Analysis of the Evolving IFC Schema

Robert Amor, trebor@cs.auckland.ac.nz
The University of Auckland, New Zealand

Abstract
The newly released (March 2013) version of the IFC standard for open BIM (IFC4) is a major step forward for this ISO standard. It is reported to represent 8 person years of development effort and to have addressed over 1200 issues and change requests. It is released 6 years after the last major version of IFC (2x3 TC1) reflecting in part the major development effort required by the industry to handle new versions of this large standard (766 entities). While buildingSMART have documented the increased domain coverage over the previous version (e.g., for structural steel and timber, GIS, 5D support, etc) there is also an impact of changes in regards to the complexity and maintainability of the IFC standard. This paper investigates the changes within IFC4 and IFC4 Addendum 1 with respect to schema-level metrics, to understand the directions that the new standard has taken from a data management and implementers point of view. This complements work undertaken in 2007 (Amor et al. 2007) which looked at the impact of the then newly release IFC 2x3. In that study it was identified that there was unnecessary and growing complexity in the schema. This analysis addresses how much effort has been put into the new version to tame the complexity whilst growing its domain coverage.

Keywords: Schema metric, IFC, longitudinal study

1 Introduction

The development of the Industry Foundation Classes (IFC) as a means of achieving interoperability in the Architecture, Engineering, Construction and Facility Management (AEC-FM) domains has been ongoing since the 1990s. This parallels, and builds upon, the efforts of ISO-STEP (Standard for the Exchange of Product model data) (1994) community to achieve interoperability in engineering and manufacturing disciplines. In particular the EXPRESS schema specification language from ISO-STEP was chosen as the specification language for IFC as well as utilization of the STEP Physical File (SPF) format as the original method of encoding instance data to be transferred between applications. The international consortium of specifiers for the IFC, originally called the International Alliance for Interoperability (IAI), now rebadged as buildingSMART, define their top level goal as follows:

> *"The Industry Foundation Classes (IFC) represent an open specification for Building Information Modeling (BIM) data that is exchanged and shared among the various participants in a building construction or facility management project..."* IFC (2015)

The first stable versions of the IFC schema, which were taken up by industry and implemented in a range of BIM and simulation tools, were published in the late 1990s and then new versions with additional coverage emerged every 1-2 years after that. Figure 1 provides an overview of the release dates of the major versions of the IFC schema since 1997. As can be seen in Figure 1 the largest time between versions of the schema being released came with IFC 4, released in 2013, a period of six years after the release of IFC 2x3 TC1. This change in release periods is one indication of the impact that changes in the schema can have on the AEC-FM industry and the software tool developers which support the industry. Software vendors have indicated that the cost to implement new versions of the IFC schema (Autodesk has previously indicated that this equates to around one person year per schema release) was becoming a major burden for them. So yearly releases of new versions of the IFC schema was impacting many developers, alongside the cost of maintaining translators for all of the previous major versions of the IFC schema potentially being used in various software tools that project partners would bring to a construction project.

