Unlocking NHS data for research: how to improve the regional Secure Data Environment network

February 2025

Contents

abpi

BIA UK BioIndustry Association

ABHI HealthTech for Life

Executive summary

Unlocking the wealth of data within the NHS for research, via a network of harmonised Secure Data Environments (SDEs), has the potential to provide major benefits for researchers, patients and the healthcare system. This ambition aligns with the government's missions of growing the economy and improving the health of the population. However, realising this ambition is dependent on designing and implementing infrastructure and services that support research requirements.

Life sciences industry investment in research and development (R&D) is a major source of revenue for the UK economy. 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. For a network of SDEs to be of value to the life sciences industry, it must provide services that meet industry's R&D needs.

This report outlines the essential requirements from the 11 regional SDEs within the Data for R&D Programme, referred to throughout the report as 'the regional SDE network', to meet industry research needs.

The report describes common requirements shared by all companies and provides an overview of how companies of different sizes and parts of the life science sector use health data for research, and how their requirements can differ as a result. The report also reflects industry's perceptions and experience of their engagement in the development of the regional SDE network to date.

Budgets for the Data for R&D Programme are tight, and the technical and operational challenges to deliver the regional SDE network are considerable. It is therefore vital that government budgets are prioritised to deliver the features considered most important by users, including industry. Involving industry in co-developing the regional SDE network is essential to maximising the value that can be derived from the network.

Insights in this report offer NHS England, the Department of Health and Social Care (DHSC), and the Department for Science, Innovation and Technology (DSIT) the opportunity to better align implementation of the Data for R&D Programme's regional SDE network with the needs of industry researchers, as key contributors to the UK life science sector and the economy. We therefore hope that, in the lead up to the Comprehensive Spending Review (CSR), this report will inform future prioritisation of the regional SDE network, helping the government achieve its vision of utilising NHS data to improve health and grow the economy.

Although this report sets out recommendations for the regional SDE network, a recurrent message from industry was the importance of making comprehensive GP data, linked to nationally collected secondary care datasets, available for research. Although a centralised national collection of GP data is outside the scope of the regional SDE network, it is vital that this pressing priority is taken forward by government if the UK is to truly realise the value of health data to benefit patients' lives, stimulate innovation and attract inward investment.

The Association of the British Pharmaceutical Industry (ABPI), in partnership with Health Data Research UK, have outlined a potential approach to delivering a health data research service providing comprehensive GP data in a recent analysis, Options appraisal to deliver a health data research service in England.

This report presents the results from a consultation with members of the ABPI, the Association of British HealthTech Industries (ABHI) and the BioIndustry Association (BIA). Members who use health data for research and have experience of the Data for R&D Programme were invited to take part in the consultation. In total, 30 member companies completed a survey, and 21 of these companies also took part in a follow-up interview. The recommendations below are distilled from the survey and interview findings.

Recommendations

1. Focus on delivering achievable outcomes to ensure that the most essential, high-value datasets and services are prioritised.

It is vital that the programme's budget is spent on key priorities that meet the needs of researchers. This report details the key service requirements of different industry user profiles, based on sector and size. This information should be used to inform the prioritisation of programme implementation to maximise return from the investment made into the regional SDE network. This prioritisation must be realistic about what is achievable within the timeframe of the programme to build confidence among researchers.
This report shows that all life sciences companies that completed the survey use datasets – such as those within the regional SDE network – for discovery science. We welcome the ambition to extend the utility of datasets and services in the regional SDE network for a range of additional research applications. However, we recommend that, within the current financial and operational constraints of the programme, the regional SDE network focuses on providing services for discovery science, AI development, precision medicine and local public health and health services improvements that are best supported by the rich datasets held within the regional SDEs.

2. Provide clarity about the types of research best supported by datasets held at a local trust or regional SDE, or at national level.

Access to data that is demographically and geographically representative of the English population is essential when research outcomes must be generalisable to the whole UK population or populations in other countries. In particular, the capability to bring together representative data is crucial for both multi-national pharmaceutical and health technology industries, where research findings are used for regulatory and health technology assessment (HTA) processes in global markets. Currently, to fulfil these requirements, our members access nationally collected datasets for regulatory and population analyses through the Clinical Practice Research Datalink (CPRD) and the NHS England SDE, which is also part of the NHS Research SDE Network. These representative population datasets are different from the rich, multimodal datasets that will be held within the regional SDEs. Our members have emphasised the value of combining datasets from across multiple SDEs, leveraging the value provided by locally led insights with the scale and diversity of multiple SDE populations. However, achieving full federation of datasets across the regional SDE network is likely to be difficult in the short to medium term. Despite this, common data standards and enabling architecture should be put in place now to realise this ambitious goal in the future. Regardless of timeframes, it is important to recognise that different datasets are used for different research purposes and as such, distinct infrastructure may be needed to support these applications. As recommended by the Sudlow Review, a broader review of national data infrastructure to match varying research requirements is needed to support the breadth of industry and non-commercial research uses.

3. Provide a unified, integrated service, with a single point of entry to the regional SDE network, to streamline user experience and reduce duplication of effort.

Industry values clarity of service offerings and efficient delivery of these services in a consistent manner. To achieve this expectation, there must be a single entry point to access all data from across the network, supported by a service catalogue of data availability. It is not realistic to expect commercial researchers to engage, negotiate and contract with 11 separate regional SDEs. The regional SDE network needs to present a combined offering, avoid internal competition and generate economies of scale. The network should collaborate to share best practice, expertise and feedback intelligence of market needs centrally to develop an integrated strategy that responds to user priorities. By focusing on provision of a unified coherent service, the regional SDE network will offer a commercially compelling user experience.

4. Adopt consistent and harmonised information governance, contracting and pricing processes within commercially competitive timeframes for data access within the regional SDE network.

Businesses operate under commercial delivery pressures. Uncertainties caused by long or inconsistent supply timeframes result in companies choosing alternative data suppliers, often in other countries where their needs can be better met. Respondents indicated that commercially competitive timeframes would require data access request decisions to be made within 25 working days of submission, with access to the data provided within 10 working days following an approved request. Respondents stated that variability in information governance, contracting and pricing processes lead to industry investing more time, money and effort into commissioning and conducting research projects than is necessary. As a consequence, companies could take their research elsewhere. Standardised data access processes must therefore be adopted across the regional SDE network if it is to offer an internationally competitive service.

5. Form an external advisory group, with industry represented, to ensure that users are placed at the heart of strategic decision-making.

User-centred design is at the heart of all well-configured programmes. Health data research services are no different and need to be designed to meet the needs of users of the service. Although there has been some engagement with industry by the regional SDE network, members felt that their input had not been taken into account in subsequent implementation. Members expressed the urgency of meaningful engagement with industry in programme design and implementation. Without this meaningful engagement, members were not confident that the network would meet their needs.

To address this, the Data for R&D Programme should implement an external strategic advisory group that has a formal role in advising the Programme Board. This formal relationship would ensure that the board has sight of discussions that take place at the group and, where appropriate, group members could present at board discussions. It is also essential that there is consistency of operational processes and technical standards across the regional SDE network. The Programme's Community of Practice should convene groups with subject matter experts, including industry representation, to advise on the technical, governance and contracting aspects of the regional SDE network. There should be transparency of how advice from these groups has been acted on by each of the regional SDEs.

Context

Access to high-quality health data at scale is a cornerstone of the development of life sciences innovations. Development of new medicines, vaccines, diagnostics and devices is reliant on data-driven approaches for all parts of the product development process, from initial discovery to clinical development, through to post-market authorisation, safety and effectiveness.

While the UK has some world-leading health data assets, overall it is a fragmented ecosystem. Professor Cathie Sudlow's 2024 review, Uniting the UK's Health Data: A Huge Opportunity for Society, described in clear terms a system where issues around data access, linkage, and quality are inhibiting crucial health research about conditions affecting millions of people across the UK. Professor Ben Goldacre's 2022 review Better, broader, safer: using health data for research and analysis and the DHSC's 2022 strategy Data saves lives: reshaping health and social care with data set out a vision to unlock the power of NHS data to improve health by bringing together health datasets within new regional SDEs. These SDEs would facilitate access by researchers while protecting patient data, generating and linking together previously inaccessible datasets from NHS hospitals and services in all regions of England. This regional SDE network was envisaged to be transformational in helping generate health insights that would support the development of new life science products and could represent a significant competitive advantage for the UK in being an attractive place for the life sciences industry to invest in health data research.

For the regional SDE network to reach its full potential, it must provide data and services to standards and timeframes that align with business needs in a commercially competitive global environment. This can only be achieved by developing and implementing these features in partnership with the life sciences companies that will be significant users of the network. If research users are absent from the design and implementation process, there will inevitably be a misalignment between the services and infrastructure that are developed and the needs of users. This misalignment will reduce the research value of the regional SDE network, inhibit health research, and cause companies to look to other countries to fulfil their health data research needs.

Funding for the regional SDE network currently runs until spring 2025, by which time there is an expectation that the network will already be offering a range of data and services for users. With the next CSR funding cycle rapidly approaching, it is a timely opportunity to review the regional SDE network's implementation from an industry perspective, with a view to inform future plans, ensuring the network can deliver for industry as well as non-commercial users.

The memberships of the ABPI, the ABHI and the BIA trade associations (TAs) include companies of all sizes across the biopharma and medtech life sciences sector. Together TA members represent most industry researchers that will potentially use data in the regional SDE network. TA members have expressed concerns that the design and implementation of the regional SDE network have been carried out without sufficient meaningful engagement with life sciences companies. As a result, there is a perceived mismatch between expectations of the data and services provided by the regional SDE network and the needs of industry research users.

In response to members' concerns, the TAs conducted a consultation exercise with members to understand their health data uses, service requirements to support commercial needs, and their expectations of the regional SDE network meeting these needs. In addition, members were asked to suggest priorities and improvements for the regional SDE network to increase its utility and attractiveness for the commercial research sector. The findings of this exercise are outlined in this report.

The recommendations in this report reflect the views of the TAs and our members on the necessary course correction to ensure the regional SDE network fulfils its ambition and is fit for purpose for life sciences companies.

Methodology

The TAs conducted a consultation exercise with their members via a survey and interviews. Companies invited to take part were existing health data research users and had some knowledge of the regional SDE network and its planned implementation. Questions in the survey encompassed key data uses and technical, governance and contracting requirements for industry-level standard data services, as well as their perceptions of the programme. Questions in the interviews explored recommendations for the regional SDE network, as well as the attractiveness of the network in a globally competitive environment. Recommendations were distilled from the survey and interview findings.

