

**The
Alan Turing
Institute**

EMBL-EBI



**Medical
Research
Council**

Who Does the Data Work?

Roles, Recognition, and Career Pathways in Biomedical Data Science

Emma Karoune, Denise Bianco, Vera Matser, Daria Sokolova, Kim Gurwitz, Cath Brooksbank, Eirini Koutsouroupa, Giulia Tomba.

STEP-UP Conference - 29 July 2026

Advancing Biomedical Data Science Careers



Aim: Documenting skills, roles and team science approaches to foster the recognition and advancement of data science careers in biomedical research

Webpage:

<https://www.turing.ac.uk/research/research-projects/advancing-biomedical-data-science-ca>



Advancing Biomedical Data Science Careers

- Supported by **MRC Biomedical Data Science Leadership Award**
- 1 of 5 projects funded in this award
 - University of Liverpool
 - University of Leeds
 - University of Oxford
 - HDRUK

‘Develop innovation and evidence to improve how biomedical research resources and recognises high-quality data science’.



ABDC Project Team

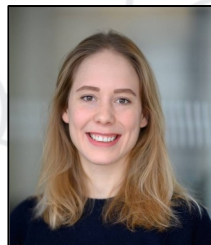
The Alan Turing Institute



Emma Karoune
Project Lead,
Principal Researcher



Vera Matser
Co-Lead,
Head of Skills



Denise Bianco
Senior Research
Community Manager



**Eirini
Koutsouroupa**
Programme Manager



Giulia Tomba
Senior Research
Associate

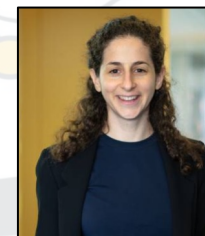


Alison Marsh
Project Management
Support

EMBL-EBI



Cath Brooksbank
Co-Lead,
Head of Training



Kim Gurwitz
Co-Lead,
Scientific Training Lead



Daria Sokolova
Scientific Project
Officer

Investigating who does the data work?

How do we do this?



Ecosystem



Organisation



Individual



Work packages

1 - Mapping the landscape - people, organisations, projects, skills and competency frameworks to increase ecosystem understanding and identify gaps

2 - Documenting case studies at organisational, team and individual level to understand roles, career pathways and team science approaches

3 - Evaluate and recommend innovative approaches and ways of working to drive forward capacity building and improve quality and standards in biomedical data science

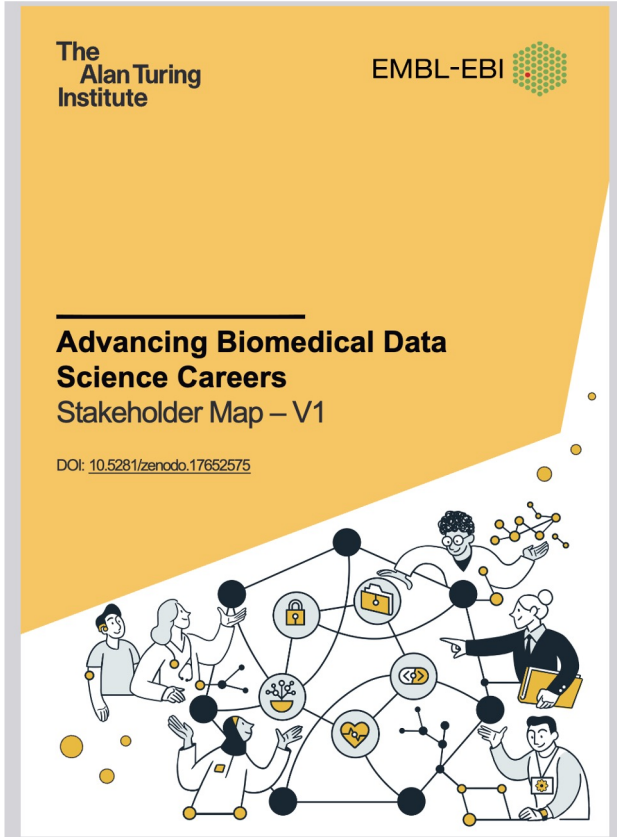
Reflecting the biomedical data science ecosystem

MRC definition of biomedical research

‘Ranging from omics to microscopy to medical imaging to population cohort to environmental data’

- Very broad definition
- Also taking a very broad view of data science roles - really anyone working in a data science team

WP1 - Landscape mapping V1



- Draft map on Miro
- Published on Zenodo

<https://doi.org/10.5281/zenodo.17652575>

Landscape mapping V2

Overview

Discussions

Advancing Biomedical Data Science Careers

The [Advancing Biomedical Data Science Careers \(ABDC\)](#) project is a collaboration between two world leaders in biomedical data science – The Alan Turing Institute and EMBL's European Bioinformatics Institute (EMBL-EBI). It began in February 2025 and is set to conclude in January 2027.

