PAPER • OPEN ACCESS

Handling huge and complex 3D geometries with Semantic Web technology

To cite this article: Peter Bonsma *et al* 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **364** 012041

IOP Conf. Series: Materials Science and Engineering **364** (2018) 012041 doi:10.1088/1757-899X/364/1/012041

Handling huge and complex 3D geometries with Semantic Web technology

Peter Bonsma[1], Iveta Bonsma[1], Anna Elisabetta Ziri[2], Ernesto Iadanza[3], Federica Maietti[4], Marco Medici[4], Federico Ferrari[4], Rizal Sebastian[5], Sander Bruinenberg[5], Pedro Martín Lerones[6]

1. RDF Ltd., Bulgaria

2. Nemoris srl., Italy

3. Consorzio Futuro in Ricerca, Italy

4. University of Ferrara, Department of Architecture, Italy

5. DEMO B.V., Netherlands, the

6. Fundación CARTIF, Spain

peter.bonsma@rdf.bg; iveta.bonsma@rdf.bg; annaelisabetta.ziri@nemoris.it; ernesto.iadanza@unifi.it; federica.maietti@unife.it; marco.medici@unife.it; federico.ferrari@unife.it; rizal@demobv.nl; sander.bruinenberg@demobv.nl; pedler@cartif.es

Abstract. In INCEPTION, a European collaborative research project, a Heritage BIM (H-BIM) ontology is being developed to store all relevant semantic data concerning cultural heritage objects. Similar to other projects dealing with storing semantics, one of the major questions is whether, and if yes, how should geometry be stored using semantic web technology. The INCEPTION cross-disciplinary research consortium chose to allow the storage of all relevant geometric information using semantic web technology. The alternative is to store geometry in a different way, or storing only the aggregated parts of geometry, for example through bounding box representations.

The geometry is originally generated by a CAD/BIM system and, as we are dealing with Cultural Heritage, in many cases it is derived from 3D point clouds. These result in a large amount of 3D data to be stored using semantic web technology. A well-known issue is that the performance of databases and inferencing engines for semantic web data drops considerably when the data grows to very large sizes. This paper explains how the performance issues on these large sets of geometric data can be solved while still being able to use the databases, inferencing engines, and the geometric data effectively.

IOP Conf. Series: Materials Science and Engineering **364** (2018) 012041 doi:10.1088/1757-899X/364/1/012041

1. Introduction

This paper describes the architecture and choices behind H-BIM approach within the INCEPTION project. It explains how the H-BIM approach can serve as the connection between all available knowledge, how it is populated and why it is structured the way it is. It also shows how all relevant BIM content (in the shape of IFC files) can be converted to both the H-BIM and the GEOM ontologies for respectively the semantic and geometric content of these BIMs. The complete set of geometrical information is processed even for large IFC files, overcoming performance issues as normally would be expected with the huge amounts of data that the INCEPTION platform should handle.

INCEPTION realizes innovation in 3D modelling of cultural heritage through an inclusive approach for time-dynamic 3D reconstruction of artefacts, built and social environments. It enriches the European identity through understanding of how European cultural heritage continuously evolves over long periods of time.

INCEPTION's Inclusive approach introduces novel solutions of 3D digital modelling:

- forever: INCEPTION "Time Machine" that represents an innovative use of time-scale for dynamic 3D reconstruction with emphasis on how the modelled cultural heritage evolves over time in association with its built and social environments;
- for everybody: portable, user-friendly and cost-effective hardware and software instruments for 3D capturing, modelling and analysis, as well as INCEPTION Semantic Web technologies and open-standard platform to creating and sharing understanding between various end-user group;
- from everywhere: INCEPTION's proposed standard procedures for data acquisition and open-standard format for cultural Heritage Building Information Modelling (H-BIM).

The INCEPTION project is funded by the European Commission under H2020-EU.3.6.3. - Reflective societies - cultural heritage and European identity.

The main technical challenge within this stage of the project is to move BIM models from a wide variety of existing commercial CAD systems towards a semantic web environment. One of the known issues is that CAD systems exports large amounts of both geometric and non-geometric data, both relevant for further use in the INCEPTION H-BIM platform. Although Semantic Web technologies are powerful, scaling towards large amounts of data is still an issue. This means that whatever the solution is, it has to be able to work with standard tools for Semantic Web technology on these large amounts of data without running into scalability issues.

