Metadata Harvesting and Quality Assurance within Open Urban Platforms

PHILIPP LÄMMEL, Fraunhofer FOKUS

BENJAMIN DITTWALD, Fraunhofer FOKUS

LINA BRUNS, Fraunhofer FOKUS

NIKOLAY TCHOLTCHEV, Fraunhofer FOKUS

YURI GLIKMAN, Fraunhofer FOKUS,

SILKE CUNO, Fraunhofer FOKUS

MATHIAS FLÜGGE, Fraunhofer FOKUS

INA SCHIEFERDECKER, TU Berlin & Fraunhofer FOKUS

During the past years, various activities and concepts have shaped and prepared the path for the development of urban environments towards smart cities across the world. One of the initial activities was relating to the opening of vast amounts of data from various public administrations and utility companies within a city, in order to create a viable eco-system of urban services and applications. Thereby, the harvested metadata needed to be verified in terms of correctness and a corresponding level of quality had to be assured. In addition, the concept of an Open Urban Platform emerged as an overall solution for smart cities ICT in the sense that an abstract reference model was established and standardized, providing an overall picture of the ICT structures within a city. Within this paper, we use the Open Urban Platform concept as the basics to describe and map our activities within the Open Data domain focusing mainly on the Open Data prototype for German Open Governmental Data – namely GovData.DE. Thereby, we describe our metadata harvesting and metadata quality assurance approach and discuss on lessons learned which flow into the definition of metadata quality metrics and have the potential to lead to a corresponding standard within the DIN German national standardization.

CCS Concepts: • **Information systems → Database management system engines**

KEYWORDS

metadata, quality assurance, open urban platforms, smart cities, open data

1 INTRODUCTION

Smart Cities are currently emerging as one of the pillars of our future digitalized world. Indeed, it took a couple of years to develop various concepts and approaches towards the structured creation and establishment of an urban ICT eco-system, and it must be stated that still many cities do not follow a holistically organized and sustainable design to the development of their Smart City ICT environments. Very often cities across the world allow service and application providers to simply deploy their products and achieve revenue without contributing to a viable overall eco-system. This creates additional silos based on the customer segments these providers have managed to reach and does not foster the involvement of multiple stakeholders and their interplay, interoperability and information/data exchange within an urban

environment. A feasible approach to supporting the development of eco-systems is to follow the path of *system openness* in its various dimensions.

One possible track which has been pushed for a lot by research, academia and public institutions during the past years is constituted by the idea of Open Data. Nevertheless, Smart Cities – when it comes to breaking silos, exploiting synergies, and fostering collaboration - have been and are currently emerging around the data that is gathered and made available within an urban environment. In this line of thought, many research and deployment projects have been financed to study and establish Open Data portals and publishing processes across different cities [1][2], countries [3][4] and continents [5]. Ideally, Open Data would be embedded in a more complex open eco-system encompassing and providing a framework for various dimensions of openness such as **open standards**, **open interfaces** and **open source**, thereby allowing for highly scalable integrative systems for urban ICT enabling Smart City applications and services. Reference architectures for such an open eco-system are given by approaches like the EIP SCC results [12], the H2020 Triangulum reference architecture [8][9], the reference architecture of H2020 Espresso [7][10] as well as the proposition for an Urban Data Space (UDS) [14] [15] [16] as worked out in recent Fraunhofer studies and publications.

In the current paper, we embed our experience on the development and operation of Open Data platforms into an Open Urban Platform, which was initially defined on European level within the EIP SCC[1] [11] [12] and further adapted to the needs of the German market in the scope of DIN SPEC 91357 Open Urban Platform [13]. Thereby, we focus on one of the main problems when dealing with a large scale collection of public data, namely the quality of the gathered, refined and offered (to the public) metadata. The metadata can describe different datasets and data streams (e.g. from IoT networks) which are openly available, harvested and catalogued within metadata catalogues such as CKAN [17], Piveu [25] or Socrata [24]. Thus, we describe our practical experiences on the automated harvesting and quality assurance for metadata within one of the large scale Open Data prototypes of the past years – the GovData.DE portal which was prototyped and operated by Fraunhofer FOKUS in its pre-production phase. The metadata quality is guaranteed by various monitoring tools and an overally established process leading to the regular interaction with the metadata providers - i.e. public institutions harvested by GovData.DE. The platform and its harvesting components are mapped to the DIN SPEC 91357 OUP blueprint of an urban ICT eco-system and operated based on sustainable design principles avoiding vendor lock-in and enabling the creation of an integrative eco-system consisting of academia, SMEs, industry and open source projects.

More specifically, we show how we solved the problem of *"How to guarantee the metadata quality within Open Urban Platforms?"* within the development and subsequent operation of the prototype of the German Governmental Open Data platform (govdata.de). The technical contribution is given by the implementation of an efficient metadata harvesting processes within this data platform, thereby realizing a set of guidelines that were systematically discussed and derived with the responsible public servants playing the role of data providers.

The rest of this paper is organized as follows: Section 2 presents the ideas of Open Urban Platforms and focusses on EIP SCC and DIN 91357 thereby explaining the various dimensions of openness which are of particular interest. Section 3 describes the metadata harvesting process

---

[1]EIP SCC stands for European Partnership Smart Cities and Communities: https://eu-smartcities.eu/

within GovData.DE and positions the relevant components in the scope of the DIN 91357 reference architecture. Section 4 elucidates on some resulting findings and recommendations relating to the quality of metadata, which have the potential to ease the establishment of the metadata quality process and the belonging delivery chains, e.g. the one emerged within GovData.DE. Section 5 summarizes other related works and puts our approach in the perspective of current developments. The final section 6 concludes the paper and presents briefly some subsequent works whilst at the same time outlining future research directions.

2   OPEN URBAN PLATFORMS

The paragraphs below introduce and elucidate on the concept of Open Urban Platforms, which has emerged across Europe during the past years and constitutes a viable blueprint for building and structuring a Smart City ICT.

2.1   The Need for Open Urban Platforms

Open Urban Platforms have emerged during the past years as a high level concept for providing design principles for ICT infrastructures in Smart Cities. Thereby, a large number of aspects of vital importance for urban ICT are captured within the OUP concept as defined within EIP SCC [12] and DIN 91357 "Reference Architecture Model Open Urban Platform (OUP)" [13], being the corresponding German national standard. These vital aspects are constituted by (1) the definition of an **abstract blue print for ICT in Smart Cities**, (2) the requirement for the utilization of **open standards** and (3) the need for the **increased utilization of open interfaces** based on such open standards towards the (4) **creation of urban ICT eco-systems** based on (5) **interoperability** and (6) **avoiding vendor lock-in**. The open standards are expected to originate from within different standardization bodies such as ISO, DIN, ETSI, IETF etc. The topic of vendor lock-in avoidance is of particular importance for future sustainable Smart Cities by establishing means for preventing the dependence of one vendor or service provider and allowing for the creation of a viable eco-system consisting of industry, local SMEs, academia and other ICT related stakeholders. Hence, similarly to the established reference models from the Internet and telecom domains (such as ISO/OSI [26] and TCP/IP [27]) an Open Urban Platform must enable the integrative creation of concepts and Smart City solutions thereby providing and describing the basic and abstract ICT structure.

Fraunhofer FOKUS together with other partners on European level has conducted substantial research during the last years towards the definition of an abstract and holistic open urban platform and reference architecture for ICT in Smart Cities. The research includes a large number of interviews conducted in many European cities (e.g. Manchester, Eindhoven, Stavanger, Sabadell, Leipzig, Prague, Emden …) towards capturing the status regarding the legacy/traditional deployed systems and identifying the needs which emerge when establishing an ICT infrastructure in an urban context.

2.2   Standardized Open Urban Platform

In order to address the above needs, the European Innovation Partnership on Smart Cities and Communities (EIP SCC) [11] set up a corresponding work stream in 2015 towards defining an ICT reference architecture for Smart Cities in Europe. The resulting document [12] has defined the basic structure, which can be instantiated by concrete technical architectures within a city. As

discussed above, this basic structure allows intrinsically for the creation of an eco-system (including large-scale industry, SMEs, academia and open source initiatives) based on the integration of single solutions into complex service architectures targeting the need of the citizens and the public administration in general. The integrative approach based on the utilization of open interfaces - building on open industrial and community standards - allows for increased interoperability involving the different stakeholders and avoiding vendor lock-in because of applying the EIP SCC principles.

