Report on Business Models for Linked Building Data

DiCtion WP G. Platform Model Enabling Global Business

17.11.2020

Authors

VTT Technical Research Centre of Finland: Rita Lavikka & Markku Kiviniemi

Aalto University: Ergo Pikas & Olli Seppänen

VisuaLynk: Seppo Törmä

Acknowledgements

The authors want to thank Business Finland and the DiCtion consortium members:

Fira, Consolis Parma, Sweco, Bonava, Trimble, Ruukki Construction

Table of Contents

Table of Figures

1 Introduction

The construction sector is beset with many well-known, long-standing problems. Many of these problems are related to or caused by poor project management, decision-making and risk management practices. For example, competing for the lowest cost contract, companies often cut costs to the minimum and hope to make money through claims rather than a good delivery of projects (Jang et al., 2019). This practice may lead to poor performance, and might even be one of the causes for the construction sector's low performance.

Recently, it has been argued that project management, decision making, and risk management could be improved with the support of data and information-based services. Furthermore, the European Green Deal steers the sector towards increased digitalisation, energy-efficient buildings, and circular economy[1]. Many parties, such as architects, engineers, and managers, already produce and consume interdependent data during a construction project. Also, data are increasingly produced automatically by sensors and information systems. This means that data are and will be more and more crucial in the construction sector. Gartner's report reveals that, in general, organisations sharing data with their partners usually generate three times more measurable economic benefits than the organisations that do not (Logan et al., 2020). Hence, it is recommended to start sharing data between organisations.

The development of advanced and novel algorithms and data analytics, which are used to improve construction projects' delivery, necessitate structured data from the various phases of the project delivery (Sacks et al., 2020). Data acquisition is a challenge because construction project data are often in the format of files and stored in project banks from where data are not readily usable for sophisticated machine-to-machine data analysis. Poor interoperability is one of the main issues in utilising advanced and novel algorithms and data analytics (Sacks et al., 2018). In addition to interoperability problems, according to Gartner's latest report, data sharing is also hampered by stakeholder resistance, data management and governance policies, lack of proper tools and technologies, perceived regulatory restraints, and risk assessments of security vulnerabilities (Jones, 2020).

Linked data technologies can be used for publishing, sharing and linking structured data inside or between organisations. Data, structured according to linked data ontologies, can be made available through semantic queries. The linked data provides a ready-made, standardised representation of building data and tools to improve the interoperability in the fragmented construction sector. Appendix – Linked Data technologies – describes linked data technologies in more detail. The linked data could support the construction sector to reach the next level in data usage, the level 3 maturity in Figure 1.

The use of linked building data will provide several benefits. It could reduce the multiple-handling of information, which would reduce costs and errors. Linked building data would provide structured data to develop and implement advanced and novel algorithms and data analytics. Additionally, this structured data could support business data analytics to design and create new services. However, although the linked building data technology has been proven to work in practice, viable business models based on linked building data are still missing. This is the motivation of this research.

[1] European Green Deal. Available from: https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en. [accessed 25.6.2020]

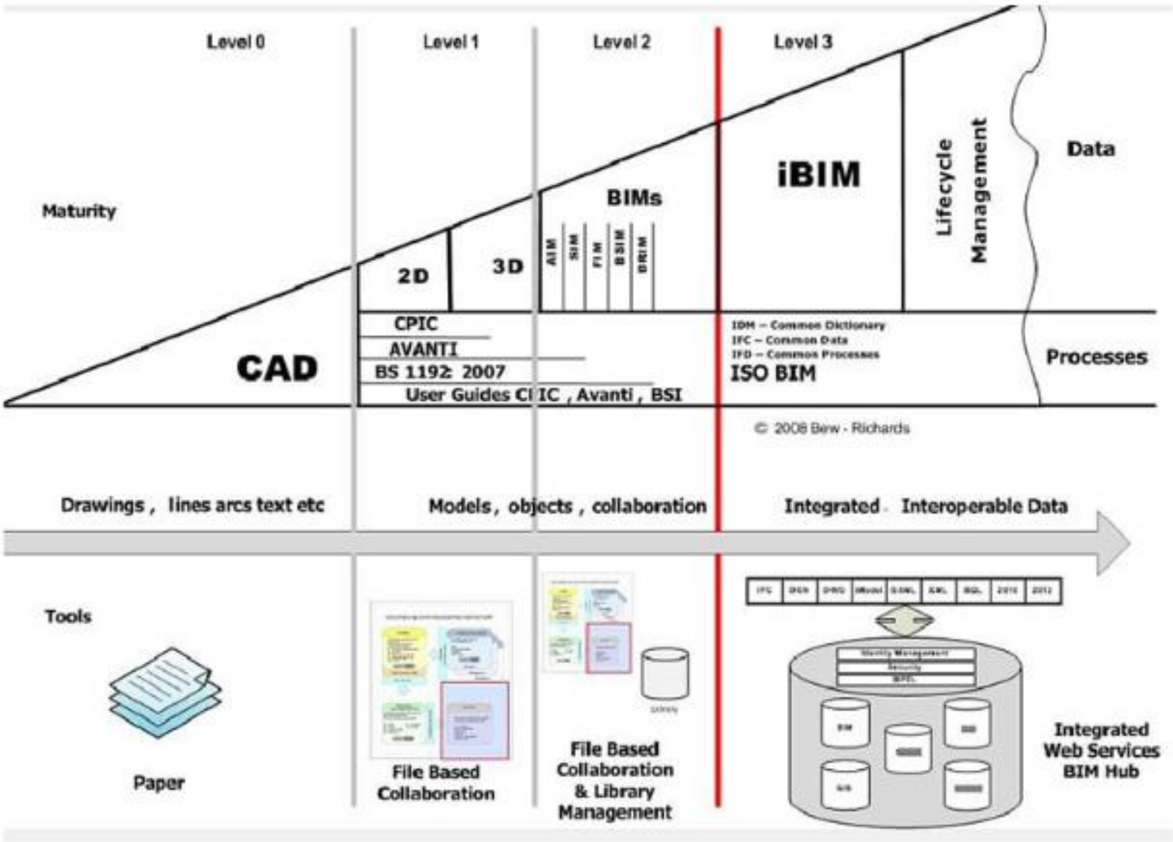

Figure 1. The BIM levels of Maturity as defined by BSI Standards Limited (PAS 1192-2:2013) Adopted from (Werbrouck et al., 2019).

This report aims to describe the findings related to linked building data, the usage of construction ontology for linked building data, and revenue models generated with the support of linked building data. The research, funded by Business Finland, is conducted as part of the work package G - Platform Model Enabling Global Business in DiCtion research project. The research organisations were Aalto University and VTT Technical Research Centre of Finland and VisuaLynk, which acted as a subcontractor to VTT. The private companies involved in the research project included Fira, Sweco, Bonava, Consolis Parma, Trimble, and Ruukki Construction.

The research work mainly consisted of analysing already published articles and reports on linked data business models. In addition, two workshops were organised to discuss the findings. The first workshop was organised on May 26th at the pre-event of WDBE[2]. In addition to two facilitators, three participants, one from the USA, one from Oman, and one from Bangladesh, took part in the workshop. The second workshop was organised on October 23rd when approximately twenty-five participants, representing the construction sector and academia, discussed the way forward with linked data.

The research questions for work package G were the following:

1. What are linked building data and ontologies?
2. How to apply a construction ontology in practical use cases?
3. What kinds of revenue models can we envision based on linked data?

