



UK BioIndustry Association

# Driving growth and patient benefit through secure data environments

*A UK life science SME perspective on  
technical and governance requirements*

December 2022

# Driving growth and patient benefit through secure data environments

## A UK life science SME perspective on technical and governance requirements

*This document was informed by the insight of members of the UK BioIndustry Association (BIA) - COHESION Medical, Jiva.ai, PrecisionLife, Benevolent.ai, Congenica Ltd, Human Centric Drug Discovery - but does not represent companies' opinions and is published by the BIA.*

## Introduction

The life sciences industry is a key pillar in the UK's innovation economy. Small to medium enterprises (SMEs) are critical to the growth of the industry and the economy; they turn our world-class academic insights into treatments for patients and products to export internationally. There are 6,548 life sciences businesses in the UK, which employ 68,900 people and generate £8.1bn of turnover<sup>1</sup>. 77% of these are SMEs which sit at the cutting edge of research and innovation. They are developing new tools and technologies that will revolutionise healthcare, save lives, and improve our health and wellbeing. They are, however, smaller and less resourced than large pharmaceutical companies, meaning they need special consideration by government and other organisations when designing policies and research infrastructure.

Global life sciences are increasingly reliant on access to health data for a variety of purposes, including drug discovery, safety testing, patient stratification and diagnostics development. Despite the importance of health data to life science research and innovation, the process of applying for access, and the data itself, is suboptimal and fragmented in the UK. Key industry requirements for health data have been published by the Association of British Pharmaceutical industry (ABPI) and the Medicines Discovery Catapult and include: data breadth, depth, and scale; speed of access; data quality; expertise; public trust; and affordability<sup>2</sup>.

All these needs can be facilitated by the appropriate use of secure data environments (SDEs)<sup>3</sup>, also called trusted research environments (TREs). SDEs are controlled environments where sensitive data can be accessed and analysed without the need to move or copy data into the researcher's data system. Well-designed SDEs can streamline data access while maintaining data security and privacy. The use of SDEs is particularly important in reassuring the public on data use where there are differing attitudes to commercial research amongst the wider public. They are also a proven model, having been shown to work successfully through infrastructure like SAILDatabank (Wales), the Scottish Data Safe Haven (Scotland) and Genomics England.

Both the public and SMEs are aligned on the need for ways to do life-saving research while protecting the privacy and respecting the choice of individual patients. SDEs support secure, trustworthy, and controlled access to data. As such, industry is supportive of them, providing they facilitate data access in a practical,

---

<sup>1</sup> OLS (2022), Bioscience and health technology sector statistics 2021: <https://www.gov.uk/government/statistics/bioscience-and-health-technology-sector-statistics-2021>

<sup>2</sup> MDC, ABPI (2019) Use of health data by the life sciences industry: a UK perspective: <https://md.catapult.org.uk/resources/use-of-health-data-by-the-life-sciences-industry-a-uk-perspective/>

<sup>3</sup> DHSC (2022) Secure data environments: <https://www.gov.uk/government/publications/secure-data-environment-policy-guidelines/secure-data-environment-for-nhs-health-and-social-care-data-policy-guidelines>

