29 MAY 2020

DIGITALEUROPE’s response to the EU Data strategy consultation

Executive summary

DIGITALEUROPE welcomes the European strategy for data and the opportunity to share our members’ views via a dedicated consultation. Increasing access to data to all will facilitate the digital transformation of our societies, leading to innovative solutions and business models.

Although supportive of the Commission’s aim to unlock the potential of data sharing, we believe that the Data strategy should take into account the following:

- **Foster a partnership culture.** Private and public sectors should be encouraged to assess data sharing opportunities and determine, on a case-by-case basis, through voluntary contractual arrangements, how they can best achieve the full potential of data partnerships.
- **Ensure legal certainty.** Companies need assurance regarding their data sharing activities, for instance that they can join data partnerships without falling under antitrust legislation. For personal data, uniform interpretation and further practical guidance is needed regarding anonymisation, consent and secondary use of data under the GDPR framework.
- **Build open and reliable data platforms.** Data spaces should be based on non-discriminatory, collaborative and transparent rules. Their governance should ensure adequate representation of the private sector. Pilots should be launched and reviewed before any major scale-up. Other marketplaces and platforms should also be open to all actors.
- **Leverage global initiatives.** Potential regulation and self-regulatory schemes should consider the global data framework and exchanges, to ensure that the European Single Market remains connected to the rest of the world. Regarding standardisation, any EU efforts should be based on existing international standards and the work carried by well-established bodies.

DIGITALEUROPE looks forward to working with the European Commission to discuss and implement the ideas and proposals outlined in the Data strategy.

Table of Contents

Data governance

DIGITALEUROPE believes that the EU data governance model should leverage the potential of data to advance the digitalisation of our societies, allowing the European economy to stay competitive, while respecting core EU values, for both individuals and companies.

Access to data

Partnership freedom

Voluntary, cooperative solutions are preferred for any exchanges of data involving the private sector, whether between companies (B2B) or towards the public sector (B2G). Voluntary solutions ensure a swift implementation of data sharing practices benefitting all.

Contractual arrangements should be encouraged to share industry data, as they do not undermine the capacity of companies to enjoy mutually beneficial data partnerships. Agreements give data partners choice and control over their data and the resulting exchanges, without undermining business models and competitiveness.

Contractual freedom ensures steady investments in innovation to collect and use data, creating economic growth. Supporting industry-driven initiatives working on voluntary models and templates for more standardised contracts would empower companies to share more data – notably between SMEs and midcaps – while ensuring freedom of contract in a competitive business environment.

Additionally, the sole collection of data is not enough to create added value and must be followed by extensive data curation and data management work. Such efforts are costly and mean that companies should be free to choose with whom and how they want to share their data.

Regarding the specific case of data created by the IoT, smart machines and devices, existing agreements and contracts work relatively well, and no major challenges have been identified. Potential issues between parties can be addressed under the current legal framework, which includes intellectual property (IP), contractual and competition law.

Legal certainty

Clear schemes are needed to allow companies to voluntarily cooperate and exchange data without falling under antitrust legislation, for example through a block exemption on data sharing and pooling. Without increased legal certainty, data sharing levels and overall uptake will only slowly grow.

Difficulties in using data from other companies can also stem from the lack of legal clarity as to which extent data from businesses can be accessed, shared or mined.

Where such data is personal data, further clarification and a uniform interpretation of the GDPR[1] is needed for certain aspects such as anonymisation, consent and secondary use of data. National divergences on the interpretation of the GDPR across Europe must be addressed.

Ultimately, the GDPR legal framework and related guidance should facilitate the anonymisation of personal data in a balanced and practicable way. Guidance is needed on the legal basis under which data anonymisation can be carried – it is unclear if article 4(2) of the GDPR provides sufficient grounds under the ‘processing’ definition. And a clear list of criteria to perform sufficient anonymisation processes GDPR-wise would provide companies with the legal certainty allowing them to fully develop their data potential.

Not sharing data is sometimes the only solution to comply with data protection obligations. Given the broad definition of personal data under the GDPR, companies may prefer not to share datasets that could directly or indirectly contain personal data[2], as anonymising or separating personal and non-personal data may prove difficult or impossible.

[1] (EU) 2016/679, General Data Protection Regulation (GDPR), http://data.europa.eu/eli/reg/2016/679/oj

[2] Datasets may “indirectly” contain personal information which could be deduced from inferred data.

