

## Response to the Public Consultation on the European Commission's

### European Strategy for Data

Submitted by Prof. Derek McAuley, Dr. Ansgar Koene and Dr. Jiahong Chen of  
Horizon Digital Economy Research Institute, University of Nottingham

29 May 2020

1. Horizon<sup>1</sup> is a Research Institute centred at The University of Nottingham and a Research Hub within the UKRI Digital Economy programme<sup>2</sup>. Horizon brings together researchers from a broad range of disciplines to investigate the opportunities and challenges arising from the increased use of digital technology in our everyday lives. Prof. McAuley is Director of Horizon and Principal Investigator of the EPSRC-funded DADA<sup>3</sup> (Defence Against Dark Artefacts) project, addressing smart home IoT network security, and its acceptability and usability issues, the ESRC-funded CaSMa<sup>4</sup> (Citizen-centric approaches to Social Media analysis) project to promote ways for individuals to control their data and online privacy, and the EPSRC-funded UnBias<sup>5</sup> (Emancipating Users Against Algorithmic Biases for a Trusted Digital Economy) project for raising user awareness and agency when using algorithmic services. Dr Koene was a lead researcher of the CaSMa and UnBias projects, is Research co-Investigator on the EPSRC-funded ReEnTrust<sup>6</sup> (Rebuilding and Enhancing Trust in Algorithms) project and chairs the working group for developing the IEEE P7003 Standard for Algorithm Bias Considerations. Dr Jiahong Chen is a Researcher Fellow of Horizon, working on the DADA project.

#### ***Introduction***

2. We welcome the Commission's initiative in putting forward a comprehensive, proactive and prudential strategy for realising the potentials of a data-driven economy and maximising its benefits for individual well-being, economic sustainability and social progress. In addition to our responses to the specific questions in the survey, we would like to highlight a number of aspects covered by the survey questions where the further specification and implementation of the Strategy could improve. We would be happy to be contacted for further discussions, and for our comments to be published.

#### ***Data interoperability and portability***

3. While we recognise that expanding data interoperability and portability among market players is crucial for promoting competition and innovation, it is equally important that when it comes to personal data, this should be subject to data protection principles, especially data minimisation. We have reservations for the statement that "it should be made easier for individuals to give access to

---

<sup>1</sup> <http://www.horizon.ac.uk>

<sup>2</sup> <https://epsrc.ukri.org/research/ourportfolio/themes/digitaleconomy/>

<sup>3</sup> <https://www.horizon.ac.uk/project/defence-against-dark-artefacts/>

<sup>4</sup> <http://casma.wp.horizon.ac.uk>

<sup>5</sup> <http://unbias.wp.horizon.ac.uk>

<sup>6</sup> <https://ReEnTrust.org>

existing data held about them, e.g. by online platform providers, car manufacturers, producers of wearables, voice assistants or smart home appliances, to new services providers of their choosing, in line with the GDPR". It is often unclear what amounts to "necessary" data uses under the GDPR in these areas, and there has been substantial evidence suggesting over-collection and over-sharing of personal data especially in sectors where services are not mutually exclusive (e.g. "multi-homing" on social media). Consumer welfare can be improved with greater data interoperability and portability only on condition that data protection rules are further specified and better observed.

#### ***Non-personal data generated by smart devices***

4. Preliminary findings in our ongoing research on the use of IoT-generated data in smart homes and smart buildings support the statement that "It is currently challenging to define solutions on the allocation of the rights to use data coming from smart machines or devices that are fair for all parties concerned". However, we have also identified other legal uncertainties inhibiting the optimal functioning of smart technologies in shared buildings, such as the ownership of devices/sensors, the collective operation and maintenance of the systems, and the distribution of economic surplus of co-produced data. These issues should also be addressed by the Strategy.
5. Also, certain types of valuable non-personal data (e.g. building energy efficiency data) are currently under-supplied due to market failure in this area. In such cases, apart from granting full control to entities producing the data, regulators should also consider non-market-based approaches, such as mandatory sharing of data combined with regulatory incentives or compensations. We therefore do not fully agree with the statement that "The EU should make major investments in technologies and infrastructures that enhance data access and use, while giving individuals as well as public and private organisations full control over the data they generate."

