ABHI Response to the Government Review into the efficient and safe use of health data for research and analysis for the benefit of patients and the healthcare sector.

Introduction

As set out in the Life Sciences Industrial Strategy, whilst the UK possesses several data sources which offer significant potential to researchers, the NHS is not able to offer access to deep, real-time data across multiple care settings. At present health data is collected by an extensive range of organisations and processes meaning researchers, commissioners and innovators are not always able to access the NHS data they need due to the fragmented nature of the datasets, the lack of interoperability and the burdensome and diverse information governance requirements.

ABHI welcomes the opportunity to respond to this review and the Data Strategy for Health and Social Care, which it will inform. There are many opportunities for legislation, system wide guidance and standardisation to deliver a coherent data strategy that will enable clinicians, patients, the health and care system, industry and wider UK society to reap the benefits of the potentially unique data opportunity with our NHS.

Response

1. **How do we facilitate access to NHS data by researchers, commissioners, and innovators, while preserving patient privacy?**
    - Protecting patient privacy must be the number one priority when designing any system which allows for the use of patient data. There is a need to define what the intent is behind facilitating access to NHS data as this could impact appropriate measures. Relevant purposes should include
        - Conducting scientific research
        - Measuring health economic impact of interventions
        - Monitoring product and service performance for reasons of patient safety
        - Support innovation in technology development and implementation
    - Before the benefit of utilising anonymised patient data for clinical research and to improve patient care can be full realised some of the barriers to interoperability and standardisation need to be overcome. The NHS could be uniquely placed to setup a National Health Hub, underpinned by a National Electronic Health Record system. However currently, health and care data in the UK is held on multiple platforms; a recent study found that, of the 117 NHS trusts using electronic health recording systems, 92 of them were using at least 21 different systems[1]. To facilitate access to data that creates intelligence about disease states and public health while providing insights to designing better and more personalised therapies this fragmentation needs to be addressed.
    - As a first step, the UK needs to aggregate and curate NHS data. The cost of aggregating and curating patient records for data driven research and innovation is

[1] https://spiral.imperial.ac.uk/bitstream/10044/1/75302/9/ImprovingDataSharingBetweenAcuteHospitals.pdf

estimated at £2-3bn[2]. Once complete, this would generate income for the NHS, reduce costs and increase the speed of clinical research, benefiting patients. It would also make the UK a more attractive location for foreign direct investment, creating 17,000 jobs and delivering £5bn in annual revenue[3].

- An international viewpoint should be taken to support trade and investment, UK rules should seek to be coherent with initiatives and legislation in other major trading blocs such as the US and EU. The UK should not lose sight of adequacy with the EU and align with the EU Recommendations on European Electronic Health Record exchange format.
- A clear and consistent legal and governance framework for accessing data, including governance mechanisms for primary and secondary use of health data, is needed. Local variances in IG interpretations and literacy impede the routine analysis of health data. A central and single IG guidance, readily applicable to NHS routine care is needed.
- The common regulatory and governance framework should be centrally arranged and should enable multiple deployments e.g. at regional/national level, depending on data types and intended use, and result in federated Health Data Spaces.
- Existing approaches of providing access to pseudonymised data (such as HES data) via licences managed by NHS Digital allows appropriate safeguards to be in place.

2. **What types of technical platforms, trusted research environments, and data flows are the most efficient, and safe, for which common analytic tasks?**
   - A well-defined, common distributed data infrastructure is fundamental to facilitate a consistent and secure use and re-use of health data.
   - Achieving technical and semantic interoperability and seamless exchange of data and information is critical to the success of the Health Data Spaces and improvements in clinical operations, patient outcomes and cost of healthcare. The interoperability of electronic health records, as well as semantic and technical interoperability should be strengthened. The governance framework should prioritize standardisation needs and improve data interoperability.
   - Security measures like pseudonymisation or de-identification are important elements to safeguard data flows. In addition, following security and privacy by design are required including ensuring encryption at rest and in transit by default, together with minimum least privileges and segregation of duties in the code. A final piece is to ensure compliance to stringent regulatory requirements and rigorous certifications such ISO 27001 and HDS.
   - There should be clarity around whether data needs to be held within the NHS N3 network or if it is acceptable for the anonymised, aggregated data to be shared outside of this under certain conditions.

