Smart Data

Smart Data Business Models

Working group on economic potential
and social acceptance

Publication data

Published by
Smart Data Accompanying Research
FZI Research Center for Information Technology
Berlin Office
Friedrichstr. 60, 10117 Berlin
www.smart-data-programm.de

Evaluation, editing and design
LoeschHundLiepold Kommunikation GmbH, Berlin

Translated by
Federal Ministry for Economic Affairs and Energy

Status
January 2016

Printed by
WIRmachenDRUCK GmbH

Picture credits
Sergey Nivens – Fotolia.com (Title page)

Supported by:

Federal Ministry
for Economic Affairs
and Energy

on the basis of a decision
by the German Bundestag

Table of contents

Foreword

Data is becoming more and more important in our modern society. The increased use of high-performance smartphones and the use of social media have become part of everyday life. The number of sensors in all forms of products and classes of equipment is also growing. In fitness wristbands, manufacturing equipment, vehicles, washing machines and heating systems, to name just a few examples, sensors are constantly recording a large number of parameters. The need to collate and process these new volumes of data has led to the widespread use of the concept of "big data" in recent years. Once the data is collected, the focus is on deploying them intelligently. This is because, if the commercial and social potential of these data is to be utilised, it is necessary to extract information for analysis so that big data can become smart data. The Federal Ministry for Economic Affairs and Energy has recognised this opportunity, and created the technology programme "Smart Data – Data Innovations" to promote 13 select flagship projects around Germany that aim to develop innovative smart data services for industry, mobility, energy and health.

The working group on economic potential and social acceptance is operating as part of this programme. Representatives from the projects talk with internal and external experts about interdisciplinary aspects of the commercial and social exploitation of data. The aim here is to formulate joint positions and thus to help shape the policy debate on smart data with a particular focus on commercial exploitation issues and social aspects.

This position paper is the group's first publication, and focuses on smart data business models. The reason why the group started its substantive work with this issue in the summer is that the underlying business models will only be successful if society accepts the technological innovations. The smart data environment offers particularly good opportunities for the development of such business models; it also poses challenges, and the working group looks into these in this paper.

I would like to thank the team of authors and the members of the group for their intensive collaboration, and I wish our audience an informative read.

Prof. Dr. Christof Weinhardt
Director at FZI Research Center for Information Technology and Head of Accompanying Research of the "Smart Data - Data Innovations" technology programme and of the working group on economic potential and social acceptance

Summary

Using smart data to create added value

Users of smart data solutions can derive economic advantages at various stages of data processing, e.g. in making accessible, storing or recognising data, or improving their quality. The providers of such solutions can generate added value by processing, enhancing or converting existing data, e.g. for use in analytical tools or in smart data infrastructures. Companies can also generate crucial added value by further developing existing solutions relating to smart data technologies. In all the forms of value creation, it is vital to gain a holistic view of complex and distributed scenarios and processes involving a large number of data sources. Depending on the desired value creation, a distinction should be made between the development of domain-specific applications and services, tailored solutions for a specific field of application, and pilot systems and infrastructures which permit the stakeholders in a field of application to obtain added value themselves.

Addressing all the participants on multisided markets

In addition to technical innovations, appropriately designed revenue models are an essential component of successful smart data business models. This is because the specific revenue mechanism ultimately determines how and where revenue is earned. In the relationship between revenues and costs, some specific features of the digital economy require attention, such as multisided markets. If smart data solutions are to succeed, it is crucial to identify the various market participants and to include them in the creation of the solution. Also - particularly in multisided markets with overarching supplier and user structures - there is a high level of dependency on data providers. If data are not obtained directly in-house, but are obtained together with other companies or provided by third parties, changes in the conditions governing the collaboration can represent a great risk for smart data business models. An early and durable resolution of questions of data availability is required in order to minimise these risks.

Organising data trade

A single set of data does not generally generate added value. However, further sets of data which can be used to enhance already available data are not always freely accessible but have to be acquired on a commercial basis. This implies that entire sales chains of data can be established. However, this trade in data is impeded by the lack of one or several (German or European) market places for the secure buying and selling of data based on clear licensing provisions.

Creating incentive systems to release data

Public acceptance plays a central role in the context of smart data, and is one of the most important preconditions for success with smart data solutions. In order to generate this acceptance on the part of consumers, it is necessary to take the public's reservations seriously. Firstly, it is necessary to promote general awareness of the smart data technologies being used, and secondly, it is necessary to formulate, communicate and uphold clear ethical principles.

If a business model involves the use of certain personal data, any generation of added value will ultimately depend on how an attractive incentive system for the release of these data can be developed and implemented by the holder of the data.

Pressing ahead with open government data

When it comes to enhancing data sets, an important role is played not only by in-house data, but also by open, freely available data sets. Government data represent a substantial proportion of the global volume of data, and it is now easier than ever before to network them with other data and to evaluate and interpret them smartly. But this raises questions about access to and use of these government data. With a view to creating the necessary incentive structures, the pub-

lic sector must play a pioneering role by providing its data in a systematic and structured way, thus giving substantial support to innovative smart data services. Open government data are an important production factor for the German economy, and must be recognised as such.

