TONY DAVIES COLUMN

A national data strategy

Antony N. Davies

SERC, Sustainable Environment Research Centre, Faculty of Computing, Engineering and Science, University of South Wales, UK

Continuing our information age/mis-information age theme, did you know that the UK has a Department for Digital, Culture, Media & Sport (DCMS)? Apart from having Secretaries of State who tend to have surprisingly short periods of occupancy in their roles, they have issued something which could build into something of great value to scientists in years to come—the National Data Strategy.[1]

The UK National Data Strategy

The strategy policy paper recognises that "data is now the driving force of the world's modern economies. Fuelling innovation in businesses large and small and has been a lifeline during the global coronavirus pandemic". But usefully it also recognises that there must be public trust in the data itself.

For all of us who want to see more of the data that is generated by projects funded through our taxes being made available for others, this is good news. Data is an essential resource. The policy paper recognises early on that "Data is a non-depletable resource in theory, but its use is limited by barriers to its access—such as when data is hoarded, when access rights are unclear or when organisations do not make good use of the data they already have." And goes on… "We will drive an approach to data that holds that all can benefit when data

is used responsibly, and that withholding data can negatively impact society." It is not only about persuading data generators, out of the goodness of their hearts, to make data more openly available. Clearly there is a need to protect people's data rights and private enterprises' intellectual property, but there is also the implication that by not making data available they are actually damaging the society which is paying for their work.

The Four Pillars of the strategy

The Four Pillars of the national strategy could almost have been written as a distillation of the messages and pleading coming from this column over many years. They clearly find resonance in our own field and include support for the FAIR data principles, standardised formats and futureproofing. They also call for increasing education in these key areas.

1) Data foundations: The true value of data can only be fully realised when it is fit for purpose, recorded in standardised formats on modern, future-proof systems and held in a condition that means it is findable, accessible, interoperable and reusable. By improving the quality of the data, we can use it more effectively, and drive better insights and outcomes from its use.

2) Data skills: To make the best use of data, we must have a wealth of data skills to draw on. That means delivering the right skills through our education system, but also ensuring that people can continue to develop the data skills they need throughout their lives.

3) Data availability: For data to have the most effective impact, it needs to be appropriately accessible, mobile and re-usable. That means encouraging

better coordination, access to and sharing of data of appropriate quality between organisations in the public, private and third sectors, and ensuring appropriate protections for the flow of data internationally.

4) Responsible data: As we drive increased use of data, we must ensure that it is used responsibly, in a way that is lawful, secure, fair, ethical, sustainable and accountable, while also supporting innovation and research.

It is not my intention to cite all 47 pages of the strategy, and much of what you will find in the document discusses data strategies around the government's own use of data which is not strictly relevant here. However, there are five Missions building on the pillars which inform the actions to be taken as a result of the data strategy.

- Mission One: Unlocking the value of data across the economy
- Mission Two: Securing a pro-growth and trusted data regime
- Mission Three: Transforming government's use of data to drive efficiency and improve public services
- Mission Four: Ensuring the security and resilience of the infrastructure on which data relies
- Mission Five: Championing the international flow of data

Now, as spectroscopists we know we have limited reference data resources available to support our work, whether industrial or academic. We can certainly sign up to the intentions of Mission One around setting the correct conditions to make data usable, accessible and available. It is also in Mission One that we get to hear about the desire to develop this policy using an evidence-based approach—and this is where you can all help—by joining in the consultations

explained below. (There is a very short deadline to the end of the next consultation!)

Data use in the coronavirus pandemic

I must admit to feeling quite sick as I read Mission Three. In this document, whilst clearly identifying that the pandemic has shown up a massive untapped potential in the use of data by government and public services, it indicates that the pandemic has set a high watermark that must be maintained. On a personal note, I lost my mother and many of her friends in the first COVID wave which swept through the care home where she was a living. There were no measures in place to protect the residents, and I hope you will excuse my very sceptical view of the way governments have handled the pandemic. Listening to most people I come into contact with, they are baffled as to why the data on infection rates in early 2020 led to some of the most extreme restrictions on civil liberties and freedoms in our lifetimes, yet data showing more than an order of magnitude higher infection rates in late 2020 and mid-2021 seem to leave our politicians struggling to explain why they are now doing so much less.

I have many friends who work in some part of the health care system. As a result, the numbers I always looked for in the regular data-driven reporting during the pandemic, and the biggest political decisions I could easily identify with and support, were around desperately trying to avoid overwhelming our hospitals and long-term care facilities. That said, as I write this column, the Netherlands have just announced that the rates there are so worrying that they are starting to introduce additional partial lockdown. This is now being closely followed by the Irish and other European governments.

I think one of my inputs into this, or any consultation on the use of data in this area, would certainly be around not using data to justify any specific political decision or pathway and then ignoring exactly the same data analysis when it becomes an inconvenient nuisance to a change in policy. As a scientist under lockdown, I found this apparent lack of

Figure 1. Cover image of the new book written by Jeremy Farrar and Anjana Ahuja. Reproduced with kind permission of Profile Books.

transparency at times extremely frustrating, as I am sure most of you have done as well. This may have led to the sporadic resignations of some of the UK's top experts from governmental advisory bodies. Unfortunately, this now includes Professor Sir Jeremy Farrar, a longstanding Director of the Wellcome Trust[2] with an exceptional global scientific reputation in the field. He has been outspoken about how expert advice has been used, and recently resigned as a member of the Scientific Advisory body for Emergencies (SAGE). He has been critical of the way scientific advice has been handled by government and has published his insights into this area this year.[3]

