BRIEFING

Requested by the IMCO committee

European Parliament

Optimal Scope for Free-Flow of Non-Personal Data in Europe

KEY FINDINGS

A debate is taking place on whether current EU restrictions on confidentiality of personal private information should be relaxed so as to include personal information in free and open data flows. However, it is unlikely that a loosening of such rules will be positive for the growth of open data. Public distrust of open data flows may be exacerbated because of fears of potential commercial misuse of such data, as well of leakages, cyberattacks, and so on

Data is not static in a personal/non-personal classification – with modern analytic methods, certain non-personal data can help to generate personal data – so the distinction may become blurred. Thus, de-anonymisation techniques with advances in artificial intelligence (AI) and manipulation of large datasets will become a major issue. Sed posuere consectetur est at lobortis. Nullam id dolor id nibh ultricies vehicula ut id elit. Praesent commodo cursus magna, vel scelerisque nisl consectetur et. Donec sed odio dui.

There are several key recommendations:

1. Promote the use of open data licences to build trust and openness
2. Promote sharing of private enterprises' data within vertical sectors and across sectors to increase the volume of open data through incentive programmes
3. Support testing for contamination of open data mixed with personal data to ensure open data is scrubbed clean - and so reinforce public confidence
4. Ensure anti-competitive behaviour does not compromise the open data initiative.

Introduction

This briefing is based on a presentation by Simon Forge (SCF Associates Ltd) and discussion at a workshop for the IMCO Committee on "Free Flow of Data – A Cornerstone of the Digital Single Market" on 20 February 2018 (rapporteur Anna Maria Corazza Bildt, MEP). The workshop was organised in order to critically assess, among other points:

(1) The legal distinction between personal and non-personal data,

(2) The definition of non-personal data proposed by the European Commission,

Policy Department for Economic, Scientific and Quality of Life Policies
Author: Simon Forge, SCF Associates Ltd
Directorate-General for Internal Policies
PE 618.988- April 2018

EN

(3) The capacity to sufficiently protect the free flow of non-personal data and the data economy in Europe,

(4) Treatment of mixed data (personal and non-personal) in order to strike the right balance between economic and non-economic considerations,

(5) The economic importance of non-personal data, with the current situation and trends in the foreseeable future,

(6) Proposals for the optimal way forward for regulation of the scope of free flow of non-personal data, to strike the right balance between the ambition of ascertaining EU wide free flow of non-personal data and legitimate security concerns,

Consequently, recommendations are presented indicating the optimal policy direction and proposed modifications to the texts that could lead towards such a policy, with measures directly concerning the current scope and beyond.

