MDMC
Master in Data Management and Curation

Pilot training programme (Master) in Data Management and Curation

Mariarita de Luca, Research Institute for Technological Innovation (RIT), Area Science Park, Trieste, Italy

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

This will be the first training course on Research Data Management focused on the implementation of **FAIR-by-design research workflows** in the involved laboratories.

The skills and knowledge of FAIR Research Data Management, Curation and Stewardship are nowadays essential to ensure **responsible and reproducible research** in the framework of the possibilities offered by EOSC (European Open Science Cloud)

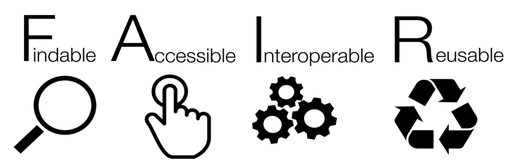
MDMC is a professional training program with six weeks of full-time live lectures in Trieste, and the following seven months of internship in the selected laboratories of the supporting projects (i.e., NFFA-DI and PRP@CERIC), in a materials science or life science experimental facility.
The innovative aspect of **MDMC** lies in its strong emphasis on practical training modules during the six weeks of lessons, coupled with the ambitious aim of implementing FAIR-by-design pipelines and workflows in the selected laboratories as part of the thesis projects.
The course structure is outlined in the following table:

	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	Project laboratory	Presentations and meetings in Trieste

Scientific coordinators: Stefano Cozzini (Area Science Park), Giorgio Rossi (University of Milan and CNR-IOM) and Stefano Ruffo (SISSA)
Master coordinator: Mariarita de Luca (Area Science Park)

RESEARCH DATA

Research data are the raw data collected and studied in research, serving as evidence for published findings. They can be **primary (generated by the researcher)** or **secondary (from existing sources).** Besides raw data, research data include necessary information for generating or replicating results, like computer code, methods, instruments, and metadata (reading.ac.uk).



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

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.

Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3.1 (2016): 1-9.

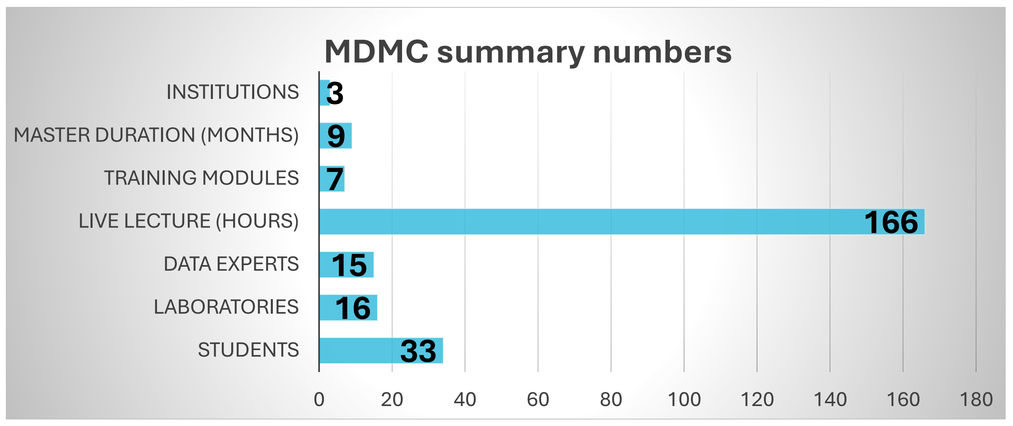
PARTICIPANT LEARNING GOALS:

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

During the seventh months of internship the students will test the implementation of workflows for FAIR-by-design data acquisition in the selected laboratories of the supporting projects.

