BIA policy submission

12 May2025

UK BioIndustry Association

BIA submission:

Data brokers and National Security

About the BIA and our members

BIA is the voice of the innovative life sciences and biotech industry, enabling and connecting the UK ecosystem so that businesses can start, grow and deliver world changing innovation. Our members include start-ups, biotechnology and innovative life science companies, large pharmaceutical companies, universities, research centres, tech transfer offices, incubators and accelerators, and a wide range of life science service providers: investors, lawyers, IP consultants, and IR agencies. We engaged with a range of organisations in collating this response, from large pharmaceutical companies to small techbio firms like PrecisionLife and Jiva.ai. We promote an ecosystem that enables innovative life science companies to start and grow successfully and sustainably. BIA have experience working closely with government, as part of the Vaccine Taskforce, on the Life Sciences Council with Peter Kyle, and on the Responsible Innovation Advisory Panel.

Life science is a growing sector of the future that poses a unique opportunity. The UK life sciences industry employs over 300,000 people, with around two-thirds of these jobs outside London and the South East. There are 6,850 life sciences businesses, 75% of which are SMEs, and combined they generate a turnover of £108.1bn.[1] The average GVA per employee is over twice the UK average at £104,000 and the sector consistently invests more in R&D than any other (£9 billion in 2022).[2]

Part 1: Definition and Services of Data Brokers

There is no universal definition of a data broker and the term 'data broker' and 'data broking' are not defined in UK law, though DSIT has developed a working definition of 'data broker' and 'data broking'. Broadly, DSIT considers data broking to be the practice of obtaining and trading or licensing data, data products and services to third parties, where the data can come from any wide range of sources and including where there may or may not be a direct relationship with data subjects. DSIT considers organisations conducting data broking activities to be data brokers, and in scope of this call for views.

[1] DSIT, DHSC, OLS: Bioscience and health technology sector statistics 2021 to 2022. (2023)
[2] ONS: Business enterprise research and development, UK: 2022. (2024)

UK BioIndustry Association

Does your organisation take part in any data broking activities?
No

Do you agree with the government's understanding of data broking as set out in Part 1 and that the term 'data broker' is best used to describe organisations that conduct this type of activity? If you disagree, what term(s) would you use to describe organisations that conduct data broking activities?
Partially agree. We are aware of different types of organisations that facilitate the sharing of data that may or may not be captured by this definition, but which are very different due to the ultimate purpose of their data sharing activities and the regulatory systems under which they operate. Including purpose within the definition, and examples of such organisations, may help interpretation.

There is a distinction between organisations that collect and share data for commercial or profit-making purposes compared to those whose primary remit is to support research and development (R&D) on a basis that may not be fully commercial or profit-making, which could include publicly-owned entities.

In health data the term "data custodian" is most frequently used to describe organisations that hold and grant access to data, especially as it is not always 'traded' for money. These may be analogous to data intermediaries described by the Department for Science, Innovation and Technology (DSIT) in the consultation. A data custodian acts as a steward, entrusted with the crucial responsibility of safeguarding, managing, and governing data according to legal and ethical guidelines on behalf of its owners. While not holding ownership, they bear significant accountability for adhering to stringent data governance policies encompassing security, access, and ethical use. Ultimately, the data custodian's aim is to facilitate responsible data use while upholding privacy, fostering trust, and ensuring regulatory compliance.

They often do this through "secure data environments" (SDEs) or "trusted research environments" (TREs) hosted and operated either by the "data custodian" or the data recipient under strict data security and governance frameworks. Examples include UK Biobank, Genomics England and NIHR Bioresource.

What distinguishes 'data brokers' from these custodians is that their primary remit is to make a profit. They operate for commercial purposes, engaging in the collection, aggregation, and sale or licensing of data, frequently without direct engagement with the individuals involved. While they generally don't own the data, their business model revolves around profiting from access to it.

Accountability tends to be more limited, often characterised by a lack of transparency where individuals may be unaware that their data has been traded or repurposed.

Both custodians and brokers play an important role in the life science sector and in our response we will refer broadly to all organisations involved in supporting data access. In general, it would be beneficial to develop clarity about these definitions so that the purpose, remit, and any responsibilities – regulatory or otherwise – of these organisations is made clear.