The European level EIP SCC document was adapted to the needs of the German Smart City market based on expert discussions within a DIN[2] specification work group including different German municipalities and industrial partners. The DIN OUP and EIP SCC key architectural illustration is provided in Figure 1 showing the structural ICT layers for concrete instantiations of technical architectures within a Smart City. The belonging documents [12] [13] contain a much more in-depth description of the layers, capabilities and design principles to be applied in modern ICT urban infrastructures, thus providing a full set of guidelines for developers, architects and municipal decision makers.

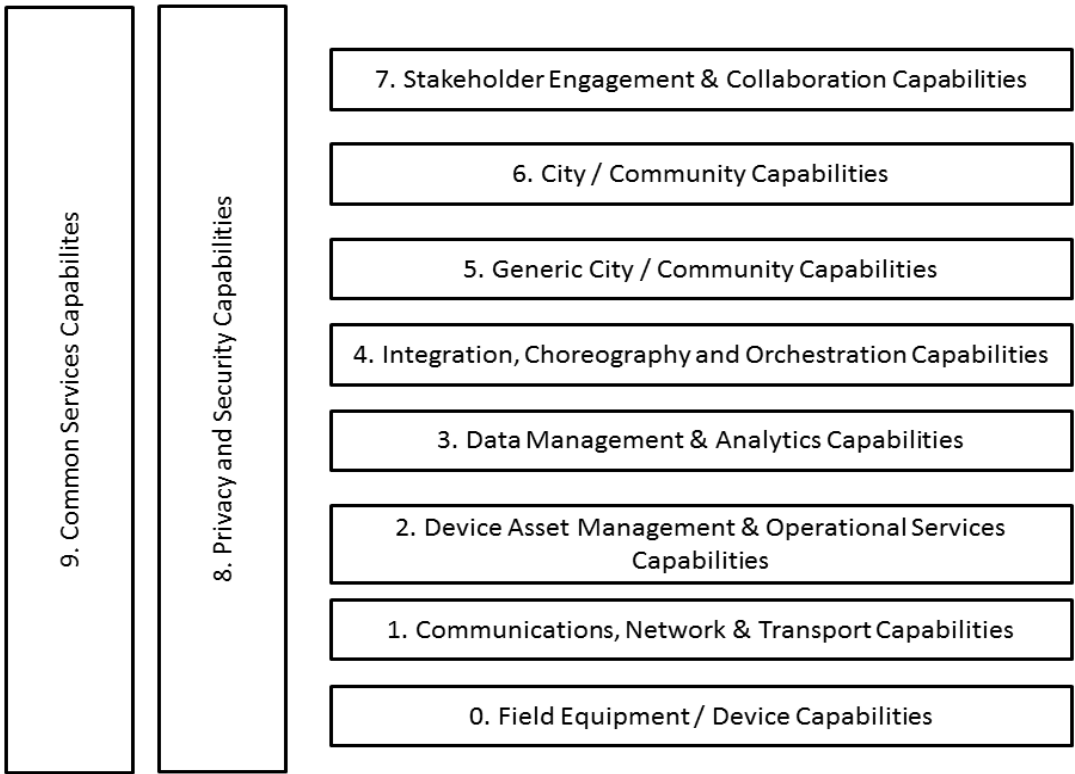
Figure 1 - The Open Urban Platform as a Reference Architecture defined within DIN OUP [13] and EIP SCC [12] and presented in [46]

---

[2] DIN stands for Deutsches Institut für Normung e.V. (in English, German Institute for Standardisation)

2.3 Open Systems in an Urban Context

Based on the above-introduced urban ICT reference architecture, a city can establish and integrate large-scale open ICT eco-systems. Thereby, the establishment of an open system is vital for the sustainable development of a Smart City. In that line of thought, a suitable understanding of the term Open System is required. We stick to a very generic definition as provided by the Oxford Dictionary of Computer Science establishing an Open System as *"any system in which the components conform to non-proprietary standards rather than to the standards of a specific supplier of hardware or software"* [29]. Hence, an open system is non-proprietary in terms of its implementation based on **open accessible standards** and communicates over **open interfaces**, which function based on standards that openly describe the belonging communication protocol/format and allow for an integrative approach towards the establishment of Smart City architectures avoiding **vendor lock-in**.

A further key aspect of openness of particular relevance for Smart Cities is given by the **Open Data** concept being at the heart of many Smart City projects for around a decade so far. Thereby, the definition provided within the above-discussed DIN OUP SPEC [13] is considered by us a straightforward and pragmatic one stating that Open Data encompasses "*datasets provided by public authorities and public enterprises to the general public*" [13]. Further definitions – as for example provided by [30] – involve more details like the property of data to be easily processable by machines/programs, the timeliness of the published data, and the licences under which the data has been made available to the public.

Another aspect of openness - which can/should be considered within the scope of Open Systems in urban environments - is constituted by the utilization of Open Source components for the creation of Smart City eco-systems. Open Source is mainly characterized by aspects such as free source code, free distribution, free possibility for modifications and free for the creation of derived works [58]. In addition, the Open Source components should be made available under corresponding licenses, which enable the free usage and further development of the code [58]. In general, the utilization of Open Source should be considered a viable path for the establishment of open urban ICT eco-systems thereby enabling the interoperability of the ICT modules and facilitating the avoidance of vendor lock-in.

Naturally, Open Systems in an urban context should be established within an Open Urban Platform (OUP) as the one presented in the paragraphs above and standardized in the scope of DIN and EIP SCC.

3 METADATA HARVESTING

Having discussed Open Urban Platforms and openness as a vital prerequisite for sustainable ICT Smart City eco-systems, the following subsections present our experiences relating to the piloting of the German governmental data portal. Thereby, we relate the developed platform and its processes to the overall DIN 91357 architecture and illustrate how the belonging design principles and requirements are fulfilled by GovData.DE.

At this point of time, the history of GovData.DE should be shortly presented as to support the reader in following the presentation. GovData.DE was initiated in 2012-2013 by the Federal Ministry of the Interior based on a study [22] that was executed in collaboration with the

Fraunhofer FOKUS institute in Berlin. In the following years Fraunhofer FOKUS was responsible for the operation of the German governmental data portal pilot until it was politically decided to proceed from the prototype phase to the production phase and hand over the platform to one of the governmental service providers. The experiences presented here are based on the pilot phase, for which Fraunhofer FOKUS was responsible and some first initial lessons learned were published in the conference articles [21] and [55].

3.1   German Open Data Portal

As previously mentioned, in this section we are presenting our activities around the official pilot of the German Federal Ministry of the Interior - GovData.de – and relate these to the Open Urban Platform of DIN SPEC 91357. The portal GovData.DE was officially started on the 9th of February 2013 under the following address - *https://www.govdata.de*.

The prototype was developed and improved in an iterative way meaning that we had several feedback rounds with the German Federal Ministry of the Interior regarding the prototype and its look and feel. Thereby, we had a couple of problems which were iteratively fixed, e.g. with respect to cybersecurity. These problems were revealed in belonging hacking workshops were we invited the community to attack our prototype and try to hack it. The main lessons are that security patches for Open Source software should be installed immediately after having been made available and that vulnerability databases (e.g. CVE MITRE) should be strongly considered when building an Open Source based system that is expected to turn operational in a wider context.

The homepage of the initial prototype can be seen in Figure 2. The landing page focuses on the two most important aspects of such an Open Data portal: a search bar for finding relevant data and a quick navigation to the available categories thereby giving end users the possibility to rapidly find data they are interested in.

Figure 2 - Screenshot of the Portal GovData.DE in its Prototype Stage

GovData.DE offers different types of data – datasets, documents and applications: Datasets encompass metadata information of historic, static and real-time data as well as their location thereby enabling end users and developers to build applications on top of these datasets. These applications can then be published on GovData.DE and can be used by other end users. Documents encompass different types of documents, such as documentations of APIs, planning documents like development plans and others. The number of published datasets is constantly growing. Currently, there are more than 36 000 datasets from all across Germany.

Almost all data that is published on GovData.DE is available under free licenses, e.g. Data license Germany (dl-de 2.0) or Creative Commons Attribution (CC-BY). Under these licenses, developers are allowed to use datasets for commercial and non-commercial use, in particular, they are allowed to copy, alter or process data and can merge data with own data or data of others in order to create new knowledge and increase the usefulness of Open Data.