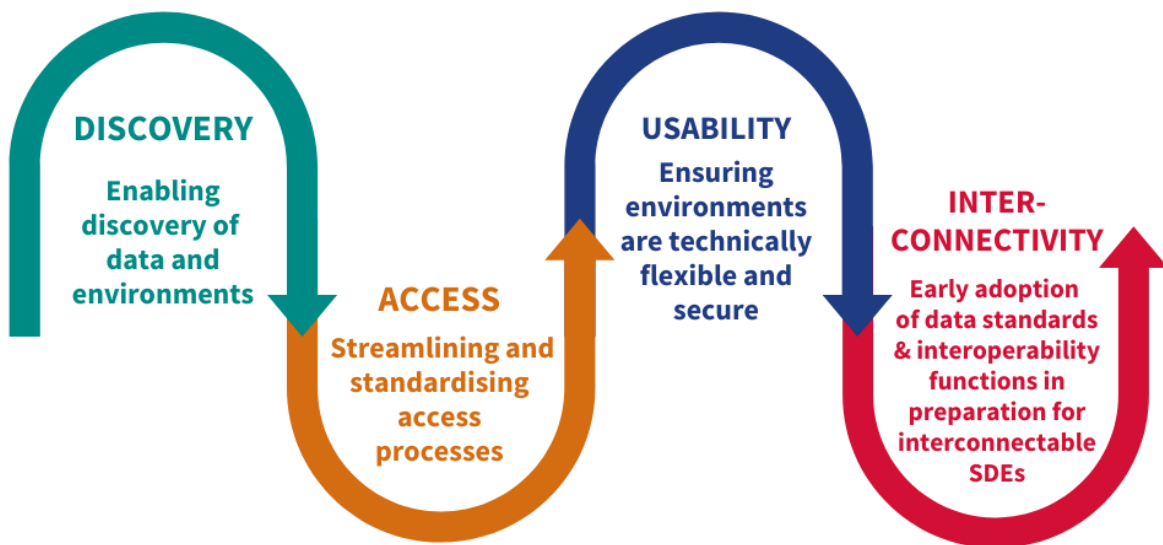
secure, and timely way. The recent Goldacre Review<sup>4</sup> as well as previous publications<sup>5</sup>, heralded SDEs as an answer to many of the concerns raised about health data access. However, the review did not cover industry access in detail.

Furthermore, SMEs face unique challenges in accessing health data, as they have more limited resources to navigate complex governance processes. These innovators also require greater consideration and resources from government and other data custodians to enable the full economic potential of research on genomic and health data to be realised. The move to SDEs presents an opportunity to address many issues associated with data access. This paper sets out what SMEs want to see from the new NHS SDE ecosystem, and as such complements the findings of the Goldacre Review and other recent data strategies. To fully realise the Government’s Life Science Vision and further grow the UK’s thriving data economy, use of SDEs should involve industry engagement. The suggestions made in this paper are for stakeholders involved in the development of NHS SDEs, although insights will also be valuable for those in academia, industry and the charity sector.

## Summary recommendations

*We recommend adopting a user-centred design approach, with users in industry having an ongoing role in the technical design of SDEs.*

*Key to this approach will be:*



<sup>4</sup> Better, broader, safer: using health data for research and analysis (2022): <https://www.gov.uk/government/publications/better-broader-safer-using-health-data-for-research-and-analysis>

<sup>5</sup> HDRUK Alliance, NHSX (2021) Building Trusted Research Environments - Principles and Best Practices; Towards TRE ecosystems: <https://zenodo.org/record/5767586#.Y4SIXhTP1Pb>

## Innovative SMEs in the life science sector

This paper is written from the perspective of SMEs in the life science sector and aims to inform SDE developers about that perspective. These companies require rich, clinically validated data and multi-omic data for novel analytical techniques, ranging from AI in diagnostics to precision medicine. Find out more about the BIA's Techbio<sup>6</sup> community from our recent reports<sup>7</sup>.

---

<sup>6</sup> BIA (2021), Techbio: <https://www.bioindustry.org/policy/strategic-technologies/techbio.html>

<sup>7</sup> BIA (2022), Techbio 2.0: unlocking data, transforming biology: <https://www.bioindustry.org/policy/strategic-technologies/techbio/tech-bio.html>

# 1. Discovery of data and environments

## 1.1 Findable data

**Problem: Data users do not know the characteristics of the data they are applying for.**

Applying for data access takes time and resource up front, therefore there should be clarity about what data will be made available and to what standard. It is unhelpful for data custodians to be optimistic about data characteristics. Clear metadata with objective measures is vital for innovators to understand what they are applying for.

To ensure clarity in what users are applying for we recommend:

### ***Being clear about data availability***

- Accurate data catalogues or dictionaries should sit outside the SDE, for example as in the HDRUK Gateway<sup>8</sup> or the National Cancer Registration and Analysis Service (NCRAS)<sup>9</sup>. A catalogue should show what data is available, for example: data points per person, number of people with full datasets and population characteristics (geographic, demographic etc). This metadata should follow an international open standard ontology (for example Mauro<sup>10</sup>).
- A representative synthetic data set, or the tools required to create such a synthetic data set, should be made available for users to check data quality and test and execute code.

