CrossMark

Towards certified open data in digital service ecosystems

Anne Immonen[1] · Eila Ovaska[1] · Tuomas Paaso[1]

Abstract The opportunities of open data have been recently recognized among companies in different domains. Digital service providers have increasingly been interested in the possibilities of innovating new ideas and services around open data. Digital service ecosystems provide several advantages for service developers, enabling the service co-innovation and co-creation among ecosystem members utilizing and sharing common assets and knowledge. The utilization of open data in digital services requires new innovation practices, service development models, and a collaboration environment. These can be provided by the ecosystem. However, since open data can be almost anything and originate from different kinds of data sources, the quality of data becomes the key issue. The new challenge for service providers is how to guarantee the quality of open data. In the ecosystems, uncertain data quality poses major challenges. The main contribution of this paper is the concept of the Evolvable Open Data based digital service Ecosystem (EODE), which defines the kinds of knowledge and services that are required for validating open data in digital service ecosystems. Thus, the EODE provides business potential for open data and digital service providers, as well as other actors around open data. The ecosystem capability model, knowledge management models, and the taxonomy of services to support the open data quality certification are described. Data quality certification confirms that the open data is trustworthy and its quality is good enough to be accepted for the usage of the ecosystem's services. The five-phase open data quality certification process, according to which open data is brought to the ecosystem and certified for the usage of the digital service ecosystem members using the knowledge models and support services of the ecosystem, is also described. The initial experiences of the

✉ Anne Immonen
anne.immonen@vtt.fi

Eila Ovaska
eila.ovaska@vtt.fi

Tuomas Paaso
tuomas.paaso@vtt.fi

[1] VTT Technical Research Centre of Finland, Digital Systems and Services, P.O. Box 1100, FI-90571 Oulu, Finland

⚛ Springer

still ongoing validation steps are summarized, and the concept limitations and future development targets are identified.

1 Introduction

Digital service providers have been increasingly interested in digital service ecosystems as the ecosystem-based service development provides several advantages, including collaborative innovation and value co-creation among ecosystem members. In a digital service ecosystem, the ecosystem members can utilize and share common assets and knowledge, nevertheless act independently. The product of a digital ecosystem, a digital service, can be anything that is intended to be entirely automated and can be delivered digitally through an information infrastructure. Recently, freely available open data has increasingly interested service providers, as this data has been identified to provide several business benefits, such as new data-based content, ideas and basic functions, increased understanding about business opportunities, improved competitiveness, and potential new customers (Immonen et al. 2014). Especially open social media data interests companies as it can provide insight into consumers' opinions, preferences, and requirements considering the company or its products/services (Bhatia et al. 2013; Antunes and Costa 2012; Fabijan et al. 2015), thus enabling the companies to achieve "customer insight" into business decision-making (Immonen et al. 2015a). Bringing open data into the context of ecosystem-based service engineering delivers all these benefits available to ecosystem members and also facilitates the utilization of open data in digital service engineering.

Open data is based on the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents, or other mechanisms of control (Auer et al. 2007). The open data concept has evolved over the 10 years since its first definitions. The tendency in many countries has been to open the administrative data (Poikola et al. 2011), and several local and global open data portals already exist that help people to create and share data and knowledge. Open data typically originates from enormous amounts of different kinds of sources, and it can be structured (with a strict data model), semi-structured (with an evolving data model), or unstructured (not associated with any data model). The utilization of this kind of data requires knowledge about its provenance, quality, and trustworthiness to ensure that the data is what it is expected to be. Data quality can be defined as data that is fit for use by data consumers (Wang and Strong 1996). The evaluation of data quality is challenging due to the facts that there are no agreed definitions of quality attributes, and the data quality cannot be judged without considering the context at hand (Nurse et al. 2011). The growing amount of semi- and unstructured data, new ways of delivering information, and users' changed expectations and perceptions of data quality (Madnick et al. 2009) further provide new challenges in data quality evaluation. At the same time, this dictates that new quality evaluation means and methods are required to verify the quality of open data. The importance of quality evaluation is emphasized in a digital service ecosystem, where poor quality of data affects several digital services and, in that way, the whole trustworthiness of the ecosystem.

Therefore, in the digital service ecosystem, data quality evaluation should be one of the key activities supported by the ecosystem's assets. The main question is how the ecosystem can guarantee the quality of open data.

According to the survey on the state of the practice among industry (Immonen et al. 2014; Immonen et al. 2013), quality assurance of open data is the biggest obstacle for its exploitation in digital service development. The contribution of this research is to specify the concept of an open data based digital service ecosystem (called the EODE concept), in which the ecosystem ensures the quality of open data utilized in digital services. In this approach, open data is provided as a service for the ecosystem's usage. The purpose of the concept is to verify the trustworthiness and the quality of open data, thus, to ensure that the data comes from reliable sources, and its quality is good enough to be accepted for the usage of the ecosystem's services. The members of the ecosystem do not have to be familiar with the metrics or techniques for data quality evaluation, but the ecosystem is responsible for certifying the quality of data that can be then utilized by the ecosystem members. The EODE concept includes the ecosystem's capability model with activities for the quality evaluation of the open data source, open data itself, and open data services, and it provides knowledge management models and ecosystem support services to enable these activities. The EODE supports the businesses of both the open data providers and the digital service providers. The open data providers reach more users (and thus more income) for the data when they pay more attention to data quality; data with poor quality is not selected for the ecosystem. The service providers receive more satisfied consumers when they provide trustworthy data via digital services. The ecosystem also provides other benefits to its members, such as finding partners and customers, and ways to deliver services and data. In addition, the EODE provides the possibilities and business potential for other support service providers as well, such as for analysis and monitoring service providers.

The EODE concept includes a quality certification process for open data, which specifies how the knowledge and support services of the ecosystem are utilized to carry out the quality evaluation. Thus, the process is a kind of instantiation and a guideline of the knowledge and the support services necessary to implement the quality certification of open data. Data certification contains several aspects, such as legal, practical, technical, and social aspects. Thus, besides data quality, data privacy, availability, and licensing aspects must also be considered when making decisions to accept the data for usage. However, data quality, i.e., the ability of the data to be fit for use by data consumers (Wang and Strong 1996), is the first aspect that must be ensured. If the quality is not good enough, there is no need to evaluate the other aspects. Therefore, this research concentrates purely on the technical quality aspects of open data. The quality certification process enables bringing open data to the ecosystem, transforming it to a usable form for the ecosystem, validating it against its intended usage, monitoring the data sources and the usage of the data, and continuously evaluating the quantified value of the open data service, thus, certifying the quality of the data for the ecosystem and its members.

This paper is organized according to the following: Section 2 presents the background for this research; the basic terminology is first defined, after which our earlier research on open data based business ecosystems, the quality evaluation of open data, and service engineering in ecosystems are presented. These are used as the basic and starting point for this research and are combined and refined to form a full open data based service

ecosystem concept. Studies related to this research are presented to understand the shortage to which this paper tends to respond, including concepts of the open data ecosystem (and the current status and development of open data), the quality evaluation of open data, and ecosystem-based digital service engineering. Finally, Section 3 introduces the concept of the EODE; Evolvable Open Data based digital service Ecosystem. The EODE is represented from two viewpoints: the ecosystem and the service providers. Section 4 introduces how the elements of EODE are utilized to implement the open data quality certification process. The process consists of five validation phases with related activities, required support services, knowledge assets and related evaluation targets, quality attributes, and metrics. Section 5 presents the analyses and discussion, consisting of the current validation of the concepts of the EODE, and limitations, open issues and future research targets. Finally, Section 6 presents conclusions drawn.

2 Background and related works

2.1 Terminology

The following terminology is used in this paper:

Data—Data that is produced by observing, monitoring, or using questionnaires, but has not yet been processed for any specific purpose.

Open data—Data that it is freely available to everyone to use and republish as they wish, without restrictions of copyrights, patents, or other mechanisms of control (Auer et al. 2007).

Information—Data that is refined and processed for assigning meaning to the data (Chen et al. 2009).

Quality of data—Data that is fit for use by data consumers (Wang and Strong 1996).

Data quality certification—Confirmation that the open data is trustworthy, and its quality has been verified according to strict quality policies.

Metadata—A standardized way to describe the semantics of data.

Policy—A collection of alternative tasks and rules, each of which represent a requirement, capability, or other property of behavior (W3C 2007).

Ecosystem policy—Description of the principles, strategies, tactics, and guidelines of the ecosystem that are common to all ecosystem members.

Organizational data policy—Description of the principles and guidelines required to effectively manage and exploit the data/information resources of a company.

Open data ecosystem—A free-formed community of organizations each of which have their own part and know-how in the data-based business.

Open data service—A service that encapsulates the open data, providing the open data as a service.

Digital service—A service that utilizes the open data, is entirely automated, and can be anything that can be delivered through an information infrastructure, e.g., web, mobile devices, or any other forms of delivery.

Digital service ecosystem—An open, loosely coupled, domain-clustered, demand-driven, self-organizing environment, in which digital services are created in value networks under the common ecosystem regulation.

2.2 Our earlier studies as a starting point for the research

The earlier studies by the authors are used as the basis and starting point for this research and are therefore presented in the following sub-sections.

2.2.1 Motives for the research

While examining the usage of open data in Finland, industry interviews were performed in 2013 pertaining to open data in business (Immonen et al. 2014; Immonen et al. 2013). It was discovered that there exists huge interest in open data and its exploitation in business. However, serious barriers were found to exist that prevent the fluent utilization of open data. These concern the lack of a standard description of data sources and APIs, as well as the uniform format for the data. Furthermore, the management of data privacy and varying licensing conditions and data quality were seen as highly important issues but have not been solved yet. However, the low quality of data and changes in data quality were seen as risks that complicate or even prevent the open data utilization in business (Immonen et al. 2014). In Immonen et al. (2014), the data broker actor is defined to include the role of data promoters that maintain "a list" of available data in the ecosystem and the quality of data, price, applied licenses, etc. Since the quality of the data was detected to be unknown, a need was identified for a data quality verification service in the ecosystem.

Now, two and a half years after the first interviews, new interviews were performed among the same industry representatives. It was detected that although data quality was seen as highly important, no significant progress had occurred in 2 years. The companies conceded that interest in open data and its exploitation in business still exists, and they have also recognized that the demand for open data from authorities, companies, and individuals has increased. Moreover, opening of data is also done in a smaller scope: contract-based exchange of data between companies is seen as a working collaboration model. The same main challenges remains: (1) the data that the companies are interested in is not available, (2) the data is not free of charge, and (3) there is uncertainty about the quality of the available data. Thus, the problems are related both to business and to used technology and raise the following questions: (i) What business reasons are there to produce open data and how can it be marketed? (ii) How can the use of open data be made profitable in service development? In summary, the use of open data has slowly progressed, nevertheless, several obstacles remain that must be removed. And, quality assurance of open data is the greatest obstacle to its exploitation in digital service development.

2.2.2 Ecosystem actors

In Immonen et al. (2014), the actors and their roles in the open data ecosystem from the business viewpoint are defined. These actors include the following: (1) data providers make data available to other stakeholders; (2) data brokers promote the data in the ecosystem, distribute it through the communication channels, and match the demanded and provided data; (3) service providers produce supporting services related to the data to be utilized in applications; (4) Application developers use the available data and services and develop applications for the data; (5) Application users are the data end-users that consume the data and services with the help of applications; and (6) infrastructure and tool providers provide utility services

to all the actors so they are able to act in the ecosystem. Furthermore, in this research on digital service ecosystem (Immonen et al. 2015b) defined the actors of the digital service ecosystem are defined from the service engineering viewpoint. These include the following: service providers that provide digital services to be used by other ecosystem members or consumers, service brokers that promote and deliver the services and match the demand with the best available services, service consumers that are the actual users of the services, and infrastructure providers that provide the utilities for acting in the ecosystem.

Open data-based digital service ecosystem merges and refines the actors both from the open data ecosystem and service ecosystem (see Fig. 1). The roles of infrastructure and tool providers remain the same. Open data service providers encapsulate the open data and provide the data as utility services, thus enabling the utilization of the open data in digital services. The data owner has data sovereignty and, thus, specifies the terms and conditions of use of the data (Boris et al. 2016). Digital service providers provide digital services that utilize the open data. Ecosystem support service providers provide services that support extracting, monitoring, and evaluating the data and, thus, assist in managing open data and its quality in the ecosystem. Finally, the digital service consumers utilize the data with the help of digital services.

