

IDEAL

Integrating Data Analysis and AI in Learning Experiences

D3.5

Datathon Report

Work Package 3: Blended Learning Offer

Prepared by

Smart Revolution Srl and Consiglio Nazionale delle Ricerche (CNR-ITD)

May 2026

Co-Funded by
the Erasmus+ Programme of
the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or of the National Agency Erasmus+ – INDIRE. Neither the European Union nor the granting authority can be held responsible for them. Project Number: 2024-1-IT02-KA220-HED-000251425



Table of Contents

Introduction **4**

1. Educational purpose and target group **5**

Institutional consultation and needs analysis 5

Intended learning 5

2. Datathon programme **7**

3. Group work presentation and analysis **8**

Cross-Group Analysis: Common Questions and Themes 11

Results Analysis 11

AI and Data Analysis Tools: Overview 12

4. Reflection and competence assessment **15**

5. Resources and preparation for delivery **17**

Preparation for delivery– lessons learnt 17

6. Adaptation and transfer **17**

Annex I Preparation Questionnaire **18**

Introduction

The IDEAL Datathon activity aims to help university staff or students develop practical, critical and responsible approaches to using data analysis and artificial intelligence in their work. Informed by participants' needs, it fosters collaborative exploration of real university datasets. Participants practise defining meaningful questions, analysing and visualising data, and translating findings into useful insights for institutional or research-oriented decision-making, while considering data quality, privacy and the limitations of AI. This document outlines the programme's objectives, activities and educational approach.

IDEAL Datathon at the University of Turin and Universitat Pompeu Fabra

The **IDEAL Datathon *"Data and AI for Higher Education: Understand, Explore, Decide"*** took place on the 21st of May 2026 in a **hybrid format**: in presence at the University of Turin (UniTo), Italy with a group of 57 technical and administrative staff members and online with 4 students from the Universitat Pompeu Fabra (UPF), Spain, for **a total of 61 participants**.

Facilitation team:

- Enrica Pautasso, Training expert, Smart Revolution
- Davide Taibi, Senior Researcher, National Research Council of Italy – Institute for Educational Technology (CNR-ITD)
- Giovanni Fulantelli, Senior Researcher, National Research Council of Italy – Institute for Educational Technology (CNR-ITD)
- Sebastian Schwäbe, AI and Training expert, Q21

1. Educational purpose and target group

The innovative learning activity addresses on one side technical and administrative university staff supporting research, innovation and institutional management and university students on the other side. Its purpose is to build a shared understanding of AI and data analysis, encourage informed experimentation and connect these subjects with participants' daily tasks.

For what the technical and administrative staff is concerned, the activity extends the IDEAL approach to professional roles beyond teaching, with competences interpreted in their administrative context. Departmental recruitment should connect participation with individual interest and relevant work responsibilities. The Turin arrangement invited departments to nominate up to five participants. Such a mixed group requires examples that are recognisable across roles, accessible explanations for less experienced users and opportunities for experienced participants to raise specific questions.

Students were selected by the project partner UPF based on their personal interest in the topic.

Institutional consultation and needs analysis

Preparation combines discussion with the university office responsible for internal training and a questionnaire for participants. Institutional discussion establishes the audience, duration, available tools, internal practices and suitable working contexts. The questionnaire examines dataset use, existing AI experience, expectations, difficulties and preferred topics. These inputs guide the examples, tasks and level of explanation.

The preparation workbook contains 54 response records and shows frequent dataset use alongside varied AI experience. Open responses identify prompting, reliability, data organisation and time for exploration as relevant concerns. This supports an introductory format with a common baseline, contextual demonstrations and guided discussion, followed by an optional opportunity for more sustained practice.

The template of the questionnaire used is available in Annex I.

Intended learning

- Knowledge: Understand basic data literacy, data quality and relevant AI capabilities and limitations in university work.
- Skills: Formulate a working question and a contextual request, examine suitable data and interpret or communicate results with appropriate checks.
- Attitudes: Approach AI critically and responsibly, recognise uncertainty and engage with colleagues in reviewing choices and outputs.