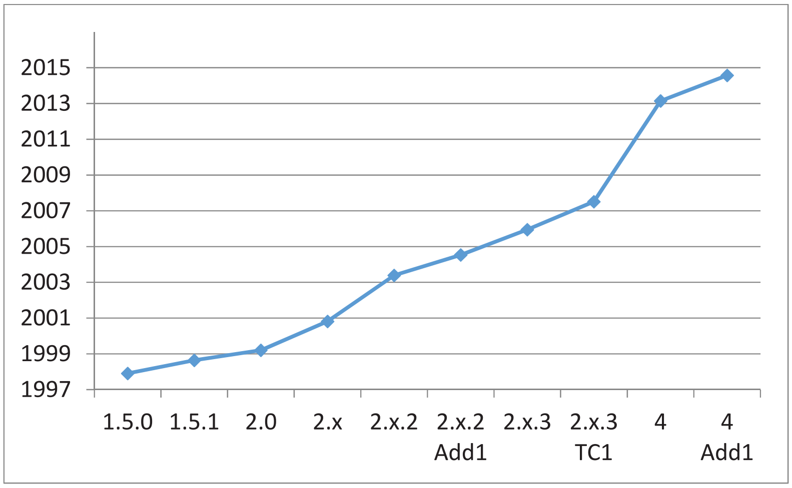

Figure 1 Time between releases of IFC schema versions

This balance between the continual development of a schema to address additional aspects of an industry's information transfer needs versus the implications of these changes is of interest in this paper. While the frequency of release of versions can be readily associated with an implementation cost for developers there are other impacts that are unlikely to be as readily identified from changes in the schema. As the schema increases in size there are questions about the effect on its comprehensibility for developers and the time required to become conversant and expert with all of the structures available within the evolving schema. This has a direct consequence on the ability to develop correct implementations of each version of the schema by each software developer, a necessary requirement if the expected level of industry interoperability is to be achieved. The software industry is well aware of the impossibility of guaranteeing correct software code on anything apart from trivial pieces of software, and hence the testing burden on software developments increases as the size and complexity of the specification increases. This is particularly true when the specification of requirements for the schema is in English language text (as is the case for the IFC schema), rather than a computer interpretable formal notation. Then for both developers and those utilizing the IFC schema as a transfer mechanism on construction projects there is the issue of the efficiency of the representation. The choices of representation in a schema can have a large impact on the size of the data files and hence the ability to ensure fast transfers of information during a project.

A number of these measures are affected by the very early choices made by the IAI in regards to the overall architecture of the IFC schema. As shown in Figure 2 the IFC schema comprises a number of layers which allows for a separation of concerns (a good software development practice) between core concepts, shared information resources, and then the domain specific representations for constituents of the AEC-FM industry.

1.1 Sources of IFC Schema Data and Metrics

The data presented in this paper is based on the IFC schemas published by buildingSMART (and previously the IAI) on their website for IFC versions 2x onwards (IFC 2015). Earlier versions of the IFC schema are from the author's repository of schemas systematically collected from the first official IFC publications. Where possible, summary data on the constituent components of the IFC is drawn from the online documentation of the various schema releases which is provided by buildingSMART (IFC 2015). To complement this data, and to validate the data published by buildingSMART, a number of scripts have been written to produce summary counts of the constituent parts of the IFC schema. The more sophisticated analysis of metrics of the IFC schema is generated by an updated version of the EvaSys system, previously described in Ma et al. (2006), and now renamed ifcEval. As noted in Amor et al. (2007) there is a large choice of metric that can apply to the IFC schema. While Piattini et al. (2002) has described many specifically for data models this work focusses on those developed more specifically for UML class diagrams as summarised by Genero et al. (2000).

