



Advancing Biomedical Data Science Careers Stakeholder Map – V1

DOI: [10.5281/zenodo.17652575](https://doi.org/10.5281/zenodo.17652575)



1. Introduction

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.

From our extensive experience in this area and working with a diverse range of existing and new partners, our project outputs will help organisations to incorporate data science skills into their teams and establish greater understanding of the common language needed for skills and careers in this domain.

To achieve its aim of documenting skills, roles, and team science approaches that foster the recognition and advancement of data science careers in biomedical research, the project is structured into three work packages:

- Mapping skills and competency frameworks to identify gaps.
- Documenting case studies at organisational, team, and individual levels to understand roles, career pathways, and team science approaches.
- Evaluating and recommending innovative approaches and ways of working to strengthen capacity building and improve quality and standards in biomedical data science.

As part of **Work Package 1 – Mapping skills and competency frameworks**, we have begun a landscape mapping exercise to better understand the biomedical data science ecosystem. This phase focuses first on **mapping stakeholders**, followed by a detailed review of available and relevant competency frameworks. From there, we will conduct a high-level competency analysis to highlight areas requiring further development, which will then be linked to training offerings.

The work presented below is the **first draft of the UK biomedical data science stakeholder map**, one of the outputs under Work Package 1.

2. Stakeholder map

Work on this draft map, developed in Miro, began in April 2025 and continued through October 2025. We will keep updating our map every six months until the end of the project in January 2027.

To guide our work, we adopted a broad definition of biomedical research, following the Medical Research Council: *ranging from omics to microscopy to medical imaging to population cohort to environmental data*.

The draft map captures a wide range of stakeholders, including organisations, projects, and teams, as well as the different domains of biomedical data science they represent. Our focus is on the UK ecosystem, with international players included only when their work directly relates to the UK.

Our approach was iterative and outward-looking. We began by mapping relevant programmes, projects and teams within the two leading institutions of the ABDC project, as well as the ABDC project's partner institutions (shown on the left side of the map). From there, we expanded outward to capture the broader UK ecosystem, incorporating feedback from the wider community.

It is important to note that while this map is intended to help organisations and individuals understand the UK biomedical data science landscape, it is not exhaustive, particularly at this first-draft stage.

2.1 Clustering

The following categories were used to cluster the different stakeholders. It should be noted that these categories were defined by the research team, and their definitions are therefore subjective and may differ from those used by others. Below, we provide a short description of each category to help navigate the map.

- **Category: The Alan Turing Institute: Grand Challenges & Mission programmes**
Description: *The Alan Turing Institute focuses on transforming health through its challenge-led approach outlined in the 5-year Turing 2.0 strategy, and its key mission programmes.*
- **Category: The Alan Turing Institute: relevant projects**
Description: *Past and present projects from The Alan Turing Institute that may be relevant to, inform, or support the work of ABDC.*
 - ◇ Sub-category: Past Turing projects to document
- **Category: ABDC Project: core team**
Description: *The Advancing Biomedical Data Science Careers project is jointly delivered by the Alan Turing Institute and EMBL-EBI. Here, we have mapped teams, projects, and programmes from both organisations that are relevant to the project.*
 - ◇ Sub-category: Other relevant Turing teams and Interest Groups
 - ◇ Sub-category: EMBL/EMBL-EBI programmes, projects, and products
 - ◇ Sub-category: Other relevant EMBL-EBI teams, projects and products
- **Category: ABDC Letter of Support (LoS) partners**
Description: *Organisations that provided a letter of support for the Advancing Biomedical Data Science Careers project.*
- **Category: Research Institutions**
Description: *Organisations that conduct scientific research across the UK, including universities, university centres and units, as well as independent or charitable organisations that apply data-driven approaches to improve understanding of biomedical systems across sectors.*
 - ◇ Sub-category: UK Universities for Health Data Science (main)
 - ◇ Sub-category: Centre for Doctoral Training (CDTs) and Doctoral Training Partnership (DTPs)
 - ◇ Sub-category: University TTOs (Tech Transfer Offices)
- **Category: Patient and Public Involvement and Engagement (PPIE)**
Description: *Initiatives, communities, and groups that work with patients and the public at all stages of research to ensure that studies reflect real-world needs, that patients' rights are fairly represented, and that outcomes are communicated in clear and accessible ways.*
- **Category: MRC list of institutes, units and centres**
Description: *MRC-funded research organisations across the UK, including long-term, multidisciplinary establishments (Institutes), focused research groups (Units), and collaborative entities that integrate various research efforts (Centres), typically based within universities or independent research organisations. To note: MRC is moving from Centres and Units to 'Centres of Research Excellence'. Current Centres/Units will shut over next few years, and new CoREs will open.*
 - ◇ Sub-category: Institutes
 - ◇ Sub-category: Centres and Units - Infections and Immunity Board
 - ◇ Sub-category: Centres and Units - Molecular and Cellular Medicine Board

