

Fraunhofer FOKUS
On behalf of the Senate Department of
Economics, Technologies and Research in Berlin

Dr. Wolfgang Both (ed.)
Prof. Dr. Ina Schieferdecker (ed.)

Executive Summary

Berlin Open Data Strategy

Organisational, legal and technical aspects of Open Data in Berlin.

Concept, pilot system and recommendations for action.

BERLIN
OPEN
DATA

 **Fraunhofer**
FOKUS

 projektzukunft mit Berlin

ServiceStadt
 **Berlin**

Fraunhofer FOKUS
On behalf of the Senate Department of
Economics, Technologies and Research in Berlin

Dr. Wolfgang Both (ed.)
Prof. Dr. Ina Schieferdecker (ed.)

Executive Summary

Berlin Open Data Strategy

Organisational, legal and
technical aspects of Open Data in Berlin.

Concept, pilot system and recommendations for action

Authors:

Dr. Nils Barnickel	Fraunhofer Institute for Open Communication Systems (FOKUS)
Dr. Matthias Flügge	Fraunhofer FOKUS
Daniel Hanke	Fraunhofer FOKUS
Dr. Edzard Höfig	Fraunhofer FOKUS
Jens Klessmann	Fraunhofer FOKUS
Florian Marienfeld	Fraunhofer FOKUS
Prof. Dr. Ina Schieferdecker	Fraunhofer FOKUS
Jan Ziesing	Fraunhofer FOKUS

Additional Contributors:

Hartmut Bömermann	State Statistical Institute Berlin-Brandenburg
Dr. Wolfgang Both (Project supervisor)	Senate Department of Economics, Technologies and Research
Julia Franke	Intern at the Senate Department of Economics, Technologies and Research in Berlin
Olaf Franke	Senate Department for the Interior and Sports
Andreas Mängel	BerlinOnline Stadtportal GmbH & Co. KG

Berlin, January 2012

Contact:

Fraunhofer Institute for Open Communication Systems (FOKUS)
Kaiserin-Augusta-Allee 31, 10589 Berlin

Phone +49 30 3463 7200
Fax +49 30 3463 8000
E-Mail info@fokus.fraunhofer.de
URL www.fokus.fraunhofer.de

© by FRAUNHOFER VERLAG, 2012
Fraunhofer-Informationszentrum Raum und Bau IRB
Postfach 800469, 70504 Stuttgart
Nobelstraße 12, 70569 Stuttgart
Phone +49 711 970-2500
Fax +49 711 970-2508
E-Mail verlag@fraunhofer.de
URL <http://verlag.fraunhofer.de>

The content of this document is licensed under the CC BY license (Creative Commons Attribution 3.0 Germany). When attributing, you must use <http://s.fhg.de/BerlinerOpenDataStrategie> as source. The license can be found at <http://creativecommons.org/licenses/by/3.0/de>.

Summary

If nothing else, the unexpected success of the Pirate Party in the elections for the Berlin City Parliament and the District Parliament made it clear that demands for open governance and administration, access to databases and the active provision of government and administrative information beyond legal texts and regulations are not just coming from the fringes of society. Public petitions, citizen participation, civic protests and participatory budgets show that these topics have reached the heart of society. The fact that these expectations are acceptable to the majority was illustrated in an SAS survey¹ in the summer of 2010 in which 88 % of the citizens interviewed were strongly in favor of non-personal information being made public by the public authorities and 81 % of them considered this to be a chance for more participation.

All democratic parties included the topics transparency and open government in their manifestos and the Berlin SPD/CDU government coalition, formed after the last election, announced the continuance and further development of the Open Data initiative of the state in their coalition agreement.

The Berlin Open Data Strategy

In this document, the Berlin Open Data Strategy has been formulated based on preparatory work, in light of a series of interviews with actors in the State of Berlin and as a further development of existing Open Data concepts. It includes the following elements:

- **Status and analysis of the potentials:** The current status and the potentials of Open Data in the state of Berlin are analyzed for the basis of the strategy. A scientifically derived assessment of the status quo in Berlin and in comparison to other regional authorities in Germany and abroad is presented.
- **Organizational analysis:** Selected organizational processes in the administration regarding the handling of public data are examined. Furthermore, the necessary changes considering the organizational conditions, the quality control, possible additional expenses and the training of employees are demonstrated. In addition, organizational barriers and risks that could be associated with the implementation of a strategy for the release of public data are identified.
- **Architectural design of an Open Data platform:** Architectural recommendations and the technical design of the Berlin Open Data platform are developed. An high-level specification for coordinating with possible data providers, processors, and consumers is given, and was used as the basis for the pilot phase of the Berlin Open Data portal.

¹<http://www.sas.de/open>.