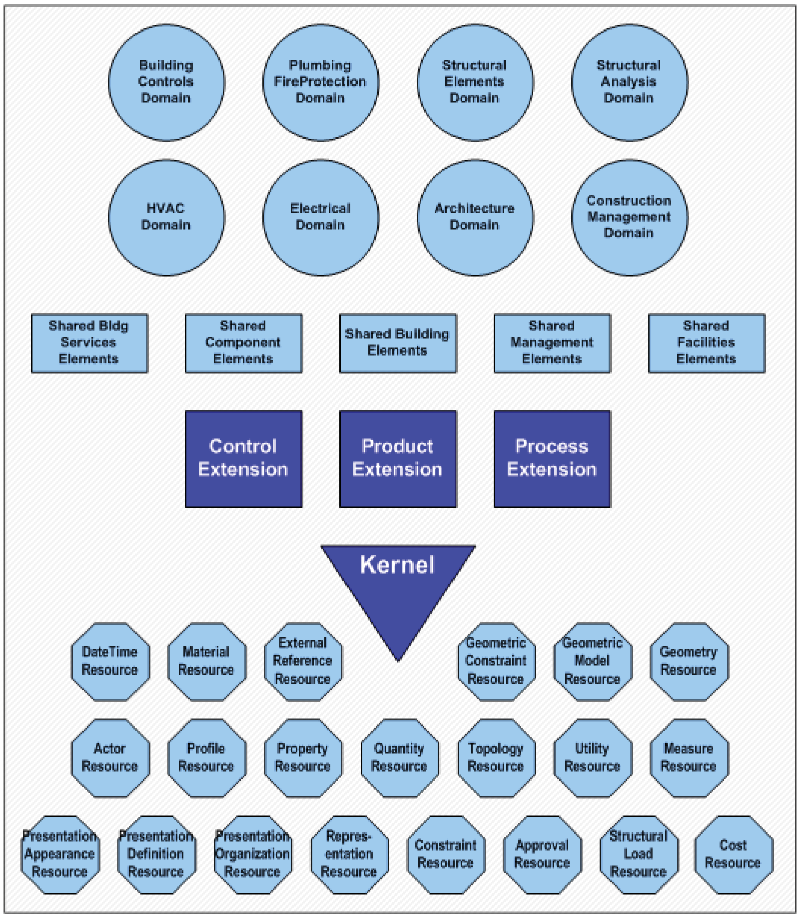

Figure 2 Layered structure of the IFC schema (IFC 2015)

2 Trends in the IFC Schema Releases

The EXPRESS longform of each IFC schema version as listed in Figure 1 has been passed through the same analysis scripts and tools to generate comparable data on the changes since 1997 as the schema has evolved. The data is classed into two general categories, those metrics which are counts of bits inside the schema as covered in Section 2.1 and those metrics which are based on measures of the graphstructure of the schema as covered in Section 2.2. Section 2.3 provides ancillary analysis of the difference between one of the MVD specifications and the complete IFC schema.

2.1 Schema Growth

The simplest sets of metrics gathered on the evolving IFC schema comprise counts of the various structures within the schema as it has evolved over time. These counts focus on the entities in the schema, attributes, types, functions, and rules as well as the Property Sets defined since IFC 2.x.2.

2.1.1 Entities

The total number of entities in the IFC schema is indicative of the growing scope of the schema and especially in later years, as the kernel layers have seen less development, they represent the structures needed to represent new concepts in the domains being added to the schema. However, the growth in number of entities also represents a bulk measure of the global complexity of the schema and the effort that would be required to understand the totality of the available schema.

As can be seen in Figure 3 there was a remarkable increase in the schema size from version 2.x to 2.x.2 and then another significant increase from 2.x.3 TC1 through to 4. These increases align with major coverage changes in the IFC schema as new disciplines are addressed. Figures 3 and 5 show that the number of instantiable entities (entities which represent real world things that can have data associated with them) have increased at a similar pace, and the ratio shown in Figure 5 has remained pretty constant across all versions of IFC.