- ◇ Sub-category: Centres and Units - Neurosciences and Mental Health Board
- ◇ Sub-category: Centres and Units - Population and Systems Medicine Board
- ◇ Sub-category: Other relevant MRC investments:
- ◇ Sub-category: MRC Biomedical Data Science Leadership Award projects

- **Category: UK Government, agencies and departments**

Description: *Departments, agencies and government-funded organisations and initiatives.*

- **Category: National Health Service (NHS)**

Description: *NHS-affiliated data science and analytics teams, projects, networks, and communities.*

- **Category: Regulators**

Description: *Organisations responsible for setting, overseeing or enforcing standards, quality, and safety within healthcare, medicines and data interoperability.*

- **Category: Funders**

Description: *Organisations that provide financial support for research, innovation, and development across health, science, and technology sectors, including government agencies, research councils and philanthropic foundations.*

- **Category: Research Infrastructures, Health Data Holders and Gateways**

Description: *Projects, facilities, knowledge-based resources, services, and networks of data, technologies, and digital infrastructures that support research and innovation communities and industry in conducting research. Some organisations and platforms also collect, curate, and provide controlled access to UK health and biomedical data, enabling secure data sharing for research through governance frameworks, metadata catalogues, and access services.*

- ◇ Sub-Category: UKRI Digital Research Infrastructures
Selected initiatives funded through the UKRI Digital Research Infrastructure (DRI) Programme.

- **Category: Strategic Technical Platforms**

Description: *Strategic investments for systematic support, training and development to promote, enable and empower the Research Technical Professional community in UK universities.*

- ◇ Sub-category: Facilities

- **Category: Societies/Professional bodies**

Description: *Membership organisations that represent and support researchers and professionals within the biomedical data science community. They promote foster collaboration, create standards, provide training and sometimes accreditation.*

- **Category: Networks/Communities**

Description: *Collaborative groups of individuals or organisations that share common goals, expertise, or interests in advancing biomedical data science. While informal or not legally incorporated, they facilitate knowledge exchange, capacity building and coordination across sectors and disciplines.*

- **Category: Industry**

Description: *Companies and commercial organisations operating in biomedical research and data science, including those developing therapeutics, diagnostics, medical devices, software, and data-driven solutions, often collaborating with universities, research institutes and healthcare providers to translate research into practical applications.*

To note: mapping industry stakeholders of all sizes, from startups to multinationals, goes beyond the scope of this work. We have therefore made a conscious decision to focus on what we consider to be larger players. Startups have been mostly excluded from the map, as the UK biomedical startup ecosystem is continuously growing, but we included established incubator and accelerator programmes to reflect this.

◇ Sub-category: Incubator/Accelerator programmes

- **Category: Charities**

Description: *Non-profit organisations that fund, conduct, or support biomedical research and data-driven initiatives, often focusing on specific diseases, health conditions or public health priorities. They typically raise funds from the public and frequently collaborate with universities, research institutes, industry partners, and healthcare providers.*

3. Community engagement

We recognise that this is not work we can do alone, so we have organised and taken part in a range of activities to gather feedback from the UK health and biomedical data science community.

The first activity was a two-hour workshop held in June 2025, where we shared our early mapping and rationale and invited feedback and contributions. This workshop was invite-only, with 23 participants attending. The ABDC core team selected participants to ensure balanced representation across the categories included in the map. Invitations were extended through our wider network of partners and collaborators (past and present) and, where necessary, through targeted research and snowballing.