#### ***Secondary use of sensitive data and data altruism***

6. Our ongoing research on the applicability of the GDPR to open genome initiatives (platforms for volunteers to upload and share their genetic data for research) has shown that both the platform operators and the data users are facing grave uncertainties in achieving full compliance with the data protection framework. One major challenge is choosing the appropriate lawfulness basis (e.g. explicit consent, manifestly public self-disclosure, research exemption, public health, etc.) and some of these legal grounds depend on a further legal basis as well as additional safeguards provided by EU or Member State laws. The lack of specialised legislation at EU level means that research projects are frustrated by complicated legal issues resulting from the fragmentation in national approaches. As part of the Commission's pledge to "Develop sector-specific legislative or non-legislative measures for the European health data space", harmonised rules should be introduced to provide a clear legal basis for international (intra- and extra-EU) research with health-related data.
7. As regards individual control over their sensitive data, the Strategy has rightly pointed out that "Citizens have the right in particular to access and control their personal health data". To further strengthen trust in reuse of health data, a dashboard-like system should be deployed as part of the Common European Health Data Space to allow individuals to trace for what purposes their data have been used and, where appropriate, to grant or revoke permission for specific uses.

### ***Cloud and edge computing***

8. Existing EU-level regulation on providers of cloud computing, notably the RIS Directive, is a clear indication of the necessity for regulatory interference in ensuring security for these data infrastructures. This is especially critical considering the wider impact of private service providers may have on the overall level of national and international cybersecurity, which is a matter of substantial public interest. While self-regulatory measures like codes of conduct certainly form an important part of a robust regulatory strategy, an irreplaceable role remains with EU and national regulators, for example, to lay down critical principles and baseline requirements, to coordinate preventive endeavours and contingency plans, and in the case of violations, to enforce applicable rules. Edge computing services, depending on their scale and impact, should be subject to a similar set of measures.

## Annex: Responses to the main body of the consultation

### Section 1: General questions on the data strategy

Over the last few years, digital technologies have transformed our economy and society, affecting all sectors of activity and the daily lives of all Europeans. Data is at the centre of this transformation, and more is to come as the volume of data produced in the world is growing rapidly.

Do you agree that the European Union needs an overarching data strategy to enable the digital transformation of the society?

Yes

No

*“More data should be available for the common good, for example for improving mobility, delivering personalised medicine, reducing energy consumption and making our society greener.”* To what extent do you agree with this statement?

Strongly agree

Somewhat agree

Neutral

Somewhat disagree

Strongly disagree

I don't know / no opinion

Do you think that it should be made easier for individuals to give access to existing data held about them, e.g. by online platform providers, car manufacturers, producers of wearables, voice assistants or smart home appliances, to new services providers of their choosing, in line with the GDPR?

Yes

No

Have you faced difficulties in recruiting data professionals (workers who collect, store, manage, analyse, interpret and visualise data as their primary or as a relevant part of their activity) during the last 2 years?

Yes

No

*‘General data literacy across the EU population is currently insufficient for everyone to benefit from data-driven innovation and to become more active agents in the data economy.’* To what extent do you agree with this statement?

Strongly agree

Somewhat agree

Neutral

Somewhat disagree

Strongly disagree

I don't know / no opinion

One area of study are difficulties experienced in accessing and use data from other companies. With the following questions we seek to further examine the importance and the nature of data access issues in business-to-business situations.

Have you had difficulties in using data from other companies?

Yes

No

A specific class of data are non-personal data generated by smart machines or devices connected to the Internet-of-Things in professional use, such as smart tractors used in farms or smart robots in use in a factory. Data generated by such machines or devices are typically 'co-generated' by several parties, such as: manufacturer of a smart device, lawful user of the smart machine or device etc.

*'It is currently challenging to define solutions on the allocation of the rights to use data coming from smart machines or devices that are fair for all parties concerned'.* To what extent do you agree with this statement?

Strongly agree

Somewhat agree

Neutral

Somewhat disagree

Strongly disagree

I don't know / no opinion

*'The EU should make major investments in technologies and infrastructures that enhance data access and use, while giving individuals as well as public and private organisations full control over the data they generate.'* To what extent do you agree with this statement?

