

The CISUC logo consists of a red square icon with a white grid pattern to the left of the text "CISUC" in a bold, red, sans-serif font.

CISUC

Café com dados

-

Processos e workflows de curadoria em Repositórios



FÓRUM
GESTÃO DE DADOS
DE INVESTIGAÇÃO

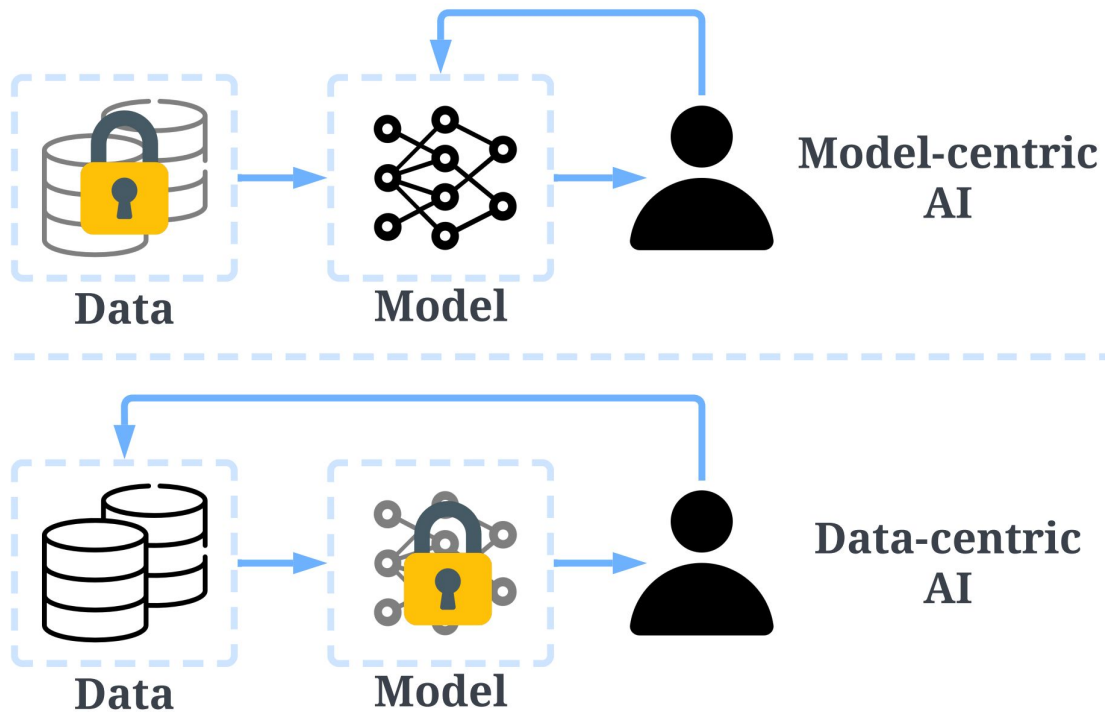
Bruno Direito,

bruno.direito@uc.pt

AI and research data

“If 80 percent of our work is data preparation, then ensuring data quality is the important work of a machine learning team.”

“When a system isn’t performing well, many teams instinctually try to improve the code. But for many practical applications, it’s more effective instead to focus on improving the data”



Importance of data curation in Repositories

- Research data curation practices in Institutional Repositories (IRs)
 - Data curation/management activities in IRs?
 - User activities in IRs?
 - What are the structures of those activities
 - Objectives, goals, data types, tools, policies, rules, norms contradictions

Curation Activities and Actions

Activities	Actions
Understanding data curation needs	Interviewing researchers Communicating with IR or library staff Consulting with researchers
Managing and sharing data	Receiving or transferring data files Cleaning data Converting data to a different file format Developing and adding metadata Validating data Packaging data Uploading and publishing data into IR
Ensuring that data is accessible and reusable	Annotating data for relevant entities Optimizing data to search engine Keeping data up to date into mirror repository
Re-evaluating data for long term preservation	Selecting dataset for long term preservation

Other Related Activities and Actions

Activities

Analyzing data usage

Creating policy and administrative infrastructure

Educating people about data management

Continuing education

Actions

Managing descriptive statistics of data usage
Providing researchers with data tracking results

Understanding local needs and creating local policies and rules

Building infrastructure component

Training librarians

Educating researchers

Providing workshops for data analysis tools

Providing outreach for data curation

Learning the best practices for research data management

Learning future technologies

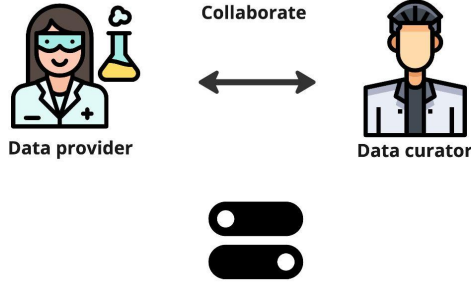
IR staff's role-related activities.

Head	Build or plan data governance structure in their IR Communicate with researchers Provide outreach for their IRs
Data Curator	Consult with data providers and connect them to metadata specialists or IR managers Facilitate communication across different entities Evaluate or view research data to see whether the dataset would continue to be maintained or whether it would be deselected Build or plan data governance structure in their IRs Outreach and educate campus community
IR Manager	Manage IRs on a daily basis Work with data providers to help add metadata and upload data into IRs Answer questions about IR use and data management Outreach and educate campus community
Metadata Specialist	Help data providers to create appropriate metadata for their dataset Design metadata schema for their IRs
Developer	Maintain and update IR software
Subject Specialist	Evaluate or view research data to see whether the dataset should continue to be maintained or whether it should be deselected Manage and approve incoming submissions to their own collections

User profiles

- Head (governance)
- **Data curator (Interface)**
- IR manager (library, IT)
- Metadata specialist (research field know-how, ontologies)
- Developer (IT)
- Subject specialist

Research data curation in IRs



Tools (software, metadata schemas, identifiers schemas, ontologies, etc.)

Policies (deposit, preservation, recommendation for metadata, file formats, workflows)

Skillsets (manage people, data)

Importance of data curation in Repositories

- Designing effective teams for IR data curation
 - communicate with data providers to understand the providers' research data
 - Once the coordination is completed, each activity and their actions are conducted throughout the data and data curation lifecycles.
 - Having the skills needed to develop software within a data curation team makes it easier to maintain the IR system.
- IR infrastructure for research data curation
 - Policies, rules, and norms enable the IR system and the curation teams to be systematic in their practices,