-
- Analysis of possible business models: Different operational concepts are defined and evaluated. In addition, possibilities for the financing of Open Data that could bring additional revenues and funding for the further development and operation of the Berlin Open Data platform are assessed, facilitating the sustainable extension of the Berlin Open Data offerings.
 - General legal conditions: Based on a thorough legal assessment of the relevant laws, a legal framework for Open Data in Berlin is defined and a current overview of the legal status is given. Suitable solutions and amendments are proposed, enabling the legal implementation of Open Data in Berlin.
 - Open Data pilot system with pilot data: Based on the architectural recommendations for a Berlin Open Data platform, a data portal for Berlin was established that already confirms the relevance and potentials of Open Data for the city, despite its short time of operation. Practical experiences in the provision and processing of public data are collected and demonstrated.
 - Recommendations for action and a step-by-step plan: The recommendations for action summarize the results of the project in the form of political, organizational, legal and technical recommendations for action and define a step-by-step plan for the further expansion of the provided data and data-based services in Berlin.

The Berlin Data portal

In line with the project, a pilot system was developed for the Berlin data portal in collaboration with the Senate Department for the Interior and Sports, the State Statistical Institute Berlin-Brandenburg and the BerlinOnline Stadtportal GmbH. The data portal acts as the central contact point for the data of the State of Berlin. The collected data sets are machine readable and freely licensed, and therefore enable greater transparency of administrative procedures and provide, for example, information about infrastructural and environmental aspects in the city and encourage further use and processing by third parties.

The portal is technically split: The surface is formed by a Drupal content management system and is implemented and operated by the BerlinOnline Stadtportal GmbH. The Comprehensive Knowledge Archive Network (CKAN) is used as data catalog and forms the backend. It provides access to the metadata about the data sets. The Berlin CKAN was designed and set up by Fraunhofer FOKUS and is for the time being also maintained by FOKUS.

The separate backend offers various possibilities, the most important of which is the federation. In regards to Open Data this means: The Berlin Comprehensive Knowledge Archive Network (CKAN) can on the one hand concentrate the contents of subordinate instances, for example CKAN data portals of the districts or even those that are used in the individual institutions. In addition, Fraunhofer FOKUS can provide the federal

government with access to Berlin's data through CKAN as soon as it starts to operate a comparable portal. Already today, the data can also be published at European level, for example on publicdata.eu. In order for this federation to function properly, an elaborated metadata scheme is necessary. That means that all involved parties have to maintain the same information about data sets with compatible formats. Therefore, Fraunhofer FOKUS developed a metadata scheme based on the experiences with existing portals of Great Britain as well as of the cities of London and Vienna and a survey conducted in European cities that will also be employed in the metropolises Barcelona, Amsterdam and Paris in line with the European project "Open Cities".

The portal currently offers access to about 60 data sets in 15 different data categories not only of public authorities but already also of other organizations. The users of <http://daten.berlin.de> may use the data for any purpose as long as it is not in violation of the terms of use or the rights of third parties. Wherever possible, data sets are covered by the license "Creative Commons Attribution 3.0 Germany" (CC BY). In addition, applications are available that build on these data sets and evaluate and/or visualize them. The strengths of the approach to make data openly available in a city are especially reflected by the applications. Shortly after opening the portal, applications were uploaded that provide new insights – for example about the current level of ozone pollution in Berlin. The Berlin data portal was equipped with a feedback option in order to enable tips, suggestions and special requests in regards to data and applications, to enable the interaction with users in general and to discuss possible further developments of the portal.

For administrative institutions, information events and software tools are offered so that their data sets can be added to the data portal in the future as well. Contributions from industry, science and civil society are welcomed and can be published using a special interface of the data registry. The pilot phase runs until the end of 2011 and after that preparations for additional development and normal operation are set to begin.

Recommendations for action

The strategic recommendations for action in regard to the Open Data in Berlin include the following areas:

- The political commitment to Open Data in and from Berlin and to the Berlin data portal. It is necessary to appoint an overall responsible person for Open Data and respective contact persons for the data. The data, which is provided through the expansion of the data portal, has to be continuously available in the long-term.
- Organizationally it is necessary to develop guidelines for choosing the data as well as defined processes for the data portal. In addition, training initiatives for data processors, data providers and editors have to be developed. The user community has to be involved through various possibilities for interaction that also include

the possibility for suggestions regarding databases, applications and tools. Data published by the Senate of Berlin should be automatically added to the data portal.

- The legal recommendations for action include uniform and simple licensing terms and terms of use as well as the direct access to these regulations in order to minimize barriers for the use of the data. The traceability and comprehensibility of the data is essential for its use.
- Technically it is recommended to set clear standards for the data portal itself and its security. Compatibility, integrability and quality control of the data are thereby especially important.