Throughout this report, 'the regional SDE network' is defined as the 11 regional SDEs established as part of the NHS Research SDE Network overseen by the Data for R&D Programme. The 'Programme', when capitalised, refers specifically to the Data for R&D Programme. Unless otherwise noted, all charts shown represent the responses of the 30 companies who responded to the survey. Charts illustrating companies' views on a Likert scale used a five-point scale from 1 (not at all important) to 5 (essential). Quotes from interviews are used illustratively and represent the views of a single company.

Survey and interview questions

Members were invited to complete an online survey. The questions in the survey were designed to gather an understanding of the profiles of companies across the three TA memberships, in terms of how they use health data, their technical, governance and contracting requirements from the regional SDE network to meet their business needs, and their confidence in the regional SDE network meeting those needs. The survey questions can be found in the report Annex.

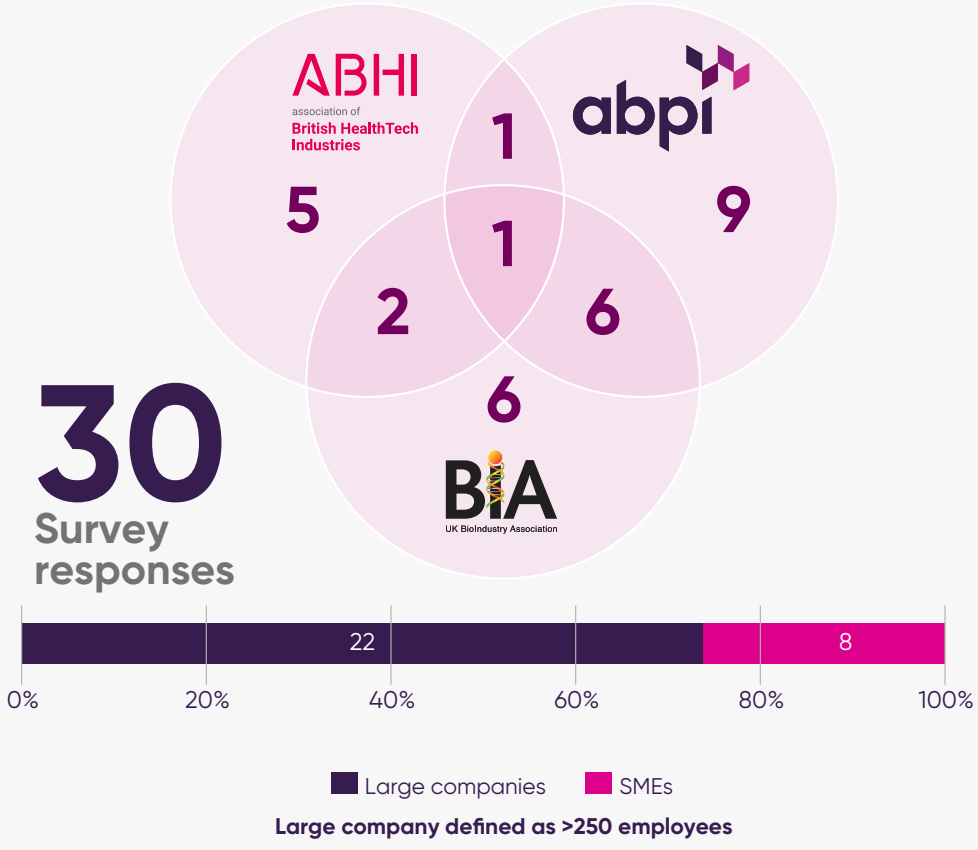
Companies who completed the survey were invited to participate in follow-up interviews. The interview questions below were designed to enable the three TAs to develop recommendations that reflect the needs of our members and ensure that the ongoing implementation of the regional SDE network would result in a fit-for-purpose and globally competitive system.

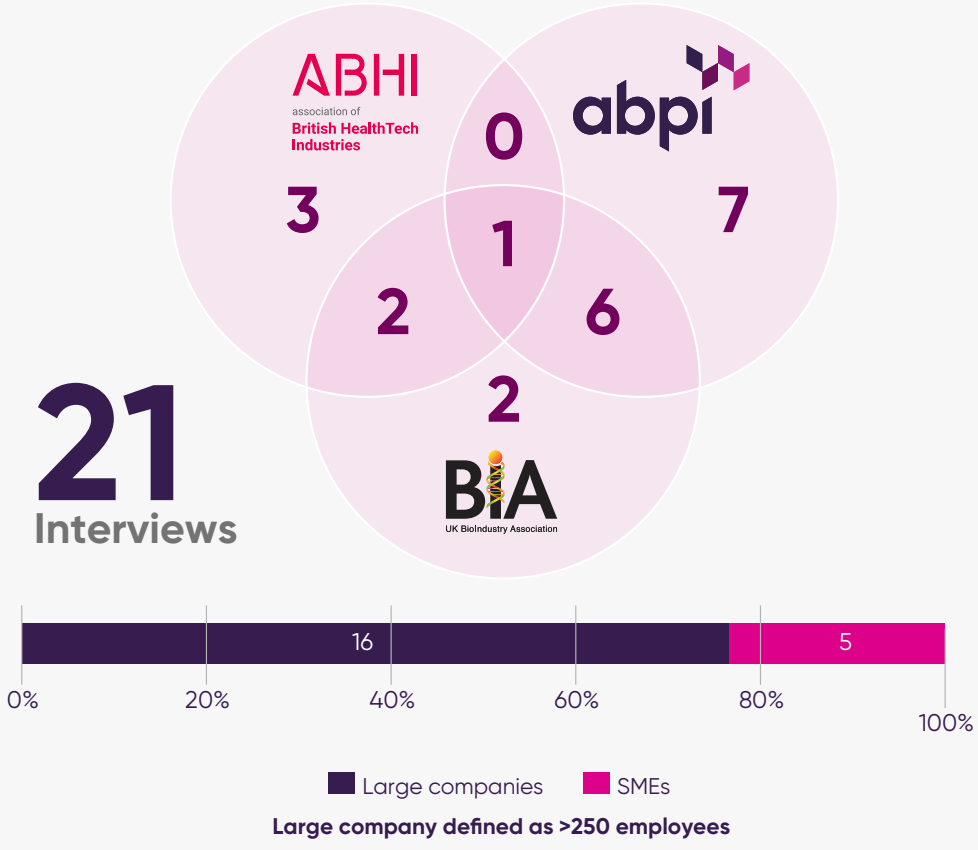
1. What recommendations would you make to improve the planned implementation of the SDE network and ensure it creates a fit-for-purpose internationally competitive offering?
2. What are the top three areas that the SDE network should prioritise?
3. Are there any aspects of the implementation of the SDE network that, if not resolved, could lead to your organisation choosing to preferentially carry out data-enabled R&D in countries other than the UK?
4. What would increase your organisation's confidence in the ability of the SDE network to meet your business needs?

Profile of respondents

Figure 1 shows a breakdown of respondents according to TA membership, including those who are members of more than one TA, and by company size. Thirty companies across the three TA memberships responded to the consultation survey. Of these companies, 23 self-reported as large companies (>250 employees), and seven as small and medium-sized enterprises (SMEs) (<250 employees). Twenty-one companies went on to take part in an interview to explore recommendations for the regional SDE network, of which 16 interviewees were large companies and five were SMEs.

Figure 1 A breakdown of the respondents by TA membership and company size (large companies defined as >250 employees).





Life sciences companies use health data for research across the product development life cycles. However, uses of data vary by company size and sector. Respondents were divided into four profiles based on a combination of self-reported company size and sector(s) (Table 1).

Table 1 Company profiles derived from self-reported company size and sector(s).

Profile	Employees	Countries of operation	Business sectors
Small and medium-sized biotechnology and pharmaceutical companies ('SME biotechs', n=3 for survey, n=1 for interview)	**<250, ranging from microenterprise** (up to nine employees), **to medium-sized** (up to 249 employees)	**Often UK-based or UK-market-focused**	**Pharmaceutical, vaccines and associated life sciences products and services**
Small and medium-sized medical technology companies ('SME medtech', n=5 for survey, n=2 for interview)	**<250, ranging from microenterprise** (up to nine employees), **to medium-sized** (up to 249 employees)	**Often UK-based or UK-market-focused**	**Devices or diagnostics, including data-driven approaches and digital solutions**
Large, multinational pharmaceutical companies ('Large pharma', n=16 for survey, n=13 for interview)	**>250, often thousands or tens of thousands**	**Global in operation and markets**	**Pharmaceutical, vaccines and associated life sciences products and services**
Large, multinational medical technology companies ('Large medtech', n=6 for survey, n=5 for interview)	**>250, often thousands or tens of thousands**	**Global in operation and markets**	**Devices or diagnostics, including data-driven approaches and digital solutions**

Consultation findings

How the life sciences industry uses health data

Figure 2 shows reported data uses across the pipeline of product R&D as a proportion of respondents in each of the four company profiles defined above. The data uses are abbreviated in Figure 2 for brevity. The full definitions of the requirements as included in the survey are in the report Annex. The findings clearly demonstrate that the uses of data vary by the size and business sector of life sciences companies.

All four company categories reported using data for the discovery and development phases of life sciences product development. This is consistent with the product development pipeline of life sciences SMEs and larger companies, described elsewhere including in the ABPI's, the BIA's and the Medicine Discovery Catapult's 2024 State of the Discovery Nation report. In particular, SME biotech and SME medtech companies indicated a focus on discovery and development use of data (Figure 10 and Figure 11 – Annex). In contrast, no SMEs reported post-authorisation safety as a current use case. Less than 50 per cent of SMEs consulted reported using data to support regulatory, HTA, real-world effectiveness, population health and health service delivery research. Further details of data use by biotech SMEs are described in the BIA's 2024 TechBio report.

In contrast, more than 50 per cent of all large companies surveyed reported the use of data across all 17 use cases surveyed. More than 75 per cent of large companies reported using data for discovery, regulatory and HTA purposes, population health and post-authorisation safety. These findings reflect data uses by large pharmaceutical companies described in the ABPI's Unlocking the promise of UK health data report.

