Data access and transfer with a focus on APIs and industrial data platforms

EC workshop

Brussels, Centre de Conference Albert Borschette

June 8 2017

10:30 – 16:15

Workshop report

Objective

In the context of a broad stakeholder consultation process announced in the Communication "Building a European data economy"[1] the European Commission organised a workshop in order to deepen its understanding on the state of the discussion on two of the ways forward described in the Communication with respect to enhancing data access in B2B constellations, notably: (1) Fostering the development of technical solutions for reliable identification and exchange of data and (2) default contract rules (paired with recommended contract terms).

Participation and selection process

Participants were invited through an open **registration process**. The rule of "first-come-first-served" was applied until the level of registration had reached 60 participants. In order to ensure participation of the widest possible range of stakeholders in light of the limited seating capacity, as of then registrations were filtered manually so as to ensure that not more than 1 person per organisation/company participates. Expecting a certain "no show" rate, the event was mildly overbooked. No organisation had thus to be refused access. **68 persons** (including speakers) participated in the event. An unknown number of persons followed the event through a public web stream.

Summary

In this workshop, data sharing in B2B constellations was discussed in greater detail, both from a conceptual, technical and legal perspective. It allowed the Commission to understand the current discussion challenges an "API economy" or "B2B data sharing economy" and prepare for potentially upcoming actions.

From a conceptual perspective, the value of B2B data sharing was explained by Christian Reimsbach-Kounatze, OECD, and confirmed by Sebastiano Toffaletti, Secretary General of the European Digital SME Alliance. It was demonstrated by two case studies: Transport for London and Telefonica. Whereas Transport for London is under a statutory obligation to make certain data it holds accessible to third parties, it was understood that this still required

---

[1] COM(2017)9 of 10 January 2017.

fundamental changes in organisational culture and in the IT architectural design. The result is impressive: Building on 200 data assets, some 12,000 developers develop 600 applications that are used by 40% of Londoners. Data interactions by Transport for London (TfL) are by no means uni-directional: TfL also collaborates with other players, including big players such as Waze or Google for mutual data exchanges.

Telefonica on the other hand demonstrated how partnering with other organisations allows Telefonica to offer a richer service: Telefonica mobile phone customers can agree to allow third parties to access data they generate as part of the usage of their mobile phone with the aim of new services being offered. This requires evidently a more selective process as regards the partners.

What mechanisms are available for companies that decide to make available data to others? The workshop presented three mechanisms: It discussed the current use made of application programming interfaces (APIs) as an established means of opening up data to a wider ecosystem. The approach is largely tested and relatively well used. Challenges include data security and licensing conditions. Industrial data spaces and data marketplace, on the other hand, are designed for individual data transactions with one or a limited number of buyers. Both are emerging technological offers having the characteristic of multi-sided platforms, requiring network effects and sufficient scale in order to become relevant for B2B data sharing. The specific concept of an industrial data space as developed in Germany also includes the possibility for data experimentation to happen inside the space, whereas Dawex currently limits its business offer to assisting in the data transaction, paired with related services such as data anonymization (for transfer of personal data).

Data transfers need a legal reflection, typically a data usage agreement. Contracts and contract law are there to reduce transaction costs. Model contract clauses, offer e.g. by Dawex, help further lowering such transaction costs. Contract law (and to some extent model contract clauses) can also achieve another objective: Give guidance on what constitutes a fair and balanced contractual relationship. European contract law rules can reduce additional transaction costs in cross-border constellations. Also the usage of an API raises a sufficient number of legal questions separate from questions on data usage, justifying model agreements just for this. An interesting stream of discussion revolved around the question to what extent automation in contracting through machine-readability of contract clauses is feasible mid-term.

Detailed account of the presentations and discussions

The scene was set by two presentations that aimed at describing the rationale for B2B data sharing.