2.2.3 Ecosystem capability and infrastructure

In Immonen et al. (2015b), the elements of the digital service ecosystem that influence service engineering in the ecosystem (see Fig. 2) are defined. The main elements, ecosystem members, infrastructure, capabilities, and digital services, are classified according to (Ruokolainen 2013). The capability of the ecosystem defines the properties of the ecosystem and how these are implemented using the infrastructure services (Immonen et al. 2015b). Thus, the capabilities define the purpose of the ecosystem, its ability to perform actions, and the rules for how to operate in the ecosystem. The actions and rules address the following:

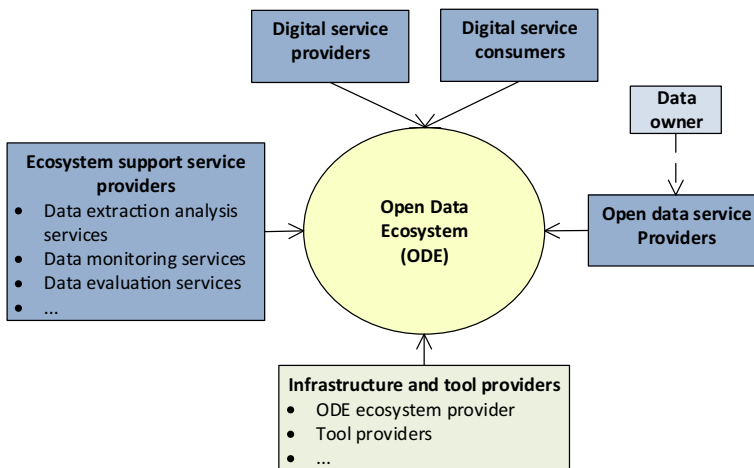1. Governance and regulation actions of the ecosystem (Immonen et al. 2015b) for

Fig. 1 The actor roles in an open data-based digital service ecosystem

- Directing, monitoring, and managing the ecosystem: these include, for example, rules of trusted collaboration establishment, interactions rules, and how to join and leave the ecosystem
- Directing and managing service engineering: these include, for example, rules for describing and delivering services and managing knowledge.

2. Service engineering-related actions

- Provide reusable assets for defining requirements (both functional and quality)
- Assist in the matchmaking of services
- Provide reusable assets for quality requirements specification, quality modeling, and quality evaluation of digital services

The infrastructure of the digital service ecosystem provides the knowledge models and services for implementing the ecosystem's capabilities. These include the following (Immonen et al. 2015b):

- A domain model: describes the concepts of the domain, their relations with each other, e.g., domain-specific quality attributes, rules, and policies
- A knowledge management model: describes the knowledge, know-how, and assets of the ecosystem
- A service engineering model: describes how the services are co-innovated and co-engineered in the ecosystem
- Ecosystem support services: implement the actions of the capability model

The elements described in (Immonen et al. 2015b) do not consider open data and the quality of data in digital service engineering. In this study, the capability model has been refined to include the open data related actions. The focus will be on ecosystem
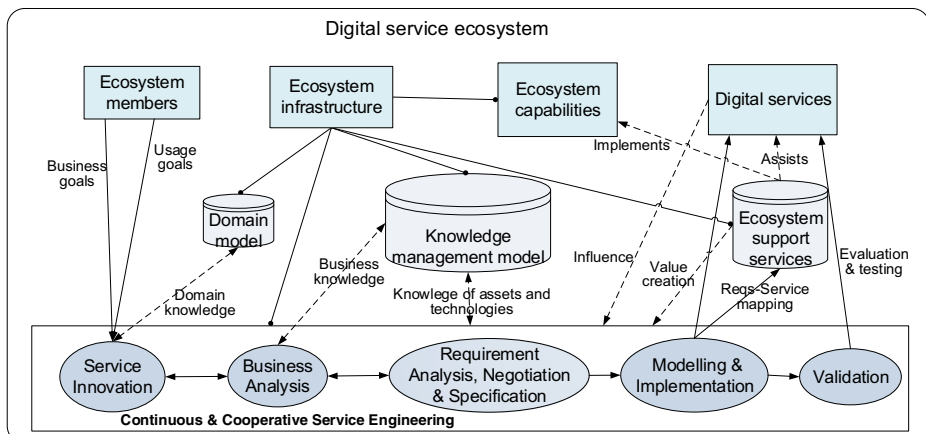

Fig. 2 Continuous cooperative service engineering in a digital service ecosystem (simplified from (Immonen et al. 2015b))

infrastructure and capabilities, including a domain model, a knowledge management model, and ecosystem support services. The content of these elements have been refined to include the activities, models, and services necessary to certify the open data and enable data utilization in service engineering in the ecosystem. Certification in this research means purely data quality certification; the other issues that concern assessing the extent of the open data (HM Government Cabinet Office 2012), such as access, licenses, and privacy, are beyond the scope of the present research.

2.2.4 Quality evaluation of open data

In Immonen et al. (2015a), the elements and phases of open data (social media data) quality evaluation in big data architecture are defined. The data is evaluated in data extraction, processing, and analysis phases with the help of organizational policies, and, finally, its value in decision-making is evaluated using a decision-making policy. The data is managed with the help of metadata, which is again managed utilizing the metadata management component/element included in the big data architecture.

The paper follows the five-star scheme (HM Government Cabinet Office 2012) and goes beyond that by defining "open data services" that make the data available for different service developers that want to utilize the data. Earlier work has focused on data quality and quality evaluation inside one company; it has not considered the ecosystem context. In this work, the purpose is to present how to validate the quality of data in the context of the digital service ecosystem. Therefore, the term "organizational policy" is not used, but it is distributed into the data filtering policy and evaluation policy that manage the data quality certification in the ecosystem. The aim is to obey the earlier definitions of open data (Fig. 3) and go further by providing the support services that are the first movement towards automated quality certification of open data in the context of digital service ecosystem (the green "future" box in Fig. 3).
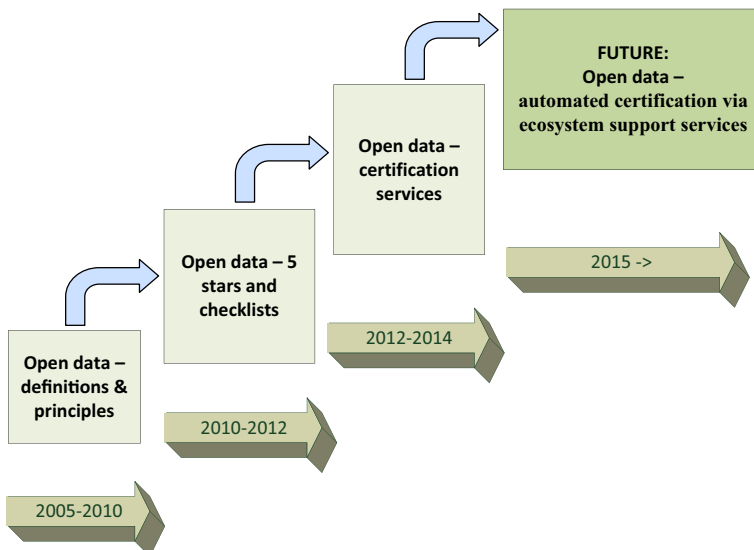

Fig. 3 Open data development stairs

2.3 Related research

2.3.1 Concepts of open data ecosystems

The concept of "open data" most notably has its roots in Great Britain, which has advanced the Open Government Data ecosystem over the past 15 years. The major breakthrough in the era of open data was in 2009 when both Great Britain and the USA launched their first data portals. Since then, the tendency in many countries has been to obligate to open the data of the public sector collected along with tax revenues. There exist many foundations and initiatives that "push" organizations to open their data, such as the Open Knowledge Foundation,[1] the Open Data Institute (ODI),[2] the Global Open Data Initiative,[3] and the INSPIRE[4] directive of the European Union. However, the "pull" mode has received less attention. Therefore, the data holders do not know the demand for the data that they own, or the possibilities that their data would provide to some other stakeholders. Some attempts already exist that tend to untangle the demand for data that is not yet opened. For example, some local groups in Finland (e.g., Helsinki and Oulu) provide the potential for organizations, companies, and individuals to demand data to be opened. They also allow users to provide feedback about the data that is already open. Thus, they meet the cyclical characteristics (Pollock 2011; Sande et al. 2013) of an open data ecosystem. In addition to data from government, institutions, and private companies, recently, different forms of social media, such as Twitter, Facebook, or Instagram, provide more and more data available online. This kind of social media data is obviously open as such, as it is based on free-formed conversations or other volunteer releases, both from communities or individuals. Due to the continuous growth of the usage of social media and the different yet increasing social media forms, the amount of this "big data" is rapidly growing. This data may not have a rational or an organized structure when compared with organizational data, but when properly treated, it can be valuable in several ways.

Open data is the main resource of the open data ecosystem. Open data and its definitions have evolved from basic definitions and principles via classifications and other kind of checklists for open data certification services (Heimstädt et al. 2014a). Figure 3 illustrates the development stairs of open data. From the first definitions, it took about 5 years before the reusability of data from the user perspective were considered. For example, the government of Great Britain proposed in 2012 a five-star scheme for assessing the degree to which the individual datasets are reusable (HM Government Cabinet Office 2012): 1 star: the data is available on the web in any format, 2 stars: the data is available in a structured format, 3 stars: the data is available in an open, non-proprietary format, 4 stars: Uniform Resource Locators (URIs) are used to identify the data using open standards and recommendations from W3C, and 5 stars: the data is linked to other people's data to provide content. A few years after that, data certification approaches emerged. For example, the Open Data Institute (ODI) provided Open Data Certificates[5] that enabled data providers to assess the extent to which open data is published according to recognized best practices. The certificate tells data users what the data is about and how to get hold of it, sharing legal (e.g., licensing, privacy), practical (e.g., discovery),

---

[1] https://okfn.org/
[2] http://theodi.org/
[3] http://globalopendatainitiative.org/
[4] http://inspire.jrc.ec.europa.eu/
[5] https://certificates.theodi.org

technical (e.g., structure, quality), and social (e.g., documentation) information. In the future, digital services will be able to automatically certificate open data.

Generally, an open data ecosystem consists of actors, i.e., the organizations and individuals with the roles of data suppliers/providers, data intermediaries, and data consumers (Heimstädt et al. 2014b). In addition to data and actors, the existing literature contextualizes open data ecosystems according to the following characteristics (Heimstädt et al. 2014b):

- Nested structure: The data ecosystem has a nested structure with micro, meso, and macro levels.
- Cyclical: After the data has been released, data consumers are able to view the data, edit, and update it and also contribute to it and provide their feedback (Pollock 2011; Sande et al. 2013).
- Demand-driven: The ecosystems are formed in response to the demand for data (Boley and Chang 2007).
- Sustainable: The ecosystem finds ways to emerge in the event of sudden changes (Boley and Chang 2007).

These characteristics are also essential for open data-based digital service ecosystems. The digital service ecosystems can exist on micro, meso, and macro levels, depending on the size and the amount of the value networks, and the size and scope of the provided digital services. The digital service ecosystems also implement a cyclical structure and data cycles, which enable data consumers to act as data providers, and vice versa. In a digital service ecosystem, the actors cooperate to fulfill a certain demand, and, thus, the ecosystem is demand-driven. Finally, the digital service ecosystem finds a new balance and substitutes in the event of changes. For example, new partners and data providers are sought in the case when certain data is no longer provided as open.

2.3.2 Quality of open data

A lot of work has been done to standardize quality attributes in the field of software engineering (ISO/IEC 2001; ISO/IEC 2003) and software architecture design (Gorton and Klein 2015; Immonen and Niemelä 2008; Ovaska et al. 2010; Niemelä and Immonen 2007; Kazman et al. 2000; Dobrica and Niemelä 2002). Although data quality has been the subject of several studies (Castillo et al. 2011; Agichtein et al. 2008; Gil and Artz 2007; Dai et al. 2008; Naumann and Rolker 2000; Nurse et al. 2013), the quality issues are not commonly brought into use in the case of data. The ISO 25012 data quality model (ISO 2008) defines 15 data quality attributes and classifies them into inherent quality and system-dependent quality. Some of the existing research on data quality uses the quality model as a basis, such as (Behkamal et al. 2014; Rafique et al. 2012). Data quality evaluation is challenging because data quality cannot be judged without considering the context or situation at hand (Nurse et al. 2011; Bizer 2007; Bizer and Cyganiak 2009). At this moment, there are neither agreed classifications for the applicability of quality attributes to certain contexts nor are there agreed definitions of quality attributes themselves. Quality assessment metrics are heuristics and are designed to fit a specific assessment situation (Bizer 2007; Pipino et al. 2005). Recently, the characteristics of big data, volume, variety, velocity, and veracity, have also been detected to define new challenges for data quality and data quality assessment (Ferrando-LIopis et al. 2013; Cai and Zhu 2015). Recent research on the quality of online data can be summarized

under three main factors (Nurse et al. 2011): (1) provenance factors refer to the source of information, (2) quality factors reflect how an information object fits its intended use, and (3) trustworthiness factors influence how end-users make decisions regarding the trust in the information.

The availability of the information in a machine-readable format with the commonly agreed metadata facilitates data cross-reference and interoperability and, therefore, considerably enhances the value of information for reuse (European Commission 2011). For example, data.gov.uk already includes basic metadata about all its data sets (HM Government Cabinet Office 2012). Currently, there are some de-facto standards for metadata, such as the Dublin Core Metadata Element Set (http://dublincore.org/) and the metadata of the CKAN data portal platform (http://ckan.org/). However, recent metadata standards do not assist in determining the quality of data from the data end-user's viewpoint. Different parties use different, informal ways to ensure the quality of data. For example, the ODI's Open Data Certificates rely on the data providers' assessment, enabling the users to decide how much to rely on the data.