In addition to data uses across the medicines and vaccines development pipeline, interestingly, 50 per cent of large pharmaceutical companies indicated that they also use data for the development of non-pharmaceutical interventions. Approximately 30 per cent of these companies indicated that they use data for the discovery and development of HealthTech-associated products (defined as devices, diagnostics and digital) (Figure 8 – Annex). This observation reflects the recent shift in approach among pharmaceutical companies in developing and using AI and digital technology approaches for discovering new targets and stratifying patient responses to treatments, and for safety signal detection, among a growing number of other applications. This finding shows that the traditional boundary between the pharmaceutical and medtech sectors is blurring, as companies utilise new technological approaches to develop new medicines and vaccines. Medtech companies are also using these data for the development of AI tools for use in therapeutic or diagnostic applications.

Similarly, approximately 30 per cent of large medtech companies indicated that they use data for pharmaceutical R&D, which may be in areas such as companion diagnostics or tools to support large pharmaceutical companies' medicines R&D pipelines (Figure 9 – Annex).

Reported data uses by company profile

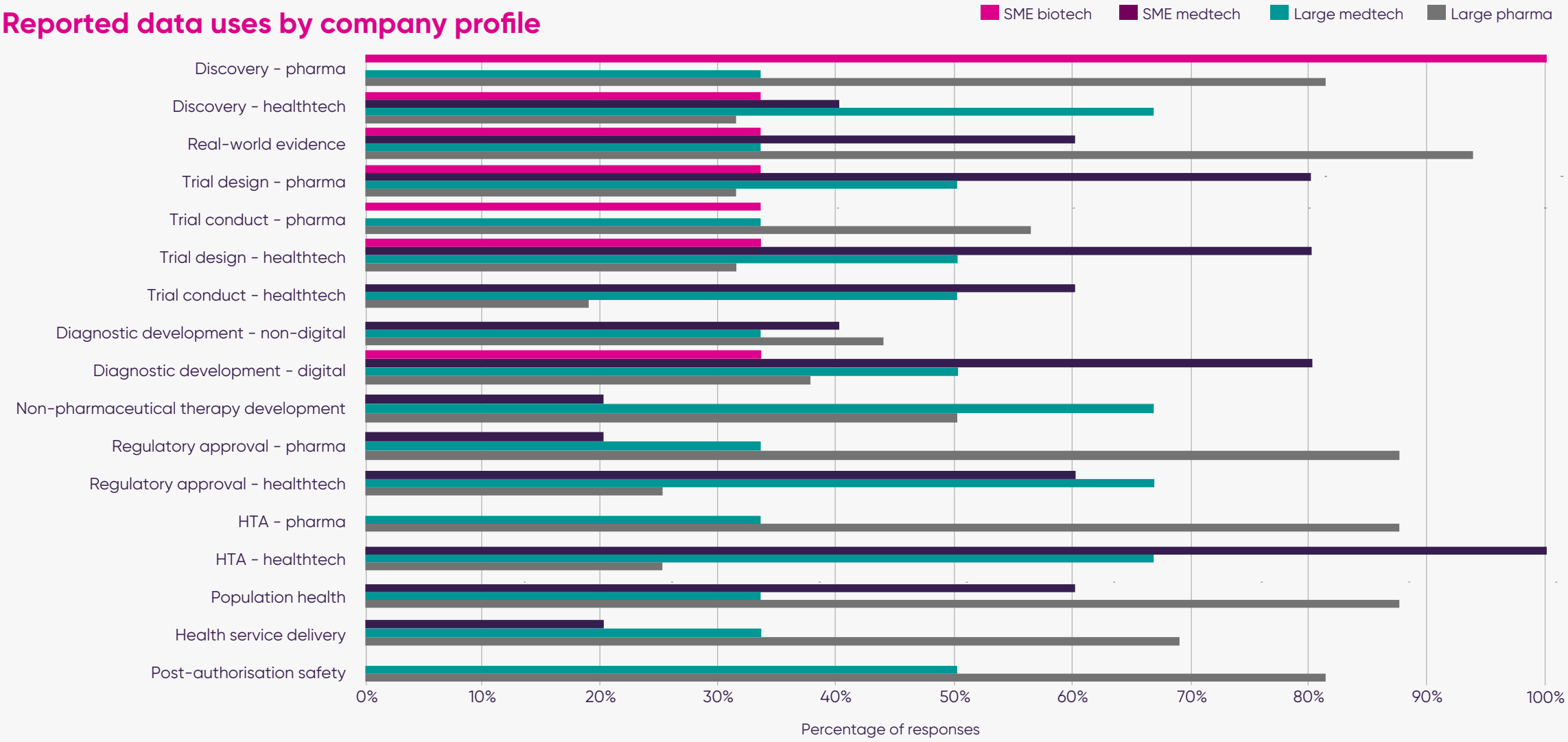


Figure 2 Reported use cases by company profile.

Volume of data required to support data uses

Different uses of health data require different types and scales of data. Many of the uses of data shown in Figure 2, in particular for regulatory, HTA and post-authorisation safety applications by large medtech and pharmaceutical companies, require scalable representative data. This regulatory requirement of national and international regulators is necessary to ensure results can be generalisable to the target population.

Each regional SDE encompasses a discrete geographical area, with populations between 2.8 million (Wessex) and 10.5 million (London). The geographical nature of SDEs means they are subject to inherent biases due to local variability of NHS product adoption or difference in local patient care provision pathways, as well as regional demographic profiles. By their very nature, no single regional SDE can provide the scale or prerequisite data representativeness that will enable generalisability of findings to whole populations for most product licencing, HTAs and post-market safety monitoring applications.

To validate this assumption, companies were asked whether the data available in a single SDE would be sufficient to meet their data use needs. As shown in Figure 3 (opposite), only SME biotechs indicated that a single SDE would be sufficient. Sixty per cent of SME medtechs indicated 'not sure'. In line with use cases in Figure 2, more than 80 per cent of large pharmaceutical and medtech companies reported that the data available in a single SDE would not meet the majority of their data uses.

Companies were also asked to rate the importance of being able to access data federated across multiple SDEs. As shown in Figure 4, most large medtech and large pharmaceutical companies and SME biotechs rated the importance as very important or essential (a 4 or 5 on the Likert scale).

Would a single regional SDE be sufficient to meet the majority of your company's business needs

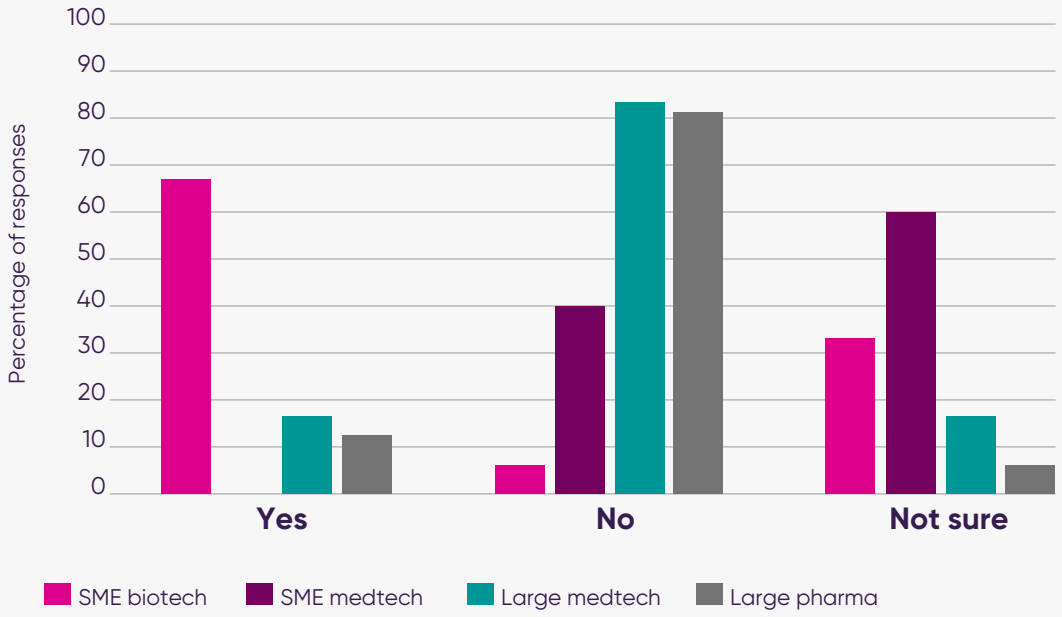


Figure 3 Companies' views on whether data within a single SDE would be sufficient to meet their business needs.

Importance of being able to access data federated across multiple regional SDEs

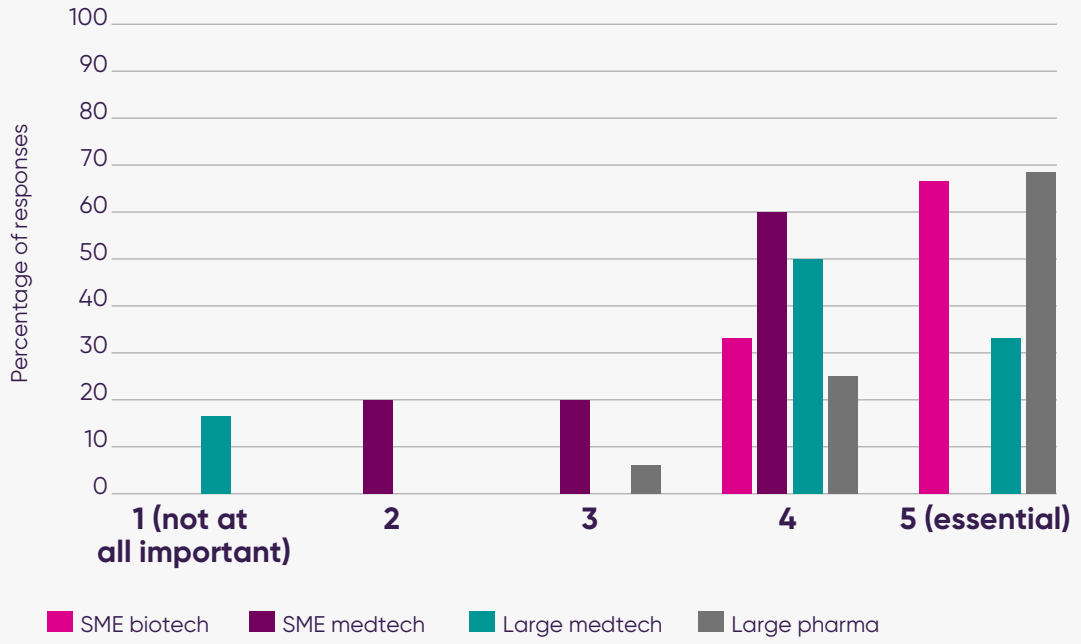


Figure 4 Companies' views on the importance of being able to access data federated across multiple regional SDEs.