Strongly agree

Somewhat agree

Neutral

Somewhat disagree

Strongly disagree

I don't know / no opinion

*'The development of common European data spaces should be supported by the EU in strategic industry sectors and domains of public interest (industry/manufacturing, Green Deal, mobility, health, finance, energy, agriculture, public administration, skills).'* To what extent do you agree with this statement?

Strongly agree

Somewhat agree

Neutral

Somewhat disagree

Strongly disagree

I don't know / no opinion

Are there general comments you would like to make about the data strategy?

300 character(s) maximum

Interoperability and portability of personal data must be subject to data minimisation principle (see para 3 attachment).

Promotion of IoT data uses should focus not just on control of data and market-based mechanisms (see paras 4-5 attachment).

## **Section 2.1 - Specific questions on future actions: Data governance**

The use of data in the society and the economy raises a series of questions of legal, ethical, organisational and technical nature. Many angles need to be looked at in order to fully reap the benefits of the use of data without harm.

With the term 'data governance' we seek to refer to the set of legal, organisational and technical rules, tools and processes that determine the use of data by the public sector, business, individuals, civil society organisations, researchers.

This may translate into establishing mechanisms for data governance at European level which may support data-driven innovation in different ways:

- At cross-sector level, it could identify the need for standards to facilitate data-sharing, including for the various actions to be taken in this regard (identification, authentication, access control). It could identify use cases in which cross-sector data re-use is supported by standardisation. It could provide technical guidance on technologies for lawful processing of data in accordance with data protection legislation, the need to protect of commercially sensitive information as well as competition law.
- At sector-specific level, data governance could be developed, building on existing structures and coordination mechanisms.

*'Data governance mechanisms are needed to capture the enormous potential of data in particular for cross-sector data use.'* To what extent do you agree with this statement?

- Strongly agree
- Somewhat agree
- Neutral
- Somewhat disagree
- Strongly disagree
- I don't know / no opinion

### **Standardisation**

Significant interoperability issues impede the combination of data from different sources within sectors, and even more so between sectors. One such issue is the absence of a consistent description of the data, including information on how it has been gathered. This can impact on data discoverability and on the capacity to evaluate data quality. Another issue is the differences between data models used for similar or identical information assets. This constitutes a barrier for re-users, both commercial and from academia. Standardisation is one of the means to respond to these challenges.

*'The re-use of data in the economy and society would benefit greatly from standardisation to improve interoperability.'* To what extent do you agree with this statement?

- Strongly agree
- Somewhat agree
- Neutral
- Somewhat disagree

- Strongly disagree
- I don't know / no opinion

*'Future standardisation activities need to better address the use of data across sectors of the economy or domains of society.'* To what extent do you agree with this statement?

- Strongly agree
- Somewhat agree
- Neutral
- Somewhat disagree
- Strongly disagree
- I don't know / no opinion

Which of the following elements do you consider to be the most useful in terms of standardisation?

- Metadata schema
- Metadata variables (semantic interoperability at the metadata level), including ontologies, controlled vocabularies and agreed thesauri
- Data formats
- Common data models
- Data exchange protocols
- Application Programming Interfaces (APIs)
- Licences or licence terms
- Other
- I don't know / no opinion

What role should EU or national government bodies take in standardisation?

- Provide necessary funding in order to ensure open standards
- Take an active role in the prioritisation and coordination of standardisation needs, creation and updates
- Be directly involved in defining standards
- Provide funding to test draft standards in practice and develop tools to implement them early-on
- Other
- Governments should not have a role in standardisation
- I don't know / no opinion

### **Secondary use of data**

Sensitive data (e.g. health data) stored in public databases has a high potential for re-use for the public interest. However, it is normally not possible to open such data as 'open data' and specific efforts are necessary on the side of the public sector data holder. Some statistical offices have put structures in place for processing of statistical micro-data for research. In some countries, similar approaches are taken for certain social services or health information. With these questions we seek to understand whether such approaches should be broadened to other types of data and what types of services should be offered.

'Public authorities should do more to make available a broader range of sensitive data for R&I purposes for the public interest, in full respect of data protection rights.' To what extent do you agree with this statement?