2.3.3 Ecosystem-based digital service engineering

The digital service ecosystem takes characteristics both from business ecosystems (Zhang and Fan 2010; Li and Fan 2011; Iansiti and Levien 2004) and software ecosystems (Bosch 2009; Jansen and Cusumano 2012; Hanssen and Dybå 2012). However, in a digital service ecosystem, the service provider shares the service taxonomy and service descriptions that enable the dynamic, behavioral, and conceptual interoperability and interactions between services (Immonen et al. 2015b; Pantsar-Syväniemi et al. 2012). Just like in a business ecosystem, the members of a digital service ecosystem share the common ecosystem regulations but are able to act independently. Partner networks are created inside both ecosystems, but there are also dependencies between the digital service ecosystem members other than business dependences. Unlike in software ecosystems, in a digital service ecosystem, the members are not bound to a shared development platform or technology. However, the software can be provided as a service to the ecosystem.

Service engineering in the digital service ecosystem can be characterized according to the following features (Immonen et al. 2015b):

- Service co-innovation: open innovation enables the potential to co-create ideas for a service with other actors of other ecosystems (Stathel et al. 2008; Chan 2013; Chesbrough and Appleyard 2007).
- Service value co-creation: the value is created inside the ecosystem in value networks formed by the ecosystem members (Stathel et al. 2008; Wiesner et al. 2012) (Kett et al. 2008).
- Enabling infrastructure: the ecosystem infrastructure supports the collaboration and cooperation of ecosystem members, providing the required services and tools (Pantsar-Syväniemi et al. 2012) (Khriyenko 2012; Ruokolainen et al. 2011; Ruokolainen and Kutvonen 2009).
- Utilization of the ecosystem's assets: the existing ecosystem assets, such as the ecosystem's rules, methods, and practices for service engineering, enable co-innovation and co-creation of the services (Pantsar-Syväniemi et al. 2012; Ovaska et al. 2012; Ovaska and Kuusijärvi 2014).

Although several methods and approaches exist that take into account some of the previous features, they do not cover all of them but concentrate on their own viewpoint and not working together. Furthermore, recent approaches to ecosystem-based service engineering do not take into account the data and data quality.

3 Evolvable open data-based digital service ecosystem

This section combines and refines the earlier work of the authors on open data based business ecosystems (Immonen et al. 2014; Immonen et al. 2013), digital service ecosystems (Immonen et al. 2015b), and the quality evaluation of open data (Immonen et al. 2015a) (see Section 2.2), and it introduces the main concepts of the Evolvable Open Data based digital service Ecosystem (EODE). Interesting and certified open data is a key enabler in the EODE. Data quality certification ensures that the quality of data is verified to be good enough for the usage of the ecosystem's services. Thus, the certified data provides added value for the whole ecosystem, its members, and customers through digital services co-created based on that open data.

Figure 4 introduces the structure and the elements of the EODE; the models and services required for establishing and operating open data based service engineering (vs. Fig. 2 in Section 2). In this work, these models and services are inspected from the viewpoint of the quality of data. The term "evolvable" refers to the abilities of the digital service ecosystem to be long-lasting and to tolerate internal and external changes; the ecosystem introduces and activates survival actions based on up-to-date knowledge and support services that exploit the knowledge to adapt digital service engineering models and practices to the present situation. The EODE core illustrates a service framework that is a common infrastructure for coordinating and managing the operation of the EODE. Thus, the core contains all the mechanisms for controlling the ecosystem, including legal, practical, technical, and social aspects. In this context, the focus is on the quality certification of open data, i.e., what kinds of knowledge and support services are required from the ecosystem to ensure the quality of open data and open data services. Although the main focus is on open data and open data services, there is a brief discussion of how open data services are exploited in digital service engineering.
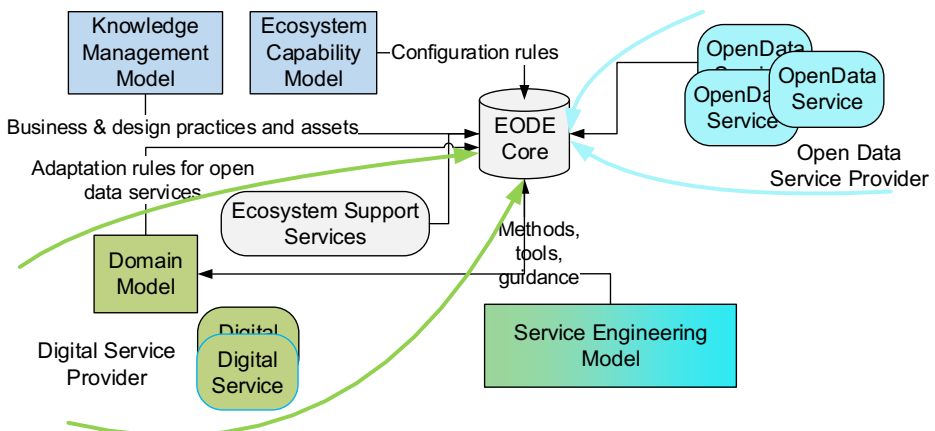

Fig. 4 Overview of an open data-based digital service ecosystem

The content of an open data based service ecosystem is specified from two viewpoints:

- *Ecosystem viewpoint* defines governance- and regulation-related actions for (a) acting in an ecosystem, (b) evaluating and monitoring the quality of open data and open data services, and (c) developing digital services on the basis of open data services. The ecosystem viewpoint has been established on the following artifacts (Fig. 4): the ecosystem capability model, the knowledge management model (KMM), ecosystem support services, and the EODE core that integrates the models and support services. The ecosystem viewpoint defines and describes how collaboration among ecosystem members is regulated, guided, and assisted. The goal is smooth collaboration among the ecosystem's members.
- *Service provider's viewpoint* defines two different viewpoints: An open data service provider's viewpoint and a digital service provider's viewpoint. The first viewpoint explains how the EODE helps open data providers to create proper open data services with the required quality. The viewpoint needs the following artifacts: ecosystem support services, the KMM, and the EODE core. The main goal is trust making among open data (service) providers and digital service providers. The digital service provider's viewpoint describes how open data services are exploited in digital service engineering. The focus is on quality evaluation of the used open data services and the digital service under development. The viewpoint exploits open data services, the ecosystem support services, the domain model, the KMM, the EODE core, and the service engineering model. The outcome is a new digital service. The main goal is to provide (personalized) digital services of high-quality.

3.1 Ecosystem viewpoint: models and services for operating in the ecosystem

This section describes the models and services common for all ecosystem members. These include the capability model, the KMM, support services, and the EODE core from Fig. 4. Figure 5 illustrates the relationships of these elements; the ecosystem support services and the knowledge management model implement the ecosystem capability model, and these are provided as services to the ecosystem through the EODE core. These elements are described in the following sub-sections from the viewpoint of the quality of open data.

3.1.1 Capability model

The capability model defines the purpose of the ecosystem, its ability to perform actions, and the rules governing how to operate in the ecosystem. The capabilities define the governance activities and regulations for directing, monitoring, and managing the ecosystem, and the activities and regulations for open data certification (including quality, availability, privacy,
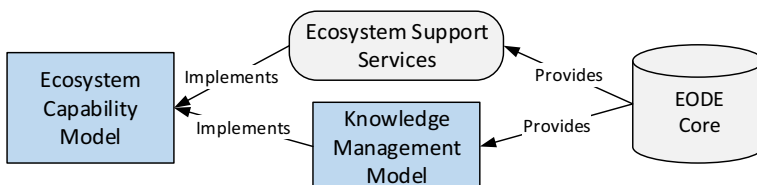

Fig. 5 The relationships of the common elements of the ecosystem

licensing) and digital service development. In this context, these capabilities will be examined from the quality evaluation of the open data and open data services perspectives.

The EODE supports community-based cooperation and collaboration among ecosystem members by providing service engineering facilities for open data service providers and digital service providers. The capabilities are implemented in the form of actions that, in EODE, are clustered according to the stakeholders' activities into three categories:

i. Quality-related activities for governance and regulation actions of the ecosystem for

- Finding reliable and trusted data sources/service providers/ecosystem members
- Contract making with ecosystem members
- The SLA specification of open data service providers and digital service providers. SLA defines the kinds of tactical rules that are used for quality evaluation. Tactical rules depend on the member's role in the ecosystem
- Supporting the bi-directional communication between digital service providers and open data service providers
- Defining an ecosystem policy that is to be followed by the ecosystem members. The ecosystem policy defines strategic evaluation regulations of the ecosystem. Examples of strategic evaluation rules are the quality criteria for open data sources
- Offering standard quality evaluation practices both for open data providers and digital service providers
- SLA contract making with open data service providers and digital service providers, i.e., defining the criteria for tactical quality evaluation
- Marketing open data services and digital services of the ecosystem

ii. Quality-related activities for open data certification that

- Find acceptable open data sources
- Extract data from different types of data sources
- Check the syntax and semantics of open data and transform them to a standard format
- Enable the quality evaluation of open data services
- Change quality policies based on changes on open data sources and/or (the quality of) open data
- Provide certification of the quality of open data services.

iii. Quality-related activities for service engineering-related actions that

- Provide reusable assets for defining data requirements with the required data quality
- Assist in matching required data quality with the provided data quality of open data services
- Provide reusable assets for quality requirements specification, quality modeling, and quality evaluation of digital services
- Test the digital services with the EODE service architecture specification
- Certify the digital services

The rest of this section concentrates on the activities of the second category since these activities guide the definition of the KMMs and support services required for achieving

certified open data to be used by the ecosystem members. The activities of the first and the third categories also influence the content of KMMs and support services and are, therefore, briefly discussed.

3.1.2 Knowledge management model (KMM)

The KMM includes common models and transformation rules for adapting specific data models to the common ones shared and accepted among ecosystem members. These models can include metadata models of (open) data, standard data models of specific application domains, and rules for how some specific data models can be adapted from a domain-specific data model to the common data model. The KMM includes the following types of quality-related knowledge:

- Ontologies that conceptualize the things related to data, quality, metrics, and services. The quality attribute ontology, e.g., reliability ontology (Zhou et al. 2011), defines the sub-characteristics of the quality attribute, metrics (García et al. 2006) for each sub-characteristic, application time, the formula used as a measuring method, value range, and target value (Immonen et al. 2015a; Niemelä et al. 2008). Context ontologies are required to identify the situation of the digital service and to carry out the situation-based service adaptation of that digital service (Pantsar-Syväniemi et al. 2011). Rules can be represented as ontologies as well.
- Design-time artifacts, i.e., architectural styles and patterns (Ovaska et al. 2010; Ovaska and Kuusijärvi 2014). It is also possible to use ontology orientation to represent the concepts of architectural descriptions and styles. In (Guessi et al. 2015), the ISO/IEC/IEEE 42010 standard of an architectural description is formalized and described as an ontology model, and further specialized to SOA architecture. Thus, the assumption is that integration architecture is represented as a common knowledge model shared among ecosystem members. Other common knowledge models may include service description ontologies, service component models, quality of service models, service composition models, and service community models (Aubonnet et al. 2015).
- Domain models that define domain-specific quality attributes, variations between the domain and the common model, and the adaptation rules for mapping the variable things to the context of the EODE.
- Policies used in quality evaluation and management (Bizer 2007; Rahman et al. 2011; Bertino and Lim 2010). The ecosystem policy defines a set of governance services that are common for all ecosystem members, and rules for how to configure and monitor these services. It also defines how SLAs for service providers are specified, configured, monitored, and adapted. Each SLA follows the same quality evaluation policy but is configured and adapted according to the service provider and the context of the used digital service and its user(s). Moreover, ecosystem policy also manages the following data quality policies:

– Data filtering policy: This policy defines which open data sources are acceptable in the ecosystem. The data sources must fulfill the quality criteria of the ecosystem.
– Data quality evaluation policy: This policy defines the quality attributes, metrics, and rules for their applicability for quality evaluation. The policy is used for data quality evaluation of the ecosystem, but it is also configurable for the specific need of each service provider.

– Decision-making policy: This policy defines how decisions are made based on the strategic or tactical operation of the ecosystem. Input for strategic decision-making is collected with the identification and analysis of changes in the ecosystem and its surroundings, e.g., related open data markets. The tactical operation of the ecosystem needs different kinds of decision-making: e.g., defining and updating actors' roles, business models, and value networks. Service engineers need online guidance realized by means of semantic wikis, a well-defined model-driven engineering environment, and continuous synchronization between the wiki and MDE-based models (Baroni et al. 2014).

3.1.3 Ecosystem support services

The purpose of the ecosystem support services is to assist in carrying out the tasks defined as activities in Section 3.1.1. The support services that evaluate the quality of open data services are common for all members. Support services provided for open data service development and digital service development are recommended, but member-specific solutions are also allowed. In that case, they need to be adapted to work in a way specified in the knowledge management model and service engineering model. In this context, the focus will be on the quality-related activities that boost open data service development and ensure the quality of open data and open data services (Section 3.1.1.ii):

A1: Defining acceptable open data sources—requires services for searching and evaluating the quality of open data sources, and monitoring the quality of data sources accepted to the ecosystem; thus ensuring that the quality remains as acceptable.

A2: Extracting data from different types of data sources—requires services for monitoring the quality of open data sources and open data, ensuring that the quality remains as acceptable.

A3: Checking syntax and semantics of open data and transforming the open data to a standard format—requires services to ensure that the data is syntactically and semantically straight, and to transform the data into the format acceptable to the ecosystem.

A4: Enabling quality evaluation of open data services—requires tailorable services that enable the evaluation of the quality of the open data service in its usage context against the required quality.

A5: Changing quality policies based on the changes of open data sources and/or (quality of) open data—requires services that enable the detection of different contexts and changes, and adapt the models and support service to the changes or in a situation-based manner.