[2] Digital Situational Awareness in Construction pre-event of WDBE 2020, available at https://kirahub.org/en/wdbe2020/pre-event-digital-situational-awareness-in-construction/ [Accessed on 15.6.2020]

Answer to the first question is provided in Chapter 2 *Linked building data*. Respectively, the answer to the second question is given in Chapter 3 *Ontology use cases for linked building data*. The response to the third question is given in Chapter 4 *Towards business models supporting linked data in construction operations*.

2 Linked building data

2.1 The rationale for the linked data technologies

Linked data refers to structured data interlinked with other, possibly remotely managed data and made available through semantic queries. To simplify a bit, when data is called "structural", it means that it is composed of data entities – such as objects, relations, attributes, collections and/or tables – that are machine-processable, instead of text and media that are meant for human consumption. Structural data always conforms to some kind of data scheme, and it can be used in a granular manner.

Linked Data, as a technology, builds upon standard Web technologies, such as HTTP, RDF, and URIs. While the regular web applies these technologies to help human readers to browse the web, Linked Data uses the technologies to allow computers to read data on the web automatically. In a nutshell, linked data refers to a decentralised environment, like the regular web, but it is meant for sharing machine-readable, structured data. Figure 2 illustrates the difference between the regular web and linked data. Linked data can thus be referred to as the web of data.

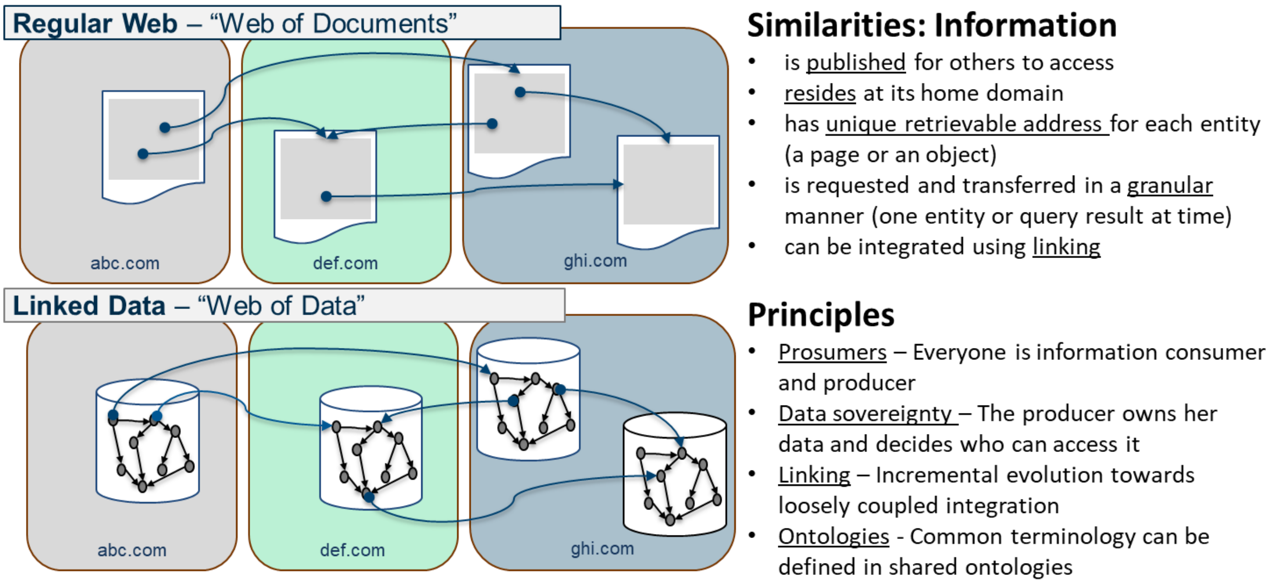

Figure 2. The difference between the regular web and linked data (©Seppo Törmä)

Linked Data technologies extend the basic technologies of the web:

- **Identifiers**: Each entity is identified with a URI (Uniform Resource Identifier), a globally unique and retrievable identifier (that is, it can be used to retrieve the description of the identified object).
- **Graph-based data**: The data is organised as a graph consisting of links between nodes. The graph format is called RDF (Resource Description Framework), in which both nodes and link types are identified with URIs.
- **Graph storage**: There are many different RDF databases available, both free (Jena, Virtuoso) and commercial (Amazon Neptune, Oracle Spatial and Graph, GraphDB, ...)

- **Serialisation**: RDF can be manually written in several different serialisation formats (rdfXML, JSON-LD, Turtle) and the contents of an RDF database can be serialised to files in those formats and exchanged with others or imported to another RDF database.
- **Query language**: RDF graphs can be queried and updated with SPARQL language, supported by all RDF databases.
- **Terminology (ontology)**: The schema for RDF graphs can be defined with Web Ontology Language (OWL).
- **Reasonin**g: Different reasoning tools are available to check the consistency of RDF graphs and to derive new data based on existing data. There are specific ontology reasoners (Hermit, Pellet, Fact++) and rule languages (SWRL, SPIN)
- **Metadata**: Linked data technologies support the representation of metadata (data about data) directly, and there are several vocabularies for that purpose.

2.2 Linked Building Data

During different periods of the building life-cycle, many different software applications by many different parties and disciplines are used to create, organise, and manage building data in the construction sector. With the advent and implementation of, for example, building information modelling and sensor technologies, the volume of data to be managed has increased remarkably. However, data creation is often dispersed between different data sources, such as discipline-specific building modelling applications. Data exchanges rely upon the exchange of standardised file formats (e.g., Industry Foundation Classes) or the development of bespoke APIs (application programming interfaces). Hence, although these technologies have alleviated interoperability issues to some extent, the creation, organisation, and management of construction project information are still fragmented.

The Linked Building Data Community Group, part of the W3C Community Group, brings together experts in building information modelling and Web of Data technologies to address the fragmented nature of managing the vast amounts of building information in the construction sector[3]. The group's scope is focused on web technologies as they may be applied to buildings (e.g., products, geometry, usage, topology) and infrastructure data (e.g., bridges, roads, railroads). The objective is to enable all stakeholders using web technologies in the building life cycle to access and query required data to support their business use cases. The group has envisioned several use cases to develop and implement linked building data, including requirements management, building permitting, code-compliance checking, different kinds of simulations (e.g., energy efficiency), facility management, indoor navigation and risk management. Linked building data is especially useful in contexts with multiple domains involved in the use case (Curry et al., 2013).

2.3 Ontologies for the construction stage

To develop a linked building data ecosystem, an ontology that covers the terminology for the construction stage is required. In the Diction project, a major effort was spent to create a suite of

- Digital Construction Ontologies (DICO): https://w3id.org/digitalconstruction/
    - Basic entities encountered in construction: Activities, Agents, BuildingObjects, MaterialBatches, Equipment, Information Content Entities (such as BIM models, drawings, notifications, issues), and their relations and attributes.
    - A general way to associate identifiers and classifications to any entities. Identifiers can be global (GUID, GTIN, ...) or local ones unique within a given scope (room number, control point

[3] Linked Building Data Community Group web page: https://w3c-lbd-cg.github.io/lbd/

number, ...). Categories can come from any available classification system (OmniClass, UniClass, CoClass, Talo2000, ...).
    - Support for multi-context data. The same object can have different values of its properties in different contexts. This can be used to model planned and actual values of time points, as-designed/as-built models, levels of detail, and so on.
    - Variables and constraints to capture the technical knowledge commonly expressed in the form of constraints, and to support the representation of partially specified plans.

