

AREA SCIENCE PARK

Stefano Cozzini, Mariarita de Luca

Trieste, 22.05.2024

INFO DAY - Master in Data Management and Data Curation (MDMC)

OUTLINE

- ✓ Why do we need MDMC?
- ✓ FAIR Data Management
- ✓ Digital Infrastructures' supporting projects
- ✓ Schedule and details of MDMC
- ✓ Open positions in Area
- ✓ Q&A



Foto di [Brian Penny](#) da [Pixabay](#)

The background is a blue-tinted collage of various electronic components and tools. On the right side, there is a large, semi-transparent image of a microscope. Scattered throughout the background are various electronic parts, including resistors, capacitors, integrated circuits, and some cables. The overall aesthetic is technical and scientific.

WHY DO WE NEED MDMC?

SQUIRRELS...

They eat too much at once

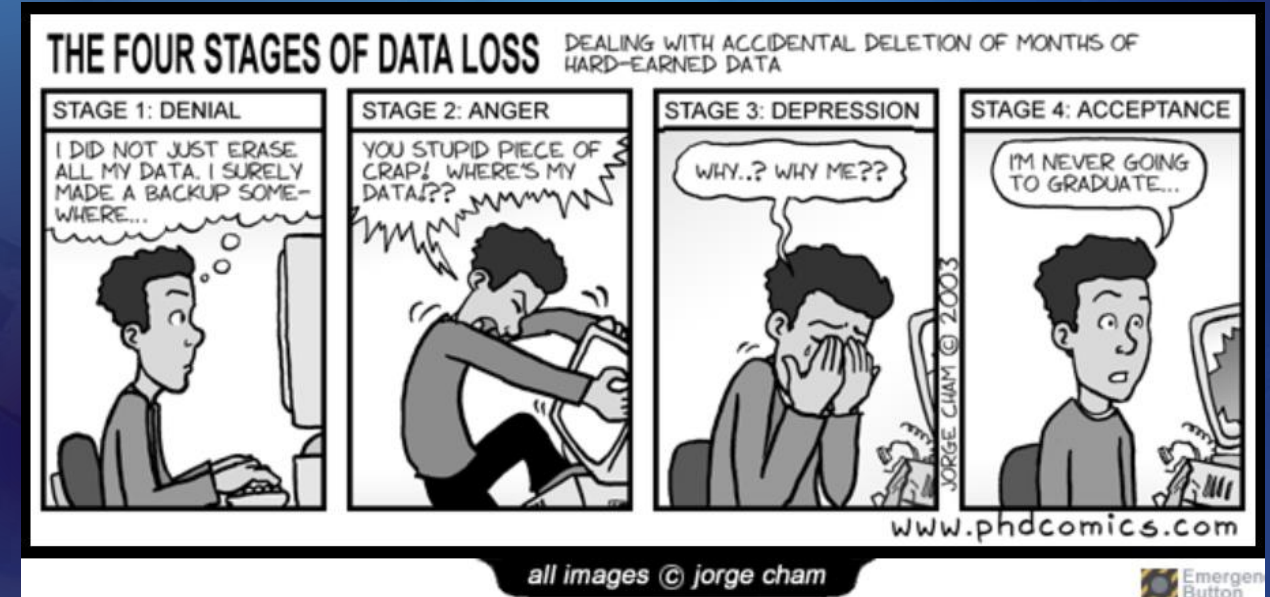


They forget where they store their food

SCIENTIST ARE LIKE SQUIRRELS

[PHD Comics: Stages of Data Loss](#)

They get so many data at once



They may forget where they store data

[QS World University Rankings by Subject 2015 - challenges and developments - QS](#)

HOW DO YOU SAVE YOUR DATA ?

