Guidance

Code of conduct for data-driven health and care technology

Introduction

Today we have some truly remarkable data-driven innovations, apps, clinical decision support tools supported by intelligent algorithms, and the widespread adoption of electronic health records. In parallel, we are seeing advancements in technology and, in particular, artificial intelligence (AI) techniques.

Combining these developments with data-sharing across the NHS has the potential to improve diagnosis, treatment, experience of care, efficiency of the system and overall outcomes for the people at the heart of the NHS, public health and the wider health and care system.

Innovators in this field come from sectors that are not necessarily familiar with medical ethics and research regulation, and who may utilise data sets and processing methods that sit outside existing NHS safeguards.

It is our duty as NHS England and central government to capitalise on these opportunities responsibly. People need to know that their data is being used for their own good and that their privacy and rights are safeguarded. They need to understand how and when data about them is shared, so that they can feel reassured that their data is being used for public good, fairly and equitably.

Our responsibility as an internationally trusted health and care system is to use all the tools at our disposal to improve the quality and safety of care, including data-driven technologies, in a safe, ethical, evidenced and transparent way. For this reason, we have developed our 10 principles in a code of conduct to enable the development and adoption of safe, ethical and effective data-driven health and care technologies.

These principles were published on 5 September 2018 along with a questionnaire for members of the public to offer feedback. In addition to this online feedback, the Department of Health and Social Care and NHS England began an extensive period of engagement with industry experts, academics, regulators and patient representative organisations over the last quarter of 2018. This version of the code of conduct reflects that engagement.

The code is designed to recognise that, while data-driven health and care technologies will undoubtedly deliver huge benefits to patients, clinicians, carers, service users and the system as a whole, it is our duty as NHS England and central government to capitalise on these opportunities responsibly. If we do not think about issues such as transparency, accountability, liability, explicability, fairness, justice and bias, it is also

possible that the increasing use of data-driven technologies, including AI, within the health and care system could cause unintended harm.

The code of conduct clearly sets out the behaviours we expect from those developing, deploying and using data-driven technologies, to ensure that all those in this chain abide by the ethical principles for data initiatives developed by the Nuffield Council on Bioethics:

1. respect for persons
2. respect for human rights
3. participation
4. accounting for decisions

The code tackles a number of emerging ethical challenges associated with the use of data-driven technologies in the NHS and the wider health and care system. We therefore expect to engage with the Centre for Data Ethics and Innovation on developing and monitoring the code to ensure it fits with the latest best practice. We will continue to engage with all members of the health and care community.

We will publish further updates at the end of 2019.

The ultimate ambition is that the code is complementary to the health research, medical device regulations and the CE mark process as well as other regulatory approvals. When used as part of an overarching strategy it will help to create a trusted environment that supports innovation of data-driven technologies while being:

- the safest in the world
- appropriately responsive to progress in innovation
- ethical, legal, transparent and accountable
- evidence-based
- competitive and collaborative
- in alignment with the NHS Constitution

The principles

1. Understand users, their needs and the context

Understand who specifically the innovation or technology will be for, what problems it will solve for them and what benefits they can expect. Research the nature of their needs, how they are currently meeting those needs and what assets they already have to solve their own problems. Consider the clinical, practical and emotional factors that might affect uptake, adoption and ongoing use.

2. Define the outcome and how the technology will contribute to it

Understand how the innovation or technology will result in better provision and/or outcomes for people and the health and care system. Define a clear value proposition with a business case highlighting outputs, outcome, benefits and performance indicators.

3. Use data that is in line with appropriate guidelines for the purpose for which it is being used

State which good practice guideline or regulation has been adhered to in the appropriate use of data, such as the Data Protection Act 2018. Use the minimum personal data necessary to achieve the desired outcomes of the user's needs and the context.

4. Be fair, transparent and accountable about what data is being used

Utilise data protection-by-design principles with data-sharing agreements, data flow maps and data protection impact assessments. Ensure all aspects of the Data Protection Act 2018 have been considered.

5. Make use of open standards

Utilise and build into the product or innovation current data and interoperability standards to ensure it can communicate easily with existing national systems. Programmatically build data quality evaluation into AI development so that harm does not occur if poor data quality creeps in.

6. Be transparent about the limitations of the data used and algorithms deployed

Understand the quality of the data and consider its limitations when assessing if it is appropriate for the users' needs and the context. When building an algorithm, be clear about its strengths and limitations, and give clear evidence of whether the algorithm you have published is the algorithm that was used in training or in deployment.