Step-by-step plan

Based on the recommendations for action and the suggestions for further development of the Berlin data portal, a step-by-step plan for the Berlin Open Data Strategy is developed that includes short-term, medium-term and long-term measures. In the short-term, Open Data has to be incorporated into the administrative regulations of the State of Berlin and the Berlin data portal has to be transitioned into regular operation. In the medium term, the sustainable development of the data services of the State of Berlin and their supplementation with tools, added-value services, training etc. have to be ensured. In the long-term, the nation-wide coordination and integration of the Berlin data services into services in Germany, in the German-speaking countries and in Europe have to be further developed.

We are sending a strong signal to administrations today. Your data is worth more if you give it away. So start releasing it now: use this framework to join the other smart leaders who are already gaining from embracing open data. Taxpayers have already paid for this information, the least we can do is give it back to those who want to use it in new ways that help people and create jobs and growth.

EU Commission vice president Neelie Kroes,
12 December 2011

1

Introduction

1.1 Background and motivation

With the increasing penetration of various areas of business and society with information technology, information as an good take center stage. Data and information are essential resources in an information society.

In line with the assignment “Open Data for Citizens, Business and Government” Fraunhofer FOKUS has developed concepts, an architecture and a pilot system for the open and reliable access to Berlin’s public data. Governments, companies, organisations as well as the citizens of the city thus receive information in order to increase transparency, to enable the collective design of municipal processes and procedures and to increase the economic attractiveness of the city. The concept that has been developed enables the efficient provision, analysis and processing of Open Data in services and applications through third parties. In doing so, transparency, participation and collaboration within the city are all strengthened.

The Berlin Open Data initiative is supported by two initiatives for making data of any kind available, Open Data in the USA, Great Britain and other countries and for making public data available, the EU guideline for the re-use of information from the public sector (Public Sector Information – PSI) and other regulations.

In the preliminary stages of these activities, Fraunhofer FOKUS has developed a preliminary study on the state of public data in Berlin in collaboration with Senate Department for Economics, Technology and Women’s Issues in Berlin. According to this study, a significant amount of Berlin-based data is available including search and filter functions,

however, this data presents mostly static information with limited possibilities for further processing. The majority of data is provided in Portable Document Format (PDF) and often presents previously accumulated, interpreted and evaluated data. Except for the statistical data and geo-data, there is a lack of raw data in Berlin which can be provided in well-defined formats and via open interfaces for further processing. In addition, restrictive licensing terms hinder the development of new business models based on Open Data in the long run.

Similar to the Global Positioning System (GPS), data that has been available for more than a decade, other data from public sources offer opportunities for successful economic use; this should be supported in the interest of the city.

1.2 Objective and approach

Berlin has been preparing to make its data available since the spring of 2010. Technical, legal and organisational questions were examined in line with a preliminary study by Fraunhofer FOKUS on behalf of the state government. In the fall of 2010, an online poll in regard to the most important and most interesting data categories provided more clarity about the interests of the citizens.

Therefore, a project “from a public to an open government” was able to be launched in the spring of 2011 by the State Secretary Committee for Administrative Modernization in line with the program “ServiceStadt Berlin”. The goal of the assignment “Open Data for Citizens, Business and Government” is to provide structured access (data catalogues) via a central access point (data portal) to the databases in the state that remain locally maintained.

The Senate Departments for Economics, Technologies and Women’s Issues and for the Interior and Sports and the State Statistical Institute Berlin-Brandenburg formed the inner project team. The study was conducted by Fraunhofer FOKUS. On the Berlin Open Data Day (BODDy) in May 2011, the initial results like the Berlin Open Data Agenda were presented. This forum also provided the opportunity to involve actors from civil society. Networks like the Government 2.0 Network and the Open Data Network as well as Wikimedia supported the activities and aim to use the databases for their own development. On the other hand, wishes were expressed that can be taken into account for development and testing. Through the involvement of the BerlinOnline Stadtportal GmbH as operator of the official capital city portal of the State of Berlin, Berlin.de, the development activities were accelerated as that company already had solutions for an internal data warehouse concept. Thus, Berlin was able to activate a data portal on September 14, 2011 as the first German city.

1.3 Results, recommendations for action and step-by-step plan

Fraunhofer FOKUS developed a concept and a procedure model for a Berlin Open Data platform in close cooperation with the principal and the actors in Berlin. Thereby, the status and potentials as well as the organisational, legal and technical framework for Open Data in Berlin were assessed and possibilities for financing Open Data in the city were evaluated. In addition, a pilot system for a Berlin data portal was developed in cooperation with the Senate Department for Economics, Technology and Women's Issues, the Senate Department for Interior and Sports, the State Statistical Institute Berlin-Brandenburg and BerlinOnline.