1. The legal Distinction between Personal and Non-Personal Data

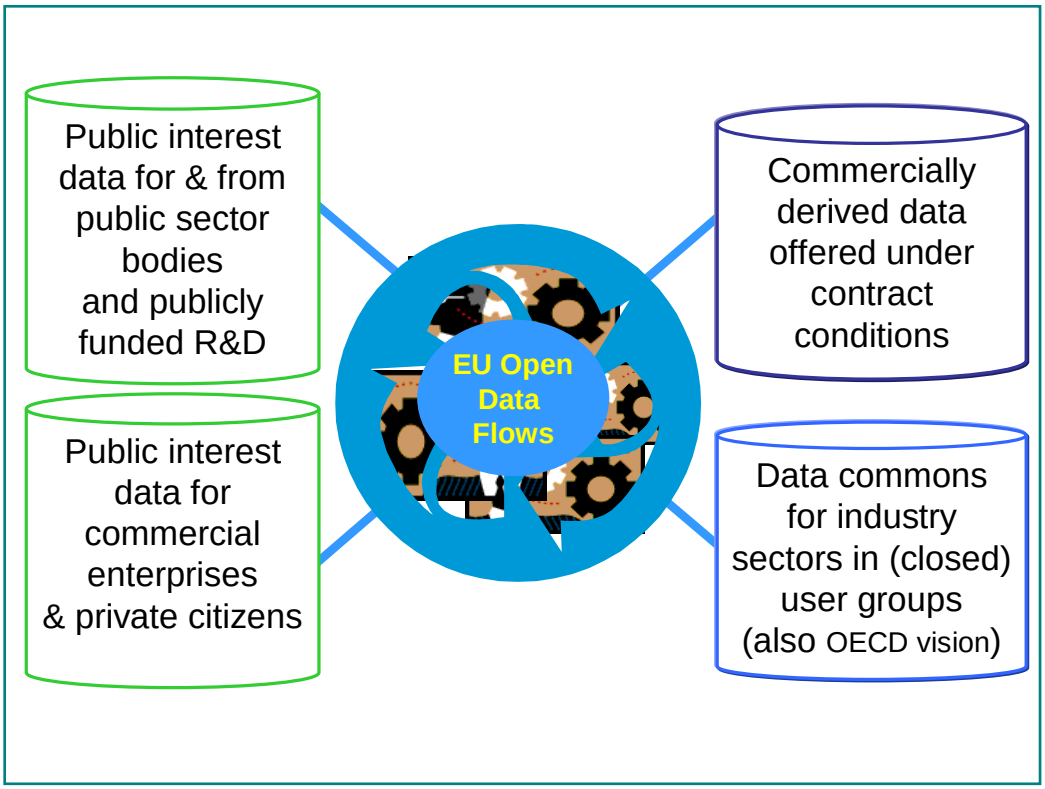
At the outset, it is necessary to clarify what falls within the definition of data permitted to flow freely across the EU. The scope of the definition encompasses various ideals. Flows should consist purely of non-personal data, whether it originates from the commercial or public sectors. Such data should naturally be useful in some way for productive and innovative output in manufacturing and commercial services and/or for government sectors.

In addition, ideally such datasets should be equally accessible to all potential users, without preferential treatment, e.g. the reservation of parts of datasets for a particular user or user group is not permitted. Access should be at low cost or free (e.g. public domain data). Its content must be machine-readable, in a common format, with the metadata. The restraints arising from intellectual property rights (IPR) and commercial confidentiality should be negligible, or have no limits at all, or at least be negotiable.

The General Data Protection Regulation (GDPR) provides the legal distinction between personal and non-personal data with its definition of personal data (Art.4), and the implication that non-personal data is simply the rest. This apparently seems simple and straightforward but begs several questions. For example, can (and do) these definitions correctly balance the economic benefits of free flows of open data with rapidly increasing security and privacy concerns?

Relaxing the Article 4 separation rules between personal and non-personal data to boost a European data economy could create greater problems, bringing resistance to open data concepts in general. Note that EU open data flows will mainly consist of four major sources, as shown in Figure 1.

Figure 1 Main sources of open data flows



Source: Author

Issues Concerning Ownership of Data and Rights of Owners

The ideal of distinctly separate worlds of personal and non-personal data does not reflect current reality. Since open data flows have varied sources and degrees of openness, future rules for end users' access and exploitation need to be flexible enough for commerce to flourish based on releasing data and exploiting data flows.

While the GPDR defines protection of individuals' privacy with more precise definition, issues related to privacy and data protection have not yet been considered in depth by the Commission in relation to open data. Moreover, it should be noted that there is no EU legislation that specifically regulates the question of ownership of data. Instead there is legislation that has an impact on data or that may confer some kind of protection to certain types of data, e.g. copyright, database rights and trade secrets. Many Member States have their own legislation that impacts data ownership and data transfer and may confer some kind of protection to certain types of data and datasets, often specifically for copyright, database rights and trade secrets.