7. Show what type of algorithm is being developed or deployed, the ethical examination of how the data is used, how its performance will be validated and how it will be integrated into health and care provision

Demonstrate the learning methodology of the algorithm being built. Aim to show in a clear and transparent way how outcomes are validated.

8. Generate evidence of effectiveness for the intended use and value for money

Generate clear evidence of the effectiveness and economic impact of a product or innovation. The type of evidence should be proportionate to the risk of the technology and its budget impact. An evidence-generation plan should be developed using the evidence standards framework published by NICE.

9. Make security integral to the design

Keep systems safe by safeguarding data and integrating appropriate levels of security into the design of devices, applications and systems, keeping in mind relevant standards and guidance.

10. Define the commercial strategy

Purchasing strategies should show consideration of commercial and technology aspects and contractual limitations. Consider only entering into commercial terms in which the benefits of the partnerships between technology companies and health and care providers are shared fairly.

Principle 1: Understand users, their needs and the context

Understanding the people and their specific needs will help with uptake and adoption of the technology or innovation being built, as well as clearly showing a commissioner or buyer the problem being solved.

In health and care, users can be patients, family, carers, staff or any combination of those roles working together to achieve an outcome. They may well have existing capacities and ways of solving their problem, which innovation might sustain or disrupt. Health and care services are for everyone, including people with different physical, mental health, social, cultural or learning needs.

The user, data subject and beneficiary of a data-driven service may be the same person. If they're not, make sure you understand the different interests and expectations of each. 'User needs' are the needs that a user has of a service, and which that service must satisfy for the user to get the right outcome for them.

User needs can be:

- clinical, such as understanding co-morbidities
- practical, such as access to technologies, or time to spend interacting with a service
- emotional, such as needs for reassurance as well as diagnosis or treatment

If you are unsure how to carry out good user research, follow the Government Digital Service manual on user research and the NHS Digital design service manual.

Principle 2: Define the outcome and how the technology will contribute to it

Demonstrate how and where the product will add value to people and the health and care system. This will help with uptake.

Our ultimate objective, as NHS England and central government, is the provision of better care and improved health outcomes. Having a clear hypothesis about how the technology or innovation will contribute to that, for example through:

- new knowledge and capabilities to apply in health and care
- a firmer evidence base, and reduced uncertainties
- more efficient ways of doing things
- improvements to the patient experience

Have a clear value proposition and make a business case highlighting outputs, outcomes, benefits and key performance indicators (KPIs). Clearly define KPIs and where the product will result in better provision and/or outcomes for people, in addition to outlining where and how cost savings or reductions are likely to be made.

Principle 3: Use data that is in line with appropriate guidelines for the purpose for which it is being used

State which good practice guideline or regulation has been adhered to in the appropriate use of data, such as the Data Protection Act 2018. Be able to explain to a member of the public why the data used was needed and how it is meeting the user need. The Data Ethics Framework, published by the Department for Digital, Culture, Media and Sport, sets out clear principles for how data should be used in the public sector.
Explain the proportionality of the data. This applies to both data use in the research or testing period and after the digital health product goes live and is used as part of standard care. Wherever possible it is preferable to use anonymised data in testing rather than identifiable patient data.
If using patient data or accessing NHS patients in order to conduct healthcare research to either develop a proof of concept or test a digital health tool (sometimes referred to as

health technology assessment), conform to the UK Policy Framework for Health and Social Care Research.
If conducting research, seek approval from the Health Research Authority (HRA) to carry out this research. This can be facilitated through the Integrated Research Application System (IRAS). For further information, see the HRA information on data-driven technology.
If developing a patient-facing data-driven app, follow the standards set out in the Digital Assessment Questionnaire (DAQ). The DAQ was developed by a group of subject matter experts across several specialist organisations and regulatory bodies with the aim to provide a clear reference point for current regulation, standards and best practice relevant to apps in health and social care.
If the patient data planning to be used has been anonymised in line with the ICO's code of conduct on anonymisation and meets the requirements of the common law duty of confidentiality, ethical review from the HRA will not be needed. However, HRA approval may still be required. If planning to use identifiable patient data in the development and/or testing of the technology, ensure that there is appropriate consent to access the data or some other legal basis, such as approval under section 251 of the NHS Act 2006 and Health Service (Control of Patient Information) Regulations 2002.
Establish if the data-driven technology that is being developed is defined as a medical device or in vitro diagnostic tool and has followed the required regulatory conformance route. Software that meets the definition of a medical device will be regulated as such.
Since May 2018 the national data opt-out allows people to opt out of their confidential patient information being used for purposes beyond their individual care and treatment. By 2020 any health and care organisation that processes and/or disseminates data that originates with the health and adult social care system in England is required to be in compliance with the national data opt-out policy. Anonymised data in line with the ICO's code of practice is exempt from this. Data flow maps will enable data controllers to be clear when the national data opt-out should be applied, as this is defined using the legal basis.