Across the survey and interviews, all companies expressed the added value of scale when conducting analyses on health data, regardless of the data use. Taken together, these findings confirm that companies will need, at a minimum, access to federated data across multiple SDEs to fulfil most of their data uses. However, except for a small number of pilot projects, the current roadmap for the regional SDE network will not enable analyses of data across multiple SDEs by spring 2025, and no timeframe has been given for when routine federated analyses will be possible across the network.

> "The top priority is to link c.3 SDEs from across the country to get a representative sample of the population."
> **large pharma company**

> "Single SDE projects might be helpful but when it comes to commercial/regulation could mean not representative and therefore not acceptable."
> **large medtech company**

In interviews, companies told us that GP data linked to secondary care data is a key resource for supporting their research needs. Longitudinal GP records, which provide an overview of the medical histories of the UK population, are a rich source of health insights and are enhanced further when linked to hospital data.

Currently, large companies source representative GP and secondary care data of the UK population through the CPRD, which has approximately 30 per cent coverage of UK GP data, linked to secondary care data for English patients. While the data within the CPRD is not comprehensive of the entire UK population, it is sufficiently representative of the diversity of the population to be used for important use cases. However, companies told us that the ambition should be to achieve nationally comprehensive GP data coverage to unlock new benefits for R&D.

We welcome the efforts of individual regional SDEs to augment the value of their datasets by incorporating GP data into their service offerings. However, the broader aspiration to provide comprehensive GP data through federation across the entire regional SDE network is unlikely to be achievable for technical and operational reasons. Instead, nationally comprehensive GP data could be better achieved through a centralised, national data collection that sits within other national data infrastructure. This could be delivered, for example, through the expansion of the CPRD, which has a proven track record for delivering services for industry, academia and government researchers alike, or through a new national health data service. The ABPI, in partnership with Health Data Research UK, have outlined a potential approach to delivering a health data research service providing comprehensive GP data in a recent analysis, Options appraisal to deliver a health data research service in England.

> "Linking primary and secondary data longitudinally is the top priority."
> **large pharma company**

When considered alongside the data uses indicated by the four company profiles, the consultation findings do suggest that regional SDEs should be able to support the discovery science needs of companies, which may be deliverable without the need for the scale and representativeness offered by multi-SDE analyses. For example, a regional SDE should be able to provide high-quality, multimodal datasets combining unstructured and structured data such as digital pathology, medical imaging, multi-omics and health records. These data can provide a richness and depth unparalleled across any individual health dataset. Deriving insights from such complex, bespoke datasets will require close partnership between the NHS and researchers, utilising clinician-led insights. This type of intensive dataset preparation is suitable for bespoke discovery research but is less suitable for other routine uses of data by industry that require representative data at scale. It is also less suitable for the training of AI models, where scale and diversity of data are essential.

In the future, if the regional SDE network is routinely able to offer federated data analyses at scale, the network is likely to be utilised for an increasing range of data uses. However, in the short to medium term, the regional SDE network may be better suited to directing efforts and limited budgets to supporting research that their data is best amenable to delivering, such as discovery science, AI development, precision medicine, and local public health and health services improvements. National data access policies will need to be adaptable to the progress of the regional SDE network and ensure alternative routes to access to data remain available. Concurrently it is critical that the investments into infrastructure, governance and services taking place now provide the platform for future growth and capabilities. In practice, this means that the regional SDE network's infrastructure, data standards and linkage methods must be future-proofed for emerging and potential future use cases and data federation.

RECOMMENDATION ONE

Focus on delivering achievable outcomes to ensure that the most essential, high-value datasets and services are prioritised.

It is vital that the programme's budget is spent on key priorities that meet the needs of researchers. This report details the key service requirements of different industry user profiles, based on sector and size. This information should be used to inform prioritisation of programme implementation to maximise returns from the investment made into the regional SDE network. This prioritisation must be realistic about what is achievable within the timeframe of the programme to build confidence among researchers. This report shows that all life sciences companies that completed the survey use datasets, such as those within the regional SDE network, for discovery science. We welcome the ambition to extend the utility of datasets and services in the regional SDE network for a range of additional research applications. However, we recommend that, within the current financial and operational constraints of the programme, the regional SDE network focuses on providing services for discovery science, AI development, precision medicine, local public health and health services improvements that are best supported by the rich datasets held within the regional SDEs.

RECOMMENDATION TWO

Provide clarity about the types of research best supported by datasets held at a local trust or regional SDE, or at national level.

Access to data that is demographically and geographically representative of the English population is essential when research outcomes must be generalisable to the whole UK population or populations in other countries. In particular, the capability to bring together representative data is crucial for both multi-national pharmaceutical and health technology industries, where research findings are used for regulatory and HTA processes in global markets. Currently, to fulfil these requirements our members access nationally collected datasets for regulatory and population analyses through the CPRD and the NHS England SDE, which is also part of the NHS Research SDE Network. These representative population datasets are different to the rich multimodal datasets that will be held within the regional SDEs. Our members have emphasised the value of combining datasets from across multiple SDEs, leveraging the value provided by locally led insights with the scale and diversity of multiple SDE populations. However, achieving full federation of datasets across the regional SDE network is likely to be difficult in the short to medium term. Despite this, common data standards and enabling architecture should be put in place now to realise this ambitious goal in the future. Regardless of timeframes, it is important to recognise that different datasets are used for different research purposes and as such, distinct infrastructure may be needed to support these applications. As recommended by the Sudlow Review, a broader review of national data infrastructure to match varying research requirements is needed to support the breadth of industry and non-commercial research uses.

Business needs for a competitive service

Across all companies consulted, we heard a common set of business service requirements, irrespective of company size or sector, to support R&D in a globally competitive environment.

1. Technical requirements for using secure data environments

Companies were asked to rate 10 different technical requirements for using secure data environments to support the development of their products, according to a five-point Likert scale, with 5 being 'essential' and 1 being 'not at all important'. The requirements are abbreviated in Figure 5 (opposite) for brevity. The full definitions of the requirements as included in the survey are in the report Annex.

All 10 criteria were rated highly across all companies (Figure 5). However, some requirements were rated more highly than others.

Relative importance of technical requirements

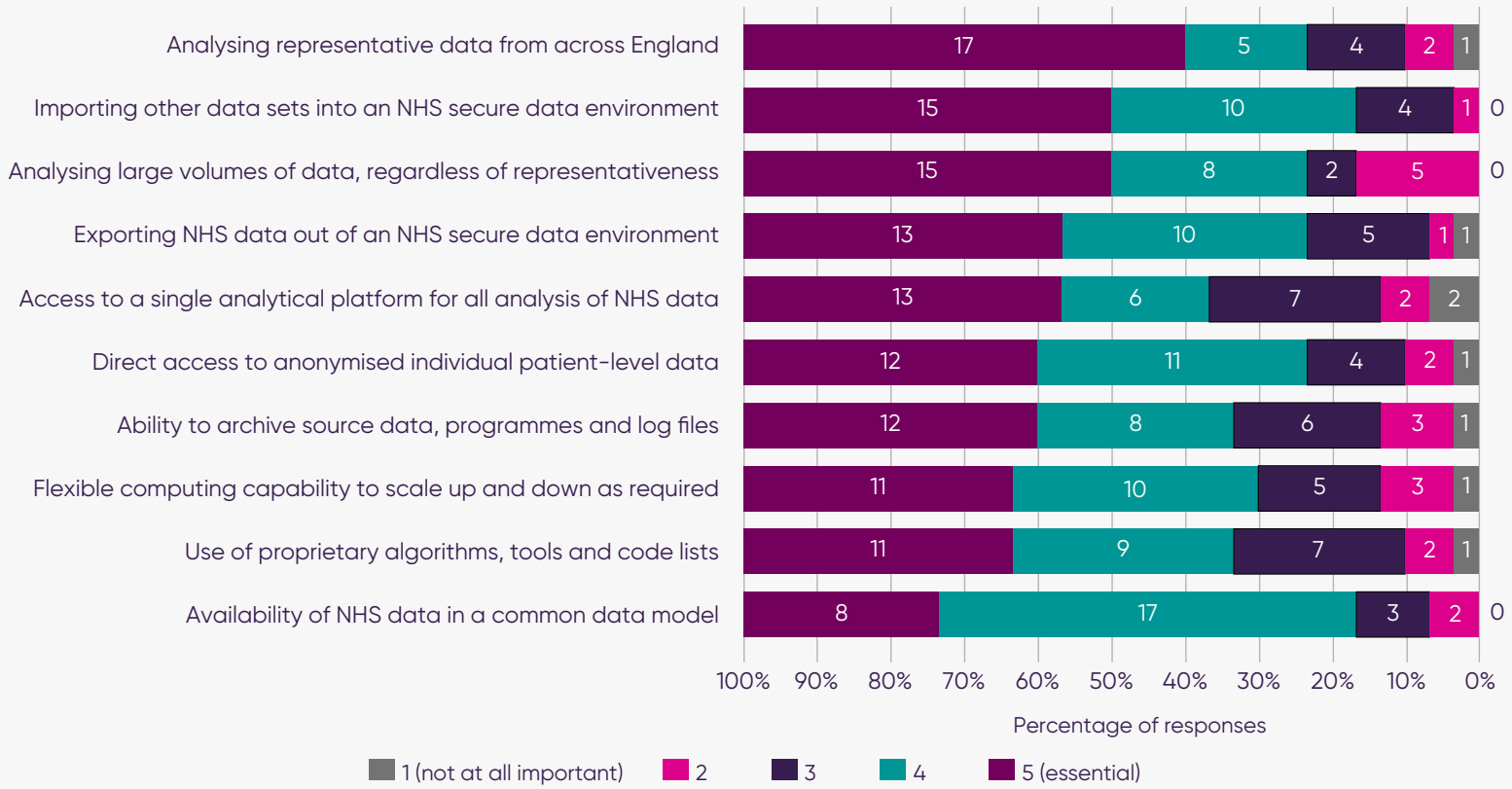


Figure 5 Importance of different technical requirements.

The single highest priority from companies was for the technical ability to conduct analyses on representative population data from across England. This is consistent with the previous finding that access to federated data from multiple SDEs would be needed to support most industry R&D data applications.