1) Access to data & data-driven innovation, Christian Reimsbach-Kounatze, OECD

In the first presentation, Christian Reimsbach-Kounatze, described why enhancing access to data, including privately-held data, deserves attention from policy-makers from an economics of data point of view: The roll-out of IoT objects, the digitisation of physical assets and the

datafication of business processes lead to ever more data being generated. Such data is by its nature a non-rivalrous resource but by no means automatically available for use by anyone as its use can be restricted (excludable resource). Consequently, data should not be compared to a depleting resource like oil but should rather be considered as an enabling infrastructure. Data in this view is also a capital with increasing returns as there may be multiple ways of re-using the same data as input for other products or services or in order to improve processes. Additionally, with Big Data, insights can be derived through data analytics which in turn can improve products or services or increase efficiency of production processes. He also argued that data had no intrinsic value. Value-creation depended on other complementary factors such as software or skills.

In certain cases, making some data available can improve interactions with customers and lead to gains both on the side of the company making data available and the customers. This can enhance the option value of investments made into data collection, in other words the value arising from the fact that there are multiple usage options for the data.

Citing forthcoming figures from the OECD Digital Economy Outlook 2017[2], 80% of respondents recognise policies for facilitating data reuse across organisations and sectors as important. In particular, more and more government authorities are recognising the added value of "open data" initiatives.

Policy-makers are thus presented with a policy dilemma of on the one hand putting in place policies that enhanced access to data also in B2B constellations and on the other hand preserving individuals' privacy, confidentialy of commercial operations and the investments made into data collection(s). He argued, that openness with respect to data is not a binary concept and that there could be degrees of openness when it comes to data access (eligible parties, conditions under which data can be accessed). He suggested using the concept of "data commons" as a description for all degrees of data openness. He said at this stage, there is still some conceptual work necessary in order to correctly assess the social and economic costs and benefits of greater openness, of designing the available mechanisms of governing such data access and selecting the most efficient approach. Some evidence is available for the costs and benefits resulting from re-using public sector information ("open government data"), but only limited conclusions could be drawn for B2B constellations.

In response to questions, Christian specified that "machine-generated data" was still a broad concept and taxonomies were necessary to identify which data are sensitive and which ones are not.

2) "What if I had access to such data", Sebastiano Toffaletti, European Digital SME Alliance

In the second intervention, Sebastiano Toffaletti, the Secretary General of the European Digital SME Alliance, made a strong plea for Europe to seize the opportunities of data-driven innovation. He put the size of major data-intensive economic operators in relation to the size

---

[2] Publication forthcoming

of European economies in order to demonstrate the importance of the subject. In his view, current initiatives such as digitising European industry ("Industry 4.0") were important but not aiming to seize the full potential of data: The data generated in the smart factory of the future is mainly used for process optimisation, but not in order to enable new business opportunities. On the other hand, major non-European players like Google are building a very strong position with respect to data, including data generated by apps developed by European developers. He pleaded for Europe to follow the path of moving into the "servicisation" of everything and take measures to ensure its "digital sovereignty". One important field in order to re-establish a level playing field was to ensure an open and accessible data market, especially when it comes to data from connected, sensor-equipped "IoT" machines, tools or devices. He took the view that SMEs would in particular benefit from this immensely from this. Currently, it is very difficult to exemplify as limitations in data access do not allow SMEs to bring new business models to the market. Under his understanding, the issue is, however, not specific to any sector, but is of concern in all segments of the economy. The European Digital SME Alliance has member companies active in all industry sectors.

He made an explicit plea to ensure that the user of a connected, sensor-equipped "IoT" machine, tool or device has the right to access and use the data generated by it and allow access and use to other interested parties.

In the open discussion, participants brought in different considerations: Both Mehdi Medjaoui (APIDays) and John Foster (Telefonica) pointed to business models that entirely transform what is today the sale or lease of an object (aircraft/car) into a service, either a paying service (Rolls Royce jet power-by-the-hour service) or a service being free, but given on the premise of full exploitation rights on the data generated in the context of the service. Any move towards enhanced openness could make such – in their view – legitimate business models impossible. A representative of FIGIEFA, an organisation that has as objective "maintaining free and effective competition in the market for vehicle replacement parts, servicing and repair", argued that leaving B2B data access and transfer entirely to the parties involved can produce satisfactory results, but only under the condition that all parties have an interest of working together. In their view, car manufacturers, however, are expanding into secondary markets in which FIGIEFA members are currently active, using the competitive edge resulting from access to data generated by the vehicles they manufacture.

3) "Why we chose to open up some of our data – and how we did it"

Subsequently, two presentations illustrated examples of companies that actively engage in B2B data sharing.