- Strongly agree
- Somewhat agree
- Neutral
- Somewhat disagree
- Strongly disagree
- I don't know / no opinion

Which of the following should public authorities do to facilitate data re-use:

- Help the re-user to identify the exact authority that is holder of a specific set of data (one-stop-shop)
- Ensure that the request for data access is processed faster, within agreed deadlines
- Assess whether the re-use of the data could potentially harm the interests of others (of the persons/companies whose data is being used) for concrete use-cases
- Be able to provide anonymisation of specific data for concrete use-cases
- Offer the possibility to process data within a secure environment it makes available, so that the user does not need to obtain a copy of the data
- Clarify from the outset the legal rules on the purposes for which the data can be used
- Provide for recourse mechanisms to challenge decisions on one or several of the above.
- Other
- I don't know / no opinion

If other, please specify

200 character(s) maximum

Develop and operate data dashboards for individuals to access and control their sensitive data (see paras 6-7 attachment).

### **Data altruism**

Data altruism is about making it easier for individuals to allow the use of the data they generate for the public good, if they wish to do so, in full compliance with the GDPR and namely on the basis of consent as a legal basis. This is sometimes referred to as 'data donation', a term that could be misunderstood to mean that the consent to the processing of such data in question cannot be withdrawn. Article 7 of the GDPR provides that consent can always be withdrawn and there is no intention to change this rule for 'data altruism'.

Do you think that law and technology should enable citizens to make available their data for the public interest, without any direct reward?

- Yes
- No
- I don't know / no opinion

For which of the following purposes would you be willing to make data available:

- For health-related research



For aspects relating to the city/municipality/region I live in, including for example improving mobility, to improve environmental challenges that can be addressed through action at local or regional level

For other public interest purposes

None of the above

I don't know / no opinion

Do you think there are sufficient tools and mechanisms to "donate" your data?

Yes

No

I don't know / no opinion

In which of the following domains do you see potential for the use of 'contributed' data:

For health-related research

For aspects relating to the city/municipality/region I live in, including for example improving mobility, to improve environmental challenges that can be addressed through action at local or regional level

For other public interest purposes

None of the above

I don't know / no opinion

What would support the usefulness of 'data altruism' mechanisms as a means to build up data pools for research and innovation:

A standard form for obtaining consent (and, where necessary, requesting data portability) from the individual in line with the GDPR

A European approach to obtaining consent that is compliant with the GDPR

Public registers of persons that are willing to make available some of their data for research or innovation purposes

The existence of intermediary infrastructures such as personal data spaces/wallets/stores controlled by each individual from which the data could be retrieved

Additional EU legislation on data altruism relating to deceased persons

Information campaigns sensitising individuals on the subject matter, e.g. via clinical practitioners

Measures to mitigate inherent bias in the data collected through this means

Other

### **Data intermediaries**

In the data economy, novel intermediaries such as 'data marketplaces' or 'data brokers' are increasingly aiming to support business-to-business data sharing with a range of services such as match-making, offering dedicated sharing platforms, including technology on controlled access and use, as well as legal support.

*'Such intermediaries are useful enablers of the data economy.'* To what extent do you agree with this statement?

Strongly agree

Somewhat agree

Neutral

Somewhat disagree

- Strongly disagree
- I don't know / no opinion

## Section 2.2 - Specific questions on future actions: identification of high-value datasets

The recently adopted [Directive 2019/1024/EU](#) (Open Data Directive) introduces the concept of high-value datasets (HVDs), defined as documents the re-use of which is associated with important benefits for society and the economy (e.g. job creation, new digital services, more efficient and evidence-based policy making). Under the directive, the Commission is required to adopt an implementing act setting out a list of specific high value datasets within the thematic categories listed in Annex I to the directive (geospatial; earth observation and environment; meteorological; statistics; companies and company ownership; mobility). The directive specifies that those datasets shall be made available for re-use free of charge, in machine-readable formats, provided via application programming interfaces (APIs) and, where relevant, as bulk download.

The answers to the questions below will help the Commission draw up an EU-wide list of specific high-value datasets.

*'The establishment of a list of high-value datasets, to be made available free of charge, without restrictions and via APIs, is a good way to ensure that public sector data has a positive impact on the EU's economy and society.'* To what extent do you agree with this statement?