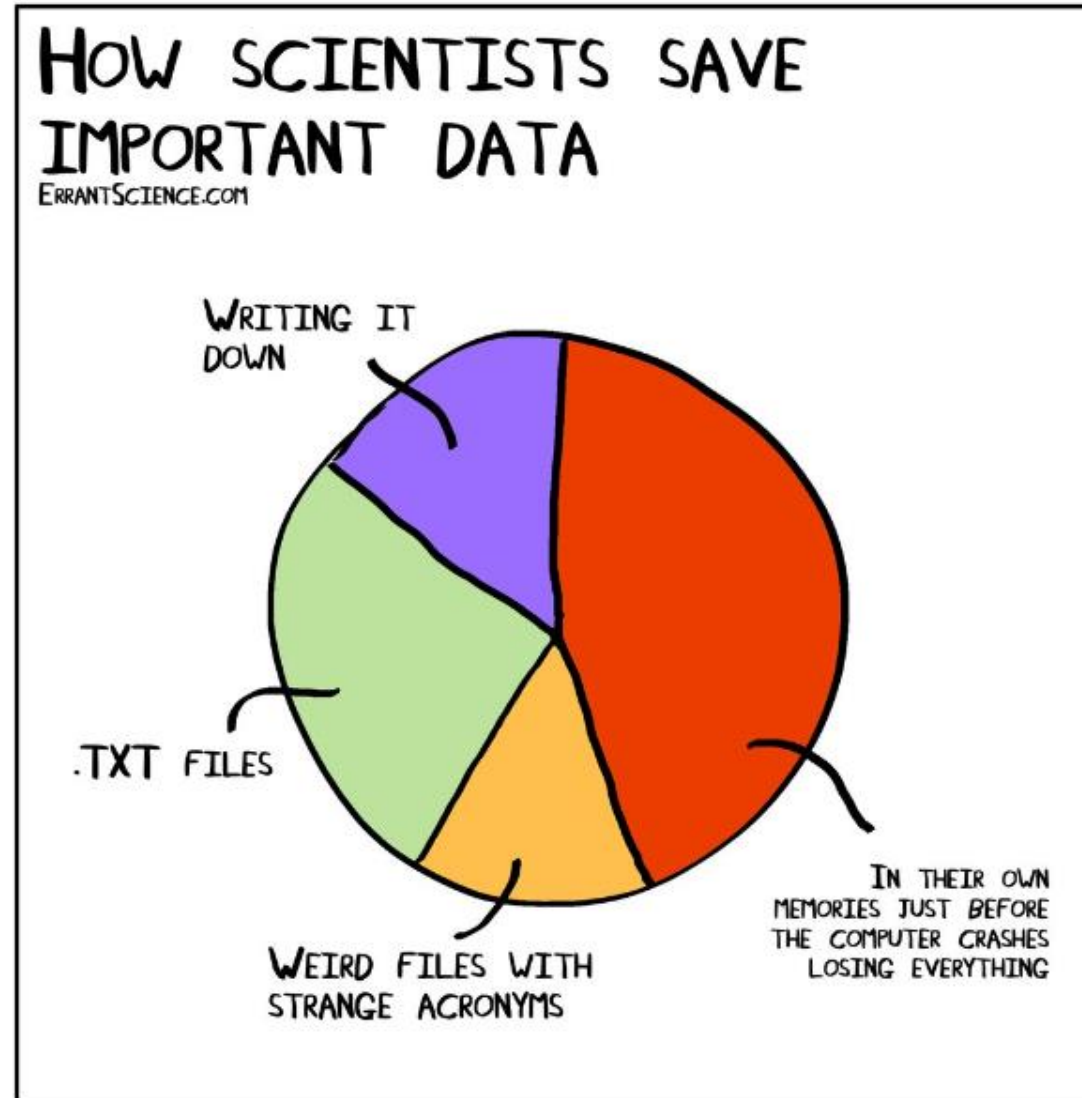
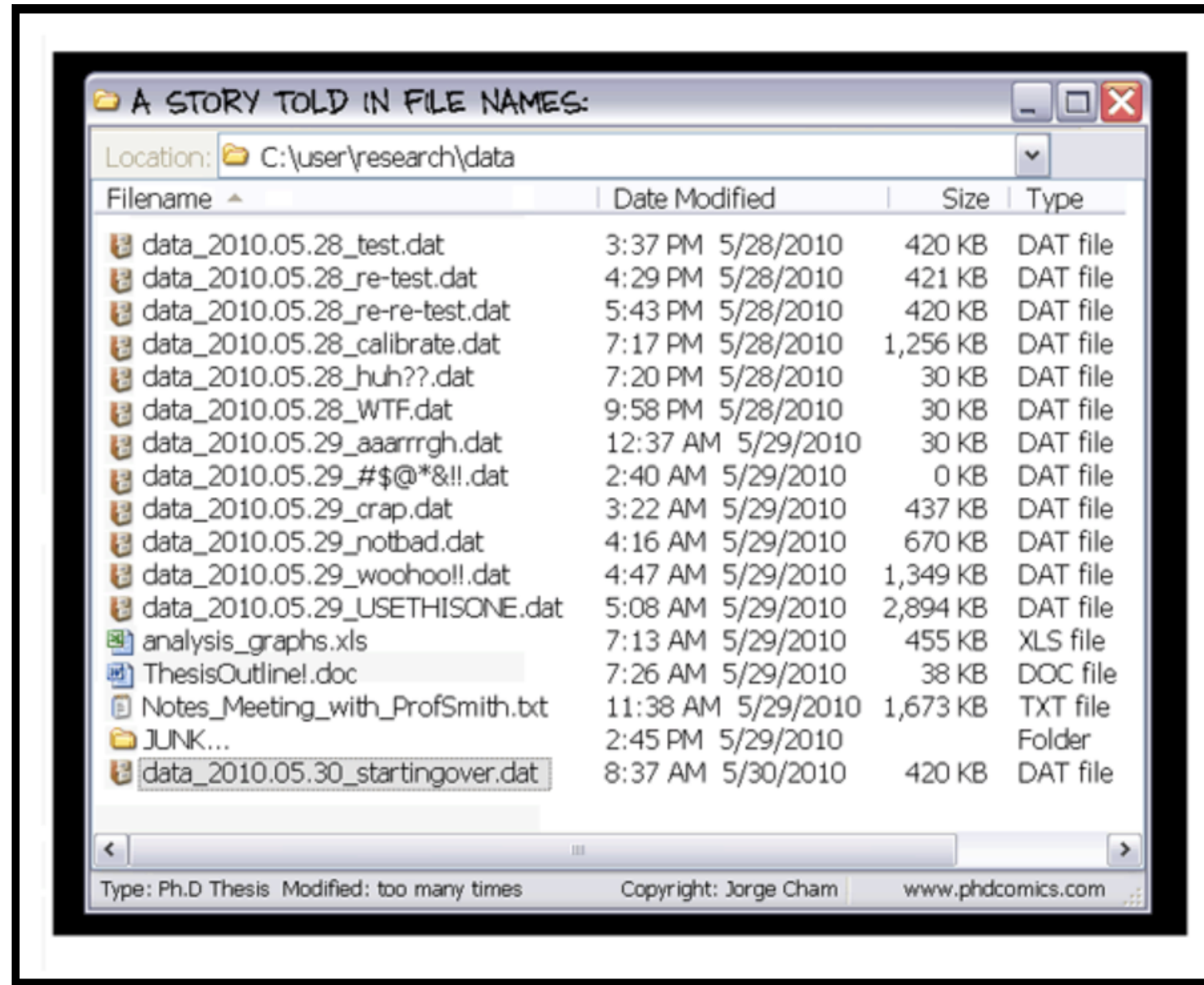