A final distinction is that of data processors within the data brokering ecosystem. These are infrastructure such as AWS Bedrock and other AI/data science service providers, which might import brokered data into their infrastructure including to build models, but should not be classified as data brokers, even if their services add value to brokered data or they partner with data brokers. This is because these organisations are not data controllers and do not make decisions about who has access to the data. These data processors still bear the responsibility of ensuring data safety and security, for which standards like ISO 27001, when properly implemented, are generally sufficient.

What social and economic impact do you consider the data broker market to have in the UK? Please consider both positive and negative effects.

Positive impacts:

1. **Improving patient access to innovation**

Access to high quality health data is vital for many purposes in the life science sector as it allows medicines and diagnostics to be developed with an understanding of real-life human biology, health and disease. This includes at various levels including development of new medicines, vaccines, diagnostics and devices and for all parts of the product development process, from initial discovery to clinical trials and clinical development, through to post-market authorisation, safety and effectiveness[3]. All medical innovations have been facilitated by access to health data in some form.

2. **Economic Value creation**

Life sciences industry investment in research and development (R&D) is a major driver of economic growth and tax revenue for the UK economy. The sector employs over 300,000 people across the UK and contributes over £108 billion in turnover to the UK economy[4] . The use of health data underpins industry R&D capability and is essential for the development of effective and safe innovations that benefit patients and improve care offered by the NHS, improving NHS efficiency.

[3] Techbio: UK leads innovation frontier 2025

[4] DSIT, DHSC, OLS: Bioscience and health technology sector statistics 2021 to 2022. (2023)

In addition, we see the use of data in the life science sector a key economic growth opportunity for the UK, in a subsector called techbio[3].

3. **Driver of research and academic excellence**

Finally, the UK's health data assets are a key driver of the UK's reputation as a centre for scientific and research excellence. For example, over 9,000 scientific papers have been published using UK Biobank data alone, including significant findings during the COVID19 pandemic[5] and the development of AI driven methods for identifying early markers of common diseases[6] [7].

Potential negative Impacts

1. **Erosion of public trust**

Given its importance to the life science sector, a negative impact – if data brokering is conducted poorly - is the risk that public trust is compromised. The sharing of NHS data for research purposes has caused concerns in the UK in the past as it was not communicated properly[8] . Public trust is also likely to be compromised in the event of a data breach, or the use of data that is not within the public's expectations. Therefore, it is crucial to always share and use data within public expectation and communicate effectively with the public about the governance and use of their data.

2. **Commercial exploitation**

Another risk of uncontrolled data brokering is that UK data is commercially exploited in a way that does not benefit the UK or UK population. For example, there are concerns within our sector that international companies with significant investment can outcompete UK companies using UK data. There is also a risk that companies might use data against the UK publics best interests. UK health data should be seen as a sovereign asset that should be exploited for national benefit. There is also a risk that without UK benefit public trust is further undermined.

3. **National security risks**

There are theoretical risks that bad actors may be able to access UK data for nefarious means, risking national security and public safety.

To address these various issues, it is crucial for the industry and government to demonstrate significant efforts in ensuring that health data is used responsibly and ethically to benefit society and drive medical research, whether through academic, public sector, or commercial entities. Ensuring that individual consent is respected, that their data is used for purposes with public

---

[5] Enabling scientific discoveries that improve public health – UK Biobank

[6] Garg, M., Karpinski, M., Matelska, D. *et al.* Disease prediction with multi-omics and biomarkers empowers case–control genetic discoveries in the UK Biobank. *Nat Genet* **56**, 1821–1831 (2024). https://doi.org/10.1038/s41588-024-01898-1

[7] AstraZeneca's new AI technology MILTON predicts more than 1,000 diseases before diagnosis

[8] Care.data: How did it go so wrong? BBC News

support, and that entities accessing the data comply with all applicable laws and data governance restrictions (e.g. GDPR) is vital, particularly in terms of geographical limitations on where the data is hosted, processed, and accessed.