Consequently, the ownership of personal data by an individual is unclear. In this situation, possible rights, including ownership, might even be attributed to those who have gathered, controlled and processed personal data; that is a premise which is sometimes put forward by such entities claiming rights over the product of their processing efforts as data with added value. In contrast, such ownership may be clearly subject to the individual's control over their own personal data in certain Member States (France and Germany especially). This right of ownership varies by Member State.

Although the European Commission's Communication (2017a), Building A European Data Economy, offered an approach to ownership of data, be it non-personal or personal data, this is absent in the proposal for a Regulation of September 2017 (European Commission, 2017b).

2. Economic Significance of Non-Personal Data Flows

As few data assets are valued in company balance sheets, open data's worth to the economy is notoriously difficult to assess. Some indications are given below.

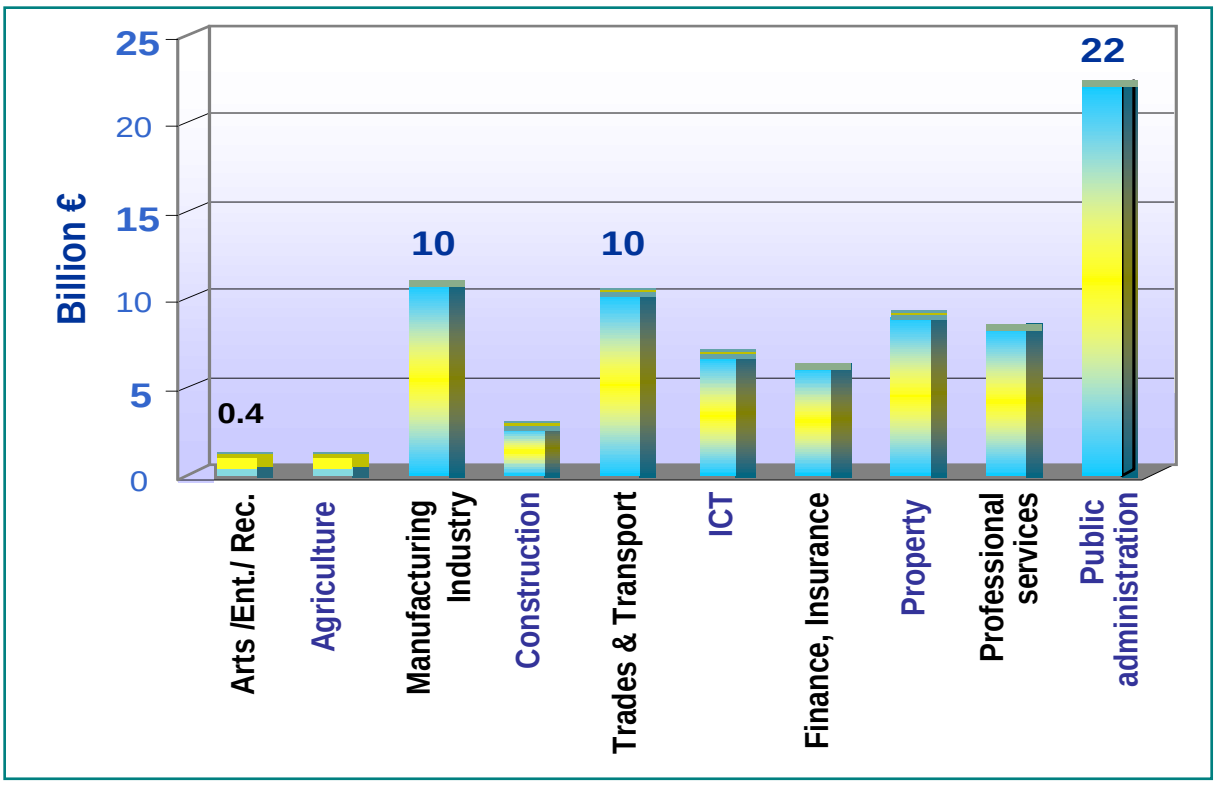
- **Macroeconomic**: G20 countries: 1.1% GDP benefit, 2014 (Lateral Economics, 2014)
- **Microeconomic**: for a global subsector: public data for digital transport: $720-920 billion (Arup)
- **Case study:** US Landsat datasets: $2.2 billion in 2011 (Landsat, 2012)
- **Case Study:** Transport for London (TfL): €148 million/year, 2017 (ODI and Deloitte, 2017, for TfL)
- **EU benefits of open data:** €55 billion in 2016 (EDP, 2015)