MDMC OUTCOMES:

1. Having a stable presence in all project 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;
2. **Creation of a new generation of experts, formed in FAIR data management and stewardship.**



COST OF NOT HAVING FAIR DATA

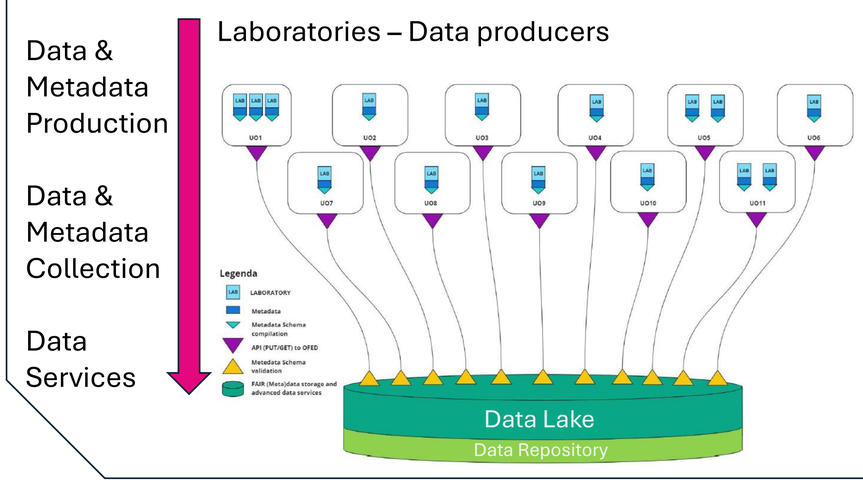
AREA	INDICATORS	COST (Million EUR per year)
Impact on research activities	1. Time spent	4500
	2. Cost of storage	5300
	3. License costs	360
Impact on opportunities for further research	4. Research retraction	4,4
	5. Double funding	25
	6. Cross-fertilization	N.A.
Impact on innovation	7. Potential economic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

European Commission

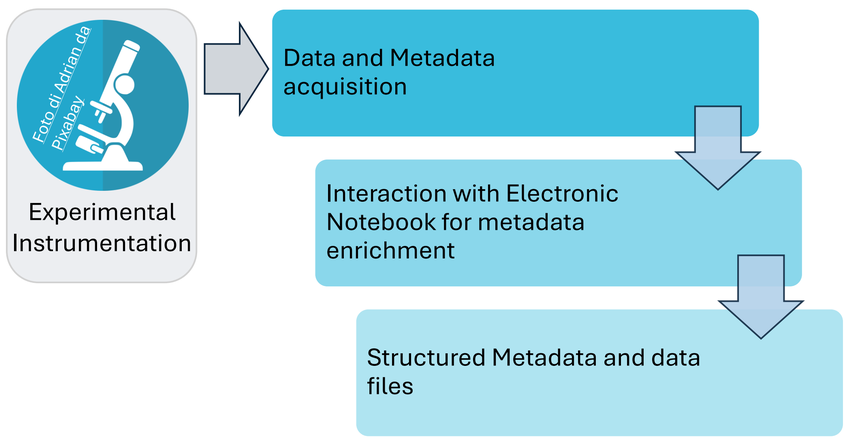
1. FAIR principles are transforming science by promoting data stewardship and openness, requiring proper infrastructure and policies to maximize research data value.
2. The absence of FAIR research increases time and storage costs, highlighting FAIR's potential to improve data handling and storage efficiency.
3. Despite the lack of data linking FAIR to innovation, it is believed that FAIR principles will significantly benefit Europe's science ecosystem and innovation.

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**

FAIR BY DESIGN IN DIGITAL INFRASTRUCTURE – CENTRAL DATA LAKE



FAIR BY DESIGN AT LABORATORY LEVEL - AUTOMATIC WORKFLOW



PNRR SUPPORTING PROJECTS *

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.

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.

SCIENTIFIC PROGRAMME – TRAINING MODULES

The training modules of live lessons have been designed to provide all the skills and competences necessary for the development and execution of the subsequent FAIR-by-design project in laboratory.

FAIR as FULLY AI READY

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

Mons, B. (2020). Invest 5% of research funds in ensuring data are reusable. Nature, 578(7796), 491-491.

***** This Pilot training activity has been funded by the European Union – NextGenerationEU within the project PNRR "PRP@CERIC" IR0000028 and «NFFA-DI" IR0000015 - Missione 4, "Istruzione e Ricerca" – Componente 2, "Dalla ricerca all'impresa" – Linea di investimento 3.1, "Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione" – Azione 3.1.1, "Creazione di nuove IR o potenziamento di quelle esistenti che concorrono agli obiettivi di Eccellenza Scientifica di Horizon Europe e costituzione di reti".