Mapping the UK Biomedical Data Science Landscape

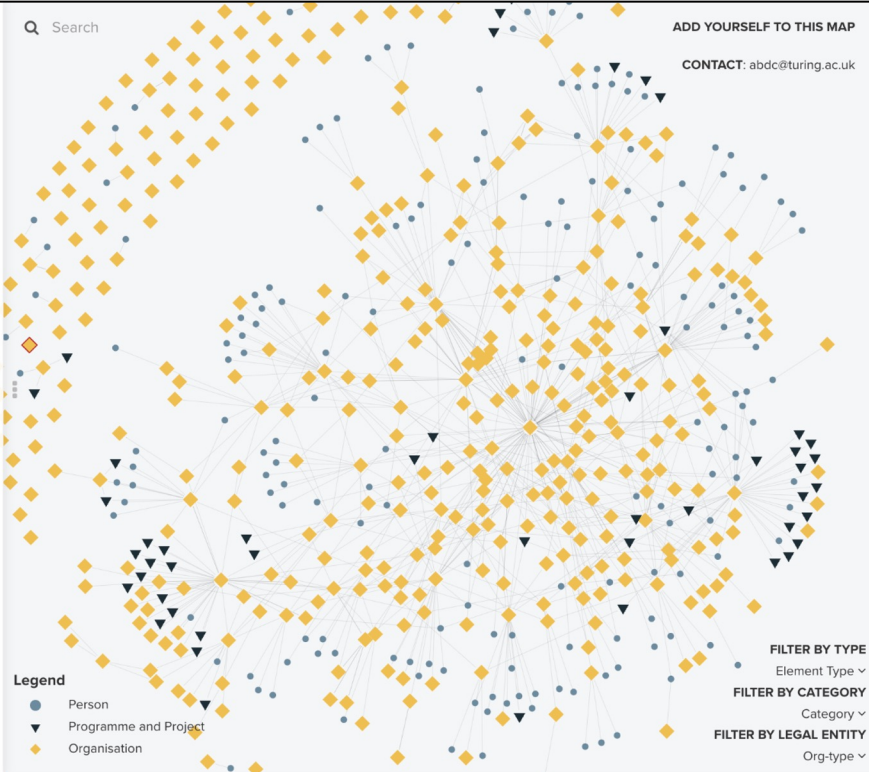
In order to achieve our goal of evaluating the skills gaps in biomedical data science and identifying priority areas for development, we first needed to understand the landscape itself and who the stakeholders in this space are.

To guide our mapping, we adopted a broad definition of biomedical research, following the Medical Research Council's strategic review (2022), as *ranging from omics to microscopy to medical imaging to population cohort to environmental data.*

Q Search

ADD YOURSELF TO THIS MAP

CONTACT: abdc@turing.ac.uk



Legend

- Person
- ▼ Programme and Project
- ◆ Organisation

FILTER BY TYPE

Element Type ▾

FILTER BY CATEGORY

Category ▾

FILTER BY LEGAL ENTITY

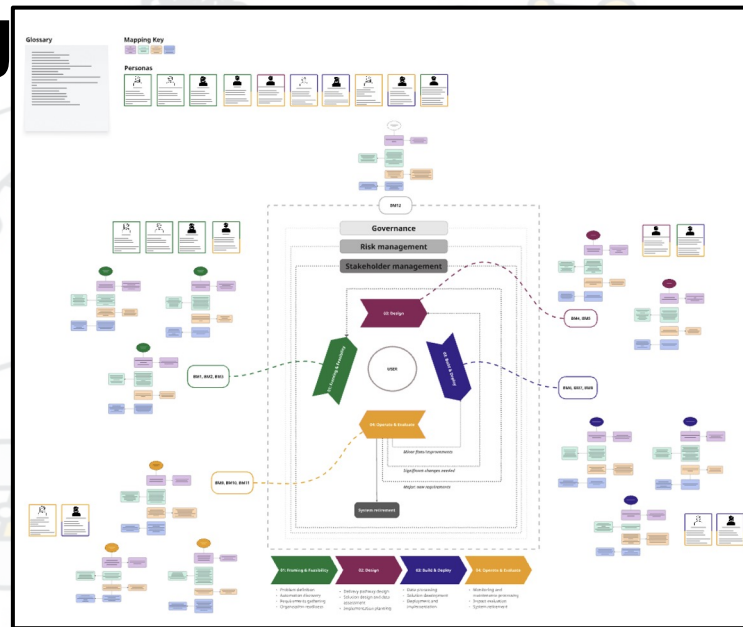
Org-type ▾



[Interactive map](#) on Kumu + [Presentation](#)