The data portal presents a central hub for data of the State of Berlin. The here collected data sets are machine-processable and are freely licensed. They thereby enable the greater transparency of administrative processes and provide for example information about the infrastructure and environmental aspects in the city and encourage re-use and processing by third parties.

In the conclusion of the study and in summary of the experiences gained from the pilot system, short, medium and long term recommendations for actions for the political, legal and technical handling of Open Data in Berlin are formulated for the next five years. A step-by-step plan provides recommendations for the next steps to stabilise and further develop the available data in Berlin.

2

Recommendations for action

Information is the resource of the 21st century. The open and structured access to freely available data sets of the public sector is an important contribution for the development of the knowledge society. It strengthens the trust between politics and civil society, between government, business and media. Government actions are becoming more transparent, comprehensible and verifiable. Furthermore, the opening of data leads to potential for more efficient administration and increased quality. In addition, the free access to and the proactive provision of information and data contribute considerably to the goal of strengthening the cooperation and solidarity of social groups. New solutions from the business sector can be developed and can contribute to the further development of the common good.

The following recommendations for action are grouped into political, organisational, legal and technical recommendations for action. They are the results from the study and the findings from the pilot phase of the Berlin data portal. In regard to chronological dimensions, the recommendations are sorted into short-term (S, ca. 1 year), short- to medium-term (SM), medium-term (M, ca. 2-3 years), medium – to long-term (ML) and long-term (L, ca. 5 years) recommendations for actions. In combination with the expansion of the Berlin data portal, the measures recommended here will lead to the final step-by-step plan for the next leg on the path to Open Data in the state of Berlin.

2.1 Political recommendations for action

2.1.1 Political commitment to Open Data in and from Berlin

The conflict of objectives between contributions to the budget reorganisation through the sale of data sets, the Berlin Freedom of Information Act and the disclosure pursuant the EU-PSI guidelines and the German Federal Act on the Re-Use of Public Sector Information (IWG) has to be resolved through a political commitment of the state of Berlin to Open Data and a legal basis for action. The allocation of Open Data to a department and the appointment of a person in charge of Open Data are thereby essential. The data processors in Berlin have to be provided with clear guidelines that make the identification and provision of Open Data possible.

→ **S1.** A clear political commitment to Open Data is necessary: A principle for the general publication of non-personal information and non-security-critical data in the public interest has to be anchored to the Berlin Freedom of Information Act.

→ **S2.** The political commitment to Open Data in and from Berlin forms the basis for concrete measures and therefore has to be implemented on very short notice for the promotion of the other recommendations for action.

→ **S3.** The financing of the Berlin data portal covering the operation of the portal, the provision of the data sets and the execution of processes relevant to Open Data in Berlin has to be ensured.

→ **S4.** Guidelines for the selection of data for the Berlin data portal according to their economic, security and privacy related eligibility have to be prepared and anchored to the governmental regulations. It is necessary to prepare a criteria catalogue for the selection of data for the Berlin data portal that characterises possible data sets according to their features, so that it is possible to assess data sets in regard to their suitability.

→ **S5.** The data preparers and editors of the data portal have to be provided with tools and training in order for high quality data to be provided in an efficient and prompt manner.

2.1.2 General information strategy

The Open Data Strategy of the state of Berlin has to be viewed in context with a general information strategy of the state of Berlin that has to be developed and that should be incorporated into the Open Data Strategy.

→ **M1.** The Open Data Strategy of the state of Berlin is an essential element of a general information strategy not yet developed and therefore should not be promoted individually but in the context of this general strategy.

2.1.3 Political commitment to the Berlin data portal

In addition to the study at hand, there already is a Berlin data portal for Open Data in and from Berlin that has been launched at <http://daten.berlin.de>. It contains more than 100 data sources from governmental authorities. The Berlin data portal served as an example for the two-stage call for proposals of the Federal Ministry of the Interior for Open Data in Germany and provides data sets as basis for the Apps4Deutschland competition, which started in October 2011.

→ **S6.** The transition towards regular operations of the Berlin data portal has to be assured. For this purpose, a department has to be appointed that will also perform the registration of employees for the data register and additionally, regular operations on the city portal have to be ensured on a long-term basis. State editors have to be assigned with the processing of the Berlin data as well as with the moderation on the data portal. An operational concept with defined roles has to be prepared.

2.1.4 Person in charge of Open Data and contact partner for the data

Responsibilities for the work in regard to Open Data in Berlin have to be arranged and responsible persons have to be appointed.

→ **S7.** In addition to the contact person for each data set, 1. a person in charge has to be appointed and 2. An interdepartmental workgroup has to be established for the further development of the Berlin data portal.

