



Institute
and Faculty
of Actuaries

IFoA response to UK National Data Strategy Consultation

The Institute and Faculty of Actuaries (IFoA) is a royal chartered, not-for-profit, professional body. We represent and regulate over 32,000 actuaries worldwide, and oversee their education at all stages of qualification and development throughout their careers.

National Data Strategy
DCMS
100 Parliament Street
London
SW1A 2BQ

2 December 2020

Dear Sir/Madam,

IFoA response to UK National Data Strategy consultation

The Institute and Faculty of Actuaries (IFoA) welcomes the opportunity to respond to the Department for Digital, Culture, Media and Sport's consultation on the UK National Data Strategy consultation.

The IFoA welcomes the general direction of the Government's proposals. At this stage, these proposals are still at a very high-level and it may be at the next stage of the consultation where the profession is able to provide more detailed input into the Government's thinking. However, we have identified a number of areas that we believe it would be useful to start thinking about at this stage to help shape the debate and policy considerations going forward. These specific considerations are set out below some introductory remarks on why data is key to the work of our members and the customers they serve.

General comments

Data science's potential comes from not only gathering ever more data, but also from the growth in analytic capability, and the ability to see people and processes in much more detail than ever before. Improved analytics are also identifying relationships which would otherwise remain hidden in data.

In addition to its impacts on modern life, data is transforming processes for financial institutions, businesses, governments, consumers and wider society. Data also has significant potential to promote innovation across a range of areas in which our members practice, including general, life and health and care insurance, as well as other emerging fields. However, as it delves ever deeper in our lives, data science raises significant questions over ethics and the public interest. These are challenges the IFoA has been considering in an insurance context for over 150 years, as actuaries are the original data scientists.

Data science and artificial intelligence (AI) have the potential to greatly enhance the tools and capacities of those working in statistics and actuarial science, providing for new approaches to solving longstanding problems and opening up new sectors and industries to our members. This great potential impact can lead to

Beijing	14F China World Office 1 · 1 Jianwai Avenue · Beijing · China 100004 · Tel: +86 (10) 6535 0248
Edinburgh	Level 2 · Exchange Crescent · 7 Conference Square · Edinburgh · EH3 8RA · Tel: +44 (0) 131 240 1300
Hong Kong	1803 Tower One · Lippo Centre · 89 Queensway · Hong Kong · Tel: +852 2147 9418
London (registered office)	7th Floor · Holborn Gate · 326-330 High Holborn · London · WC1V 7PP · Tel: +44 (0) 20 7632 2100
Oxford	1st Floor · Park Central · 40/41 Park End Street · Oxford · OX1 1JD · Tel: +44 (0) 1865 268 200
Singapore	163 Tras Street · #07-05 Lian Huat Building · Singapore 079024 · Tel: +65 6906 0889

www.actuaries.org.uk

difficult ethical challenges for practitioners and calls for a need for strong ethical values and professionalism if the public interest is to be supported. The profession has considered both the opportunities and challenges that the increased prevalence of data poses at great length; [our 2017 policy briefing](#) goes into further detail.

A key objective of our work on data science is to provide an actuarial perspective on the ethical and wider public interest issues relating to data science in insurance and other actuarial practice areas. We have collaborated with the Royal Statistical Society (RSS) to produce a paper, '[A Guide for Ethical Data Science](#)', setting out the practical and ethical implications of data science which continues to reflect our views. Our work focuses not only on the potential risks to the public interest associated with the increased use of data science, but also how data science can harness innovation in the financial sector.

Response to consultation questions

Question 1

We are supportive of missions and pillars of the NDS as described which we believe provides a framework focused on economic growth balanced against the public interest. For example, from a life insurer/reinsurers viewpoint being able to readily access information such as medical information for underwriting (for Health & Care and life insurance) will be transformative to enable the industry to grow the market and to make it easier for customers to obtain cover quickly. Also, if we can price the risks more accurately through better access to data then we should be able to reduce capital requirements and deploy capital more efficiently which will ultimately lead to lower premiums to the customer.