Allowing regulatory sandboxes for companies to freely test new data management and processing tools would also help companies develop and try innovative data uses without legal uncertainty risks.

Data security should be promoted through security by design at development phase of products and services, and then during the entire product and service lifecycle, for both new and existing/legacy data management systems (brownfield and greenfield).

Finally, Member States are encouraged to coordinate, and further harmonise their national data strategies to avoid any standalone legislation and initiatives – which could lead to fragmentation in the Digital Single Market and more uncertainty.

Awareness and guidance

Many businesses do not share their data because they lack proper guidance. Perceived concerns (security, privacy, liability and competition issues) outweigh the identified benefits of sharing data.

Clear, simple and user-friendly guidance tools would provide more certainty to businesses concluding data sharing contracts, particularly SMEs. Increased support would allow to reduce fears regarding data sharing and better explain the advantages for the business models of many companies.

Data sharing advice could be provided by further developing the concept of Support Centres for Data Sharing as one-stop shops. EU Member States should develop their own support centres. Ideally, national and local centres should form a network connected to the EU support centre, providing sound, non-conflicting and practical advice. Differing (legal) interpretations would defeat the purpose of these centres and lead to further legal uncertainty.

Non-binding models and templates could be developed to facilitate data exchanges, notably by leveraging existing industry-driven initiatives.

Strengthening the implementation of the PSI Directive[3] to ensure effective enforcement of its provisions would allow companies to try new business models based on public data and then expand to private sector arrangements. National support centres should provide guidance for both PSI and B2B data exchanges.

Data quality and information

Access to large quantities of data is ultimately of no use if data quality cannot be ensured. This includes data reliability, but also surrounding documentation, for instance information on the origin of the data, how it was prepared and how to use it. Data curation and annotation is crucial as only accurate and reliable data provides added value through data sharing.

Tagging and description of data should be clear to ensure an efficient use, particularly when data is shared in large volumes. Dataset structures and related information should

[3] Directive (EU) 2019/1024 on open data and the re-use of public sector information http://data.europa.eu/eli/dir/2019/1024/oj

then be based on generally accepted taxonomies at EU or global level, without any semantic ambiguity.

Data curation is critical to enable data access and notably secondary use of data, but this often comes with significant time delays. Technology solutions – notably AI-driven – should be leveraged to facilitate data processing and reduce delays from data collection to data (re-)use.

DIGITALEUROPE has developed recommendations on B2B data sharing to increase data generation, cooperation and exchanges between companies[4].

Data spaces, platforms and other intermediaries

DIGITALEUROPE supports the development of platforms allowing private and public sectors alike to compile, curate, share, sell, trade and access quality datasets. This could have a major positive impact on the European economy.

Common European Data spaces

Creating Common European data spaces would support the objective of making more data available for AI applications to thrive. It is however important to ensure that the development of such data space schemes is based on a robust and market-friendly governance framework, ensuring voluntary participation to the schemes.

Voluntary participation, associated with a bottom-up approach based on what the industry can offer through those data spaces, would ensure a satisfying take-up. Incentives should be found to encourage both public and private sectors to increase data generation, data cooperation and exchanges via such platforms.

A proof of concept should first be developed to ensure that the data spaces would support and enable innovative, promising and market-driven business models. In practice, small-scale pilots should be launched and reviewed before investing in the supporting technical infrastructure as planned in the Data strategy, to make sure that governance and business models are viable.

After the data spaces have been launched, regular reviews should be carried by independent third-party bodies. Such reviews and resulting reports should notably assess the impact of the spaces on the data markets and whether data sharing between companies and across sectors increased.

Data spaces should be open to all actors and based on non-discriminatory rules. Governance structures should ensure adequate representation of the private sector in

[4] https://www.digitaleurope.org/resources/digitaleurope-key-recommendations-to-support-business-data-sharing-in-europe/

advisory and governing bodies. Decision-making processes should be open, collaborative and transparent to ensure industry participation.

Data sharing between companies should be based on the principle of contractual freedom and should therefore be the result of individual negotiations between market participants: activities related to the common data spaces should not affect the general functioning of the existing market for data.

Security of data and data exchanges should be of the utmost importance to ensure that the industry could safely take part in any data spaces initiatives. Data spaces should abide by the strictest cybersecurity rules and standards, and allow the use of privacy-preserving machine learning and confidential computing solutions.