John Foster presented on behalf of **Telefonica** the company's efforts to allow secondary usage of data held by a mobile phone operator. Examples are almost infinite and range from more personalised services linked to the original service offer (such as warnings if consumption patterns vary from usual patterns) over additional personal "Big Data" services (such as: "your credit card company signals cash withdrawals with your card in place X whereas your mobile phone data suggest you were in place Y") to "Big Data" services that are useful for society at large ("there has been a Ebola outbreak in this region and people

movements from mobile phone data suggest that it may spread to these other regions which may help preparing the emergency response"). For this, Telefonica has developed a new platform called Aura. One of the objectives is to incentivise the individual to share data in a trusted and secure manner. Users can see at one glance what data they share and with whom in a much more digestible way than traditional consent notices.

Ultimately this platform shall not only connect mobile phones, but all connected objects. It is also designed to cater for non-smart mobile phones (in usage in major markets of Telefonica).

Telefonica has begun to design an open ecosystem of trusted partners that connect to the Aura platform in order to consume customer data in return for services with the user's consent. The ecosystem is open in the sense that it conceptually aims at a wide range of partners. The partnership programme is only beginning to develop. The question whether the ecosystem is also open in the sense of allowing competing services providers to become partners has not occurred very often, with one exception: Telefonica partners with competing retail banks as it is clear that customers have their respective account with different banks. Partners are chosen hoping to develop a mutual win-win, providing value for the customer. They are chosen on the basis of trust, sign bespoke contracts, receive data through APIs. They have to undergo a screening process, also in order to ensure respect of data protection legislation. Partner organisations receive only the relevant insights they need for the specific service in question – no bulk access to customers data. John Foster was somewhat sceptical as to whether the right to data portability under article 20 GDPR is really good news in this respect. Whereas under Telefonica's programme partners receive only the minimum of data necessary, under the GDPR customers may ask a bulk download from Telefonica and "port" the data to any third party they desire. This could, however, prove harmful for individuals if they simply ask to port too much.

Part of the business drive results from costs savings (notably reducing calls with simple requests to the call centre). Additionally, Telefonica considers the collection of the data in question already to be amortised through the mobile phone subscription to which it is ancillary.

In the second presentation, **Rikesh Shah** from **Transport for London (TfL)** explained the creation of an open ecosystem around the multiple travel offers proposed by Transport for London (underground, overground trains, buses, trams and parts of the London street grid). Transport for London is a local government body.

Its open data policy is guided by the principles of transparency (as a local government body), reach (enabling the relevant information to reach the customer in the best possible way), facilitation of niche products, creating additional economic activity and facilitating innovation.

TfL departs from the premise that it would be commercially not wise to expect customers to use one of the TfL apps as a channel of information. Also, before the adoption of the Open Data policy, some developers tried to screenscrape information from TfL websites. This resulted in multiple errors in the data which lead to sub-optimal results.

Therefore 200 data elements are made available through an API to some 12,000 developers producing some 600 apps that 40% of Londoners use.

TfL has formed partnerships with major IT players such as Apple (for mobile payment, rental of bikes), Twitter (for pushing alerts out), a two-way data-sharing agreement with Waze (enriching the app with data from the road network TfL manages while itself benefiting from data collected through Waze) and Google (enriching the maps application with real-time data).

In terms of creation of additional economic activity, it has been calculated that this policy generates GBP 100m of direct value and has enabled some 1,000 jobs.

The data can be consumed under the terms of the UK Open Government Licence with some minimal additions for free. This is done under a statutory requirement as part of UK legislation. Mechanisms are in place to ensure that consumption remains at an acceptable level. There is one single set of data at the base that are both consumed by TfL for its purposes and by third party developers. Developers must give attribution to TfL for the fact that their app includes TfL data.

Where personal data are at the source, e.g. data generated by users touching touchpoints at access gates for payment purposes,

For data acquired by a third party, e.g. Waze data, restrictions resulting from the partnership agreement apply.

All data made available are data that TfL collects anyway for its own purposes. TfL is not collection additional data merely in view of making them available to third parties.

Mashing data provided by TfL with privately-held data can bring additional insights (e.g. "Are there correlations between rainfall and collisions involving cyclists?).

