D6.8

Intermediate report on improved data integration platform

Deliverable Report: 6.8 Final version

D6.8

Intermediate report on improved data integration platform / open server for data and tools

Main author	Léon van Berlo (TNO)			
Co-authors	Pim van den Helm (TNO),			
Dissemination	Public			

Document history

| Version	Date	Status	Produced by	Comments
	2017-02-17	Draft	Pim vd Helm	
	2017-02-23	Final draft	Léon van Berlo	
	2017-02-27	Reviewed	Karl-Heinz Haefele; Joachim Benner	
	2017-02-28	Reviewed	Cartwright John	
	2017-03-01	Final	Léon van Berlo	Reviewers comments processed
	2017.04.25	Approved	Freek Bomhof	
	2017.04.25	Submitted / Published	Annemarie Mahieu	

Colophon

Publishable executive summary

Within the STREAMER project the challenge is to provide design teams with advanced design tools by improving the open interoperability between Building Information Modelling (BIM) and Geographical Information System (GIS) by enabling model-based analysis of the energy performance.

It is the objective to effectively manage information flow, knowledge integration, communication and decision-making in the participatory semantics-driven design process according to the principles of Integrated Project Delivery.

In the design phase of a complex project like a healthcare district, the data varies widely. These data are highly distributed and heterogeneous, thereby implying the use of multiple models and resources. This report will define the base for the final solution on how BIM and GIS-based assessments tools can be coupled with the product lifecycle models in order to be straightforward and easy to use.

Given the situation and analysis we propose a combination of technologies to be used in STREAMER for an improved data and tools integration. The concept is sketched in the following image:
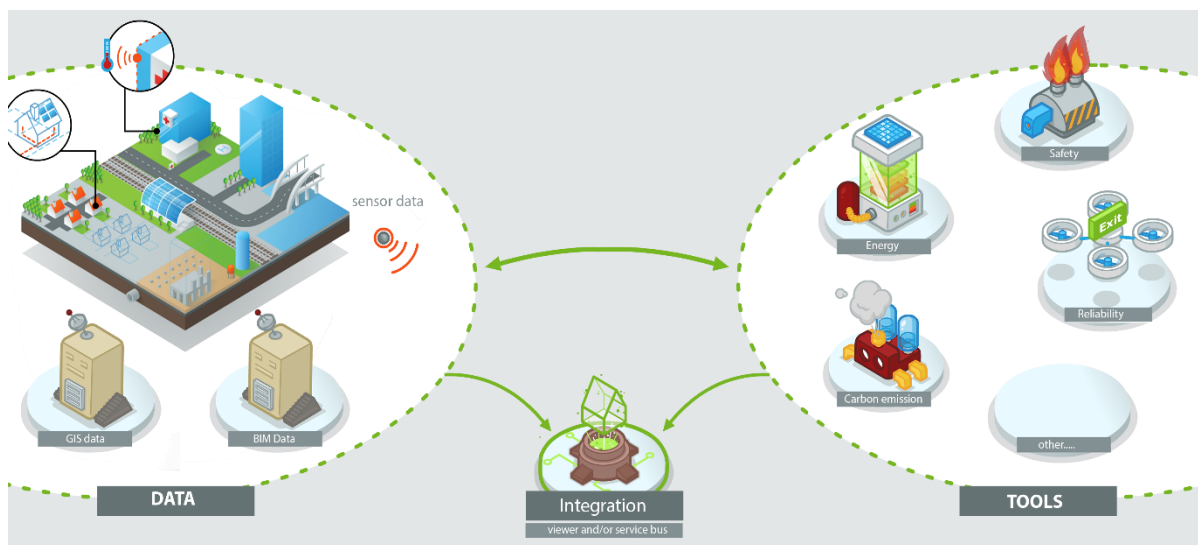
Generally the STREAMER data can be split into three categories:
- The tools (simulations)
- The data (BIM, GIS and sensor data)
- The integration platform

To limit the number of one-to-one connections the concept of an 'Enterprise Service Bus' (ESB) is introduced as a concrete solution to the 'Information Logistics' layer in the PLM architecture. An ESB is implementing a communication system between mutually interacting software applications in a service-

oriented architecture (SOA). In the next phase of the STREAMER project the focus will be on implementing the integration concepts described in this report.