[2] EY report, Indicative estimates of market value for NHS datasets, March 2019

[3] Data from OLS

3. **How do we overcome the technical and cultural barriers to achieving this goal, and how can they be rapidly overcome?**

   - Consideration of the need for industry, to access the data and research environments needs to be an integral part of the solution rather than an after-thought. Currently it appears that industry access to a range of NHS data is becoming increasingly difficult to navigate. This includes even basic, non-clinical data such as costing and payment data.
   - Numerous research pieces have highlight patients and citizen reticence to share data with commercial (for profit) entities. As regards cultural barriers; Government and the NHS can help with communications on the importance of data sharing, the privacy and security standards employed and the role of the private sector in developing innovations utilised in the NHS.
   - Technically a critical part of the issue is overcoming the legacy of older systems, a national data health record system that is centrally managed could support this. For new implementations a regime of common standards would overcome technical barriers that facilitate data portability and it will be important to build trust and transparency in the process. Conformity to interoperability standards should be a go/no-go part of procurement processes and be reflected in reimbursement decsions.
   - Data portability and interoperability would be further supported by common standards and templates for data collection. The reality at the moment is that different Trusts collect different data and there is no common standard.
   - The technology used to input data needs to be made more user-friendly for clinicians. Technology must be easily integrated into clinicians' daily routine and viewed as a tool to improve care rather than a burdensome bureaucratic exercise.
   - The Centre for Data Collaboration should provide guidance to NHS organisations around the appropriate patient consent or other appropriate mechanisms for projects, specifically around whether secondary use is appropriate or not. Currently interpretation and implementation can vary widely. The way that the EU General Data Protection Regulation (GDPR) is being applied makes it difficult to move research forward and to share data. There is little or no evidence of patient data ever becoming de-anonymized during clinical research and there is a need to be able to access public databases. But the fear of personal details being released has led to overregulation, resulting in a reduction in research[4].
   - Ensure adequate training for the workforce by implementation of the Topol's reviews recommendations on skills such as data science, leading to a much more ambitious NHS People Plan. This should be supported by investment to speed up the progress on IT infrastructure

[4] https://perspectives.esmo.org/latest-edition/featured-content/does-academic-research-really-need-so-much-administrative-work

4. **Where (with appropriate sensitivity) have current approaches been successful, and where have they struggled?**

   - There are many examples at different levels, national, supra-national, or for specific disease. We accept that international best practices cannot always easily translate to the UK/NHS due to different organisational, cultural and political circumstances, but we believe learnings can be gained nonetheless.

     - The Covid-19 Data Store was set up to provide system leaders with data-driven evidence to inform decision making during the Covid-19 pandemic. Working with commercial organisations, data has been used to monitor the spread of the virus and ensure services and support are available to patients; for example by assessing bed capacity in hospitals and the number of ventilators available in particular areas. This centralised approach to data management and analysis has been a notable step forward in many regards.
     - The Genomics England Trusted Research Environment has worked well in terms of access to data and clarity around how it works. The technology does however require updating, which is underway.
     - The EU Gaia-X initiative creates the foundation for a federated, open data infrastructure based on a credible, trustworthy and trusted infrastructure for the hosting of the Health Data Space, including health data spaces at regional/national level, leading to federated Health Data Spaces. This initiative holds the potential of achieving the scale required to access and share health data securely and confidently. Such federated cloud ecosystem would enable the creation of a competitive marketplace for cloud services and help avoid dominant vendor 'lock-in'.
     - The eICU Research Institute, a non-profit institute established by Philips and governed by its customers. It is a platform built from a repository of data that is used to advance knowledge of critical and acute care providing participants with access to the most comprehensive database of ICU care in the world with opportunities for sponsored research.
     - The MIMIC dataset, an openly available dataset developed by the MIT Lab for Computational Physiology, comprising deidentified health data associated with ~60,000 intensive care unit admissions. It includes demographics, vital signs, laboratory tests, medications, and more.
     - Research Data Assistance Center (ResDAC) is a Centers for Medicare and Medicaid Services (CMS) contractor that provides free assistance to researchers interested in the CMS data set.
     - Findata, a 'one-stop shop' set up with co-operation between the public and private sectors to make the secondary use of social welfare and healthcare data easier and make the data more available, as well as to promote its secure use for more extensive purposes. With new enabling legislation that came into force in May

2019, Finland became the first country to successfully enact a law on the secondary use of well-being data that met EU GDPR requirements[5].