Image by [Errant Science](https://errantscience.com)

HOW DO YOU NAME YOUR DATA ?



[PHD Comics: A story in file names](http://www.phdcomics.com)

IMPORTANCE OF DATA MANAGEMENT IN SCIENCE

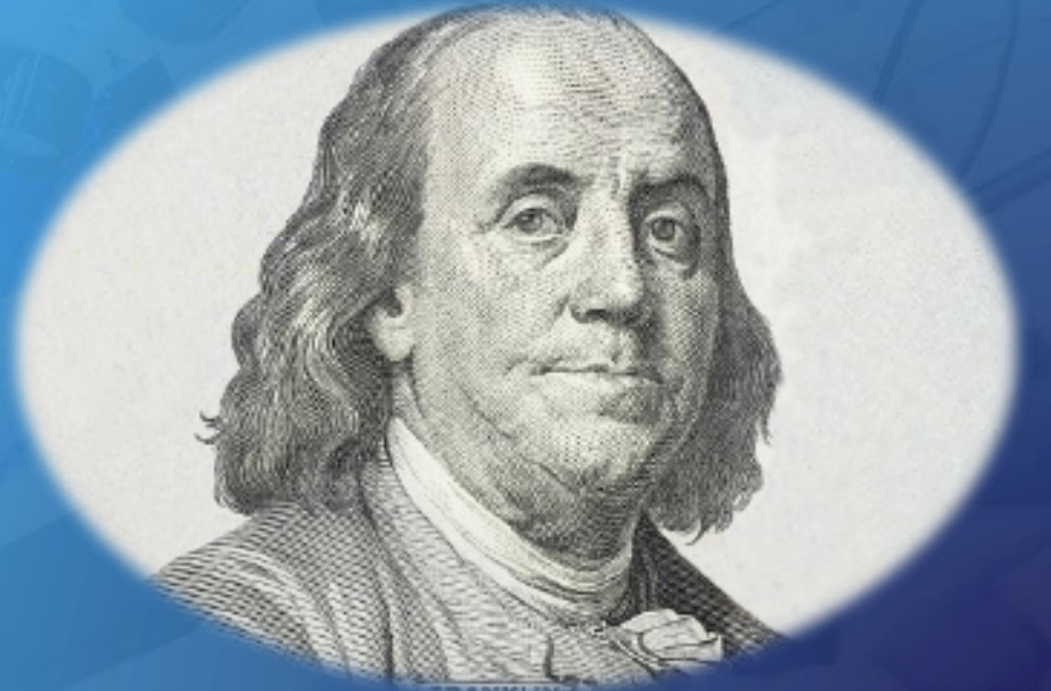


- *"Research cannot flourish if data are not preserved and made accessible. All concerned must act accordingly".*
- *"Data management should be woven into every course in science, as one of the foundations of knowledge"*

'Editorial: Data's Shameful Neglect'

(10 September 2009) in Nature 461, p. 145, doi:10.1038/461145a.

**“If you think education
is expensive,
Try ignorance”**



Benjamin Franklin

**“If you think ~~education~~
data management is expensive,
Try ~~ignorance~~ without it”**

subject: External hard disk lost
Organization: S.I.S.S.A.
Date: Wed, 4 Jul 2018 13:54:48 +0200
From: Students' Secretariat <XXXX@sissa.it>
To: SISSA Users;;

An external hard disk has been lost, most probably on the 4th floor, black, in a white box.

It contains **a lot of work data of a SISSA PhD student.**

If you happen to find it, please leave it at the reception desk or at the students' secretariat. Alternatively you can leave it in the Students' Secretariat mailbox in the lower level.

Marconi: scratch is almost full – quota imposed

16 May 2024

Dear Marconi Users,

we inform you that the scratch space has reached the occupation of more than 87% today. This may cause malfunctions to the filesystems. To avoid reaching a 100% occupancy, we temporarily set a quota of 20 TB on the scratch folder of each user. We encourage you to clean your scratch folders by removing useless data or by moving data to work and dres spaces. We will inform you as soon as normal occupancy will be restored and the quota removed.