The findings also show that a range of other technical solutions would need to be implemented if the regional SDE network is to provide a viable service offering to industry. More than 80 per cent of companies rated the need to import other data, which includes non-NHS or propriety datasets, into an SDE as very important or essential. The regional SDE network's current roadmap for the implementation of technical features does not include the ability to import data as a short-term priority, suggesting a mismatch of prioritisation between the network and users.

More than 70 per cent of companies indicated that it was essential or very important to be able to export NHS data from an SDE. This reflects the need for large companies to comply with regulatory requirements or to combine NHS data on patients who have consented to participate n clinical trials into a united, often multi-country, trials database. Recent updates on the DHSC's data access policy suggest that data export will be supported where there is a technical or regulatory reason for doing so.

Importantly, no requirement listed in Figure 5 was considered less than very important by 70 per cent of all companies, regardless of their size or sector.

In interviews, some companies suggested additional technical requirements not included in the survey, such as the provision of synthetic datasets and technical pilot projects that demonstrate the practical feasibility of conducting complex projects.

> "SDEs should develop synthetic datasets to enable data analysis feasibility testing within the SDE, and to allow users to develop analytics and algorithms freely that can then be applied to the real data."
>
> **SME biotech company**

Overall, it is clear that many of the technical requirements companies need to use secure data environments for their projects are not current priorities for the initial launch of the regional SDE network in spring 2025 and may not be in place for years to come. It is likely that a lack of consultation with users has led to this disparity and means that, regardless of other aspects of the network being ready, many companies may be unable to use the network. This will inhibit valuable research and calls into question the financial sustainability of the network if it cannot provide services for the commercial sector.

2. Commercially competitive timeframes for data access

Companies were asked to indicate what they considered to be commercially acceptable timeframes for each of the three key steps involved in conducting research within a regional SDE.

The median time companies expect a data access request decision to be made is within 25 working days (30 mean working days) after submission of a request. Following an approved request, companies expect a maximum

of 10 working days (17 mean working days) between the decision being made and being able to access the data to begin the research project. Once analyses have been completed, companies expect a maximum of 10 working days (15 mean working days) for results to be released from an SDE.

In interviews, companies stressed that current timeframes to access health data across the UK are not viable for businesses working within commercially competitive environments and undermine their ability to conduct research. We heard of examples of data access applications taking many months, or longer, leading to projects being delayed or cancelled outright. Professor Cathie Sudlow's 2024 review, Uniting the UK's Health Data: A Huge Opportunity for Society, similarly found that accessing UK health data can often take months or years.

In addition, we heard how important predictability and consistency of timeframes were for business planning. Protracted or inconsistent timelines harm the competitiveness of the UK system compared to other countries. Companies told us that the regional SDE network offers the opportunity to speed up data access compared to existing processes used by other health data infrastructure.

> "If the regional SDE network cannot get data access speed to the times [discussed], [the company] could choose to deliver data R&D in other countries."
>
> **large pharma company**

> "Speed of access is key. Industry cannot work with six-month data access delays and prefers clarity/consistency on timelines."
>
> **large pharma company**

3. Governance, contracting and pricing requirements for the regional SDE network

In creating 11 new parts of health data infrastructure, the regional SDE network may be perceived as adding to the complexity of an already fragmented UK health data ecosystem. This is exacerbated by a lack of harmonisation of processes across the network, leading to different approaches emerging across the SDEs. As each regional SDE is at a different state of maturity, some have well-established practices that they are reluctant to change, and others are building processes from scratch. This approach is leading to duplication of effort across the regional SDE network, with parallel processes needlessly developed independently of each other.

While the Programme is working to ensure some consistency and harmonisation across the network, this may fall short of what is needed by companies to effectively navigate and engage with the regional SDE network. To understand this risk, companies were asked to rate six different governance, contracting and pricing requirements according to a five-point Likert scale, with 5 being 'essential' and 1 being 'not at all important' (Figure 6).

Figure 6 clearly reflects industry's unified view of the criticality of streamlined and consistent governance and contracting processes for accessing data across the regional SDE network. The need for consistent pricing across the regional network for the same types of data and services was also ranked as very important or essential by more than 80 per cent of large pharmaceutical and medtech companies.

In interviews, companies expressed that they wanted the interface with the regional SDE network to look and feel like engaging with a single entity wherever possible, to avoid duplication of effort both for themselves and for the regional SDEs. If each SDE adopts its own governance and contracting processes in an uncoordinated and independent manner, industry users will face an unreasonable burden of time and money to navigate these differences across multiple SDEs. In practice, this will risk companies looking elsewhere for a more integrated globally competitive offering to support their research. This is especially true for larger companies, which may be conducting multiple projects with the regional SDE network simultaneously.

Relative importance of governance, contracting and pricing requirements

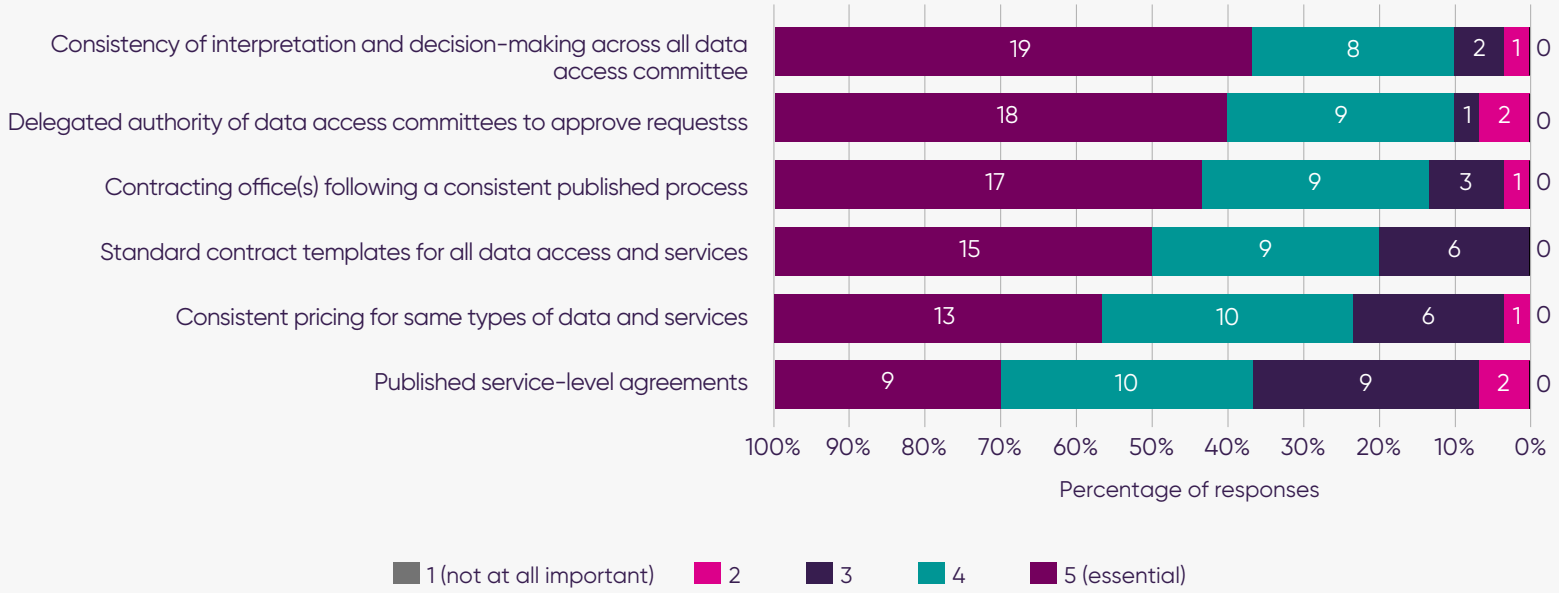


Figure 6 Importance of different governance, contracting and pricing requirements.

"We need interoperability across SDEs in some way, but current governance and architecture suggest this will not be possible"

SME biotech company

To rectify this, companies called for a single point of entry into the regional SDE network via a centralised service that can support project feasibility, governance and contracting engagements with the regional SDE network. Such a centralised service would also be able to coordinate efforts across the regional SDE network, reducing duplication, preventing competition and identifying opportunities for regional SDEs to work together to provide enhanced data and services. The quality of the service provided is critical, with expectations that this centralised service will have all the requisite expertise in data science and analytics, business development and legal teams to ensure an efficient, industry-level standard of service.

> "The promise of single point access for regional SDE network is a great idea. [It would be] very onerous to go to 11 different SDEs."
> **large pharma company**

> "Multiple contracts, environments, governance processes won't work for the customer as it means they can't work out if the overall project will actually happen at all or in time." **large pharma company**

The regional SDE network's current planned approach to a single point of entry does not include many of the aspects that companies need to use the network. It is essential that the pricing is consistent, transparent and competitive. We welcome the efforts made by the Data for R&D Programme to harmonise data access processes and support data discoverability and engagement with individual regional SDEs.

However, the commercial models for the network are still in their infancy and unlikely to be resolved by the time the network launches in spring 2025. This will be a major barrier to the viability of the service in the short and medium term. The Programme should prioritise establishing a centralised service supporting project feasibility, governance and contracting engagements, with all associated necessary expertise and resourcing. Such a service is critical to making the network attractive to companies who place their global investments where their projects are most likely to be successful.

RECOMMENDATION THREE

Provide a unified, integrated service with a single point of entry to the regional SDE network, to streamline user experience and reduce duplication of effort.

Industry values clarity of service offerings and efficient delivery of these services in a consistent manner. To achieve this expectation, there must be a single entry point to access all data from across the network, supported by a service catalogue of data availability. It is not realistic to expect commercial researchers to engage, negotiate and contract with 11 separate regional SDEs. The regional SDE network needs to present a combined offering and avoid internal competition and generate economies of scale. The network should collaborate to share best practice, expertise and feedback intelligence of market needs centrally to develop an integrated strategy that responds to user priorities. By focusing on provision of a unified coherent service, the regional SDE network will offer a commercially compelling user experience.

RECOMMENDATION FOUR

Adopt consistent and harmonised information governance, contracting and pricing processes within commercially competitive timeframes for data access within the regional SDE network.

Businesses operate under commercial delivery pressures. Uncertainties caused by long or inconsistent supply timeframes result in companies choosing alternative data suppliers, often in other countries where their needs can be better met. Respondents indicated that commercially competitive timeframes would require data access request decisions to be made within 25 working days of submission, with access to the data provided within 10 working days following an approved request. Respondents stated that variability in information governance, contracting and pricing processes lead to industry investing more time, money and effort into commissioning and conducting research projects than is necessary. As a consequence, companies could take their research elsewhere. Standardised data access processes must therefore be adopted across the regional SDE network if it is to offer an internationally competitive service.