Marketplaces

EU regulators should create a framework supporting the growth of marketplaces and other platforms to strengthen data sharing, notably for sectors that would not be covered within the scope of the Common European data spaces. Those marketplaces should be open to all actors and based on non-discriminatory rules.

Marketplaces should have the possibility to be integrated into Common European data spaces on a voluntary basis, allowing marketplace vendors to make available datasets to a larger cross-sector public. In this context, public sector data made available on marketplaces should comply with relevant legislation, notably on charging costs to access datasets[5].

Data trusts

Data trusts may be interesting tools in a B2C context to empower citizens with their data and allow them to easily share it with different service providers. In a B2B perspective, data trusts may be a useful instrument for SMEs to help them managing and sharing their data.

With no wide implementation yet, there is still a lack of clarity about the data trust concept, its functioning in practice and its compatibility with existing legislation, notably the GDPR (e.g. delegation of data rights and consent to the data trust).

Secondary use of data

Developing secondary use of data for purposes deemed of societal value would foster innovation and lead to scientific breakthroughs.

Health, social and mobility are domains where developing secondary use of data would prove particularly beneficial, notably through innovative AI uses. For health, secondary use of data would facilitate the implementation of clinical trials and more generally health research, and help improving the efficiency of healthcare policies by enabling evidence-based decision-making, leading to better care for patients.

[5] Cf. Directive 2019/1024 on Open data and the re-use of public sector information https://eur-lex.europa.eu/eli/dir/2019/1024/oj

Further legal certainty is needed for stakeholders. Public authorities have a key role to play to facilitate access to such data by improving, promoting and clarifying the use of effective legal agreements, data sharing agreements and governance tools.

While protection of personal data is essential, consent should not be the default legal basis and public or legitimate interest may be an alternate ground to consent for processing of secondary data, in compliance with the GDPR. This approach is reflected in the position of the EDPB in the context of clinical trials[6]. As obtaining and maintaining consent from patients for data re-use may be challenging or even impossible, especially where there is no direct relationship with the patient or when the patient is in a vulnerable state, further EDPB guidance on such alternate processing grounds would be welcome. Clarifying when legitimate interest and public interest can be used would benefit the further processing of personal data without the need for consent.

Common, acceptable pseudonymisation or anonymisation processes, tailored to the circumstances, would enable and safeguard secondary use of data. It should be possible to use a “relative” anonymisation model based on robust standards and governance model, providing traceability back to the source records without risking subject identification by the parties involved, all considering the policy and contractual requirements as well as the security measures applied. When relative anonymisation is insufficient and there is an acknowledged risk that data subjects could be re-identified (e.g. for health, patients with rare diseases) yet further de-identification would impact the ability to use the data, secondary use may be based on an opt-out model, meaning that data subjects would notify that they oppose their data being re-used.

While this should not be the norm, it should be possible to process data within a secure environment when this is the only way to access sensitive data. This ensures that data would not be moved out of the secure environment.

We invite Member States to establish one-stop shops to facilitate the secondary use of data. Centralised authorities could be set up to handle data requests for research or other purposes of public interest. This model works well for secondary use of health data, for instance via the Clinical Practice Research Datalink (CPRD) in the United Kingdom.

Public data intended for re-use should not contain private sector data or IP, unless the data rightsholder agreed to share the data or legitimate commercial interests would not be impacted. Public data should be distributed under a licence that allows commercial re-use and derivatives should fall under the same licence as the original dataset.

DIGITALEUROPE has developed recommendations on health data processing, which notably address secondary use of health data[7].

[6] EDPB, Opinion 3/2019 on the interplay between the Clinical Trials Regulation (CTR) and the GDPR, https://edpb.europa.eu/our-work-tools/our-documents/avis-art-70/opinion-32019-concerning-questions-and-answers-interplay_en

[7] https://www.digitaleurope.org/resources/digitaleurope-recommendations-on-health-data-processing/

Data portability

We recognise the importance of data portability in a B2C context, to ensure wider data access. B2C and B2B data portability contexts should not be confused. Individuals and companies have different portability needs, which require tailored solutions.