Question 2

Good quality pandemic data is important for assessing long-term liabilities in several areas. This includes Life insurance business (mortality, longevity, critical illness, income protection) as well as non-life business (business continuity, motor insurance, travel insurance etc).

In practice the most significant 'data delivery' improvement that the government could provide is to ensure that data is made available in seriatim + attribute record format so that it can be readily linked to other data sources that have the same seriatim + attribute data fields.

Providing data in a format that includes key seriatim fields – specifically for infection cases – should always include Deemed Date of Infection, and Date of Birth. Where any government agency collects data, it should add the key seriatim field. For example:

- HOSPITAL admission – Deemed Date of Infection, Date of Birth, Date of Hospital Admission
- ICU Admission – Above plus - Date of ICU Admission
- ICU Discharge – Above plus - Date of ICU Discharge
- HOSPITAL discharge – Above plus - Date of Hospital Discharge
- HOSPITAL Death – Above relevant + Date of Death
- REACT antigen incidence – Date of Test, Deemed Date of Infection
- REACT antibody incidence – Date of Test, Deemed Date of Infection (although we acknowledge the challenges with this).

Further detail is provided in Annex A.

Where data is provided in aggregate groupings, for example age bands, it is recommended to have uniform reporting groups across cases, infections, deaths and for different regions. Where individual-level data sharing is not practical, it is best to provide as many important stratifiers as possible, for example: death data by region, age, gender, socio-economic status and comorbidities.

A scalable database with real-time and a consistent data capture is essential to allow its users to monitor the true impact of the pandemic. More broadly, the data ecosystem should be upgraded and made transparent to reduce the risk of data failure, such as those arising from Excel data format.

Technologies such as the [OpenSAFELY](#) platform, a secure analytics platform storing pseudonymised electronic health records, should be made available more widely (e.g. to actuaries and researchers) to allow timely access to data and to deliver urgent morbidity and mortality insights. Such data would enable agile analysis of infections, hospitalisations, treatments, deaths, risk factors, vulnerable groups and emerging trends. This helps to inform data-driven actions/mitigations with quantifiable benefits.

Furthermore, more up-to-date information on mental health would be useful to the industry and other stakeholders, particularly as we emerge from the Covid-19 pandemic and seek to understand its wider impacts on the population. To provide one example, there is currently very little up-to-date or detailed information on the number of suicides or data on people who are at risk or seeking treatment. This information would ideally be available at a granular level so we can understand the drivers for mental health issues and modify our products and services accordingly.

Over and above health data, and the publicly available google mobility data (which is flawed), any provision of national or preferably regional data on key economic metrics would be useful. Any metrics on economic activity, transport passenger volumes, motor vehicle numbers, import/export application numbers provided on a weekly regional basis would also provide valuable insights. Ultimately the impact on economic activity, employment, insurance provision is about human movement rather than about Covid itself.

For maximum value this data could be compared to the similar period in previous years. Here it is important to compare like for like – e.g. comparing Easter week which is a different week each year. Something like the google mobility data which calibrates off an average from January does not allow for the ‘normal’ fluctuations during the year, and therefore invalid to use for economic activity impact.

Question 3

A risk of the increased use of Big Data and automated AI algorithms to society may introduce bias. This may in turn lead to bias in the system disadvantaging people with protected characteristics, and potentially other vulnerable groups. Our paper referenced above, [‘A Guide for Ethical Data Science’](#), sets out a framework that could form the basis for guarding against these unintended consequences. Although the risk could be regulated against to some extent, another key ingredient is the diversity of the workforce creating the output. Proactively promoting diversity across the full data ecosystem, including its senior leadership, from its initiation should lead to better outcomes in terms of product design, marketing and overall service for underrepresented groups.