Best regards,
HPC User Support @ CINECA

31/10/94: ALL SG1 + 2/10s on EUCLID.

MARK 1

12/9/94: EUCLID. →

MARK 2.

FAIR DATA MANAGEMENT



RESEARCH DATA

Research data are the **raw materials** collected, processed and studied in the undertaking of research. They are the evidential basis that substantiates published research findings.

They may be **primary data** generated or collected by the researcher, or **secondary data** collected from existing sources and processed as part of the research activity.

In addition to the 'raw' data, **research data include information about the means necessary to generate data or replicate results**, such as computer code, experimental methods and instruments used, and essential interpretive and contextual information, e.g. specifications of variables.

[Research data defined \(reading.ac.uk\)](https://www.reading.ac.uk/research-data-defined)

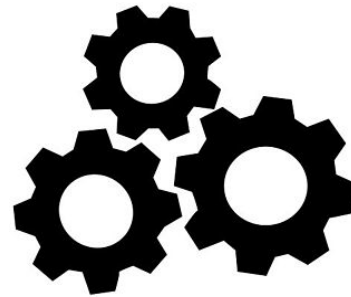
F_{indable}



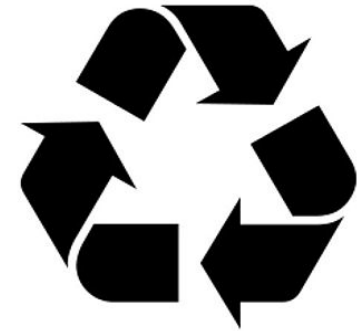
A_{ccessible}



I_{nteroperable}



R_{eusable}



Door SangyaPundir - Eigen werk, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=53414062>

FAIR PRINCIPLES

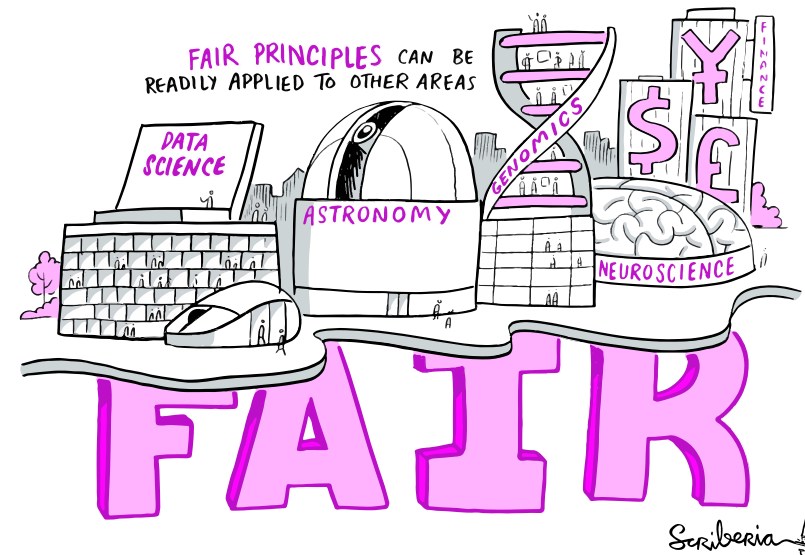
Good data management is not a goal in itself, but rather **is the key conduit leading to knowledge discovery and innovation**, and to subsequent data and knowledge integration and reuse by the community after the data publication process.

The emphasis placed on FAIRness being applied to **both human-driven and machine-driven activities**, is a specific focus of the FAIR Guiding Principles that distinguishes them from many peer initiatives.

These high-level FAIR Guiding Principles **precede implementation choices**, and do not suggest any specific technology, standard, or implementation-solution;

Moreover, the Principles are not, themselves, a standard or a specification.

They **act as a guide** to data publishers and stewards to assist them in evaluating whether their particular implementation choices **are rendering their digital research artefacts Findable, Accessible, Interoperable, and Reusable.**



FAIR PRINCIPLES APPLY DURING ALL THE PHASES OF THE RESEARCH DATA LIFECYCLE

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards



Harvard Biomedical Research Data Lifecycle, [10.5281/zenodo.8075933](https://doi.org/10.5281/zenodo.8075933)