Article 20 of the GDPR sets a wide portability right over personal data and sufficient control over who can access and use such data. Thus, we did not identify a need to expand such portability right for machine-generated data. Current need is not legislative or regulatory, but to:

- Raise awareness of individual users regarding data portability.
- Facilitate portability in a practical way for individuals, for this right to achieve its full potential. For instance, by supporting the development of standardised, secure and interoperable personal data sharing mechanisms, with real-time access when possible and relevant.
- Ensure trust by allowing users to control their portability right.

Such portability right should also allow individuals to retrieve or move data about them stored in public institutions repositories. For instance, citizens should be able to access data on their skills/education stored in university repositories, e.g. to transfer it to job search platforms.

DIGITALEUROPE members support and develop solutions that help users to move their data securely and seamlessly between service providers. Trustworthy portability tools and solutions should be further developed and promoted, e.g. MyData in Finland[8], the Data Transfer Project[9], etc. To support such activities, existing interoperability standards should be leveraged, and further work should be carried when needed and relevant.

Data donation

Data “altruism” and donation schemes are welcome to give clear, easy and secure ways for citizens to give access to their data for the public good. Data donation could have a major positive impact in some domains, notably for health-related research, but also for environmental purposes (e.g. mobility and energy data).

A good solution may be a European standard form for obtaining consent (and, where necessary, requesting data portability) from the individual, in line with the GDPR.

Depending on the data access needs, contracts may be more suited than consent as legal basis for regular or continued data donation, as consent without any kind of contract means that users can withdraw their consent at any point. Data processing would need to be stopped and a request to delete previously acquired data could be made.

[8] https://mydata.org/

[9] https://datatransferproject.dev/

Data altruism could be encouraged via model contractual clauses or data sharing agreements to which individuals would agree.

While contracts would ensure continuity of processing, consent may work better in specific situations, particularly for “small” donations needing quick approval from data donors.

User information and awareness is crucial to ensure that potential data donors understand how their data can be used and why donations are important. This can be supported by developing clear use cases while upskilling individuals to give them an understanding of how aggregated data is used to advance research and innovation for society as a whole.

Standardisation

Standardisation is essential to ensure that data can be protected yet accessible, and shared easily between different actors, analysed, compiled, and merged into additional standardised datasets.

There are already many different standards for defining semantics (common taxonomies, data formats, models, etc.), APIs and interoperability protocols. Efforts should be made to build upon current best practices.

Where new standards are needed, the role and ongoing activities of global standards developing organisations should be leveraged to avoid duplication and encourage the development of voluntary, consensus-based and industry-driven standardisation efforts. For new standards, we stress the importance of semantic data modelling specification addressing conceptual, physical and logical data models in addition to ontologies or controlled vocabularies.

European standardisation efforts should be based on existing international standards considering adoption of the work carried by well-established standardisation bodies, such as ISO, W3C and IEC. Global standardisation activities are preferred over European-centred activities, and even more over national activities, as domestic standards create further market fragmentation and technical trade barriers.

To ensure wide uptake and adoption of common interoperable standards, standardisation should be stakeholder-driven, supporting and expanding public-private partnerships. Where relevant, government agencies should participate in the standards-setting process as one of the stakeholders, with equal membership.

Literacy and skills

To make sure that data-related innovation benefits all, the general data literacy of the population should be improved. As awareness around digital technologies plays a critical role in digital transformation, data literacy should be at the centre of European efforts on digitalisation.

More data being made available is of no use if there is a lack of skilled data science professionals. Existing innovative technical solutions cannot always be implemented successfully or efficiently in organisations due to a lack of specialists with competences in critical areas such as AI, machine-learning, data analytics, and cloud computing. Organisations need to upskill their workforce and obtain support to achieve that.

Increased use of data by companies also impacts and changes employees' tasks in the workplace. The EU and Member States should facilitate and support activities to prepare professionals of all levels to adapt and update their skill set to carry data-mapping tasks and develop data-driven business models.

EU digital skills instruments and strategies should therefore have a stronger focus on data science and data literacy. Addressing data skills shortages through targeted education, including medium and advanced training, will ensure that individuals can manage and use data effectively.

Cloud computing governance

European market for cloud services

DIGITALEUROPE members represent most major cloud services providers as well as users with extensive use cases.

We generally believe that the cloud market currently offers the technological solutions needed to develop businesses and innovate. European users should have access to the widest range possible of competing cloud services. New initiatives should not lead to limitations on cloud service offerings.