### ***Being clear on what quality control has occurred***

- If data has been manipulated or adjusted to improve its quality, non-identifiable raw data should be available, with versions of cleaned data available with annotation from point to aggregated.
- Objective measures of data quality or details of any quality assurance run (i.e., amount of checking and what the check was) should be provided. Listing the dataset as ‘research ready’ or ‘quality checked’ is not sufficient. Data made accessible should have undergone basic quality control.
- Level of data completeness, either average or by field should also be shown.

### ***Being clear about what the data has been used for***

- The data should be referenced and tagged to demonstrate the prior usage of the data, see Health Data Research UK (HDRUK)’s recommendations for a health data use standard<sup>11</sup>.
- The usage logs should be searchable across SDEs, so that reviewers can find out if research has been performed before, leading to faster decision making by review boards.

### ***Being clear about what the data can be used for***

- Details of patient consent, including if consent for research involving private companies was obtained. This should link to consent forms, information sheets, and ethics approval numbers.
- Further usage restrictions such as location of analysis, small number suppression, and permitted use or users should be displayed.

<sup>8</sup> HDRUK gateway: <https://healthdatagateway.org/>

<sup>9</sup> National Cancer Registration and Analysis Service (NCRAS):

[https://www.datadictionary.nhs.uk/nhs\\_business\\_definitions/national\\_cancer\\_registration\\_and\\_analysis\\_service.html](https://www.datadictionary.nhs.uk/nhs_business_definitions/national_cancer_registration_and_analysis_service.html)

<sup>10</sup> Mauro: <https://maurodatamapper.github.io/about/introduction/>

<sup>11</sup> HDRUK (2022), Improving transparency in the use of health data for research: Recommendations for a data use register standard: [https://zenodo.org/record/5902743#.Y4S\\_gXbP1PZ](https://zenodo.org/record/5902743#.Y4S_gXbP1PZ)

## 1.2 Environment capabilities

**Problem: Users do not know what functionality and performance to expect from environments.**

The SDE ecosystem should be transparent and accountable. As such there should be mechanisms for measuring the success or value of a SDE, to foster a dynamic ecosystem. Data users should be able to understand what the environment offers in terms of functionality and performance.

To show the capability of any given SDE, we recommend publishing the following key performance indicators (KPIs)

- **Data** – including volume, velocity, variety, variability, veracity, visualisation, and value.
- **Capabilities** – including what features and functions are available.
- **Certifications** – including what compliance and standards have been achieved.
- **Services** – descriptions of what services are available including any platform downtime. For example, Genomics England displays this publicly online<sup>12</sup>.
- **Accessibility**, including
  - The number of biotech companies making access requests.
  - The turnaround time to a positive or negative decision.
  - The proportion of biotech companies gaining access.
- **Satisfaction/Feedback** – For example, net promoter score (NPS) of biotech companies
  - On access procedures.
  - On delivery of access.
- **Measures of impact** on UK health and wealth, including
  - Medical advances achieved via data access.
  - Company growth driven via data access.

---

<sup>12</sup> Genomics England status page <https://genomicsengland.statuspage.io/>

## 2. Data access

### 2.1 Governance

**Problem: The process of applying for access to health data is confusing and opaque.**

Complex access governance models for controlled access data are particularly challenging for SMEs. Whereas larger companies make use of legal or regulatory professionals to navigate this process, in SMEs, this is often left to senior staff due to resource constraints. Gaining access to the environment or the data can frequently involve filling out several forms (in some cases requiring physical signatures), delays while waiting for the data access committee to meet, and requests for additional information, extending the timeline further. A survey of UK SMEs found that only 25% of data access attempts were successful and slow turnaround times can incentivise companies to go abroad for data<sup>13</sup>. Cancer Research UK has also highlighted access time discrepancies, showing two similar projects experienced a 5-month difference in data access approval times. Delays in data access have held up Cancer Research UK-funded projects for up to 2 years<sup>14</sup> and this is representative of the SME experience. It is therefore hard for organisations with limited time and resources to confidently schedule work and is burdensome to make and manage the requests.