For the future, growth in the value of EU open data by 2020 is put at 37% compared to 2016. That implies cumulative direct benefits of EU open data over 2016-2020 of €325 billion (EDP, 2015).

Source: The Open Data Institute (ODI) and Lateral Economics, 2015: https://medium.com/@ODIHQ/the-economic-impact-of-open-data-what-do-we-already-know-1a-119c1956a0 b; European Data Portal (EDP) report: Creating Value through Open Data, 2015; ODI and Deloitte, 2017, Report for Transport for London.

In addition, it should be noted that by sector, the net worth of free flows are highly variable. The diagram below gives an estimate of the value of free flows of data to 10 vertical sectors in 2020.

Figure 2 Future estimates of EU direct market size for open data in 2020, by market sector for the EU28 (€ billions)



Source: European Data Portal (2015): Creating Value through Open Data, Study on the Impact of Re-use of Public Data Resources, 2015

What exactly is the volume of free and open data flows today, as a proportion of all EU data? Very little today can really be considered as open data. Most data is confidential and not released at all, used just within its source organisation. Only about 2% of data is currently open, as shown in Table 1.

Table 1 Proportions of data that are open and closed in the EU

Data classed by openness	% of total EU data
Open data (1%) and open community shared data (1%)	2%
Commercially shared data, from joint ventures, data purchase, company acquisition and data resale	20%
Closed data	78%

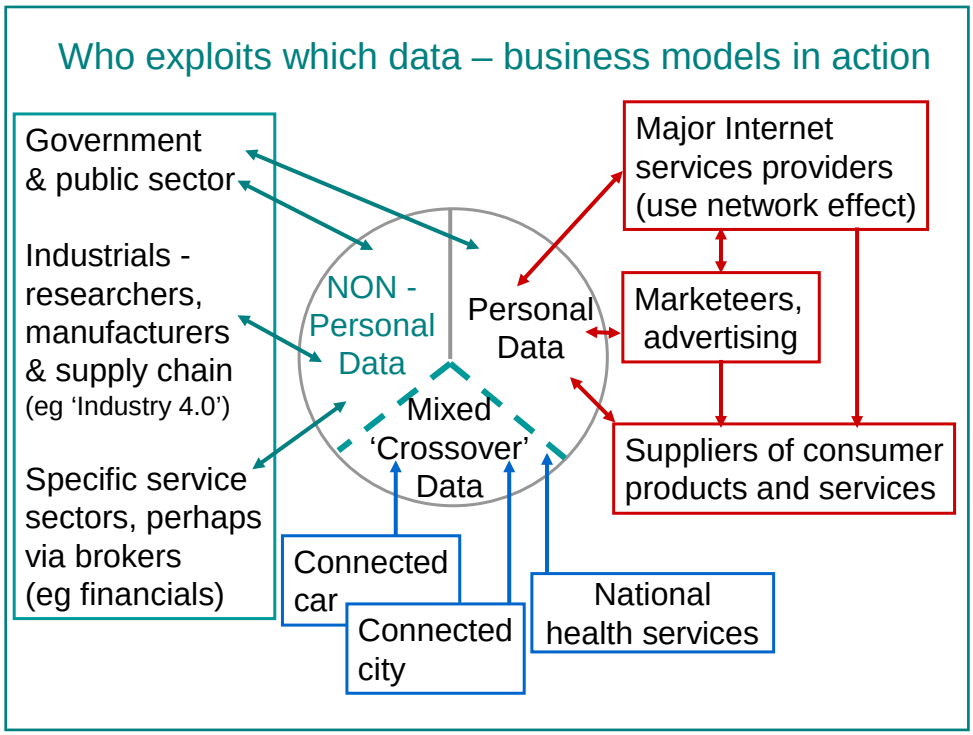
Sources: European Commission staff working document on free flow of data, COM (2017) 9; and Deloitte, Impact assessment support study on emerging issues of data ownership, interoperability, (re)usability and access to data and liability, First interim report.