The lack of public procurement rules fit for cloud services prevents public-sector organisations from taking full advantage of cloud-based innovation. Cloud public procurement frameworks should take into account the following principles:

- Support commercially available versions of cloud services, via multi-tenant cloud architectures. As cloud business models rely on economies of scale, custom-made requirements impact prices and efficiency for users.
- Shared responsibility is the basis of cloud environments: cloud service providers ensure the protection of the overall multi-tenant architecture and the continuity of service, whereas users are responsible for the content hosted in the cloud and the potential additional applications built over the architecture.
- Any contractual framework should be limited in scope and cater for shared services and facilities, which is a fundamental aspect of cloud ecosystems. Existing terms of services of cloud providers have been developed with a deep understanding of how cloud services operate in practice and factor in their constant technological evolution and innovation.

Unjustified limitations to the use of cloud services, including data residency, hinder European organisations' abilities to innovate.

Clarity around governments' access to data in the cloud (in EU Member States but also in the US and other relevant countries) would help cloud users to make risk-based decisions and alleviate concerns that hinder cloud-driven innovation. A potential EU-US agreement would facilitate cross-border access to electronic evidence in a mutually beneficial way and bring certainty to the market.

National standards on cloud computing should be avoided as they create market barriers, affecting particularly smaller cloud providers trying to scale up.

Self-regulatory schemes

Schemes awareness

Self-regulatory schemes include voluntary standards, codes of conduct and other private schemes. Countless schemes already exist for cloud services, related to data protection, security, portability, energy efficiency, etc. DIGITALEUROPE members participate in many of these schemes.

These schemes are usually based on ISO standard frameworks and other international/regional bodies, and are widely accepted. They are most successful when developed in organisations with well-established processes and IPR policies, and should be kept within these kinds of settings.

The market is already largely aware of, and driving many such schemes, which is why providers have obtained corresponding certifications. Market awareness of such schemes seems rather satisfactory, expect maybe for smaller organisations such as SMEs. Encouraging not only third-party certifications but also self-certification would be a way to raise awareness on these schemes and encourage adoption, particularly for SMEs.

We believe that there is a real market demand for schemes, as they offer common requirements and operational processes to provide certified services to users. Greater awareness should be driven by greater application from providers and stronger demand of these schemes by users.

Further awareness-raising could be done via industry forums and alliances, and by encouraging the public sector to integrate relevant schemes in their processes, for instance via public procurement. Raising awareness of existing schemes and standards is preferable to calling for new, potentially duplicative, schemes to emerge, but market awareness itself is not necessarily a problem.

When new schemes are developed, they should use existing global standardisation work rather than run parallel (and counter) to standardisation processes. Ensuring that schemes are built on recognised international work and do not create duplicates is essential to avoid disadvantaging smaller cloud providers which do not have the capacity to assess and conform with new certifications that would be local and/or redundant compared to existing schemes.

Cloud rulebook

We welcome the creation of an EU Cloud rulebook compiling existing regulation and industry-recognised standards and schemes into a single document.

The rulebook should be easy to read, to understand and to navigate, to help cloud users to prepare their cloud usage projects, plan their risks assessments, etc. Such document should not lead to re-written rules and different interpretation. Finally, the rulebook should not transform successful self-regulation into legislation.

High-value datasets selection

DIGITALEUROPE’s members welcome the 2019 revision of the PSI (Open Data) Directive[10] and the new provisions creating the specific category of public datasets deemed of “high-value”, to 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. Such datasets offer a significant potential for citizens and businesses, to help address societal challenges and to develop innovative services and products.

We hope that a successful implementation of the high-value datasets (HVDs) concept will encourage the inclusion of more categories of HVDs in the future, for instance in the health and energy sectors.

Identification factors

Regarding the main characteristics to prioritise for the selection of the HVDs, we believe that the added value of the datasets would be maximised thanks to the use of APIs and if those datasets were to be available under uniform conditions across the entire EU. When it comes to factors to select HDVs, ensuring the access to datasets from previously unavailable thematic areas and free of charge is useful, but less desirable in comparison.

APIs

The availability of datasets via APIs allows companies to automate datasets collection, which is particularly useful for datasets regularly accessed, for instance for data updates to improve the accuracy of services and products. By reducing human operations on the data, companies can ensure that data is not accessed only once or a few times, but

[10] Directive (EU) 2019/1024 on open data and the re-use of public sector information http://data.europa.eu/eli/dir/2019/1024/oj

repeatedly or even real-time, when possible. This permits to develop solutions and services that would not operate or be viable without APIs.