A6: Certification of the quality of open data services—requires services to enable the validation of the open data service in the usage of the ecosystem.

The initial taxonomy of ecosystem support services (Fig. 6) defines an evolving set of services selected for the common use of ecosystem members. The ecosystem support services include eight main categories; utility services enable the management of the ecosystem policy and also the policies for data filtering, data quality, and decision-making. Matchmaking services assist in verifying the syntax and semantics of the data, thus, ensuring that the data is in the right form and is usable for the ecosystem members. Monitoring and evaluation services monitor the open data sources, the open data itself and the open data services, and
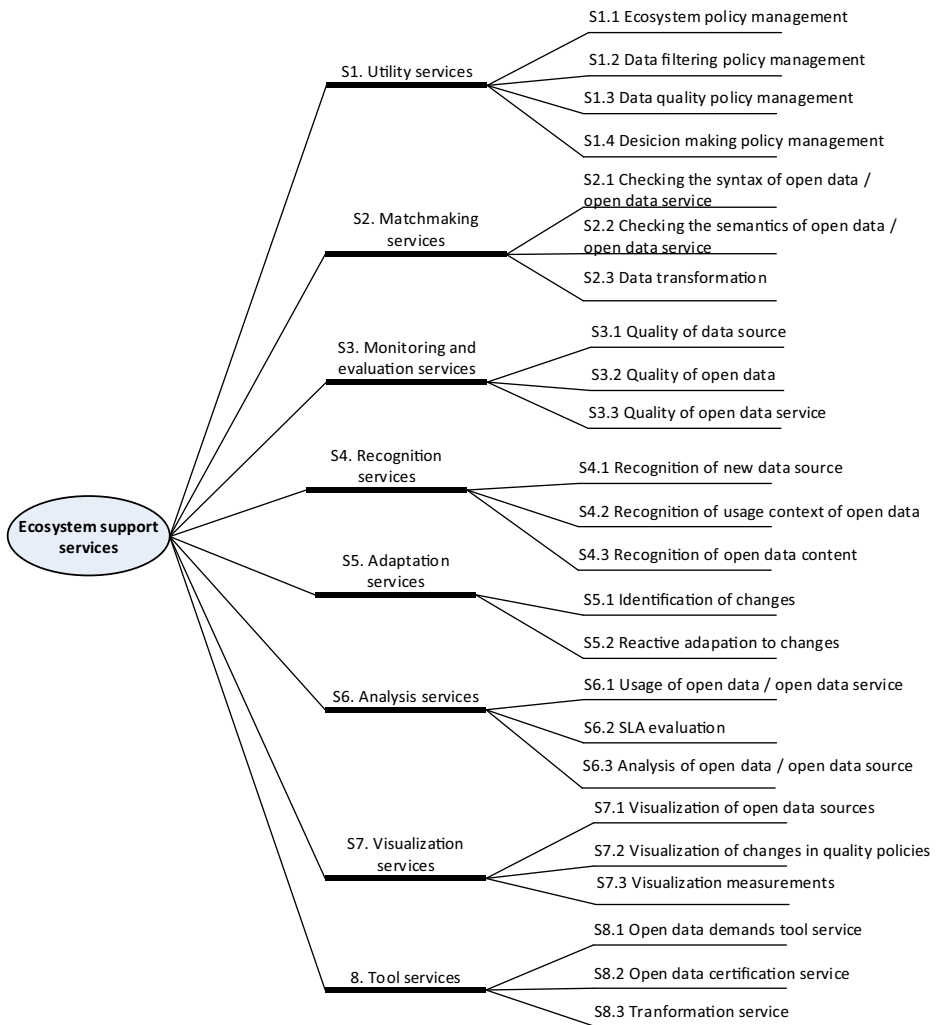
Fig. 6 The taxonomy of ecosystem support services

detect changes in their quality. Recognition services recognize the changes in the context (e.g., a new data source or changed usage context). Adaptation services adapt to the recognized changes according to policies. Analysis services perform the data quality evaluation according to the quality policies and also evaluate the SLAs between the open data service provider and the digital service provider. Visualization services provide views of the open data and open data services. Finally, tool services assist in all activities of the ecosystem members.

3.1.4 EODE core

The EODE core is an integration framework for combining models and support services for developing open data services and digital services based on them. The integration framework registers open data services, digital services, and support services and provides knowledge

management, service engineering, and domain models also as services. Thus, in addition to being an integration framework, the EODE core also acts as a means of knowledge sharing. The core also provides mechanisms to control and manage open data certification; in this context, the focus will be only on quality certification.

The core is a centralized system for maintaining a list or catalogue of the digital services of the ecosystem and additional information, such as service user feedback and rating, access management, availability information, and service logging. Service registration (see Fig. 7) is a process in which the necessary information for using and discovering the service is published in a uniform way. First, the service provider registers the services and receives a unique ID (within this registry) for the service. Second, the service provider adds the required service descriptions. At the end, the URLs of service endpoints are linked with the service resource description. Service discovery (see Fig. 7) is based on the registered service descriptions. Basic service discovery is enabled by the human readable service description and additional information associated with the service description. For more intelligent service discovery and, in particular, intelligent service matching, a semantic service data description is required. Semantically enriched descriptions support (i) multilingual searches, (ii) matching different data elements that describe the same thing, and (iii) using the relations of data elements in searches. Semantics also support interoperability between different services. These are discussed in more detail in Section 4.2.

3.2 Service providers' viewpoint: models and services for collaboration

Two main types of service providers collaborate and co-create in the evolvable open data-based digital service ecosystem: open data service providers and digital service providers. The EODE forms a two-directional communication channel between open data providers and digital service providers (Fig. 8). Trust making among service providers is supported by a common open data service model and quality assessment services provided as ecosystem support services. Due to the common knowledge management model, the service engineering model and ecosystem support services, the collaboration among diverse actors is smooth and interoperable at the levels of business, technology, and processes.

Open data service providers encapsulate their offerings (open data) with the service model and the quality policy adapted according to the current situation, thus, utilizing the knowledge management models of the ecosystem. Ecosystem support services are used in service development and in ensuring that the service is interoperable by the ecosystem members.
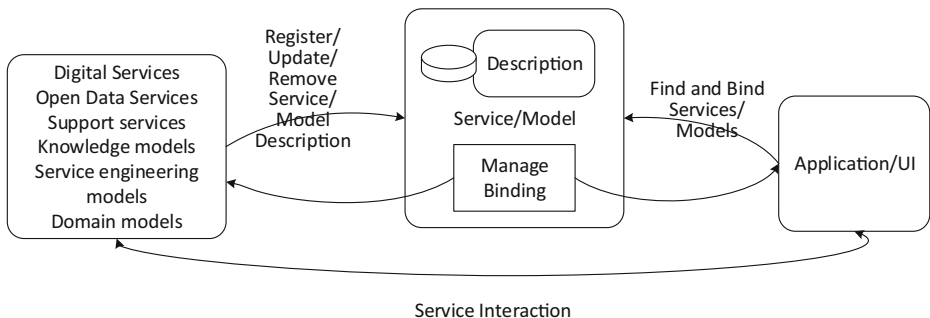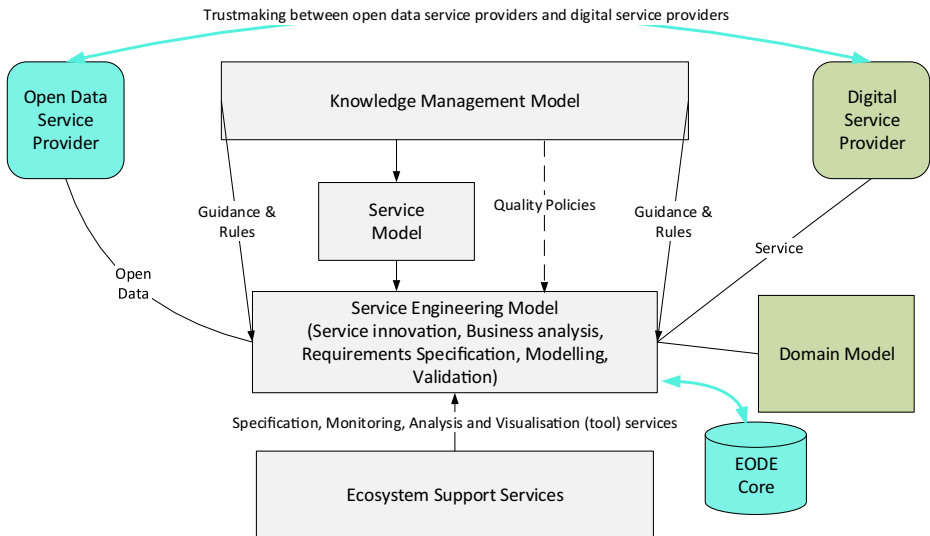

Fig. 7 Digital service framework

Fig. 8 Collaboration between open data service providers and digital service providers

These new open data services are provided for markets (i.e., to ecosystem members and outsiders) through the EODE core.

Digital service providers use open data services as building blocks in digital service development, and provide digital services that can be (1) domain-specific services to global markets, (2) support services to ecosystems, or (3) tool services or technology enablers for open data providers. Digital service providers utilize the support services and knowledge management models in their service development activities, and they utilize the EODE core for searching for applicable data and for registering the digital service into the EODE core registry, from where it can be searched.

Next, the service model, domain model, and service engineering model are introduced.

3.2.1 EODE service model

Open data providers encapsulate the open data and provide it as an open data service with a standard service interface, including syntactical and semantic definitions. Each open data provider can have their own data model or they can utilize the common EODE service model. The open data service interface must be implemented as a common standard, such as REST-API or/and SOAP interface.

A generic service model is defined for all kinds of digital services. The KMM can include several digital service models. A service provider can also utilize their own service model. In that case, (being an accepted member of the ecosystem), this service model can also be included as an acceptable service model for the ecosystem. The digital service model defines a common digital service interface that includes

- Interface description according to the selected architecture style, e.g., as a REST-API
- Service capabilities as a service ontology, e.g., (Kantorovitch and Niemelä 2008)
- Utility services for monitoring service availability and data quality management
- Related rules defined as policies

Ecosystem support services are internal services used by ecosystem members as part of the service engineering of digital services. The EODE core can be used for marking these digital services to customers and service users. However, other market places may also be exploited. In this context, the EODE core will be examined as a means to market open data services and digital services as well.

Open data services are categorized according to the purposes of usage and application domains. Generic open data services that can be used in any application domains include, e.g., open data from sensors and location[6] or information concerning culture and up-to-date activities.[7] Domain-specific data is categorized according to the application domain, e.g., traffic, transportation, and health.

3.2.2 Domain model

The domain model provides configuration rules for adopting open data services to match the quality requirements of the digital service under development. The digital service engineering context specifies how the open data service should be adapted. Domain-specific adaptation rules form a means to perform reactive adaptation according to the situation at hand. For example, the data format alignment service is used to adapt open data to the common data model of the ecosystem.

3.2.3 Service engineering model

The service engineering model provides the methodology and tools for developing open data services and digital services. It supports service innovation, business analysis, requirements identification, negotiation, and specification. The modeling of digital services exploits the SOA integration architecture described in the KMM and the related tool services used to describe the functional and non-functional capabilities of a new open data service or a new digital service. The service engineering model is described in more detailed in (Immonen et al. 2015b).

4 Open data quality certification process

This section describes how the elements of the ecosystem specified in the previous section are used to implement the quality certification of open data. The purpose of the certification process is to verify the trustworthiness and quality of open data, i.e., that the data comes from a reliable source, and its quality is good enough to be accepted for the usage of the ecosystem's services. The certification is a continuous process; the quality of data sources and the data itself is evaluated and monitored, and its exploitation in the ecosystem and its value is continuously evaluated. Sub-section 4.1 describes the certification process in more detail. Sub-section 4.2 provides an example of the usage of the quality policies in connection with the certification process in order to help understand how the quality certification is managed with the help of the policies in the EODE.

---

[6] http://www.paikkatietoikkuna.fi/web/fi
[7] http://www.hri.fi/fi/

4.1 Description of the certification process

The quality of open data is certified in five phases in the ecosystem (see Fig. 9). Some of the phases are continuous processes in the ecosystem, controlled by the ecosystem quality policies, whereas some of the phases are triggered by an event. The phases are described in the following:

P1—Acceptance: The demand for open data comes from the ecosystem. The search for new data sources is either triggered after a certain time period by the ecosystem policy or the search is triggered by an ecosystem member that demands new data. The open data source and the open data itself are evaluated to be accepted for the ecosystem with the help of the quality evaluation. The evaluation target is, first, the data source, then, the data content, and, finally, the data quality. The data extraction is a continuous process, and the quality of the (accepted) data source and the open data itself is evaluated in connection with data extraction.

P2—Adaptation: The accepted open data is modified to follow the interoperability requirements (e.g. related to format, syntax, and semantics) of the ecosystem. Thus, the open data is transformed or adapted to an open data service that can be used as a building block in digital services.

P3—Validation: The digital service provider validates the open data against its intended use, i.e. whether the data is fit for use within the certain context and situation of the digital service provider. The service provider configures the quality evaluation policy according to its own organizational data policy.