Based on previous experiences and first experiments, we recommend use of WSO2 Enterprise Service BUS and Zapier Developers platform to open unlock the 'Information logistics' layer of the proposed PLM architecture.
A solution to connect the several different entities is with a so called 'event driven' approach.

List of acronyms and abbreviations

- API: Application Programming Interface
- CAAD: Computer Aided Architectural Design
- GIS: Geographic Information System
- GML: Geography Markup Language
- GUI: Graphical User Interface
- OGC: Open Geospatial Consortium
- OSM: OpenStreeetMap
- STREAMER: Semantics-driven Design through Geo and Building Information Modelling for Energy-efficient Buildings Integrated in Mixed-use Healthcare Districts
- TNO: Netherlands Organisation for Applied Scientific Research
- XML: Extensible Markup Language
- XSLT: Extensible Stylesheet Language Transformation

Definitions

Web Service – A web service is a service offered by an electronic device to another electronic device, communicating with each other via the World Wide Web .

Contents

1. Introduction

Within the STREAMER project the challenge is to provide design teams with advanced design tools by improving the open interoperability between Building Information Modelling (BIM) and Geographical Information System (GIS) by enabling model-based analysis of the energy performance.

It is the objective to effectively manage information flow, knowledge integration, communication and decision-making in the participatory semantics-driven design process according to the principles of Integrated Project Delivery.

In the design phase of a complex project like a healthcare district, the data varies widely. These data are highly distributed and heterogeneous, thereby implying the use of multiple models and resources. This report will define the strategy how BIM and GIS-based energy performance assessments tools can be coupled with the product lifecycle models in order to be straightforward and easy to use.

2.   Integration of Data

On the topic of integration of data, the most basic integration of data is using a single data format or standard. All data can then be stored in a single location in a single defined format. The choice of the right data model is always difficult.

IFC is the only standardized model for buildings, which claims to cover the complete life cycle. Accordingly, the IFC model is very comprehensive and supports nearly 770 entities and nearly 410 property sets (IFC4). With IFC4, the building model can have arbitrary geospatial references and generalized civil engineering (e.g. pavements) and geographical elements (e.g. roads, trees). Usually, not all the geometry is described explicitly. Hence, the importing application needs basic geometrical functions at least. The import and export of IFC is supported by free and commercial tools, but needs more experience than e.g. XML. Currently, IFC is supported by all major CAAD vendors and most of them have already been certified or are in an on-going certification process performed by BuildingSMART International (bSI).

Theoretically, IFC might be able to cover all processes, which are planned within STREAMER. One important reason why this approach does not work in reality, is that when generating a digital representation of buildings, the complete life cycle is not fully considered in an integrative way. There are many actors in many domains with many software applications, which have their own view of the building. All want to have models, which exactly represents the information required and speak their languages. This leads to domain-specific, specialized data models, like gbXML or ILCD. On the other hand, it is not easy to synchronize two major domains, such as the GIS and BIM worlds. The responsible standardization organizations (OGC and bSI) started nearly at the same time (20 years ago) to independently develop standards for their domains. Of course, software applications of the GIS world most likely support OGC standards, while BIM applications provide better support for bSI standards. For designing buildings, it is essential to consider the neighbourhood to obtain a holistic view on the project.

STREAMER Deliverable 6.5 gives an overview of "Advance Mapping Structures and Standards". In Chapter 2 modelling languages, data models, data formats, and standardised web services, which might be relevant to STREAMER are introduced. The D6.5 report also concludes that STREAMER has to consider more than one standard. Depending on data availability and project phase, different data models might be deployed. STREAMER has to consider using different data models in a combined manner in order to realize the intended workflow.

If a process requires more than one data model, for instance, when considering a detailed building model in its neighbourhood, a so-called hybrid approach is required. In this approach, data are kept in the original models and:

- enriched by other information sources,

- converted into the requested target models,
- integrated into a single software application, or
- linked, by the corresponding software tools, if necessary.

In order to bridge the gap between GIS and BIM worlds, models can be converted. A typical example is to generalize a detailed IFC model to a certain LoD and save it as a CityGML model. If an application covers more than one domain and is able to import different models, further analysis and simulations can be performed on the integrated model.