COST OF NOT HAVING FAIR DATA



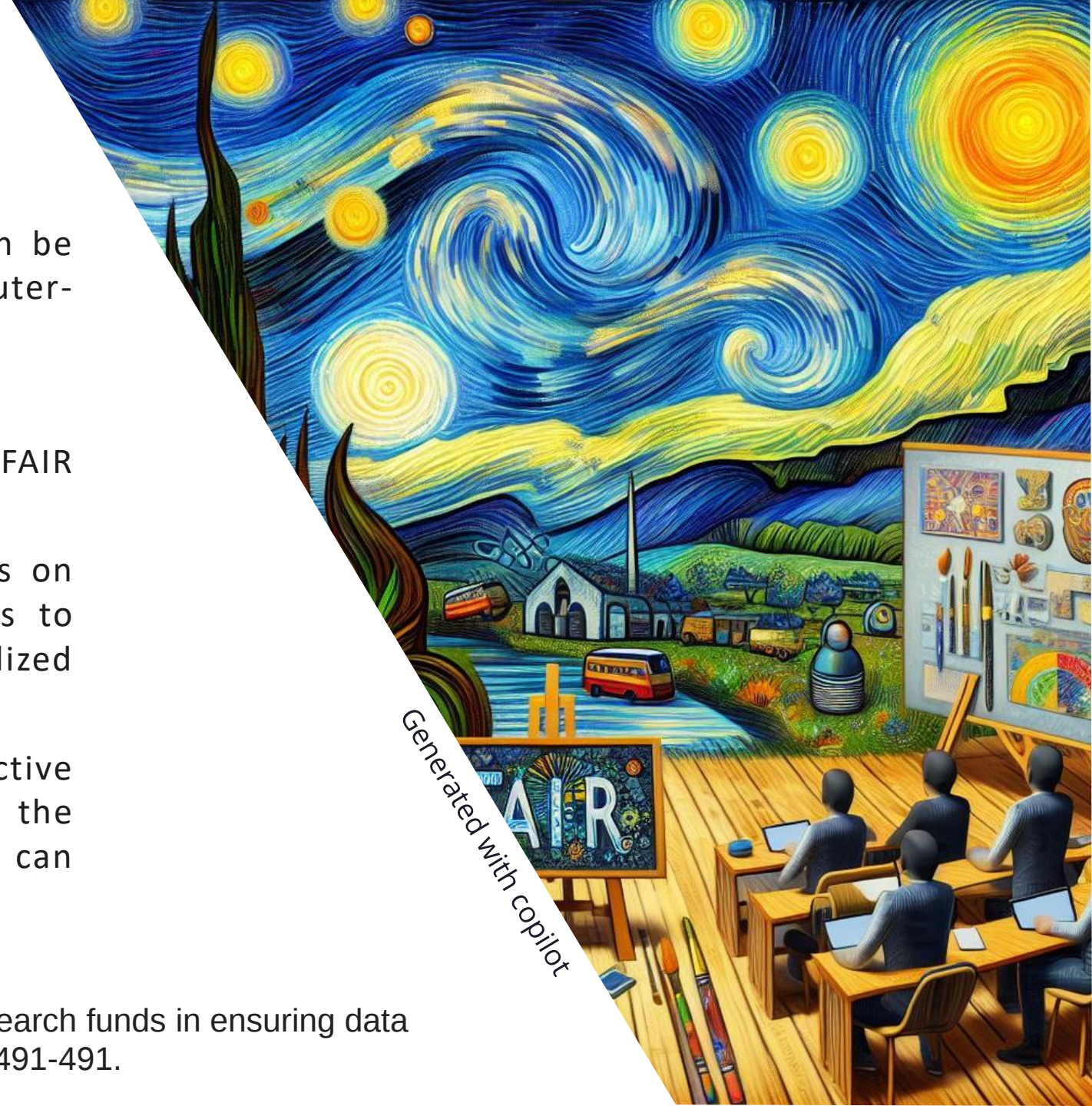
AREA	INDICATORS	COST (Million EUR per year)
Impact on research activities	1. Time spent	4500
	2. Cost of storage	5300
	3. License costs	360
Impact on opportunities for further research	4. Research retraction	4,4
	5. Double funding	25
	6. Cross-fertilization	N.A.
Impact on innovation	7. Potential economic growth (as % of GDP)	N.A.
		10189,4

FAIR AS FOR AI READY

Many of the world's hardest problems can be tackled only with data-intensive, computer-assisted research.

From this perspective, the investments in FAIR data management offer a dual advantage:

- They prevent the waste of public funds on redundant research and provide access to properly processed data that can be utilized by artificial intelligence algorithms.
- FAIR data allow much more effective artificial intelligence and playing with the acronym, **Barend Mons** claims that FAIR can be interpreted as “**Fully AI Ready**”.



The background is a blue-tinted collage of various electronic components and tools. On the right side, there is a large, semi-transparent image of a microscope. The rest of the background is filled with smaller, overlapping images of circuit boards, wires, and other electronic parts.

DIGITAL INFRASTRUCTURES OF SUPPORTING PROJECTS

PNRR SUPPORTING PROJECTS*

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.