P4—Exploitation: The open data sources and the usage of the open data are monitored, and feedback from users is collected. The users of the data are the digital service providers that use the data in their services, and the service consumers that utilize the data via digital services. The ecosystem enables reactions to changes and allows decision-making, based
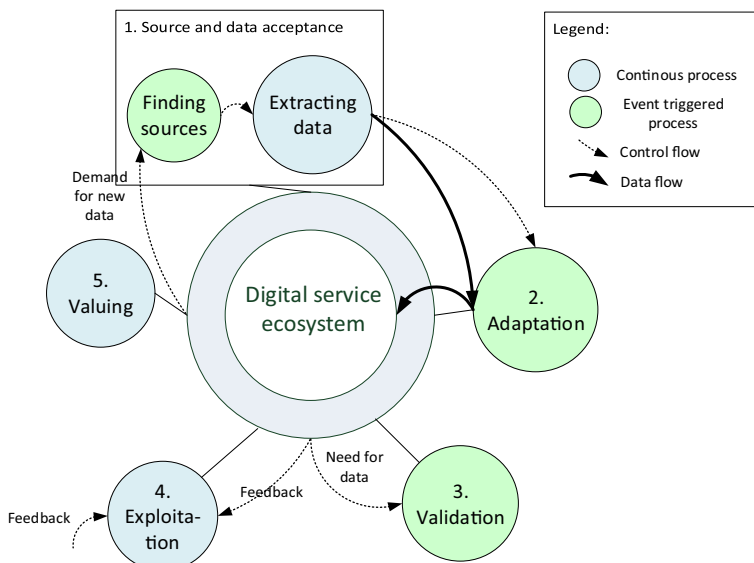

Fig. 9 The phases of the open data quality certification process

on the collected open data, by visualizing the alternatives and enabling the configuration of parameters.

P5—Valuing: The quantified value of the open data service is continuously evaluated. This includes the value comparison of open data services and the decision to keep the service or substitute it with another service.

Table 1 describes the quality attributes with the metrics and measurement approaches that were identified to be applicable in the ecosystem context. The measurement approach is defined as a sequence of operations aimed at determining the value of a measurement result, being a measurement method, a measurement function, or an analysis model (García et al. 2006). The measurement method is a logical sequence of operations that is used to quantify an attribute with respect to a specified scale (defining a base measure). The measurement function is an algorithm or calculation performed to combine two or more base or derived measures (defines a derived measure). The analysis model is an algorithm or calculation that combines one or more measures with associated decision criteria (defining an indicator). The quality attributes are defined by the knowledge management models, i.e., policies, and are evaluated with the related support services.

Table 2 maps the activities (A1–A6 described in Section 3.1.3) to the certification phases (P1–P5), and summarizes the support services and knowledge assets that are required to implement each activity. Table 2 also maps the derived quality attributes for evaluating the quality of open data/open data services to each activity.

Data filtering, data quality evaluation, and decision-making policies have different purposes in each evaluation phase in the ecosystem. Open data can originate from different kinds of source types, and each type can have different kinds of properties relating to, for example, the data content, structure, and size. Therefore, the first thing that the data filtering policy must define is acceptable data source types. Each data set is then classified into these types. The open data source types can be, for example, web pages (free-formed), Facebook, Twitter, Instagram, customer feedback, analyses and reports, or other semi-structured documents. The filtering policy and data evaluation policy must define the quality properties and rules specific for each data source type for data quality evaluation. These include the following:

- The attributes for data source/open data/open data service evaluation
- The metrics of which the attribute consists and which are used in the assessment
- The value range for each metric
- The formula for achieving the metric value from measured value
- The acceptable value for each metric
- The rules that define which attributes/metrics are taken into account and which weights are assigned to the metrics

The filtering policy uses the quality metrics and rules in evaluating whether or not to accept the data set to the ecosystem. Thus, the data filtering policy defines the quality criteria for open data sources, i.e., the strategic evaluation criteria. The data evaluation policy is used by the ecosystem in data extraction, in data monitoring, and in decision-making. Also, service providers have their own evaluation policies when searching data for a certain purpose (in phase 3). The quality evaluation policy is utilized as criteria for SLA specification and tactical quality evaluation of the data itself. The decision-making policy defines the criteria for actions based on quality evaluation, such as how to adapt to changes or what actions to take based on evaluation results.

Table 1 Metrics and measuring methods used in quality attribute evaluation

Quality attribute	Description	Metric	Measurement approach
Believability	The extent to which data is regarded as true and credible; credibility of the data source, comparison to a commonly accepted standard, and previous experience (Pipino et al. 2002)	The lowest number of three properties; the believability of the data source, believability against a common standard, and believability based on experience. E.g., the source has a verified account.	Analysis model
Completeness	The degree to which data is not missing; schema level: all of the required classes and properties are represented, data level: no missing values of properties with respect to the schema (Behkamal et al. 2014)	E.g., completeness of data within a data set (A/B); A = number of data required for the particular context in the data set, B = number of data in the specified particular context of intended use	Measuring function
Consistency	Consistency of the values of data; implies that two or more values do not conflict with each other (Mecella et al. 2002) (Wand and Wang 1996)	Data consistency checks	Measuring function
Corroboration	The same data comes from different sources (Dai et al. 2008)	The amount of data sources	Measuring function
Coverage of data	The extent to which the volume of data is appropriate for the task at hand (Pipino et al. 2002)	The minimum of two ratios: The ratio of the number of data units provided to the number of data units needed, and the ratio of the number of data units needed to the number of data units provided	Measuring function
Popularity	The data has a number of followers, and/or the data is liked and repeated by others	Number of data users; Number of re-tweets	Measuring function
Relevancy	The extent to which the data is applicable and helpful for the task at hand (Pipino et al. 2002)	The number of occurrences of relevant key words	Measuring function
Semantic accuracy	Data entities must reference a real world correspondent and must have faultless attribute values (Behkamal et al. 2014).	Detection of outliers; Detection of inaccurate values	Measuring function, Analysis model
Syntactic accuracy	Structural validity of a dataset, such as compliance with the RDF/XML standard (Behkamal et al. 2014)	E.g., A/B; A = number of records with the specified field syntactically accurate, B = number of records	Measuring function
Timeliness	The freshness of the data	A timestamp: data set creation date	Measuring function
Uniqueness	The non-redundancy characteristic of the entities, classes, properties, and values of properties in a dataset		Analysis model

Table 1 (continued)

Quality attribute	Description	Metric	Measurement approach
	(Behkamal et al. 2014)	Number of redundant data classes; ratio of similar properties; ratio of redundant instances; ratio of functional properties with different values	
Validity	The likelihood that the data in an appropriate format and the values are still valid (Ramaswamy et al. 2013)	Data syntax and semantic checks; Data set creation date	Analysis model
Verifiability	The degree and ease with which the data/information can be checked for correctness. E.g., checking the data source names from the actual source of the data, or whether the data points to a trusted third party source where the data can be checked for correctness (Naumann 2002)	Resource identifier, Relation, Cross references	Analysis model

Table 2 Mapping between the activities, support services, knowledge assets, and quality attributes

Phase	Activity	Support services	Knowledge assets	Quality attribute	Evaluation target
Acceptance (P1)	Finding out relevant open data sources (A1)	Recognition and acceptance of new data sources (S4.1) Content recognition of open data (S4.3) Quality evaluation of open data source (S3.1)	Filtering policy	Believability Popularity Verifiability Timeliness	Data source
	Extracting open data from different types of data sources (A2)	Quality monitoring of open data sources (S3.1) Content monitoring of open data(S4.3) Quality evaluation and monitoring of open data (S3.2)	Quality evaluation policy	Believability Popularity Verifiability Timeliness Uniqueness Coverage of data	Data source, open data
Adaptation (P2)	Checking syntax and semantics of open data and transforming these to a standard format (A3)	Checking syntax of open data (S2.1) Adaptation of open data to the standard format (S2.2) Checking semantics of open data (S2.2) Semantic alignment of open data (S2.3)	Service model	Syntactic accuracy Semantic accuracy	Open data
Validation (P3)	Enabling quality evaluation of open data services (A4)	Recognition of usage context of open data (S4.2) Quality evaluation of open data services (S3.3)	Quality evaluation policy	Validity Completeness Relevancy	Open data
Exploitation (P4)	Changing quality policies based on the changes of open data sources and/or (quality of) open data (A5)	Monitoring the quality of open data sources and open data (S3.1, S3.2) Monitoring the use of open data (S6.1) Analyses of open data sources and (quality of) open data (S6.3) Visualizing quality measurements (S7.3) Configuring quality policy (S1.1–S1.4) Visualizing changes in quality policy (S5.2, S7.2)	Quality evaluation policy Decision-making policy	Popularity Corroboration Coverage of data Relevancy Validity	Open data service
Valuing (P5)	Certifying open data services (A6)	Checking syntax of open data services (S2.1) Checking semantics of open data services (S2.2) Quality evaluation of open data services (S3.3)	Service model Quality evaluation policy	Syntactic accuracy Semantic accuracy Uniqueness Completeness Consistency	Open data service

The policies are expressed using event-condition-action (ECA) rules. ECA rules take the form "when Event occurs and Condition holds, then execute Action," in other words, the ECA rules are composed of event definitions, triggering conditions, and the actions to be taken.

4.2 An example of the usage of policies in data quality certification

Table 3 provides an example of a detailed description of how policies and support services are related to the two activities of phase 1 of the certification process. Data filtering policy enables the selection of reliable data and data sources for the ecosystem. The description of the policies of the example concentrates on the data source type "Twitter." Table 3 describes two activities. Activity 1 is described in more detail below.

4.2.1 Introduction of policies

According to Table 3, the data filtering policy is used in phase 1, activity 1 ('Finding out relevant open data sources'). The content of the policy is described in more detail below:

A) *Acceptable data source types and the acceptable content for each data source type*: The filtering policy defines the list of acceptable data source types and the content for each data source type, e.g.,

$$Data\ source\ type = Twitter, content\ type = tweet$$

$$Data\ source\ type = Youtube, content\ type = videos$$

$$Data\ source\ type = Facebook, content\ type = text, pictures, videos$$

Table 3 The policies and support services that implement the activities

Policy	Policy content	Support services
Phase 1 activity 1: finding out relevant open data sources		
Data filtering policy	A) Acceptable data source types and the acceptable content for each data source type	Content recognition (S4.3)
	B) Quality attributes and evaluation metrics for each data source type	Quality evaluation of data source (S3.1)
	C) Value range and rules for acceptance for each data source	Data source recognition service (S4.1)
Phase 1 activity 2: extracting data from different types of data sources		
Data filtering policy	Description of acceptable data sources	Data source recognition service (S4.1)
Quality evaluation policy	Definition of quality attributes and metrics for each data source type	Quality evaluation of open data (S3.2)
	Definition of rules for data set acceptance	

B) *Quality attributes and evaluation metrics for each data source type*: Table 4 describes an example of the attributes, metrics, and their value range included in data filtering policy for the data source type "Twitter".

C) *Value range and rules for acceptance for each data source*: The following describes the filtering policy rules for the data source type "Twitter":

Event: Finding out relevant open data sources
Condition:

IF the data source type = Twitter AND the content type = tweet AND
IF (Data publication date > 1.1.2015) AND (Popularity > 0.6 OR Believability > 0.9) AND Verifiability = 0.7

Action: The data source is assigned as an acceptable data source for the ecosystem.

4.2.2 Usage of the policies

Figure 10 illustrates how support services exploit policies in phase 1, activity 1 and in phase 1, activity 2. The usage of the policies is described in more detail below:

Table 4 An example of the elements of the data filtering policy

Attribute	Metric	Formula	Value range (% of the estimated maximum value of the topic; normalized between 0...1)	Acceptable value
Timeliness	Data publication date	Retrieving the date	1.1.2015–today	Newer than 1.1.2015
Popularity	Number of followers in a day	Followers count	<0.2 = "not relevant"; <0.4 = "not acceptable"; ≥0.6 = "acceptable"; ≥0.8 = "relevant"; ≥0.9 = "highly relevant"	≥0.6
Believability (of source)	Registration age of the author in days	Retrieving registration age	<0.3 = "not relevant"; ≥0.5 = "acceptable; ≥0.8 = "relevant"; ≥0.9 = "highly relevant"	≥0.6
	Author frequency: the number of tweets/ comments/questions at posting time	Counting all tweets/ comments/questions	<0.4 = "not relevant"; ≥0.5 = "acceptable; ≥0.8 = "relevant"; ≥0.9 = "not relevant"	≥0.4 and <0.9
	The source's account is verified	Is verified	1 = yes; 0 = no	1
Verifiability (of content)	Resource identifier	Resource identifier search	1 = yes; 0 = no	1
	Number of cross references	Cross references search	<0.4 = "not acceptable"; 0.4 = "1"; 1.0 = "estimated max"	≥0.4 ≤ 1.0

Phase 1, activity 1: finding out relevant open data sources The ecosystem evaluates whether or not the new data source can be accepted to the ecosystem. The trigger for this activity can come from an ecosystem member, or the activity can be triggered after a certain time period according to ecosystem policy.