The health data research UK (HDRUK) alliance has established a data access and governance steering group<sup>15</sup> to work on this area. Following consultation, the alliance has published a harmonised data access request form<sup>16</sup>. We recommend the adoption of such standard processes to improve data access. The introduction of SDEs allows for the introduction of clear access criteria, application system, expected time frames and requirements. These application processes include those to gain access to the SDE and those between the data providers (e.g., NHS trusts) and the applicant. Clear **service level agreements** (SLAs) outlining the application process will allow monitoring of access times.

To improve the process of applying for data and SDE access we recommend ensuring the process is **predictable, fair and time-bound** by:

- **Using service level agreements** so that users are clear about what service they can expect.
- **Publishing the decision-making process**, ensuring applicants can see clear details of the basis for the decision and examples of why applications would be or have been refused.
- **Using a flow diagram** to show the application process with timelines and decision-makers. This should be linked to application document templates. Data custodians should publish escalation processes if timelines are not met by the data provider. See example flow diagram and page layout used by UK Biobank<sup>17</sup>.
- **Transparency of data access approval meetings**, by both broadening attendees to include patients, public and users (as recommended in the Goldacre review), and by keeping and

<sup>13</sup> MDC, ABPI (2019) Use of health data by the life sciences industry: a UK perspective: <https://md.catapult.org.uk/resources/use-of-health-data-by-the-life-sciences-industry-a-uk-perspective/>

<sup>14</sup> Cancer Research UK (2020) Unlocking the potential of Data to Transform Cancer outcomes: [https://www.cancerresearchuk.org/sites/default/files/cruk\\_vision\\_for\\_data\\_jan\\_20\\_1.pdf](https://www.cancerresearchuk.org/sites/default/files/cruk_vision_for_data_jan_20_1.pdf)

<sup>15</sup> HDRUK alliance (2022): Data access and governance: <https://ukhealthdata.org/projects/data-access-and-governance/>

<sup>16</sup> HDRUK alliance (2022): Five safe data access request application form: <https://zenodo.org/record/5946892#.Y4nQZsvP1Pa>

<sup>17</sup> UK Biobank access process <https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access>

publishing notes. Commercially sensitive information will need to be redacted or protected by agreements signed by attendees.

- **Using standardised agreements** where data is not centralised. This would mean that where an applicant is establishing access with several organisations, e.g., NHS trusts, one application process can be used.

## 2.2 Accreditation

***Problem: The process of applying for access to health data is time-consuming and repetitive.***

Applying for data access can lead to duplication of information, with the same information being requested by different stakeholders at different stages of the same process, or when applying for different data sets to perform the same analysis. To streamline access processes and following the ‘five safes’ model<sup>18</sup>, we advocate people, organisations and projects being accredited where possible, based on agreed criteria. The Government’s Data Saves Lives strategy<sup>19</sup> also references accreditation of the SDEs themselves, which would support a full framework of accreditation across the system as referenced in the Digital Economy Act<sup>20</sup>.

To avoid repetitive and timely application processes, we recommend:

- **Accreditation of users** and organisations similarly to the authentication of data safe haven users (see the UCL data safe haven <sup>21</sup>). This would show that an individual has had the appropriate training and experience to perform the analysis in any given environment. This would also facilitate streamlined access to multiple data collections.
- **Risk-based project approvals**, that focuses more resource on uncertain or novel projects. Projects which are low-risk or fit certain criteria would undergo a streamlined approval process. Public involvement would be key to this process.
- **Centrally controlling this process** so that consistent information is collected once and then supplied to individual SDEs and data access committees. This would also facilitate general suspension or revocation of a user or organisation’s access from all SDEs where needed (for example: data breaches, changes in rules and change of consents). Where central control is not possible, standardised information should be collected so that applicants can utilise the same information at different places.