The competence focus draws mainly on Type 3, Ethical, Inclusive and Data-Literate AI Use, and Type 4, AI-Driven Professional Development and Collaboration of the IDEAL Competence Framework. It includes critical AI literacy, responsible data use, fairness awareness, reflective development and collegial exchange. A short workshop introduces and practises selected elements; it does not establish comprehensive proficiency across these areas.

2. Datathon programme

The Datathon on Thursday 21 May provides two scheduled participation windows of four hours each, 08:30 to 12:30 and 13:30 to 17:30. The table offers an illustrative thematic structure for either window.

Morning	Afternoon	Topic	Content and educational purpose
08:30 to 09:00	13:30 to 14:00	Working/Research context and challenge	Identify a university task, the intended practical result and working/research conditions. Define the Action Field and a focused challenge.
09:00 to 09:30	14:00 to 14:30	Learning focus and suitable data	Identify what participants need to understand, practise and judge. Define the Learning Field and review the available data or examples.
09:30 to 10:30	14:30 to 15:30	Exploration and experimentation	Explore patterns and questions, trial an approach and refine requests with support. Use challenge-based and design-thinking methods to encourage inquiry and revision.
10:30 to 11:30	15:30 to 16:30	Analysis and verification	Develop analysis or visualisation, check calculations and source information, and explain methods and limitations.
11:30 to 12:00	16:30 to 17:00	Interpretation and presentation	Share the challenge, approach, findings and implications. Provide facilitator and peer feedback and space for clarification.
12:00 to 12:30	17:00 to 17:30	Reflection and transfer	Identify learning, remaining uncertainty and a realistic next step for university work/research. SPIDER questionnaire to assess and orient learning development.

At UniTo the activity took place in person, the participants divided themselves between morning and afternoon according to their preferred window. 33 participants attended the morning session, while 24 attended the afternoon session. During each session, participants divided themselves in groups of 4-5 people, based on their office/department or similar tasks and challenges.

At the start of each topic/sub-session the team of facilitators provided a brief introduction and examples of AI and data analysis tools that could be used. Throughout the group work, facilitators supported discussions, encouraged the exchange of experiences and provided practical guidance on the use of the selected tools.

At the end, each group presented to the audience their challenge and their solution, through the designated template (Annex II). Facilitators and participants raised questions and provided feedback.

At UPF, students attended a two-hour online introductory session on 19 May, delivered by one of the facilitators, Davide Taibi. The session introduced the objectives and scope of the Datathon, presented the templates to be used during the exercise, and provided an overview of selected AI and data analysis tools. On 21 May, in parallel with the Datathon held in Turin, four students participated in the exercise remotely and asynchronously, working in two groups. The presentations on the designated template were sent directly to the team of facilitators who provided feedback.

3. Group work presentation and analysis

Participants elaborated eight group presentations that are here summarised with the name of the team, the title of the group work, the area of focus, the main question (challenge) and the key findings. The full presentations are presented in Annex III.

1. NO LOGO — PhD Enrolment Data for FFO Purposes

Area: Staff supporting research (PhD administration)

Main Question: To obtain an analysis of PhD enrolment data with a focus on the criteria that affect the calculation of the FFO (Ordinary University Funding Formula).

The group worked with three internal Excel datasets covering academic years 2024, 2025, and 2026, containing data on enrolment, internationalization rates, financial support coverage, and PhD course characteristics. Personal identifiers were removed before processing. ChatGPT was used for prompt-engineered text analysis and Julius AI for quantitative trend identification. They attempted to use Napkin for data visualization, but the tool failed to produce usable results.

Key findings: the internationalization quota grew from 15.7% (2024) to 26.5% (2026); financial coverage of PhD students remained consistently above 92%.

2. SOS RDA — Expense Eligibility Across Competitive Funding Lines

Area: Research, Innovation and Internationalization Directorate

Main Question: What is the correct cost imputation for expenses to be reported across the different types of competitive funding (Regional, Ministerial, and European)?

The dataset consisted of the official financial reporting manuals for Regional (CSR Piemonte), Ministerial (PNRR), and European (Horizon Europe) funding in PDF format. Using ChatGPT, the group extracted eligible cost categories from the three documents and requested a comparative overview. NotebookLM was then used to generate a structured comparison chart, resulting in a "Comparative

Guide to Expense Eligibility: CSR, Horizon Europe and PNRR." An attempt to use Napkin for visualization failed, as the tool could not accept PDF or Word inputs.