DICO does not define all concepts in the construction domain by itself. Rather, it uses other existing ontologies to provide important domains of concepts that are necessary or useful in construction:

- Basic Formal Ontology (BFO): https://buffalo.app.box.com/v/bfo-iso-owl-cl
    - The fundamental categories to organise the information, such as division of objects into those occurring in time and those having a continuing existence. It is a standard top-level ontology (ISO/IEC 21838-2).
- Time (OWL-Time): https://www.w3.org/TR/owl-time/
    - The terminology for temporal entities: time intervals, time points, durations, relations between intervals, and various time value datatypes. Web Standard by W3C.
- Units of measurement (QUDT): http://www.qudt.org/release2/qudt-catalog.html
    - A comprehensive representation of quantities, units, unit dimensions and related datatypes, developed by NASA.
- Agents (FOAF): http://xmlns.com/foaf/spec/
    - A commonly used terminology for agents and organisations. From W3C.
- BIM models (ifcOWL): https://technical.buildingsmart.org/standards/ifc/
    - An IFC ontology to represent BIM models in a standard way as Linked Data. Due to the complexity of IFC, ifcOWL also has hundreds of classes and relations. Published by buildingSmart.
- Building Topology Ontology (BOT): http://www.w3id.org/bot
    - A simple model for the spatial topology of buildings, to support construction and maintenance related use cases where all design details of BIM models are not needed.
- Semantic Sensor Networks (SSN/SOSA): https://www.w3.org/TR/vocab-ssn/
    - A layered ontology for sensor observations (SOSA) and sensor devices (SSN).
- Smart Appliances (Saref): https://saref.etsi.org/
    - A shared model that facilitates the matching of existing assets in the smart applications domain.

The classes and properties in the ontology are abstract entities. The real-world entities, instances, perceivable through human experience can be represented in terms of the classes and properties. Figure 3 presents the relationship between ontology and instances. For simplicity, the ontology used in the example is the Building Topology Ontology. In DICO, it is possible to represent more flexible location structures.

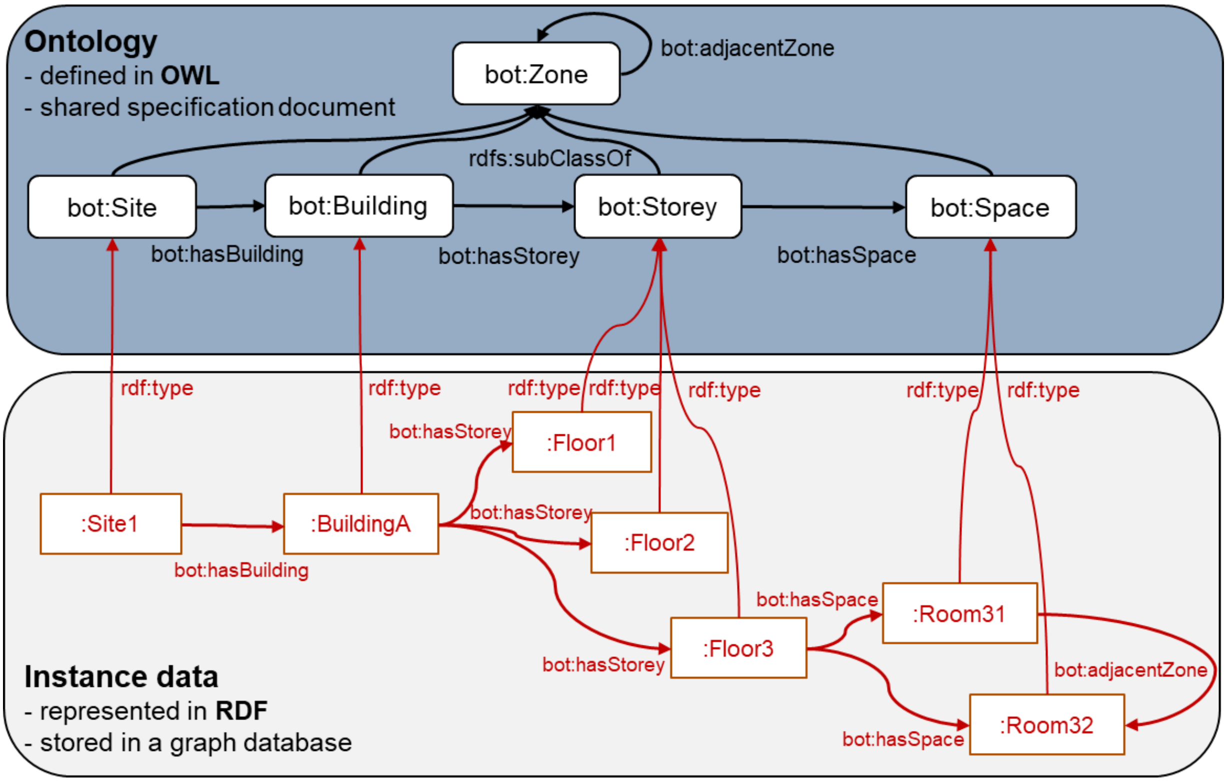

Figure 3. The difference between ontology and instance data

3 Ontology use cases for linked building data

For understanding the usefulness of the building ontology, one use case is presented in detail, and several others are mentioned generally. The following use case presents the ontology for monitoring trade contractors (Figure 4).

3.1 Ontology use case – Subcontract monitoring

Background: Recent availability of new data from construction processes – especially positioning data and inspection issues generated from mobile devices – creates opportunities to monitor subcontracted work's productivity and quality. How much time trade contractors spend in the locations of their tasks and what is the volume of quality deviations resulting from their work?

However, different pieces of data need to be organised and connected to derive meaningful information about the execution of processes, and their productivity and quality. As of now, the systems managing the data are separate: in this example, Fira's SiteDrive is used for construction scheduling, C4 systems for material logistics, Congrid for quality inspections, and AIForSite systems for positioning of people and materials within a construction site. Each system stores and organises its information in its idiosyncratic way. Initially, the systems did not directly share information—consequently, people's method to connect the diverse information to combine interrelated information from different systems manually.

Problem: Originally, there is no role or task assigned to any person in a construction project to use more than one of the systems mentioned above. The only way to combine diverse information from different systems is the discussions in project meetings. Compared to the volume of data and frequency with which it changes, the opportunities to make connections are insufficient. There is plenty of room to improve data integration's accuracy and frequency, ultimately to support real-time control.

Goal: To establish a common model of the domain that enables efficient sharing of data across the systems. Since standard solutions for technical interoperability between systems are readily available, the challenges are at the level of semantic interoperability, concerning the terms used in the data: object types, relations, and attributes. The model should be based on an ontology covering the terminology related to site activities, locations, agents, positioning data, material batches and kits, and quality issues. In particular, the ontology should represent the relations between different entities. The shared model enables the monitoring of the productivity and quality of subcontracts by gathering data about the time spent on the locations of tasks and the set of quality issues generated from the results of the tasks.

Benefits: By using a common ontology, the systems can access and interpret data from each other, without ad hoc conversions between different concepts. With a shared ontology, different parties and systems could interoperate with each other in different projects. In effect, a project could use any system for quality tracking, which supports the shared ontology, and it could work in a plug and play fashion.

The linked data approach means that companies can directly refer to each other's data objects using the URIs as addresses for those objects. There is no need to copy, transform and synchronise data between systems explicitly. The links implement the connections between data of different parties.

Ontology description: The figure below shows the entities and their relations between the entities created by the different systems involved in the scenario. Fira SiteDrive maintains information about location breakdown structure, schedules, task plans, and actual execution data. Alforsite generates information about the real-time position of assets (labour, equipment and materials) which can be mapped to shared locations. Carina4 handles material logistics and material orders, and the same material batch information is used in Aiforsite system, which is linked to activities in SiteDrive. Similarly, Congrid system manages issues which can be connected to locations or material batches.