As IFC is the only open data model for BIM which covers the whole life cycle of the building and which can also be used in the construction process, it is recommended for use as a basis for all STREAMER processes. In general, this means that IFC is used for designing the new or refurbished building. Depending on the deployed software application and on additionally required information, models can be enriched, converted, or linked.

If different data models describe the same real world object (e.g. a building), it is likely that similar or even the same information is stored multiple times. This does not mean that the information is redundant ithin the process.

The management of different data models including data linking, web services, and preserving data consistency, is difficult to perform manually. Therefore, the next chapter will highlight options for the integration of tools.

3.   Integration of tools

Tools can be offline (desktop applications) or online. Integration of desktop tools is not a topic related to the STREAMER project, so this chapter will only deal with the integration of online tools.

As Deliverable D3.3 of STREAMER already stated, the objective of this project is to bring together the BIM and the simulation worlds. The principle that engineers would continue to use the software they use for their performance and quality rather than change for software they have not yet mastered still stands. So, the proposed approach aims to foster a "smooth" transition to integrate the practices of engineers into a BIM workflow rather than imposing a particular tool.

The biggest challenge for integrating tools is the integration of the different data structures.

There are several options to integrate online tools:
- Centralisation into one tool
- Centralized data with distributed tools (one to one connections between data and tools)
- Distributed data connected via a central tool (APIs and BUS systems)
- Distributed data and distributed tools (Standardised API)

3.1   Centralisation in one tool

Using the same analogy as with data integration, the most basic way to integrate tools is to integrate them into one (new) tool. This creates a black box with as much features as possible.

This single tool or group of tools from the same family is often seen as the most efficient way to programmers. The conditions are very controllable and the result of the tool is very predictable.

Experience (and research) has learned us that the AEC industry is so fragmented and specialized, that a tool (or even group of tools) can never serve the bulk of the industry.

This approach is not feasible in the industry and therefore also not feasible for STREAMER.

3.2   Centralized data with distributed tools

There are many misconceptions about BIM. Maybe the biggest misconception is that BIM is centralizing all the data of a project into a single data repository. Because most BIM software tools work with a central database that is being used for all features, this concept is being copied on a project scale. The recent rise of online BIM collaboration platforms is feeding this concept. Many online BIM platforms are being developed at the moment. All of them have specific features and both positive and negative points.

A well-known initiative is the open source BIMserver.org project. This initiative started in 2008 when the industry was still unaware of the concept of online BIM. The stability of the BIMserver.org platform is proven enterprise ready. Many commercial applications build on the stable base of BIMserver. STREAMER partner TNO is heavily involved in the development and use of BIMserver.

Using BIMserver (or any other online database for that matter) as a single source of data for several online applications gives many advantages. The data is always in sync between partners and disciplines; users always work with the latest version of the data; etc.

However, BIMserver research has shown that collaboration processes using a central data repository are actually decreasing productivity of the project. Working efficiently with distributed data storages is more effective for the project than trying to centralize everything.

Within the STREAMER project the diversity of the data formats and standards also limits the flexibility of a solution based on a single data model.

Therefore this solution is not seen as feasible within the STREAMER solution space.

3.3 Distributed data connected via a central tool

Different online BIM services from Autodesk, Bentley, Tekla, Graphisoft, Nemetschek, Onuma and others support different APIs. Connecting BIM applications requires n*(n-1)/2 custom interfaces.

To limit the number of one-to-one connections the concept of an 'Enterprise Service Bus' (ESB) is introduced. An ESB is implementing a communication system between mutually interacting software applications in a service-oriented architecture (SOA). It implements a software architecture like this:
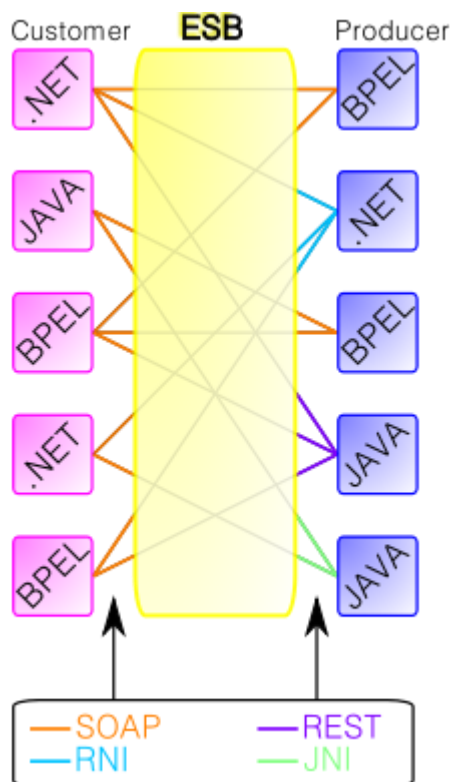

