Building Information Deduced

State and potentials for Information query in Building Information Modelling

Martin Tamke, Morten Myrup Jensen, Jakob Beetz, Thomas Krijnen, Dag Fjeld Edvardsen

1,2 CITA - Royal Academy of Fine Arts, Schools of Architecture, Design and Conservation
3,4 Design Systems Group Department of the Built Environment Eindhoven University of Technology
5 Catenda Norway

1,2 cita.karch.dk 3,4 http://www.ds.arch.tue.nl/ 5 http://catenda.no
1,2 {martin.tamke|morten.myrup}@kadk.dk 3,4 {J.Beetz|T.F.Krijnen}@tue.nl
5 dag.fjeld.edvardsen@catenda.no

In recent years, Building Information Models have become commonplace in building profession. The extensive use and increasing experience with BIM models offers new perspectives and potentials for design and planning. A recent stakeholder study conducted by the authors of this paper show that in practice models are no longer solely observed as culmination of knowledge in a 3d representation of future built structures, but as a source of information in itself. Experienced users of BIM want to Find Information within a model or across a set of these and Compare models in order to evaluate states of a model, differences in separate models or models from different point of time. Current BIM tools support both modes only in a rudimentary form. This paper discusses current modes of information query within and across BIM models, shows beneficial scenarios for building and planning practice through customised queries and exemplifies these on the base of a scripted tool. This customized approach is used to test approaches for a machine-based assessment of Level of detail and BIM-readiness in BIM models.

Keywords: Building Information Modelling, BIM, IFC, openBIM, Information query, Data extraction

INTRODUCTION
In recent years, Building Information Models have become commonplace in building profession. The extensive use and increasing experience with BIM models offers new perspectives and potentials for design and planning. A recent stakeholder study conducted by the authors of this paper show that in practice models are no longer solely observed as culmination
of knowledge in a 3d representation of future built structures, but as a source of information in itself. Experienced users of BIM want to find information within a model or across a set of these and compare models in order to evaluate states of a model, differences in separate models or models from different point of time. Current BIM tools support both modes only in a rudimentary form. This paper discusses current modes of information query within and across BIM models, shows beneficial scenarios for building and planning practice through customised queries and exemplifies these on the base of a scripted tool. This customized approach is used to test approaches for a machine-based assessment of Level of detail and BIM-readiness in BIM models.

**BIM IN BUILDING PRACTICE - SCANDINAVIA**

The paper is based on a survey among stakeholders of the building industry in Scandinavia, conducted in the frame of the European research project DURARRAK (Tamke 2014). The stakeholders provided an extensive dataset of over 150 BIM models in the Industry Foundation Classes (IFC) format, which we investigated in order to report on the current use of 3d information processing in architectural practice. Scandinavia was chosen, as the building industry here can be seen as forerunner in the digitalisation of their processes, as the government pushed for usage of open standards and the large public building owners required usage of open BIM in new projects (wong 2009). Stakeholders are engaging in a practice where BIM models are shared and edited collectively and geometric data is enriched with additional metadata. The conducted study shows, that among those stakeholders, that actively use BIM, all hand over or is handed BIM models from partners (Tamke 2014). The practices investigated in the study show a professional environment that uses a wide set of BIM tools. (100% of landsurveyors use Revit to build BIMs from point data, 60% of Architectural offices use Revit, 20% use Bentley Microsystem, 20% use ArchiCad. The interviewed construction company use a wide range of BIM software. ArchiCad is the preferred one for architectural modelling. 100% of the large building owners and facility managers use Revit for maintaining models, while they demand Revit and IFC files from their contractors.)

This paper focuses on BIM files in the IFC format, as our investigations show, that this format is predominantly used for the exchange of BIM data amongst partners in the building profession. This might be due to the fact, that open standards (open-BIM) are enforced in scandinavia through building legislation. OpenBIM proves in practice to provide the versatility needed to accommodate the requirements of a wide range of stakeholders. More than 50% of the interviewed stakeholders use IFC as exchange format with other partners (Tamke 2014). The versatility of the IFC format means on the other hand, that 3D objects can be modelled and described in many not prescribed ways. The data produced is hence inconsistent and not on standardised level of quality. To overcome this issue, data producers and receivers make written agreements on structure and content of BIM models. However, in order to actually exchange data, stakeholders need to validate whether the agreed standards, in respect to quality and status of development, are met in the models.