The widely supported Open-BIM standard IFC is able to contain both geometric and non-geometric information like classification, relations and properties. Additionally, IFC is supported by all major CAD systems, scales very well and is very rich in the sense of BIM information.

The solution found and discussed within the project scope is to split semantic geometric knowledge from all other semantic knowledge, i.e. non-geometric semantic knowledge. Both parts are converted to semantic web technology. For the geometry the GEOM Ontology has been developed. All non-geometric knowledge is converted and stored in the H-BIM Ontology. Part of this H-BIM Ontology is directly generated from the IFC schema. This generation can be performed for all relevant IFC schema's, i.e. IFC 2x3 TC1, IFC4 ADD2 and IFC 4x1 Final. The generation is IFC-dependent but similar for all named IFC schemas and therefore different compared to existing Semantic Web BIM solutions like ifcOWL.

All generated semantic web content is loaded into the INCEPTION platform. To ensure performance, the geometric information stored in the GEOM Ontology can be used separately from the non-geometric information. This way it can be excluded from Fuseki-2. Visualization can be requested on demand by a dedicated service directly on top of the semantic web content. This dedicated service for geometry generation has limited functionality on top of the semantic web content and is therefore able to overcome the scalability issues intrinsic to all current developments for semantic web.

As a result in case, if there is a limited amount of information contained in the BIM it can be generated and transferred to the Fuseki 2 triple store. In the case that the amount of data is larger, all geometrically related data can still be converted but can be excluded from the Fuseki 2store. This structure still allows all benefits from using semantic web technology as having distributed data and connecting information from several different sources.

In addition to the conversion from IFC conversion from Collada, OBJ and DWG is also technically possible and is implemented step-by-step within the developed tools/applications and later on in services as part of the project. All converted data can be used in an integrated manner and visualization of 3D geometrical data is possible as sub-selection of all converted content in an integrated 3D view. Early tooling already available is able to visualize this integrated 3D data on virtually any modern mobile device.

2. IFC

IFC and BIM are often named in one sentence, although there are many BIM standards (both open and closed). The main idea behind IFC is to enable open exchange within and between all available disciplines within the Building & Construction sector. This holds for both non-geometrical data, i.e. relations, classification, properties, as well as for geometrical data. In practice, both geometric and non-geometric data are stored as a snap-shot of the data (IFC). Complex parametrical information is not included in the stored information. Since the geometrical information is used by different technical disciplines for different purposes, the definition of the geometry as an entity is essential. IFC has a strong focus on product information, storing this in an object based manner, although it has (limited) process related capabilities it is rarely used for supporting processes. The IFC 2x3 TC1 version of the standard already dates back a decennium, although it still is the main IFC exchange format used. Until very recently this 10-year-old IFC 2x3 TC1 standard was the only format software products could get certified for. While IFC 4 ADD2 has been available for a few years only a limited number of companies support this standard. The companies that do support the standard still struggle to export and import each other's data correctly. Within INCEPTION we found that although we were planning to use IFC4 for a few demonstration cases and the used CAD software was capable of exporting IFC4, the quality of the exported data was not good enough to be used even for use within the prototype software. We have to note here that the way CAD software is used is typical and different from normal architectural design work.

3. ifcXML / ifcOWL

There are two different ifcXML variants. The first variant is specific for IFC2x3 TC1 and although it really is 100% valid XML against a dedicated XSD schema, the style is very close to the original structure as found within STEP. The fact that it is valid XML / XSD enables many tools to be used for such a type of content. The amount of (developer) tools available, like XSLT, is much larger than the amount of tools available for SPFF (Step Physical File Format). The same counts for software developers that need to implement solutions.

One lesson learned from applying and using this XML serialization of IFC is that while it is technically valid XML / XSD, the fact that the structure still follows STEP ideas means the real practical use is somewhat limited compared to other XML standards. To overcome this for IFC4 a different XML serialization has been defined. The outcome is still 100% correct XML against a dedicated XSD, containing more than 99.9% of the knowledge as stored in SPFF, while the structure is much closer to what is expected by tooling in XML. This potentially enables the same XML enabled tooling to work more effectively on the content. The new XML serialization used for IFC XML serialization is called SimpleXML.