The vast proportion is privately held, often considered confidential, with implications of being essential to an organisation's core activities and success.

3. Treatment of Mixed Data (Personal and Non-Personal)

With high power computing, the division between personal and non-personal data is becoming blurred, creating a third, or crossover, grey area. This is shown in Figure 3, with the major data generators and potential users of personal, non-personal and crossover data. In the latter category is data from smart cities and future connected cars, whose data can be used to track users as well as to report on technical performance.

Figure 3 Treatment of mixed data (personal and non-personal)



Source: Author

Recent advances in processing massive amounts of data using novel analytical techniques will tend to make this grey area expand, as more personal data is mined from non-personal data records. That mining of non-personal data can be used to build identifiable personal profiles.

The trend is towards identifying personal data from apparently anonymous non-personal data with algorithms using AI, via low-cost supercomputing to drive the leakage. Often these profiling algorithms incorporate other external datasets. So effectively "mixed data" has no clear limits. The capability for de-anonymisation blurs the boundary.

This raises several issues. For example, if a company extracts personal data from an open data source using its own algorithms, the company may gain competitive advantages. The body charged with data protection under the Commission's policy would not be aware of this. This implies that there cannot be rigid rules that work. How should the obligations of open data licences be assured? How can the conditions that are prescribed in the licence be enforced, to prevent de-anonymisation?

Some possible solutions may follow from recent advances in data processing and also, perhaps, in cybersecurity. In the same way as it is possible to de-anonymise data and deduce citizens' personal profiles from non-personal data by merging with other datasets, it may also be possible to use similar kinds of identification technology on the profile of an enterprise, but aimed at understanding whether there is possible misuse of apparently non-personal data. Whether that should be done is an ethical question. But technically, it may be possible, with various types of cognitive science tools and the information that is likely to be available, to try to understand if the licence conditions are being infringed.

Ultimately it means controlling what processing is permitted for a given dataset, i.e. using conditions in an open data licence to prohibit mining to reveal personal data. Also, such data should be cleaned ("scrubbing") for any remaining personal-related content before being designated as open data for free flow. Removing specific data types is conventional practice already for large datasets, so various tools already exist to enable this from the major software and database providers such as Oracle, etc, as well as bespoke tools.

4. Possible Regulatory Actions Needed

In the light of the mixing of data types and confusion over ownership, greater policy clarity is needed to balance the economic benefits of free flow of data with the need for protection of personal data. Some direction has already been expressed in regulation, such as the GDPR. To encourage the free flow of data, the obvious strategy would be to exploit a specific type of contract law that provides handover agreements for transfer of non-personal data, with conditions prohibiting de-anonymisation. However, the principal options currently proposed for contractual solutions all present difficulties, e.g.:

- **Do nothing** – use existing EU law, e.g. on copyright and contracts, etc, but this approach may not be optimal for datasets and data flows

- **Rights in data** – based on ownership rights, may be complex, possibly have gaps but could use data traceability safeguards (e.g. distributed ledger technology, DLT)

- **Soft law** - such as guidelines, likely to be too weak for real commercial transactions.

The above approaches might clarify some issues of ownership and rights over the data and so could ease transfer of data under contract but none of these options is entirely satisfactory. An agreed EU framework is necessary to promote open data, and suggested minimum objectives are proposed in Table 2.

