THE ROYAL SOCIETY

Creating resilient and trusted data systems

The public perspective and recommendations for action

Background

This note provides a summary of a public dialogue commissioned by the Royal Society to explore public attitudes in the UK relating to the sharing of data in emergency and non-emergency situations. In addition, it sets out recommendations for creating trusted and resilient data systems informed by the public dialogue report; previous Royal Society initiatives; an online workshop in October 2021; and a workshop held at the Society in October 2022. The public dialogue was conducted by Hopkins Van Mil and supported by an informal steering group chaired by Professor Chris Dye FRS FMedSci.

Introduction

As evidenced throughout the COVID-19 pandemic, scientists and decision-makers benefit from rapid access to high quality data in a fast-changing, emergency environment. There were a number of data-driven initiatives[1] set up in response to the pandemic. For example, the RECOVERY trial used secure, linked datasets to rapidly enrol patients onto a clinical trial that enabled the discovery of dexamethasone as a life-saving treatment in a matter of months. A longitudinal study which linked health data with 2011 Census data revealed that minority ethnic groups suffered disproportionately greater levels of mortality from the first two waves of COVID-19 in the UK[2]. A joint review by the Royal Society and the Ada Lovelace Institute identified some of the factors which enabled the success of such initiatives: Having a clear purpose and specific mission from the outset; generating public confidence in data and how it is used; creating shared processes and data alliances; and providing support to data and evidence-led responses to the pandemic[3].

However, there were some instances where challenges in accessing and sharing data hampered the ability of advisers and decision-makers to understand the situation and recommend actions. Much progress has been made over the past two years to try to understand and overcome these challenges. There is now a critical opportunity to understand whether the systems we have in place now will be effective in supporting a data-led response to emergencies of different types and whether data can be used in a trusted way during both emergency and non-emergency situations.

Led by Professor Christopher Dye FRS FMedSci, *Creating Resilient and Trusted Data Systems* positions the public's perspectives, priorities, and concerns at the heart of these issues. Reflecting on the findings of a UK-wide public dialogue commissioned by the Royal Society, this statement sets out five recommendations to help ensure that data systems generate or retain public confidence. Mistrust in, or lack of engagement with, data could hinder the benefits of a truly effective data system – one which delivers value to individuals and communities, drives innovation, and responds to challenges. To prevent this, it is important to understand what the public expect from data systems, their priorities and how this may change across different emergency and non-emergency situations.

1 Health Data Research UK. Five data-driven insights from our urgent COVID-19 projects. See: https://www.hdruk.ac.uk/news/five-data-driven-insights-from-our-urgent-covid-19-projects (accessed 13 December 2022).

2 Nafilyan *et al*. 2021 Ethnic differences in COVID-19 mortality during the first two waves of the Coronavirus Pandemic: a nationwide cohort study of 29 million adults in England. MedRxiv. (https://doi.org/10.1101/2021.02.03.21251004).

3 Ada Lovelace Institute and the Royal Society. 2021 Learning data lessons: data access and sharing during COVID-19. See: https://royalsociety.org/-/media/policy/Publications/2021/learning-data-lessons-data-access-and-sharing-during-COVID-19.pdf (accessed 13 December 2022).

Both the public dialogue and the recommendations set out in this statement are focused on the data-sharing environment in the UK. It is important to note, however, that global lessons[4, 5] and initiatives[6] on data-sharing in emergencies will also be critical. Data in this context refers to statistics (eg population density by accommodation type), facts (eg temperature), forecasts (eg predictions based on modelling), and insights (eg social media intelligence). Adopting the definition set out in the Civil Contingencies Act 2004, emergencies refer to 'events that could cause or threaten serious damage to human welfare or the environment'[7].

The Royal Society's public dialogue on 'creating resilient and trusted data systems'

The Royal Society commissioned the public facilitation agency Hopkins Van Mil to deliver a public dialogue to explore the public's views on data systems during emergencies and non-emergencies. The dialogue format was chosen to facilitate an immersive and informed discussion, where a full range of viewpoints could be shared, exploring nuanced views, trade-offs and 'least-regret' options. The full report of the dialogue (which includes its methodology) accompanies this statement.

The public dialogue addressed the following questions:

a) Do the current systems in place support a trusted and effective response to emergencies?

b) Have the systems been established in ways that enable them to be used in a trusted way outside of emergencies?

c) Are we any better placed to put in place a data-led response to other emergencies?