Establishing technical standards

Before raw data can be used in smart data applications, they often need to be laboriously structured, enhanced and processed. These pre-processing stages are labour- and cost-intensive. Data need to be digitised, and their seamless technical integration needs to be made possible. If data are to be used for targeted analyses or new services, they must be accurate in terms both of syntax and, particularly, of content. Given a high diversity of data sources, some of which are new, unstructured or semi-structured, however, it is often difficult to determine the quality of the data. There is currently a lack of standards which offer criteria and benchmarks for the assessment of data quality. For this reason, the establishment of such standards should be promoted - on more than a purely national basis.

Training skilled personnel

The growing demand for smart data solutions is leading to new jobs being created in the IT sector, but these are proving difficult to fill. The problem is that the requisite skills refer to a very young field of computer science, and the appropriate courses of vocational training and higher education are only just being designed. Domain-specific skills to evaluate and interpret data and to provide data-based decision-making support should not only be taught at higher education institutions, but also, for example, via in-house training courses.

1 Added value and the generation of added value

Dr. Martin Memmel, German Research Center for Artificial Intelligence (DFKI), Smart Regio

Business models in the smart data context can be designed in as many different ways as the research projects of the smart data programme itself. In this publication, the concept "business model" follows Stähler[1], i.e. a combination of the three components Value Proposition ("What benefit does the company create?"), Added Value Architecture ("How is the service provided and in what configuration?") and Revenue Model ("How is money earned?").

Depending on the application scenario and the relevant target group, many different types of added value can be generated in the smart data environment. Here, added value is understood to mean characteristics which distinguish a product or a service from others and thereby create value. This added value can play a part at various stages of data processing: in the storage of data, in making the actual data accessible, in the access to data, in the quality of data, in their analysis and in the creation of durable ecosystems for data and applications.

The following sections make a fundamental distinction between the types of added value for the users of smart data solutions and the types of added value for their providers.

1.1 Added value for the users of smart data solutions

Data storage
In the context of many different processes, companies accrue very large volumes of data which can be used in many different ways. Often, however, the sheer volume of information which is generated e.g. by sensors in industrial manufacturing processes exceeds the capacity of the systems in use to store these data. Here, smart data infrastructures can help, by offering appropriate storage structures and interfaces.

Access to data
Irrespective of the specific field of application, the decision-making by end-users and companies depends on what data and information are available to them. Here, a central role is played both by availability of data and simplicity and speed of access, and also by legal issues, e.g. relating to data protection. Leaving aside the challenges posed to data privacy law by big data, smart data technologies actually offer the potential to solve some of these problems by making possible the legally compliant use of data by anonymising them. In an ideal case, smart data solutions offer users the possibility to select - out of a potentially very large, distributed and heterogeneous volume of data - precisely those data which are of relevance for the specific context.

Networking data
Both in open and in closed scenarios (in which for example only data sources within a specific company or jointly in a process of networked companies are used), relevant information is frequently distributed across a large number of different sources. Firstly, many of these sources are unknown; secondly, accessing them often requires specific technical skills. Furthermore, the data tend to be in heterogeneous formats, so that laborious conversion processes are required before they can be used. Here, smart data solutions can make a major contribution by accessing the diversity of the data better, so that full use can be made of them for analyses and strategic decisions.

Data quality
The mere existence and availability of data is a necessary precondition for the generation of added value from them. But the quality of the data is also crucial. Alongside formal aspects, such as adequate representation and availability, content issues are of particular relevance here. Smart data technologies can help with recognising and improving the quality of the data, e.g. by combining them and enhancing them with information from networked sources (news, social media, etc.).

Data analysis
When relevant data sources are made accessible in an appropriate form, new findings can be extracted from the data. Here, a variety of smart data services can be used to select, analyse and visualise data. This can enable findings to be generated automatically in many cases. Also, the users themselves can draw on a wide range of ways to gain fresh insights. Finally, the networking of data makes it possible to answer questions that have not previously been asked.

Creation of sustainable ecosystems for data and applications
If the market for big data applications is to be developed, it is necessary not only to implement specific smart data technologies, but also to design new methods, processes, standards and business models. These help producers and users of data to implement their own processes whilst complying with legal, data protection and socially relevant aspects.

1.2 Added value for the providers of smart data solutions

Providers of smart data solutions can generate added value in many different ways. In addition to the provision of advisory services, added value can particularly be generated via

- the provision of data in a suitable form,
- the enhancement or conversion of existing data (e.g. for use in analytical tools or in smart data infrastructures),
- the further evolution of existing solutions relating to smart data technologies,
- the provision or adaptation of smart data services, particularly on a sector-specific basis, and
- the operation of smart data infrastructures which can be used by companies or by other providers to apply their own smart data services.