Perceptions of engagement with industry users

In the first two years of the Programme, engagement with prospective users has been largely through one-way communication channels such as press releases, newsletters, webinars and engagement events or ad hoc stakeholder meetings. It was not until the past 18 months, when, at the request of the TAs, a Stakeholder Advisory Group was convened to allow the TAs and other representative stakeholder bodies to collectively engage with the Programme and provide formal stakeholder input into implementation plans. Although the Stakeholder Advisory Group meets quarterly, it has no role in holding the Programme to account for adopting its advice and no reporting mechanism back to the Programme Board.

In the previous absence of a formalised cross-sector advisory group, it is unclear how the regional SDE network was designed from the outset to meet user needs.

To understand members' views on their engagement with the Programme to date, companies were asked whether they felt that their input into the Programme since its inception has led to meaningful changes to its implementation (Figure 7). One hundred per cent of SME biotech companies indicated 'not at all'. Only one company out of 30 respondents indicated 'mostly', with none expressing 'entirely'. Large pharma and large medtech companies were mixed in their views, with approximately 50 per cent indicating 'not at all'.

How much of your feedback about the regional SDE network, provided through engagement opportunities, has led to meaningful changes to the programme?

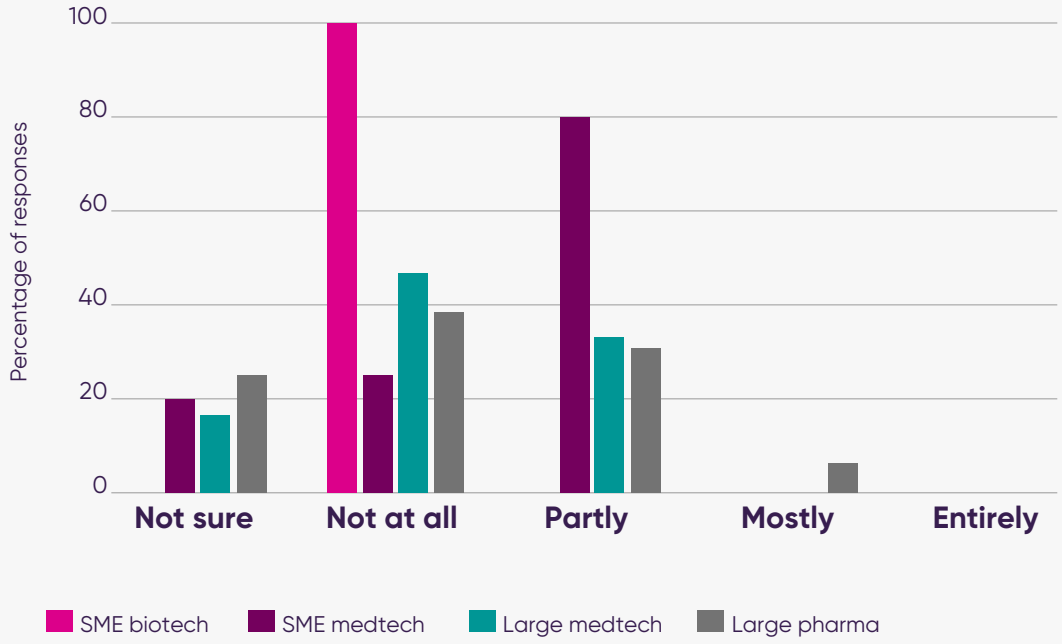


Figure 7 How companies feel about whether their engagement with the Programme to date has led to meaningful change.

In interviews, companies told us that to date, current and potential industry users of the regional SDE network do not feel that the Programme has been developed in partnership with industry, and they lack confidence in the network's ability to deliver their business needs. They stressed the importance of effective future engagement, involvement and communication as tools for building confidence and as vital means of designing services around user needs and requirements. Proposed solutions varied, but companies were universal in the desire for users to be closer to the Programme and to have a bigger role to play in the strategic steer of the Programme.

"The real work is happening behind closed doors, meaning there's a gap between messaging and delivery." **large medtech company**

"The Programme needs to be clearer about the roadmap from where we are now to the final end state with corresponding timelines and points at which external engagement (i.e. with industry) will be sought and decisions expected." **SME medtech company**

Taken together, these reflect a clear need for users of the regional SDE network, including industry, to be more embedded into the design and implementation of the Programme, with an opportunity to shape prioritisation and input into strategic discussions. Companies were not set on the mechanism for this, but they made it clear that the status quo would lead to a continuing decline in confidence of industry users. As a minimum, external advisory groups, including industry user representatives, should have a clear formal remit to provide strategic advice to the Programme, with a reporting mechanism to understand whether their advice has been taken into account.

To achieve financial sustainability, which is a core aim of the Data for R&D Programme, the regional SDE network will be reliant on users paying for services and access to data. The financial reliance on users means that it is essential that the regional SDE network offers commercially competitive services with a realistic pricing model. This will be difficult to achieve without close involvement of users in the design of services.

To enable the regional SDE network to realise its potential, we recommend the establishment of an external stakeholder strategic advisory group that includes industry representation and has a formal reporting mechanism to the Data for R&D Programme Board. This formal role would ensure that the views and proprieties of users are represented in key strategic decision-making and would allow users to play a more active role in supporting the Programme to deliver its objectives.

"Industry [should be] partners in development of national strategy for SDEs." **large pharma company**

"We recommend that the Programme management includes an industry representative embedded in the team."

large pharma company

Companies also recognised that implementing all of the technical, governance and contracting requirements for a health data service operating to industry standards is a significant undertaking. To support the implementation of these various aspects of the regional SDE network, the Programme should convene technical advisory groups, made up of subject area experts, to advise on implementation. The advice and insights of these groups should be disseminated to the regional SDE network to ensure local decision-making is well-informed and user-centred.

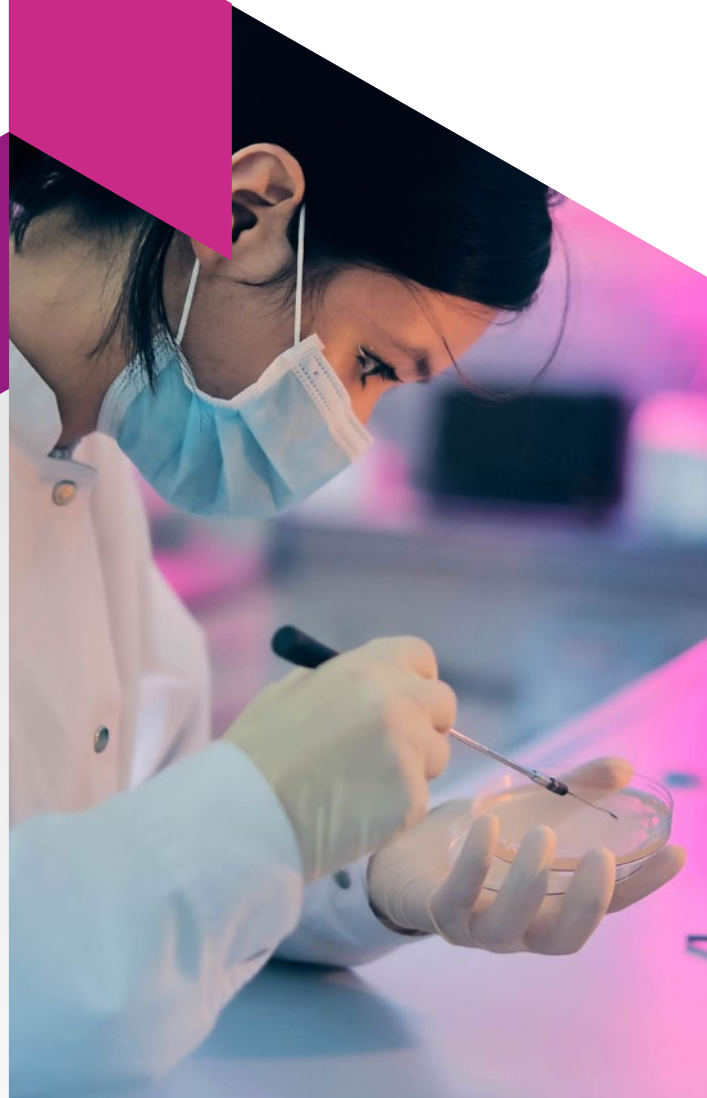
We and our members believe that had these two-way engagement mechanisms been in place from the outset of the Programme, the divergence between user requirements and Programme implementation plans outlined throughout this report could have been avoided.

RECOMMENDATION FIVE

Form an external advisory group, with industry represented, to ensure that users are placed at the heart of strategic decision-making.

User-centred design is at the heart of all well-configured programmes. Health data research services are no different and need to be designed to meet the needs of users of the service. Although there has been some engagement by the regional SDE network with industry, members felt that their input had not been taken into account in subsequent implementation. Members expressed the urgency of meaningful engagement with industry in programme design and implementation. Without this meaningful engagement, members were not confident that the network would meet their needs.

To address this, the Data for R&D Programme should implement an external strategic advisory group that has a formal role in advising the Programme Board. This formal relationship would ensure that the board has sight of discussions that take place within the group and, where appropriate, group members could present at board discussions. It is also essential that there is consistency of operational processes and technical standards across the regional SDE network. The Programme's Community of Practice should convene groups with subject matter experts, including industry representation, to advise on the technical, governance and contracting aspects of the regional SDE network. There should be transparency of how advice from these groups had been acted on by each of the regional SDEs.



Conclusion

For the first time, this report provides an overview of how different companies, according to their size and sector, use health data for R&D, the scope and scale of data requirements underpinning these R&D applications, and a set of clearly defined expectations for what constitutes a viable data service that meets industry needs.

Throughout this consultation exercise, members expressed the view that they hadn't been sufficiently involved in the design and implementation of the regional SDE network. They described the design process in the first two years of the Programme as opaque, with communications from the Programme predominantly one-way and overly simplistic. As demonstrated by the findings of this consultation, the result is a clear misalignment between what industry needs from an effective data service and what is being prioritised and delivered.

The TAs and our members believe that the vision for implementing a regional SDE network to unlock access to high-quality, rich, multimodal hospital datasets is a laudable ambition. However, any ambitious programme, supported by public funding, needs to be realistic with what can be achieved within the available timeframes and budgets. This is particularly important if the resulting services are to transition to financial sustainability from customer revenue in the medium term.