Question 4

In terms of regional inequalities, we believe the provision of data on the 343 local authority levels is critical. National and UK-wide data tends to mask any underlying trends and developments. National figures are typically a blended average of the local authority figures and could therefore be misleading and not granular enough for decision-making at a local level. Heat maps of ranges of data at local authority level is useful, and data feeds of such data is then useful for analysis.

When reporting on regional inequalities, it is important to account for any confounding factors and make an effort to disentangle them for actionable insights. For example, identifying the extent to which regional inequality is driven by the mix of socio-economic status, ethnicities, age groups, population density etc needs to be reported and considered alongside the observed inequalities. In short it is critical that it is possible to distinguish between association and causal effects.

Relating to the reach and impact of the proposals outlined in the consultation, it is prudent to consider different regional preference of communication needs and channels, education level and availability of resources.

Should you want to discuss any of the points raised please contact Henry Thompson, Head of Public Affairs (henry.thompson@actuaries.org.uk) in the first instance.

Yours sincerely,

Scott Reid
Chair, Health & Care Research Committee
Institute and Faculty of Actuaries

Annex A

As mentioned above, in practice the most significant “data delivery” improvement that the government could provide is to ensure that data is made available in seriatim attribute record format so that it can be readily linked to other data sources that have the same seriatim attribute data fields. This would radically improve the analytics and ability to extract critical information across many industries and social situations. A key seriatim field for this purpose would be “Deemed Date of Infection”. If attribute data is then collected and included with this on records it then becomes possible to generate highly informative analytics. Data points/sets that it would be useful include are:

Exposed to risk Population records are likely available (or readily derived) as follows:

Deemed date of infection
Date of birth (to derive age, etc) +
Attributes:
Sex
Ethnicity
Socio/economic class – Deprivation Index
Country
Region
Local Authority
...

If Event Information is collected and made available with consistent link fields (including seriatim date) as well as other attributes – rich trend and incidence information can be derived – e.g.

Random antigen surveys (Like REACT):

Date of antigen Test
Antigen Test Result
Deemed date of infection
Date of birth (to derive age, etc) +
Attributes:

Sex
Ethnicity
Socio/economic class – Deprivation Index
Country
Region
Local Authority
...

Similarly with Hospital Admissions

Date of Hospital Admission
Deemed date of infection
Date of birth (to derive age, etc) +
Attributes:
Sex
Ethnicity
Socio/economic class – Deprivation Index
Country
Region
Local Authority
Fields for co-morbidities (Cardiac condition, Cancer categories, diabetes, stroke.....
...

Hospital discharge

Date of Hospital Discharge
Date of Hospital Admission
Deemed date of infection
Date of birth (to derive age, etc) +
Attributes:
Sex
Ethnicity
Socio/economic class – Deprivation Index
Country
Region
Local Authority
Fields for Treatment (PAP, intubation, drugs)
Fields for co-morbidities (Cardiac condition, Cancer categories, diabetes, stroke etc)
...

Hospital Deaths

Hospital Date of Death
Date of Hospital Admission
Deemed date of infection
Date of birth (to derive age, etc)
Attributes:
Sex
Ethnicity
Socio/economic class – Deprivation Index
Country
Region
Local Authority
Fields for Treatment (PAP, intubation, drugs)
Fields for co-morbidities (Cardiac condition, Cancer categories, diabetes, stroke)
...

Similarly for **ICU Admission/discharge**.

Data made available in this way can be associated with a range of other data – travel data, restaurant/pub visits, google movement data, etc.

Data is collected by different groups of people for different purposes. These disparate groups are not coordinated, but if they have collected data with “relevant” seriatim and other attributes it is possible to combine their data with data from other sources. Mostly the attribute information (age/sex/ethnicity) is common, but the seriatim information (date of Birth, Deemed Date of Infection, Date of Hospital admission, Date of Discharge, Date of Death), etc is not being made available or consistently collected – e.g. data collected for hospital admission is typically not recording Deemed Date of Infection.