<sup>18</sup> ONS (2017) The ‘five Safes’ – Data privacy at ONS (2017): <https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/>

<sup>19</sup> DHSC (2022) Data saves lives - reshaping health and social care with data <https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data/data-saves-lives-reshaping-health-and-social-care-with-data>

<sup>20</sup> Digital economy act: <https://www.gov.uk/government/publications/digital-economy-act-2017-part-5-codes-of-practice/research-code-of-practice-and-accreditation-criteria>

<sup>21</sup> UCL data safe haven: <https://www.ucl.ac.uk/isd/it-for-slms/research-ig/articles/data-safe-haven-assurance>



## 3. Usability

### 3.1 Supporting proprietary analytical pipelines

***Problem: Limiting the technical ability of data environments will prevent advanced analysis techniques being executed in the data environments.***

The Government has published its policy principles for SDEs<sup>22</sup>. These guidelines set out expectations for how SDEs will be used to access NHS health data. One of the guidelines states: “Secure data environments must be able to support flexible and high-quality analysis for a diverse range of uses”. SDEs should therefore be optimal for innovators in the life science sector, including SMEs. If an SDE limits how data can be used or analysed by design they become barriers to innovation and not fit for purpose as enablers of innovation. Access to environments should be under a clear **service level agreement**, which outlines the technical service provided.

The ability to support the deployment of innovative proprietary analytical pipelines (PAP) on flexible computational platforms should be available to all users. PAPs require analytics (and third-party data) to be docked within SDEs to deliver innovation which does not impact reproducible analytical pipelines (see figure 1).

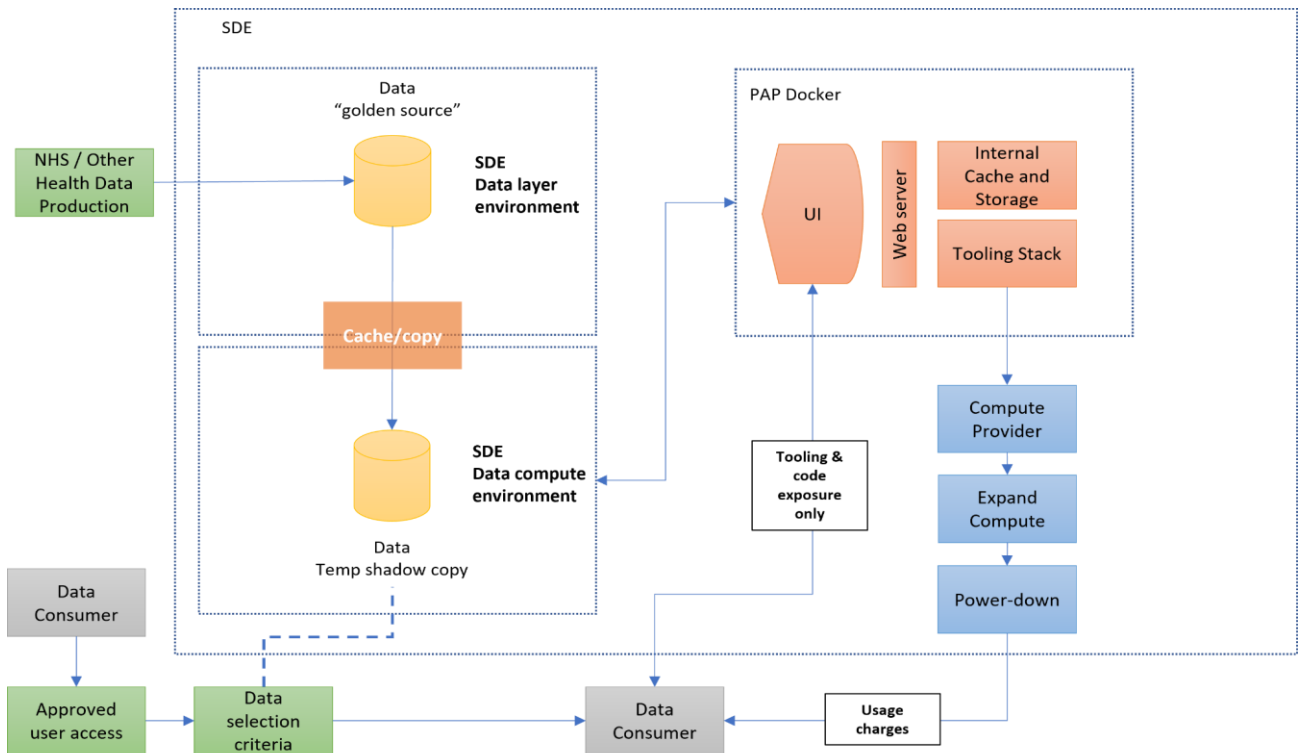
To support the use of proprietary analytical pipelines we recommend:

- **Flexibility of compute environments;** many modern analytical techniques require different processing units such as graphics processing units (GPU), central processing units (CPU) and tensor processing units (TPU). We recommend supporting different types of compute, as well as providing an environment (e.g. Docker) where software could be deployed.
- **Elasticity of compute resources** to allow for uneven compute load across the SDE at any point in time. For example, AI training requires high performance compute, but the compute requirement is less once the project moves to validation (see case study 2). In addition, the types of analysis that will come online in the future cannot be designed into the system now. This will need professional resource deployment management under a service level agreement. Costs for these services should be agreed up front. Expandable and elastic compute infrastructure should be done via a multi-cloud provider. A single cloud provider should be sufficient at the proof-of-concept stage so long as there are the API keys to grow compute resources on demand. Access to compute should be within an NHS-approved firewall.
- **Open standards and operating systems** such as linux environments, with either a process for ensuring that OS level libraries are made available and maintained, or that users of the environment can install standard packages/libraries as required.
- **Clear governance for proprietary tools,** outlining the method for import, use and vetting of software. SDEs should provide assurance that the system will only be accessed or used by those authorised by the contributor. An agreement on the availability of the output for egress from the

<sup>22</sup> DHSC (2022) Policy guidelines for NHS SDEs: <https://www.gov.uk/government/publications/secure-data-environment-policy-guidelines/secure-data-environment-for-nhs-health-and-social-care-data-policy-guidelines#secure-data-environment-guidelines>

system should be used. A timely review and release process is particularly important for AI developers (See case study 2).

- Support for end users**, even experienced data analysts will need support when using new environments. This includes both at an administrative level, to support with payments and access and at a technical level, to support with functionality. Support should be in the same time zone as users and be covered by a service level agreement.



**Figure 1: A high level solution for a typical PAP environment.** The strategy by which data is cached and copied is up to the SDE administrator; importantly, the required selection of data might be too large for typical egress/ingress, moreover the data is required not to leave the security of the SDE. Therefore, the SDE should provide the means for PAPs to both expose their technical capability via a UI and API (which would be assessed during the certification phase) as well as the ability to expand and contract compute.

# The need for proprietary analytics

## Case study 1: Genomics data analysis

*There are a multitude of ways to understand and dissect the human genome. Genomic data can be in the form of genome wide association studies (GWAS) array genotyping, whole exome sequencing (WES), whole genome sequencing (WGS) as well as next generation DNA sequencing applied to measure gene expression and epigenetic regulation. Genomic data can also be linked to other high content omics data types such as metabolomics, proteomics and microbial genome analysis. As the cost of sequencing and other multi-omic technologies continues to fall, these types of data sets will become more common. Full analysis of the genetic and genomic data requires integration with the clinical phenotyping, ultimately to establish gene to disease and trait relationships. This supports future discovery of diagnostics, underlying molecular mechanisms of disease and potential drug target hypotheses and is directly linked to UK innovation in Biotech.*

*While many studies make use of GWAS and polygenic risk scores (PRS) other genomics analysis techniques are under development. Therefore, any genomics data environment should provide for a full range of analytical methods as described in supporting proprietary analytics. Linking genomics to clinical phenotyping data will be of particular value.*

### 3.2 Security

***Problem: there is a risk of proprietary code or personal data being accessed by unapproved third parties.***

The SDEs must be able to demonstrate that they are secure to allow maximum use by industry and endorsement by the public. Commercial innovators will be uploading commercially sensitive information and intellectual property (IP) into the environment and as such risk losing their competitive advantage should insights be lost or shared. Adequate protection for commercially sensitive code and other intellectual property should therefore be given to SDE users.

To provide industry with assurance on security we recommend:

- **Adopting high standards of cybersecurity:** Industry would expect SDEs to be separately audited by an independent body (e.g. Cyber Resilience Centre for Wales or the National Cybersecurity Centre) against best practice. SDEs should undergo periodic CREST approved penetration testing. Environments should also show that they have cybersecurity measures in place without compromising their systems.
- **Governance for IP protection;** including contracts in place to provide guarantees that IP is protected within an SDE. Providers should also give assurance that systems will only be accessed or used by those authorised by the contributor. Auditable logs to demonstrate who has accessed project environments should be available.