Table 2 Proposed framework for open data with objectives and corresponding measures required

Objectives		Possible Ways Forward
Improve access to anonymous machine-generated data	Non-legislative measures	Guidance on incentivising businesses to share data
Facilitate and incentivise the sharing of such data		Foster development of technical solutions for reliable identification and exchange of data
Protect investments and assets		Model contract terms
Avoid disclosure of confidential data	Legislative measures	Default contract rules
Minimise lock-in effects		Access for public interest and scientific purposes
Clarify who owns what, and who is a data producer, eg a person's image is owned by the person; a consumer owns their identity and consumption data		Well-defined data producer's rights
Enforce disclosure of a) holding personal data; b) use of personal data		Make part of data producer's rights and data user's obligations
Enforce rights to remuneration of consumers for use of their data (less need for open data flows)		Access must be set against remuneration

Sources: Blackman and Forge, 2017; Van Asbroeck, Debussche, and César, 2017.

This combination of goals has been discussed in the past but, given the current lack of clarity, more effort is required to define the status of digital property and data ownership, preferably founded on this set of objectives. The list above has been derived from extending the Commission's 2017 Communication, Building a European Data Economy and so could form a future potential framework for open data access, based on licensing datasets.

Open data licences offer a way forward

A specific contractual approach based on the model from open source software (OSS) could provide the simple, flexible solution needed for open data licences for specific datasets. It would also control the extent and types of subsequent processing. That should offer a pragmatic option founded on well-known and accepted principles, which have been effective in the past. This is the solution that the OSS licence model offers, a model which today forms the basis of much of our current digital infrastructure, especially for Big Data (e.g. OSS framework Hadoop, which came from a collaboration between the largest Internet and cloud service

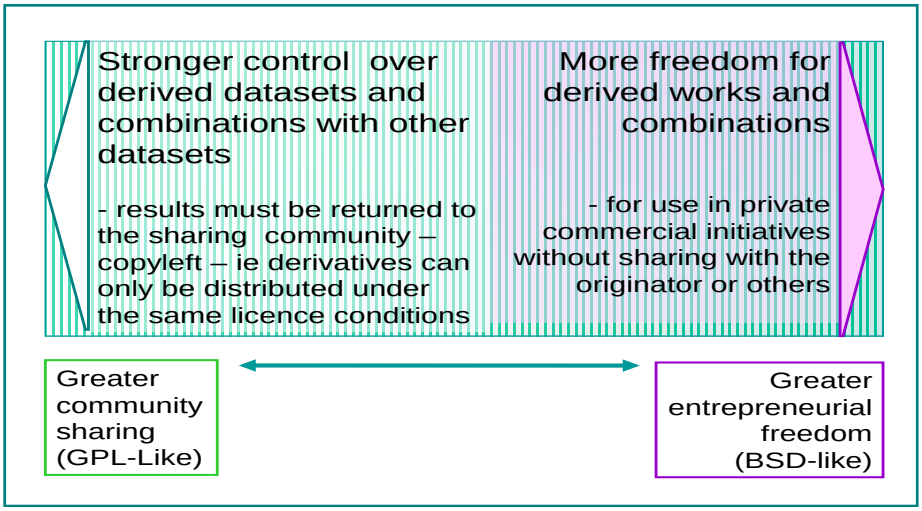
providers such as Yahoo, Google and IBM, etc, who gain mutual benefits from sharing code and development efforts). Open source licensing is used for many software components, from operating systems to web servers and browsers to office systems and device drivers (e.g. Linux, iOS, Apache, Firefox, Libre Office and the VP9 streaming codec). For datasets, it could be reinforced for traceability with a ledger system to register the multiple users of the dataset, providing a historical log of who uses what, when and how, with what obligations (perhaps using DLT).

Open data licensing is a more subtle approach and may be essential to convince the owners of 78% of the privately held data to release it in some form. In this approach, it is accepted that completely accessible and free-from-payment data flows are only one end of the spectrum of possible agreements – and many others are possible. Flexible contractual arrangements should expand the options if they open up to more commercial aims but in well-understood ways.