Examples of applications

- Health: better, customised patient care[2], technical upgrading of medical equipment[3]
- Energy: security of supply[4], efficiency, conservation, grid stability, improving energy management[5]
- Transport: more efficient mobility[6], logistics, ecologically sustainable traffic planning, prevention of transport and travel risks[7]
- Industrial context: improvement of processes[8] (quicker, cheaper, more sustainable), optimal use of resources, early error detection[9], near-time error response

1.3 Forms of added value generation

There are many different potential approaches to attaining the desired added value in the smart data context. They are all based on a holistic view of complex and distributed scenarios and processes involving a large number of data sources and an explicit targeting of SMEs. Depending on the desired added value, a distinction can be made between the development of

1. specific and domain-specific applications and services which deliver tailored solutions for a focused field of application (direct added value generation),
2. pilot systems and infrastructures which enable the stakeholders in a field of application to generate added value themselves, and
3. methods and business models which provide a sound basis for further developments.

Specific applications and services

The focus here is on services which enable the integration, aggregation, networking and enhancement of data. In particular, technologies from the field of information processing and artificial intelligence are used to develop solutions to the following characteristic problems in the field of smart data deployment scenarios:

- heterogeneity of structures and forms of representation,
- distributed data sources with differing access mechanisms,
- storage and processing of large volumes of data

which in many cases are constantly being produced,
- differences of data quality,
- a lack of data classification and
- problems with access speed.

Pilot systems and infrastructures

Smart data infrastructures aim to offer not only basic functionalities to integrate and store data but also advanced services, and also to permit the services of other providers to be integrated. They permit adaptation to various sectors and greater granularity within these fields of application.

Methods and business models

These can basically be developed from scratch in a top-down approach, or else they can emerge on a bottom-up basis via new infrastructures and data market places which are implemented via model smart data projects in the fields of industry, mobility, energy and health.

2 Smart Data Revenue Models

Dr. Eva Anderl, FELD M GmbH, ExCELL
James Löll, data experts GmbH, SAHRA

In addition to the technical innovations themselves, revenue models are an essential component of successful business models: the specific revenue mechanism ultimately determines how revenue is earned. The types and sources of revenue contrast with the cost structure on the other side of the balance sheet. The fundamental tension between revenues and costs is related to certain special features and challenges in the case of business models for smart data.

When identifying sources of revenue for smart data services, the main challenges are caused by the fact that

- in addition to the traditional scenarios (customer and supplier), several parties are often involved (multisided markets),
- end-users do not always pay directly for the service, and
- the benefit for the end-user is not always expressed directly in terms of sales or profits, but can lie in non-monetary incentives such as reduced risk or improved productivity or efficiency.

2.1 Multisided markets

In many cases, smart data solutions address different target groups simultaneously, e.g. private users and companies. One example: health insurance funds provide anonymised information about their insurees for regional and supra-regional analysis. The more participants from the healthcare system that provide their data, the more attractive it is for research establishments to use the data for various studies. Also, the completeness of the data makes it more attractive for doctors or other solution providers to draw on these data when treating patients.

If a smart data solution is to succeed, it is crucial to identify the various market participants and to rapidly integrate them into the creation of the solution. Here, the question of “volume” plays a central role in creating the added value for the paying user.

2.2 Direct and indirect revenue sources

Due to the positioning on multisided markets, it is necessary to distinguish between direct and indirect revenue sources when developing revenue models for smart data services.

Direct revenue sources

In the case of direct revenue sources, the user either pays directly for the use (service) or the provider integrates smart data technologies as part of a holistic solution. Instead of isolated hardware, a combined service is sold.

An example of direct use is the contracting of a smart data provider for research studies or market analyses. The provider is contracted to undertake an analysis of the available data on the basis of a hypothesis, and is paid directly for this service. A research institute could also be given temporary access to the data in order to carry out research and analyses as part of a study.

Furthermore, it is possible for providers of smart data technologies to transform markets. For example, a manufacturer of production facilities could alter its product range by using smart data to move from traditional hardware sales to service provision. These services could be transaction-based or use-based, and supplemented by further services like systems or process monitoring. The process monitoring enables the service provider to proactively plan the interruption of production processes with the customer in the event of potential problems. This enables the customer to reduce unexpected production losses which can cause more work and thus higher production costs.

New, integrated services could also be provided e.g. for crisis or epidemic management, contingency planning, risk assessment in real estate trading, or city route planning

Indirect revenue sources

Some smart data solutions require a certain quantity of data and information if they are to generate added value for the customers. The problem is that collecting the data is expensive if it requires the creation of a dedicated technical infrastructure (for example) or if data are not publicly available. When addressing this problem, it is possible to distinguish between paying and non-paying users. Free applications serve to reach as many users as possible. In return for using the solution free of charge, the users automatically provide important data. This data can then be used to charge commercial customers for a service. The economic viability of the solution depends on the paying customers helping to finance the non-paying customers.