- France's Health Data Hub model, where it aggregates data from different sources and makes it possible to share some data with public and private institutions for those specific cases such as studying diseases and using artificial intelligence to improve diagnoses.

- Consideration should also be given to innovations in business models that support public/private collaborations, do not restrict use of the data and ensure that appropriate value is return to all parties involved. An example of this would be NHS Trusts having an equity stake and entitled to receive a share of any revenues generated as a result of the commercialisation of research.
- More broadly, approaches have been successful when there has been political will and financial incentives to set up appropriate organisations.

5. **How do we avoid unhelpful monopolies being asserted over data access for analysis?**
   - Patient data is a national asset and patients, the NHS and the UK taxpayer should have a stake in the value of scientific insights unlocked by combining large datasets. In order to avoid unhelpful monopolies, the NHS should remain a key stakeholder in the use of data.
   - The UK has the opportunity to become a world leader in ethical data analysis and medical discovery and we cannot afford to let this opportunity slip through our grasp. The risk is that value, jobs, and clinical breakthroughs will be commercialised outside of the UK. We need a national health and care data strategy that embraces a partnership model to ensure public trust in the use of patient data and that the UK does not lose its competitive advantage.
   - The strategy should include the ethical application of clinical artificial intelligence and the creation of a trusted data community between patients, clinicians, healthcare providers and life science companies that is transparent, fair and effective.
   - Conditions and fees for re-use of public sector data should be non-discriminatory, proportionate, objectively justified and not restrict competition. Enforcement and compliance mechanisms should be put in place to avoid disproportionate access for those who are not compliant. The existing work of the Centre for Data Collaboration picks up on a number of these themes. A variety of commercial models should be available to those wishing to access data.
   - Data monopolies can be addressed by creating a health data system that provides strong incentives for companies to participate. This might include the ability for participants to benchmark their data/performance against alternatives, validate findings/products in real-world NHS practice, or otherwise put their proprietary data into a broader health/health economic context. In addition, the development of open, non-proprietary standards for data, data access, and interoperability can also be an

[5] 'A Finnish Model for the Secure and Effective Use of Data' https://www.sitra.fi/en/publications/a-finnishmodel-for-the-secure-and-effective-use-of-data/

important tool to ensure an alternative to data monopolies. A dialogue needs to take place between the contributors and users of data to develop a common dictionary and interoperability of data and put in place a common set of standards for how healthcare data is accessed, shared and stored.

- Clear policies and an appropriate governance framework led by national authorities that support the sharing and use of data for the public interest is critical. A recent EU proposal, the Data Governance Act, aims to define such a framework and could provide a reference point for UK policymakers in this regard.
- The NHS has a patient dataset with world leading potential for medical discovery and drug development. This is an attractive proposition for organisations in the health tech sector, therefore the question of monopolies is an important one for the UK's approach to data driven innovation. The non-rivalrous nature of data means it can be difficult to establish the rights to develop intellectual property from data, which is important for driving new discoveries. As the National Data Strategy makes clear, we need a sophisticated approach that encourages innovation and competition while maintaining public trust and ensuring that the NHS and the public see the benefits of health data innovation.
- More data sets should be made available under open license, this could help prevent unhelpful monopolies. This could be supported through procurement frameworks making it mandatory that all agreements on access to data must be non-exclusive.

6. **What are the right responsibilities and expectations on open and transparent sharing of data and code for arm's length bodies, clinicians, researchers, research funders, electronic health records and other software vendors, providers of medical services, and innovators? And how do we ensure these are met?**

   - As first principle, data should only be used in the best interest of the patient and wider public health. It will be important to maintain medical confidentiality while following common standards to facilitate more open and transparent sharing.
   - Consider introducing guidelines for the ethical use of AI. Such guidelines could be based on the work of the EU High-Level Expert Group which in 2019 adopted Ethics Guidelines for Trustworthy AI that included a set of 7 key requirements that AI systems should meet in order to be deemed trustworthy[6].
   - Consideration should be given to development of joint codes of ethics between relevant NHS body (possibly the Centre for Data Collaboration) and organisations representing user groups.
   - All stakeholders should see value and have an incentive to participate. Importantly, this includes ensuring protection of intellectual property rights as well as providing a clear understanding of whether/how data may be used for regulatory, reimbursement, or legal actions/decision-making in the future. For example, data that is contributed for the purpose of research or specific clinical investigation may not be properly contextualized for use in subsequent regulatory or health technology assessments. Stakeholders

[6] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

should have assurance that incomplete data contributed as a public service to the research or public health communities won't be used later, without consent, as a basis for business-impacting decisions.