**Current modes of data inquiry in practice - From CAD to Information Modelling**

Queries within BIM models are currently executed within planning environments, as Autodesk Revit or Bentley Microstation, in specialised tools, as Solibri Model Viewer or Dalux BIM checker and in online based BIM platforms as Bimserver. All methods work predominantly with predefined queries that cannot be adopted to the stakeholder's individual BIM processes. Domain-, stakeholder- and project-specific queries would however be beneficial in order to accommodate the stakeholders specialising processes, which are guided by internal requirements and workflows:

Architects and engineers generate several domain-specific models based on internal libraries.
and enrich these in a collaborative way. The investigated stakeholders use domain specific tools to model the data, but use web based services such as bimsync [11] and byggeweb [12] in order to coordinate their efforts. The models show big differences in information density, modelling approach, Level-of-Development and -Detail. The stakeholder’s BIM processes are shifting with ever-changing client demands and nature of projects. Exchange and collision control of BIM are done via IFC exports whenever the used software differs between stakeholders. However, whenever the software environments are identical, exchange takes place in native formats (Tamke 2014).

Construction companies are taking over the data and refine these to a level ready for fabrication and assembly. Large stakeholders use fully 3D-based software applications such as TEKLA [13] and have developed standardized internal processes. As they are responsible for the materialization of the design they need to implement rigorous quality checks for incoming data and often engage in re-modeling to comply with constructability. A seamless data exchange via IFC is crucial for the stakeholders vision of BIM. In order to assure interoperability of their internal software solutions, the interviewed company NCC [x] validated different BIM software and chose, despite trends of the general profession, ArchiCAD as internal modelling tool.

Building Owners and Administrators - Large corporations dominate the market in Scandinavia. Their operation is engaging with digital Facility Management processes (FM) that incorporate control and maintenance of their building stock. FM is in its essence operating in database, highly specific to the building portfolio of the stakeholders. Hence tools such as Dalux FM [15] offer highly customizable implementations. The requirements of IFC models for FM are different to those in the planning domain, where stakeholders have a hard time to comply with FM tools’ logic and level of development. Typical information kept in FM models are besides the geometric properties: U-values for estimation of energy consumption, expected life span of building parts, warranty periods, building classification codes, and external links to product specifications [10].

Building Owners and Building Legislation pursue in Scandinavia a quest for higher productivity in the building sector through digitalization. They demand the use of IFC for as-built documentation and competition proposals. Here, delivered IFC models are inquired for quantities, volumes as well as sustainable performance through energy simulation tools.

The disperse set of perspectives and approaches towards models necessitates validation and monitoring efforts through the query within or between IFC models and every step of a buildings development.

**Current means to derive Information from BIM - Quality Assurance and Quality Control**

We find that stakeholders query their model in order to:

- Find Information within a model or across a set of these.
- Compare models in order to evaluate states of a model, differences in separate models or models from different point of time.
- Queries on BIM models require two steps in order to create information meaningful for the user: data-extraction and data-analysis (Fig. 1).

Data Extraction is today obscured to the user behind preset tools and queries. Alternatives for custom queries (Mazairac & Beetz 2012) as BimQL (BimQL [X] is a query language implemented on top of Bimserver by Wiet Mazairac. Like its name suggests it acts like SQL/Sparql, and less like other DSLs (Domain specific languages), which are designed to feel similar to human language.) or the programming language connected to the IFC importing plugin for SketchUp exist. (The DSL connected
to the IFC importing plugin for SketchUp [x] is another attempt to make navigating a BIM more human friendly is. Using it, it is possible to write requests like "Sketchup.active_model.layers.collect[](l.name)" to get a collection of all the names of the layers in the model). The two mentioned approaches are however not directed to end-users, nor are they embedded in the common BIM environments.