4) What role for model licences for data consumption through APIs?

In the **third session**, **Andreas Krohn**, consultant at Dopter, reported on an initiative funded by the Swedish innovation agency VINOVA in 2013 aiming at lowering legal costs in relation to licensing the use of data through APIs. This led to the "Swedish API licence", a modular legal document tackling common questions but also allowing the licensor to make choices. This allows for flexibility as not all data are created equal. It was drawn up with both data publishers and users in mind. It builds on Swedish law, documented in an Annex to the license, which, as Dirk Staudenmayer noted in comments, is for many parts an implementation of EU Directives, making porting into other jurisdictions possible. It is meant to be useable also for non-lawyers. Main elements covered: IP protection, source reference, personal data, limitations, warranties, liability, payment and termination, modification of the API. On personal data, he acknowledged that more detail work would be necessary.

It also includes a customisation wizard allowing any interested party to generate a customised license.

The use of the license itself is licensed under open terms and free to use. This makes listing use cases difficult. Andreas is aware on some examples, however.

It is required, however, that attribution is given to the original source, mainly so that any licensee understands that it builds on the Swedish API license project and allows seeing changes in comparison to the model.

In terms of lessons learnt, an in-built flaw results from the funding being project-based. This did not cater in particular for maintenance and for dissemination (marketing). Also, in future exercises, data delivery and data use should be separated. For data usage licenses, the Creative Commons licensing suite provides already a good basis.

Dirk Staudenmayer, Head of Unit "Contract Law" in DG Justice of the European Commission, built on the presentation by Andreas Krohn and expanded on the potential ways forward mentioned in the Communication "Building a European Data Economy" of 10 January 2017 (COM(2017)9). The Communication discussed essentially two complementary ways forward: Setting down in legislation default contract rules complemented by a set of recommended standard contract terms designed by stakeholders. Default contract rules would set a benchmark of what constitutes a balanced solution in order to reduce the negative effects of the imbalance in bargaining positions while allowing at the same time a large degree of contractual freedom. Such default contract rules, which could be modified or derogated from by the parties, could be paired with an unfairness control mechanism: unfair contractual clauses which deviate grossly from good commercial practices and from the default rules should be eliminated by the unfairness control mechanism, while the remainder of the contract would stay intact. These default contract terms could be complemented by a set of recommended standard contract terms designed by stakeholders. The objective would be two-fold: Achieve greater fairness of data licence terms and legal certainty across borders.

In the discussion, Christina Reimsbach-Kounatze, OECD, wondered whether it is opportune to combine two elements of licensing: The conditions under which data are being made available ("how you get the data") and the conditions under which the data can be use ("what you can do once you have the data"). Andreas considered that there was a case to build one license with those two parts. He also believes that there were already a number of model licenses available that tackle the latter question. Mehdi Medjaoui wondered about the potential benefits of having such a license available in machine-readable terms. Andreas Krohn considered that reaching some level of machine-readability was inevitable, but that it would be hard to predict by when this would be achieved. With more funding this could have been achieved already in the Swedish API licence project.

5) The future of B2B data-sharing – APIs & Industrial Data Spaces: Panel discussion

In the **final session**, different technical mechanisms for B2B data sharing were presented and discussed.

Erik Wilde, API Evangelist at CA Technologies' API Academy started with a discussion on the use of **application programming interfaces (APIs).** APIs are a technical enabler of (data) interactions within one organisation and business partners. Their exact usage depends