In this context, the declining willingness of private and commercial customers to pay for services is a problem. Providers like Apple gave their customers the feeling that the software comes free of charge when they launched the iPhone with its inclusive (and regularly updated) software package.[10] This is possible due to cross-financing, e.g. from hardware sales and the App Store. But the lack of willingness to pay creates the challenge for providers of smart data solutions of financing a free-of-charge solution via various add-ons. In some sectors and solutions, it is possible to include advertising as an additional revenue source.

2.3 Highlighting non-monetary incentives

The advantages offered by smart data solutions cannot always be expressed directly in terms of sales or profits. In many instances, smart data solutions help companies to boost their productivity or to reduce the risks posed by investments or decisions. The products of smart data solutions can help them to notice and respond quickly to potential problems. Not least, smart data solutions can help to avoid production failures, to implement process changes efficiently, and to make data more easily accessible in the interest of better and quicker decisions.

For example, the statutory health insurance funds could invest in the establishment of an eHealth smart data solution and grant doctors, clinics and perhaps patients the right to use it. Whilst the solution is financed by the health insurance funds, the participating parties would boost added value for everyone by providing data quickly, even if this added value cannot necessarily be directly quantified. In this case, the added value for doctors and clinics is to be found in the support for patient care, e.g. in the form of tailored treatment indications or recommendations, or information about the regional spread of disease. The added value for the patient takes the form of improved healthcare, fewer misdiagnoses and possibly a more rapid identification of serious illnesses. Finally, the health insurance funds can benefit since early identification and treatment of illnesses is much cheaper for them.

Costs

On the cost side, providers of smart data business models are also facing challenges. It is very difficult for many companies to assess how much it will cost them to deliver their service in future. Firstly, it is difficult to predict how the costs of data storage and processing will develop. Whilst recent decades were characterised by “Moore’s Law”, according to which the complexity of integrated circuits with minimal component costs regularly doubles, there is increasing uncertainty about how long this development can be continued.[11] Secondly, not even the European Union has a common legal framework on data protection at present, and this represents a substantial problem for the scaling of smart data business models. The European General Data Protection Regulation could provide a reliable basis for planning, but would ultimately only represent one more step forward in the global data ecosystem.

At the same time - particularly on multisided markets with overarching provider and user structures - there is a high level of dependency on data suppliers. If data

are not obtained directly in-house, but are obtained together with other companies or provided by third parties, changes in the conditions governing the collaboration can represent a great risk for smart data business models. If there are changes to the monetisation schemes for supplying data, substantial additional costs can accrue. If data sources are entirely lost, other sources need to be developed wherever possible, and this can result in additional costs both for providing and for integrating data. If data are unique and cannot be substituted, this can represent a risk to the viability of smart data business models. In order to minimise such risks, questions about data availability, and particularly about remuneration models, need to clarified at an early stage and on a long-term basis (this is discussed further in the following chapter).

Revenue models are crucial for the success of smart data. The value of smart data derives not from the technology itself, but from the successful integration into an innovative business model. Experience shows that the technology that wins on the market is frequently not the best one, but rather the one which is integrated into a more accepted business model.[12] For this reason, providers of smart data technologies should consider potential revenue models at an early stage so that they can market their innovations well.

3 Barriers and Challenges

Dr. Martin Voigt, Ontos GmbH, SAKE
Dr. Sonja Zillner, Siemens AG, KDI

The successful development of smart data business models depends on four aspects in particular:

- Business value: How can actual revenue be generated from data (and the underlying smart data technologies)?
- Data access: What data sources need to be processed and combined in order to generate value?
- Skills: Specific skills are required from people able to evaluate and interpret data. How can these domain-specific skills be fostered in the field of big data?
- Public acceptance: How can public concerns be alleviated and public acceptance and incentives to release data be achieved for data-driven services and products?

3.1 Business value

The smart data business is virgin territory for most companies. As they start out, they are often unclear about what data are actually available at what quality, what conditions attach to the use of the data, what new knowledge can derive from the smart processing of the data, and how (commercial) value can be generated from this.

But smart data business models are based on the value of the underlying data and the potential findings that can be derived from them. As long as the value and quality of the data is unknown, the business potential will also tend to be uncertain.

As stated in Chapter 2, the data-based business models already established on the markets reflect a change in the underlying logic governing how money is earned. For example, traditional one-to-one transactions are supplanted by business models on multi-sided markets in which the consumer of the service no longer pays for the service, but rather himself becomes part of the product.

The value of the smart data application rises in relation to the number of different data sources, which can be integrated and interrelated across company or sectoral boundaries and thus realised in comprehensive data ecosystems. For this reason, the dynamism of the underlying network has to play a central role in the development of new business models.

3.2 Data access

Before data have been processed, their value and quality are rarely known. A systematic analysis and pre-processing of the data - which in turn frequently necessitates a lot of investment in time and money - is required to reveal what (business) potential rests in the data and what additional data sources need to be tapped.

