Open Knowledge Foundation
86-90 Paul Street
London EC2A 4NE
United Kingdom
okfn.org

Friday 12 July 2019

To whom it may concern,

This submission is in response to the open call for evidence for the UK government's National Data Strategy published on 10 June 2019.

It has been prepared by Stephen Abbott Pugh on behalf of the Open Knowledge Foundation, a UK not-for-profit dedicated to the pursuit of open knowledge.

This response does not need to be kept confidential and we are happy for it to be published or shared. We are happy to be contacted by anyone from the Department for Digital, Culture, Media and Sport about this submission via emails to info@okfn.org.

The Open Knowledge Foundation

The vision of the Open Knowledge Foundation is for a future that is fair, free and open.

Our mission is to create a more open world – a world where all non-personal information is open, free for everyone to use, build on and share; and creators and innovators are fairly recognised and rewarded.

Since our inception in 2004, we have pioneered the understanding and open licencing of information and data by creating the Open Definition; enabled anyone - including the UK government - to publish data openly by developing the CKAN open source data portal platform; and evaluated the availability and quality of government open data worldwide using our Global Open Data Index.

We live in a knowledge society where we face two different futures: one which is open and one which is closed.

An open future means knowledge is shared by all – freely available to everyone, a world where people are able to fulfil their potential and live happy and healthy lives.

A closed future is one where knowledge is exclusively owned and controlled leading to greater inequality and a closed future.

The UK National Data Strategy must emphasise the importance and value of sharing more, better quality information and data openly in order to make the most of the world-class knowledge created by our institutions and citizens. Without this commitment at all levels of society, British businesses, citizens and public bodies will fail to play a full role in the interconnected global knowledge economy of the future.

Questions answered

2.6. How important are basic data skills for employment in today's economy? What is the basic level of data skills needed and what kinds of skills are needed?

Learning data skills can prove hugely beneficial to individuals seeking employment in a wide range of fields including the public sector, government, media and voluntary sector.

At present there is often a huge amount of work required to clean up data in order to make it usable before insights or stories can be gleaned from it. This is why Open Knowledge Foundation launched the School of Data which has trained thousands of people in person across four continents since 2012. Their Data Pipeline methodology - applicable to many fields - teaches people how to define, find, get, verify, clean, analyse and present data with online courses to help people develop the essential skills needed to work with data.

We have learned that "data is a team sport" and that everyone can benefit from developing different levels of data skills. We believe there are fundamental approaches which help people make sense of data, no matter their field of expertise. Our deeper engagement with those working in government, civil society and the media lead to us distilling lessons about those fields of work into the Open Data Handbook, a leading free resource now translated into 20 languages, and the Data Journalism Handbook.

We believe that the UK government could help empower the wider workforce by instigating or backing a fundamental data literacy training programme open to local communities working in a range of fields to strengthen data demand, use and understanding. Such a programme could form part of a government response to the Augar review's call for improved technical education and increased investment in adult learning. It could be jointly coordinated by the Department for Education and Department for Digital, Culture, Media and Sport.

Without such training and knowledge, large numbers of UK workers will be ill-equipped to take on many jobs of the future where products and services are devised, built and launched to address issues highlighted by data. Empowering people to make better decisions and choices informed by data will boost productivity but not without the necessary investment in skills.

4.8. Should government encourage businesses and non-profit organisations to make more of the data they hold open? If so, how?

The most important things government can do to help businesses and non-profit organisations best share the data they hold is to promote open licencing and interoperability.

Open licences are legal arrangements that grant the general public rights to reuse, distribute, combine or modify works that would otherwise be restricted under intellectual property laws. Licences are open, if they enable anyone to use works for any purpose, both commercially and non-commercially. Without an open licence, users face legal grey areas.

Our research has shown that the "proliferation of [non-standard] licences continues to be a major challenge for open data". Our most recent Global Open Data Index, an international survey of the state of open government data carried out by civil society and thematic experts, found that only 40% of Great Britain's datasets were fully open, as defined by the Open Definition.

If the UK government continues with a policy of regulatory alignment with the European Union in key areas, it should back the call to use "open standard licences" for all key datasets as laid out in the revised 2019 directive on open data and the re-use of public sector information.