In the following subsections the high-level architecture, operational aspects as well as reporting strategies will be presented.
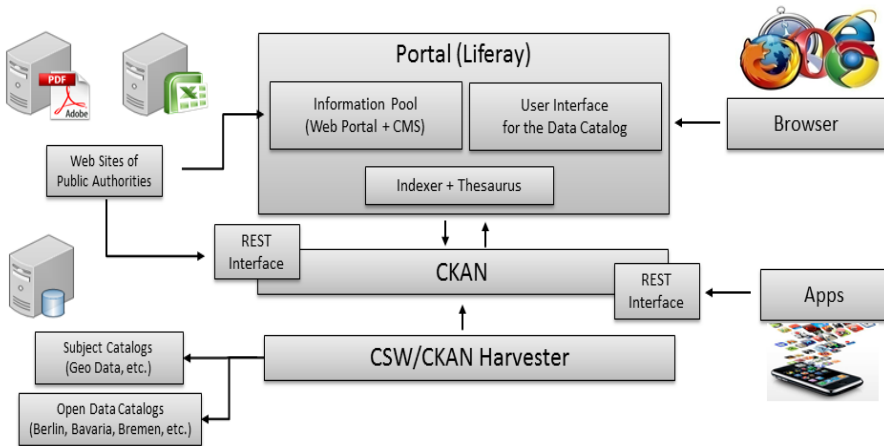

Figure 3 - High-Level Architecture of the GovData.DE Platform as presented in [44]

3.1.1   General Architecture

The high-level architecture of the GovData.DE platform is depicted in Figure 3 and will be elaborated in more detail in the following paragraphs. This architecture can be clearly placed in the *"3. Data Management & Analytics Capabilities"* of DIN SPEC OUP as presented in Figure 1. Thereby, the data is consolidated using the telecom/Internet connectivity in the layer below (*"1. Communication, Network & Transport Capabilities"*). The sources of Open Data for GovData.DE are either static data sets provided by public institutions or sensor data streams origination from the lowest layer (layer *0.*) of DIN SPEC OUP and EIP SCC. A number of detailed capabilities can be listed here that describe the functionality and properties of an Open Data Portal within an DIN 911357, however for the sake of keeping a comprehensible level for the reader we omit the detailed naming of the belonging capability description from the identified layers.

It is also important to remark that we considered various aspects of openness in our design, specification and implementation, in order to avoid a potential vendor lock-in. The selected

software components were Open Source, run on Open Source operating platforms (i.e. Linux) and in an Open Source webserver (nginx, Tomcat und Apache) utilizing open interfaces and standards (http(s), DCAT, CSW, JSON …).

Statistik für:
www.govdata.de

Zusammenfassung
Wann:
Monatliche Historie
Tage im Monat
Wochentage
Stunden (Serverzeit)
Wer:
Länder
⊞ Gesamte Liste
Rechner
⊞ Gesamte Liste
⊞ Letzter Zugriff
⊞ Unaufgelöste IP Adressen
Robots/Spiders (Suchmaschinen)
⊞ Gesamte Liste
⊞ Letzter Zugriff
Navigation:
Aufenthaltsdauer
Datei-Typen
Downloads
⊞ Gesamte Liste
Zugriffe
⊞ Gesamte Liste
⊞ Einstiegsseiten
⊞ Exit Seiten
Betriebssysteme
⊞ Versionen
⊞ Unbekannt
Browser
⊞ Versionen
⊞ Unbekannt
Verweise:
Herkunft
⊞ Suchmaschinen
⊞ Websites
Häufigkeit
⊞ Suchausdrücke
⊞ Suchbegriffe
Sonstige:
Verschiedenes
HTTP Fehlercodes
⊞ Nicht gefundene Seiten

Monat	Unterschiedliche Besucher	Anzahl der Besuche	Seiten	Zugriffe	Bytes
Jan 2014	5,524	10,493	205,523	1,549,880	13.70 GB
Feb 2014	6,005	11,091	245,507	1,835,888	15.71 GB
März 2014	5,066	9,644	221,793	1,446,772	13.47 GB
Apr 2014	5,324	9,786	240,857	1,622,484	14.10 GB
Mai 2014	5,791	10,566	235,330	1,794,979	15.42 GB
Juni 2014	5,116	9,515	201,747	1,552,021	13.54 GB
Juli 2014	8,499	19,038	461,851	3,207,023	25.47 GB
Aug 2014	5,245	14,798	275,532	1,742,407	15.78 GB
Sep 2014	6,130	12,899	266,818	1,937,528	16.07 GB
Okt 2014	6,741	15,145	298,042	2,128,061	18.24 GB
Nov 2014	3,970	8,070	160,170	1,171,110	10.29 GB
Dez 2014	0	0	0	0	0
Total	63,411	131,045	2,813,170	19,988,153	171.78 GB

Tage im Monat

01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 Durchschnitt
Nov Nov Nov Nov Nov Nov Nov Nov Nov Nov Nov Nov Nov Nov Nov Nov **Nov** Nov Nov Nov Nov Nov Nov Nov Nov Nov Nov Nov Nov Nov

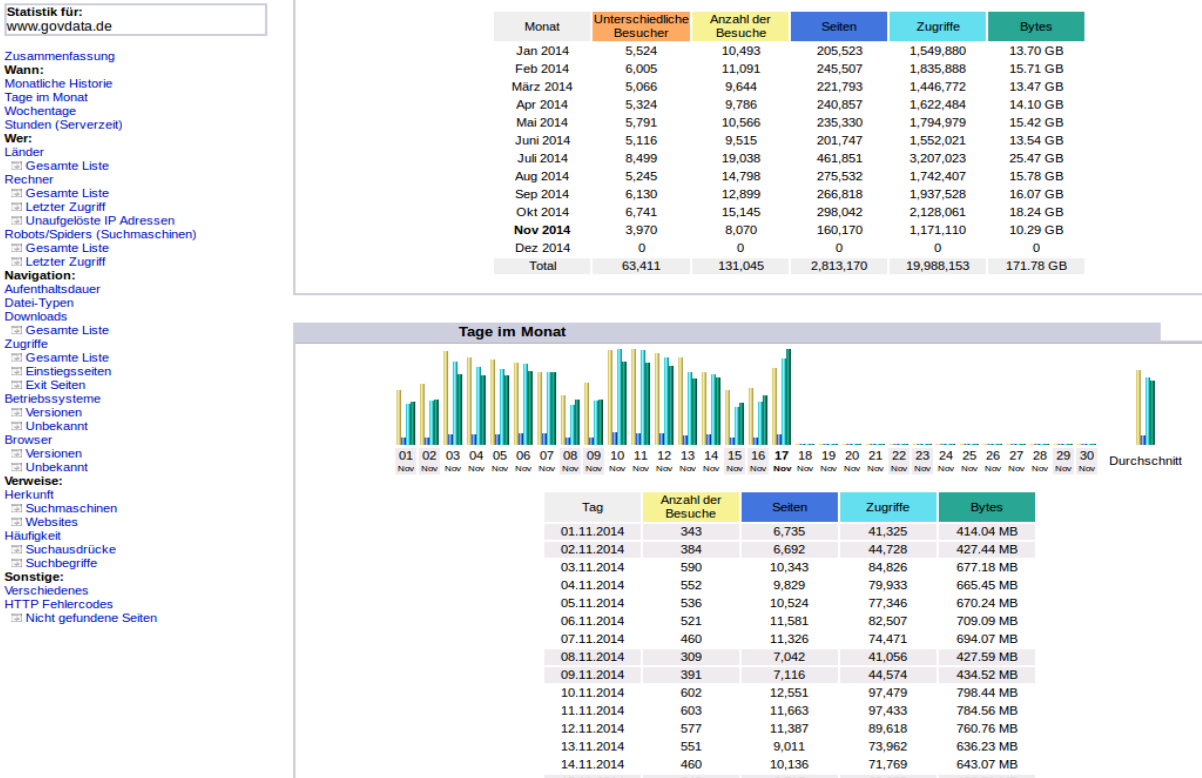| Tag	Anzahl der Besuche	Seiten	Zugriffe	Bytes
01.11.2014	343	6,735	41,325	414.04 MB
02.11.2014	384	6,692	44,728	427.44 MB
03.11.2014	590	10,343	84,826	677.18 MB
04.11.2014	552	9,829	79,933	665.45 MB
05.11.2014	536	10,524	77,346	670.24 MB
06.11.2014	521	11,581	82,507	709.09 MB
07.11.2014	460	11,326	74,471	694.07 MB
08.11.2014	309	7,042	41,056	427.59 MB
09.11.2014	391	7,116	44,574	434.52 MB
10.11.2014	602	12,551	97,479	798.44 MB
11.11.2014	603	11,663	97,433	784.56 MB
12.11.2014	577	11,387	89,618	760.76 MB
13.11.2014	551	9,011	73,962	636.23 MB
14.11.2014	460	10,136	71,769	643.07 MB