2.1.5 Stable and long-term provision of the data

The successful use of the services of the data portal by the businesses as well as the increased transparency of government activities are mainly based on a stable and long-term provision of the Berlin data. The solution has to be implemented in a reliable and high-quality manner in order to allow sustainable actions using this data, for example in economic, civic and journalistic areas.

→ **SM1.** The data has to be provided in open, machine-readable, well-defined and documented formats and data catalogues via a central and user-friendly Berlin data portal that complies with defined quality criteria for the available data.

2.1.6 Further development of the data portal

For the active provision of Open Government Data (aside from data that is subject to privacy and security restrictions) to anybody (citizen, businesses, media, government, etc.) for information purposes as well as an increasing transparency in the city that allow

a combination and integration with data from other sources, the following measures are recommended. Exceptions must be substantiated.

→ **SM2.** Increase of contents: Further machine-processable data sets from the government and other sources must be continuously added to the Berlin data portal.

→ **M2.** Organisational development: The necessary legal foundations and adaptations of the administrative regulations in Berlin have to be established.

→ **ML1.** Technical development: Further projects to the development of added-value services and tools for the provision of Open Data in Berlin should be carried out.

2.1.7 Federation of the Berlin data portal with other portals

Berlin is doing pioneering work with the Open Data concept and the data portal. Only a few European metropolises offer such a service. However, without joint agreements or even standards, the individual activities will lead to isolated solutions that will hinder the future interconnection for integrated services.

→ **ML2.** Berlin should demand at domestic and European committees that the pressing question about the standardisation of formats, metadata and licenses is put to a vote and should become actively involved in those processes.

2.2 Organisational recommendations for action

2.2.1 Integrability of the Open Data Strategy

The state of Berlin has already done essential preliminary work for the provision of administrative information like the data pool at the State Statistical Institute Berlin-Brandenburg or the geo-data infrastructure of the Senate Administration for Urban Development. The further development of the Berlin data portal should be coordinated with existing projects.

→ **SM3.** Further development of the Berlin data portal should be coordinated with existing projects for the provision of administrative information of the state of Berlin.

2.2.2 Guidelines for the selection of data

Difficulties arise in the institutions in regard to the selection of possible data sets for the Berlin data portal in line with the identification of the data within the limits of the legal framework. Even if non-personal and non-security-critical data is supposed to be published, the assessment of the situation might be difficult in a concrete case. Privacy issues are anticipated.

→ **S8.** It has to be clarified, which limits have to be observed in regard to data security, for example, regarding organisational information of the institutions like names, addresses or dates.

2.2.3 Well-defined processes for the data portal

Well-defined processes have to be described for the operation, development and maintenance of the Berlin data portal that define the responsibilities also in interaction with the users of the data portal.

→ **S9.** A responsible editor needs to be appointed who runs and coordinates the editorial department of the Berlin data portal.

→ **S10.** A responsible person must be selected amongst the editors of the Berlin data portal who is responsible for the interaction with the users and coordinates the reaction to user inquiries and recommendations in a timely manner.

→ **M3.** The editorial department of the Berlin data portal has to be supported by a reliable technical operation that not only pays attention to the reliability and accessibility of the service but also to the IT-security and the back-up and the archiving of the provided data.

2.2.4 Training measures concerning Open Data

According to the current assignment, institutions publish data for a specific purpose with appropriate restrictions and corresponding (isolated) structures. However, Open Data requires a paradigm shift in various areas: Anything that is not specifically marked as secret is public. All data that is not subject to warranted privacy or security restrictions is proactively and entirely published in a timely manner.

→ **M4.** To implement this paradigm shift, appropriate training measures have to be developed and offered, in order to convey the resulting advantages for the institutions and to enable organised acting in relation to Open Data.

2.2.5 Various interactions with the users of the Berlin data portal

The provided data in Berlin profits from the diverse interactions with the users, from civil society, the organisations and businesses of the city as well as from the government. For example, thoughts on the portal itself or about concrete or desired data can be expressed and commented on in an idea portal. The users can express on a wish list which data and which applications they desire. The user community should be able to vote for suggested data sets and to prioritise them. In forums, “communication threads” can be started for various topics. Thereby, users can help each other aside from the pure discussion and in such a way that a discussed solution will be available for uninvolved users afterwards.

Also and especially in the forum area, moderation is necessary. Here, editors can merge topics that appear twice or refer to similar topics. In addition, mechanisms for the removal of entries in violation of the terms of use have to be implemented.

→ **M5.** The existing possibilities for interaction like comments and feeds should be supplemented with further possibilities like an idea portal or a wish list.

2.2.6 Possibility for suggestions for data sets

Often, the already selected data sets like geo-data or data of the public transportation system are asked for by users of the Berlin data portal.

→ **M6.** The available data on the Berlin data portal has to be continuously assessed in regard to possible and necessary expansions and be upgraded accordingly.