The upcoming CSR provides a timely opportunity to review the strategic and operational implementation of the regional SDE network and ensure it is aligned with user needs and is realistic and deliverable. This report provides a blueprint for course-correcting the implementation of the network, to ensure it can provide globally competitive services that are of use to industry.

The ABPI, ABHI and BIA hope that this report provides useful insights that will support the UK in becoming the best place to use health data for R&D. We look forward to working with the Data for R&D Programme and the regional SDE network in this next phase of the Programme.

Background to this report

The inception of this report stems from the November 2023 Life Sciences Council (LSC) meeting, where industry expressed concerns about the implementation of the regional SDEs. A request was made that the Health Data Industry Sub-Group (HDISG) of the LSC should consider the design and implementation plans of the SDEs and make recommendations, as necessary. In the absence of this request being taken forward by the HDISG, the TAs initiated this consultation exercise with their members to understand their needs and to make recommendations for the implementation of the regional SDE Network.

Throughout this exercise, the TAs have engaged with government and NHS England colleagues to keep them informed about progress and the findings and recommendations arising from the industry consultation. We believe that the information in this report, which outlines for the first time different industry sector use of data and requirements to meet their needs, will be of immense value as the government considers future plans for the Data for R&D Programme, implementation of the Sudlow Review recommendations and creation of a National Data Library.

Annex

Survey questionnaire

1. What sector(s) does your organisation operate in? (check all that apply)

a. Biotech
b. Pharmaceuticals
c. Digital health
d. Diagnostics
e. Medical devices

2. How many employees does your organisation have? (check one option)

a. Microenterprise: 1 to 9 employees.
b. Small business: 10 to 49 employees.
c. Medium-sized company: 50 to 249 employees.
d. Large company, entirely UK-based: 250 or more employees
e. Large company, multinational: 250 or more employees

3. For what purposes does your organisation currently access and use health data for research and development (R&D) in England?

a. Discovery science – pharmaceutical
b. Discovery science – health tech
c. Real-world effectiveness, unmet need and drug usage
d. Clinical trial feasibility and design – medicines and vaccines
e. Clinical trial conduct and follow-up – medicines and vaccines
f. Clinical trial feasibility and design – devices (including digital and AI)
g. Clinical trial conduct and follow-up – devices (including digital and AI)
h. Diagnostic development and validation – non-digital (pre-market)
i. Diagnostic development and validation – digital including AI (pre-market)
j. Non-pharmaceutical therapeutic interventions development and validation – including AI algorithm training (pre-market)
k. Regulatory approval – pharmaceutical
l. Regulatory approval – devices, diagnostics and digital including AI
m. Health Technology Assessment – pharmaceutical
n. Health Technology Assessment – devices, diagnostics and digital including AI
o. Population health
p. Health service delivery
q. Post-authorisation safety studies
r. Other (please describe)

4. At what points to date has your organisation engaged with NHS England about the SDE network? (check all that apply)

a. Input into the initial design
b. Regular and ongoing engagement since programme initiation
c. Involvement in standing advisory groups
d. Involvement in ad-hoc or time-limited advisory meetings
e. Through your trade association
f. No proactive involvement, but receive NHS England communications about SDE network

5. In your view, how much of your feedback about the SDE network, provided through engagement opportunities, has led to meaningful changes to the programme? (check one option)

a. Entirely, or almost entirely (75%-100%)
b. Mostly (50%-75%)
c. Partly (25%-50%)
d. Not at all (0-25%)
e. Not sure

Technical requirements

6. How important are each of the following technical requirements for use of data for R&D purposes? Please rank each of the following criteria from 1 to 5, with 1 being 'not at all important' and 5 being 'essential'.

a. Use of proprietary algorithms, tools and code lists (would need to be imported into a secure analytical environment)
b. Combining other data sets with NHS data, which could be analysed in an NHS secure data environment (would need to import data sources into an NHS secure analytical environment)
c. Combining NHS data with other external data sets that could not be analysed in an NHS secure data environment (would need to export data from an NHS secure analytical environment)
d. Analysing representative data from across England (for example primary care data representative of populations across all English regions)
e. Analysing large/sufficient volumes of data regardless of regional or demographic representation
f. Access to a single analytical platform for all analysis of NHS data, regardless of where the source data is held
g. Ability to archive source data, programmes and log files (for example to meet regulatory requirements)
h. Availability of NHS data in a common data model
i. Direct access to anonymised individual patient-level data
j. Flexible computing capability to scale up and down as required
k. Other essential criteria (please specify)

7. Based on your current understanding, do you believe the SDE network will be able to deliver your organisation's technical requirements? (check one option)

a. Entirely, or almost entirely (75%-100%)
b. Mostly (50%-75%)
c. Partly (25%-50%)
d. Will not deliver most important requirements (0-25%)
e. Not sure

8. Would availability of data in a single SDE be sufficient to meet the majority of your organisation's R&D needs? (check one option)

a. Yes
b. No
c. Not sure

9. How important is it for your organisation to be able to access data federated across multiple (3 or more) SDEs? Please score between 1 and 5, with 1 being 'not at all important' and 5 being 'essential'. [rate from 1 to 5]

Governance processes

10. How important are each of the following governance processes for accessing data for R&D purposes? Please rank each of the following criteria from 1 to 5, with 1 being 'not at all important' and 5 being 'essential'.

a. Access committees have delegated authority to approve requests on behalf of all data custodians
b. Mechanisms in place to ensure consistency of interpretation and decision-making across all NHS data access committees
c. Published service-level agreements
d. Other essential criteria (please specify)

11. What is a commercially acceptable length of time in working days for the following:

a. Submitting a data access request to receiving a decision [free text box]
b. From approval granted to physically accessing data [free text box]
c. Time from completion of analysis to release of results [free text box]

12. Based on your current understanding, do you believe the SDE network will be able to deliver your organisation's governance requirements? (check one option)

a. Entirely, or almost entirely (75%-100%)
b. Mostly (50%-75%)
c. Partly (25%-50%)
d. Will not deliver most important requirements (0-25%)
e. Not sure

Contracting and pricing requirements

13. How important are each of the following contracting and pricing criteria to your organisation for accessing datasets for R&D purposes? Please rank each of the following criteria from 1 to 5, with 1 being 'not at all important' and 5 being 'essential'.

a. Standard contract templates for all data access and services
b. Contracting office(s) following a consistent published process
c. Consistent pricing for same types of data and services
d. Published cost for accessing data including computational cost
e. Other essential criteria (please specify)

14. Based on your current understanding, do you believe the SDE network will be able to deliver your organisation's contracting and pricing requirements? (check one option)

a. Entirely, or almost entirely (75%-100%)
b. Mostly (50%-75%)
c. Partly (25%-50%)
d. Will not deliver most important requirements (0-25%)
e. Not sure

Abbreviated data uses

Abbreviated title	Full title in survey
Discovery – pharma	Discovery science – pharmaceutical
Discovery – healthtech	Discovery science – health tech
Real-world evidence	Real-world effectiveness, unmet need and drug usage
Trial design – pharma	Clinical trial feasibility and design – medicines and vaccines
Trial conduct – pharma	Clinical trial conduct and follow-up – medicines and vaccines
Trial design – healthtech	Clinical trial feasibility and design – devices (including digital and AI)
Trial conduct – healthtech	Clinical trial conduct and follow-up – devices (including digital and AI)
Diagnostic development – non-digital	Diagnostic development and validation – non-digital (pre-market)
Diagnostic development – digital	Diagnostic development and validation – digital including AI (pre-market)
Non-pharmaceutical therapy development	Non-pharmaceutical therapeutic interventions development and validation – including AI algorithm training (pre-market)
Regulatory approval – pharma	Regulatory approval – pharmaceutical
Regulatory approval – healthtech	Regulatory approval – devices, diagnostics and digital including AI
HTA – pharma	Health Technology Assessment – pharmaceutical
HTA – healthtech	Health Technology Assessment – devices, diagnostics and digital including AI
Population health	Population health
Health service delivery	Health service delivery
Post-authorisation safety	Post-authorisation safety studies

Data uses by company profiles

We asked companies to indicate how they use data across 17 categories of use cases.

Reported data uses for large pharmaceutical companies

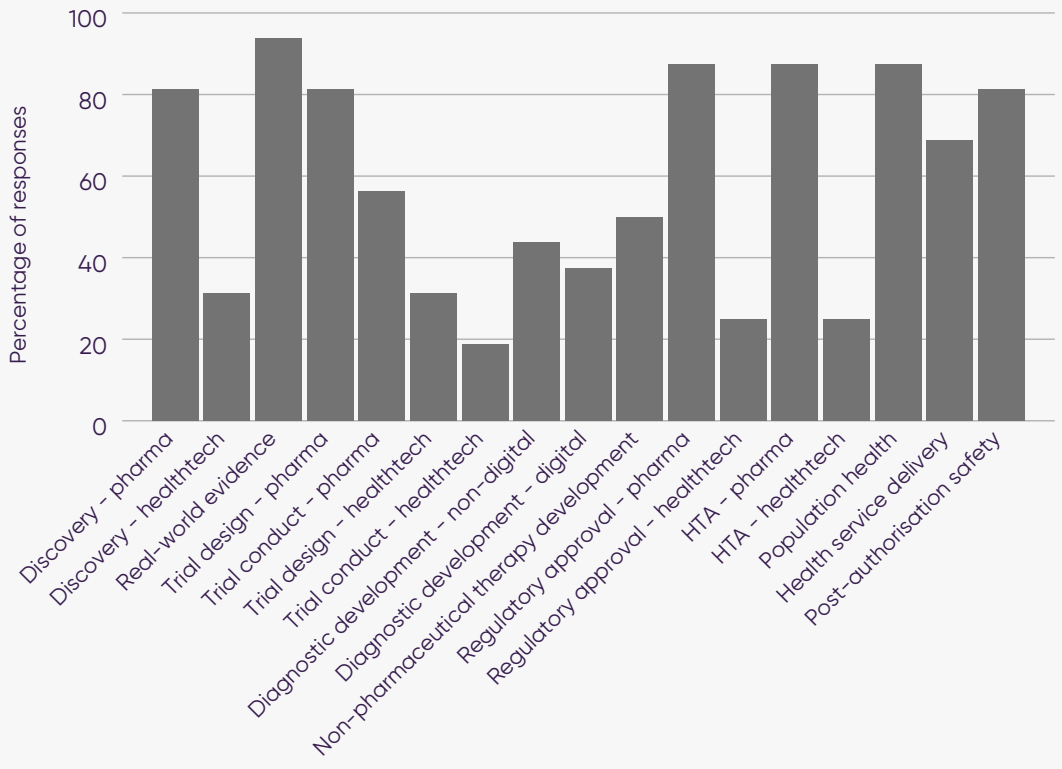


Figure 8 Reported use cases for large pharmaceutical companies.