- All data should be anonymised in accordance with Recital 26 of the GDPR. For the data to be considered anonymous, one must not be able to identify a natural person by any means such as singling out, 'linkability' or inference. Thus, the determination of whether or not the data is anonymous is based on a risk-based evaluation and anonymisation is therefore defined as a risk threshold.
- There should be strict rules on when anonymised data can be accessed based upon the Caldicott Principles for data sharing and all relevant data governance legislation such as GDPR.
- The NHS Organisation should remain the data controller as defined in Article 4 paragraph 7 of the GDPR regulation and ensure data is handled according to national privacy and security legislation.
- Any external organisation processing anonymised patient data should be required to ensure a high standard of internal cyber security protections.
- Sharing of code presents greater challenges due to intellectual property restrictions, however by defining standard interfaces between applications and incentivising open source code and library sharing, we can seek to address these challenges.

7. **How can we best incentivise and resource practically useful data science by the public and private sectors? What roles must the state perform, and which are best delivered through a mixed economy? How can we ensure true delivery is rewarded?**

- We recognise that when it comes to data sharing there needs to be an exchange of value - it's important that health services see a benefit, not least to inspire public confidence in data sharing. Whilst considerable weighting should be given to the value of data for the NHS and patients, there should also be consideration given to the overall value of any partnership to the advancement of healthcare and the positive economic impact on the life sciences industry in the UK.
- To encourage private sector to participate in data sharing and data management there need to be certain safeguards in place, which will include
  - Value should be correlated to the extent to which the analysis can create system-wide benefits.
  - IP assignment and protection should be transparent and aligned with value creation.
  - Commercial in confidence information should be respected.
  - Provision should be made for "safe harbour" discussions on data handling.
  - There needs to be greater clarity on how data will be assessed by payers and what their minimum expectations are.
- 'True' risk sharing approaches, where each party puts in the required resource and is remunerated accordingly should be encouraged. The Data Strategy for Health and

Social Care must enable this approach and ensure there is a clear recognition of value beyond the short term, monetary value for the health service.

8. **How significantly do the issues of data quality, completeness, and harmonisation across the system affect the range of research uses of the data available from health and social care? Given the current quality issues, what research is the UK optimally placed to support now, and what changes would be needed to optimise our position in the next 3 years?**

   - The UK is lagging behind several jurisdictions in the quality, completeness and interoperability of its national data. For example, whether the patients' consent covers the intended use of the health data, whether they are legally able to use the dataset in the way they intend, what the costs of access are, and how long it will take to get permission and practically access the dataset[7]. This digital gap could be closed by building a robust IT infrastructure for data storage and sharing, implement a national electronic health record across primary and secondary care and addressing interoperability challenges. It is recognised that given the diversity of secondary care system that this is a long-term goal.
   - The establishment of the Health Data Research Innovation Gateway as part of the HDRUK Research Alliance is a positive step However, challenges exist with the Alliance - organisations are not required to make specific commitments to data access, and membership of the group is not comprehensive, with some relevant bodies not signed up8. The Government should continue to support the Gateway and help greater completeness and transparency of quality of the data in the available datasets.
   - Allowing data harmonisation across the system as part of Integrated Care Systems will allow the impact of innovation and the total cost of care to be better quantified across the budgetary siloes which exist.
   - The propagation of standards will help reduce variety of data formats and structures and will make it easier to achieve higher data quality standards as (software) solutions will then focus on interoperability, improving on 'first-time-right' data capture/entry.