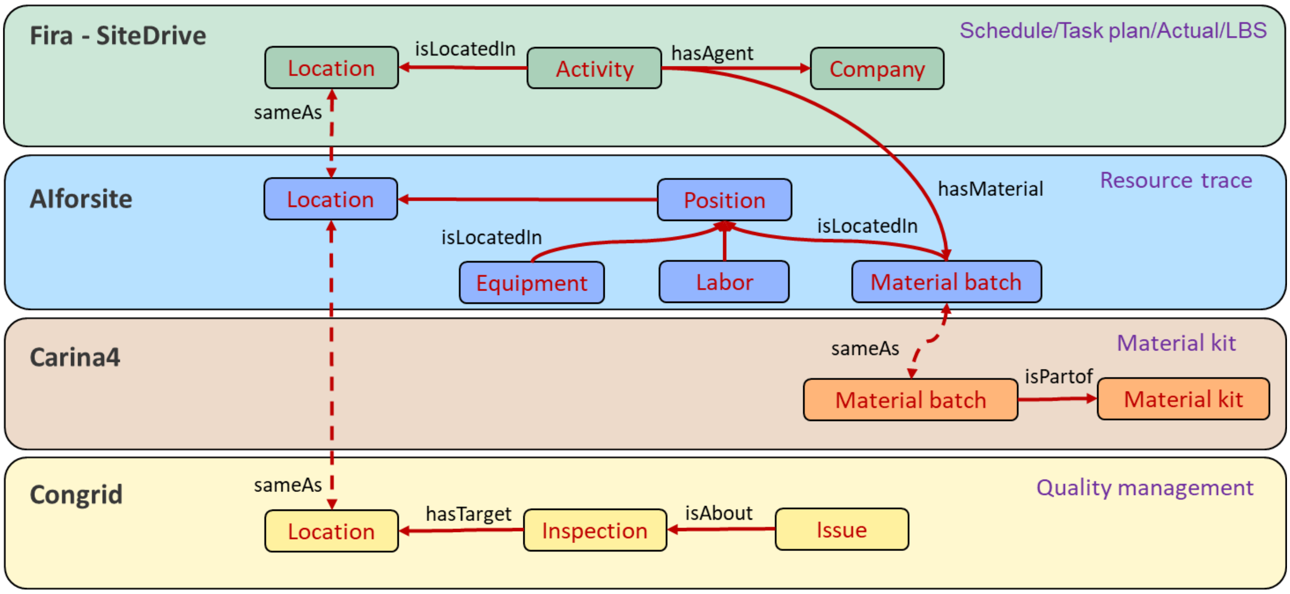

Figure 4. Ontology use case 1 – Subcontract monitoring

In current business models, the management of business relationships with trade contractors is essential for general contractors. The success of construction projects often depends on how well or poorly are these relationships managed. There are many elements related to the management of the trade contractor relationships, including the scoping and generation of work packages, carrying out the bidding process, levelling the bids, instructing trade contractors related to work organisation and safety, verifying the products and materials used by trade contractors, tracking the progress of work and costs, and evaluating the trade contractor performance. In this example, the focus is particularly on the monitoring of work progress.

Typically, trade contractors' work is monitored manually or by using some software applications (possibly also mobile devices), and actual and planned information is compared manually. In the first case, trade contractors are monitored by superintendents who walk the construction site daily and by trade contractors reporting their progress of work at the end of every day and/or at weekly planning meetings. In the second case, some software application or a mobile application may be used, where trade contractors are required to report their work progress in near real-time or at the end of every day. Also, mobile phones and emails are used to communicate, collect and share information. Project file repositories are often used to communicate design project information and document construction processes, products and materials.

However, these processes are often not well managed because processes, methods and different technologies and systems are not properly aligned, resulting in manual double handling of information. Also, as many sources of information are used, it is quite common that design information communicated to trade contractors on site is not the latest, resulting in errors and re-work. Furthermore, data collected through monitoring trade contractors' work is compared to planned data manually. Of course, that means human judgment is necessarily part of this process, which may have limitations.

If instead a linked data would be used to network the data across different data sources, then a lot of this manual double handling of information could be reduced. This would be achieved by avoiding interoperability issues, common in the current construction industry. Furthermore, it would be possible to automate many processes related to managing trade contractor relationships. For example, it would be possible to track the fabrication, transportation and installation of construction products and materials. Having the situational awareness of construction processes is required for the industry to increase its performance, which in turn, together with digital twins of the construction process could enable platform economies in the construction industry. New and novel services could be developed, which would help improve productivity and value creation. It would also benefit software development companies and vendors because there would be limited issues related to exchanging data between different applications.

3.2 Other use cases for linked building data

In addition to the uses case above, there are many other use cases. Linked building data enabled by common ontology is especially useful in the context of when multiple domains are involved in delivering the business value (Curry et al., 2013). For example, some other use cases include, but are not limited to:

<u>Sustainability, circular economy and energy efficiency</u>

All these application areas require collaboration and exchange of information between various disciplines and stakeholders. Linked building data can support the assessment of the sustainability and energy efficiency of designs. It could also be used to assess and manage the processes of circular economy in the construction sector.

<u>Collective and across-project learning</u>

Construction is beset with many problems, of which many take the shape of errors and defects in the end product. With the linked building data, knowledge across projects could be shared, which would make possible to close the gap between the already completed and in use facilities and newly planned and designed facilities (Lee et al., 2016).

Construction platforms and plug-in apps/services

Due to the highly fragmented data environment and interoperability issues in the construction sector, platform technologies' uptake and development have been slow. With the common linked building data environment, building platforms' vision could come to a viable business case.

Digital twin

The vision of the digital twin of a building or construction process requires a common data environment. Again, the linked building data would provide the necessary data structure to facilitate a digital twin development for construction. (Jian et al. 2018, Journal of Computing in Civil Engineering)

Facilities and operations management

Automated multi-domain and category mapping of data for information retrieval during building operations and management, would support facility managers in maintaining the building in order.

Request for Information and Change

It requires the retrieval of data from multiple sources to compile and evaluate the request for information and change. During our benchmarking study in February 2020 in Bay Area, California, we had the opportunity to visit Skanska USA, the division in San Francisco, which is currently working on developing the Request for Information solution based on the idea of linked building data.

4 Towards business models supporting linked data in construction operations

In this section, the intent is to address what kind of business and revenue models can be generated with the linked building data and construction ontology.

4.1 Business model change is imperative

According to the workshop participants of the WDBE pre-event (on May 26th 2020), the older workforce in the construction sector is not tech-savvy, although differences exist between individuals. This older generation has not been educated to utilise digital tools in the management of construction operations. As a result, not all incumbents have realised that eventually, software will be an integral part of construction operations, and long-term investments are needed to digitalise operations. Thus, some companies will fall behind, and some will go forward. However, the younger generation is accustomed to using digital tools, and that generation is already investing in new technologies and having transformation projects aiming at productivity leaps.

The construction sector is highly fragmented, and technologies in the current form, such as the BIM technologies relying on IFC, do not support the exchange of data in a seamless way related to life-cycle processes. Also, implementing the linked building data in the context of traditional business models may face challenges. Thus, a business model change is imperative in light of increasing digitalisation in the construction sector.

According to Osterwalder, the creator of the Business Model Canvas, a business model is an abstract representation of a company's business logic. The model is an abstract comprehension of how a company makes money; in other words, what it offers, to whom it offers and how it can accomplish this. (Osterwalder et al., 2005). Business Model Canvas (BMC) has become an essential tool, or arguably the most important tool, for entrepreneurs to continuously reflect and develop business models.

Figure 5 illustrates the BMC that specifies nine elements and their relationships (Osterwalder, 2004). Key partners perform activities with resources to create value for customers. The company should segment its customers and maintain excellent customer relationships. The customers should be served through appropriate channels. In the best case, the company always creates more revenue than costs. However, growth companies are often expected to invest all their revenues, and thus, these companies tend to create more costs than revenues during the period of rapid growth.