Piano Nazionale di Ripresa e Resilienza

LIFE SCIENCE PATH

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

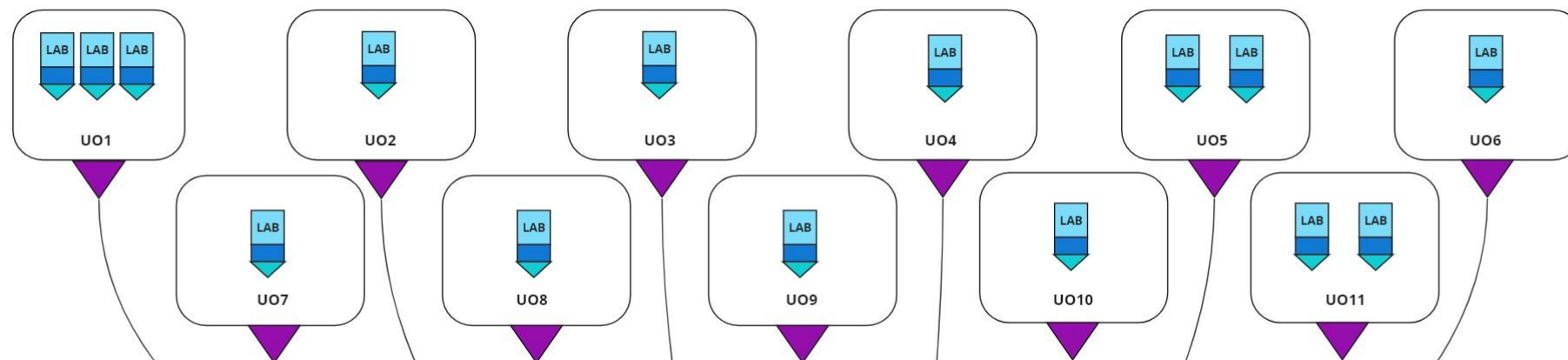
PRP@CERIC focuses on developing and implementing platforms and tools to address pandemics, including tools for diagnostics, early intervention, treatment development, and prevention approaches.



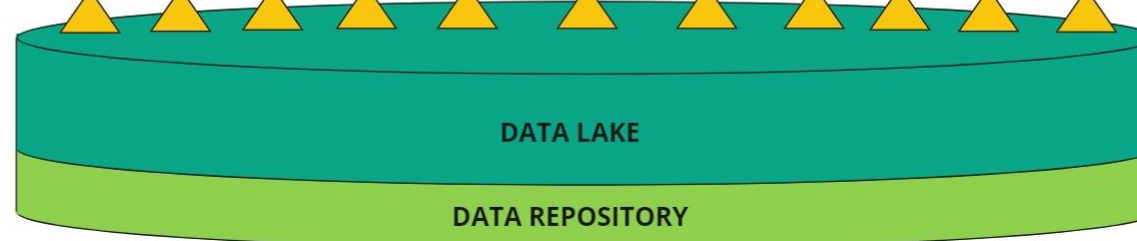
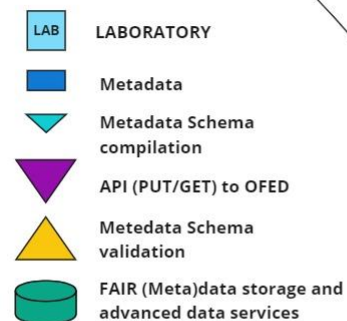
* Funded by the European Union through the National Recovery and Resilience Plan (NRRP), part of Next Generation EU, as part of Mission 4 “Education and Research”, Component 2 “From Research to Business”, Investment Line 3.1 “Fund for the creation of an integrated system of research and innovation infrastructures”.