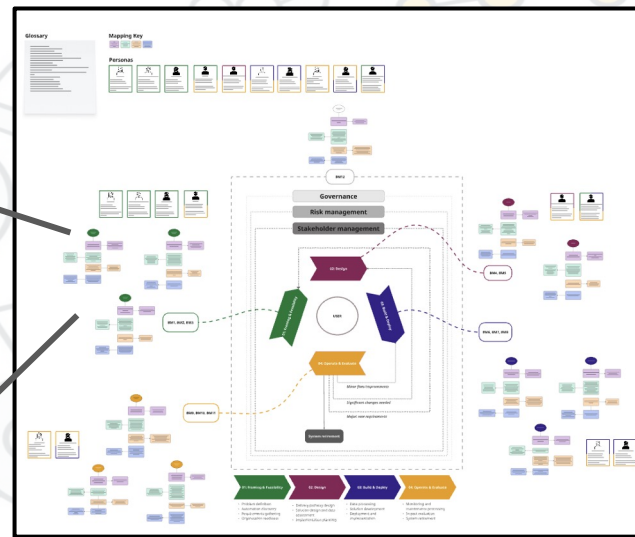
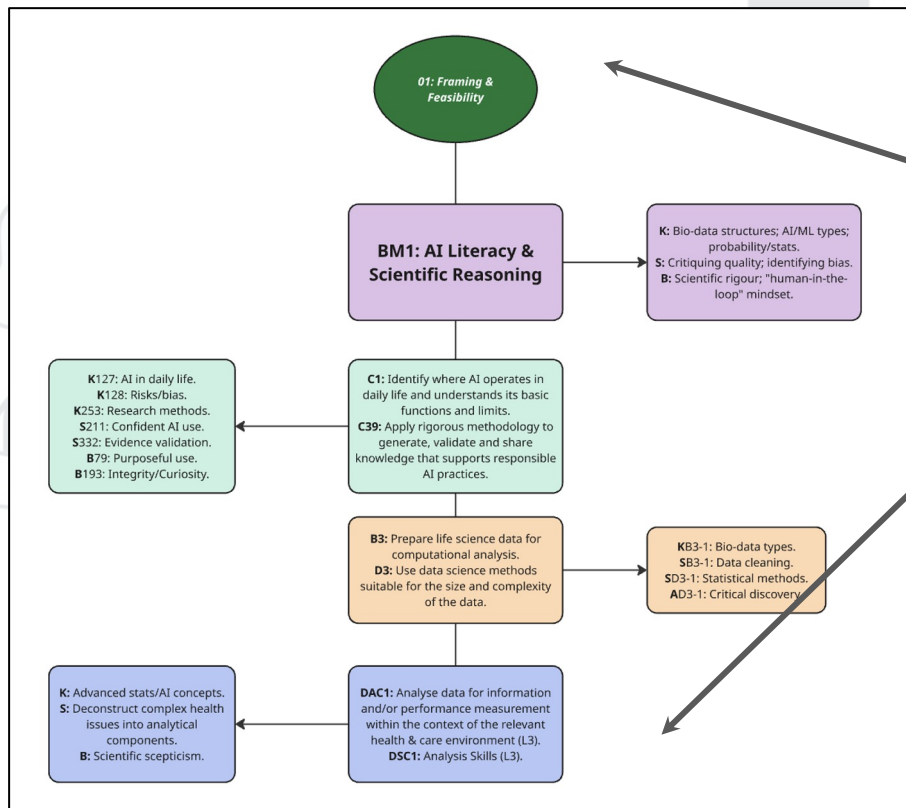
WP1 - Competency mapping

- **The challenge:** no competency frameworks for UK biomedical data science
- Benchmarking and mapping of existing trusted standards to **define a "Minimum Standard"**
- The **three pillars** of our mapping:
 - AI Skills for Business (AISfB)
 - NHS National Competency Framework (V2)
 - ISCB (International Society for Computational Biology) V3
- **Results:** 12 minimum standards, 4 proto-personas, 10 specific personas



Access through
EMBL-EBI
Competency hub

Competency mapping



Mapping Baseline Competencies to the Project Lifecycle

Competency mapping



Academic PI

Sector: Academic Research
Core Competencies: BM1, BM3, BM12

Sets research direction and ensures team compliance with scientific standards.