Even though it may initially appear contradictory in the context of big data, the availability and the selection of data is a major problem in smart data business models. Whilst the volume of data keeps rising around the world, only a small portion of the data is actually accessible or available for immediate use.

Before raw data can be used in smart data applications, they often need to be laboriously structured, enhanced and processed. These pre-processing steps are needed in order to make the data available both technically and in the right quality for further analysis. But the steps are also labour- and cost-intensive.

Technical availability of data

The availability of big data also creates a variety of problems on the technical side. Data need to be digitised, and their seamless technical integration needs to be made possible. In view of the quantity, formats and speed, the data can only be used via programming, smart user interfaces, special frameworks like Apache Spark or Flink and adequate hardware resources. In the case of unstructured data, e.g. images and texts, domain-specific algorithms need to be developed so that the information contained in the data can be ex-

tracted in a processable format. People without programming skills but with a knowledge of the domain can use special tools to gain access. But this still requires a certain degree of analytical expertise, making it necessary to train staff (more on this in Chapter 3.3).

Quality of data sets

If data are to be used for targeted analyses or new services, they must be accurate in terms both of syntax and, particularly, of content. Even if this issue has been the subject of research for some decades, big data is introducing new dimensions and thus new challenges. The volume of data is so great that it is difficult to assess the quality at short notice. At the same time, the data are subject to rapid change, and often have a short lifetime, which necessitates a rapid evaluation. The high diversity results from data sources which may be new, unstructured or semi-structured, the quality of which is often difficult to assess. These technical problems result in a great need for research and development work in order to create methodologies and tools to enable even non-experts to assess the data quality quickly enough. But there is also a lack of standards which offer criteria and benchmarks for the assessment of data quality. One such standard is ISO 8000,[13] which aims to distinguish between companies and software in terms of the quality of the data supplied, but which has yet to be finalised and issued.

Access to open data sets

A single set of data rarely generates added value on its own. When it comes to enhancing data sets, an important role is played not only by in-house data, but also by open, freely available data sets, e.g. weather data,[14] local public transport,[15] map material[16] or structured Wikipedia data[17]. It is true that an increasing volume of public-sector data is being published (Germany now ranks 9th in the Open Data Index,[18] which assesses how countries make public data accessible), but this is chiefly true of data at federal level. The situation is not so simple at the level of the Länder, districts or municipalities.[19] On the one hand, there is often a lack of awareness or acceptance that open data adds value and generates commercial potential. On the other, there is a lack of money and technical skills to publish the data. Also, the federal system impedes the establishment of uniform legislation on the obligation to publish data (cf. Chapter 4 for details of the use of open government data).

Another issue is the provision of data from research and development. Projects in natural, biological and medical sciences which are supported by public-sector grants can easily be encouraged to publish not only the findings (articles in journals and conference documents) but also the data collected. In principle, this permits third parties to continue re-using them in other contexts and to create added value. In addition to the question of standardised access, possibly via a nation-wide portal, other organisational and technical issues also play a role.

In various situations, businesses can also benefit from making parts of their data publicly available. This can result in new ways to use the data and, subsequently, new marketing opportunities. Innovations can also occur in domains where this was not previously expected. However, there is a lack of clear licensing rules for the data and for the findings derived from them. Also, it is desirable to provide outside stimulation, e.g. via public funding, in order to encourage companies to publish their data, with a view to hopefully inspiring others to follow suit.

Lack of data sets that are closed but can be purchased

Not all data which can be used to create commercial or social added value will necessarily be open and freely available. For example, anonymised floating point data from an operator of navigation equipment could be used to extrapolate traffic volumes, which could in turn be used for logistics scenarios or infrastructure planning. This permits entire sales chains

of data to be established. However, one barrier is the lack of a (German or European) market place for the secure buying and selling of data on the basis of clear licensing provisions. This should not be left in the hands of large U.S. companies, as is the case with the Microsoft Azure Marketplace[20]. Further problems are the lack of awareness in companies about the market value of their data, licensing questions, data protection and anonymisation.

Data ownership

In addition to the question of what can be done technically with the data, there is the question of what can be done legally: Who may use the data, under what conditions and for what purposes? Data ownership is a concept which stipulates who may access and use what data and in what manner. When access and exploitation rights are stipulated, a distinction is made between various categories of data (private/public; personal/non-personal, etc.).

Data protection

The implementation of the right to data privacy covers all the measures to protect personal data against misuse in their storage, transfer, alteration and deletion.[21] This means that data protection is a key priority, not least in terms of gaining the trust of potential users.

However, the implementation of data protection legislation is frequently a very laborious process. One reason is that Germany's various Länder have differing rules and regulations. This means a lot of work for providers of smart data-based solutions, since the solutions have to be adapted to fit the various rules. Harmonisation of the data protection rules at national and at EU level is therefore a key precondition for the promotion of data-based business models.