Additional feedback was gathered through the ABDC Advisory Board at its first meeting in September 2025. The board is made up of 9 individuals with an almost equal gender split, representing different sectors, from industry to research and academia, and spanning career stages from early career to senior leadership.

The ABDC team also collected internal feedback from colleagues at The Alan Turing Institute and EMBL-EBI and presented the work in progress at various conferences and events, including ISMB/ECCB 2025 (July), STEP-UP (July), CaSDaR (September), the HDR UK Conference 2025 (October), and the MRC Biomedicine Conference 2025 (November).

4. Defining biomedical data science

As shown above, the definition of biomedical data science guiding our mapping is very broad. A key focus of our stakeholder engagement activities has therefore been to explore the different definitions in use and consider where the boundaries of the discipline should be drawn.

Below are some of the comments we received during our workshops, which are helping us define the next steps.

- *Biomedical research has two main dimensions: the research side, which seeks to understand how the human brain, body, and mind function across large populations; and the clinical side, which focuses on developing treatments and running trials for specific diseases or groups. I see people working within either of these domains or bridging the space between them.*
- *Difficult... I work with omics, healthcare, social and economic data related to public health. Most data scientist are interdisciplinary in nature and will have worked with datasets coming from multiple domains, i.e. physics, chemistry, biology, engineering, etc.*
- *We define biomedical data scientists as someone who applies computational, mathematical or*

- *statistical skills in a biomedical research field this includes (and is not limited to) all researchers, educators, data stewards, research infrastructure or software engineers contributing or enabling biological, medical or health research that improves our understanding of health and disease.*
- *Biomedical data science is the use of data-driven methods to solve problems in biology and medicine, with a strong focus on human health. Its boundaries are defined by intent (health-focused), data type (biomedical), and method (quantitative and computational).*
- *It's difficult to draw a boundary, as the discipline is going to continue developing – new frontiers, new approaches, new capacities. It should include all people who work with data in any profession/ use case touching on biology. This is because animal research impacts healthcare, agricultural research touches healthcare as a food source, ecological research impacts healthcare as environmental factors.*

5. Next steps

With this first draft now finalised, we are moving the map to [Kumu](#), a tool that organises complex data into clear, interactive visualisations, ideal for mapping ecosystems, their connections, and their dynamics. Kumu will enable us to better capture the multiple layers of intersection between actors in the ecosystem, as well as the plurality of relationships that extend beyond the plain, two-dimensional representation of this initial draft.

Once complete, the map will be published and made openly available to disseminate our work and benefit the broader community. The public Kumu map will include organisations, relevant projects and programmes, and individuals where possible. It will not contain any sensitive information; names and roles will only be included if they are already public or if individuals have given explicit consent.

As this is an ongoing piece of work, we welcome feedback, comments and suggestions for improvement. **If you'd like to contribute to this work, please email your comments to: abdc@turing.ac.uk.**

6. Authors

- Bianco, Denise (Senior Research Community Manager for ABDC, The Alan Turing Institute)
- Tomba, Giulia (Senior Research Associate for ABDC, The Alan Turing Institute)
- Sokolova, Daria (Scientific Training Officer, EMBL-EBI)
- Koutsouroupa, Eirini (Programme Manager for ABDC, The Alan Turing Institute)
- Gurwitz, Kim (Training Manager and Project Co-Lead for ABDC, EMBL-EBI)
- Matser, Vera (Head of Skills and Project Co-Lead for ABDC, The Alan Turing Institute)
- Karoune, Emma (Principal Researcher and Project Lead for ABDC, The Alan Turing Institute)

We would like to acknowledge the contributions of the wider biomedical data science community and thank everyone who contributed to this work, including the ABDC Advisory Board and the other project teams working under the MRC Biomedical Data Science Leadership Awards.

7. Funding acknowledgment

This project is funded by the **MRC Biomedical Data Science Leadership Awards** (grant number MR/Z50662X/1). The views expressed are those of the author(s) and not necessarily those of the Medical Research Council. We would also like to acknowledge the in-kind contributions of The Alan Turing Institute and EMBL-EBI.