Uniform conditions

Ensuring the availability of datasets under uniform conditions is also crucial. Access regimes differ from one Member State to another. Depending on the country, the processes to access datasets may be very different, resulting in difficulties to collect similar datasets across all EU Member States. This prevents the creation of EU-wide datasets.

Thematic areas

Similarly, there are some thematic areas which do not have many available datasets, due to only few EU-level requirements for opening up data, leading to country differences. Some Member States have extensively developed open access policies on part of the public data but did not for the rest, which may still be locked under restrictive and differing licensing systems. Thus, access to data may be facilitated in certain cases yet rather complicated in others, depending on countries and thematic areas.

However, the fact that a category or thematic area of data is not widely available does not necessarily mean that corresponding datasets should be selected as HDV only based on this criterion. Such datasets could be made available under the regular provisions of the Open Data Directive.

Charging costs

While accessing data free of charge is important, particularly for citizens and SMEs, this should not be the driving factor in the selection of HVDs. Costs for licences to re-use public sector information are often expensive, which means that large firms may hesitate to contract such licences and that SMEs will not be able to. However, if the costs of making available the data are too important for the public sector, it may have negative side effects, such as badly collected data, rendered useless.

Reducing licences costs to only charge a small price (‘marginal cost’) may be appropriate, particularly when this includes proper data preparation and delivery costs (e.g. fast servers). Marginal costs can help public institutions to cover their costs related to the making available of data and should ensure that access to data is provided in good conditions, notably if they do not receive additional national or EU support for such activities. Providing data free of charge is then advisable if it does not lead to low-quality datasets and/or undersized access infrastructure.

IP protection

Finally, selected datasets should not contain IP elements from the private sector or lead to the disclosure of confidential data (e.g. rail datasets including suppliers’ data or information permitting to easily isolate such data). We encourage public authorities to contact companies to find a mutually acceptable solution if they believe that datasets to be released publicly may contain private IP or data related to legitimate commercial interests.

For data resulting from public procurement or public-private partnerships of any kind, we encourage public institutions to discuss with their private partners and carefully review what the existing contracts permit. There should be an agreement between partners and a possibility to opt-out if there is a lack of clarity as to whether datasets resulting from public-private interactions can be released under the PSI Directive framework (either as HVDs or “normal” datasets), notably if contract terms do not explicitly foresee sharing with the wider public.

Eased re-use

Identified high-value datasets (HVDs) must comply with the provisions of the PSI Directive (offer access to APIs, be free of charge, etc.). Besides, to ensure wide re-use of public sector data, we believe that the HVDs and other datasets falling under the scope of the Directive should also conform with good practices in the field.

Taxonomies

To be easily findable by re-users, dataset structures should be based on generally accepted taxonomies at EU or global level, without any semantic ambiguity. Currently, even when available on platforms such as the European Data Portal or national equivalents, datasets may be difficult to find because they have not been properly filed and tagged.

Whenever relevant and possible, existing standards should be used for defining semantics, formats, metadata and interoperability protocols.

Licences

Companies need legal clarity to operate. For this reason, public sector datasets should be made available under clear licences, easy to understand and to use, allowing datasets to be compiled and merged, even with different licences. Some open data licences such as Creative Commons may be a simple solution for public authorities.

Formats

Restrictive formats are one of the main barriers to the re-use of public sector data. Some datasets may be almost unreadable because they cannot be easily accessed, opened and analysed due to format restrictions. Extra operations to open and convert illegible formats reduce the effective use of data, especially when it comes to dynamic data. Institutions should support the release of machine-readable and ‘user-friendly’ datasets.

Different institutions or departments within the same public institution may also be releasing data in different formats, which complicate even more its re-use. Further harmonisation between Member States is needed to address such disparities at EU, national and local levels.

Documentation

Additional documentation on the datasets should be provided whenever possible, especially when restrictive formats are used. This would help re-users to make the most of the data they have access to, by facilitating the operations needed to use such data. Ideally, documentation should include tutorials that users can follow to get hands-on experience with data.

Documentation should not only cover the explanation about the use of the data delivered, but also include basic information on how the data was collected. Data re-users need to understand when the data was created, the methodology used to create it, how to interpret the values contained in it, and if there are any licences that may limit how the data can be used. Documentation should also include a contact point able to answer questions about the data.