The European Data Portal indicates that there could be up to 90 different licences currently used by national, regional, or municipal governments. Their quality assurance report also shows that they can't automatically detect the licences used to publish the vast majority of datasets published by open data portals from EU countries. If they can't work this out, the public definitely won't be able to: meaning that any and all efforts to use newly-released data will be restrained by unnecessarily onerous reuse conditions.

We would urge the UK government to explicitly recommend the adoption of Open Definition compliant licences from Creative Commons or Open Data Commons in its National Data Strategy.

On interoperability, organisations working alongside us in the "open" sector have developed a wide range of tools and standards for publishing open data aiming at general data publication or specific use cases. Stronger government endorsement and promotion of these solutions would help position the UK as a leader on open data while ensuring that data released could be more easily used or combined with other datasets from across the world.

Robust existing standards including the EITI standard for extractive industry transparency; IATI standard for international aid and humanitarian efforts; 360Giving standard for grants data; OpenOwnership standard for beneficial ownership data; Open Contracting data standard for public procurement; Fiscal Data Package for government spending or budget data; as well as our own Frictionless Data specifications for publishing data packages or tabular data which have recently been endorsed by the French government's open data team.

Several of these standards are endorsed by commitments in the latest (delayed) UK national action plan submitted to the Open Government Partnership but more work needs to be done to raise awareness, use and publication of data in line with these standards especially given the "lack of visible high-level political support for the open government process in the UK".

The CKAN open-source data portal platform, first developed by the Open Knowledge Foundation, gives any organisation the ability to publish and licence their data openly with tools and a powerful API allowing data to be combined with that from other CKAN instances including those from the UK government, USA, EU and UN amongst others. To date, more than two million datasets have been published using the platform.

Open-source platforms like CKAN should be recommended or included in government guidance to non-profit organisations looking for the best way to publish and share their data.

5.1. How effectively are government and the wider public sector collecting, sharing, analysing and storing the data it holds? What does good practice look like? What does bad practice look like?

The UK government may have adopted the principles from our colleagues at the Open Data Charter but evidence to date demonstrates a backsliding on the UK's commitment to being "open by default" since we first started our Global Open Data Index in 2013 (see the Open Data Charter Measurement Guide - which we contributed to - for more on how to measure progress towards the Charter principles).

Despite the good example set by government with its use of the CKAN platform to publish datasets on data.gov.uk since 2009, only 40% of Great Britain's key datasets were found to be fully published open in 2017, as surveyed by our Global Open Data Index and defined by the Open Definition.

Our report - The State of Open Government Data - set out a series of good practice recommendations for governments and civil society for how best to foster the production and use of meaningful data. This was based on insights gleaned from convening a dialogue about government data across more than 90 countries.

Recommendations included: giving data files comprehensible names; ensuring that machine-readable formats are used; understanding the needs of data users; publishing accurate, precise, processable raw data; adding clear metadata; and the use of standardised open licencing.

Further research commissioned by our team also found that governments should support open data champions in government agencies and strengthen their capacities as well as providing government agencies with diverse incentives to publish open data and engaging with external data users.

These recommendations become more and more relevant as the number of data publishers and users grows exponentially due to the compounding of issues they face where more and more data could lead to more and more confusion or siloing of data rather than contributing to a growing living library of complementary and interoperable datasets.

Beyond government data, we also recently produced a report on Citizen-Generated Data and the Sustainable Development Goals in collaboration with King's College London, Public Data Lab and the Global Partnership for Sustainable Development Data. This found that good citizen-generated data efforts can "open up new types of relationships between individuals, civil society and public institutions" to fill in data gaps or provide useful information to test or complement government data.

In the UK, we would like to see the public sector pioneering new ways of producing and harnessing citizen-generated data efforts by organising citizen science projects through schools, libraries, churches and community groups. Empowering these community through teaching data skills (as recommended above) could reap huge benefits in terms of regenerating community spaces and multiplying the value of data training by building on existing networks. These local communities could help the government to collect high-quality data relating to issues such as air quality or recycling while also leading the charge when it comes to increasing the use of central government data. Our Frictionless Data project could form part of the training for

these local groups as our open-source tools empower anyone to create and publish high-quality open data with a current focus on reproducible research.