The types of licence envisaged could also create the formation of an ecosystem, as seen in the open source software world, where members of a software application or operating system community share the advances by returning code enhancements to the community. In this case, it could be the processed data – so the owner is rewarded by access to the refined outputs from the users of the initial raw dataset. That could offer far more opportunities for data owners, and respond to needs for collective working, under a range of conditions on the final exploitation.

Open data licences would be similar to current commonly used forms for open source software. A range of licence restrictions and freedoms would offer flexibility, ranging from community shared data to a single commercial user, as shown in Figure 4.

Figure 4 Open data licensing is flexible



Source: Author

Although custom licences are possible, a system of standard model open data licences would encourage competition, as putting data in the public domain in a controlled fashion would become straightforward, and would reduce concentration of data assets through simplified sharing.
But more is required. Unlocking the open data hidden inside business and government demands effective, specific measures for management and to provide incentives.

A framework to manage and encourage free flow of data

The difficulty in any form of legal contract is governance, that is, enforcement of the data licence with its contractual conditions. Consequently, some regulatory measures will be needed that includes the detection of infractions of obligations, for application of enforcement actions, which would form a part of the regulatory framework outlined in Table 2.

Two key pillars of this framework could be regulation, in order to:

- **Introduce open data licences** - note that the EU already has a model, in open source software generally, and has its own forms of open source software licences for controlling distribution of the IPR from its research programs, such as FP7/FP8.
- **Establish promotional measures** to encourage release of datasets. Promotion would be aimed at encouraging the freeing of data by business and the public sector (at EU, national and municipal levels) with two major targets: first more generally, for any users, be they citizens, businesses or public sector bodies; and second, for vertical sectors to enable industry-specific data to be accessed by interested parties (probably via open data licences with usage controls and payments for use).

This form of promotion would require a planned campaign to encourage opening up of datasets, within an agreed framework, for instance:

- **EU sectoral projects could aggregate datasets to build new "data commons" in specific vertical sectors** (pharma, automotive, aerospace, agriculture, mining, etc) which could mix public sector and commercial non-personal data, with access via open data licences. Sectors as diverse as insurance or agriculture could benefit from this.
- **Conditions in funding contracts for EU R&D projects** should ensure release of all open datasets, as being in the public domain.
- **Public sector bodies** might be required to release all publicly funded non-personal data on EU/national/local activities that do not have security or privacy/personal risk implications (e.g. environmental data such as rainfall and sea temperatures). This could significantly expand the sources of public interest data.
- **Create incentives** for enterprises to release data - probably from the vast data silos in many industries that they might consider has low IPR value. Such persuasion is challenging, as all data is considered not only from the perspective of potential value to the originating organisation but from the viewpoint of possibly strengthening the data owner's competitors. That means tangible benefits must be seen to return to the data owner, either in mutual exchange of different datasets, or in the return of the final processed datasets to the original owner of the "raw" dataset (as in the copyleft practice for open source software). In this way, relationships and transactions between a data donor and multiple data users should be seen as positive for both parties, i.e. a symbiotic, not a parasitic association.

A second avenue for promoting open data flows may be to examine motivations for data owners to give consent to data processing that may include some personal data. Under certain quite limited and specific circumstances, some personal data might be released for use by third parties in specific cases. For instance, medical data is already shared to some extent but might be more so, if anonymised and destined for healthcare professionals only, with the aim of benefiting the patient directly and possibly all those with similar conditions. However, this would be a limited opening of data and would be beneficial for highly defined applications. It would still require various forms of anonymisation and safeguards over who can access the dataset.

5. CONCLUSIONS

The debate on whether current EU restrictions on confidentiality of personal private information should be relaxed is unlikely to result in the growth of open data. More likely, public distrust over possible misuse of data, leakage, cyberattacks, etc, may be exacerbated.