<table>
<tr><td rowspan="2">Key Partners</td><td>Key Activities</td><td rowspan="2">Value Proposition</td><td>Customer Relationships</td><td rowspan="2">Customer Segments</td></tr>
<tr><td>Key Resources</td><td>Channels</td></tr>
<tr><td colspan="2">Cost Structure</td><td colspan="3">Revenue Streams</td></tr>
</table>

Figure 5. A schematic Business Model Canvas (BMC) template (Osterwalder and Pigneur, 2010).

Traditional business models are based on the idea that the competitive edge depends on withholding information and sharing with others as little as possible. In this context, companies are creating their own information systems, which often duplicate the information (classes, properties and relationships) in the original source by either manually transferring data or by using exchange file formats. This has often led to large, closed IT systems that try to solve the whole problem (e.g. ERP systems).

Implementing linked building data with construction ontology in this context would result in a situation that a decision needs to be made: which sources of data are more authoritative for specific classes (e.g., location, actor), properties (e.g., name, type) and relationships (e.g., contains, is_a subclass). Furthermore, if the information in respective dispersed software applications is changed, the question becomes which data are accurate.

To solve the problem of multiple data sources and versions, ISO 19650 defines the concept of a common data environment. This concept combines current thinking, technologies, mostly BIM-based, and workflows for managing project and asset information. Specifically, the ISO 19650-1 defines the common data environment as the "agreed source of information for any given project or asset, for collecting, managing and disseminating each information container through a managed process"[4]. This would alleviate the already mentioned problems but raises several other questions. Who is responsible for setting up the common data environment, who is the owner of the data, and how is the authorship of the data protected? According to the standard, the appointing party (client/owner) should establish the common data environment for the construction project or the asset to be delivered and managed.

However, a better solution would be a common data environment utilising the linked open data and construction ontologies. This would help address all issues related to multiple data sources, versioning of data, ownership and authorship of project and asset data during the construction project delivery, and handover to the client. In other words, we need a digital twin of construction processes and assets. However,

[4] ISO, 2018. ISO 19650-1, Organization and Digitization of Information about Buildings and Civil Engineering Works, Including Building Information Modelling (BIM) — Information Management Using Building Information Modelling — Part 1: Concepts and Principles. International Standards Organisation.

such a technology does not yet exist, and this is where the startups may play a significant role. With the help of startup companies, a technological platform would have to be developed (Sacks et al., 2020).

4.2 Linked building data services and solutions

Technology service providers are probably the first ones to create services based on linked data in the construction sector. We can envision several types of services, which support both data providers and users in utilising linked data:

- **Linked Data Hosting** – provide server capacity to store and access Linked Data.
- **Linked Data Brokering** – provide metadata about datasets to direct users to relevant data
- **Linked Data Linking** – generate, host and serve link sets between existing datasets
- **Linked Data Backup Service** – backup/cache link data to ensure uninterrupted access
- **Linked Data Integrity Service** – check data sources for broken, missing or changed links, validate data against data shape specifications
- **Linked Data Access Control** – manage user roles, authentication, authorisation, single sign-on, encryption, etc.
- **Linked Data Aggregation** – gather data into merged datasets for easier analytics (acknowledging IPR)
- **Linked Data Analytics** – use Linked Data for data analysis or machine learning
- **Linked Data Adapters** – tools to expose existing data as Linked Data through conversions or on-the-fly adapters
- **Linked Data Consultancy** – help companies to utilise Linked Data technologies

Currently, only a few linked building data providers exist. The ones that have already been developed, such as the Trinity by GraphMatrix in the USA and Platform of Trust in Finland, are utilising the concepts and principles of linked data and construction ontologies. These technologies can be used to create and provide micro-services for different use cases, such as predictive analytics, automated tendering and facilities management. As has been already stated, linked building data and construction ontologies are especially useful in conducting activities and solving problems related to multiple domains, such as between trades, and data therein at the same time.

GraphMatrix exploits the Solid Foundation's technology to develop their own services. Solid is a project led by MIT Prof. Tim Berners-Lee, the World Wide Web inventor. The objective is to change the way Web applications work today by providing true data ownership as well as improved privacy. It is a technology for organising data, applications, and identities on the web by building on existing web standards. Based on this technology, GraphMatrix provides several services in four application areas, including services for construction information exchange, productivity, facilities management and logistics handling. Figure 6a illustrates a conceptual model of the GraphMatrix system and Figure 6 b shows the service for interlinking project documents.

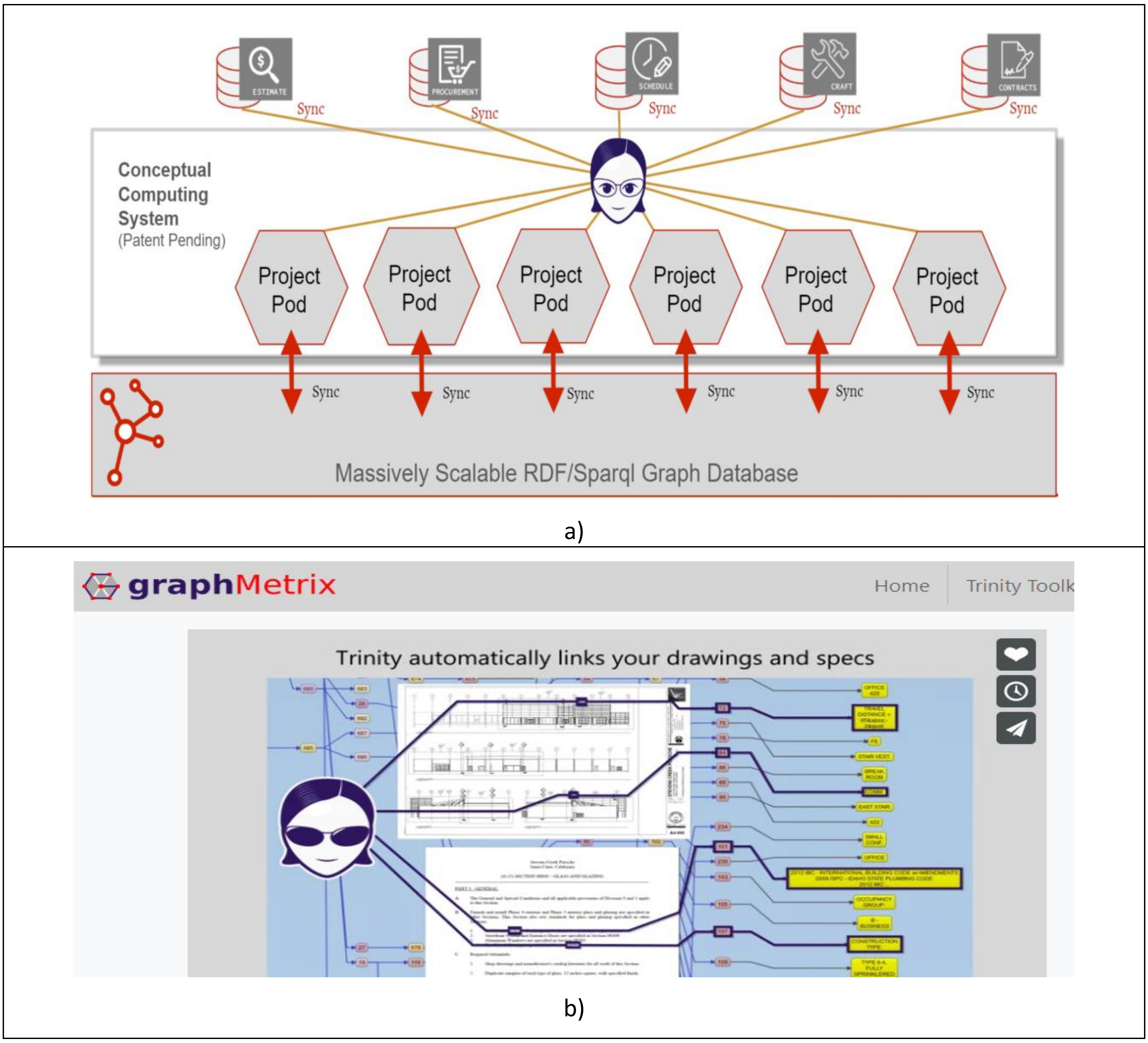

Figure 6. GraphMetrix's Trinity: 2) conceptual model of the GraphMatrix System; and b) Trinity services for interlinking related documents in the project database (adapted from https://graphmetrix.com/).