If data is not documented, its audience will be inherently limited. There are times when re-users will do the 'detective' work required to interpret poorly documented data, but a lack of documentation will usually frustrate users to the point that they will simply not trust the data.

Availability

When available, real-time data ('dynamic data') should be made accessible for re-use. This is particularly needed to develop services relying on data from specific sectors, such as transport, energy, health and environment. The cost of giving access to real-time data may be important, particularly to send data to third parties wanting to re-use it: data delivery costs may be integrated into marginal costs.

Data available for re-use is often outdated. Sometimes, it takes more than a year for sets of data to be updated, even if there are no manual changes to be made. We understand that removal of confidential information, personal information for privacy reasons, etc. may extend the processing time for data to be made available. However, when there is no justification, data should be made available for re-use as soon as possible after its collection. AI-driven technologies should be leveraged to support the process of providing updated data.

Data previously made accessible should still be available in the future. Ensuring that data will be obtainable on an ongoing basis provides certainty to businesses which can develop services without the risk of investing and then having to discontinue a service due to datasets not being available anymore. Developers will not make significant effort to create tools or applications based on data if they have no assurance that the data will still be available in the future.

Access

Public sector datasets should be easily accessible, preferably on dedicated EU-wide or nationwide platforms such as the European Data Portal – or the potential Common European data spaces.

We encourage public institutions to use cloud solutions, particularly when sharing large volumes of dynamic data. Cloud services provide great platforms for storage, low-latency access and transfer of data. Users can retrieve data directly from the source, which assures them that they can reliably access a trustworthy copy of the data.

Quality

Data curation work is of the utmost importance, to ensure that the datasets are reliable and can be used for all purposes, even sensitive activities.

Sensor data ensure high levels of reliability, in comparison to man-made data, which can contain mistakes during creation or copy to datasets. Sensor or machine-generated data can also be made automatically available as real-time data. When technically possible and financially feasible, we thus advocate to avoid human interventions on datasets – or at least reduce them to the strict minimum.

EU support

EU funding programmes have a key role to play to accompany public authorities in their data collection, curation, management and delivery processes. The future Digital Europe programme should support the availability of real-time data and of documentation on the datasets and help public authorities to find solutions for storage, low-latency access and transfer of data. Horizon Europe should fund the development of innovative solutions to help public authorities to collect and manage data. Structural funds like the ERDF should assist local and regional authorities that may struggle with the implementation of open data policies.

Proposed datasets

Find below our proposals for high-value public datasets:

Dataset type	Specific datasets
Geospatial	1. GPS data, 3D mapping (building / object mapping). 2. Maps (national and local maps, cadastres/land registry, land usage, terrain form, postcodes, topography, city 3D models). 3. Real-time data on government infrastructure (roads, railways, mobile communications / Internet), water and electricity supply, construction sites, traffic signage, etc.
Earth observation and environment	1. Environmental science: Air quality (pollution levels), land and water quality, etc. and environmental disasters (including manmade, such as nuclear explosions and waste, gas leaks and explosions, etc.), biophysical parameters monitoring (status and evolution

	of the land surface, including vegetation growth, water cycles). 2. Energy data: Energy consumption, energy performance of buildings, and renewables data: Optimal locations for solar energy plants (solar panels), wind power plants (wind turbines), tidal and wave power plants, etc. 3. Agriculture, forestry, mining and fishing data, e.g. for mining: rare materials location sources; for agriculture: parcel boundaries, 3D digital model of the soil and restricted areas for fertilisers (agriculture).
Meteorological	1. General weather forecast: temperature, wind, rain, humidity, atmospheric pressure, amount of sunshine, etc. 2. Disaster and outbreak: tsunami warning, fire monitoring (urban and wildfires), seismic and volcanic activity, storms, etc. 3. Space-related, including sunspots and solar flares.
Statistics	1. National, regional and local statistics, including economic indicators and forecasts, wealth, demography/population (incl. census: population growth, age, mortality and other medical data; income), infrastructure, skills (available, in training, needed/shortages). 2. Banking, currency data. 3. Private sector-related stats: procedural data of the tax authorities, data on real estate/property purchases.
Companies and company ownership	1. Annual financial statements of companies (including annual reports, solvency, etc.). 2. Address of companies. 3. Company/business register, information on which sectoral codes companies are listed under in public sector data.
Mobility	1. Public transport timetables and real-time updates, usage, reach, intermodality. 2. Traffic updates (congestion, etc.), construction works, public gatherings and other events/activities affecting transport, including temporary and permanent traffic signage (traffic signs, road markings, lane barriers, lane markings, traffic lights), anonymised vehicle flow (urban planning and organisation) and air traffic management, airport and air traffic control, traffic models.