→ **SM4.** Even today, a clear demand for a freely available spatial reference system for Berlin can be recognised. It has to be determined in what way such a system could be provided via the Berlin data portal.

2.2.7 Adding published data to the portal

The Berlin authorities and its institutions regularly publish comprehensive numerical data as well as data and facts in form of press releases, reports and compilations from the various operational areas. The responses to parliamentary questions often contain interesting data sets as well. The material is noticed, processed and discussed by media, businesses and citizens.

→ **S11.** Press releases, reports and other publications with numerical data have to be supplemented with machine-readable tables that are based on the original data and be made available with the notice or the report on the website of the relevant authority. The press offices are responsible for the entry into the data register of the Berlin data portal.

2.2.8 Adding resolutions and protocols to the data portal

Aside from raw data, resolutions and protocols of the Berlin government and the regional politics are relevant in line with Open Data and should be made available via the data portal.

→ **M7.** In the medium-term, all resolutions and protocols of senate, city council, parliament and committee meetings should be published in an open format like for example Office Open XML (OOXML) or ODF.

2.2.9 Adding data of state-owned companies to the data portal

The portal should not only provide raw data from the administrative institutions of the state of Berlin but also raw data from companies and organisations, especially the data of state-owned companies, foundations, etc.

→ **M8.** State-owned companies, foundations, etc. should add their raw data to the data portal in the medium-term.

2.3 Legal recommendations for action

2.3.1 Uniform and simple licensing terms and terms of use

The specifications for uniform and simple licensing terms and terms of use that are based on the open definition that support the further processing, use and distribution of the data of the state of Berlin by anybody and for any purpose, incl. commercial purposes are recommended. Minimal fees should only be charged in exceptional cases (for example for the technical provision).

→ **S12.** Data sets should be provided under the Creative Commons Attributions License (CC BY). Exceptions from this basic regulation should be explained.

→ **M9.** For data sets with limited commercial use, a licensing and a re-use proposal should be developed that for example allows a proportional cost-recovery in case of commercial use. This needs to be coordinated on both domestic and European levels.

2.3.2 Direct access to the licensing terms and the terms of use

In order to support the re-use of Open Data, it is essential to provide the appropriate rights of use in a simple manner.

→ **SM5.** The licensing terms and terms of use are provided in a highly visible way and are accompanied by an explanation. Inquiries from users are answered in a legally compliant and timely manner.

2.3.3 Traceability of the data sources

The provided data has to – if possible – be clearly labelled in regard to their authorship and in addition – if possible – in regard to the occurred aggregation of data through third parties.

→ **M10.** An attribution license for the data sets on the Berlin data portal is essential.

→ **ML3.** It has to be assessed if the traceability of processing steps can be provided for data and information that was derived from the raw data.

2.4 Technical recommendations for action

2.4.1 Standards and security of the Berlin data portal

Only the provision of Open Data in defined, standardised formats in line with various content criteria allow for the simple re-use in new applications that, for example, create visualisations or other data processing.

→ **M11.** Appropriate technical standards for Open Data have to be defined and classified according to standards and architectures for eGovernment applications (SAGA), if applicable.

→ **M12.** Relevant standards have to be specified and used in the long-term in order to ensure the necessary security of the Berlin data portal.

2.4.2 Compatibility and integrability of the data

The following measures are recommended in order to ensure an optimal compatibility and integrability with Open Data of other services through the use of international standards, for example of the World Wide Web Consortium (W3C), International Organization for Standardization (ISO) and Comité Européen de Normalisation (CEN).

→ **L1.** The Berlin data portal will only provide data in machine-processable formats that in addition comply with domestic, European or international standards.

→ **L2.** The available data of the Berlin data portal has to be enhanced with formats with higher semantic quality like the Simple Query Language (SQL), Resource Description Framework (RDF), etc. and supported with suitable tools for the review, analysis, assessment and presentation of semantic data.

→ **ML4.** Separate metadata schemes for the characterisation of the data sets and the characterisation for the applications for the Berlin data portal have to be defined. These have to be coordinated in Germany, in the German-speaking world and in Europe.

2.4.3 Quality management of the data

Measures for the quality management of the data and, if applicable, of the processed data, have to be established. The data quality should be assessable with the help of a documented method. Thereby, the traceability of the sources, the currentness, the correctness and precision of the raw data and of the processing steps should be assessed.

→ **ML5.** A quality management process has to be specified as part of the operational concept for the Berlin data portal and has to be applied consequently. The high quality of the available data has to be ensured continuously.

→ **M13.** The users of the Berlin data portal and its data sets have to be provided with various feedback options that for example enable a lively exchange in regard to current and future available data of the portal.