Platform of Trust supports the integration, i.e., merging and utilisation, of incompatible data coming from various sources. The objective is to support the development of smart services and more comprehensive analyses for decision-making and new business opportunities by reducing cost and time related to the market. Platform of trust does not collect data, but only provides the integration services using APIs. Customer can decide where data are stored, who can use the data, and for what purpose.

Currently, Platform of Trust supports data integration across various data products, such as humidity, temperature, CO2 sensors, and data sources, such as Schneider EcoStruxure™, Granlund Manager, and AccuWeather. Platform of Trust is planning to incorporate DICO ontology[5] (partly developed in the DiCtion project) for the construction phase.

[5] Digital Construction Ontologies (2020). URL: https://w3id.org/digitalconstruction. Retrieved 20.10.2020.

4.3 Ecosystem and revenue models for providing linked data services

The provision of linked data services would enable the development of a 'Linked Data Ecosystem', consisting of various actors, such as linked data providers, data consumers, data brokers and regulatory entities (Archer et al., 2013). Linked data providers provide their data as linked data. In the context of a construction project, any actor could be considered as a data provider. Data consumers, also any actor in the construction project, reuse linked data for their and project's benefit. Thus, in a construction project, an actor can be both a data provider and a data consumer. (Archer et al., 2013)

Data brokers are usually third-party organisations that provide linked data services, such as data brokering, analytics, backup, integrity and access control. Data brokers may also provide data services, such as linking product data to BIM, requests for information or changes, deviation/error data exchange, process and product analysis or information exchange to buildability and maintainability. Data brokers can provide market places for these services as well, often in the form of cloud-based service platforms (See, e.g., Lavikka et al., 2018). Finally, regulatory entities are usually national public administrations or transnational institutions, such as the European Commission, which regulate data exchange, through GDPR and other similar types of rules and regulations. (Archer et al., 2013) Figure 7 illustrates a linked data ecosystem that consists of key actors: regulatory entities, data brokers, linked data providers and consumers.

In general, a business ecosystem's actions are based on trust and mutual benefit (Moore, 1996). In the construction sector context, the linked data ecosystem's shared goal could be, e.g., better quality construction processes and end-products (materials, buildings etc.). In the best case, the resources and competences of the ecosystem actors are complementary. The linked data business ecosystem would be founded on linked data technologies and tools, defined processes, plans, norms and rules. The ecosystem would be based on the interaction between organisations. The interaction would aim to create value to the customers; in this case, the construction supply chain actors. However, it must be noted that linked data technologies can also be used only to improve organisations' internal processes (Pellegrini et al., 2013). Although, we claim that the use of linked data between organisations would bring more significant benefits to the whole construction sector.

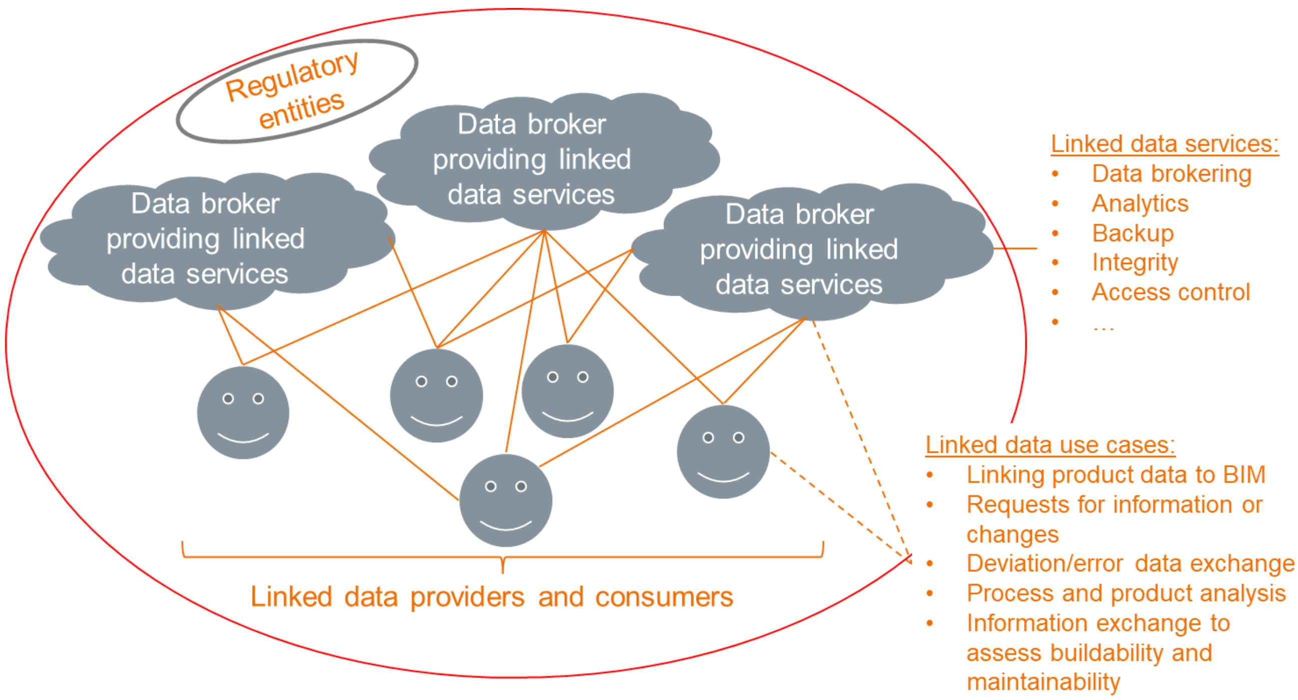

Figure 7. Linked data ecosystem

Algorithms, Internet and cloud computing were the building blocks for the consumer web's revenue models. First, in the early days, the web content was free for the users, but the users had to get accustomed to the several ads included in the web pages they were browsing. This 'free for the users' business model was based on the processing of personal data. In sum, companies paid for the web page owner/service to advertise their products and services for the users. Later on, users had to get accustomed to subscription-based business models practised by web services, such as Yousician, Spotify, and Netflix, based on trust and consent-based data sharing. The benefit of subscription-based business models has been that the user does not have to watch ads and receive personalised services. For example, algorithms are used to recommend content based on the users' previous service actions.

The previous linked-data deployments, such as Linked Open Data Cloud, have been based on the use of encyclopaedical, governmental, geographical, life sciences or cultural data. However, most building data can only be shared among authorised users because there are physical security, data security, competition, IPR and privacy concerns (Novak and Tjoa, 2019). Linked building data requires role/actor-based authentication, authorisation, access control and single sign-on methods. Advertisement is not an applicable business model in machine-to-machine systems. Currently, no "proven" linked building data revenue models exist. One challenge is the extraction of value in a decentralised system. Another challenge is the definition of a control point/role that allows value extraction.