In consequence, regulatory changes may be needed to expand the range of datasets that are available for open access in data flows and also to ensure equal and fair access for all users. Such measures should also be aimed at preventing anti-competitive practices (as the High-Level Policy Hearing with regulatory lawyers and economists emphasised).

In connection with this, a theory of harm focuses on anti-competitive behaviour in B2B relations over data flows, particularly the control of data as a source of market power, including cross-subsidisation by data. Moreover, there are several further issues to consider. For example, for personal data, the boundaries are vague between consumers, prosumers (pro-actively seeks to buy) and micro-SMEs– but whether and how that might be factored in is unclear today.

Whether EU citizens' e-privacy concerns are more about private enterprises or governments obtaining access to personal data is a further possible issue. Eurobarometer surveys point to strong concerns over both in connection with privacy invasions and possibilities for surveillance with potential misuse. Eurobarometer (2017) surveys showed that 71% of EU citizens polled in favour of privacy rather than convenience, to prevent enterprises sharing personal data and 83% of citizens want e-privacy rules).

In conclusion, the main areas for attention are likely to concern regulatory amendments with specific actions to open up free flows of data:

- **An EU data licensing structure is needed** – to promote use of open data licences, by extending existing contract mechanisms to a data-driven economy. There could be a set of EU model open data licences for different forms of exploitation ranging from user community to single enterprise uses.
- **An EU framework for promotional activities** - for open data flows should be considered with EU-wide programmes, to incentivise data owners to release more data openly for business and public sector uses.
- **Reviews of competition policy** - to promote the free flow of data, as market power can hold it back; hence extending EU competition regulation to data flows is a possibility to be considered.

6. RECOMMENDATIONS

In consequence, the Recommendations ranked according to likely impact are:

1. **Promote the use of open data licences to build trust and openness.**
2. **Promote sharing of private enterprises' data within vertical sectors and across sectors to increase the volume of open data through incentive programmes.**
3. **Support testing for contamination of open data mixed with personal data to ensure open data is scrubbed clean - and so reinforce public confidence.**

References

- Blackman, Colin, and Forge, Simon (2017), Data Flows - Future Scenarios, In-Depth Analysis, ITRE Committee, IP/A/ITRE/IC/2017-04, PE 607.362, European Parliament, November, http://www.europarl.europa.eu/RegData/etudes/IDAN/2017/607362/IPOL_IDA(2017)607362_EN.pdf.
- European Commission (2017a), "Building A European Data Economy", Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and The Committee of the Regions, Brussels, COM(2017) 9 final, 10 January, http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=41205.
- European Commission (2017b), "Framework for the free flow of non-personal data in the European Union", Proposal for a Regulation from the European Parliament and of the Council, Brussels, COM(2017) 495 final, 13 September, http://ec.europa.eu/transparency/regdoc/rep/1/2017/EN/COM-2017-495-F1-EN-MAIN- PART-1.PDF.
- Hern, Alex (2017), "'Anonymous' browsing data can be easily exposed, researchers reveal", The Guardian, 1 August, https://www.theguardian.com/technology/2017/aug/01/data-browsing-habits-brokers.
- Van Asbroeck, Benoit, Debussche, Julien, and César, Jasmien (2017), Data Ownership: A New EU Right in Data, Supplementary Paper, Bird & Bird, 31 March, https://www.twobirds.com/en/news/articles/2017/global/data-ownership-a-new-eu-right-in-data

Contact: Poldep-Economy-Science@ep.europa.eu

This document is available on the internet at: www.europarl.europa.eu/supporting-analyses

IP/A/IMCO/2018-09

Print ISBN 978-92-846-2837-7 | doi: 10.2861/762257 | QA-04-18-369-EN-C

PDF ISBN 978-92-846-2836-0 | doi: 10.2861/873820 | QA-04-18-369-EN-N