- Strongly agree
- Somewhat agree
- Neutral
- Somewhat disagree
- Strongly disagree
- I don't know / no opinion

Apart from the potential to generate socio-economic benefits, please indicate the relevance of the following additional factors to be taken into account when selecting datasets for the future list of high value datasets:

	Very relevant	Relevant	Neutral	Not relevant	Not relevant at all	I don't know / no opinion
The re-use of the dataset would increase if it was provided free of charge.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
The dataset belongs to a thematic area in which there are few EU-level	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

requirements for opening up data.						
The re-use of the dataset would increase if its availability under uniform conditions was ensured across the entire EU.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The re-use of the dataset would increase if it was available via an application programming interface (API).	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If other factors: please specify

200 character(s) maximum

Under the Open Data Directive, specific high-value datasets will have to be available free of charge, in a machine-readable format, provided via APIs and, where relevant, provided as a bulk download. Please indicate the relevance of each of the other arrangements indicated below to improve the re-usability of specific high-value datasets.

	Very relevant	Relevant	Neutral	Not relevant	Not relevant at all	I don't know / no opinion
Licensing and other terms applicable to re-use	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Standardised formats of data and metadata	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Possibility of user feedback	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Specific technical arrangements for dissemination	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If other arrangements, please specify:

200 character(s) maximum

Catalogues of projects using the dataset and facilities for data users to publish or disseminate processed or linked datasets.

Please specify which specific technical arrangements for dissemination:

200 character(s) maximum

EU programmes may provide funding to enhance the availability and re-use of high-value datasets across Europe. For each of the following activities, please indicate how relevant it is to support them.

	Very relevant	Relevant	Neutral	Not relevant	Not relevant at all	I don't know / no opinion
Improving the quality (e.g. machine-readability) and interoperability of the data/metadata	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ensuring sustainable data provision via application programming interfaces (APIs)	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Engaging with re-users (promoting the data, co-defining use-cases)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

If other activities, please specify:

200 character(s) maximum

According to your experience and the expected potential of concrete datasets, indicate up to three specific datasets that should be listed in each of the thematic categories of high-value datasets, as referred to in Article 13(1) of the Open Data Directive:

	Specific datasets
Geospatial	
Earth observation and environment	
Meteorological	
Statistics	
Companies and company ownership	
Mobility	

### Section 2.3 - Specific questions on future actions: the (self-/co-)regulatory context of cloud computing

Data processing that underpins data sharing, in particular in common European data spaces will increasingly be performed in cloud and edge infrastructures. Such infrastructures can present cost-effective alternatives and, in the case of edge computing, enable real-time data processing in a connected device. Cloud and edge services are covered by European rules like the General Data Protection Regulation and the Free Flow of Non-personal Data Regulation, and implementation of these rules can happen by means of self-regulatory schemes like codes of conduct. To ensure that cloud and edge services offered in Europe are secure, user friendly and compliant with the rules, the applicable rules should be implemented by service providers in an appropriate and transparent way.

Does your organisation use and/or provide cloud or edge services?

- Yes, my organisation **uses** cloud or edge services
- Yes, my organisation **provides** cloud or edge services
- None of the two

Do you believe the cloud market currently offers the technological solutions that you need to grow and innovate your business?

- Yes
- No

Do you feel that your organisation's sensitive data is adequately protected and secured by the cloud services you use?

- Yes
- No

Have you experienced problems in the context of the current functioning and constitution of the market for cloud services in Europe?

- Yes

No

Do you perceive risks emerging from the current functioning and constitution of the market for cloud services in Europe?

Yes

No

Does your organisation have flexibility to procure/adopt new and innovative cloud solutions if they emerge on the market?

Yes

No

Is your organisation aware of self-regulatory schemes for cloud/edge services (for example, codes of conduct or certification schemes)?

Yes

No

How do you believe market awareness of these schemes could be raised?

*300 character(s) maximum*

Do you believe a self-regulatory approach is appropriate to identify best practices to apply EU legislation or self-regulation?

Yes

No

Please explain why:

*200 character(s) maximum*

Cloud and edge services are strategic infrastructures that will impact significant public interest in relation to cybersecurity (see para 8 attachment).

Would it be beneficial for your organisation if applicable rules for cloud and edge would be bundled and corresponding information made available by the European Commission?

Yes

No

I don't know / no opinion