	3. Personal and commercial mobility, including autonomous vehicles performance (i.e. confidence in collision detection, lane identification etc.) and commercial aviation.

Additional categories

For further revisions of the categories of datasets via delegated acts[11], we encourage the Commission to also include the following categories for high-value public datasets:

Dataset type	Specific datasets
Health	Pseudonymised data, incl. treatment data (e.g. which type of patient gets treated, with what drugs and therapies, for what medical condition), prescription data, performance data (state of health of patients following care).
Social mobility and welfare	Housing, health insurance and unemployment benefits.
Real estate	Dynamics in the real estate market, anonymised information for potential sellers/buyers: ⏩ Land value of an area (average value derived from property sales). ⏩ Number and prices of properties on sale (via land registry changes and real estate tax transfers data). ⏩ Land register extracts and property plans.
Identification	Digital ID cards (eID) access data, to enable secure identification using existing eIDs (e.g. for a company to accept employees' eIDs to access the company's premises) via an open data interface/API.

FOR MORE INFORMATION, PLEASE CONTACT:

Julien Chasserieau

Policy Manager

julien.chasserieau@digitaleurope.org / +32 492 27 13 32

[11] Cf. article 13, Directive (EU) 2019/1024 on open data and the re-use of public sector information http://data.europa.eu/eli/dir/2019/1024/oj

About DIGITALEUROPE

DIGITALEUROPE represents the digital technology industry in Europe. Our members include some of the world’s largest IT, telecoms and consumer electronics companies and national associations from every part of Europe. DIGITALEUROPE wants European businesses and citizens to benefit fully from digital technologies and for Europe to grow, attract and sustain the world’s best digital technology companies. DIGITALEUROPE ensures industry participation in the development and implementation of EU policies.

DIGITALEUROPE Membership

Corporate Members

Accenture, Airbus, Amazon, AMD, Apple, Arçelik, Bayer, Bosch, Bose, Bristol-Myers Squibb, Brother, Canon, Cisco, DATEV, Dell, Dropbox, Eli Lilly & Company, Epson, Ericsson, Facebook, Fujitsu, Google, Graphcore, Hewlett Packard Enterprise, Hitachi, HP Inc., HSBC, Huawei, Intel, Johnson & Johnson, JVC Kenwood Group, Konica Minolta, Kyocera, Lenovo, Lexmark, LG Electronics, Mastercard, METRO, Microsoft, Mitsubishi Electric Europe, Motorola Solutions, MSD Europe Inc., NEC, Nokia, Nvidia Ltd., Océ, Oki, Oracle, Palo Alto Networks, Panasonic Europe, Philips, Qualcomm, Red Hat, Roche, Rockwell Automation, Samsung, SAP, SAS, Schneider Electric, Sharp Electronics, Siemens, Siemens Healthineers, Sony, Swatch Group, Tata Consultancy Services, Technicolor, Texas Instruments, Toshiba, TP Vision, UnitedHealth Group, Visa, VMware, Xerox.

National Trade Associations

Austria: IOÖ
Belarus: INFOPARK
Belgium: AGORIA
Croatia: Croatian Chamber of Economy
Cyprus: CITEA
Denmark: DI Digital, IT BRANCHEN, Dansk Erhverv
Estonia: ITL
Finland: TIF
France: AFNUM, Syntec Numérique, Tech in France
Germany: BITKOM, ZVEI
Greece: SEPE
Hungary: IVSZ
Ireland: Technology Ireland
Italy: Anitec-Assinform
Lithuania: INFOBALT
Luxembourg: APSI
Netherlands: NLdigital, FIAR
Norway: Abelia
Poland: KIGEIT, PIIT, ZIPSEE
Portugal: AGEFE
Romania: ANIS, APDETIC
Slovakia: ITAS
Slovenia: GZS
Spain: AMETIC
Sweden: Teknikföretagen, IT&Telekomföretagen
Switzerland: SWICO
Turkey: Digital Turkey Platform, ECID
Ukraine: IT UKRAINE
United Kingdom: techUK