9. **If data is made available for secondary research, for example to a company developing new treatments, then how can we prove to patients that privacy is preserved, beyond simple reassurance?**

   - The government and the health service must lead a national conversation with the public about the benefit of using data and the safeguards which are in place. This should provide clear and simple information on how data is used and protected, the meaning of 'de-identification', and the benefits to patients and society of using health data.
   - The phraseology of 'secondary use for research' is often used. However, it should be clear that there are a wider variety of secondary uses that could be appropriate, such as

[7] ABPI, Unlocking the promise of UK health data, Supporting innovation in the development of new medicines. Spring 2020

[8] HDR UK, UK Health Data Research Alliance, accessed 6 December 2019

- Prototyping, validation and verification of algorithms
- Post market surveillance for patient safety
- Clinical and economic impact modelling for innovations (technology, clinical or pathway) where access to datasets such as PLICS would be beneficial

- It is important that transparency around the use of the data for secondary purposes, and what those purposes are, is set out to patients as comprehensively as possible.
- The Government should seek to establish a framework to articulate the balance of protecting personal data with the benefits to society in using de-identified data. This should ensure privacy is preserved at the point the data is shared, via proper de-identification/anonymization and encryption.
- A further technical method could include federated learning, where the data is not shared, but instead the algorithm training is done where the data is located without exposing the data. This way the source data does not need to be exchanged and can remain 'on premises'.
- A national approach to data-sharing with a common consent model for data use, a system of certifications and data privacy authorisation/approval provided by national data protection authorities or competent bodies to the relevant parties accessing and using the data could be an important tool assure patients.
- Any data sharing should have clear protocols including definitions of the use, further sharing, retention times and deletion procedures for the anonymised patient data.

10. **How can data curation best be delivered, cost effectively, to meet these researchers' needs? We will ensure alignment with Science Research and Evidence (SRE) research priorities and Office for Life Sciences (OLS) (including the data curation programme bid).**

11. **What can we take from the successes and best practice in data science, commercial, and open source software development communities?**
    - The advantages of Open Source Software development have been established including free and more access to data, fostering the speed of innovation, and better interoperability, integration and standardization. Privacy and data security must be integrated into the design and structure of these open platforms.
    - See also question 4 for some best practices on co-creation on data & AI with clinical partners.

12. **How do we help the NHS to analyse and use data routinely to improve quality, safety and efficiency?**
    - The Health Technology industry has a range of skills, resources and tools that could support the NHS (for example the HealthSuite Digital Platform). More use should be made of this resource and NHS policy and practice should encourage and facilitate private-public partnerships to help the NHS realise the full potential of the data asset, which will only be recognised if data is properly curated and linked at scale, rather than developed in an ad hoc way by small groups of Trusts.

- Increasingly connected and powerful devices provide data that can produce significant clinical and system efficiency in sights. The NHS should allow the data to be analysed, processed by equipment manufacturers and returned to the NHS.
- The Government should prioritise the aggregation and curation of NHS data so that it can be used to its full potential. The cost of aggregating and curating patient records for data driven research and innovation is estimated at £2-3bn9. Yet, once complete, this would generate income for the NHS, reduce costs and increase the speed of clinical research, benefiting patients. It would also make the UK a more attractive location for foreign direct investment.
- Trusts should also receive funds to upgrade their IT systems and ensure they have the capacity and infrastructure in place to collect and analyse data. In addition, Trusts should receive appropriate financial return for the use of their data.

About ABHI and HealthTech

- The Association of British HealthTech Industries (ABHI) is the leading health technology (HealthTech) industry association in the UK. We are a community of over 300 members, from small UK businesses to large multi-national companies. We champion the use of safe and effective medical devices, diagnostics and digital health technologies. The work of our members improves the health of the nation and the efficiency of the NHS.
- The HealthTech industry makes a vital contribution to economic growth in our country. The industry employs over 127,400 people across 3,860 companies, mostly small and medium sized enterprises (SMEs). Many companies are working closely with universities and research institutions. The industry is generating a turnover of over £24 billion and has achieved employment growth of greater than 5% in recent years. ABHI's members account for approximately 80% of the value of the sector as measured by sales to the NHS. As the most highly regarded universal healthcare system in the world, the NHS in turn is dependent on technology produced by the industry to enhance the efficiency of services and drive continuous improvement in their delivery. The NHS has grown and developed partly on the basis of the UK's historic 'can do' approach to engineering and problem solving.
- HealthTech is accordingly an engineering-based industry, characterised by rapid, often iterative product design and development, and a large number of SMEs. It is one of two distinct subsectors of the broader Life Sciences. Future growth and success will mean the HealthTech sector being recognised in its own right. The sector has evidence, regulatory and adoption needs that differ significantly from those of the other, biopharmaceuticals.

[9] EY report, Indicative estimates of market value for NHS datasets, March 2019