Communication improvements and data infrastructures

Rant over. The pandemic has shown the importance of some of the core tenants of Mission Four around the security of data infrastructures and the protection of data-driven services. Looking back to March 2020 and the UK entering lockdown, our communication providers have made enormous strides in adapting their services and software solutions to meet our multi-person, simultaneous

communication desires. Watching people's lips move in a conference call, speaking words I had heard five seconds earlier are fortunately a thing of the past. For the most part, our behaviours have also changed for the better (the infamous British local council meetings viewable on YouTube being the notable exceptions!) with a significantly larger degree of self-discipline required in contentious meetings. People trying to aggressively shout down discussion when they think they are losing an argument is mostly a thing of the past. Speaking louder or over the top of an existing speaker usually just results in nobody hearing anything either party has to say. If we could replicate those enormous improvements in the software and underlying IT infrastructure, we have observed in the last 22 months to, say, the ability to securely and robustly get data in a timely manner from a chemical plant running in an Industry 4.0 type environment, that would greatly enhance our ability to deliver on the energy and resource savings promised by such innovations. Guess what, in Mission One we also get to hear about the need for further developing an Information Management Framework establishing a common language for Digital Twins and to quote one author on this subject "Digital twins (DT) are the key enablers for transformation to Industry 4.0 (I4.0)".[4] The communication of data from plant-based sensors such as spectroscopic instruments monitoring reaction progress within industrial plants is, of course, a key component within an I4.0 chemical manufacturing scenario. For more information on similar topics there are a number of excellent articles in this journal as well as in the area of the Theory of Sampling with Figure 2 showing an example taken from an expert in the field.[5]

Call for evidence on metrics

Now, if like me you missed the original call for consultation, which was only open for 13 weeks last year when many people's focus was understandably elsewhere, the questions that were asked are available[6] and the consultation report has been published.[7] This seems to be
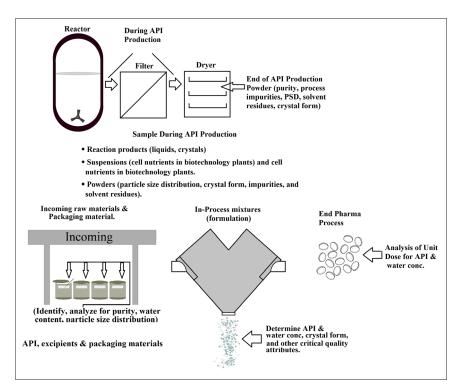
TONY DAVIES COLUMN

Figure 2. Flow path of the generic pharmaceutical manufacturing process with the principal sampling locations indicated taken from Reference 5 with permission.

an excellent initiative and has received some very good inputs from the original strategy paper that are relevant to our industrial and academic base. I think we should support it wherever possible.

Now they are specially looking for input from different groups which include:

- technology and data-driven, or data-rich companies
- investors in technology and data companies
- academics, and research and policy organisations with a particular interest in the role of data in the economy and society
- international data standards, regulation and governance bodies

As they point out for the National Data Strategy to be successful, they need to have a clear grasp of the real-world context for our action, and as they developed the monitoring and evaluation framework, they kept coming up against the "difficulties of measuring the impact of data specifically aimed at evidence on metrics to assess how the missions are delivering".

So now there is an invitation to suggest concrete ways of assessing their success. Again, this has a very short window of opportunity, closing on Friday 3 December 2021. If you believe you have something to contribute especially on these two key topics:

1) Are you aware of any metrics which currently exist which we could use as indicators for the pillars or opportunities? If so, please provide a hyperlink or further information on where they are recorded. We would particularly value input where we have no or low confidence measures.

2) Do you have any thoughts on metrics which are being developed, or could be developed, which we could use as indicators for the pillars or opportunities?

use the DCMS's online survey platform to submit your response to get your views heard![8]

References

1. https://www.gov.uk/government/publications/uk-national-data-strategy
2. https://wellcome.org/news/jeremy-farrar-and-sage
3. J. Farrar and A. Ahuja, *Spike: The Virus vs. The People - the Inside Story*. Profile Books (2021). ISBN: 1788169220, Buying Options
4. Z.M. Cinar, A.A. Nuhu, Q. Zeeshan and O. Korhan, "Digital Twins for Industry 4.0: A Review", in *Industrial Engineering in the Digital Disruption Era*, Ed by F. Calisir and O. Korhan. GJCIE 2019. Lecture Notes in Management and Industrial Engineering. Springer (2020). https://doi.org/10.1007/978-3-030-42416-9_18
5. R.J. Romañach and K.H. Esbensen, "Sampling in pharmaceutical manufacturing—Many opportunities to improve today's practice through the Theory of Sampling (TOS)", *TOS Forum* **4**, 5 (2015). https://doi.org/10.1255/tosf.37
6. https://www.gov.uk/government/consultations/uk-national-data-strategy-nds-consultation/uk-national-data-strategy-consultation
7. https://www.gov.uk/government/consultations/uk-national-data-strategy-nds-consultation/outcome/government-response-to-the-consultation-on-the-national-data-strategy
8. https://dcms.eu.qualtrics.com/jfe/form/SV_1FA8nNe0lx1vjQW

Tony Davies is a long-standing *Spectroscopy Europe* column editor and recognised thought leader on standardisation and regulatory compliance with a foot in both industrial and academic camps. He spent most of his working life in Germany and the Netherlands, most recently as Lead Scientist, Strategic Research Group – Measurement and Analytical Science at AkzoNobel/Nouryon Chemicals BV in the Netherlands. A strong advocate of the correct use of Open Innovation.
ⓘ https://orcid.org/0000-0002-3119-4202
antony.n.davies@gmail.com