3.3 Staff training

There is a high level of demand for trained staff on the IT market.[22] The growing demand for smart data solutions is leading to new jobs being created in this sector, but these are proving difficult to fill.[23] The problem here is that this is a very young field of computer science, and the appropriate courses of vocational training and higher education still need to be designed or adapted. This is chiefly a matter for the higher education institutions. If skills shortages are to be tackled in the near future at all levels, from developers to decision-makers, research-related institutions like the Fraunhofer IAIS Institute will need to hold more workshops and seminars.[24] In order to reach a wide audience, thought could also be given to organising and funding online courses (e.g. MOOCs[25]).

3.4 Public acceptance - creating incentives to use smart data

As is the case with all advances in technology, public acceptance plays a central role in the context of smart data, and is one of the most important preconditions for success with smart data business models. Big data technologies open up potential for strong social progress visible to every individual. Examples include smart internet searches, which can now be enhanced with additional information (e.g. Google's Knowledge Graph[26]), or start-ups like Uber[27] or AirBnB,[28] which can be called part of the "sharing economy". People are less aware of the use of the technologies for research into areas like cancer.[29] There is need for a smart incentive system for all of these solutions, since additional benefits can only be generated when the information is provided and integrated.

However, when it comes to big data, the public debate is mainly - rightly - focused on public fears about potential disadvantages from big data for individuals or society in general. For example, data theft by hackers or secret services can result in the targeted observation of individuals. This can result in great risks not only, but particularly for social minorities, journalists, or human rights activists. A problem is also posed by the acquisition of knowledge about individuals from

analyses which the people involved do not wish to know or share, e.g. predicting future illnesses or when people will die. Of course, the digital recognition of disease patterns, e.g. by Google Flu Trends, represents a major technological advance.[30] But excessive trust in such services can also have serious consequences, e.g. in the case of erroneous forecasts.

In the case of all of the problems described here, it is clear that the use of big data always needs to observe ethical principles. A start here can be provided by the "Ten Commandments of Computer Ethics"[31], but more is needed.[32] It is therefore necessary for clear ethical principles to be formulated and communicated at European level too. Should companies and society not automatically adopt such principles on a voluntary basis, they could then be cast in legislative form.

Public acceptance can ultimately only be improved by thorough information campaigns which move the debate away from problems and fears, towards the added value and opportunities of smart data, naturally whilst observing individual interests.

4 Open Government Data

Ingo Schwarzer, DB Systel GmbH, SD4M

Apart from the use of private-sector data, innumerable data sets are in the possession of public institutions or in companies and bodies mainly owned by the federation, Länder and municipalities. Companies working on behalf of the public sector also administer large stocks of potentially beneficial data. Initiatives from recent years to publish these data are generally dubbed "open data" (or, more specifically, "open government data").

The public debate in Germany on this issue is dominated by political aspects like boosting transparency, greater control, and thus ultimately the legitimation of measures taken by the state and its agencies. In contrast, the economic dimension of open data still plays a subordinate role. This is despite the fact that these data are a significant production factor in an ever-expanding information and knowledge economy. Back in 2011, a study on behalf of the European Commission put the economic potential of improved access to and use of government data within the European Union at €40 billion.[33] Also, the potential commercial uses of open (government) data vary widely. In some cases only a fraction of the uses or the scope of use is understood, but the data contain a great deal of potential to boost Germany's economy.

Existing business models can be evolved and improved, and new products or services can be designed and realised. This affects all areas of social life - potential exists e.g. in the fields of healthcare, infrastructure, transport, logistics, energy and industrial production. In particular, services can be provided as a sort of basic infrastructure for general use.

4.1 Open government data: definition

Basically, open government data is understood to mean all data stocks which may be freely used, re-used and passed on, either free of charge or only at the cost of provision.[34] If these data are held by governmental and administrative institutions, they are called "open government data".

The openness of data refers to various aspects, and authorities pay varying degrees of attention to these in the practical implementation. For example, the German Federal Government defines open (government) data as the "practice of providing machine-readable data by governments and administrations [...] for re-use by third parties."[35] The stress here is on machine-readability, which is intended to prevent technical restrictions due to the provision of formats like scans or pdf files. In the past, such formats have made it more difficult to re-process the data.[36]

The definition embraces not only the re-use by third parties but also the way the data may be used. This also covers the social and economic entitlement to use the information via a licensing system that imposes as few restrictions as possible.

4.2 Relationship between open and smart data

Big data methods and technologies have created ways for civil society and companies to collect, process and evaluate data which appeared impossible just a few years ago. Developments like social media, increasingly automated sensors, and cloud computing have contributed to an enormous rise in the global volume of data. If the data are put into a beneficial, high-grade and secure state, they are called smart data.

