi-cores.github.io)

**FAIR data management by design** is an approach that ensures the application of FAIR principles from the inception of a project, making the data Findable, Accessible, Interoperable, and Reusable (FAIR) throughout all stages of research data lifecycle:




- plan and design
- collection and creation
- analysis and collaboration
- evaluation and archive
- sharing and dissemination
- publication and reuse


Harvard Biomedical Research Data Lifecycle, 10.5281/zenodo.8075933

AREA SCIENCE PARK

Book: Open Science by Design: Realizing a Vision for 21st Century Research (2018) doi 10.17226/25116 ©

# TIMELINE OF MDMC

| | Part I | Part II | Part III | Part IV |
|---|---|---|---|---|
| Duration | 6 weeks (~ 166h) | ~ 2-3 days | 7 months | ~ 2-3 days |
| **Dates** | **September 16th - October 25th 2024** | **October 28th - 30th 2024** | **November 2024 - May 2025** | **end of May 2025** |
| Topic | Introduction to Data Management and tools | Definition of FAIR-by-design approach in the labs | Implementation of FAIR-by-design approach in the labs | Thesis Discussions |
| Location | Training in Trieste | Presentations and meetings in Trieste | OU and labs | Presentations and meetings in Trieste |

AREA SCIENCE PARK

Google Calendar - Week of September 16, 2024

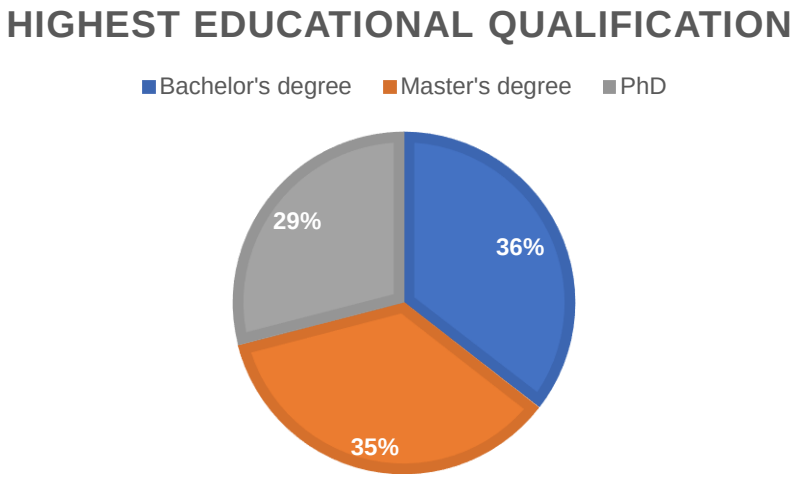
# PARTICIPANTS

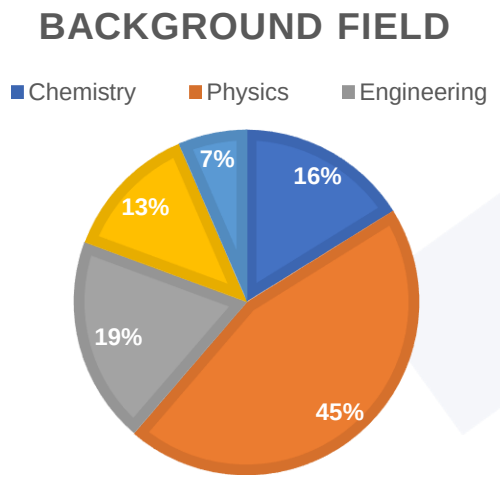
In this pilot version, during the academic year 2024/2025,

we have **31 participants** selected by the various operational units of the supporting projects.

**Two categories of students:**

- **23 Full participants:** Students enrolled in the entire training program (six weeks of live lecturers and seven months of laboratory internship);
- **8 Auditors:** Students enrolled only in the six weeks of live lectures, 5 of which decided to continue after the six weeks of lesson with a small project dedicated to FAIR/Open Science topics.



# PARTICIPANT LEARNING GOALS

- Open Science principles and practices, within the context of Horizon Europe programme and EOSC;
- FAIR principles: data FAIR-by-design approach and FAIR-ification of data processes;
- Basic knowledge of data infrastructure and cloud data infrastructure
- Tools and software for data acquisition and metadata enrichment;
- Tools and methods for data and metadata analysis.


Foto di Dirk Wouters da Pixabay

AREA SCIENCE PARK

# TRAINING MODULES

**List of the seven training modules:**

- Introduction to Open Science (OS)
- Scientific Programming Environment (SPE)
- Cloud Data Environment (CDE)
- Python for data management (PY)
- Data Infrastructures (DI)
- Data Management Tools (DMT)
- Introduction to Statistical Data Analysis and Machine Learning (SDA&ML)

**All training materials (slides and additional materials) is openly available on the Zenodo Community of MDMC :**
**Master Data Management and Curation (MDMC) (zenodo.org)**

# MDMC OUTCOMES

- Having a stable presence in all RI laboratories of FAIR data experts both at the level of implementation, maintenance and constant upgrade of the FAIR-by-design technology, interoperability with the centralized headquarters, and stewardship with respect to the users;

- Creation of a new generation of experts, formed in FAIR data management and stewardship.