CENTRALIZED DIGITAL INFRASTRUCTURE – FAIR BY DESIGN

Data producers



Legenda



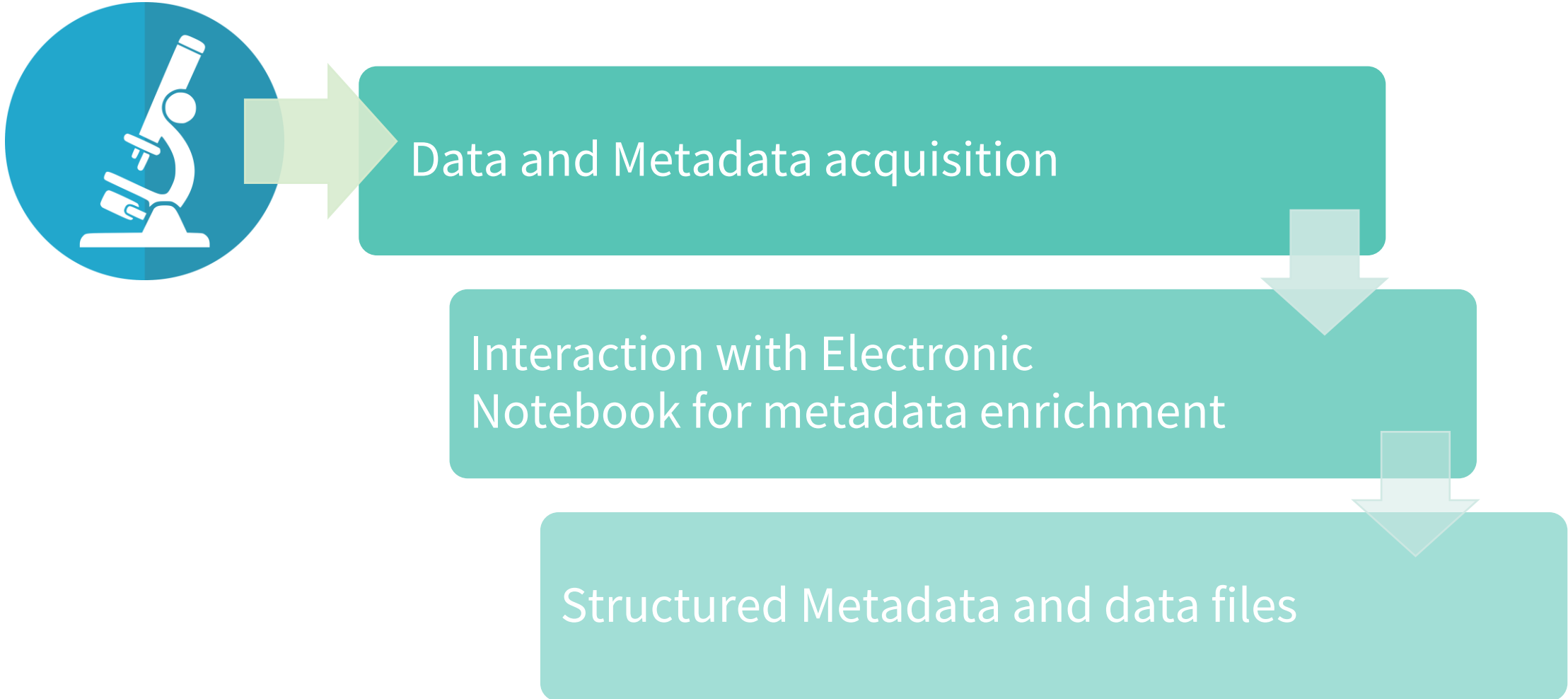
Data &
Metadata
Production

Data &
Metadata
Collection

Data
Services

AT LABORATORY LEVEL – AUTOMATIC WORKFLOW

Foto di Adrian da Pixabay



SCHEDULE AND DETAILS OF MDMC



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 FAIR Research 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.



Consiglio Nazionale
delle Ricerche



SISSA

Scuola
Internazionale
Superiore di
Studi Avanzati



MASTER IN DATA MANAGEMENT AND CURATION (MDMC)

Learning goals:

- Open Science principles and methodologies, within the context of Horizon Europe Framework programme and EOSC (European Open Science Cloud);
- FAIR principles: Data FAIR-by-design and FAIRification of data;
- Tools and software for data acquisition and metadata enrichment
- Tools and methods for preliminary data and metadata analysis

TIMELINE OF MDMC

	Part I	Part II	Part III	Part IV
Duration	6 weeks (~ 160h)	~ 1 week	7 month	~ 1 week
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	UO and labs	Presentations and meetings in Trieste

TARGET PARTICIPANTS

In this first edition participants are selected by the projects partners

- Students that hold at least bachelor's degree (laurea triennale or equivalent);
- Students still enrolled in a university master's course (laurea magistrale or equivalent) , in science, engineering, or informatics are also eligible.
- Staff of supporting PNRR projects (researchers, technologists, PhD students)

PARTICIPANTS

- STUDENTS ENROLLED AS SISSA STUDENTS
- 70% OF LESSON ATTENDANCE IS MANDATORY



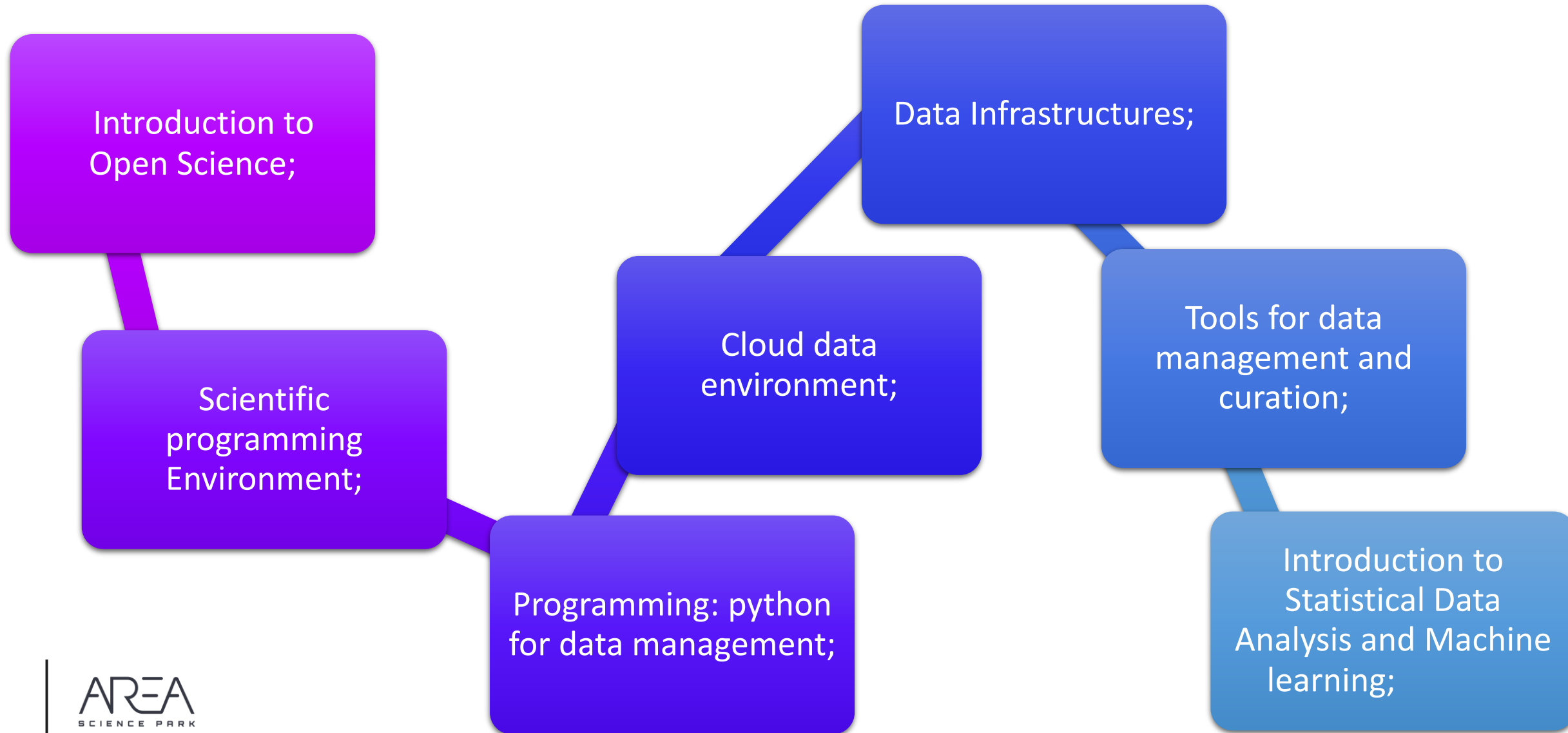
2 MONTH PROGRAMME (LESSON PARTICIPANTS)