The key findings from the dialogue are as follows:

1. Data is a complex, personal and often emotive issue for people. The dialogue unearthed varying perspectives on data, often changing through the course of the discussions as participants considered advantages and risks of their data being used. Initial perceptions on data centred around fears from data breaches and personal data unknowingly being used within the private sector. These perceptions shifted as awareness grew around the uses of data for public benefits and for emergency responses. By the end of the dialogue, participants had a strong belief in the wider public and community benefit of data, and its importance in emergency situations.

2. Trust in data systems comes from there being clarity of purpose for that data, transparency in data flows, and knowledge of who owns, controls and governs that data. The public need to have confidence in the existence of mechanisms that ensure compliance with the law and standards or opt-out mechanisms. Participants considered the features of good governance to be i) independence, ii) effective sanctions against misuse, iii) scrutiny, and iv) safeguarding.

3. Participants expect robust systems to be in place to enable swift data-led responses to emergencies while retaining security and privacy of data.

4 OECD. 2022 Public communication and engagement in science: lessons learned from COVID-19. See: https://www.oecd.org/science/inno/public-communication-engagement-in-science.htm (accessed 13 December 2022).

5 UK Government. 2022 Intelligent open science: viral genomic data sharing during the COVID-19 pandemic. See: https://www.gov.uk/government/publications/intelligent-open-science-viral-genomic-data-sharing-during-the-covid-19-pandemic (accessed 13 December 2022).

6 The Global Pandemic Data Alliance. See: https://gpdahub.org/ (accessed 13 December 2022).

7 UK Government. 2011 Keeping the country running: natural hazards and infrastructure. See: https://www.gov.uk/government/publications/keeping-the-country-running-natural-hazards-and-infrastructure (accessed 13 December 2022).

4. Participants expect society to be prepared for emergencies. Participants feel that the data that would be needed in an emergency response should be collected before it happens. During an emergency, participants expect to see the data previously collected used to save lives, protect those at risk and communicate what's happening to those affected. There is also an opportunity to identify and fill gaps in the system, where there may be missing data. Following the emergency situation, participants expect all the data gathered before and during the crisis to be assessed, reviewed and evaluated so that that whole system can become a learning tool and continue the cycle of preparing for the next emergency. Emergencies, and the limits in which emergency powers for processing data exist, should be clearly defined.
5. Participants concluded that connected data systems are of value to society in both emergency and non-emergency situations. They call for more co-ordination and inter-operability between systems, particularly those which deliver public benefit such as health, care, housing, and education.
6. Participants proposed that data systems should be designed with inclusion and diversity in mind including:
    - Involving a diversity of people in the design of data systems
    - Standardising the design of systems, particularly those in the public sector so that it is easier to move from one to another
    - Having dedicated and specialist teams responsible for system accessibility.
7. Participants proposed the following features to ensure the system is trusted, effective and resilient in the event of future emergencies:
    - Conducting stress tests focused on potential risks, checking readiness for a range of emergency and non-emergency situations
    - Future proofing to anticipate likely challenges ahead and ensure data is available on relevant topics
    - Finding innovative and creative ways to make full use of existing data considering new data sources such as big data and anticipating future data needs
    - Learning from expert and vetted staff who provide the best expertise available to support learning, development and technical innovation within data systems.

Recommendations

The following high-level recommendations reflect on the findings of the public dialogue as well as discussions held during a workshop at the Royal Society on 19 October 2022 with experts in data policy from across UK Government departments and agencies, and leading civil society organisations. A list of organisations represented at the workshop can be found in the annex.

Taken together, the recommendations seek to embed public engagement on data as a routine activity, to provide clarity for data controllers in emergency situations, and to strengthen the resilience of data systems in the public and private sector.

Area of action: Public engagement

Public acceptability is key to the beneficial use of public data, in both emergencies and non-emergency situations. Transparency and building a shared understanding of how data is used is a key element of building a trusted system. However, the public awareness of data systems, how data is used and how data is governed is relatively low. There have been a number of initiatives seeking to address this, including the Understanding Patient Data programme[8] and the UK Government's use of data dashboards during the COVID-19 pandemic.

Recommendation 1

Public engagement should form a core part in the development of new data policy. To achieve this, prior engagement initiatives should be evaluated in order to inform a new, continual programme of engagement focused on increasing public understanding of how sensitive data and statistics are used, scrutinised and regulated.

Area of action: Data protection guidance

Uncertainty exists for how and when data protection exemptions apply for emergency situations. For example, there may be data which exists that would indirectly be useful during an emergency yet not qualify for exemptions under UK GDPR. There is a risk that the perceived or actual risks of breaching data protection law, intellectual property rights or regulatory requirements may be deterring organisations from using or accessing data.