Figure 1: principle of an ESB

As it implements a software architecture for distributed computing, it also implements a special variant of the more general client-server model. Whereas in general any application using ESB can behave as server or client in turns. ESB promotes agility and flexibility with regard to high protocol-level communication between applications.

An ESB limits the number of one to one connections since every application is connected to the ESB and the communication is streamlined by the ESB. Advanced ESB systems can also validate, aggregate, check and transform the data before forwarding or distributing it to connected tools.
STREAMER Deliverable 3.3 identified the need for intermediate tools to transfer data from a BIM format to the simulation tool by:
• checking / controlling data (including geometric data)
• assigning properties from external data sources BIM (material properties, use and occupation of the areas, equipment parameters)
An ESB could be used to facilitate this.

The use of an ESB to implement an improved data integration platform and open server for data and tools is very feasible and we are exploring the use in an experimental way.

3.4 Distributed data and distributed tools (standardised API)

Different online BIM services from Autodesk, Bentley, Tekla, Graphisoft, Nemetschek, Onuma and others support different APIs. Connecting BIM applications requires n*(n-1)/2 custom interfaces.
Where ESB solutions are very useful to connect legacy systems in a given environment, the one-to-one connections to the ESB are still very labour intensive and limited to the API competences of the tools.

Implementing effective BIM and GIS data in the cloud requires automated (machine-to-machine) interaction between different online BIM services. A standardized web services API for online BIM services would enable machine-to-machine exchange of BIM data and enable innovation in the industry. BIMsie is intended for this, being the open web services API for online BIM services [BIMsie2017].
The recent developments around BIM Bots use a simplified version of the original BIMSie API to connect online tools.
Chapter 2.4.3 of STREAMER Deliverable 6.5 recommends the evaluation of the BIMSie API and that is what has been done. The latest developments around BIM bots seem to be very promising since they create the possibility to automate connections between online tools.

The BIM Bots (updated BIMSie concept) therefore seems very feasible for a STREAMER solution for improved data and tools integration. However, STREAMER has to consider using different data models in a combined manner in order to realize the intended workflow. In Chapter 2.4.2 of D6.5 the WFS is indicated as important for STREAMER to access the data of the hospital district itself and the surrounding district.

The BIM Bots concept therefore has to be evaluated against the use of both BIM data formats, as GIS formats and the combination of BIM Bots API and WFS API.

The next phase of the STREAMER project will focus on the use of the ESB and BIM Bots concept in combination. This will automatically require the combined use of IFC and (City) GML as data formats.

4.   Proposed solution

4.1   STREAMER context

Within STREAMER the PLM server plays a central role in the connection between several systems. PLM (Product Lifecycle Management) consists in centralizing all information regarding an industrial product. The PLM is intended to be the central server between all the different tools and data. All the documentation, models, computation results will be stored on the server, will be accessible at any time by any actor and will be visible through any solution that will implement the PLM API.