Part 2: National Security Risks

Direct acquisition of UK data in an open market can be used as a pathway by hostile actors to harm UK national security. Harms can include gaining access to sensitive information, which can reveal insights pertaining to individuals, organisations or to government assets or UK data enabling hostile actors or strategic competitors to develop technologies that give them a strategic advantage.

To what extent are you/is your organisation aware of hostile exploitation of UK data and the extent to which this contributes to the risks outlined in Part 2?

- Aware

To what extent are you concerned about the collection and use of UK data by organisations conducting data broking?

- Concerned

Part 3: Security and Regulatory Frameworks

The government recognises that data brokers are already in scope of a range of security and privacy legislation. However these exist in a privacy context and are designed to protect individuals' personal data rights and were not designed to mitigate potential national security risks. Therefore, the government is considering what tools, beyond existing legislation, may be appropriate to strengthen itself against emerging data-related national security risks. This includes what processes, practices or policies already exist within the data broker industry to ensure the security of the data they handle.

Do you consider current legislation and regulations to sufficiently protect UK data from misuse? Please explain the reasoning for your answer.

While there is an abundance of regulation intended to protect individuals from being identified and data exploited (e.g. UK GDPR and Data protection Act 2018), this is sufficient primarily to protect individuals from being identified at an individual level. This leaves the potential that bad actors may gain access to population level UK data that has been uncontrolledly shared and use this for malicious means. Current data regulation was not designed to address the risks posed by large-scale data aggregation which are uniquely sensitive due to their identifiability, permanence, and population-specific insights. When aggregated and combined with other datasets, this data

could theoretically be used by hostile actors or strategic competitors to develop technologies that offer military, economic, or a biotechnological advantage, or to erode trust in UK systems. Even the claim that this could theoretically happen undermines public trust, demonstrating that robust oversight is needed. As an example, in 2024-2025 UK Biobank featured in a series of articles published by the Guardian illustrating several risks related to the potential misuse of health data by pseudo-science groups and foreign companies operating outside of the UK[9][10].

We acknowledge the need for regulatory control in this space, especially when data are being used outside the UK. Future regulations need to balance streamlined data access for bona fide researchers with the protection of research freedom while simultaneously addressing the aforementioned concerns. The UK is uniquely positioned with its valuable research resources, such as UK Biobank, NIHR, Genomics England, and East London Genes and Health, which rely on volunteer participation. It would be detrimental to the UK life sciences sector if future regulatory frameworks either create a perception of "uncontrolled" data access, thus discouraging public participation in similar future programs, or excessively restrict opportunities for research organisations to explore data and in doing so limit UK innovation.

In addition, it may also be worth considering a mechanism to incentivise data access and the ability to securely store copies of the UK health data for responsible and ethical companies operating in the UK which present a low risk from a national security perspective. This priority could also encourage companies to process data within the UK, thereby creating future job opportunities and spill-over benefits for the UK economy.

Do you believe there are sufficient standards within the data broking industry to ensure UK data is shared safely?

While there is an abundance of governance and regulation placed on the users of health data in the UK, which can prove a barrier to many well intended innovators[11], there are currently few standards – outside of GDPR - in place to regulate the practice of health data brokering. We advocate for improved standards of accreditation for SDE/TREs so that data can be stored and shared in a safe, secure and regulated manner. As an example, Our Future Health will only provide access to their data hosted within their centralised SDE/TRE, unless you can demonstrate that you can host your own accredited SDE/TRE for data transfer[12] . This mechanism means that the data never leaves an accredited secure environment but allows innovators to operate their own

[9] US startup charging couples to 'screen embryos for IQ', The Guardian, 2024
[10] Concerns raised over access to UK Biobank data after 'race scientists' claims, The Guardian, 2024
[11] The Sudlow Review 2024
[12] Accreditation process for trusted research environments open for applications, Our Future Health

SDE/TREs once they have been certified by an independent body. This gives essential flexibility to the life science sector, whilst upholding the highest standards of data security.

Part 4: Customer Base, Consumer Awareness and Transparency

The data broker industry is a complex ecosystem, and there is a lack of publicly available information profiling the industry's customers and main beneficiaries. Therefore, the government is keen to learn more about the customers buying UK data.

Have you ever been a customer of a data broker? If yes, what product(s) or service(s) did you use and for what purpose?