3. CLE-SIPE — Research Project Deadline Chronogram (2026–2028)

Area: Research and Innovation Area (CLE Pole) + Digital Services

Main Question: Which research and innovation projects managed by the CPS Department are expiring between 2026 and 2028, and how can their deadlines be visualized in a Gantt-style chronogram?

The group worked with a large internal Excel export of 2,472 rows and 24 columns covering projects managed by the Department of Culture, Politics and Society (CPS), including project registry data, funding bodies, multiyear financial forecasts, and budgets. ChatGPT (extended thinking mode) generated a textual deadline schedule organized by funding body (MUR, EU Commission, Foundations). ChatGPT (standard mode) then converted it into an Excel Gantt chart; the first version was rejected for poor readability and a reworked iteration was produced.

4. Indomiti GPT — Extracting Culture & Creativity Projects from IRIS

Area: Research, Innovation and Internationalization Directorate

Main Question: Which Culture and Creativity projects are currently active at the University? Which research groups and departments should be more actively involved in the promotion of specific calls in this thematic area?

Their dataset was an export from IrisAP (UniTo's research repository) covering active projects from 2019 to 2026. The group built a custom GPT configured with the IrisAP dataset to enable thematic search queries. Initial results suggested approximately 548 projects attributable to the Culture and Creativity domain, with the Dipartimento di Studi Umanistici having the highest concentration of active projects, followed by DISAFA, Studi Storici, Chimica, and Informatica.

5. DATA-NINJA — Trending Topics in European Research Funding

Area: Research & Innovation Offices (Directorates RII, SUM, MED)

Main Question: What are the trending topics in European research funding under Horizon Europe, and how does UniTo perform compared to European trends?

The group integrated two datasets: the CORDIS dataset of funded HEU projects (2021–present) and the EU Commission Dashboard (2021–present), both in XLS format. ChatGPT was used for dataset integration, analysis, and presentation generation. Julius AI required a paid upgrade; NotebookLM was judged less powerful. Key findings: the top three funded areas are ERC (€10.63B), Climate/Energy/Mobility (€9.30B), and Digital/Industry/Space (€8.45B) — totalling 53.1% of HEU funding. UniTo's strongest areas are ERC, Health, and Agrifood. Three strategic recommendations for UniTo were produced.

6. S.L.N.S.G. — Academic vs. Industrial Patent Filing Trends

Area: Research, Innovation and Internationalization Directorate

Main Question: What are the similarities and differences between the academic and industrial worlds in patent filing trends over the period 2023–2026?

Because real UIBM (Italian Patents and Trademarks Office) datascraping would have exceeded the Datathon timeframe, the group made an explicit methodological choice to use two synthetic datasets generated by ChatGPT — 100 records each (university patents and company patents) in CSV format, with anonymized inventors. NotebookLM was used to analyse the synthetic datasets; Gamma was used for the final presentation slides. Julius AI ran out of credits; Canva was judged unconvincing. Based on the synthetic data, the group identified overlapping areas including quantum computing, fintech/blockchain, and greentech.

7. AI2Design — GenAI-Enabled Learning Design Summary

Institution: Universitat Pompeu Fabra / TIDE, 2026

Challenge type: Summarization — To summarise a given learning design adapting it to a learner-facing text style in English and evaluate it, reviewing if it aligns with the structure of the given learning design and Bloom's Taxonomy.

Source materials included lesson plans from the TIDE research group, a custom system prompt in Markdown, and Bloom's Taxonomy documentation. Two AI tools were used: ChatGPT 5.5 with Extended Thinking to generate a simulated interactive UI that adapts a given learning design and calculates a pedagogical alignment score; and Gemini 3.5 Flash as an evaluator, producing quantitative alignment scores (0–100) with improvement feedback. The output was a two-column simulated dashboard displaying the learner-facing lesson plan alongside actionable feedback.

8. TeamPulse — AI-Powered Teamwork Analytics for Collaborative Learning

Institution: Universitat Pompeu Fabra / CSIM Master Program, 2026

Challenge type: Summarization — To generate an automatic summary of students' answers to the teacher, from a collaborative learning activity that followed a Pyramid script. The collaborative task given to students was to debate and define the most valuable non-digital skill a student can learn in an increasingly AI-driven educational space.