on how a firm or organisation seeks to organise production processes, namely insourcing versus outsourcing/partnering. There are three types of behavioural patterns describing companies' choice to open up data through an API: Companies can feel to be pushed to use APIs as a means to remain competitive in an ever-changing business environment, to interact in a smarter way with partners, pushing out relevant data in return for new business opportunities. Thirdly, APIs are used now by multiple companies all over the world. Companies can thus feel to be pulled into an emerging "API economy" in which business can benefit from network effects among those having adopted APIs for the interactions. Finally, companies can regard APIs as a safe and tested technology that can be adapted without major business risks. The usage of APIs in the economy is according to him best described by four key elements: *Decentralisation* as the guiding principle of the architectural design, allow *diversity* in concrete API design so as to ensure flexibility, ensure APIs *discovery* and ensure *description*, also as a means of ensuring discoverability. The EU could contribute to the further development of the API economy by support those companies and organisation who use APIs to do it well in order to reduce friction. Helping those companies who already think of opening up data to partner companies may at this stage be more impactful compared to working on regulation that could oblige companies to make data available. Concretely, sharing of best practice, also through designing a labelling system for APIs could help companies. This is also the main objective of the API Academy that should not be considered as an extended front office of CA Technologies only created to bring new clients, but an independent partner. Best practice guidelines would preferably come from companies. Big organisations in his view are oftentimes quite ready to share such guidelines. Good guidelines are a reflection of good API practice. Working on API licensing questions as brought forward in the previous presentation could be another area of work. Preserving data security when opening up data through an API remains an issue, but there are technical solutions on the market He made a call on companies to start small somewhere inside the firm, forming a small ecosystem of internal and external players around an API. Finally, making data available through an API is one form of practising "open data". Erik suggests that the two discussions should not be disconnected. This also means that there is a lot of potential for the public sector to lead by example as it is already making available a lot of data as open data.

The second presentation in this session (by **Jan Jürjens** of the Industrial Data Space Association) described the concept of an "Industrial Data Space" as a platform for B2B data exchange and experimentation. One of the key objectives is to combine the objectives of wider data sharing in B2B contexts with maintaining what he referred to as "data sovereignty", i.e. the capacity for a data holder to stay in control over access to and use made of data. He described the Industrial Data Space initiative is a user-driven initiative. Its objective is to enable a trusted data exchange and processing between multiple actors in a given data ecosystem. Its main design principles are: a decentralised, federated architecture and openness to players/neutrality from any player, governance through self-regulation (rules elaborated in a collaborated manner). It would provide for secure data exchange, for trust among participating players through certification mechanisms and sovereignty over data and services, in particular allowing control over usage of data brought into the data space by defining usage conditions (licence agreement between data holder and data user). The

Industrial Data Space is thus not meant to create a huge data lake, but rather an exchange platform. It also goes beyond a mere data marketplace and allows for data services to operate inside the Industrial Data Space, e.g. data aggregation, mapping, anonymisation, alerting, data quality management. The potential application areas include but are by no means limited to: Personalised medicine, smart traffic management, supply chain transparency, smart manufacturing. Currently, some 80 companies, including major European data economy players, are members of the industrial data spaces association.

Responding to questions, Jan Jürjens expanded on the governance model. Currently the experimental platform is developed and operated by a publicly registered not-for-profit organisation under German law (*eingetragener Verein*). This entity develops the common governance rules for data management and exchange. For the time being, license agreements are developed on an ad hoc basis. There are no model licences for data exchange developed yet, but work has begun. He explained the features of the architecture with respect to cybersecurity, in particular the fact that data are only transmitted in an encrypted way and data are only communicated to the extent and to the level of granularity required. This makes it more difficult to allow analytics on the encrypted data at the current state of technology, but advances possible. Another solution could be to bring the analytics to the data and only allow the results of the analytics to be released. Data can be exchanged according to needs in batches or as a stream. The platform allows also for participating entities to perform the function of a data broker.

In the third presentation in this session, **Fabrice Tocco** from the French start-up Dawex introduced the notion of a data marketplace as an independent third party intermediary. Building on a market potential for data trading platforms estimated at $ 500m, Dawex intends to provide a secure and easy-to-use place for sellers and buyers to meet and make data transactions. He described it as a mixture of Ebay, Amazon and AirBnB for data. It aims to overcome the following three challenges: Get into contact with potential buyers or potential sellers, agree on a data usage license and arrange for the data transfer in a secure way. Dawex thus operates a platform through which data can be exchanged (e.g. data locked-in after first inspection by potential buyer so as to prevent tampering by potential seller while negotiations are on-going). It provides for a modular licensing template that can be easily customised. Fabrice Tocco believes that the models allow for balanced deals. Active since 2015 and live since November 2016, it operates on a global level, mainly in Europe and the US. At this stage, Dawex has some 600 companies from some 20 different sectors of the economy on the platform and allows trading of any kind of data, allowing one-off deals as well as streaming on a subscription-type of basis. Importantly, it allows for enabling access to data generated by IoT devices. It has entered into a strategic partnership with Engie for smart city data. It is financed through a transaction fee of 5.5% on each data transaction.