PPIE Lead

Sector: Academia-NHS/Mixed
Core Competencies: BM2, BM11, BM12

Promotes inclusive dialogue to build community confidence in AI benefits and risks.



Clinical Bioinformatics Data Scientist

Sector: Academia-NHS/Clinical
Core competencies: BM4, BM5, BM9

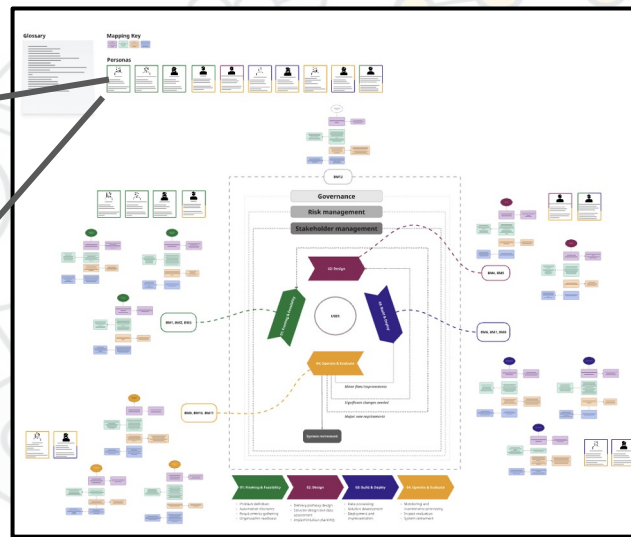
Stewards data capability, ensuring systems are socially attuned and technically valid.



Public Ethics & Trust Representative

Sector: Community/NHS Oversight
Core Competencies: BM2, BM4, BM11

Audits the technical blueprint to ensure privacy, equity, and transparency are built into the design before data processing.



Creating Personas

WP2 - Documenting case studies

We want to learn from the ground what helps (and what doesn't) those who work in biomedical data science



INDIVIDUAL

What was your career pathway?

...

Formal and informal training

...

How did you overcome interdisciplinary barriers?



TEAM

What roles make up the team?

...

Inclusion, diversity and PPIE in the team

...

Inter- multi- or trans-disciplinary?



ORGANISATION

Competency/skills frameworks

...

Integrated vs separate data science service teams

...

Fixed term or permanent contracts?

Documenting case studies

- **Interview sample**

- Individual (roles and career journeys): 19 (including 5 ECRs)
- Team (team composition and structure): 14
- Organisational (strategy): 7

** These dimensions often overlap in our interviews*

- **Almost equal gender distribution (17 - 23)** – more demographic data will be collected during the second round of consent request (Turing EDI monitoring form)
- **6 industry, 6 NHS, 20 research institutions** (of which 4 came through an open call on our webpage)
- Across Genomics, Proteomics & Integrated Multi Omics, Population Health, Epidemiology & Public Health, Precision Medicine,...

Documenting case studies

Biomedical data science careers are non-linear, contingent, and emergent

Professional identities are multiple and fluid, shaped by both individual agency and structural positioning

Data scientists as context-switchers who bridge across specialisms

Unstructured career spaces create isolation, making peer communities essential for support and learning/upskilling

Navigating the boundaries between dry and wet lab cultures requires skills that are learnt informally and rarely valued institutionally

Generalist skills are structurally undervalued even when they're essential

Why is this work important?

- **Rapidly changing research landscape** - embedding of data science and intro of AI
- Better understanding needed to **make research more efficient and more equitable** for everyone working in research
 - Building an **evidence base** is the start!
 - Working with other projects to build stronger evidence
 - **Making recommendations** for organisations, funders and policy makers to support new roles, career pathways, team science approaches.

Enter the Hidden REF Competition!!!



Acknowledgements and thanks



Thanks to the ABDC team

Thanks to our advisory board members and wider biomedical data science community for their input and support.

Contact us

abdc@turing.ac.uk



Contacts

- ABDC inbox
abdc@turing.ac.uk
- Denise Bianco
dbianco@turing.ac.uk
- Emma Karoune
ekaroune@turing.ac.uk