## The need for proprietary analytics

### Case study 2: Artificial Intelligence and machine learning

*Machine learning and AI techniques are becoming widely used across the life science sector and within healthcare. These techniques require high performance compute at variable load throughout the development process. For example, AI training requires high performance compute but once the project moves to validation then the compute requirement is less. The SDE needs to provide this sort of load to individual projects during the training phase and then reallocate it for other users. In addition, different processing units such as GPU and CPU are needed as outlined above. Raw clinical data should be quality checked or level of quality control displayed up front.*

*For AI development many iterations of code upload and export may be needed during the training period. Any lengthy review and release processes will severely hinder SMEs working on AI. Therefore, an efficient, appropriately automated ingress and egress system should be in place (see ingress and egress section).*

## 4. Interconnectivity

### 4.1 Ingress and Egress

***Problem: Too restrictive a data environment will not facilitate timely exchange to and from environments.***

The value of data is enhanced considerably when it is supplemented with data from outside the original data source. Third parties should be able to add data and value to the SDE from data sources that are not traditionally captured by the NHS or social care environments, such as a charity or trade body, and potentially directly by citizens themselves. In addition, external stakeholders may be able to enrich or supplement the data in other ways, for example by improved data curation or return of results. Users will also need a clear mechanism for uploading proprietary code.

Once data users have finished analysis, results or outputs will need to leave the SDE. It should be clear before work has started what data outputs will be permitted to be removed, to ensure that organisations can realise the value they expect from projects. The airlock mechanisms should be sufficiently automated for low-risk projects to prevent delays in return of results. This process is particularly important for the use of AI (see case study 2).

To facilitate ingress and egress to and from the environment, we recommend:

- **Providing governance mechanisms** for data to be added, linked or downloaded from the SDE, for example organisational accreditation. Clear rules on what type of outputs are allowed out of the system should be agreed up front.
- **Providing technical mechanisms** for data or outputs to be added, linked or removed from the SDE, for example connection ports. There should be sufficient automation of the airlock process on low-risk projects.
- **Building an industry-driven ecosystem of SMEs** which can support third party organisations by providing tools for connecting, analysing and managing data contributors to SDE compliant standards.

### 4.2 Standards and interoperability

***Problem: unstandardised datasets and environments cannot be easily analysed or linked, leading to fragmentation.***

To support large scale data analysis, researchers require access to pan-UK data, as well as the ability to search across data sets to assess data availability. Combining data from different sources allows more insights to be generated from existing data assets. The NHS, government and other stakeholders should work towards **standardisation, interoperability, and federation across** SDEs. Establishing local or subnational data environments will only work if there is a clear path to interoperability. Many efforts are already happening internationally, such as through ELIXIR<sup>23</sup> and The Global Alliance for Genomics and

---

<sup>23</sup> ELIXIR: <https://elixir-europe.org/>

Health<sup>24</sup>. There are also plans to create a more standardised European Health data space<sup>25</sup>, so this interconnectivity is vital for the UK to remain competitive.

Enabling pan- UK data analysis through federation or other means, will bring huge opportunities to the UK. Most importantly it will improve patient outcomes through the ability to conduct more complex analysis or to study rarer diseases. This approach leads to competitiveness driven by a marketplace, driving improvements in quality and standards. This would facilitate industry- academic collaborations, provide greater flexibility, remove duplication, and take advantage of the benefits of scale and uniformity.

The higher statistical power gained from pan-analysis will have a direct impact on UK SMEs. The UK could become more competitive in this area and harness existing strengths by bringing in these standards to make this field more investible. In addition , NHS trusts engaging with research opportunities will have the chance to improve revenue and healthcare outcomes. Stimulating research and innovation through the potential of the NHS data assets will therefore bring benefits to the NHS and beyond.