Government data also represent a part of this volume of data, and it is now easier than ever before to network them with other data and to assess and interpret them. But this raises questions about access to and use of these government data. This was doubtlessly one reason why the EU has stated in its Digital Agenda that the re-use of public-sector data and information is a key precondition for the successful development of the European Digital Single Market.[37]

In particular, a comprehensible classification of the data is required in order to overcome prejudices and reluctance to release the data.

4.3 Current situation in Germany and Europe

At European level, a study by the European Commission from 2011 shows that EU businesses and individuals do not believe they have sufficient access to information from public institutions. The reasons given were excessive fees for using the data and incomprehensible rules on re-using the data. Complaints were also made regarding things like lack of transparency about the identity of the possessors of data and the issuing of exclusive exploitation licences which place other competitors at a disadvantage.[38]

In response, the European Commission's open data strategy in 2011 defined measures and goals to overcome Europe's failure to promote open government data. For example, the Commission aims to play a pioneering role and now provides its information to the public free of charge via a new data portal[39]. Also, the Commission aims to establish a level playing field for open data across the EU.

In the European comparison, Germany is generally said to be lagging behind on the provision of open government data. For this reason, representatives of government, commerce, science and administration agreed at the 5th National IT Summit in Dresden in December 2010 to establish a central open data platform[40] by 2013 which would interlink the platforms of the Federal Government, the Länder and the municipalities and give users access to the data in the hands of the public administration.[41]

In order to assess the current state of legislation regarding open government data in Germany and Europe, it is necessary to consider the implementation of the PSI Directive of the European Parliament and of the Council from 2003.[42] This was implemented in Germany by the Act on the Re-use of Information (2006). The aim of the Directive is to ensure possibilities to re-use public-sector data. The PSI Directive was revised in 2013 in the light of new developments in technology, and the German Act was updated correspondingly in July 2015.

The main changes referred to the provision of data in machine-readable formats in order to make them easier to use. Also, there is now a basic right to re-use data from the public sector. Further to this, the fees for the re-use of the data were restricted to the costs of reproduction, provision and transfer.[43] Whilst the revision of the Act does send positive signals to German firms, critics do not believe that the changes go far enough. For example, the IT business association BITKOM complains that the rules on the prohibition of exclusive agreements grant existing agreements too long a transitional period up to 2043. Also, the exemptions from the prohibition are not clearly defined, and thus weaken the rules.

It is true that the public sector itself is paying increasing attention to the potential of open data for business models in Germany, but there are no signs that it will make full use of it. In view of the massive possibilities for structured and non-discriminatory access to public-sector data for start-ups, SMEs and large firms, the efforts by the public sector to foster open data in Germany should therefore be continued vigorously.

Outlook

The use and commercial deployment of smart data innovations have long since become part of an intensive public debate in Germany focusing not only on technical, but also and particularly on legal, economic and moral/ethical arguments. As the first publication by the working group on economic potential and social acceptance, this position paper marks the launch of our substantive consideration of urgent issues relating to the economic and social handling of smart data. The technology programme “Smart Data - Data Innovations” will now continue and deepen the debates raised in the position paper.

One central substantive priority for the working group will be the specific arrangements and challenges of smart data business models in the B2B sector. Many of the questions and problems described here are to be found in a heightened form in this sector. Especially when data owners are not the direct beneficiaries of smart data services or products, the realisation of incentive systems and the persuasion of all the stakeholders involved in the transaction represents a particular challenge which requires special consideration.

The next meeting of the group in Berlin on 26 January 2016 will therefore focus on this aspect. People outside the programme who are interested in the accompanying research on smart data can register and participate in the meeting and the work of the group. The necessary information and further assistance can be obtained from kontakt@smart-data-programm.de.

References

1 Stähler, P. (2002): Geschäftsmodelle in der digitalen Ökonomie: Merkmale, Strategien und Auswirkungen, Electronic Commerce, vol. 7, pp. 41f.

2 http://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/SmartDataProjekte/smart_data_projekt-gesundheit_kdi.html and http://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/SmartDataProjekte/smart_data_projekt-gesundheit_sahra.html

3 http://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/SmartDataProjekte/smart_data_projekt-gesundheit_innoplan.html

4 http://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/SmartDataProjekte/smart_data_projekt-energie_smartenergyhub.html

5 http://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/SmartDataProjekte/smart_data_projekt-energie_smartregio.html

6 http://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/SmartDataProjekte/smart_data_projekt-mobilitaet_sd4m.html and http://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/SmartDataProjekte/smart_data_projekt-mobilitaet_excell.html

7 http://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/SmartDataProjekte/smart_data_projekt-mobilitaet-itesa.html and http://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/SmartDataProjekte/smart_data_projekt-mobilitaet_sd-karma.html

8 http://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/SmartDataProjekte/smart_data_projekt-industrie_pro-opt.html and http://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/SmartDataProjekte/smart_data_projekt-industrie_smart-data-web.html

9 http://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/SmartDataProjekte/smart_data_projekt-industrie_sake.html and http://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/SmartDataProjekte/smart_data_projekt-industrie_sidap.html

10 http://www.horizont.net/medien/nachrichten/Umfrage-Zahlungsbereitschaft-fuer-Online-Inhalte-ist-nach-wie-vor-gering-134479

11 http://www.zeit.de/2005/16/Moore_s_Gesetz

12 Chesbrough, H. (2007): Business model innovation: It's not just about technology anymore. In: Strategy & Leadership 35 (6), pp. 12–17.