However, based on the consumer web's revenue models, similar types of revenue models could also be valid in case of linked data services. The data brokers could provide linked building data services applying the following three types of revenue models:

1. Service fees
    - Subscription in the form of periodical payments
    - Usage-based, for example, depending number of accesses, the number of links generated/served, and the amounts of data stored or transferred.
2. Freemium
    - Provide basic Linked Data services for free and do business with high availability/quality (backup/cache, reference integrity checking, usage statistics) and/or analytics applications based on linked data for a fee.
3. Outcome/value-based fee
    - Customer pays for a measurable outcome, e.g. "20% increase in productivity". Note that in this revenue model, the risk is moved from the customer to the supplier. However, the parties can agree that compensation is provided to the supplier if he exceeds expectations.

Service fees and freemium types of revenue models are probable models for sharing and using data in the construction sector. It may be that the outcome/value-based fees will not be that popular, since a study revealed that the current business models of construction companies are too similar to enable outcome/value-based competition (Pekuri et al., 2015).

5 Conclusions

This report discusses the rationale, ontology use cases, services and revenue models for linked building data. The implementation of linked data into the construction sector will take time. In general, new technology implementation is challenging as the parties who should use it, do not often understand the technology on a detailed-enough level. Also, for some companies, technology's cost is an essential factor in deciding whether to implement it.

One obvious solution is that data brokers, i.e., linked data service providers, show the construction sector parties how linked building data benefits them. One benefit of linked data is that data stays in the authors' ownership, although other parties can apply the data for the smart supply chain management. Another benefit is that linked data enables structuring data in relevant ways, allowing the use and creation of algorithms, thus, enabling automation and better data analysis. Still, viable business cases for linked building data in the construction context are missing. However, GraphMatrix and Platform of Trust, as examples, are data brokers that may start the transformation process towards applying linked data. The future will show us.

This report claims that a business model change is imperative for digitalising construction operations. Companies need to change from 'guarding their information' towards 'benefiting from sharing and integrating linked data' to stay competitive in the era of digitalisation. The design and construction actors in the sector could collaborate with startups and accelerators of the data brokers, technology providers, in the digitalisation efforts. However, most of the construction sector companies are at the beginning of their digitalisation road; these companies are only considering how to digitalise their operations. On the other hand, other companies are already trying and failing with business processes and business models. Only a few companies in the construction sector have changed their business model to better benefit from digitalisation.

In practice, several parties and actions are needed to boost the utilisation of linked building data and construction ontologies. Contractors could help create viable use cases by expressing their supply chain management challenges, which could probably be solved through better data analysis. Also, owners and general contractors need to start demanding linked building data for better software interoperability. On the other hand, software providers need to start using the linked building data technologies and construction ontologies. Co-creation between the software providers and construction-sector companies is needed to create software that fits the construction sector's needs. The report states that a linked data ecosystem should be created where we have data brokers (linked data service providers), data providers and consumers (one party can be both). Regulatory entities are also needed in the ecosystem, ensuring that business is fair, and data is used according to GDPR.

References

Archer, P., Dekkers, M., Goedertier, S. and Loutas, N. (2013), "Study on business models for Linked Open Government Data", available at: https://joinup.ec.europa.eu/sites/default/files/document/2013-11/Study_on_business_models_for_Linked_Open_Government_Data_BM4LOGD_v1.00.pdf.

Curry, E., O'Donnell, J., Corry, E., Hasan, S., Keane, M. and O'Riain, S. (2013), "Linking building data in the cloud: Integrating cross-domain building data using linked data", *Advanced Engineering Informatics*, Vol. 27, pp. 206–219.

Jones, L.C. (2020), *Flip 'Don't Share Data' Mantras — Introducing Gartner's 'Must Share Data Unless' Data Sharing Model*, https://www.gartner.com/en/documents/3989750/flip-don-t-share-data-mantras-introducing-gartner-s-must, available at: https://www.gartner.com/en/documents/3989750/flip-don-t-share-data-mantras-introducing-gartner-s-must.

Lavikka, R., Kallio, J., Casey, T. and Airaksinen, M. (2018), "Digital disruption of the AEC industry: technology-oriented scenarios for possible future development paths", *Construction Management and Economics*, Vol. 36 No. 11, doi:10.1080/01446193.2018.1476729.

Lee, D., Chi, H., Wang, J., Wang, X. and Park, C. (2016), "A linked data system framework for sharing construction defect information using ontologies and BIM environments Automation in Construction A linked data system framework for sharing construction defect information using ontologies and BIM environments", *Automation in Construction*, Vol. 68 No. August, pp. 102–113.

Logan, D., Rollings, M., Duncan, A.D., Jones, L.C. and Toncheva, A. (2020), *Survey Analysis: Fifth Annual CDO Survey — Growth Must Continue in Order to Achieve Real Impact*, https://www.gartner.com/en/documents/3982574/survey-analysis-fifth-annual-cdo-survey-growth-must-cont, available at: https://www.gartner.com/en/documents/3982574/survey-analysis-fifth-annual-cdo-survey-growth-must-cont.

Moore, J.F. (1996), *The Death of Competition: Leadership & Strategy in the Age of Business Ecosystems*, Harper Business, New York.

Novak, N.M. and Tjoa, A.M. (2019), "Towards a business value framework for linked enterprise data", *RIVF 2019 - Proceedings: 2019 IEEE-RIVF International Conference on Computing and Communication Technologies*, IEEE, pp. 1–6.

Osterwalder, A. and Pigneur, Y. (2010), *Business model generation: a handbook for visionaries, game changers, and challengers*, John Wiley & Sons, John Wiley & Sons.

Osterwalder, A., Pigneur, Y. and Tucci, C.L. (2005), "Clarifying business models: Origins, present, and future of the concept", *Communications of the Association for Information Systems*, Vol. 16, pp. 1–25.

Pauwels, P. (2014), "Supporting decision-making in the building life-cycle using linked building data", *Buildings*, Vol. 4 No. 3, pp. 549–579.

Pekuri, A., Pekuri, L. and Haapasalo, H. (2015), "Business models and project selection in construction companies", *Construction Innovation*, Vol. 15 No. 2, pp. 180–197.

Pellegrini, T., Dirschl, C. and Eck, K. (2013), "Linked Data Business Cube – Modelling Semantic Web business models", *Proc. of Share-PSI 2.0 Krems Workshop: A self sustaining business model for open data*.

Sacks, R., Eastman, C., Lee, G. and Teicholz, P. (2018), *BIM Handbook: A Guide to Building Information Modeling for Owners, Designers, Engineers, Contractors, and Facility Managers*, John Wiley & Sons, New Jersey, USA.

Sacks, R., Girolami, M. and Ioannis, B. (2020), "Building Information Modelling, Artificial Intelligence and Construction Tech.", *Developments in the Built Environment*, Vol. 100011, available at: https://doi.org/10.1016/j.dibe.2020.100011.

Werbrouck, J., Pauwels, P., Beetz, J. and van Berlo, L. (2019), "Towards a decentralised common data environment using linked building data and the solid ecosystem", *ICT in Design, Construction and Management in Architecture, Engineering, Construction and Operations (AECO) 36th CIB W78 2019 Conference*, Northumbria, UK, pp. 113–123.

Zhang, C. and Beetz, J. (2016), "Querying linked building data using SPARQL with functional extensions", *eWork and eBusiness in Architecture, Engineering and Construction - Proceedings of the 11th European Conference on Product and Process Modelling, ECPPM 2016*, Limassol, Cyprus.