One of the problems with having these IFC dedicated different versions of XML serialization is that each new serialization will take considerable time to be implemented. Although IFC is a mature and widely supported open standard and officially promotes these new ifcXML serializations, the fact that the implementation cost is at the individual client and / or toolbox provider is often a reason not to support these serializations.

IOP Conf. Series: Materials Science and Engineering **364** (2018) 012041 doi:10.1088/1757-899X/364/1/012041

The Semantic Web serialization variant of IFC is called ifcOWL. First efforts were made over 10 years ago, however only the past few years it got serious interest and also an official serialization definition was proposed for BuildingSMART. While, like IFC, this first ifcOWL standard is perfectly following Semantic Web technology and can be processed by all available tools , however the solution still very close to the original structure of STEP. This means although technically working the content is semantically not in line wityh how somebody would define an ontology representing the same content.

In case of INCEPTION the official ifcOWL definition was not deemed to be a viable solution. Partly because of the architecture within INCEPTION where geometric knowledge is separated from non-geometric knowledge, and in part also because both parts are required to be closer to the original idea behind Semantic Web.

4. Cultural Heritage within H-BIM

Concerning the H-part of the H-BIM ontology, the peculiarity of 3D models representing historical buildings is that they are not conceived to be a footprint for construction and maintenance but more a representation of a valuable existing or reconstructed environment. These models has to represent all the complexity of the related context in term of ancient architectural techniques and components, important events related to the history of the building / site and all the data that define the actual value of the monument in term of historical heritage.

According to Historic England (2017) and Arayici et al. (2017) custom content creation is a requirement for H-BIM projects, since usual construction object libraries are related to new-build rather than historic building components. However, the inclusion of BIM objects from very different sources (creators, skills, purpose and context) is problematic for information consistency. Adopting standards is strongly advised and only content that has been checked for compliancy with these standards should be allowed in the project BIM object library and subsequently in the models. Thus to facilitating the international exchange of information ISO 128 (graphical representation of objects on technical drawings), ISO 16739:2013 (IFC for data sharing), ISO 1302:2002 (surface texturing in technical product documentation) and ISO 21127:2014 (reference ontology for the interchange of cultural heritage information) needs to be considered.

The H-BIM ontology is based on very well constructed and complete ontologies, while also integrating the BIM part of the H-BIM ontology.

5. The Architecture of the INCEPTION H-BIM approach

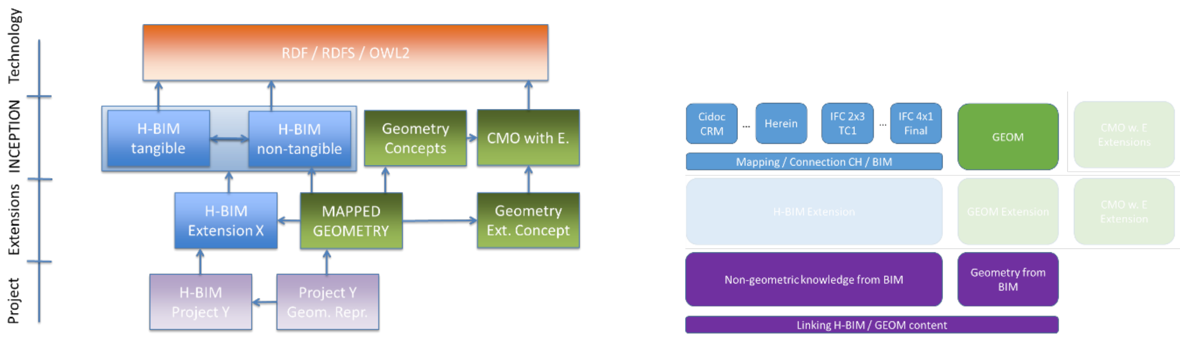

Figure 1, Architecture of the INCEPTION H-BIM approach

The image on the left represents the overall architecture of the INCEPTION H-BIM approach, while the image on the right is the detailed technical version of the same picture. In the technical picture the important parts for this paper are highlighted. In this diagram the blue and green parts are project independent, while the purple parts are project dependent.