The dataset consisted of written responses from eight students collected via PyramidApp. Three AI tools were tested comparatively: Gemini (primary tool, judged best at structured categorical output), ChatGPT (good at identifying blind spots, less consistent in format), and Grok AI (used for forward-looking intervention suggestions). All three were prompted to produce a four-part structured summary: Consensus Map, Blind Spots, Friction Points, and Orchestration Nudge.

Cross-Group Analysis: Common Questions and Themes

Across all eight presentations, the following questions and themes were central to the groups' work, shaping both the design of each exercise and the learning that emerged from it:

- Can AI help make sense of large, messy institutional datasets? Several groups (NO LOGO, CLE-SIPE, Indomiti GPT) started from the question of whether AI tools could extract meaningful patterns from complex internal exports — enrolment records, project registries, research repositories — that would normally require hours of manual analysis.
- Can AI reliably compare documents and regulations across funding types? SOS RDA explored whether AI could serve as a practical guide for administrative staff navigating the different cost eligibility rules of regional, ministerial, and European funding lines — a question with direct implications for research management.
- Can AI support the identification of strategic opportunities in European research? DATA-NINJA asked whether publicly available Horizon Europe data, processed through AI, could produce actionable intelligence about funding trends and institutional positioning — a question relevant to research strategy offices across universities.
- Can AI bridge academic and industrial knowledge through data? S.L.N.S.G. posed the question of whether AI-assisted analysis of patent filing trends could reveal meaningful convergence and divergence between university and industry innovation patterns, pointing to potential collaboration areas.
- Can AI make learning designs more accessible and pedagogically sound? AI2Design (UPF) addressed the question of whether AI can transform teacher-facing lesson plans into learner-facing summaries while maintaining alignment with Bloom's Taxonomy — a question at the intersection of educational design and AI-assisted content adaptation.
- Can AI extract structured pedagogical insights from unstructured student responses? TeamPulse (UPF) asked whether AI tools can produce meaningful, categorized summaries of student-generated text from collaborative learning activities, turning raw qualitative data into actionable input for teachers.

Results Analysis

The eight groups collectively demonstrated that AI tools can provide genuine value in university administrative and research support contexts, while also surfacing important caveats about accuracy, tool maturity, and the skill required to use them effectively.

Most successful outcomes were achieved when the task was clearly bounded (a specific question, a well-defined dataset), prompt engineering was applied iteratively, and the output was treated as a starting point for human review rather than a final product. The research project deadline chronogram (CLE-SIPE), the expense eligibility guide (SOS RDA), and the Horizon Europe trend analysis (DATA-NINJA) are concrete examples of outputs that would have required hours of manual work but were produced in minutes — even if requiring verification and refinement.

Less successful cases were those where the source data was too complex or ambiguous (Indomiti GPT's thematic classification), where the tool chosen was not fit for purpose (Napkin with PDF inputs), or where AI fluency was mistaken for accuracy without verification.

A key cross-cutting finding is that AI tools significantly accelerated tasks involving text transformation, summarization, comparison, and visualization — but they did not eliminate the need for domain expertise. In every group, the quality of the final output depended on the user's ability to formulate precise questions, recognize errors, and apply professional judgment to the results.

AI and Data Analysis Tools: Overview

Across the eight presentations, 11 distinct AI or data analysis tools were used (eight AI tools and three data sources).

Tool	Category	Used by	Main use	Outcome
ChatGPT (incl. GPT-5.5 Extended Thinking)	Generative AI / LLM	NO LOGO, SOS RDA, CLE-SIPE, Indomiti GPT, DATA-NINJA, S.L.N.S.G., AI2Design, TeamPulse	Text analysis, summarization, dataset building, slide generation, custom GPT	Widely used; effective but requires careful prompting and verification
Julius AI	Data analysis / AI	NO LOGO, DATA-NINJA, S.L.N.S.G.	Quantitative trend analysis, dataset integration	Partially effective; abandoned in 2 groups due to paywall or credit exhaustion
NotebookLM	Document AI / knowledge management	SOS RDA, DATA-NINJA, S.L.N.S.G.	Comparative chart generation, dataset analysis, trend identification	Effective for document-based analysis; rated best for structured dataset review
Napkin	AI data visualization	NO LOGO, SOS RDA	Visual data representation	Failed in both cases: could not handle PDF/Word inputs or complex tabular data
Gamma	AI presentation builder	S.L.N.S.G.	Slide deck creation	Successful; rated best presentation tool tested