The BIA as an individual organisation has not been a data broker customer. However, many BIA members (biotech SMEs, pharmaceutical and diagnostic companies) apply for and access health data and use these to create innovative therapeutic and diagnostic products. There are various types of data used across the sector, including but not limited to: Electronic health records (EHR), Hospital episode statistics (HES), Longitudinal health data, Clinical trial data, Primary care data, Genomics data, Imaging data, Pathology data and social care data. This sits across a variety of brokers/custodians[13] including but not limited to: Genomics England, UK Biobank, Academic sources such as universities, the NHS, and Clinical Practice Research Datalink (CPRD).

As referenced before, access to high quality health data is vital for many purposes in the life science sector. This includes at various levels including development of new medicines, vaccines, diagnostics and devices and for all parts of the product development process, from initial discovery to clinical development, through to post-market authorisation, safety and effectiveness.

How aware are you of the data brokers industry and the role it plays in the data ecosystem?

- Aware

How much trust do you have in organisations conducting data broking for marketing, research or other purposes? Would this trust differ if you had more transparency about how your data is used?

Levels of trust in health data sharing is variable, as outlined by a recent DHSC public deliberation exercise[14]. In general, UK data custodians such as the NHS and UK Biobank enjoy high levels of public trust, but this is easily lost if the sensitive health data is not managed according to best practices and the subject's wish expressed through consent. There is less clarity on levels of trust in the practice of brokering in commercial organisations.

[13] Models of access to health data in the UK, ABPI, 2022
[14] National engagement on data: Cohort 1 report, 2025

In general, to maintain high levels of trust, there is an expectation that health data custodians publish public data use registers to maintain transparency (for example, UK Biobank's approved research register[15]).
This is important in maintaining public trust and awareness. However, these lists are not universal, not always easy to find, and do not always outline what happens to the data once a research project is completed.

How can organisations who conduct data broking improve transparency or awareness of their operations to instil confidence in consumers about data security?

Data brokers should be required to publish a comprehensive register detailing how data is used and for what purposes, including information on the researchers involved, their affiliations, and their locations. These registers should be easy to find and well-publicised to improve transparency. Additionally, data analysis should be conducted within accredited SDEs, with the option for centralised and locally operated SDEs/TREs, with thorough records maintained for auditing and monitoring to ensure data is used safely and legitimately.

SDEs are designed to facilitate the secure analysis of sensitive data, incorporating robust security measures such as restricted access, data encryption, and audit trails to protect against unauthorised access. Techniques like de-identification and anonymisation further safeguard individual privacy. By keeping data within a secured perimeter, SDEs significantly reduce the potential for data breaches and ensure compliance with data protection laws like GDPR.

Within the health data sector, where information is highly personal and confidential, SDEs build public trust and enable advanced research. The role of accredited user hosted SDEs is crucial in balancing the imperative of protecting sensitive information with the need to enable valuable and innovative data analysis techniques including AI. The life sciences sector supports both SDEs for secure and controlled data access. Developing clear regulatory frameworks in collaboration with academia, the public sector, and industry is essential to ensure that future data access models are effective and not obstructive to using data for patient benefit.

It is important to recognise that different research organisations have varying needs. Overemphasising restrictions and exclusive use of centralised SDEs diminishes the value of data and could lead to secure data being underutilised for life-saving research. Thus, establishing clear data governance and accreditation requirements is recommended. This would allow accredited organisations to host their own SDEs within the UK, facilitating necessary research freedom. The

[15] https://www.ukbiobank.ac.uk/enable-your-research/approved-research

accreditation framework should be technology-agnostic and align with existing industry standards to provide flexibility and reliability.

It is also worth noting that placing too many constraints on genuine researcher access can have the negative effect of reducing the scientific value of the resource which would go against the intentions of the consented participants who want to see advances in medical research emerge from the donation of their data. If we reflect on the resources that have had the greatest impacts on medical research advances and public health, they are significantly increased to those that do a better job at getting the data into the hands of genuine researchers in the UK and across the globe.

For further information about the contents of this submission, please contact Emma Lawrence, Head of data tech policy and public affairs, via elawrence@bioindustry.org