In some instances, building an AI model may require the persistence of data about individuals who do not have a particular disease as well as those who do have it, in order to learn about the process of the development of the disease, for example. In cases such as these, the drivers, requirements and benefits of the data that is needed should be set out and justified in an application for data.

Principle 4: Be fair, transparent and accountable about what data is being used

The NHS particularly has a number of safeguards in place to assure patients that their data is managed safely and securely and their rights to privacy and confidentiality are upheld, such as the NHS Constitution and the Data Sharing and Privacy Toolkit. These requirements are particularly strict when processing health data, which is considered special category data under the Data Protection Act 2018. Other legal requirements also need to be considered if dealing with NHS patient data.

The Data Protection Act 2018 replaced the Data Protection Act 1998, introducing new obligations that require integration of data protection concerns into every aspect of processing activities. This approach is known as ‘data protection by design and by default’. From a practical perspective, the important documents underpinning this are data flow maps, data protection impact assessments (DPIA) and privacy notices.
A good data flow map identifies the data assets (data at rest) and data flows (exchanges of data) that enable the relevant objective or initiative to be delivered. Where data flow mapping identifies instances where data is processed by a data processor on behalf of a data controller, a legally binding written data processing contract is required. This should include clauses appropriate to the processing risks identified (highlighted in the DPIA), as well as mandatory clauses for all data processing contracts.
The data flow map will then influence the DPIA as the vehicle by which proposed flows of personal identifiable data are governed, and the controls developed to ensure lawful processing. The vast majority of data processing in a health and social care context will involve special categories of data and it is therefore recommended that a full DPIA is carried out. A DPIA is intended to be a ‘living document’ and should be regularly reviewed and updated by programmes. If a risk is identified that cannot be mitigated, the ICO must be consulted before processing commences. They will normally provide advice within 8 weeks, or 14 weeks in complex cases.

Individuals have the right to be informed about the collection and use of their personal data. This is a key transparency requirement under the Data Protection Act 2018. A privacy notice should identify who the data controller is, with contact details for its data protection officer. It should also explain the purposes for which personal data is collected and used, how the data is used and disclosed, how long it is kept and the controller’s legal basis for processing.

Data sharing agreements or contracts are only valid between data controllers and are strongly recommended. They set out specific concerns relating to the data to be shared, as identified through data flow mapping and DPIA exercises.

Principle 5: Make use of open standards

Within the English health and social care system, information standards cover the specifications used to collect and extract data from information technology systems. Collections and extractions are defined data sets that can then be used to measure or conduct analysis of particular areas of interest.

NHS Digital currently hosts a range of data, clinical and interoperability standards for the health and social care network:

- information standards
- data collection
- technology clinical safety standards
- interoperability toolkit

Other data, clinical and interoperability standards:

- NHS England interoperability standards
- InterOpen current FHIR, Care Connect, HL7 and PRSB standards
- UK government open standards

Principle 6: Be transparent about the limitations of the data used

The data used must be well understood and reviewed for accuracy and completeness. Accuracy is the closeness of agreement between a data value and its true value. Completeness is the presence of the necessary data. NHS Digital publishes a quarterly data quality and maturity index, which provides data submitters with transparent information.

A 2-stage approach is suggested when applying analytics to any data. Algorithms should be trained to understand the levels of data quality first and then achieve their objective by using the variables given. This 2-stage approach should be built in so that high fluxes in data quality are handled appropriately.

Assessment of data quality should not be a one-off check – continuous anomaly detection should be in place to provide alerts to changes in a data source. NHS England and the UK Statistics Authority have produced guidance on data quality, which should be referred to. Be aware of potential biases in the data used for training algorithms – consider the representativeness of the database used for training the algorithm. If the data provided for the AI to learn is limited to certain demographic categories or disease areas, this could potentially limit the applicability of the AI in practice as its ability to accurately predict could be different in other ethnic groups.