Recommendation 2

In its planned review of data protection legislation, the UK Government should consider providing greater clarity on what types of data can be collected for emergency responses. This should cover both predictable emergency events (eg annual flooding) as well as potential black swan events. Following any changes to the UK's data protection legislation, the Information Commissioner's Office should develop guidance on this for key stakeholders in both public and private sector organisations who may have to share data during an emergency situation.

Area of action: Stress testing

As highlighted in the public dialogue, there is an expectation that useful data for emergencies is collected in advance and used to save lives and livelihoods. Furthermore, it is expected that data used is both accurate and representative. This will require continuous preparation and data systems in both the public and private sector will need to be regularly reviewed as a core part of emergency planning functions.

Recommendation 3

The UK Government should conduct regular stress tests of the data readiness[9] of public and private sector organisations in relation to the risks and hazards described in the National Risk Register and incorporate this into the assessment of risks in the National Security Risk Assessment. This could be led by the National Situation Centre (SitCen) and may require the adoption of privacy enhancing technologies for commercially sensitive data[10]. An important factor which should be considered in stress tests is the implications of power outages for data access.

8 Understanding Patient Data. See: https://understandingpatientdata.org.uk/ (accessed 13 December 2022).

9 The Royal Society DELVE Initiative. 2020 Data Readiness: Lessons from an Emergency. See: https://rs-delve.github.io/reports/2020/11/24/data-readiness-lessons-from-an-emergency.html (accessed 13 December 2022).

10 The Royal Society. 2023 From privacy to partnership: The role of privacy enhancing technologies in data governance and collaborative analysis. See: https://royalsociety.org/topics-policy/projects/privacy-enhancing-technologies/.

Area of action: Standardisation

There was strong support amongst the public dialogue participants for secure, linked data systems across multiple sectors, particularly within emergency situations and non-emergency healthcare settings, as long as protections are in place to preserve privacy and prevent misuse or data breaches. Given the diversity of datasets and their controllers within the public sector, there is a need for standardisation of formats, processes, and governance mechanisms to ensure that organisations can share and utilise data effectively.

Recommendation 4

The UK Government should review the technological capabilities and data standards across government departments, devolved administrations, local resilience forums, and arms-length bodies to inform a strategy for enabling consistent and secure sharing of data across key public sector organisations.

Area of action: Trusted research environments

Participants highlight the importance of secure systems, particularly in times of crises. They also expect the right data to be available to the right people when needed during emergencies. The Trusted Research Environment (TRE) model is an established and effective model for health data research, which enables accredited researchers secure access to datasets held within a platform. Examples include the Office for National Statistics' Secure Research Service[11], the Secure Anonymised Information Linkage (SAIL) Databank[12], and the Electronic Data Research and Innovation Service[13] (eDRIS).

Recommendation 5

The UK Government should explore mechanisms for expanding the use of TREs for data likely to be useful during an emergency and for strengthening their resilience (eg through the development of remote access to TREs and addressing skills shortages).

11 Office for National Statistics. About the Secure Research Service. See: https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice/aboutthesecureresearchservice (accessed 13 December 2022).

12 Welsh Government. 2020 SAIL Databank – health data research during a global pandemic. See: https://digitalhealth.wales/sail-databank-blog (accessed 13 December 2022).

13 Edinburgh International Data Facility. eDRIS: Enabling research access to Scottish health datasets. See: https://www.ed.ac.uk/edinburgh-international-data-facility/updates-events/electronic-data-research-and-innovation-service (accessed 13 December 2022).

Annex

This work builds on a number of Royal Society data-led initiatives that are described below:

Data Evaluation and Learning for Viral Epidemics (DELVE)

The Royal Society's DELVE group[14] was formed in response to the COVID-19 crisis with a remit to explore specific policy questions in a data-driven manner. The DELVE Initiative was established with the ambition that data science could play a role in helping develop policy responses to the COVID-19 pandemic, by identifying lessons from the responses of other countries or by combining datasets to generate novel insights.

DELVE produced a 'Data Readiness: Lessons from an Emergency' report which laid out some recommendations following its experience of gathering and evaluating data during the Covid-19 crisis[15]. The recommendations included updating the statutory objective of the Office for National Statistics (ONS) to accommodate trustworthy access to 'happenstance' (circumstantial and unplanned) data and that the ONS should collaborate closely with the Information Commissioner's Office to formulate a standardised qualification for data access, equivalent to a 'data driving license' that would demonstrate trustworthiness and ensure that qualified experts can get rapid access to different data types with the appropriate standardised ethical and legal training in place.