Appendix – Linked Data technologies

Technologies and principles of linked building data

Moving building data to the cloud requires designing cloud-based data services from an interoperability perspective. Linked building data provides the mechanism for data publishing and integration "separate from the applications they respectively reside in and to rely on a data representation in an open data format that is commonly agreed upon" (Pauwels, 2014:559). The main elements of the linked building data approach include: (1) Universal Resource Identifiers (URIs) to name things, (2) Resource Description Framework (RDF) for representing data, (3) Linked data principles, and (4) common vocabularies to establish and share understanding (Curry et al., 2013).

Identifiers

For avoiding duplicated identifiers (i.e., globally unique, retrievable identifiers) in denoting different things, coordination between participants is required. Web technology and its infrastructure could be used to create the naming scheme. URIs, and more recently, Internationalized Resource Identifiers (IRIs), have been implemented to identify resources globally and associate disjoint pieces of data. URIs can be dereferenced via the Hypertext Transfer Protocol (in Web browser or application). A resource could be a file (e.g., a digital document, a digital image, or a digital video), a real-world entity (e.g., a person, a company, a device, or an event), or an abstract concept (e.g., building space, building system or a type of a relationship).

Graph structure

The Resource Description Framework (RDF) is the schema-less (not following a predictable structure) and self-describing (the labels of the graph describe the data itself) machine-readable representation format and method for graph-based data used on the Semantic Web to represent information. Data and facts are specified as statements and are expressed as triplets of a subject, predicate, and object. For example, the statement "Person A is the Occupant of Room 201'" is expressed in the triple format: Subject – "Person A"; Predicate – "is the Occupant of"; Object – "Room 201". In RDF, URIs are used for identification and expression of the triple statement, including the relationships: Subject – http://www. deri.ie/ about/ team/ member/ edward_curry#me; Predicate – http://www.vocab.deri.ie/rooms#occupant; Object – http://www.lab.linked data.deri.ie/2010/derirooms#r202e.

The triples of URIs may be joined together to form a graph of information. In the graph, subjects and objects are represented as nodes and predicate as an arc. RDF can therefore be understood from two different perspectives simultaneous: a set of simple statements (the truth of each of which can be verified separately) and a graph of interrelated information. The advantages of graph model of RDF are the following:

- Merging: Information can be combined from multiple sources.

  If there are other data sources about a building - for instance, related to occupant comfort or energy efficiency - the triples from those two data sources could be combined to form a larger graph.

- Linking: Non-local references across data sources can be made in a natural manner.

  Since the URIs are globally unique and retrievable identifiers, the subject and object of a triple can be objects residing in different hosts anywhere on the net. The overall graph can thus span multiple different hosts.

- Databases: The efficient storage of RDF graphs

  Multiple RDF database management systems are available, both free and commercial ones.

- Standardisation: A standard graph-based data format.

  RDF is the only standardised graph-based data format and RDF databases provide only standard-based NoSQL databases. There are standard serialisation formats, conceptual models, and query languages. This improves tool chain interoperability and prevents the vendor lock-in in data management solutions.

Querying linked building data

RDF and linked building data provide an open and common environment, traditional data modelling methods are limited by the scope of their underlying schemas, for sharing, integrating and linking data different data sources and domains. SPARQL is the technology that allows federated queries across heterogeneous linked data sources (Zhang and Beetz, 2016). SPARQL Protocol and RDF Query Language (SPARQL) are W3C standards for querying RDF data. It has a SQL-like syntax and could be used to query RDF triples, hosted in local files or triple stores, via standardised interfaces using the HTTP protocol (Zhang and Beetz, 2016). However, SPARQL does not define specific vocabularies or functions to support the many use cases common in the construction domain. For example, to query useful relationships and properties (e.g., typing, properties, spatial relations etc.), explicitly defined or implied in building models, is either difficult or impossible with the standard SPARQL (Zhang and Beetz, 2016).

Terminology

Linked building data needs a common terminology, construction ontology of common classes, properties and relationships to be effective. ifcOWL has been developed to provide a Web Ontology Language (OWL) representation of the Industry Foundation Classes (IFC) schema. IFC is the open standard for representing building and construction data (see BuildingSMART). However, IFC-based digital information management in construction has focused on physical products. Less effort has gone into process automation and how physical elements relate to workflows, schedules, and resources. As a result, it is difficult to automatically collect data from various sources to form a real-time situational picture of a construction site. To be able to collect data from various sources and combine it in a meaningful way, and without manual interaction, we need a shared data model. DiCtion has created a data model, an ontology, especially for construction processes. An ontology is a formal description of our knowledge. It ensures a common understanding of how we represent the data of things and processes and resources and the relationships between them in a certain domain.

Data exchange

A serialisation format is needed whenever RDF data is written manually, returned as a result of a URI access or SPARQL query, or exported from an RDF database to another. A serialisation format represents an RDF graph or some fragment of it as a string, possibly a long one.

To avoid vendor lock-in, open serialisation for data transfer should be a favoured, such as the JSON-LD, Turtle or XML. JSON-LD is a lightweight Linked Data format, easy for humans to read and write. It is based on the JSON format and provides a way to help JSON data interoperate at Web-scale. It is an ideal data format for programming environments, REST Web services, and unstructured databases such as Apache CouchDB and

MongoDB. Terse RDF Triple Language (Turtle) is a syntax and file format for expressing data in the Resource Description Framework (RDF) data model. Turtle syntax is similar to that of SPARQL, an RDF query language. It is a common data format for storing RDF data, along with N-Triples, JSON-LD and RDF/XML. RDF/XML is a syntax, defined by the W3C,[2] to express (i.e. serialize) an RDF graph as an XML document.

Reasoning

It is the reasoning enabled by RDFS and OWL that is essential to enable usage of the Linked Building Data as "one huge database". Based on existing triples in an RDF graph, a reasoning engine may can check the consistency or validity of data and can also derive additional triples. There are three types of reasoning possible with RDF:

- Ontology reasoning. The utilisation of the information about the relations of terms, such as subclasses, subproperties, inverse properties, and equivalence or disjointness of terms. There are a number of different ontology reasoning engines available, such as Hermit, Pellet, and Fact++.
- Rule-based reasoning. Application-specific rule-based reasoning systems using standard rule formats, such as SWRL (Semantic Web Rule Language).
- Data validation. Checking if data conforms to expected data shapes, using the SHACL (Shapes Constraint Language) and related reasoners.

Ontologies

In order to develop a linked building data ecosystem, the construction sector needs a common vocabulary, an ontology of classes, properties and relationships. Ontology has roots in the philosophy, particularly in the field of metaphysics, and means the study of the nature of being, categorise of existence. Simply, it is the question about what does exist and what can be said to exist. In the context of construction, ontologies considered for structuring the linked building data include, but are not limited to:

- **General**
  - Basic Formal Ontology (BFO): https://buffalo.app.box.com/v/bfo-iso-owl-cl
  - Time (OWL-Time): https://www.w3.org/TR/owl-time/
  - Units of measurement (QUDT): http://www.qudt.org/release2/qudt-catalog.html
  - Agents (FOAF): http://xmlns.com/foaf/spec/
- **Design/structure of a building**
  - Building Topology Ontology (BOT): http://www.w3id.org/bot
  - BIM models (ifcOWL): https://technical.buildingsmart.org/standards/ifc/
- **Construction stage**
  - Digital Construction Ontologies (DICO): https://w3id.org/digitalconstruction/: Activities, locations, physical entities, information entities, agents, roles, labelings, ...
- **Operational stage**
  - RealEstateCore (REC): https://www.realestatecore.io/
  - The Platform Of Trust ontology: https://standards.oftrust.net/v1/
- **Sensor data**
  - Semantic Sensor Networks (SSN): https://www.w3.org/TR/vocab-ssn/
  - Smart Appliances (Saref): https://ontology.tno.nl/saref/