Data-Analysis on BIM information can use a range of approaches that each have their limitations:

A) **Basic Arithmetric Operations**
Results are produced through counting in filtered datasets or the addition of object’s numeric data, for instance for quantity take off. This approach is highly stakeholder- and model-agnostic, as metadata on 3D objects is neither standardised nor obligatory. Simple queries, which add numbers from specific property definitions, are hence in many cases not providing consistent results through sets of models. Research based tools as the IFC analyzer by Robert Lipman (Lipmann 2010) investigate the IFC on the code level in order to determine their content (Fig. 2). Further analysis steps for instance for compliance check are done external to the tool.

B) **Comparison of Data**
The presence of information is verified against a set of data, i.e. through Boolean operations for Quality Control of BIM, Model Revision Comparison or clash detection (Fig.3). Tools validate the structure of the model and the correct assignment of properties to 3d objects.

C) **Interpretation of Data Through Rule Based Inquiry**
The relation of data to each other is evaluated based on rule sets (Fig.4) that can be imported from external sources or created for project specific purposes. E.g. the physical safety of building designs can be assessed through measurements between object coordinates. Algorithms check whether all components are contained by a floor or columns are positioned on load bearing structures.

Figure 2
Quantity take off performed in Tekla Structure [1].

Figure 3
Clash detection in Autodesk NavisWorks [2].

Figure 4
Rule based model checking with Solibri [3].

Figure 5
Ecotect Analysis of energy consumption based on Revit export in gbxml format [4].
**D) Interpretation of Data Through Information Fusion.**

Information that is not directly given from a single descriptor in the 3d object is derived from the merge of information from several sources. This approach is able to inquire values for parameters that are more abstract as they relate information from within a model to external tools and information, such as the simulated energy consumption values (Fig. 5), the estimated economic building costs or the experienced based timeframe for building phases.

**APPLICATION AND FINDINGS**

Our investigation show a focus of existing tools on predefined queries in single models. Users are not able to create their own data queries, nor are they able to conclude on qualitative and process-related aspects or compare and detect distinctions between different models. A monitoring and validation of single or multiple models could provide new perspectives on data and process if implemented in the architectural workflow:

Validation: the automated assessment could provide stakeholders with insights into the qualities and content of models. A uniform, reproducible tool for the objective measurement of quality and properties IFC models on a large scale can emerge.

BIM Maturity: The assessment of multiple Ifc files can provide an quantitative indicator for the BIM maturity (Bilal Succar et al). Administration on government or similar stakeholder level can collect and use the Ifc files that they constantly receive to extract statistics and get a better picture of the current state of Ifc models in practice. The DURAARK project used this approach to assess the state of Ifc implementation in the Danish building industry.

Performance Management: The work progress of workteams can be visualized through the inspection of the collection of BIM models and the project hence more efficiently managed.

Archiving: Digital Archiving of BIM files requires the constant migration of these to the newest software version (Beetz 2013). Multiple models can be compared and validated in batch mode, providing users with insights whether an upgrade of their software version changes files in an illegitimate way.

Facility management: FM software imports specific object from Ifc files. These have to be present in Building services in predefined quality and setups and stem from multiple domain models of a building. Changes to the building stock require often the re-import of Ifc models besides an initial import. This currently tedious process would be eased through the assessment of all domain models and across multiple models that describe larger building complexes.

Building Performance assessment: building portfolios can be automatically assessed in order to run energy simulations in them and evaluate their energy or other performances in comparison to the local average or the measured performance in reality. This evaluation allows for targeted actions.

**TECHNICAL APPROACH**

In order to investigate the potentials we developed an IFC-extractor to retrieve metadata from selected IFC files stored in bimsync as well as a batch friendly extractor script using the Bimserver API [5]. It gathers basic information as number of floors, spaces, areas, volumes, etc. This information is written into CSV files and further processed using formulas in spreadsheet applications.