Tool	Category	Used by	Main use	Outcome
Canva AI	AI-assisted design	S.L.N.S.G.	Slide creation	Attempted but judged unconvincing
Gemini (incl. Gemini 3.5 Flash)	Generative AI / LLM	AI2Design, TeamPulse	Pedagogical evaluation, structured thematic analysis of student responses	Effective; judged best at producing categorized, structured output
Grok AI	Generative AI / LLM	TeamPulse	Forward-looking pedagogical intervention suggestions	Partially effective; did not match Gemini's structured output format
IrisAP	Institutional research repository	Indomiti GPT	Source data for custom GPT	Data source; used as basis for building a custom GPT on UniTo project data
CORDIS / EU Commission Dashboard	Open research data	DATA-NINJA	Source data for HEU trend analysis	Data source; required integration due to structural differences between datasets
PyramidApp	Collaborative learning tool	TeamPulse	Structured data collection from students	Data source; provided input dataset for AI qualitative analysis

Tool Summary

- Generative AI and Data analysis tools tested: 5 (ChatGPT, Gemini, Grok AI, Julius AI, NotebookLM)
- AI-assisted creative and visualization tools tested: 3 (Napkin, Gamma, Canva AI)
- Data platforms and repositories used as sources: 3 (IrisAP, CORDIS/EU Dashboard, PyramidApp)
- Total distinct tools or platforms across all 8 groups: 11

Tools Recommended for Re-use

- ChatGPT – recommended by almost all groups, with the caveat of careful verification of outputs
- NotebookLM – recommended for document analysis and structured knowledge extraction
- Gamma – recommended for AI-assisted presentation creation
- Gemini – recommended for structured qualitative analysis tasks

Tools Not Recommended as Tested

- Napkin – insufficient file format support for the use cases attempted
- Canva AI – output quality judged insufficient for professional presentations
- Custom GPT on IrisAP – too slow and imprecise for generic thematic search

Facilitator support – lessons learnt

Facilitators have the key task to support the working question before focusing on tool operation and adapt the topic blocks to group needs and task complexity. Cross-departmental exchange can reveal different assumptions. A clearer question, a checked result or an account of limitations can be useful even when a complete solution is not feasible. Finally, it is important to preserve space for presentations and feedback and distinguish group outputs from individual reflection where individual learning is later assessed.

4. Reflection and competence assessment

Reflection is embedded in the programme through questions, discussion of outputs and a short account of what participants would use or change. The IDEAL SPIDER provides a manageable option for awareness and self-reflection after a condensed workshop. It can help participants recognise areas of confidence, uncertainty and further interest, and can support a conversation with a facilitator. Participants shall assess themselves on a scale from 1 to 5 on a set of AI and data analysis topics/competences.

Both UniTo and UPF participants were offered to self-assess their learning development through the IDEAL SPIDER questionnaire at the end of the Datathon exercise. In total, 49 participants filled in the IDEAL SPIDER and 27 of them decided to share it with the team of facilitators.

The questionnaire results, collected from 27 respondents, indicate an overall positive level of self-assessed knowledge and confidence across the topics at the end of the Datathon. The highest scores were recorded for checking information generated by AI (4.60) and learning for an AI-enabled future (4.56), followed by understanding the implications of AI for future work (4.40) and the responsible use of AI in professional contexts (4.32). These results suggest strong perceived confidence in evaluating AI-generated information and adapting to the changes brought about by AI.

Relatively lower scores were reported for using AI in everyday work (3.40) and recognising bias (3.56), followed by trusting numbers (3.76) and working with data (3.92). These areas may therefore offer opportunities for further capacity building, particularly through practical exercises focused on the effective use of AI tools, data interpretation and the identification of potential biases.

Overall, the results suggest that participants felt relatively confident about understanding and critically assessing AI, while showing greater scope for strengthening their practical application skills.