The principle is the following: a PLM client sends a request to this central BUS. Whatever the language or the type of demand, the Information Logistics can answer by soliciting the good data server in the right language. The localization of the data is known by the directory server and the translation between BIM and GIS data is made possible thanks to the semantic contained in this configuration server
The technology used by the PLM is the same as the ESB technology described in chapter 3.3 of this document. The standardisation of the API described in chapter 3.4 is comparable with the standard API of the PLM.
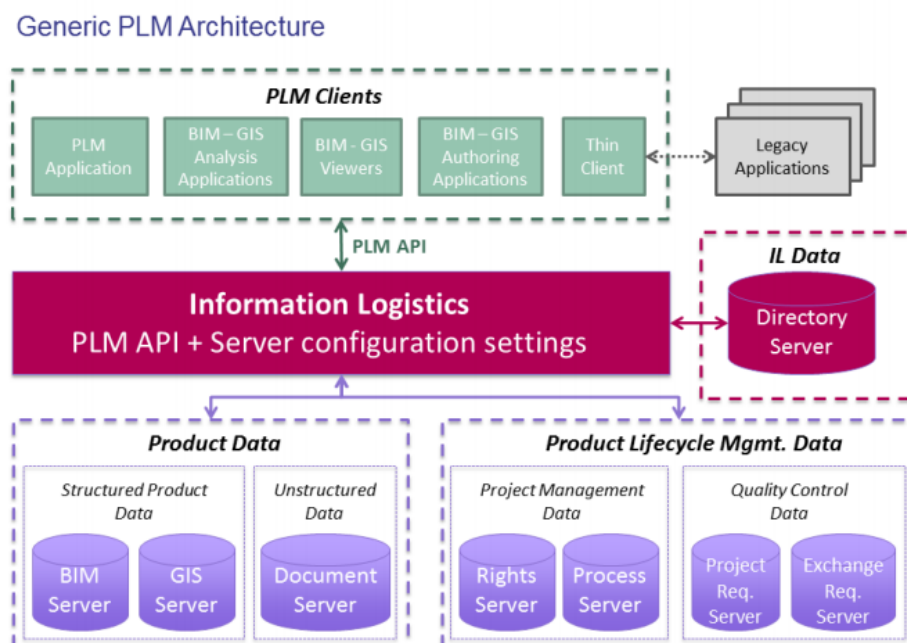

Figure 2: Generic PLM architecture in STREAMER

A PLM system coupling BIM, GIS and PLM concepts allows to formalize and to streamline the exchanges between several companies around the digital model BIM, the GIS spatial data and around all the data of the project, by protecting the management of the various actors' rights during the whole life of the project.

The PLM solution proposed in STREAMER D5.3 has been described regarding the inputs and outputs of the other workpackages and its functionalities are described via an API detailed D5.3. The objective of this API is to be as generic as possible in order to be able to connect to any PLM client in entrance and any type of data server in output. This intention is comparable with the concept described in chapter 3.4 of this document.

Nevertheless there are some challenging issues:
- Being as generic as possible through the PLM API
- Being able to exchange data within various servers and to translate the data from a server to another one.
- Determine a good PLM that will enable to use the PLM API

There is a risk that no fully generic PLM API can be implemented. We think the solution will be to reach directly the good server that contains the data instead of going through the upper layout called generic directory server.

4.2   SOA approach to PLM

In practice a PLM is a new system that tries to centralize all data and connect tools to a single server. An SOA approach to PLM must make PLM information available to any person or system that needs information and allow information flows to be orchestrated to support company-specific business processes. Ultimately, an SOA environment for PLM allows relationships between information in multiple different applications to be defined and managed to provide insight that has only been possible before then all data resided in a single system (IBM 2008).

Making the PLM available to other systems on the Information Logistics layer (the ESB part) creates the option to directly interact with other systems and data stores. This means the BIM and GIS integration can be one-to-one ór via the ESB and PLM. This creates flexibility in the solutions and prevents a single point of failure in the whole network.

4.3   Overview solution approach

Generally the STREAMER data can be split into three categories:
- The tools (simulations)
- The data (BIM, GIS and sensor data)
- The integration platform