The approach is based on the Bimserver API, where the IFC model is mapped to Java classes. Relevant information can be extracted through querying the model for entities, properties and relationships. A big amount of information is not readily available in IFC files, but has to be deduced. The overall floor area of a building can for instance be extracted by first collecting all entities of type "IFC-SLAB" which is of "FLOOR," type. and then for each of these through a walk on the relationship graph to find the connected "IFCAREAMEASURE". As "IFCSIUNIT" is as well collected takes into account that area could be measured in for instance square meters or square millimeters. Other measures extracted from the model are based on geometrical calculations on...
the model (for instance getting the size of the bounding box in world coordinates) or simply on counting the number of building components in the model.

**USE CASE: VALIDATION OF MODELS**

The coupling of batch processing of IFC files and data-fusion based approaches for the assessment of models provides theoretically means to evaluate properties that are not directly provided by the models. The automated detection of Levels-of-Development (LOD) provides the case to evaluate this claim [6].

LOD can be seen as indicator of scales, as the physical limits of information per spatial dimension that determined traditionally the architectural scale do not exist for BIM. A considerable effort is put into the definition of amount of detail in distinct LODs. As LOD does not have a single descriptor, references to diagrams and descriptions (FIG 7) are made in order to describe a certain state in the digital work process. While some authors suggest an LOD definition on base of single objects, our approach summarizes LOD as an amalgamation of information per spatial unit. A simple formula for the estimation of the LOD would hence be the amount of information \( i \) per amount of space \( s \) of the model: LOD \( = \frac{i}{a} \). Information is in this context understood as: the object itself, the objects properties, the objects relationships and its geometrical representation. Space can be the area or volume of the overall model or spatial subdivisions, as floors or building program related partitions. The following paragraphs describe the technical approach and the underlying considerations.

Several characteristics are used to judge the LOD of the model:

**AMOUNT OF INFORMATION: PROPERTIES SETS AND VALUES**

Firstly, the amount of property sets and values is assessed. The resulting metric is the average amount of property sets and property values per product. This provides an indication of the amount of semantics that is attached to the objects, or the extent to which the model constitutes more than merely a set of geometries. However, this metric is susceptible to inconsistent, redundant or empty property values. Some authoring tools for example emit property values with place-holder text along the lines of 'Enter address here', which does not constitute any meaning whatsoever. No attempt is made to filter out these property values. The amount of property values, for a single product, is the sum of all IfcSimpleProperties related to the object either directly or via either a number of IfcComplexProperty instances or via an IfcTypeObject that is related to the object.

**AMOUNT OF INFORMATION: AMOUNT OF RELATIONS BETWEEN ENTITY INSTANCES**

Secondly, the amount of relations in general between entity instances is counted. This gives an indication of the parametric intelligence associated with the model. It is assumed that this value increases throughout the refinement of a model. For this metric, all subtypes of IfcRelationship are considered, except for IfcRelDefinesByProperty instances, as these are already considered as part of the metric outlined above. Examples of the relationships typically encountered in IFC models describe concepts like containment, decomposition, connectivity and the assignment of materials.

**AMOUNT OF INFORMATION: DEDUCED FROM AVERAGE OBJECT SIZE**

On a geometrical level, the average object size is checked to give an indication whether the model consists of mainly coarse object or whether objects are subdivided based on their covering, material or other qualities.
**AMOUNT OF INFORMATION: GEOMETRY - DEDUCED FROM AMOUNT OF VERTICES**

In addition, the amount of vertices is assessed to indicate in what level of detail objects are represented in the file. This is estimated based on a tessellated rendering of the model as provided by IfcOpenShell [x]. Hence it is a function of the deflection precision for the meshing of curved surfaces.