→ **L3.** An culture of fault tolerance has to be developed for the available data on the portal that allows for the possibility of errors and includes well-defined correction methods and correction regulations, amongst others.

2.4.4 Services and tools for the handling of the data

Services and tools have to be provided on the portal for a practical, efficient and barrier-free use of the data by the general public.

→ **ML6.** The available data on the portal has to be supplemented with services and tools for the aggregation, assessment and visualisation of the raw data as well as the data and information derived thereof.

2.5 Step-by-step plan

The step-by-step plan groups the recommendations for actions into phases and provides recommendations for the next steps to stabilise and further develop the available data in Berlin:

- Short-term (ca. 1 year): Anchoring of Open Data in laws and administrative regulations of the state of Berlin and transition of the Berlin data portal towards regular operations. Thereby it is especially important to establish regulatory, organisational, technical and legal frameworks. This step includes the short-term recommendations for action S1 to S12 and the short to medium term recommendations SM1 to SM5.

- Medium-term (ca. 2 to 3 years): Long-term development of the data of the state of Berlin and supplementation with tools, higher-quality services, training, etc. This step includes the medium-term recommendations for action M1 to M13 and the medium to long-term recommendations ML1 to ML6.
- Long-term (ca. 3 to 5 years): Coordination and integration of the available data in Berlin with other available data in Germany, in the German-speaking world and in Europe. This step includes the recommendations for action L1 to L3. It has to be taken into account that some of the recommendations for action fall into the transitional area between short- and medium-term or medium and long-term and that their time-frame for implementation can be adjusted accordingly, if necessary.

Abbreviations

BODDy	Berlin Open Data Day
CC BY	Creative Commons Attributions License
CEN	Comité Européen de Normalisation, European committee for standardization
CKAN	Comprehensive Knowledge Archive Network
GPS	Global Positioning System
ICT	Information and communication technology
ISO	International Organization for Standardization
IWG	Informationsweiterverwendungsgesetz, German Federal Act on the Re-Use of Public Sector Information
ODF	OASIS Open Document Format
OOXML	Office Open XML
PDF	Portable Document Format
PSI	Public Sector Information, EU-Directive to PSI
RDF	Resource Description Framework, popular format for Linked Data
SAGA	Standards and architectures for eGovernment applications
SQL	Simple Query Language for Databases
W3C	World Wide Web Consortium

Glossary

Distinction Open Government – eGovernment

Open Government includes any activity in public administrations that supports the transparency, participation and cooperation, no matter if it was carried out using modern Information and communication technology (ICT). Provided that Open Government is conducted using modern ICT, it is also a special form of eGovernment.

Blog, Weblog

Diary-like internet page; the entries (postings, posts) can usually be commented.

Forum

A platform (a »virtual place«), where users can exchange and archive knowledge, experiences, opinions and thoughts.

Government 2.0

The terms Open Government and Government 2.0 are often used synonymously. Government 2.0 is an umbrella term for the use of Web 2.0 technologies in the internal and external communication of governmental administrations. Government 2.0 is not necessarily connected to Open Government.

Costs/payment obligations

The term costs is a business variable; for this terminology, the umbrella term payment obligations will be used. Payment obligations for the use of data can accrue in form of fees (under public law) or charges (under private law). The re-use of data is especially supported if the user incurs no payment obligations. Exceptions can be caused by budgetary requirements. A preferably simple payment system has to be set-up.

Linked Data

The term Linked Data describes structured information that does not have to be interpreted in an isolated manner but that, where ever possible, refers to external generally recognised objects. Linked Data is a variety of the Semantic Web, which concedes that terms possibly have no universal meaning but a semantic context. Example: To label a field in a table with the string “district” would not be Linked Data and has a different meaning in, for example, Germany and Austria. If however a URL like [http://de.dbpedia.org/resource/Bezirk_\(Berlin\)](http://de.dbpedia.org/resource/Bezirk_(Berlin)) is selected, the meaning is unambiguous and we speak here of Linked Data.

Licensing

Terms of use (licenses) determine what kind of use is permissible under defined circumstances. Clear, easy to handle and uniform terms of use promote the re-use and distribution. These terms of use are independent of possible payment regulations. The standard licensing agreements of the Creative Commons can provide guidance.

Machine-readable

Basically, all data that can be interpreted by software is also machine-readable. In relation to Open Data this especially includes those data formats that allow re-use. The underlying data structure and the according standards must be accessible and should be completely published and be available free of charge.

Mashup

Simple web application that combines different data sources.

Microblogging

Blogging with short messages (for example with the service Twitter).

Ontology

A system of concepts, in which concepts and their relations are defined using a distinct vocabulary.