Figure 4 - Screenshot of a GovData.DE Report

As can be seen in Figure 3, the platform consists of two main components – a Liferay [57] based portal serving the webpage and a CKAN [17] storing metadata information. Liferay is an open source enterprise portal with which information, data and applications can be personalized under a common user interface in order to facilitate business processes. Liferay comes out-of-the-box with CMS features such as blog entries that are used on GovData.DE as well (see Information Pool in Figure 3). CKAN is a web-based open-source management system for the storage and distribution of data that is nowadays mainly used by public authorities to share their data with the general public. The metadata within CKAN is described in a DCAT aligned metadata scheme. Another example of this DCAT aligned format can be found in Ukrainian data portal [75] – see the following Listing 1 for an example containing a data set description with its belonging properties in an RDF form. Both components including their automated data harvesting and consolidation procedures can be mapped to relevant capabilities of the *"3. Data Management & Analytics Capabilities"* of DIN SPEC OUP as presented in Figure 1.

Listing 1: DCAT-AP Example Catalogue from the Ukrainian Open Data Portal

```
<rdf:RDF>
….
  <dcat:Catalog rdf:about="https://data.gov.ua">
   <dcat:dataset>
    <dcat:Dataset rdf:about="https://data.gov.ua/dataset/d0af9ba0-08b3-4bca-8508-
     02cffeaae8fd">
    <dcat:keyword>зброя</dcat:keyword>
    <dct:identifier>d0af9ba0-08b3-4bca-8508-02cffeaae8fd</dct:identifier>
    <dcat:contactPoint>
     <vcard:Organization rdf:nodeID="N34dff469ee8f45969a8c4440e0870cfc">
       <vcard:fn> Желєзний Андрій Петрович</vcard:fn>
       <vcard:hasEmail rdf:resource="mailto:andrii.zhelieznyi@mvs.gov.ua"/>
     </vcard:Organization>
    </dcat:contactPoint>
    <dct:publisher rdf:resource="https://data.gov.ua/organization/13d9803c-5c4d-47e8-
     b89f-3d6017c78ef9"/>
    <dct:language>ua</dct:language>
    <dcat:keyword>особа в розшуку</dcat:keyword>
    <dct:title>Інформація про викрадену, втрачену зброю</dct:title>
    <dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2020-
     04-30T14:46:58.441971</dct:modified>
    <dcat:distribution rdf:resource="https://data.gov.ua/dataset/d0af9ba0-08b3-4bca-
     8508-02cffeaae8fd/resource/5214a063-2679-4565-aafb-e4703aeed9c7"/>
    <dcat:keyword>зброя в розшуку</dcat:keyword>
    <dct:description>Інформація про викрадену, втрачену зброю</dct:description>
    <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2018-08-
     17T05:50:15.330697</dct:issued>
   </dcat:Dataset>
  </dcat:dataset>
  ….
 </dcat:Catalog rdf:about="https://data.gov.ua">
</rdf:RDF>
```

The Liferay based frontend retrieves data from CKAN via a REST[3] API, which can also be used by authorized users to add new datasets to GovData.DE and by public authorities to push data directly and in an automated way to CKAN. Of course, public authorities are able to publish data via a web form on the portal as well which will make the necessary API calls on behalf of the authority. With these two push approaches, both tech-savvy and technology-averse end-users are able to contribute to the portal and enrich the Open Data community.

---

[3] REST stands for Representational State Transfer and is a software architectural style for creating web services [31].

Pulling data into the portal was another possibility to add data. Special Catalogue Services for the Web (CSW) and CKAN Harvesters were implemented thereby enabling the automated integration of data. These harvesters – in total 22 - ran daily and contributed to the majority of data added. The harvesters with their key aspects will be explained in more detail in section 3.2. Since it was necessary to ensure a high availability of the portal, several different monitoring tools were continuously running, which will be explained in the next subsection.
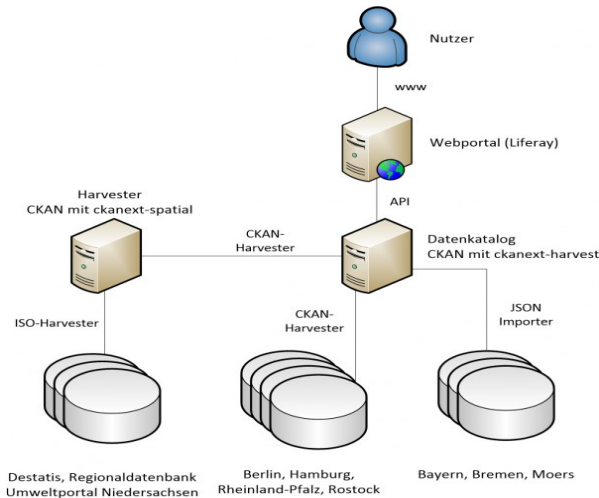
Figure 5 - Metadata Harvesting Process, initially depicted in [21]

3.1.2   Operational Aspects

As presumed, high availability and nearly faultless functioning of the portal was needed due to its importance to the German Federal Ministry of the Interior. In order to achieve these requirements, the prototype was thoroughly tested before an update went live and key components were deployed redundantly. Functional and GUI tests ran continuously on Jenkins [32]. Jenkins is an open source automation server that accelerates parts of the software development process. In addition, we sent alerts via e-mail and SMS (using Internet based SMS Gateways) whenever an error arose, in order to be able to act promptly and accordingly. Additionally to functional and GUI tests, load and performance tests were run with Apache JMeter [33] to verify the healthiness and proper functioning of the platform. With these tools we enabled 24/7 management support and a 99% availability of the portal was achieved.

Key components such as Liferay, CKAN and their databases were critical for proper functioning of the portal and therefore, multiple instances of these components were running simultaneously. A load balancer - we used *nginx* [34] for this task - received HTTP requests and distributed these requests to the next available instance thereby decreasing the individual load on each instance.

3.1.3   Reporting

In addition to regular quality assessments, a reporting process was also put in place. Advanced statistics reflecting on operation of the platform or user interactions with GovData.DE were provisioned in an easy to understand way. These reports were on an hourly basis but it was

possible to generate aggregated reports on a daily, weekly or monthly basis as well, thereby enabling trend analyses.

Schemaprüfer

Grundlage für die Schemaprüfung ist immer die aktuellste Version des OGPD JSON Schema.

Schemaverletzungen nach Datenbereitsteller

Die Anzahl der Datensätze mit Regelverletzungen des Schemas pro Datenbereitsteller.

Datenbereitsteller		Datensätze mit Regelverletzungen
http://daten.rlp.de		2276
www.regionalstatistik.de		819
http://datenregister.berlin.de		401
numis.niedersachsen.de		300
http://suche.transparenz.hamburg.de/		94
http://www.offenedaten.moers.de/		84
www.opendata.service-bw.de		53
null		45
http://daten.bremen.de/sixcms/detail.php?template=export_daten_json_d		23

Figure 6 - Quality Checks UI within the German Open Data Portal

A few selected information - unique visitors, page clicks, uptime of server, time that a visitor stayed on the platform and others - included in such a report is depicted in Figure 4. It can be seen that conclusions such as most important datasets or categories can be derived based on this report and thus, these reports give critical insights into optimization strategies for such an Open Data portal.

3.2 Metadata Harvesting Process

As already mentioned, the harvesting process was responsible for the majority of data added. In the scope of the project different source portals, namely CSW [59] sources, CKAN sources or portals that provided JSON [60] dumps of their data were supported and thus, we were able to harvest data from over 20 different source portals. The general harvesting process is depicted in Figure 5.