Good data linkage will avoid reducing data quality

There is a range of approaches for linking data, which can provide differing levels of accuracy and data loss. It is often necessary to strike a balance between good matching accuracy and loss of too many records. Consideration should be given to the effects of a selected linkage procedure on data quality. In particular, if the process could cause systematic loss or mismatching of a particular type of record, this could have downstream implications in model assumptions and algorithm training.

Linking datasets may require those carrying out the linkage procedure to use identifiable data to match the data. It is therefore important to ensure that anyone with access to the identifiable data has a legal right of access. Similarly, the process of converting an identifiable dataset into an anonymised one, if conducted by a person, will need to be carried out by someone with a correct legal basis.

Where you can access data sets

There is a range of sources of health data:

- Public Health England collects a range of data, made available in different formats, for example their fingertips tool
- the Office for National Statistics collects a range of health-related microdata at their ONS virtual microdata lab
- UCI have built an open source training dataset for machine learning (Health Data Research UK are in the process of building further training datasets)
- Health Data Finder
- NHS Digital Data Access Request Service

Access must be requested for data that is not already in the public domain. The process for this varies depending on the organisation providing the data, and this should be detailed on the organisation's website. NHS Digital holds the responsibility for standardising, collecting and publishing data and information from across the health and social care system in England.

Training vs deployment

Be clear on the strengths and limitations of the training versus deployment data set. If the algorithm has been built on a training set and not yet deployed in a real-world clinical implementation, transparency should be shown to that effect. Demonstrate whether the algorithm is published in a real-world deployed environment or a training environment.

Principle 7: Show what type of algorithm is being developed or deployed, the ethical examination of how the data is used, how its performance will be validated and how it will be integrated into health and care provision

Consider how the introduction of AI will change relationships in health and care provision, and the implications of these changes for responsibility and liability. Use current best practice on how to explain algorithms to those taking actions based on their outputs.

When building an algorithm, be it a stand-alone product or integrated within a system, show it clearly and be transparent of the learning methodology (if any) that the algorithm is using. Undertake ethical examination of data use specific to this use-case. Achieving transparency of algorithms that have a higher potential for harm or unintended decision-making, can ensure the rights of the data subject as set out in the Data Protection Act 2018 are met, to build trust in users and enable better adoption and uptake.

Work collaboratively with partners, specify the context for the algorithm, specify potential alternative contexts and be transparent on whether the model is based on active, supervised or unsupervised learning. Show in a clear and transparent specification:

- the functionality of the algorithm
- the strengths and limitations of the algorithm (as far as they are known)
- its learning methodology
- whether it is ready for deployment or still in training
- how the decision has been made on the acceptable use of the algorithm in the context it is being used (for example, is there a committee, evidence or equivalent that has contributed to this decision?)
- the potential resource implications

This specification and transparency in development will build trust in incorporating machine-led decision-making into clinical care.

Principle 8: Generate evidence of effectiveness for the intended use and value for money

When building or developing the technology, consider what function the product delivers – this will guide what sort of evidence generation plan should be put in place. A functional categorisation of digital tools is emerging, whereby the functions can be graded in terms of clinical impact and risk of harm. The expected evidence standards will increase as the risk associated with the technology grows. For those products that perform more than one function, the evidence requirements for the function that carries the highest potential impact and highest potential harm should be met.

NICE has developed an initial version of an evidence standards framework for digital health technologies, working in close collaboration with NHS England, NHS Digital, Public Health England, MedCity and other stakeholders. These standards inform technology developers and evaluators about which types of evidence should be expected, taking into account the functions and intended use of the product (as specified in principles 1 and 2) and its overall economic impact. A downloadable template is available to help with budget impact analysis.

Principle 9: Make security integral to the design

A core element of at-scale adoption and uptake is to ensure that security and data protection methodology have been incorporated. NHS Digital has launched a new Data Security and Protection Toolkit to replace the previous Information Governance Toolkit to ensure that patient information is kept safe. All organisations that have access to NHS patient data and systems – including NHS trusts, primary care and social care providers, and commercial third parties – must complete the toolkit to provide assurance that they are practising good data security and that personal information is handled appropriately. NHS Digital's Data Security Knowledge Library also contains a number of documents that are relevant to helping secure medical devices.