1.1. The user (or the search monitor) that follows the ecosystem capability model/ecosystem policy wants to add a new data source to the ecosystem. The user defines the data source to the EODE system through a user interface.
1.2. The recognition service checks the content of the data.
1.3. After the content has been verified, the recognition service notifies the monitoring and evaluation service to evaluate the quality of the data source.
1.4. The monitoring and evaluation service identifies the data source type, and it evaluates the quality of the data source utilizing the data filtering policy. The data filtering policy defines the quality attributes and metrics for the data source type at hand.
1.5. The monitoring and evaluation service returns the evaluation results to the recognition service.
1.6. The recognition service checks the value range for the quality attributes and compares them with the evaluation results. The value ranges are defined in the data filtering policy.
1.7. The recognition service returns the decision whether to accept or reject the data source. In this case, the data source is accepted to the ecosystem.

Phase 1, activity 2: extracting data from different types of data sources The data is brought to the ecosystem from an acceptable data source. The data is evaluated at the time of extraction.
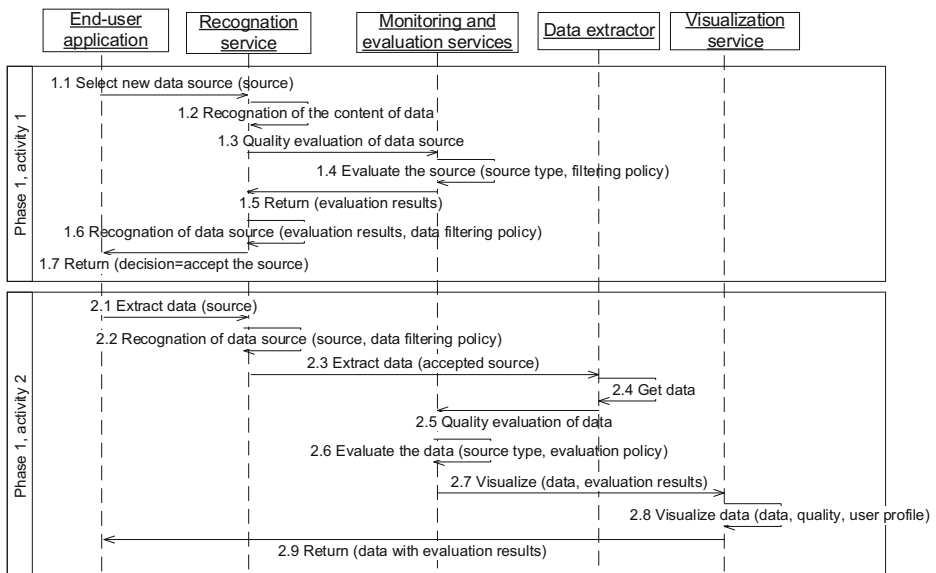

Fig. 10 The usage of quality policies at run-time

2.1. The user/monitor wants to extract data from a data source and to see the quality of the data.

2.2. The recognition service checks with the data filtering policy whether the source is an acceptable source for the ecosystem.

2.3. If the data source is accepted by the ecosystem, the recognition and adaptation service permits data extraction service to extract the data.

2.4. The data is extracted to the ecosystem.

2.5. After the data extraction, metadata is created (beyond the scope of this paper, see Immonen et al. 2015a) and the evaluation and monitoring service is requested to evaluate the quality of the data set.

2.6. The monitoring and evaluation service evaluates the quality of the data set according to the quality attributes and metrics defined in the quality evaluation policy for this kind of data source type. (Some of the quality attributes that have already been evaluated in the case of activity 1 are now reevaluated at the time of extraction.)

2.7. The visualization service is requested to illustrate the data with its quality for the user.

2.8. The visualization service illustrates the extracted data with its quality attribute values for the user according to the user profile.

5 Analyses and discussion

This section describes the maturity analyses of the main elements of the EODE concept and discussion of ongoing and future work.

5.1 Concept development and validation

In this paper, all our earlier works related to open data-based ecosystems, service ecosystems, and data quality evaluation have been combined, adapted, and extended; the concept of an evolvable open data-based digital service ecosystem is introduced. The concept specifies the capability model with the required support services and knowledge models to implement the actions related to the quality certification of open data. The development and validation of the EODE concept has been carried out incrementally in several international and national research projects. The research described in this paper was conducted in co-development in the ODEP, N4S, and DHR projects in 2015–2016. The validation of the work remains in progress in the N4S and DHR projects until 2017. The development and validation of the parts of the EODE concept are described in the following sub-sections, including the status and the maturity of the evaluation.

5.1.1 Main elements of the service ecosystem and ecosystem-based service engineering model

The digital service ecosystems were researched in Innovative Cloud Architecture for the Real Entertainment (ITEA2-ICARE)[8] project during in 2011–2014. The term digital service

---

[8] https://itea3.org/project/icare.html

ecosystem was relatively new at that time and not properly defined, and, therefore, a comparative definition of the properties of the business ecosystem, digital service ecosystem, and software ecosystem were first presented. Based on this state-of-the-art review, it was detected that methods for how to take the digital service ecosystem elements into account in service engineering were missing. The main requirements for ecosystem-based digital service engineering were identified, and the main elements of a digital service ecosystem and an ecosystem-based service engineering model were specified (Immonen et al. 2015b). The service engineering model included a requirements engineering (RE) method for digital service ecosystems, and it included two document templates for requirements elicitation and identification and for communication, knowledge sharing, negotiation, and decision-making. The RE method was validated in use with the ecosystem concept in two different ecosystems. The first case took place in the ITEA2-ICARE project, when the ecosystem concept and the RE method were applied to specifying the digital services and related support services for an interactive multi-screen TV services ecosystem. The goal for applying the RE method was to collect and analyze requirements from the ecosystem members for a shared service-oriented platform, which would enable the provisioning, integration, and use of services among the members of the ecosystem. The second case took place in the Connecting Digital Cities (EU-EIT-CDC)[9] project, in connection with the open service platform offering open real-time data from several data providers. The goal of the RE method application was to extract high-level user and business requirements for the open real-time data platform. Altogether, the method was used by 32 European partners that collected 298 requirements, including functional, non-functional and business requirements, and constraints (Immonen et al. 2015b). A feedback collection among the partners that were involved in the requirement engineering in the ICARE and CDC projects was performed to obtain user experiences and opinions about the ecosystem concept and the method and to find out any advantages, shortcomings, and development targets (Immonen et al. 2015b). The RE method was seen as valuable and useful in the beginning of the service engineering process when the long-term development of new service architecture was started for digital ecosystem-based services. The service RE method was especially useful for describing, documenting, and communicating the capabilities of the digital services and the service architecture they required. The method was also seen as useful in the analysis phase, where the different stakeholders work together. However, the definition of quality requirements was identified as a development target; special skills and knowledge on quality attributes are required and should be present in the innovation and requirements analysis and in the negotiation and specification phases.

5.1.2 Concept of open data-based business ecosystem

In our initial research on open data (Immonen et al. 2014; Immonen et al. 2013), the first draft of an open data ecosystem was defined from the business viewpoint. The work was performed in 2012–2014 on the national strategic research project, ODEP (Open Data End-user Programming), funded by the Finnish Funding Agency for Technology and Innovation (TEKES) and VTT Technical Research Centre of Finland. The purpose of the ODEP project with the research theme "Open data and analytics" was to create new technology and business potential utilizing open data. The subject was, at the time, relatively new and the utilization of open data in business by companies was at the outset. The requirements of such an ecosystem were

---

[9] EIT ICT Labs project No. 14465

collected with the help of interviews of industry representatives and the motives and challenges of acting in the open data ecosystem were identified. Altogether, 11 industry representatives participated in the interviews, including ecosystem actors such as data providers, application developers, infrastructure providers, and application users. Companies were selected from different application domains to be interviewed, and they differed in company size and service types. The interviewees, for example, product developer managers, customer and development managers, and finance and administration managers, were selected based on their knowledge of the business viewpoint of their company. The interviews provided valuable insight and requirements for the concept of an open data-based ecosystem and enabled a response to the actual needs of the data-based industry. Furthermore, the interviews enabled identifying the challenges and opportunities of open data, and applications and services of open data, and enabled evaluating the feasibility of the open data ecosystem (Immonen et al. 2014).

5.1.3 Solution for quality evaluation of open data

The evaluation of the open data quality was the main concern in the work in (Immonen et al. 2015a), in which the elements and phases of quality evaluation of open data in big data architecture were defined. The research was conducted under DIMECC's Need for Speed (N4S) program[10] funded by TEKES, in 2013–2016 and will continue until 2017. A solution for quality evaluation of open data was developed, which based on data quality policies, defines the evaluation metrics, extracts and evaluates data from Twitter, and visualizes the data for users weighting the relevant quality attributes of the user. The solution was validated with the help of an industrial case example; the solution provided a major data consulting company insight into customer needs, facilitating the R&D of the company. The solution evaluated the quality and trustworthiness of data and, thus, provided verified data for the company's business decision-making. The data in the case study was social media data (mainly from Twitter), but the approach can be extended to other data source types as well. The validation showed that although the evaluation succeeded within the case example, much more work remained to be done to extend the solution to be applicable to other data source types as well.

The current implementation of the quality evaluation of open data supports quality evaluation inside a single company; the quality policies must be implemented to be applicable to the ecosystem context. In EODE, the evaluation is done with the help of quality policies on two levels; ecosystem and service providers. The purpose of the quality evaluation performed by the ecosystem is to ensure that the quality of data is good enough to be accepted to the ecosystem. The purpose of the quality evaluation performed by the service provider is to ensure that the data fits the intended use of the provider. The service provider must first identify the intended use of data. This should be defined in the company's strategy. The ecosystem assists the service provider in specifying their own quality policies. After that, the service provider evaluates the data quality with the help of quality attributes and metrics applicable to their own context. These should be defined in the company's own quality policies. The service provider does not necessarily have to know anything about data quality evaluation methods or techniques. The provider can specify their data requirements with the help of the evaluation policy "template" provided by the ecosystem. The "template" of policy assists in selecting the evaluation attributes applicable, and, depending on the type of data source, the applicable metrics and techniques can then be selected automatically. The

---

[10] http://www.n4s.fi/en/

evaluation can also be automated, i.e. the service matchmaking algorithm can perform the quality evaluation with the help of the quality policy defined by the service provider.

5.1.4 Semantic data model

Previous attempts at the application of semantic data structures to the presentation of data, and thus bringing necessary understanding to the data, have not become widely adopted because of the additional work required. The presentation of data semantics would require modification not only to the data structure itself but also to all the applications that produce and use the particular data. To tackle this problem, the authors have made several contributions to data semantics. In (Pantsar-Syväniemi et al. 2012), a generic adaptation framework for developing situation-based applications for smart environments is described that embodies a novel architecture and general ontologies that solve the semantic, dynamic, behavioral, and conceptual interoperability problems of most physical environments. The semantic models (in the form of ontologies) ensure interoperability beyond communication and the interoperability of the information exchanged, the interoperability of context and its changes and the interoperability of application behavior. The applications developed based on the framework can use and apply semantic information in different kinds of smart places (for example, in homes, offices, and cities). A presentation of how the applications are developed based on the ontologies is provided in (Ovaska and Kuusijärvi 2014), in which the run-time quality adaptation is also described. The developed approach was applied to the development of a semantic facility data management system (Niskanen et al. 2014) that was incrementally developed and validated with four industrial pilots that were carried out in 2011–2014. The semantic models included domain-specific parts, which complicated their usage in different domains, but the security and context ontologies were generic and applicable to any domain.

Commonly, open data is published as a REST (REpresentational State Transfer)[11] resource without any description of the data structure. Thus, the content and structure of the data remain unknown, complicating the utilization of the data. Even in the cases when a data structure is presented in a commonly accepted way, e.g., XSD,[12] the XML Schema definition, the nature of the data and the purpose of the attributes cannot be detected. As a response to this problem, the service data description with the semantic data model was developed in the Digital Health Revolution (DHR)[13] project funded by TEKES in 2015. The service data description of the model enables linking the service data description to commonly available schematics, e.g., schema.org or domain-specific ones, or optionally to a service-specific dictionary. The solution allows the presentation of heterogeneous data from different domains using a common description format that still allows the presentation of the domain-specific semantic information. Linkage to commonly available schemas minimizes the need for service-specific ontology development. The structure of the data remains unaffected, which allows the data to be presented using multiple schemas, e.g., linking to different vocabularies depending on the attribute. This kind of data description model is an efficient tool for relevant data discovery; the data search can be targeted directly to the service data descriptions using the terminology of the schema. For example, the search can be targeted to find all health data that provides continuous heart rate information. Different data structures from different data services

---

[11] http://www.service-architecture.com/articles/web-services/representational_state_transfer_rest.html
[12] http://searchsoa.techtarget.com/definition/XSD
[13] http://www.digitalhealthrevolution.fi/

that all provide the same information, e.g., the heart rate, are linked to the same concept regardless of the parameter names in the original data structure and the presentation method of the data. To aid the creation of the open data service descriptions, a service framework implementation[14] provides graphical tools for service registration. The digital service registry implementation[15] provides the service description database, related REST-interfaces, and data models as open source implementation.

In summary, interoperability on the data level is achieved by the service data description, dataset descriptions, and their relations to the Concept Schema and the concepts defined in it. Service level interoperability is realized by using common service interface descriptions based on the REST architecture style. The further development of the semantic data model is still ongoing in the DHR project, and will continue until 7/2017. At this moment, tools for creating the data descriptions and linking to applicable dictionaries, and service matchmaking, based on the data descriptions, are in progress. Online testing environment deployment is also ongoing with project partners, where data and service providers can publish data sources and services that produce additional value from data to service consumers. The tool for creating a semantic data description for a service uses the normal data model as input, asks the service developer for the parameters, e.g., heart rate or pulse, and possible links to domain-specific schemas and, thus, transforms the non-semantic data model to the semantic data model.