The three aforementioned supported source platform types are displayed in Figure 5. It is worth mentioning that the CSW harvester did not import directly to the main data catalogue (see "Datenkatalog" in Figure 5) but rather to a separate CKAN. We used this approach in order to achieve a high degree of metadata quality without risking the data integrity of the main data catalogue. This was not needed in the other harvesters due to better control of the information flow.

In the following, the general process of Metadata Harvesting will be described based on the CKAN harvester but the process is in principal the same for the other harvester types. Each harvester was triggered at a specific time. When the harvester started, it checked the source portal for changes, meaning that it searched for added, updated or removed datasets. Whenever a harvester found a change, it started to retrieve only the delta and updated the destination portal accordingly. With this process, minimal resource consumption with respect to e.g. bandwidth was ensured.

It should be mentioned that information was only added to the destination portal if it conformed to the metadata scheme of the destination portal thereby ensuring a first level of quality. Due to importance and high visibility of the project, a thorough check of provided information was needed in order to deliver the most accurate and up to date information. Therefore, additional quality checks were conducted which will be described in the following subsection.

3.3   Metadata Quality Assurance Project

In the scope of the project, a production level process for quality assurance within the harvesting procedures was established to achieve and maintain a high degree of quality. The most important checks and subtleties will be described in the next paragraphs.

The harvesters that were presented in section 3.2 were not just running on the production environment, they were running on a test system as well. The test environment was a pre-production stage that consisted of exactly the same components and the same setup but was only internally available. So, whenever a new harvester was implemented and was ready to be deployed, it was first tested in our test environment to ensure proper functioning. After verification of proper functioning, the metadata harvesters went through another phase of quality assurance: Succeeding a successful import of new data to the portal, additional quality checks were conducted. These checks ran automatically and emitted alerts whenever problems were found in the datasets. A typical check and quality report can be found in Figure 6. Each dataset was evaluated with respect to completeness of required information, availability of each provided URL and general conformance to the schema of the platform. The results were prepared and presented to the belonging data provider in regular audio conferences, thereby enabling an iterative approach to increase the quality of data. Whenever a data provider has fixed faulty entries, we reran the corresponding harvester, which updated the information accordingly and afterwards, the quality checks were run on the most current version in our test system again. Once every error was fixed, the final version of metadata was imported into the production system and made available online for the community across Germany.

The abovementioned quality checks that were run on our test system – successful import of datasets, conformance to metadata scheme, completeness of information, availability of URLs [76] – were also run on the production system but at a lesser frequency thereby ensuring that all provided information is still valid on the production system. The same report (see Figure 6) was available to the data providers as well so that they were always able to proactively react in the rare case of errors on the production system.

4   METADATA QUALITY RECOMMENDATIONS

Based on the above described experiences, a number of recommendations can be derived which can improve the general metadata provisioning and increase the effectiveness of the metadata harvesting process. These recommendation and findings have paved the path for a research direction at Fraunhofer FOKUS that is currently in the discussion of shaping a standard at the DIN standardization body. It is important to remark that given the ongoing nature of the DIN discussions, we can only reveal the high level recommendations, which are to be characterized as very practical guidelines for public servants and officers towards the quality provisioning of metadata. Next, these basic recommendations are derived and discussed.

Regardless of the actual content data or its context, metadata should fulfil three major functions: (1) ensure findability of the data, (2) improve comprehensibility of the data and (3) enable further processing of the data. For each function, several requirements and recommendations for metadata can be derived, which are introduced in the following.

4.1 Ensure Findability of the Data

As data is only of use for others, if it can be found easily, findability is a common quality indicator for data. It is, for example, one of the four dimensions within the prevalent FAIR-principles[4]. Findability describes the discoverability of the data. At best, the data is as easy to find as possible not only for humans but also for machines. The findability of the data increases with the quality and completeness of its metadata. Consequently, bad and incomplete metadata hamper the findability of data. Therefore, the first action publisher should do is to fill in all metadata fields, not only mandatory ones. By this, more context is given, e.g. by providing keywords, adding themes etc., which makes it easier for the user to find the data. However, when publishing Open Data, it is unclear to the publisher at this point, who will use the data later on. Therefore, the publisher should make sure, that the data is findable for everyone independently of his or her professional background by using synonyms in addition to technical terms. Following recommendations to ensure findability of the data are explicitly listed:

-    Metadata should be as complete as possible.
-    Publisher should use controlled vocabularies wherever possible, especially for keywords and themes.
-    The title should be as precise as possible.
-    Temporal and geographical references should be part of the title. Example: "Number of Tourists in Berlin for September 2019".
-    In addition to technical terms, publishers should use synonyms as well.

4.2 Improve Comprehensibility of the Data

Comprehensibility or understandability describe how easy it is for a user to understand the data. Data and metadata are understandable, if they are clear and comprehensible to the user. After studying the data and metadata, no ambiguities should remain. At best, users already have a clear idea about the content data when viewing the respective metadata. All kinds of additional

---

[4] https://www.go-fair.org/fair-principles/

information about the data will make it easier for the user to create a context and to understand the data. Good metadata will help the user to classify the data thematically and evaluate its importance and relevance for his or her own purpose. Therefore, it is important to describe the data in a way that everyone can understand it, also non-professionals.

Following recommendations to improve comprehensibility of the data of the data are explicitly listed:

- The title of the data should be meaningful.
- Data publishers/providers should use free-text fields in the metadata to describe the content data. The description should be as precise as possible and provide a realistic depiction of the content data. The description should fit with the title of the data.
- Metadata should be understandable for both, professionals and non-professionals. Therefore, publishers should use easy-to-understand language and no abbreviations.
- Additional information about the data, e.g. survey methodology, should be described in the metadata.

4.3 Support further processing of the data

Some users not only want to have a look at the data, but rather use the data to create new value-added applications or services. Various metadata quality aspects may support the further processing of the data. From a technical point of view, users are able to process the data best when it is available in machine-readable formats. However, trustworthiness of the data is also of great importance for the decision of data users whether they want to further use the data or not. Trust can be created through credibility. Credibility can be gained for example through information about the data origin, through consistency and correctness. However, regarding the correctness it might occur that there are some minor issues within the content data. If the publisher knows about these issues, he or she should point this out in the metadata to make it transparent to data users. Following recommendations to support further processing of the data are explicitly listed:

- Metadata should be up-to-date. If contact information is outdated, users are not able to address feedback and questions. This might in turn effect their trust in the data negatively.
- To give users an idea of the reliability and correctness of the data, publishers should assign status information for each resource in the metadata.
- Metadata should be consistent, e.g. a modification date must not be earlier than the creation date.
- Publishers should state the provenance of the data.

The above discussed findings give clear indications of how to improve the metadata quality when providing data and complement the approach of data harvesting and automated quality checks with feedbacks (audio conferences) as implemented within the GovData.DE prototype. A large set of datasets where such automatic quality checks run from the start of the harvesting activities can be found within the European Open Data portal (www.europeandataportal.eu), where many of the above recommendations are implemented and automatically tested thereby significantly improving the efficiency and quality of the harvesting procedure from the very beginning. The following related work elaborations draw the larger picture and put our presented researcher in a broader perspective.

5 RELATED WORK

The current section reviews some recent research efforts relating to the key areas in the focus of this paper. We start with referring to some important recent activities from the domain of Open Data and proceed to the topics of urban ICT architectures, Smart City applications and services, and metadata quality.