In response to questions, Fabrice Tocco specified that Dawex is a B2B platform which excludes individuals making their data available "for sale". Data providers have to ensure respect with applicable data protection legislation. The platform helps data providers both in terms of the license statement and with anonymization of the data. The approach has been developed in collaboration with CNIL, the French data protection authority.

The panel discussion was opened by **Mehdi Medjaoui**, founder of APIDays, a conference series specialised on aspects of B2B data sharing. He pleaded for much more automation B2B data sharing. Companies hold vast amounts of data. The Swedish API licence is one of the building bricks. Lowering transaction costs needs to be one key objective. If transaction costs are too high, then companies will internalise operations. APIs are an important technical enabler as they allow introducing granularity in the data sharing. In the future we will need to manage both data shared in some sort of commons and entirely private data. Yet, he recognises a clear trade-off as reflected in the patents system as a way of managing the two competing interest of protecting an invention against free-riding while enabling innovation by others.

In the discussion, Christian Reimsbach-Kounatze argued that next to the technical and legal questions around APIs and their deployment, there would also be a need for a change in culture in organisations. He pleaded for a positive vision of a "data commons" which should be understood as a way of conceptualising two things: (a) data cannot be over-used due to its nature as a non-rival resource (and contrary to the tradition/medieval use of the concept of commons for a plot of land on which all villagers could let their animals graze with the danger of over-grazing) and (b) data are most often the result of actions of different persons or entities (generators, collectors, processors, enrichers etc.) which all have some sort of stake that should be reflected by specific rights on the data. It is evidently a contestable concept as a result of certain connotations it has. However, it would be important to understand that "data commons" would conceptually not have to mean free access to any possible party.

Dirk Staudenmayer pointed out that such a concept would – very much like model contract clauses – rely on the willingness of entities to share data.

Mehdi Medjaoui wondered about the desirability to make API design copyrightable. Linking to a recently closed dispute between Google and Oracle on what constitutes fair use of an API design he wondered whether APIs were "the new data" in software patents.

Representatives from organisations representing companies in the automotive aftermarket made the case for open data markets: Automation, standardisation and potentially even legislation may be necessary, in particular to see the interests of the SMEs. Also, it may prove very difficult to achieve balanced contractual relations when the control rights are not clearly allocated. Such allocation could be an opportunity to re-balance power in certain markets. Christian Reimsbach-Kounatze supported the view that SMEs would deserve more attention. Individuals have a right to data portability as a result of the GDPR. Larger business can – depending on their position on a given market – negotiate favourable terms and conditions. SMEs, however, usually are not in that position. Possibly, the law on unfair business practices would need to look at this.

Two participants intervened with respect to sharing of personal data. If model licenses for sharing of data were to be created that would include personal data, the legal base layer of the right to data privacy as an inalienable right would need to be respected. Machine-readability should be aimed for. APIs can be potentially powerful also with respect to respecting data protection legislation: The fact that consent was given could be documented through an "API

handshake", i.e. a request for consent and a consent receipt documenting the scope of the consent given in exchange for permission to access data. Transparency about the future data processing would be given through the API as a "pre-defined dataset". APIs could provide for data security by using encryption, pseudonymisation & tokenisation standards. One participant called, however, to go beyond analysing merely B2B constellations when looking at future data sharing: additional constellations include: B2C, C2B, B2G, C2G, C2C etc. That would plead for a more encompassing framework of data exchange. This would suggest deepening the discussion also on sharing of personal data. Dirk Staudenmayer wondered what certainty a consent receipt would give any data processor in an environment in which consent can be withdrawn at any time. It would seem to be very difficult to build a business model on this basis.

Yvo Volman in his conclusion thanked all the participants and discussants. He concluded that the discussion had shown the importance of APIs in B2B data interactions. As part of the follow-up to the Communication "Building a European Data Economy" of 10 January 2017 the EC was now exploring what measures it should take over the coming years. This workshop had been highly instrumental to deepen the EC's understanding on B2B data interactions.