5.1.5 EODE core

The core is based on the authors' work on a Digital Service Registry. The Service Registry is a centralized system for maintaining a list or catalogue of any digital services reachable through a URL and additional description information, such as technical and human readable descriptions, location (geographic and endpoint URL), service user feedback and rating, access management, availability information, and service logging.

The EODE core, and the Digital Services Hub as its initial implementation, was developed within the scope of the ITEA2-ICARE project. The implementation was done completely using a model-based software development method. An instance of the core was published as an open service platform[16]—Digital Services Hub—that is free to use for research and innovation purposes. The Digital Services Hub was used and demonstrated in the ITEA2-ICARE project, where project partners registered their services and used the Digital Services Hub for authorizing and visualizing service connections. The Digital Services Hub provides a user interface with which the service providers can register their services. The Digital Services Hub fulfilled its purposes well in the multimedia domain of eight international service providers. However, the context was closed, and, therefore, more validation is required. Further development of the Digital Services Hub is in progress in the N4S and DHR projects. Practical experiences of the applicability of the knowledge management and capability models in digital service ecosystems are especially required.

The Digital Services Hub can contain any kind of digital entities that have a digital API, e.g., open data, support services, and digital services. The service description can include information about the service characteristics, such as the throughput and latency, with which the service consumer can evaluate whether the service is working correctly and is good enough

---

[14] www.digitalserviceshub.com
[15] https://github.com/digitalhealthrevolution/serviceregistry
[16] https://www.digitalserviceshub.com/registry/

for the consumer's purposes. The Digital Services Hub also handles service availability in the following ways: (1) the reported availability; the registered service notifies the services Digital Services Hub regularly when they are active. (2) Requested availability; the Digital Services Hub continuously queries its services for availability and represents the status to the service users. The service provider must implement the management API that responds to the queries.

The trust between services and data privacy is implemented in the Digital Service Hub in the context of personal data; MyData. MyData is a service in which own personal data is provided in an exploitable format. Trust and privacy management is implemented on two levels (see Fig. 11):

> *Level 1*: *The trust making*. The Digital Services Hub manages the trust between two digital services by granting a binding key between two services, e.g. service A can use service B (Fig. 11). The key ensures that the service is a recognized and trusted service, and both services are controlled though the service registry.
>
> *Level 2*: *Consent management*. The data owner consents to whether or not service B can use their personal data, and also consents to whether or not their data can be transformed over the binding between services A and B. By giving consent to the binding, the data owner also must accept the terms of data usage of service A.

The implementation of consent management is a part of the ongoing work with the test environment deployment of the DHR project. The data owners can give consent about the usage of their personal data using the MyData dashboard UI (see Fig. 11).

5.2 Limitations, open issues, future research items

The validation of the last two phases of the data certification process; "Exploitation" and "Valuing," requires a real environment. These two phases are the focus of future work: an ecosystem of trusted, interoperable, and legally compliant cloud services will be set up, and the focus will be on developing the required mechanisms to register, discover, compose, use, and assess the services. The Digital Services Hub framework may be used as the core of the ecosystem to register and monitor services, but it must be extended to monitor more quality attributes. The quality evaluation approach (Phase 1) must be adapted to work in the connection with the Digital Service Hub in the way that they can utilize each other.
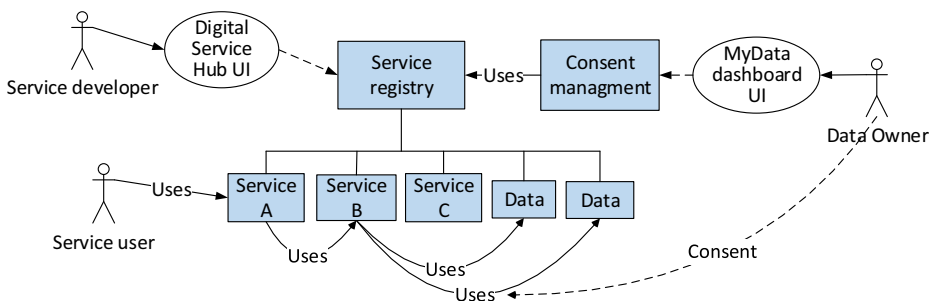

Fig. 11 The trust making and privacy management in the EODE core

Data quality was detected as the most important issue for data utilization (Immonen et al. 2014), and, therefore, in this work, data certification focused on data quality; the evaluation targets were, first, the data source, then, the data content, and, finally, the data quality itself. However, when the data quality has been ensured to be good enough, there are also many other issues that affect data selection and utilization, such as the data licenses that must be evaluated and selected. Originally, the licenses grant the "baseline rights" to distribute copyrighted work, and most licenses still contain some elements that restrict the utilization of data, such as Attribution, Non-Commercial, No-Derivatives, and Share Alike.[17] The different restricting elements can be mixed and matched, and, therefore, a huge amount of customized licenses exist for data. The selection of a license must, therefore, be done carefully and must be applicable to the situation at hand.

On the ecosystem level, there are two ways to deal with the data licenses. When the data is verified for the ecosystem based on data quality, the ecosystem can

1. Accept all data without considering the data licenses
2. Restrict data acceptance for the ecosystem based on the licenses of the data, e.g., data with certain licenses are accepted w.hile others are rejected.

In both cases, the final license selection is the responsibility of the data user, i.e., the service provider. The data provider may provide several licenses for the data. Furthermore, the service provider may utilize data from several sources in the same service. The ecosystem must bring all the license options available to service providers and assist in license selection. The service provider must be able to compare the licenses and to select the most applicable one. In the event that data from several data sources is combined to the same service, the service provider should be explicitly informed about the licenses attached to each data and the cumulative effect of the all the licenses merged together. Some assistance for the license selection has been provided for data providers. For example, Creative Commons currently provides two methods for integrating license selection into applications: the Partner Interface and the web service API. In (Daga et al. 2015), an ontology-based tool is presented for a data provider to select the license for their data. However, the selected licenses must obey the wish and intention of the data owner who, in the end, owns the rights to the data (Boris et al. 2016). In the same way, a "tool" is required for data users to compare and select licenses and to integrate several licenses of different data. This tool, "a license selection tool," could be provided as a support service of the ecosystem for service providers.

The aim is to extend the EODE with the other aspects of data certification; legal aspects, i.e., license selection and data privacy, and practical aspects (the availability of open data, open data services, and support services). These are our upcoming research topics to be examined in the upcoming international project scheduled to begin in the autumn, 2017. The data privacy aspects are domain-specific and often regulated differently in each country. Final license selection is the responsibility of the service provider; the ecosystem only assists in license comparison, selection, and integration. The concept of the EODE will be refined to include services and knowledge management models to assist the service provider in data certification based on legal aspects. Furthermore, the ecosystem's filtering policy will be used to ensure that the licenses applicable for the data are applicable also for the ecosystem. The data filtering policy may contain restrictions on license conditions that may prevent data selection for an

---

[17] https://creativecommons.org/licenses/

ecosystem even if the quality of the data has been ensured to be good enough. The domain knowledge model must include knowledge of how to take data privacy issues into account separately in each domain, e.g., healthcare, traffic, or financial. Thus, when applying the EODE in different domains, privacy issues must be solved case by case. The work on transforming the non-semantic data model to semantic data model is still ongoing in the DHR project, and the aim is to continue with the research topic in other research projects (not yet defined). In the next step, the concept schema will be extended and linked with the license ontologies in the same way that it is already linked with the data ontologies.

Furthermore, the transformation of the solution of the data quality evaluation to the ecosystem context is work the authors plan for the future, and it will be the focus of the upcoming international project planned to start in 2017. The quality policies will be refined to be applicable in the ecosystem context, both for evaluating and accepting data for an ecosystem, and for enabling users to configure policies specific to their own purposes.

The generic EODE concept described in this paper can be applied to certain domains when content of the ecosystem elements is to be specified on a more detailed level as the domain of the ecosystem becomes known. The domain model specifies the domain/application-specific knowledge used together with the generic knowledge management models to adapt the service engineering and digital services to the case at hand, e.g. to the healthcare, energy, and traffic domains. The generic EODE concept with the KMM and service engineering models, however, assists in defining these domain-specific models. The application of the EODE to different application domains and business fields is naturally the next step in the validation of the concept.

6 Conclusions

Poor and unknown quality has widely been recognized as one of the major obstacles for open data utilization. The main motivation of this paper is how to guarantee the quality of open data in the service ecosystem context. This paper combined the authors' earlier research on open data ecosystems, data quality evaluation, and service engineering in a digital service ecosystem, and it introduced the concept of an Evolvable Open Data-based digital service Ecosystem (EODE), which defines the kind of knowledge and services that are required for validating open data in digital service ecosystems. Open data is brought into EODE as an open data service that encapsulates the open data for the usage of the digital service developers. This paper introduced the main concepts of the EODE; the capability model that provides activities for quality evaluation of the open data, and the knowledge management models and the ecosystem support services that support these activities, thus enabling the quality evaluation of the data source, the open data itself, and the open data service. The EODE concept is general and applicable to any domain through domain models that describe the domain-specific concepts and rules.

This paper also introduced an open data certification process, which implements data quality certification with the help of EODE's knowledge management models and support services. The open data quality certification consists of five phases, which enable bringing validated open data to the ecosystem from trustworthy sources, transforming it to the acceptable form of the ecosystem, validating it against its intended usage of each service provider, monitoring the data sources and the usage of the data, and continuously evaluating the quantified value of the open data service. Some of the phases are continuous processes controlled by ecosystem quality policies, whereas some of the phases are triggered by an event in the ecosystem. The certification process is generally described, and must be adapted

and specialized using the domain model that defines domain- and application-specific extensions, replacements and adaptation rules, and regulations. Although several validation experiments of the EODE elements have been carried out during the last three years, the whole EODE concept still requires more experimental tests in different application and business fields in order to guarantee that generic and domain-specific knowledge can be maintained separate but smoothly exploited together using the generic capability and knowledge management models during the run of digital services.

References

Agichtein, E., Castillo, C., Donato, D., Gionis, A. & Mishne, G. 2008. Finding high-quality content in social media, in: *International Conference on Web Search and Data Mining WSDM '08*, Palo Alto, USA.

Antunes, F. & Costa, J.P. 2012. Integrating decision support and social networks. *Advances in Human-Computer Interaction 2012*(Article 9).

Aubonnet, T., Henrio, L., Kessal, S., Kulankhina, O., Lemoine, F., Madelaine, E., et al. 2015. Management of service composition based on self-controlled components. *Journal of Internet Services and Applications* 6(15).

Auer, S. R., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). DBpedia: A nucleus for a web of open data. Semantic web. *Lecture Notes in Computer Science, 4825*, 722–735.

Baroni, A., Muccini, H., Malavolta, I. & Woods, E. 2014. architecture description leveraging model driven engineering and semantic Wikis, in*: IEEE/IFIP Conference on Software Architecture (WICSA)*, Sydney, NSW.

Behkamal, B., Kahani, M., Bagheri, E., & Jeremic, Z. (2014). A metrics-driven approach for quality assessment of linked open data. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(2), 64–79.

Bertino, E. & Lim, H.-S. 2010. Assuring data trustworthiness—concepts and research challenges, in: W. Jonker, M. Petković (Eds.), *Secure Data Management*. SDM 2010. Lecture Notes in Computer Science 6358, Berlin, Heidelberg.

Bhatia, S., Li, J., Peng, W. & Sun, T. 2013. Monitoring and analyzing customer feedback through social media platforms for identifying and remedying customer problems, in: *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Niagara, Canada.

Bizer, C. 2007. Quality-driven information filtering in the context of web-based information systems. Ph.D. Thesis, Berlin.

Bizer, C., & Cyganiak, R. (2009). Quality-driven information filtering using the WIQA policy framework. *Web Semantics: Science, Services and Agents on the World Wide web Archive, 7*(1), 1–10.

Boley, H. & Chang, E. 2007. Digital ecosystems: principles and semantics, in: *Inaugural IEEE International Conference on Digital Ecosystems and Technologies* (DEST 2007), Cairns, Australia.

Boris, O., Auer, S., Cirullies, J., Jürjens, J., Menz, N., Schon, J., et al. 2016. Industrial data space: digital souvereignity over data, technical report Fraunhofer-Gesellschaft, doi:10.13140/RG.2.1.2673.0649.

Bosch, J. 2009. From software product lines to software ecosystems, in: *The 13th International Software Product Line Conference* (SPLC'09), San Francisco, USA.

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal, 14*(2), 1–10.

Castillo, C., Mendoza, M. & Poblete, B. 2011. Information credibility on twitter, in: *The 20th International Conference on World Wide Web*, Hyderabad, India.

Chan, C.M.L. 2013. From open data to open data innovation strategies: Creating E-Services Using Open Government Data, in: *The 46th Hawaii International Conference on System Sciences (HICSS)*, Wailea, USA.

Chen, M., Ebert, D., Hagen, H., Laramee, R. S., Van Liere, R., Ma, K. L., et al. (2009). Data, information, and knowledge in visualization. *Computer Graphics and Applications, 29*, 12–19.