Open (Government) Data

Open (Government) Data describes the publically accessible provision of public sector data, usually in form of raw data for further use, especially through re-use and distribution. Excluded are personal data as well as data that are otherwise worthy of protection (safety-related data, company and business secrets). Open Data can be relevant for the three sections of Open Government. The provision of data can increase the transparency (for example the use of budget resources can be visualised), encourage participation (for example through an improved information base for participation in line with participatory budgeting) and enable new forms of cooperation (for example the development of apps based on geo-data). Open Data is especially relevant for the section of transparency – whereas transparency could also be possible without Open Data. The special added value of Open Data is the possibility to show new information contexts through the connection of machine-readable data. Therewith, Open Data also has potential in the areas of for example innovation promotion and business development.

Open Data criteria

The discussion about Open Data is based on the ten so called Open Data criteria¹.

- completeness
- primacy
- timeliness
- ease of access
- non-discrimination
- permanent availability
- machine-readability
- open standards/open formats
- free license
- free of charge

The following aspects based on the Open Data criteria seem especially important for the implementation of Open Data; the definitions take the special framework conditions in Germany into account.

¹Criteria were established in 2007 by the non-governmental organisation Sunlight-Foundation as maximum requirements for data and have since been developed further.

Open Government

Open Government describes the further opening of governmental administrations for the citizenry, business and science, whereby the goal is to promote transparency, participation and cooperation. This opening is carried out with the help of modern information and communication technologies (including social media). The different sections of Open Government can be more or less considered in line with Open Government projects. The boundaries can be blurred. Open Government needs to be understood as a generic term for the three sections transparency, participation and cooperation that are defined below. Raw data (primary source) Raw data is present as long as the collected data has not been accumulated and has not yet been assessed. In individual cases, data might have been processed to make it easier to handle or to comply with legal requirements (privacy, statistical confidentiality).

SimpleSearch

“SimpleSearch applications are simple mini-applications that turn a tabular data base into a program with little effort. It can be set-up with the SimpleSearch construction kit for Imperia. For example, Excel tables or other structured data can be quickly transformed into a searchable data base with Web Frontend.”²

Social networks

Platforms for the cultivation of social relationships: People can virtually connect to each other (network) and thus contacts can be established or cultivated. Social networks usually also enable the exchange of messages using an integrated mail service.

Participation

Participation in the sense of Open Government stands for improved participation in governmental decision-making processes. The basis goal is to involve the society more in the decisions of politics and government. The form of participation can thereby be more or less pronounced; from a survey of the citizens in regard to a topic and (e-)consultation to legislative proposals and involvement in the budget in line with the so-called participatory budget. These forms of participation are supposed to increase the acceptance of decisions.

²<http://support.berlin.de/wiki/index.php/SimpleSearch>.

Transparency

Transparency stands for the further publication of information, opinion-forming, assessment and decision-making processes and the availability of public sector data. It is the basic goal to make the basis of decision-making, the use of financial resources, the responsibilities of the decision-makers, etc. freely available in order to present the activities of politics and the government in an easily understandable and comprehensible manner and to promote well-informed social discussions.

Web 2.0 tools, social media

The tools of Web 2.0 are characterised by their ability to allow the communication from many to many people. They are easy to use and interactive. Every user can participate. Popular tools are Wikis, Weblogs, and social networks. The term social media is mainly used synonymously.

Wiki

Internet page that users can read and change, allows joint editing of texts.

Cooperation, collaboration

Cooperation describes the improved cooperation of different public authorities as well as between the government and the citizenry, business and science while performing tasks (for example electronic reporting systems). The goal is to improve the service quality of governmental functions, to use the widespread knowledge in the society and, if applicable, to relieve the government from tasks.

Differentiation of the sections: The individual aspects are distinct in the individual Open Government projects. In part, there are several aspects involved, in part they are mutually dependent (for example in line with an electronic reporting system, the involvement of the citizens is requested (cooperation); the conclusions that the government is drawing from the notifications are made transparent). A clear distinction is not always possible.

The state of Berlin commissioned the Fraunhofer Institute for Open Communication Systems FOKUS with the assessment of the organisational, legal and technical basis for Open Data in the capital in line with the eGovernment program "ServiceStadt Berlin". Hereby, a concept study was prepared and implemented in an online pilot system, the first Open Data portal in Germany.

This study documents the framework conditions as well as the concept for implementation, the prototype and initial experiences from the pilot phase. Fraunhofer FOKUS, the Fraunhofer Institute for Open Communication Systems develops and researches communication and information technologies at its facility in Berlin.

The research activities focus on the development of IT infrastructures and networks for the communication across domains and organisations. With more than 20 years of expertise, Fraunhofer FOKUS is an important IT partner for the telecommunication, automotive and media industry as well as governmental administrations. On a scientific level, Fraunhofer FOKUS is a well-known partner in research programmes for the European Commission as well as of domestic sponsors.