5.1 Open Data

The current subsection reviews recent research activities relating to the domain of Open Data. [35] deals with Open Data from the perspective of biological research conducted on the level of molecular analysis. Thereby, different databases containing biological research data are analysed and corresponding conclusions are drawn regarding their accessibility and reliability as data sources. Furthermore, a set of recommendations regarding how to provide linked Open Data for molecular research and how to utilize it properly are defined which aim to improve the procedures regarding Open Data in the domain. [36] describes the situation around Open Data and the sharing of data between psychologists towards the improved reproducibility and verification of reports and empirical results. Thereby, the authors argue that not only the data but also the belonging program code (SPSS, R, Matlab …) - based on which the statistical analysis was conducted – should be shared and made available to the community. [37] elucidates on sharing and re-using Open Data in the domain of astrophysics, in order to increase the reproducibility of scientific results and provide means for error identification and results' analysis. The current situation in the astrophysics community is analysed in a structured manner and a comprehensive study of the shared artefacts (data, code, explanations …) is presented. [38] relates Open Data to the modern concept of Blockchain, which represents a distributed, immutable and verifiable P2P storage for data entries. Hence, the interplay between both concepts is reviewed in the light of various basic requirements for Open Data such as trustfulness and transparency. [39] handles the utilization of Open Data in the area of education and describes the role of so-called Open Educational Resources (OER). [40] discusses on the use of OERs for educational purposes and trainings within the public administration. Furthermore, [41] provides a discussion and arguments on how Open Data and related Smart City technologies, including participative technologies, can disintermediate the government as a central cornerstone handling the city business/activities and can enable the direct connection between citizens and the belonging assets, thereby moving the role of the local government to the background of urban processes. In addition, [42] presents open innovation strategies, which have the potential to boost Open Data based use cases and Smart City scenarios – thereby, the city of Singapore is looked at as a case study. In this context, the analysed strategies are viewed with regard to their implications to practice as well as their implications to research, and guidelines are provided for the engagement of a larger community towards the utilization and creation of Open Data based use cases. Finally, [44] and [55] present our experiences in developing Open Data portals over the past years and handling large amounts of data with respect to data and metadata harvesting. In this line of thought, [43] demonstrates the creation of a large scale urban data handling system consisting of an Open Data catalogue such as CKAN [43] and a modern distributed data store component based on Hadoop and HDFS.

5.2 Smart City Platforms

With respect to Smart City Platforms and conceptual frameworks, a variety of models were proposed during the last years with the intention to provide the structures and means for

efficiently implementing Smart City solutions and services. At the beginning of this paper, we discussed broadly about Open Urban Platforms and their development in different settings such as EIP SCC [11] and the German DIN standardization body. Smart City ICT architectures in close relation to DIN 91357 are given by the reference model of the H2020 Espresso project [10], the H2020 Triangulum ICT reference architecture [9][6][8] and the concept of Urban Data Spaces [14][15][16] as defined in a recent study conducted by the Fraunhofer FOKUS institute in collaboration with the German Federal Ministry of Education and Research. Moreover, Fraunhofer FOKUS works on a generalization of the EIP SCC, DIN 91357 and the Triangulum reference model – this generalization is denoted as oupPLUS [18] [46] and combines complementary concepts from all three above architectures including the vital SAPs (Service Access Points), which are understood as abstract artefacts containing various open communication protocols (e.g. MQTT, CoAP, HTTP, TLS, IPv4/v6 …) for data and information exchange between the architectural layers of an urban platform. Additional activities of increased relevance – at least in the European context – are provided by FI-WARE [61] and the International Data Spaces [62].

FI-WARE originates from a PPP (Public Private Partnership) that was funded by the European Commission in line with the developments around the idea of Future Internet (FI) that were determining the research agenda during the past decade. FI-WARE initially started with the concept of Generic and Specific Enablers which were representing various components that could either execute some basic or some tailored function in an eco-system of software components interacting over the Internet. Initially a large number of products and components were taken into account towards defining the FI-WARE eco-system, which has been further defined and currently contains a minimal number of (often C/C++ based) software components, including the Orion IoT broker [73] that integrates different sensor sources and represents them in the belonging ETSI NGSI (Next Generation Service Interfaces) format [64] for enabling a generic access to IoT information. Based on such principles, FI-WARE [61] aims to position itself in various domains such as automotive, the agricultural domain and Smart Cities in general. Further initiatives which relate to FI-WARE and Smart Cities are provided by OASC (Open Agile Smart Cities) [71] in close relation with the Synchronicity project and platform [72]. OASC and FI-WARE based components are often deployed in European cities (e.g. Eindhoven) and communicate over interfaces such as ETSI NGSI-LD prelim API [66], OMA NGSI [68], ITU-T SG20*/FG-DPM [70] embedded in an overall FI-WARE eco-system as the one discussed at the TeleManagement Forum (TMF) .

The International Data Space (IDS) [62] has started some years ago as an Industrial Data Space inspired by US activities such as the Industrial Internet [69] and European/German initiatives like Industry 4.0. The IDS aims at facilitating the exchange of information and data between various industrial partner in a P2P fashion. Thereby, the core functionality is implemented by a so-called Connector that realizes the connection between the internal data bases/storages of an enterprise and the external world. Thereby, the Connector implements the security standards and possesses the intelligence to implement data governance and privacy preserving activities when transferring data between the different scopes. IDS is currently extending its scope to domains beyond the pure industrial scope, including mobility (Mobility Data Space), eHealth (Medical Data Space) and further.

5.3 Applications & Services

Based on the published urban data, it is expected that new services and applications would emerge for a given city context. Indeed, such services & applications were increasingly developed during the past years and resulted in a large number of quality of life improvements in various city areas. These areas include mobility [49] [48], energy [66], emergency communications [65], education and further domains of importance for the ordinary citizen. Since the number of use cases is indeed overwhelming, we briefly refer to some examples which are close to or part of our research activities: [40] deals with the application and utilization of Open Data in the domain of remote online educational resources. [48] presents the interplay between various infrastructure providers (telecom network, electric grid, transport and mobility assets provider) towards the realization of electric mobility in Smart Cities. Furthermore, [49] presents the ideas of a mobility data cloud that integrates different types of data – including Open Data – and facilitates the provisioning of advanced services towards optimized urban mobility and increased utilization of electrical vehicles.

5.4 Metadata Data Quality

The quality of the metadata describing the Open Data sets is another vital aspect of the current paper. Related work in this area appears to be rather limited in comparison to the topics before. [50] elucidates on the metadata which is stored and managed within the Springer Nature research publishing platform. Thereby, a methodology for the quality assessment of the provided research metadata is described based on a systematic classification and curation of the metadata to a standardized level of detail. The article provides additional insights regarding the types of data described by the metadata and elaborates on recommendations regarding the provisioning and curation of the entries within the portal. [51] continues this work by defining the editorial flows for the above said metadata towards increasing and assessing its quality in the context of describing and managing research data and publications. This includes key lessons learned regarding the annotation of the datasets and the presentation of a metadata curation system. [52] presents an Open Data quality measurement framework and applies it to governmental Open Data on different levels (e.g. municipality) within the context of Italy. A set of metrics is defined that allows to calculate particular metadata quality rankings and to compare the quality of different datasets, which is also described as a case study within the paper. [53] presents the introduction of DCAT-AP as a metadata management format within the scope of Czech national Open Data catalogue. DCAT-AP [63] is the latest standard for metadata descriptions which has been worked out on European level and is currently utilized within the European Open Data portal. [53] analyses among others the impact on the metadata quality achieved through the introduction of DCAT-AP in the Czech national Open Data platform. The quality includes aspects such as the usability and the performance when it comes to ETL (Extract, Transform, Load) pipes execution for required transformations and data harvesting. [54] discusses on how to improve the findability and discoverability for open government data through the enrichment of the belonging metadata with corresponding vocabularies capturing key semantic aspects. Finally, [74] defines a large number of metrics (e.g. level of completeness, email information …) for the CKAN metadata format utilized within the prototype of GovData.DE and provides various numerical evaluations of different datasets as well as Open Data portals at their belonging development stage.

6 CONCLUSIONS

The current paper presented a variety of topics closely connected to the Fraunhofer FOKUS Smart City activities of the past years. We presented our work on Open Urban Platforms and their standardization, which constitute a cornerstone for the establishment of urban ICT eco-systems involving multiple different stakeholders such as industry, academia, SMEs and open source initiatives. The basic idea is to structure the ICT within a city in a layered manner similar to the proven approach undertaken by other legacy reference architectures such as TCP/IP and ISO/OSI. Furthermore, components which are placed/mapped within different layers should follow the principles of openness in terms of implementing *open interfaces* based on *open standards*, which are freely accessible of and well described API specifications. Moreover, the open systems can include aspects of Open Data and Open Source towards the creation of a truly scalable integrative urban ICT eco-system. This would allow for breaking silos, pushing vendors and service providers to better structure their solutions and allow the combination of various modules based on the interoperability and standard conformance of the above mentioned open interfaces.