Chesbrough, H. W., & Appleyard, M. M. (2007). Open innovation and strategy. *California Management Review, 50*, 57–76.

Daga, E.; d'Aquin, M.; Motta, E.; Gangemi, A. 2015. A bottom-up approach for licences classification and selection, in: *Proceedings of the International Workshop on Legal Domain and Semantic Web Applications (LeDA-SWAn) Held during the 12th Extended Semantic Web Conference* (ESWC 2015).

Dai, C., Lin, D., Bertino, E. & Kantarcioglu, M. 2008. An approach to evaluate data trustworthiness based on data provenance, in: W. Jonke, M. Petkovic (Eds.), SDM 2008. Lecture Notes on Computer Science 5159.

Dobrica, L., & Niemelä, E. (2002). A survey on software architecture analysis methods. *IEEE Transactions on Software Engineering, 28*(7), 638–653.

European Commission. (2011). *Open data an engine for innovation, growth and transparent governance, COM/2011/0*. Brussels: European Commission.

Fabijan, A., Holmström Olsson, H., & Bosch, J. (2015). Customer feedback and data collection techniques in Software R&D: A literature review. *Lecture Notes in Business Information Processing, 210*, 139–153.

Ferrando-LIopis, R., Lopez-Berzosa, D. & Mulligan, C. 2013. Advancing value creation and value capture in data-intensive contexts., in: *IEEE International Conference on Big Data*, Silicon Valley, USA.

García, F., Bertoa, M. F., Calero, C., Vallecillo, A., Ruíz, F., Piattini, M., et al. (2006). Towards a consistent terminology for software measurement. *Information and Software Technology, 48*, 631–644.

Gil, Y., & Artz, D. (2007). Towards content trust of web resources. *Web Semantics: Science, Services and Agents on the World Wide Web, 5*(4), 227–239.

Gorton, I., & Klein, J. (2015). Distribution, data, deployment: software architecture convergence in big data systems. *IEEE Software, 32*(3), 78–85.

Guessi, M., Moreira, D.A., Abdalla, G., Oquendo, F. & Nakagawa, E.Y. 2015. OntolAD: a formal ontology for architectural descriptions, in: *30th ACM Symposium on Applied Computing (ACM/SAC'2015)*, Salamanca, Spain.

Hanssen, G.K. & Dybå, T. 2012. Theoretical foundations of software ecosystems, in: Proceedings of the Forth International Workshop on Software Ecosystems (IWSECO), Cambridge, MA, USA.

Heimstädt, M., Saunderson, F., & Heath, T. (2014a). From toddler to teen: growth of an open data ecosystem. *eJournal of eDemocracy & Open Government (JeDEM), 6*(2), 123–135.

Heimstädt, M., Saunderson, F. & Heath, T. 2014b. Conceptualizing open data ecosystems: a timeline analysis of open data development in the UK, in: *Proceedings of the International Conference for E-Democracy and Open Government* (CeDEM2014), Krems, Austria.

HM Government Cabinet Office 2012. *Open data white paper: unleashing the potential*, Retrieved, London, UK.

Iansiti, M. & Levien, R. 2004. Creating value in your business ecosystem. *Harvard Business Review* 2004, 68–78.

Immonen, A., & Niemelä, E. (2008). Survey of reliability and availability prediction methods from the viewpoint of software architecture. *Software and Systems Modeling, 7*(1), 49–65.

Immonen, A., Palviainen, M., & Ovaska, E. (2013). Towards open data based business: survey on usage of open data in digital services. *International Journal of Research in Business and Technology, 4*(1), 286–295. doi:10.0001/ijrbt.v4i1.197.

Immonen, A., Palviainen, M., & Ovaska, E. (2014). Requirements of an open data based business ecosystem. *IEEE Access, 2*, 88–103. doi:10.1109/ACCESS.2014.2302872.

Immonen, A., Paakkonen, P., & Ovaska, E. (2015a). Evaluating the quality of social media data in big data architecture. *IEEE Access, 3*, 2028–2043.

Immonen, A., Ovaska, E., Kalaoja, J., & Pakkala, D. (2015b). A service requirements engineering method for a digital service ecosystem. *Service Oriented Computing and Applications, 10*(2), 151–172.

ISO. (2008). *ISO/IEC 25012—software engineering—software product quality requirements and evaluation (SQuaRE)—data quality model*. Geneva: International Organization for Standardization.

ISO/IEC. (2001). *ISO/IEC 9126–1: Software engineering—software product quality—part 1 : quality model*. Geneva: International Organization for Standardization.

ISO/IEC. (2003). *ISO/IEC TR 9126-2: software engineering—software product quality—part 2 : External metrics*. Geneva: International Organization for Standardization.

Jansen, S. & Cusumano, M. 2012. Defining software ecosystems: a survey of software platforms and business network governance, in: *The 4th International Workshop on Software Ecosystems*, Cambridge, USA.

Kantorovitch, J. & Niemelä, E. 2008. Service description ontologies, in: Mehdi Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology*.

Kazman, R., Klein, M. & Clement, P. 2000. ATAM: method for architecture evaluation, *The 4th IEEE International Conference on Engineering of Complex Computer Systems TECHNICAL*, Carnegie Mellon University, Monterey, USA.

Kett, H., Voigt, K., Scheithauer, G. & Cardoso, J. 2008. Service engineering in business ecosystems, in: *Proceedings of the XVIII International RESER Conference*, Stuttgart, Germany.

Khriyenko, O. 2012. Collaborative service ecosystem—step towards the world of ubiquitous services, in: *Proceedings of the IADIS International Conference Collaborative Technologies*, Lisbon, Portugal.

Li, S. & Fan, Y. 2011. Research on the Service-Oriented Business Ecosystem (SOBE), in: *The 3rd International Conference on Advanced Computer Control* (ICACC), Harbin, China.

Madnick, S. E., Wang, R. Y., Lee, Y. W., & Zhu, H. (2009). Overview and framework for data and information quality research. *Journal of Data and Information Quality, 1*(1), 1–22.

Mecella, M., Scannapieco, M., Virgillito, A., Baldoni, R., Catarci, T. & Batini, C. 2002. Managing data quality in cooperative information systems, in: *Proceedings of the Confederated International Conferences DOA, CoopIS and ODBASE*.

Naumann, F. 2002. Quality-driven query answering for integrated information systems, 2002, Berlin Heidelberg, New York.

Naumann, F. & Rolker, C. 2000. Assessment methods for information quality criteria, in: *The 5th International Conference on Information Quality, Boston, USA*.

Niemelä, E., & Immonen, A. (2007). Capturing quality requirements of product family architecture. *Information and Software Technology, 49*(11–12), 1107–1120.

Niemelä, E., Evesti, A. & Savolainen, P. 2008. Modeling quality attribute variability, in: *Proceedings of the 3rd International Conference on Evaluation of Novel Approaches to Software Engineering*, Funchal, Madeira, Portugal.

Niskanen, I, Purhonen, A., Kuusijärvi, J. & Halmetoja, E. 2014. Towards semantic facility data management, in: *INTELLI 2014, The Third International Conference on Intelligent Systems and Applications*. https://www.thinkmind.org/index.php?view=article&articleid=intelli_2014_4_50_70105 (accessed November 9, 2016).

Nurse, J.R.C., Rahman, S.S., Creese, S., Goldsmith, M. & Lamberts, K. 2011. Information quality and trustworthiness: a topical state-of-the-art review, in: *International Conference on Computer Applications and Network Security* (ICCANS), Male, The Maldives.

Nurse, J.R.C., Agrafiotis, I., Creese, S., Goldsmith, M. & Lamberts, K. 2013. Building confidence in information −trustworthiness metrics for decision support, in: *The 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications* (IEEE TrustCom−13)., Melbourne, Australia.

Ovaska, E., & Kuusijärvi, J. (2014). Piecemeal development of intelligent smart space applications. *IEEE Access, 2*, 199–214.

Ovaska, E., Evesti, A., Henttonen, K., Palviainen, M., & Aho, P. (2010). Knowledge based quality-driven architecture design and evaluation. *Information and Software Technologies, 52*(6), 577–601.

Ovaska, E., Salmon Cinotti, T. & Toninelli, A. 2012. The design principles and practices of interoperable smart spaces, in: L. Xiaodong, L. Yang (Eds.), *Advanced Design Approaches to Emerging Software Systems: Principles, Methodologies and Tools*.

Pantsar-Syväniemi, S., Kuusijärvi, J. & Ovaska, E. 2011. Supporting situation-awareness in smart spaces, in: *Grid and Pervasive Computing Workshops*, Volume 7096 of the Series Lecture Notes in Computer Science, Oulu, Finland.

Pantsar-Syväniemi, S., Purhonen, A., Ovaska, E., Kuusijärvi, J., & Evesti, A. (2012). Situation-based and self-adaptive applications for the smart environment. *Journal of Ambient Intelligence and Smart Environments, 4*(6), 491–516.

Pipino, L., Lee, Y., & Wang, R. (2002). Data quality assessment. *Communications of the ACM, 45*, 211–218.

Pipino, L., Wang, R., Kopcso, D. & Rybold, W. 2005. *Developing measurement scales for data-quality dimensions*, New York.

Poikola, A., Kola, P., & Hintikka, K. A. (2011). *Public data—an introduction to opening information resources, Ministry of Transport and Communications*. Helsinki: http://www.scribd.com/doc/57392397/Public-Data.

Pollock, R. 2011. Building the (open) data ecosystem. Open Knowledge Foundation Blog, Retrieved from http://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/on July 10, 2013.

Rafique, I., Lew, P., Qanber Abbasi, M., & Li, Z. (2012). Information quality evaluation framework: extending ISO 25012 data quality model. *World Academy of Science, Engineering and Technology, 65*, 523–528.

Rahman, S.S., Creese, S. & Goldsmith, M. 2011. Accepting information with a pinch of salt: handling untrusted information sources, in: *Security and Trust Management*, Lecture Notes in Computer Science Volume 7170.

Ramaswamy, L., Lawson, V. & Gogineni, S. V 2013. Towards a quality-centric big data architecture for federated sensor services, in: *IEEE International Congress on Big Data*, Santa Clara, CA.

Ruokolainen, T. 2013. A model-driven approach to service ecosystem engineering (PhD Thesis), University of Helsinki, Department of Computer Science, Helsinki, Finland.

Ruokolainen, T. & Kutvonen, L. 2009. Managing interoperability knowledge in open service ecosystems, in: *The 13th Enterprise Distributed Object Computing Conference Workshops*, Auckland, New Zealand.

Ruokolainen, T., Ruohomaa, S. & Kutvonen, L. 2011. Solving service ecosystem governance, in: *IEEE 15th International Enterprise Distributed Object Computing Conference Workshops*, Helsinki, Finland.

Sande, M. Vander, Dimou, A., Colpaert, P., Mannens, E. & Van de Walle, R. 2013. Linked data as enabler for open data ecosystems, in: *Proceedings of the W3C Workshop on Open Data on the Web*, London, UK.

Stathel, S., Finzen, J., Riedl, C. & May, N. 2008. Service innovation in business value networks, in: *The 18th International RESER Conference*, Stuttgart, Germany.

W3C 2007. Web services policy 1.5—framework (W3C Recommendation).

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM, 39*(11), 86–95.

Wang, R., & Strong, D. (1996). Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5–33.

Wiesner, S., Peruzzini, M., Doumeingts, G. & Thoben, K.D. 2012. Requirements engineering for servitization in Manufacturing Service Ecosystems (MSEE), in: *Conference on Industrial Product Service Systems* (CIRP IPS2 2012), Tokyo, Japan.

Zhang, J. & Fan, Y. 2010. Current state and research trends on business ecosystem, in: *IEEE International Conference on Service-Oriented Computing and Applications* (SOCA), Perth, Australia.

Zhou, J., Ovaska, E., Evesti, A., & Immonen, A. (2011). OntoArch Reliability-aware Software architecture design and experience. In A. Dogru & V. Bicer (Eds.), *Modern software engineering concepts and practices: advanced approaches*. New York: USA.

Anne Immonen received her MSc degree in 2002 from the University of Oulu. Since 2002, she has been working as a research scientist at VTT Technical Research Centre of Finland, and the last 2 years as a senior scientist. Her main research topics include quality-driven service engineering, digital service ecosystems, and data quality evaluation. Her current areas of interest are quality in digital service ecosystems, quality of data, and data certification for the data-based services. She is currently finalizing her doctoral dissertation in the University of Oulu.

Eila Ovaska received a PhD degree from the University of Oulu in 2000. Prior to 2000, she was a software engineer, a senior research scientist, and leader with the Software Architectures Group at the VTT Technical Research Centre of Finland. Since 2001, she has been a research professor with VTT and an adjunct professor with the University of Oulu. Her current areas of interest are self-adaptive service architectures and knowledge-oriented service engineering in digital service ecosystems. She has acted as a workshop and conference organizer and as a reviewer for scientific journals and conferences. She has co-authored over 150 scientific publications.

Tuomas Paaso (M.Sc. EE) is a research scientist at VTT Technical Research Centre of Finland. He joined VTT in 2005 and has since worked on various research topics concentrating on service architectures, metadata management, and service composition. His current research is focused on service composition and service data models on various domains.