6. Non-geometric Semantics from BIM

IOP Conf. Series: Materials Science and Engineering **364** (2018) 012041 doi:10.1088/1757-899X/364/1/012041

6.1. IFC Classification

The classification of BIM objects is a large task, with early estimations calculating over 25 person-years of work for a basic starting set of concepts. Within BuildingSMART, the organization behind IFC, bsDD and ISO 12006-3 (organization of information about construction works) an initial development in this direction was undertaken, are estimated to have used much more effort, while they still do not have a (nearly) finished classification available.

Within INCEPTION it is not viable to reinvent the wheel for BIM classification, and as such to keep things 'simple', the basic set of concepts as available in IFC and agreed on by most CAD vendors will be used. The classification is stored in the schema in the shape of different entities (read classes). Looking at 'instances' with a geometric representation we can restrict ourselves to any instance from entity IFCPRODUCT.

Making use of inheritance this adds up to approximately 100 different entities; depending on the IFC schema selected a few more or less. These entities include IFCWALL, IFCWINDOW, IFCDOOR, IFCCOLUMN etc. Within the conversion towards H-BIM the IFC part in the name will be removed and we will follow CamelCase naming conventions, resulting in Wall, Window, Door, Column, etc. This can be considered a good starting point to enrich classification for more specific historical definitions using the flexibility of Semantic Web approach. Some of the entities have a further refinement on IFC schema level by using enumerations. At the moment the conversions towards H-BIM do not contain these refinements, but if required it can be added easily in new versions of the INCEPTION H-BIM ontology.

6.2. IFC Relations

The IFC schemas donot only contain classification information, but also many different relations. Within the H-BIM ontology three relations are especially of interest:

- Decomposition
- Containment
- Grouping

Both decomposition and containment are defined on top of each IFCPRODUCT entity. Starting from an IFCSITE instance it is expected in a correct IFC file to find any IFCPRODUCT instance defined walking over these 2 relations (except for void instances). This knowledge is integrated into the H-BIM ontology, and as every available IFC schema works in the same way the implementation in the H-BIM ontology is the same for every IFC schema.

The grouping mechanism allows users of CAD systems to create their own grouping instances where every group is allowed to contain a subset of instances of the IFCPRODUCT entity. Allowing this specific information to pass through to the H-BIM ontology enables designers to create relevant groups, which are similarly represented in the H-BIM ontology.

6.3. IFC Properties

The object based structure of IFC is often used to 'carry' a vast amount of semantics (read information) in the form of property sets containing other property sets and / or properties. As IFCPRODUCT instances can be typed (completely in IFC4 and IFC4x1, partly in IFC2x3 TC1) different instances can share different property sets and properties.

Properties and property sets can be defined towards an agreed definition (containing +/- 3000 different properties), using naming convention PSet_xyz for the name of the property set. However every user can also create its own or application dependent properties and property sets.

Independent of the source of the properties and property sets, all available knowledge will be able to be stored against the H-BIM ontology. The structure of the properties is not defined in the H-BIM ontology other than that unit, and quantity type information can be stored.

7. Geometric Semantics from BIM - Alternative possible Geometric Ontologies

7.1. OntoBREP Ontology
OntoBREP is an ontology for CAD data and geometric constraints as a link between product models and semantic robot task descriptions. This ontology comes close to what is required within INCEPTION for being able to store geometrical data from BIM. One of the draw backs of this ontology is that it focusses on visualization exchange formats rather than geometrical definition exchange formats. In that sense it is much closer to open standards like Collada, X3D and older standards like VRML.

The real semantics within geometry representations like Boolean Operations, clipping, feature modelling within design trees cannot be stored using this standard. Another issue is that to be able to visualize any content defined against this ontology currently the results need to be converted back to STEP to be able to generate any type of visualization, which would make it complex to define a service generating geometric representations, enabling third party applications to visualize geometry.

7.2. ifcOWL Ontologies
Every ifcOWL ontology is generated by a piece of software that in itself is independent from the converted schema. This means also all geometric related entities within every IFC schema are converted towards classes almost 1-to-1. Already within IFC2x3 TC1 there are almost 200 entities somehow storing / describing geometrical information. It is important to note that IFCOBJECT entities, and its sub entities like IFCWALL, IFCDOOR etc. are not part of these 200 entities. Within IFC4 this grows to well over 200 entities where several of the entities are not 1-to-1 comparable with the entities in IFC2x3 TC1. IFC4x1 includes the alignment extension to support GIS related geometry and therefore is even larger than IFC4.