Rapid Assistance in Modelling the Pandemic (RAMP)

The Rapid Assistance in Modelling the Pandemic (RAMP) initiative set up by the Royal Society attracted 18,000 volunteers and supported rapid-response, peer review of publications and new cross-sector modelling teams[16].

Science in Emergencies Tasking – COVID-19 (SET-C)

The Royal Society established SET-C to draw on the expertise of our Fellows and others to respond to requests for rapid science advice on topics relevant to tackling the pandemic[17]. The group responded rapidly to convene data, evidence and advice on priority topics, ranging from masks to vaccines, and relied extensively on streamlined access to research and data.

UK and international context

We are now in a critical time of reviewing our data systems and assessing whether the systems we have in place now put us in a strong position for a data-led response to future emergencies. The Sendai Framework for Disaster Risk Reduction 2015–2030 report[18] called for a 'data revolution' alongside 'rigorous accountability mechanisms and renewed global partnerships.'

The UK's National Data Strategy[19] aims to unlock the power of data across the economy, which includes improving productivity, supporting innovation, transforming the public sector and the delivery of public services and creating a fairer and more inclusive society. The National Data Strategy draws on examples of the value of data in responding to the pandemic and commits to the creation of an appropriately safeguarded, joined-up and interoperable data infrastructure to support future data-led responses across government.

14 The Royal Society DELVE Initiative. See: https://rs-delve.github.io/about.html (accessed 13 December 2022).

15 The Royal Society DELVE Initiative. 2020 Data Readiness: Lessons from an Emergency. See: https://rs-delve.github.io/reports/2020/11/24/data-readiness-lessons-from-an-emergency.html (accessed 13 December 2022).

16 Rapid Assistance in Modelling the Pandemic. See: https://royalsociety.org/topics-policy/Health%20and%20wellbeing/ramp/ (accessed 13 December 2022).

17 Science in Emergencies Tasking – COVID. See: https://royalsociety.org/topics-policy/projects/set-c-science-in-emergencies-tasking-covid/ (accessed 13 December 2022).

18 United Nations Office for Disaster Risk Reduction. 2015 Sendai Framework for Disaster Risk Reduction 2015-2030. See: https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030 (accessed 13 December 2022).

19 UK Government. 2020 National Data Strategy. See: https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy (accessed 13 December 2022).

Public attitudes to data

This work takes place in the backdrop of other reviews of public attitudes to data. The Ada Lovelace Institute reviewed evidence around public attitudes to data regulation in May 2022[20]. One of the main findings of this review was that 'the UK public want data-driven innovation, and they expect it to be ethical, responsible and focused on public benefit.' The Ada Lovelace Institute's July 2022 report, 'The rule of trust: Findings from citizens' juries on the good governance of data in pandemics'[21], found seven overarching principles for trustworthy data governance which echoed with the Royal Society's findings, particularly around transparency, communication and clarity, accountability, equity, inclusiveness and non-discrimination and emergency preparedness. DARE UK's 'Building a trustworthy national data research infrastructure: A UK-wide public dialogue'[22] recommended similar themes, including transparency, public involvement and engagement and awareness, within a data for research context.

The Royal Society workshop

On 19 October 2022, the Royal Society hosted a workshop on Creating Resilient and Trusted Data Systems.

Representatives from the following organisations were in attendance:

Workshop participants
Ada Lovelace Institute
Centre for Data Ethics and Innovation
Central Data and Digital Office
Department for Digital, Culture, Media and Sport
Department for Environment, Food, and Rural Affairs
Environment Agency
Government Office for Science
Health Data Research UK
Hopkins Van Mil
Information Commissioner's Office
Met Office
National Situation Centre
Office for National Statistics
UK Health Security Agency
UK Research and Innovation

Royal Society staff

Royal Society secretariat
Areeq Chowdhury, Head of Policy, Data
Alexandra Wakefield, Senior Policy Adviser (until October 2022)

20 Ada Lovelace Institute. 2022 Who cares what the public think? UK public attitudes to regulating data and data-driven technologies. See: https://www.adalovelaceinstitute.org/evidence-review/public-attitudes-data-regulation (accessed 13 December 2022).

21 Ada Lovelace Institute. 2022 The rule of trust: Findings from citizens' juries on the good governance of data in pandemics. See: https://www.adalovelaceinstitute.org/report/trust-data-governance-pandemics (accessed 13 December 2022).

22 DARE UK (Data and Analytics Research Environments UK). 2022 Building a trustworthy national data research infrastructure: A UK-wide public dialogue. See https://dareuk.org.uk/public-dialogue-building-a-trustworthy-national-data-research-infrastructure (accessed 13 December 2022).