A certificate of attendance with a statement of the topics covered in the training modules attended and the skills acquired will be released.

9 MONTHS PROGRAMME (FULL PROGRAMME PARTICIPANTS)

A certificate of attendance will be issued upon presentation of a thesis agreed with one or more professors of the course and the supervisor at the laboratory of origin of each participant.

TRAINING MODULES



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)



The background is a blue-toned collage of various scientific and technological items, including a microscope, test tubes, a beaker, a computer monitor, and a target symbol. The text is centered over this collage.

OPEN POSITIONS IN AREA SCIENCE PARK

AVAILABLE POSITIONS



AREA

- FOUR POSITIONS FOR AREA-ELETTRA STAFF
- TWO POSITIONS FOR **AREA** (1LADE, 1LAGE)
- ONE Position for **ICGEB**

- TWO POSITIONS FOR **CNR** (1 IOM, 1 IC)
- ONE POSITION FOR **UNISALENTO**
- ONE POSITION FOR **UNISALERNO**
- ONE POSITION FOR **UNINAPOLI**

OTHER PROJECT PARTNERS



AREA AVAILABLE POSITIONS

TWO OPEN POSITION WITH EXPRESSION
OF INTEREST AVAILABLE AT THE LINK:

[Master in Data Management and Curation:
training opportunities in Area Science Park
for two STEM graduates » Blog Archive »
Area Science Park](#)

**15000 EURO OF FLAT RATE
REIMBURSEMENT FOR SELECTED
PARTICIPANTS**

Other project partners (**CNR, UniMI e PoliMI**
have other position available)

For information send and email to mdmc@sissa.it



Piano Nazionale di Ripresa e Resilienza

DEADLINE MAY
31th

***AVAILABLE FOR FULL
PROGRAM STUDENTS***

[Master in Data Management and Curation: training opportunities in Area Science Park for two STEM graduates » Blog Archive » Area Science Park](#)

[Home](#) \ [Senza categoria](#) \ **Master in Data Management and Curation: training opportunities in Area Science Park for two STEM graduates**

Master in Data Management and Curation: training opportunities in Area Science Park for two STEM graduates

A nine-month course of in presence lessons and experimental training in the laboratories, in Area Science Park, in Trieste. A flat-rate reimbursment is provided for participation. Info day 22 May at 10.00 a.m.

DEADLINE MAY 31th



Open Science methodologies, FAIR-by-design data management and data FAIR-ification, use of tools and software for metadata acquisition and enrichment and tools and methods for preliminary analysis of data and metadata; these are the main skills that the participants of the **Master Data Management and Curation (MDMC)** will acquire at the end of the course organized by Area Science Park, the Scuola Internazionale Superiore di Studi Avanzati – SISSA and the Consiglio Nazionale delle Ricerche – Istituto Officina dei Materiali (CNR-IOM), as part of the activities of supporting projects NFFA-DI and PRP@CERIC, funded by the PNRR* to enhance digital research infrastructures in materials science and life science.

FOR FURTHER INFORMATION

MDMC mdmc@sissa.it

Webpage: [Master in Data Management and Curation \(MDMC\) \(sissa.it\)](https://www.sissa.it/programmi/master-in-data-management-and-curation)

THANK YOU FOR YOUR KIND ATTENTION

AREA SCIENCE PARK

PADRICIANO 99, TRIESTE-ITALIA

WWW.AREASCIENCEPARK.IT

INFO DAY - Master in Data Management and Data Curation (MDMC) © 2024 by Mariarita de Luca is licensed under [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) 