In the spirit of the above presented Open Urban Platforms and their intrinsic features, we presented our experiences on the operation of the prototype of the German Open Data platform GovData.DE, including the belonging data harvesting processes resulting in the consolidation of metadata across the various platforms on German federal level. The GovData.DE platform is designed according to the principle of DIN SPEC 91357 and its data harvesting processes have evolved to a sophisticated quality assurance process for metadata, which is elaborated in the current paper. The lessons drawn from these experiences have led to a number of insights defining rules and recommendations regarding the provisioning of metadata in urban data platforms. These recommendations were correspondingly presented and discussed.

With respect to future work, we plan the standardization of the initial recommendations defined in this paper and currently follow the path of establishing a specification group within the German national standardization body, i.e. DIN. Furthermore, the experience gained from the operation of the German national Open Data Platform are continuously reviewed and contributed to a similar effort of Fraunhofer FOKUS and corresponding partners – namely the development, operations and data harvesting for the European Open Data portal [56].

REFERENCES

[1]   Berlin Open Data portal: https://daten.berlin.de/, as of date 12.09.2019
[2]   Hamburg Transparency Portal and Open Data: http://transparenz.hamburg.de/open-data/, as of date 12.09.2019
[3]   UK Open Data Portal: https://data.gov.uk/, as of date 12.09.2019
[4]   Open Data Denmark: http://www.opendata.dk/, as of date 12.09.2019
[5]   Open Data for Africa: http://dataportal.opendataforafrica.org/, as of date 12.09.2019
[6]   H2020 Triangulum, https://www.triangulum-project.eu/, as of date 12.09.2019
[7]   H2020 Espresso, http://espresso-project.eu/, as of date 12.09.2019
[8]   Ina Schieferdecker, Nikolay Tcholtchev, Philipp Lammel, Robert Scholz, Evanela Lapi: Towards an Open Data Based ICT Reference Architecture for Smart Cities. CeDEM 2017: 184-193
[9]   Nikolay Tcholtchev, Robert Scholz, Philipp Lämmel, Rene Richter, Evanela Lapi (Fraunhofer FOKUS), Kai Tepe (TüV Süd), D 6.1: ICT Reference Architecture, http://www.triangulum-project.eu/wp-content/uploads/2017/12/D6.1_ICT_Reference_Architecture.pdf, as of date 12.09.2019
[10] Espresso Deliverable, D4.2 – Definition of Smart City Reference Architecture, http://espresso.espresso-project.eu/wp-content/uploads/2017/03/D4-17579.2-Smart-City-reference-architecture-report.pdf, as of date 12.09.2019
[11] European Innovation Partnership Smart Cities and Communities, EIP SCC, https://eu-smartcities.eu/, as of date 12.09.2019
[12] EIP SCC work stream 2 – Main Deliverable, Reference Architecture, Draft, version 0.98 (Work in Progress), 22.06.2017, Main editors: Lutz Heuser, Jeroen Scheer, Pieter den Hamer , Bart de Lathouwer
[13] DIN SPEC 91357, Reference Architecture Model Open Urban Platform (OUP), https://www.din.de/en/wdc-beuth:din21:281077528?sourceLanguage&destinationLanguage, as of date 12.09.2019
[14] Ina Schieferdecker, Lina Bruns, Silke Cuno, Matthias Flügge, Karsten Isakovic, Jens Klessmann, Philipp Lämmel, Dustin Stadtkewitz, Nikolay Tcholtchev, Christoph Lange, Benedikt Imbusch, Leonie Strauß, Alex Vastag, Florian Flocke, Volker Kraft, "Handreichung zur Studie: Urbane Datenräume - Möglichkeiten von Datenaustausch und Zusammenarbeit im Urbanen Raum", publication date: 2019/4/5, oninle: https://cdn0.scrvt.com/fokus/702aa1480e55b335/bc8c65c81a42/190311_Handreichung_UDR_02.pdf, as of date 12.09.2019
[15] Volker Kraft, Ina Schieferdecker, Lina Bruns, Silke Cuno, Matthias Flügge, Karsten Isakovic, Jens Klessmann, Philipp Lämmel, Dustin Stadtkewitz, Nikolay Tcholtchev, Christoph Lange, Benedikt T. Imbusch, Leonie Strauß, Alex Vastag, Florian Flocke, "Urbane Datenräume – Möglichkeiten von Datenaustausch und Zusammenarbeit im Urbanen Raum", publication date: 2018/6, pp. 1-275, online: https://cdn0.scrvt.com/fokus/774af17bdc0a18cd/1f5c791aa727/UDR_Studie_062018_final.pdf, as of date 12.09.2019
[16] Cuno, S.; Bruns, L.; Tcholtchev, N.; Lämmel, P.; Schieferdecker, I. Data Governance and Sovereignty in Urban Data Spaces Based on Standardized ICT Reference Architectures. *Data* **2019**, *4*, 16.
[17] CKAN, Comprehensive Knowledge Archive Network: https://ckan.org/, as of date 12.09.2019
[18] oupPLUS web presence: https://www.fokus.fraunhofer.de/en/sqc/smart_cities_oupPLUS, as of date 30.01.2018
[19] oneM2M, http://www.onem2m.org/, as of date 30.01.2018

[20] FI-WARE, https://www.fiware.org/, as of date 30.01.2018

[21] Florian Marienfeld, Ina Schieferdecker, Evanela Lapi, Nikolay Tcholtchev, "Metadata aggregation at GovData.de: an experience report", OpenSym 2013: 21:1-21:5

[22] Klessmann, J. ; Denker, P. ; Schieferdecker, I. ; Schulz, S., 2012. Open government data Deutschland. Eine Studie zu Open Government in Deutschland im Auftrag des Bundesministerium des Innern. Berlin: Bundesministerium des Innern (Germany).

[23] Karimova Y., Castro J.A., Ribeiro C., "Data Deposit in a CKAN Repository: A Dublin Core-Based Simplified Workflow", In: Manghi P., Candela L., Silvello G. (eds) Digital Libraries: Supporting Open Science. IRCDL 2019. Communications in Computer and Information Science, vol 988. Springer, Cham

[24] Socrata: https://socrata.com/, as of date 12.09.2019

[25] Piveau Datenharvester Consus: https://www.fokus.fraunhofer.de/5abfb604919bb3f5, as of data 12.09.2019

[26] ITU-T Recommendation X.200 (07/1994), "Data Networks and Open System Communication, Open Systems Interconnection – Model and Annotation, Information Technology – Open Systems Interconnection – Basic Reference Model: The Basic Model"

[27] RFC1122: Requirements for Internet Hosts -- Communication Layers, https://tools.ietf.org/html/rfc1122, as of date 12.09.2019

[28] Clements, Paul; Felix Bachmann; Len Bass; David Garlan; James Ivers; Reed Little; Paulo Merson; Robert Nord; Judith Stafford, "Documenting Software Architectures: Views and Beyond", 2010, Second Edition. Boston: Addison-Wesley. ISBN 978-0-321-55268-6.

[29] Oxford Dictionary of Computer Science. Open System definition.

[30] Open Data principles, Sunlight Foundation: https://sunlightfoundation.com/policy/documents/ten-open-data-principles/, as of date 12.09.2019

[31] Fielding, Roy Thomas (2000). "Chapter 5: Representational State Transfer (REST)". Architectural Styles and the Design of Network-based Software Architectures (Ph.D.). University of California, Irvine.

[32] Jenkins, Continuous Integration Server: https://jenkins.io/, as of date 12.09.2019

[33] Apache JMeter: https://jmeter.apache.org/, as of date 12.09.2019

[34] Nginx: https://www.nginx.com/, as of date 12.09.2019

[35] Droege G, Karsch-Mizrachi I, Barker K, Coddington J, Seberg O, "Towards Linked Open Molecular Data: Recommendations for researchers, collections, infrastructures and publishers", 2019, Biodiversity Information Science and Standards 3: e37204. https://doi.org/10.3897/biss.3.37204

[36] Obels, Pepijn, et al. "Analysis of Open Data and Computational Reproducibility in Registered Reports in Psychology." PsyArXiv, 23 May 2019. Web.