Training

PhD courses

Pre-PhD Fellowships

Professional Master Courses

Master Courses

Visiting Student program (ViS)

# Master in Data Management and Curation (MDMC)

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

## PILOT TRAINING COURSE IN DATA MANAGEMENT AND CURATION

Area Science Park, CNR-Istituto Officina dei Materiali and SISSA organize the first edition of the Master in Data Management and Curation (MDMC).

In the digital and data-driven paradigm promoted by Open Science, data is at the core of the scientific process and its production grows at ever-increasing rates. The skills and knowledge of Scientific Data Management and Curation are nowadays essential to ensure responsible and reproducible research in the framework of the possibilities offered by the European Open Science Cloud (EOSC)
Having EOSC compliant Research Infrastructures and **FAIR-by-design Research Data Management** is among the objectives of the two supporting projects:

- NFFA-DI (Nano Foundries and Fine Analysis Digital Infrastructure)
- PRP@CERIC (Pathogen Readiness Platform for CERIC-ERIC Upgrade)

nffa-di Piano Nazionale di Ripresa e Resilienza

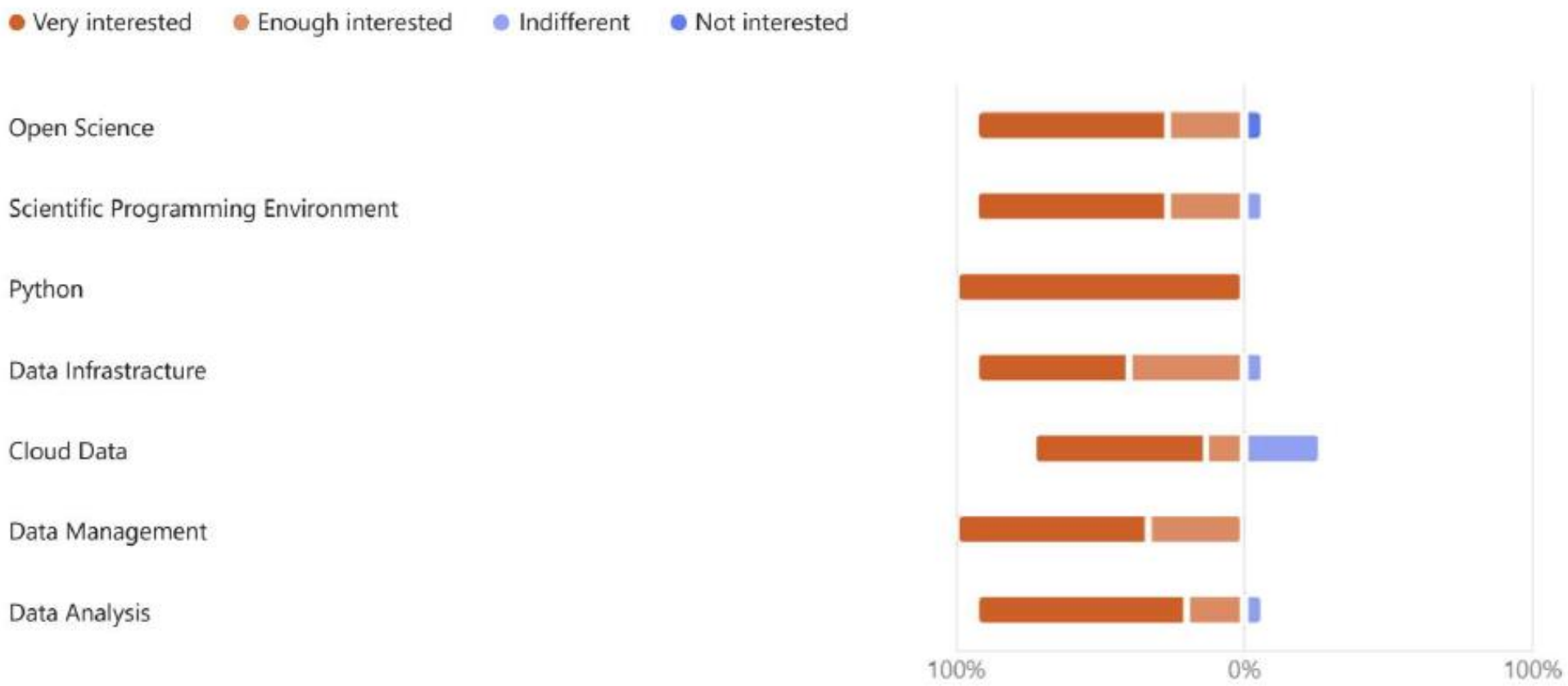
prp

Master in Data Management and Curation (MDMC) (sissa.it)

# FIRST EVALUATION

MDMC
Master in Data Management and Curation

Define your level of interest in the topics covered




AREA SCIENCE PARK

A heartfelt thanks to **Stefano Cozzini and Federica Bazzocchi** without whose contribution it would not be possible to realize MDMC.

Another special thank to all the other MDMC lecturers who are now following the students in the FAIR-by-design activities in the laboratories:
**Ruggero Lot, Marco Prenassi, Tommaso Rodani and Matteo Biagetti.**

# Thank you for your attention