13 http://www.iso.org/iso/catalogue_detail.htm?csnumber=50798

14 http://www.dwd.de/cdc

15 http://daten.berlin.de/kategorie/verkehr

16 https://www.openstreetmap.org

17 http://dbpedia.org/about

18 http://index.okfn.org/place/

19 http://www.kommune21.de/meldung_21140_Kommunale+Krux.html

20 https://datamarket.azure.com/home

[21] http://www.olev.de/it/a-k.htm

[22] http://www.roberthalf.de/id/PR-04129/IT-Trends-2020

[23] http://www.forbes.com/sites/louiscolumbus/2014/12/29/where-big-data-jobs-will-be-in-2015/

[24] http://www.iais.fraunhofer.de/big_data_consulting.html

[25] http://www.forbes.com/sites/jamesmarshallcrotty/2012/05/06/mitx-harvardx-edx/

[26] https://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html

[27] https://www.uber.com/de/

[28] https://www.airbnb.de/

[29] http://www.forbes.com/sites/bernardmarr/2015/06/28/how-big-data-is-transforming-the-fight-against-cancer/

[30] http://www.nejm.org/doi/full/10.1056/NEJMp0900702

[31] https://en.wikipedia.org/wiki/Ten_Commandments_of_Computer_Ethics

[32] http://bdes.datasociety.net/council-output/ethics-codes-history-context-and-challenges/

[33] Vickery, G. (2011): Review of Recent Studies on PSI Re-use and Related Market Developments, p. 3.

[34] http://europa.eu/rapid/press-release_IP-11-1524_de.htm

[35] http://www.bundesregierung.de/Content/DE/Artikel/2015/02/2015-02-11-kabinett-informationsweiterverwendungsgesetz.html

[36] http://www.heise.de/newsticker/meldung/Bundestag-ebnet-Weg-fuer-Verwendung-oeffentlicher-Informationen-2638583.html

[37] http://europa.eu/rapid/press-release_IP-11-1524_de.htm sowie http://ec.europa.eu/digital-agenda/en/open-government

[38] http://europa.eu/rapid/press-release_IP-11-1524_de.htm

[39] https://open-data.europa.eu/de/data

[40] https://www.govdata.de/ zuvor http://daten-deutschland.de/

[41] http://www.bmi.bund.de/SharedDocs/Downloads/DE/Themen/OED_Verwaltung/ModerneVerwaltung/opengovernment.pdf?__blob=publicationFile

[42] Re-use of Public Sector Information Directive (2003/98/EC, 2013/37/EC)

[43] http://www.gesetze-im-internet.de/iwg/BJNR291300006.html

Paper Contribution
„Smart Data Business Models“

Authors:
Dr. Eva Anderl
FELD M GmbH, ExCELL

James Löll
data experts GmbH, SAHRA

Dr. Martin Memmel
DFKI, SmartRegio

Ingo Schwarzer
DB Systel GmbH, SD4M

Dr. Martin Voigt
Ontos GmbH, SAKE

Dr. Sonja Zillner
Siemens AG, KDI

Contributors:
Dr. Simon Becker
DSA Daten- und Systemtechnik GmbH, PRO-OPT

Uwe Gabriel
travelbasys GmbH & Co. KG, iTESA

Dr. Christian Geiß
DLR, sd-kama

Dr. Norbert Hansen
Karl Storz GmbH & Co. KG, InnOPlan

Astrid Hellmanns
FZI Research Center for Information Technology, Smart Data Accompanying Research

Dr. Holmer Hemsen
DFKI, SD4M, Smart Data Web

Johanna Hillenbrand
FELD M GmbH, ExCELL

Dr. Stefan Jäger
geomer GmbH, sd-kama

Thomas Meiers
Fraunhofer Institut für Nachrichtentechnik – Heinrich-Hertz-Institut, sd-kama

Marko Meißner
inquence GmbH, iTESA

Dr. Dorothea Pantförder
Technische Universität München, SIDAP

Eyk Pfeiffer
travelbasys GmbH & Co. KG, iTESA

Nico Rödder
FZI Research Center for Information Technology, Smart Data Accompanying Research

Dr. Martin Sedlmayr
Friedrich-Alexander-Universität Erlangen-Nürnberg, KDI

Björn Stecher
Initiative D21

Prof. Dr. Volker Tresp
Siemens AG, KDI

Siegfried Wagner
in - integrierte informationssysteme GmbH, Smart Energy Hub

Prof. Dr. Christof Weinhardt
FZI Research Center for Information Technology, Smart Data Accompanying Research