The Department for Digital, Culture, Media and Sport has published a code of practice for consumer 'internet of things' (IoT) security, which sets out practical steps that will help manufacturers to improve the security of consumer IoT products and services and ensure products are secure by design. It is applicable to consumer IoT generally and is not specifically targeted towards health devices and systems.

The National Data Guardian's 10 data security standards are in place and form part of the NHS Standard Contract that goes out to all providers. The standards are set out in full in the National Data Guardian's Review of data security, consent and opt-outs.

When developing an application, ensure the product meets the OWASP Application Security Verification Standard, which is used to establish a level of confidence in the security of web applications.

The new EU regulations on medical devices and in vitro medical devices will apply from May 2020. It will give the Medicines and Healthcare Products Regulatory Agency (MHRA) increased oversight of connected medical devices, and will improve the cyber security of diagnostic equipment (and other connected medical devices) in the longer term. The UK government has publicly stated its desire to retain a close working partnership in respect of medicines regulation after the UK leaves the EU.

Principle 10: Define the commercial strategy

The foundation of any commercial structure should be to ensure that the terms of the engagement fairly allocate the benefits between the parties based on their respective contributions, roles, responsibilities, risks and costs. When the basis of the commercial arrangement is NHS data, it must adhere to the revised guiding principles described in Creating the right framework to realise the benefits of health data. We want to hear views from patients, the public and partners on these principles as we develop the final, full policy framework to be published later this year.

Before entering into any commercial arrangement, the problems that need solving, and who for, including any long-term vision, should be fully understood by all parties. The relationship between the parties will be set out in a binding legal contract that will impact on all parties both during and long after the lifetime of the contract. This will require legal advice.

Before engaging with the legal teams, the following should be considered:

- proportionality
- scope
- exclusivity
- value
- ownership of intellectual property
- liability
- audit
- bias
- roles

Proportionality

The extent to which parties can use an off-the-shelf solution for their project needs to be understood as much for the choice of commercial structure as for the choice of technology. For commercially available (or modified) off-the-shelf solutions that are remote from patients and patient data, current NHS commercial arrangements may be sufficient. For more complex data-driven technology engaging directly with patients or patient data, such as AI, this principle should be referred to in full.

Scope

Based on the objectives of the health and care providers, and the wider health and care system, it is important to determine if a local or system-wide approach should be taken.

Exclusivity

Careful consideration should be given before granting exclusivity of access to data, as exclusivity can limit benefit to the health and care system.

Value

Financial value can be realised for the NHS through numerous models, such as simple royalties, free or reduced payments for products, equity shares in the business and improved data sets that can be offered (at a price) to others. However, that is only a part of the value equation. Technology providers derive significant value from the NHS beyond access to unique data sets – through medical and clinical involvement, test beds and pilots – and this value should be captured within the commercial arrangement.

Further, technology providers may be in a position to generate significant value from any products (or even just from the learning) outside the scope of the initial project, and this value opportunity must be recognised too. Above all the value to patients, clinicians and the wider health and care system should be clearly articulated.

Ownership of intellectual property

There are numerous models in use across this area, broader medical and pharmaceutical research, and across academia. The most appropriate model depends on a host of factors. This is particularly important where the same algorithm is used in multiple applications; the participant (as data source) in a particular application cannot hope to own the underlying algorithm but might expect to share in the increase in value.

Liability

If the intended application is such that someone (an individual or institution) can suffer loss or harm if something goes wrong, clear terms around who bears the liability should be established.

Audit

There needs to be clarity on the extent to which the product will be transparent and can be audited. Both parties need to be sure that such auditing is allowed and required but also that both have the capability and capacity to do this.

Bias

The single greatest threat to reliance on data-driven technology is the actual or possible presence of bias. Any commercial arrangement will need to identify where this manifests and how it is managed, by whom, and at whose cost.

Roles

Traditional research collaborations (of which there are many examples across the NHS) have an established operating model around what each party contributes. Partnerships for the development of data-driven technologies may require completely new roles. Achieving clarity in advance of what contribution is expected of each party is of central importance

This is by no means a complete list. The range of engagements with data-driven technology providers is such that there is no single set of commercial considerations. Consider types of partnerships and commercial models as outlined in Making NHS data work for everyone, Figure 10.

The final element of a complete commercial structure is a system that ensures all levels of governance of both parties fully understand the technological, commercial and legal implications of the engagement. Systems need to be developed and maintained that enable frequent reporting, review and challenge of data-driven technologies, which are ever-evolving.