Looking at the 600+ entities in all relevant IFC versions there is a lot of redundancy and a large difference in aggregation level of the geometrical descriptions. Another issue is that there is no library available that understands the geometrical part of ifcOWL directly. Even if we would use ifcOWL there still would be no software components available to enable visualization and development of such components would be far too time consuming to develop it within the INCEPTION project.

7.3. CMO with Extensions
The open standard CMO with Extensions contains a small set of ‘primitives’, i.e. geometrical concepts to work with. While the architecture from this ontology has proven to be working, the geometry that can be represented with this set of geometrical concepts is limited. All typical geometrical concepts for GIS, such as Clothoids, vertical and horizontal alignments are missing as well as Bezier curves, B-Splines and NURBS are not available.

However this is still the most promising solution that could be found and its base is used and extended into a separate GEOM ontology. The GEOM ontology contains solutions for Splines, NURBS, 3D / 2D Boolean Operations, Boundary Representations, materials, textures and many other concepts.

8. H-BIM ontology Architecture

8.1. Technology Layer
Within this layer Semantic Web technology is represented. This means use of the RDF, RDFS and OWL2 as top layers of the H-BIM Ontology.

8.2. INCEPTION Layer
The INCEPTION layer of the H-BIM model contains the real knowledge from the specialists within the INCEPTION project. This is knowledge about Cultural Heritage but also knowledge from existing state-of-the-art open BIM and open GIS standards.

8.3. Extension Layer
Even with all the specialist and knowledge available within INCEPTION, there will always be structures missed, incorrect or incomplete. The Extensions layer allows any site, group or even

IOP Conf. Series: Materials Science and Engineering **364** (2018) 012041 doi:10.1088/1757-899X/364/1/012041

INCEPTION partners itself to extend the defined H-BIM ontology in a way appropriate. This way different views and/or more specialized knowledge can be integrated in the INCEPTION platform by every party, enabling optimal flexibility of the platform itself. As an example the INCEPTION H-BIM standard has embedded parts of the semantic structure of IFC (and therefore ifcOWL). This means classes Wall, WallStandardCase and CurtainWall exist. Adding a super class Walls and the knowledge that above named classes inherit from this new class Wall is an extension that works on all available content. A SPARQL query can be created to get all instances of new class Walls and it will directly have content for the majority of the Cultural Heritage H-BIM models stored in the INCEPTION platform.

8.4. Project Layer

Within the project layer the real content is defined, this content is arranged according to the layers above. All content can be queried according to the SPARQL queries defined on top of the INCEPTION layer. It is also possible to create solution specific queries as well as queries dedicated to certain extensions as defined in the extension layer.

9. Conclusion

One major challenge within the INCEPTION project is to create support for geometric representations using Semantic Web technology, taking into account the fact that in the area of Cultural Heritage geometry is often derived from point clouds coming from different 3D surveying techniques, and representing this information is therefore often even more space demanding than normal BIM models.

Within INCEPTION it was decided not to escape the issue itself but to look for a solution where this data rich content can really be converted towards content complying with Semantic Web technology. The distinction between geometric and non-geometric semantics as well as the structured setup to still keep all relevant knowledge allows a system in which data-heavy parts can be excluded from the core data stores, resulting in a minimal performance loss, even when handling vast amounts of data.

While such an approach is already interesting from a theoretical perspective, the INCEPTION project is making good progress and a working solution of this platform is already available and running in the cloud. It has proven to work on several of our demonstration cases and both performance and scalability seem to deliver the results expected from the theoretical solution. The amount of data in combination with the power of semantic web enables INCEPTION to deliver a platform that is working, scalable, powerful and useful.

References

[1] Arayici, Y.; Counsell, J.; Mahdjoubi, L.; Nagy, G. A.; Hawas, S.; Dweidar, K. (Eds.) (2017): Heritage Building Information Modelling. Routledge (UK).

[2] Historic England (2017): BIM for Heritage: Developing a Historic Building Information Model. Swindon, Historic England [accessed 25 January 2018]. Available at: <https://historicengland.org.uk/images-books/publications/bim-for-heritage/>.