**SPATIAL AREA: DEDUCED FROM AREA INFORMATION STORED WITHIN OBJECTS**

To further qualify the LOD, the aim is to relate the number of objects and points within those objects, to the spatial area that the building encompasses. In order to do so, several alternatives have been investigated. On the one hand, information could be derived from the property sets in the model. For example, the GrossAreaPlanned on a project level, or to aggregate respective values for GrossArea or NetArea for all floor slabs or the GrossFloorArea or NetFloorArea for interior spaces. However, neither of these properties are populated consequently enough, in the set of test models, to rely on them for comparative analysis. Therefore, an attempt is described to infer this information from the geometrical representations of the products in the model.

**SPATIAL AREA: DEDUCED FROM BOUNDING BOX**

The most trivial approach would be to describe a single bounding volume for the entire project. The upside of such an approach is that various aspect models can easily be combined, since the union of two bounding boxes is easily computed and remains fairly constant as long as the various aspect models agree on a similar building envelope. Therefore such an approach conforms to the intuition that merging various aspect models increases the level of development. On the other hand for the bounding volume to be optimal, in the sense that is the minimal volume that contains all parts of the model, is not as necessarily trivial. Hence, often the bounding volume provides an overestimation as it is susceptible to geometric outliers. In the explorations outlined in this paper only a simple Axis-Aligned Bounding Box is constructed.

**Spatial area: deduced from amount of products and building floor area**

An alternative approach is to compare the amount of products to the floor area of the building. However, determining this also is not unambiguous, and rather difficult, if no explicit semantic property set data is available. Calculating the floor area from geometry can be convoluted when considering all the different types of geometric shape items, clipping representation and opening subtractions, that can constitute the geometric representation of a product in an IFC model. Furthermore, false positives due to separating floor layers into structural floors and finishes, combining different aspect models with structural and architectural floors, or mislabelling roof slabs as floors, overestimate the total floor area. Most of these objections can be countered with tailored heuristics and arbitrary threshold values. For example for this purpose, in an attempt to remove floor finishes, a minimum height of 5cm is enforced in order for something to qualify as a floor. But, other than that, the impact of this aspect is deemed negligible in this context.

**SPATIAL AREA: DEDUCED FROM AREA PROVIDED BY IFCSpaces**

Alternatively, a way to derive floor area from the representation of IfcSpaces is presented, but also using this approach, as models with intersecting spaces are encountered, an overestimation of the building area is to be expected.

**Spatial area: deduced from solid volume of the product representation**

To obtain the floor area, from the shape representation of a three-dimensional slab or space, a generic approach is used based on the solid volume of the product representation. A boundary representation (BRep) is obtained by using IfcOpenShell [x], an open source implementation of the IFC file format that has conversion functions for most types of IfcRepresentationItems. From this boundary repre-
sentation, the solid volume is calculated, which is divided by the magnitude of a vector that spans the two vertices of the longest vertical edge, to give the approximation of the horizontal area. Since this approach operates on the BRep rendering of the product, it is agnostic to the way the shape of the product is defined, using for example a faceted BRep or swept solid, and to the boolean operations that operate on the volume as a result of clipping or opening subtractions.

**EVALUATION OF APPROACH ON A DATASET FROM STAKEHOLDERS**

The described computational approach was tested on the in the beginning described dataset of 97 ifc models from stakeholders. It comprises of IFC files in a mix of merged and individual domain models. The models Level of Development on the levels from 0-5 were determined through a manual inspection by an expert (see Fig. 7). Most models were positioned on information level 3, 1 on level 0 and none above level 4. We correlated the information level to the extracted data from the dataset in order to evaluate the data fusion based query of Ifc files (Fig. 8). The result is incoherent. In general a higher linear correlation (0.09) between the extracted values and the assigned information level for the numbers can be observed in the information "Deduced Number of objects per cubic meters" than from any other one. However the diversity in modelling approach, amount of properties given per object and other factors is reflected in the datasets large spread in values. Potential outliers are present in every calculation of the linear correlation. A more conclusive result would likely appear in a more homogeneous dataset. We find that the LOD is rising throughout the process of a group of stakeholders from design to fabrication. This logic is however broken, when a model shifts domain. Another finding is that the models have a relative consistent level of information per stakeholders, this is however very different from that of other stakeholders, that claim to work in the same level of development. The profession is far from having a homogeneous understanding of level-of-detail. Apart from the possible influences of the algorithms that are underlying the extractor tool, the main reasons for unexpected results, and generally low data content of the models, might be because of:

- Inconsistencies in the IFC model: Stakeholders do not use the IFC format in a consistent way and the user's modelling skills not always sufficient to create optimal IFCs.. The usage is not dictated by IFC, hence it is possible to make inconsistent models.