SPIDER responses are self-reported reflections. They should not be presented as equivalent to independently demonstrated competence or a complete LEVEL5 validation. If change is to be examined, the programme must specify the initial and later assessment arrangements and the basis for comparison. A single questionnaire alone does not establish development.

Where a longer pathway includes practical application and appropriate evidence, the LEVEL5 reference system can support a more detailed assessment across Knowledge, Skills and Attitudes. That process requires a selected competence, suitable evidence and explained judgments. Participation in this programme does not automatically lead to certification.

Completion of reflection activities and permission to share individual responses should be explained separately. Intended outputs include a learning reflection, a relevant professional question and, for Datathon participants, a brief account of the approach and result. Participation figures, feedback and assessed outcomes belong in the experience report.

5. Resources and preparation for delivery

Resource	Use in the programme
Questionnaire and institutional consultation	Select relevant examples, identify starting points and clarify organisational conditions.
IDEAL introduction and slides	Provide accessible input, practical cases and a route into challenge identification.
Contextual prompting materials and RCTF sheet	Support a demonstration and provide templates participants can adapt.
Traffic light discussion aid (Mentimeter) and university guidelines	Support discussion of suitable inputs, institutional conditions and verification.
Prepared examples and tested demonstration outputs	Provide continuity if a tool is slow, inaccessible or unavailable.
IDEAL SPIDER and competence resources	Support reflection and, where planned, a longer assessment pathway.

Preparation for delivery– lessons learnt

Before delivery, confirm tools, accounts, connectivity and suitable materials with the host institution. Practical upload examples should be consistent with the applicable university guidance. Product capabilities and account conditions change, so the presentation should be checked before reuse. Prepared sample or synthetic data should be available where participants cannot use their own material.

6. Adaptation and transfer

The programme can be adapted by retaining the connection between institutional needs, a shared introduction and supported work on professional questions. Change the examples and level of input to match the audience. Provide dedicated time for questions, group presentations and feedback, and agree how participants can continue after the event.

Annex I Preparation Questionnaire

The questionnaire supports preparation of training for university staff. It explores working contexts, existing experience and learning needs. The questions below are an English translation of the questions recorded in the preparation workbook.

Dear participant,

The training course on data analysis and AI will take place in a few weeks. To develop content that matches your starting competences and expectations, please complete this short anonymous questionnaire.

It will only take a few minutes. Let's get started!

See you soon,

Use of data at work

1 How frequently do you use datasets, meaning structured and organised collections of data, in your daily work?

☐ Never

☐ Rarely

☐ At least once a week

☐ Daily

2 For which activities do you mainly use datasets? Select all that apply.

☐ Analysis or reporting

☐ Decision support

☐ Research

☐ Monitoring activities or projects

☐ Teaching

☐ I do not use data

3 Which tools do you usually use to work with data? Select all that apply.

☐ Spreadsheets such as Excel or Google Sheets

☐ Statistical software such as SPSS, R or Stata

☐ Dashboards or visualisation tools

☐ Databases or information systems

☐ I do not use specific tools

Familiarity with artificial intelligence

4 Have you used generative artificial intelligence tools, such as chatbots or AI assistants?

☐ Never

☐ Rarely

☐ At least once a week

☐ Daily

5 For which activities do you use AI in your work? Select all that apply.

☐ Writing or revising texts

☐ Analysing or summarising data

☐ Programming support

☐ Preparing teaching materials

☐ Automating administrative tasks

☐ None

6 Which AI tools do you use? Select all that apply.

☐ ChatGPT

☐ Gemini AI

☐ Claude AI

☐ Notebook LM

☐ Canva

☐ Other _____

7 How would you rate your competence level in using AI?

☐ No experience

☐ Basic

☐ Intermediate

☐ Advanced

Needs and expectations

8 Which aspects would you most like to explore during the workshop? Select a maximum of two.

☐ Understanding how AI tools work

☐ Using AI to analyse data

☐ Summaries and reports

☐ Creating presentations and multimedia objects

☐ Ethical issues and responsible use of AI

9 What is the main difficulty you encounter when using datasets or AI in your work?

Open response

_____

_____

10 What do you expect to gain from this workshop?

Open response

_____

_____



Co-Funded by
the Erasmus+ Programme of
the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or of the National Agency Erasmus+ – INDIRE. Neither the European Union nor the granting authority can be held responsible for them.