Given the situation and analyses we propose a combination of technologies to be used in STREAMER for an improved data and tools integration. The concept is sketched in the following image:
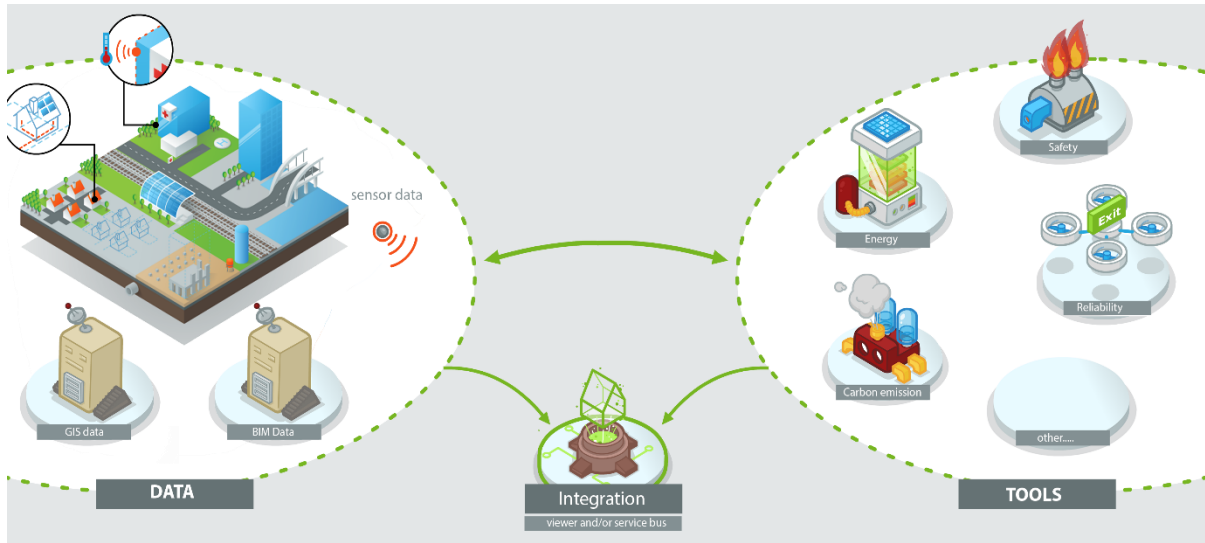
Figure 3: Integration principles of data and tools

The main difference between the conclusions of D5.3 and this report is that the use PLM and a generic API to create a single point of integration comes with effects that are prohibitive to detailed integration of data.

Experiments suggest the use of PLM combined with a more open ESB approach when necessary to integrate the data and tools. The most promising solution to connect the several different entities is with a so called 'event driven' approach. This suggestion is seen as an additional functionality on top of the PLM approach.

4.4   Event Driven architecture

The basic principle of this concept is that the sensors, the BIM data and the simulations are not lined up in a predefined workflow. The interaction between these tools is 'event driven'.
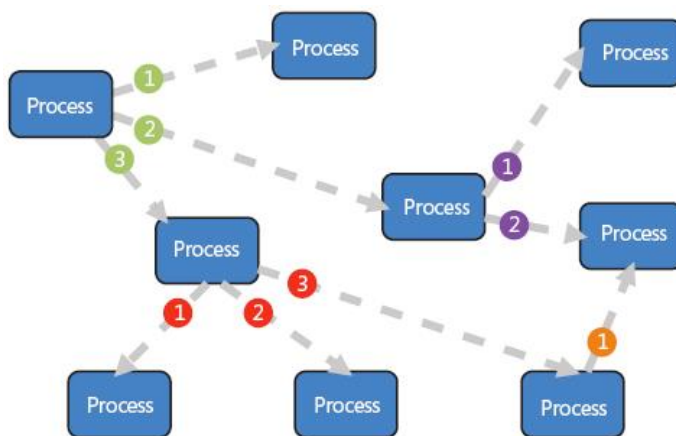

Figure 4: principle of event-driven architecture (source: Microsoft)

The principle of 'even driven' interaction is not new in IT architecture but has only been applied experimentally in the BIM and GIS applications. It facilitates the use of small '*microservices'* that seamlessly connect and integrate. These small services "do one thing and do it well". It is described as follows:

- The services are small - fine-grained to perform a single function.
- The organization culture should embrace automation of tasks.
- The culture and design principles should embrace failure and faults, similar to anti-fragile systems.
- Each service is elastic, resilient, composable, minimal, and complete.
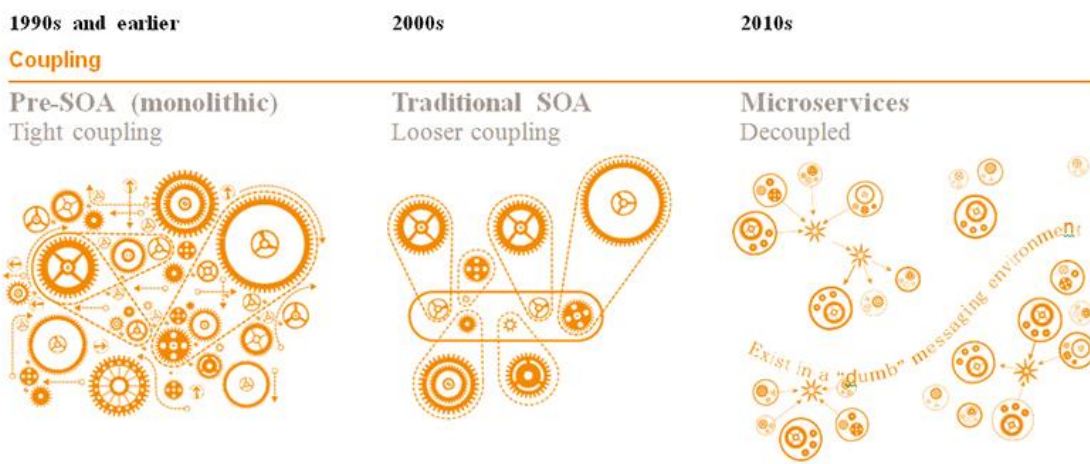