To ensure UK data assets can be used optimally we recommend:

- **Standardisation of data**, through using internationally recognised standard data models (e.g., OMOP, openEHR and Mauro).
- **Standardisation of functionality** to support a trusted framework of interoperability. For example, through the introduction of standard APIs like FHIR (Fast Healthcare Interoperability Resources)<sup>26</sup>
- **The ability to search or link data across SDEs** though interoperability and the use of identifiers.
- Ensuring that analyses can be performed across the whole of the UK population.
- Ensuring that data can interconnect with international data sets.

### 4.3 Reproducible analytical pipelines

***Problem: Standard reproducible analyses are repeated by different stakeholders using different methods.***

As outlined in the recent Goldacre review, ensuring reproducible analytics pipelines (RAPs) are available to users will reduce duplication of work and support more standardisation in healthcare analysis. RAPs work well for NHS analysts as they create more efficiency and standardisation, however they are not as valuable for commercial companies whose competitive advantage is their own analytics. Innovators performing analysis using proprietary analytical techniques should not be forced to share code or IP in the same way. For commercial users these analytical techniques are the source of their competitive advantage, and they will therefore need to be protected (see security section).

<sup>24</sup> G4GH: <https://www.ga4gh.org/>

<sup>25</sup> The European health data space: [https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space\\_en](https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en)

<sup>26</sup> NHS Digital, FHIR (Fast Healthcare Interoperability Resources): <https://digital.nhs.uk/services/fhir-apis>

To promote efficiency and standardisation in data analysis we recommend:

- Managing RAPs by an overarching data body or SDE - so that there is no difference in the analytical capability between the various SDEs. Reusing existing platforms such as Orcha<sup>27</sup> will further facilitate standardisation and interoperability.
- Providing an **optional** ability to share proprietary analytics.

## Exceptions to SDE use

Even if environments are developed to the highest technical and governance standards, there will still be use cases that may require data to be exported from the secure environment. An example of this includes international clinical trials, where data from many countries need to be linked and patients have consented for this to happen. Patient and public views on the acceptable compromise between patient impact and SDE use should be heard and acted on. Environments should be of the above technical standard to limit the number of exception cases. The introduction of too many exception cases may lead to an erosion of public confidence in the system.

## Summary

The life sciences industry is an important part of the health ecosystem, developing tools and technologies to support health and social care. Innovators rely on timely access to health data to develop these cutting-edge innovations, in the same way as they rely on access to finance or an adequate skills base. To fully realise the Government's Life Science Vision and further grow the UK's thriving data economy, access to health data should be facilitated through industry co-developed SDEs.

Industry is supportive of the use of SDEs to enable secure, controlled, and trustworthy data access. However, these environments should be developed in collaboration with SMEs to ensure they are fit for purpose, and do not become a barrier to using data for patients' benefit. **If the focus of attention and development is solely on restricting access and use, then there is a risk of investing millions in totally secure data that cannot be used for life-saving research.** Clear access governance and accreditation requirements along with the highest technical specification will facilitate timely data access and public confidence.

---

<sup>27</sup> Orcha Health: <https://orchahealth.com/>


## About the BIA

The BioIndustry Association (BIA) is the trade association for innovative life sciences and biotech industry in the UK, counting over 500 companies including start-ups, biotechnology, universities, research centres, investors and lawyers among its members. Our mission is to be the voice of the industry, enabling and connecting the UK ecosystem so that businesses can start, grow and deliver world-changing innovation.

BIA's members are at the forefront of innovative scientific developments targeting areas of unmet medical need. This innovation leads to better outcomes for patients, to the development of the knowledge-based economy and to economic growth. Many of our members are small, pre-revenue companies operating at the translation interface between academia and commercialisation.

BIA has a growing number of members working at the interface of patient and health data and innovation, including using analytic models, machine learning and AI. The ability to work on cutting edge innovations with minimal bureaucracy is of real importance to these companies.

For further information, please visit [www.bioindustry.org](http://www.bioindustry.org)

A large, thick, teal-colored curved graphic that starts from the left edge of the page and curves upwards and to the right, ending at the right edge. It is positioned above the contact information.

**Contact Dr Emma Lawrence, Senior Policy and Public Affairs Manager at the BIA, on [Elawrence@bioindustry.org](mailto:Elawrence@bioindustry.org) or 07880009251 for more information.**