Reported data uses for large medtech companies

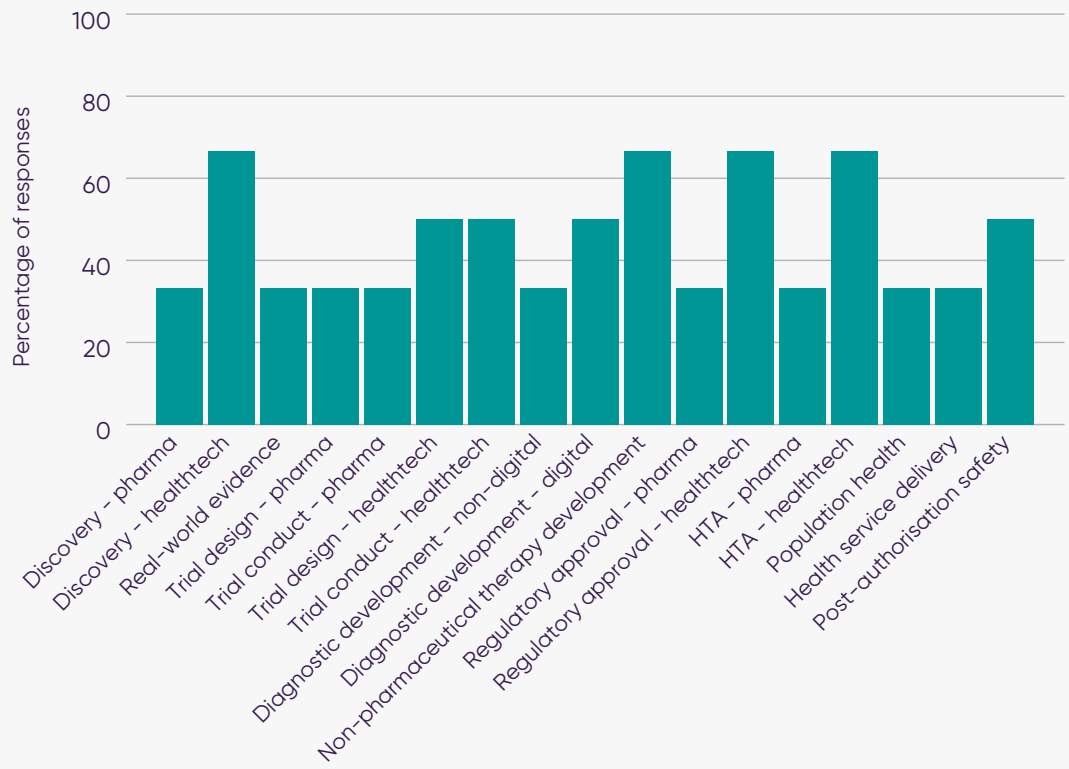


Figure 9 Reported use cases for large medtech companies.

Reported data uses for SME biotech companies

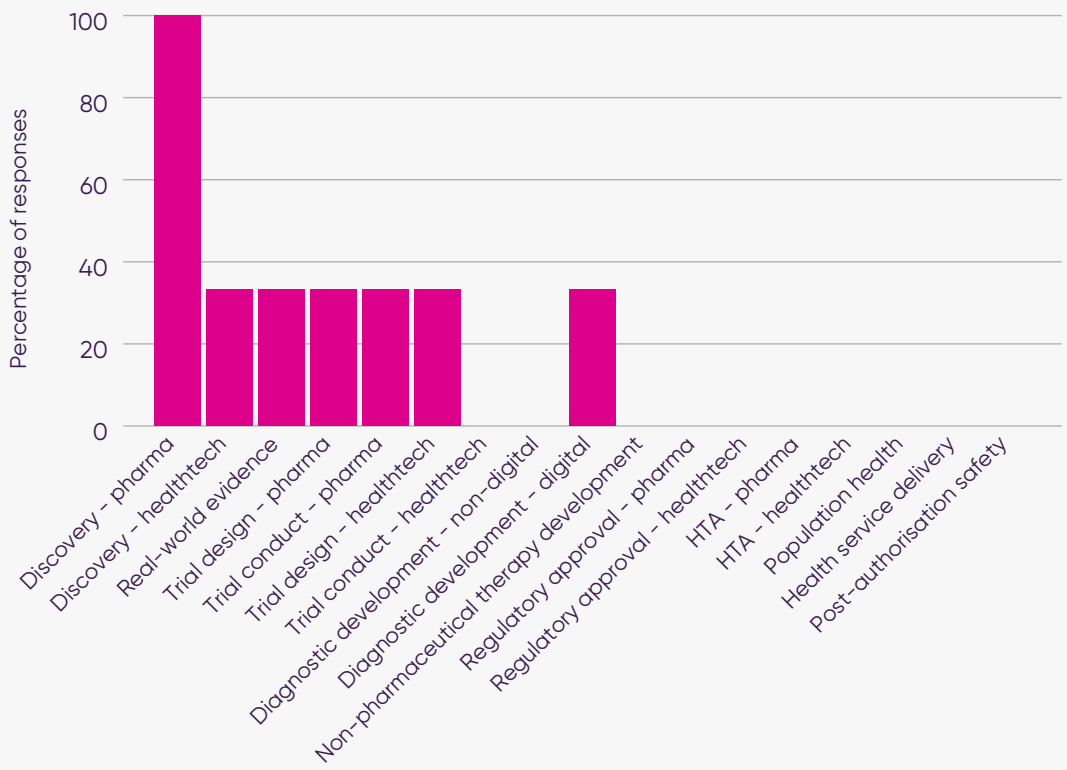


Figure 10 Reported use cases for SME biotech companies.

Reported data uses for SME medtech companies

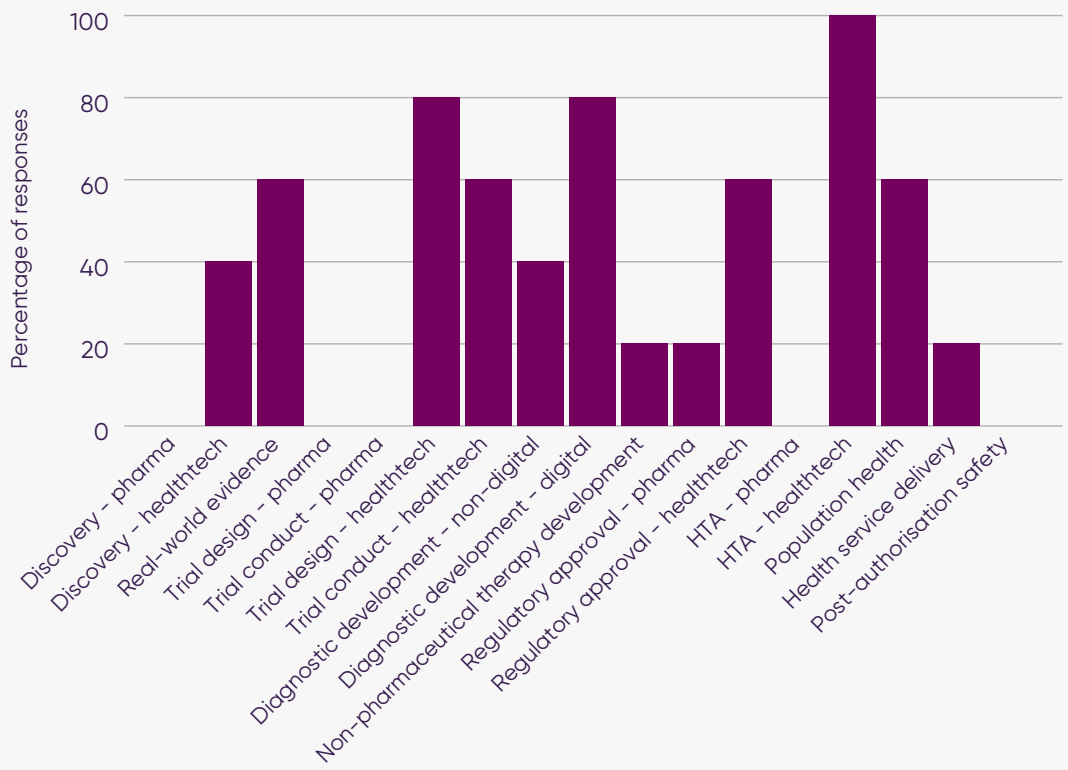


Figure 11 Reported use cases for SME medtech companies.

Additional data related to the other survey questions is available on request.

Abbreviated technical requirements

Abbreviated title	Full title in survey
Use of proprietary algorithms, tools and code lists	Use of proprietary algorithms, tools and code lists (would need to be imported into a secure analytical environment)
Importing other data sets into an NHS secure data environment	Combining other data sets with NHS data, which could be analysed in an NHS secure data environment (would need to import data sources into an NHS secure analytical environment)
Exporting NHS data out of an NHS secure data environment	Combining NHS data with other external data sets that could not be analysed in an NHS secure data environment (would need to export data from an NHS secure analytical environment)
Analysing representative data from across England	Analysing representative data from across England (for example primary care data representative of populations across all English regions)
Analysing large volumes of data, regardless of representativeness	Analysing large/sufficient volumes of data regardless of regional or demographic representation
Access to a single analytical platform for all analysis of NHS data	Access to a single analytical platform for all analysis of NHS data, regardless of where the source data is held
Ability to archive source data, programmes and log files	Ability to archive source data, programmes and log files (for example to meet regulatory requirements)
Availability of NHS data in a common data model	Availability of NHS data in a common data model
Direct access to anonymised individual patient-level data	Direct access to anonymised individual patient-level data
Flexible computing capability to scale up and down as required	Flexible computing capability to scale up and down as required

The Association of the British Pharmaceutical Industry

A company limited by guarantee registered in England & Wales number 09826787

Registered office 2nd Floor Goldings House, Hay's Galleria, 2 Hay's Lane, London, SE1 2HB

RMI-0185-0125