Figure 5: Linking 'microservices' in an event driven ecosystem (source: PWC)

The concept of stand-alone services is proposed on multiple events and deliverables in STREAMER.
Every simulation tool should be developed as an online service that is minimal and complete.
This brings a challenge for the individual simulation tools, but is the most durable architecture for long-term innovations and business models.

5.   Recommendations

In the next phase of the STREAMER project the focus will be on implementing the integration concepts described in this report. This needs to be done using tools and API's.

Within STREAMER D3.3 a list of energy performance simulation tools is gathered. Deliverable 6.5 had a strong focus on mapping data structures and standards. Deliverable 5.3 describes the PLM architecture and has a strong focus on the Lascom solution.
Deliverable 5.3 also identified challenges of being as generic as possible through the PLM API, and being able to exchange data within various servers and to translate the data from a server to another one. To lower the risk of these challenges we propose to intensify the use of the 'Information Logistics' layer (the ESB) of the PLM.
Connecting tools directly to this layer of the PLM architecture improves the changes to integrate the data in a more semantic way.

At the moment we are trying to use WSO2 Enterprise Service BUS and/or the Zapier Developers platform to connect with BIMserver.org for IFC data (with BIM Bots compliant API), Deegree WFS server (with WFS API).

To increase the speed of prototyping, the developer platform of Zapier was used. Zapier creates the possibility to connect the STREAMER ecosystem with undreds of tools like Google Spreadsheet, Gmail, Slack, dropbox, etc. This can also be used for rapid development, realizing that this creates dependence to this single online system.

Additionally API managers and other tools could be used. The following tools connect quite well with the previous stated tools:
-   WSO2 API manager
-   Zapier and Zapier developers platform
-   WSO2 Identity and Access management

6. References

[bSI2016] buildingSMART home page, http://www.buildingsmart.org/, last access February 2016

[CityGML2012] Gröger, G., Kolbe, T. H., Nagel, C., Häfele, K.-H.: OGC City Geography Markup Language (CityGML) Encoding Standard, Version: 2.0.0, Open Geospatial Consortium, OGC 12-019, 2012

[Haefele2015] Häfele, K.-H., Benner, J.: D6.5 Advance Mapping Structures and Standards, STREAMER Deliverable of work package 6, February 2015

[Helm2015] Van den Helm, P., Häfele, K.-H., Benner, J.: D6.3 Energy modelling interface IFC – CityGML, STREAMER Deliverable of work package 6, August 2015

[IFC-4Spec2014] IFC4 – Addendum 1 (Final Standard), Specification, 2014, Description for IfcDistributionPort, http://www.buildingsmart-tech.org/ifc/IFC4/Add1/html/schema/ifcsharedbldgserviceelements/lexical/ifcdistributionport.htm, last access March 2016

[ISO16739] ISO 16739:2013, Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries, 2013

[ISO19101-1] ISO 19101-1:2014, Geographic information — Reference model — Part 1: Fundamentals, 2014

[ISO 19136] ISO 19136:2007, Geographic information -- Geography Markup Language (GML), 2007

[Tag2016] Tag (metadata), Wikipedia, https://en.wikipedia.org/wiki/Tag_(metadata), last update 6. June 2016, last access June 2016

[WebService2017] Web service, Wikipedia, https://en.wikipedia.org/wiki/Web_service, last update 31. January 2017, last access February 2017