- Wrong declaration of IFC classes for objects: Exports from proprietary formats are not always true to the IFC classes.

- Flawed modelling approach, e.g. some object types do not translate directly into the correct IFC property sets.

- Minimal amount of object properties: Stakeholders minimize the amount of attributes and metadata to their own need.

- Domain models: BIM files that are produced for a specific purpose (domain models or models made for a specific kind of analysis) and hence do not contain the usual building components.

**DISCUSSION OF APPROACH**

The inconsistencies within and between Ifc models are a major challenge for the proposed approach. It is the result of the versatility of the Ifc format. This allows on the other hand a project specific and hence efficient approach towards the use of models in BIM. It allows furthermore for the description of buildings through domain specific models, typically Architecture, Building (structural), HVAC and others. Each of these models is characterized by domain specific objects that are owned by that domain, but some are of course shared (Structural engineers and architects both care about load bearing walls for instance). However the analysis of BIM models, as construction

Energy performance analysis can be used to exemplify the challenges that versatile standards pose in relation to the validation of models. For a energy performance simulation, certain information has to be present in the BIM (thermal values for each of the building components composing the exterior layer and also a quite large set of other set of data). In addition it is important for the localisation of this data through algorithms, that the information is stored in the BIM in a pre-agreed or even standardised way. This is hard to achieve in the fragmentised building industry in the of for instance the European Economic Area. Here countries have to implement the building energy directive, the exact way how to do this differs between the member countries [8].

In order to tackle this problem the Ifc consortium developed the concept of Information delivery manual (IDM) [9]. For each implementation of the energy directive a (slightly) different IDM should not only be developed, but also agreed upon and supported by the widely used software applications and specialized applications for energy analysis. IDMs have been developed, but as far as these authors know none of them are broadly adopted by BIM software. If a set of IDMs were agreed upon it would then be possible to identify whether and to which of the IDMs an open BIM was compliant with.
POTENTIALS

The creation of Building Information Models throughout the lifecycle of buildings results in an ever-growing resource of knowledge. The query of the models through project specific algorithms provides new perspectives on data and create project and workflow specific information. This information can be beneficial for a wide set of use cases from the validation of data, monitoring of processes and building portfolios, to the support of design decisions. This paves the way for novel forms of collaboration that capitalize on the precise description provide for architectural projects as the linkage of domains (Sheil 2014). The investigation of the current implementation of Ifc models in building practice, shows however that the created data is highly volatile in quality and content, even for very similar projects. It can be expected that this fact will prevail in the future due to the versatility of the IFC format. IFC however the only format that is information rich, collects the building domain on a global level and is vendor neutral and open. Rather than seeing the versatility of Ifc as a problem, that has to be counteracted with standards and mandatory fields, the profession should observe Ifc as a chance to engage in computational approaches to validate and create data specific to projects, rather than projects compliant to norms. The validity of this approach is demonstrated in this paper, though lacking in detail.

REFERENCES

Beetz, J., Berndt, R., Dietze, S. and Tamke, M. 2013 ‘Towards the Long-Term Preservation of Building Information Models’, Proceedings of the 30th International Conference on Applications of IT in the AEC Industry, Beijing, China


Lipman, R. 2010 'Developing Coverage Analysis for IFC Files', Proceedings of the CIB W78 2010: 27th International Conference


Tamke, M., Myrup, M. and Evers, H.L. 2014, Current State of 3d Object Processing in Architectural Research and Practice, DuraArk deliverable 7.7.1, EU 7th Framework Programme


[18] https://aectechtalk.wordpress.com/2012/03/19/project-lod-for-bim/