[37] Anneke Zuiderwijk, Helen Spiers, "Sharing and re-using open data: A case study of motivations in astrophysics", International Journal of Information Management, Volume 49, 2019, Pages 228-241, ISSN 0268-4012,https://doi.org/10.1016/j.ijinfomgt.2019.05.024

[38] Joseph B.K., "Blockchain for Open Data – Exploring Conceptual Underpinnings and Practice", 2019, In: Rodríguez Bolívar M., Bwalya K., Reddick C. (eds) Governance Models for Creating Public Value in Open Data Initiatives. Public Administration and Information Technology, vol 31. Springer, Cham

[39] Atenas, Javiera and Havemann, Leo, "Open Data Sectors and Communities: Education", 2019, In: Davies, Tim; Rubinstein, Mor; Walker, Stephen B. and Perini, Fernando eds. The State of Open Data: Histories and Horizons. Cape Town and Ottawa: African Minds and International Development Research Centre, pp. 91–102.

[40] V. Damjanovic, D. Glachs, N. Tcholtchev, E. Ras, E. Tobias, "EAGLE – Open Data and Linked Data Architecture of an Enhanced Government Learning Platform", In: The 9th European Conference on

Technology Enhanced Learning (EC-TEL 2014) Open Learning and Teaching in Educational Communities, Graz, Austria, 16-19 September 2014

[41] Peter Johnson, "Disintermediating Government: The role of Open Data and Smart Infrastructure", ScholarSpace for the University of Hawaii, http://hdl.handle.net/10125/59723, ISBN:978-0-9981331-2-6

[42] Chan, Calvin M. L., "From Open Data to Open Innovation Strategies: Creating E-Services Using Open Government Data", 2013, 46th Hawaii International Conference on System Sciences (2013): 1890-1899.

[43] Robert Scholz, Nikolay Tcholtchev, Philipp Lämmel, Ina Schieferdecker, "From Metadata Catalogs to Distributed Data Processing for Smart City Platforms and Services: A Study on the Interplay of CKAN and Hadoop", CLOSER (Selected Papers) 2017: 115-136

[44] Florian Marienfeld, Ina Schieferdecker, Evanela Lapi, Nikolay Tcholtchev, "Metadata aggregation at GovData.de: an experience report", OpenSym 2013: 21:1-21:5

[45] Sotirios Paroutis, Mark Bennett, Loizos Heracleous, "A strategic view on smart city technology: The case of IBM Smarter Cities during a recession", Technological Forecasting and Social Change, Volume 89, 2014,Pages 262-272, ISSN 0040-1625

[46] Nikolay Tcholtchev, Philipp Lammel, Robert Scholz, Wojciech Konitzer, Ina Schieferdecker, "Enabling the Structuring, Enhancement and Creation of Urban ICT through the Extension of a Standardized Smart City Reference Model". UCC Companion 2018: 121-127

[47] Silke Cuno, Lina Bruns, Nikolay Tcholtchev, Philipp Lämmel, Ina Schieferdecker, "Data Governance and Sovereignty in Urban Data Spaces Based on Standardized ICT Reference Architectures", Data 4(1): 16 (2019)

[48] Nikolay Tcholtchev, Lena Farid, Florian Marienfeld, Ina Schieferdecker, Benjamin Dittwald, Evanela Lapi, "On the interplay of open data, cloud services and network providers towards electric mobility in smart cities", LCN Workshops 2012: 860-867

[49] Nikolay Tcholtchev, Benjamin Dittwald, Thomas Scheel, Begüm Ilke Zilci, Danilo Schmidt, Philipp Lämmel, Jurma Jacobsen, Ina Schieferdecker: The Concept of a Mobility Data Cloud: Design, Implementation and Trials. COMPSAC Workshops 2014: 192-198

[50] Grant, Rebecca, Smith, Graham, Hrynaszkiewicz, Iain, "Assessing metadata and curation quality: a case study from the development of a third-party curation service at Springer Nature", D 2019, R 10.1101/530691, bioRxiv,https://www.biorxiv.org/content/biorxiv/early/2019/01/27/530691.full.pdf

[51] Salatino, Angelo; Osborne, Francesco; Birukou, Aliaksandr and Motta, Enrico, "Improving Editorial Workflow and Metadata Quality at Springer Nature", 2019, In: The 18th International Semantic Web Conference (ISWC 2019), 26-30 Oct 2019, Auckland, New Zealand, (In Press).

[52] Vetrò, Antonio;Canova, Lorenzo; Torchiano, Marco; Orozco Minotas, Camilo; Iemma, Raimondo; Morando, Federico, "Open Data Quality Measurement Framework: Definition and Application to Open Government Data", In:GOVERNMENT INFORMATION QUARTERLY, ISSN 0740-624, 33:2(2016), pp. 325-337, Elsvier, DOI:10.1016/j.giq.2016.02.001

[53] Jakub Klímek, "DCAT-AP representation of Czech National Open Data Catalog and its impact", Journal of Web Semantics, Volume 55, 2019, Pages 69-85, ISSN 1570-8268, https://doi.org/10.1016/j.websem.2018.11.001.

[54] Petr Křemen, Martin Nečaský, "Improving discoverability of open government data with rich metadata descriptions using semantic government vocabulary", Journal of Web Semantics, Volume 55, 2019, Pages 1-20, ISSN 1570-8268, https://doi.org/10.1016/j.websem.2018.12.009.

[55] Evanela Lapi, Nikolay Tcholtchev, Louay Bassbouss, Florian Marienfeld, Ina Schieferdecker: Identification and Utilization of Components for a Linked Open Data Platform. COMPSAC Workshops 2012: 112-115

[56] European Open Data portal: https://www.europeandataportal.eu, as of date 13.09.2019

[57] Liferay: https://www.liferay.com/, as of date 13.09.2019

[58] Open Source Initiative, 2007, *The Open Source Definition,* available at: https://opensource.org/osd, accessed 27 May 2019

[59] Open Geospatial Consortium (OGC), "OpenGIS Catalogue Service Implementation Specification", 2007-02-23.

[60] RFC 7158: "The JavaScript Object Notation (JSON) Data Interchange Format", online: https://tools.ietf.org/html/rfc7158, as of date 13.09.2019

[61] FI-WARE: https://www.fiware.org/, as of date 13.09.2019

[62] International Data Spaces: https://www.internationaldataspaces.org/, as of date 13.09.2019

[63] DCAT-AP: https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe, as of date 13.09.2019

[64] ETSI NGSI-LD: https://forge.etsi.org/rep/NGSI-LD/NGSI-LD, as of date 13.09.2019

[65] Yacine Rebahi, Kin Tsun Chiu, Nikolay Tcholtchev, Simon Hohberg, Evangelos Pallis, Evangelos K. Markakis, "Towards a next generation 112 testbed: The EMYNOS ESInet", IJCIP 22: 39-50 (2018)

[66] Rangelov, D., Tcholtchev, N., Lämmel, P., & Schieferdecker, I. (2019, October). Experiences Designing a Multi-Tier Architecture for a Decentralized Blockchain Application in the Energy Domain. In 2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT) (pp. 1-7). IEEE.

[67] ETSI NGSI-LD prelim API: https://www.etsi.org/newsroom/press-releases/1519-2019-01-etsi-cim-group-releases-full-feature-specification-for-context-information-exchange-in-smart-cities, as of date 13.09.2019

[68] OMA Next Generation Service Interfaces Architecture, online: http://www.openmobilealliance.org/release/NGSI/V1_0-20120529-A/OMA-AD-NGSI-V1_0-20120529-A.pdf, as of date 13.09.2019

[69] Industrial Internet Consortium: https://www.iiconsortium.org/, as of date 13.09.2019

[70] SG20: Internet of things (IoT) and smart cities and communities (SC&C), https://www.itu.int/en/ITU-T/studygroups/2017-2020/20/Pages/default.aspx, as of date 13.09.2019

[71] Open Agile Smart Cities (OASC): https://oascities.org/, as of date 13.09.2019

[72] H2020 Synchronicity: https://synchronicity-iot.eu/, as of date 13.09.2019

[73] Orion IoT Brocker: https://fiware-orion.readthedocs.io/en/master/, as of date 13.09.2019

[74] Konrad Johannes Reiche, Edzard Höfig: Implementation of Metadata Quality Metrics and Application on Public Government Data. COMPSAC Workshops 2013: 236-241

[75] DCAT aligned metadata scheme used by Ukrainian's data portal: https://data.gov.ua/catalog.xml, as of date 15.04.2020

[76] Reiche, Konrad Johannes, et al.: Assessment and Visualization of Metadata Quality for Open Government Data. Master's Thesis. Free University Berlin, 2013.