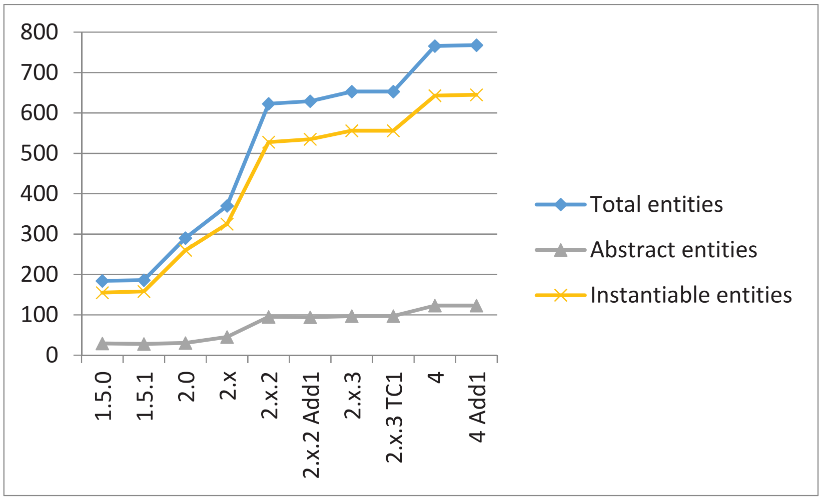

Figure 3 Instantiable and abstract entities

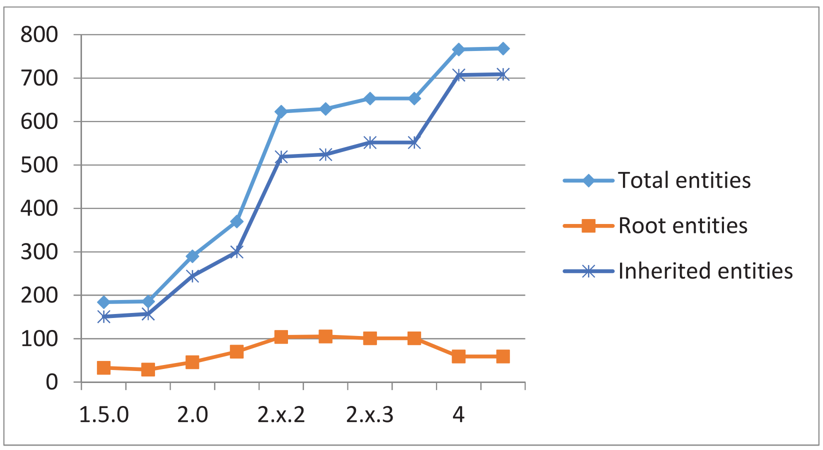

Figure 4 Root and inherited entities

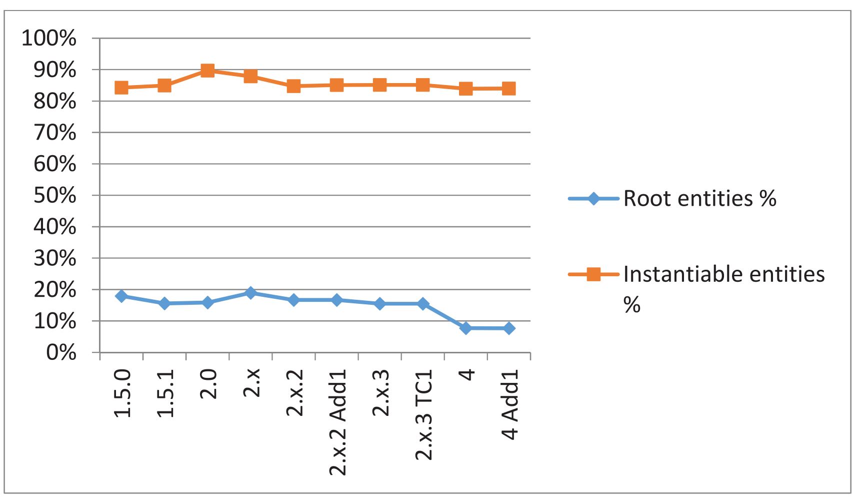

Figure 5 Percentage of root and instantiable entities

The number of root entities (those which do not inherit from any other entity), as shown in Figure 4 and as a percentage in Figure 5, also followed a similar trend of increases up until IFC 4 when it dropped significantly. This would be indicative of attempts to simplify the structure of the IFC schema, reducing the number of inheritance paths means that there are fewer root structures which have to be understood and coded by implementers of the schema. It also segments the schema into a smaller number of sets (helping comprehensibility), though increasing the number of entities which would exist in each set.

2.1.2 Attributes

As would be expected, the number of attributes in the IFC schema has grown as the number of entities has grown. Figures 6 and 7 document this increase in attributes over time, though as can be seen in Figure 7 the ratio of attributes per entity has fluctuated over time and is following a slight downwards trend. The number of attributes per entity is remarkably low for a schema reflecting a physical domain such as AEC-FM which would seem to indicate that entities in the IFC schema are being used as a way of classifying and segmenting the instances which are created for a building more than capturing the properties of these instances. In Figure 11, which documents the number of unique properties as being greater than the number of attributes, it becomes obvious that more of the attribute information for the IFC resides in the Property Sets than in the attributes assigned to entities.

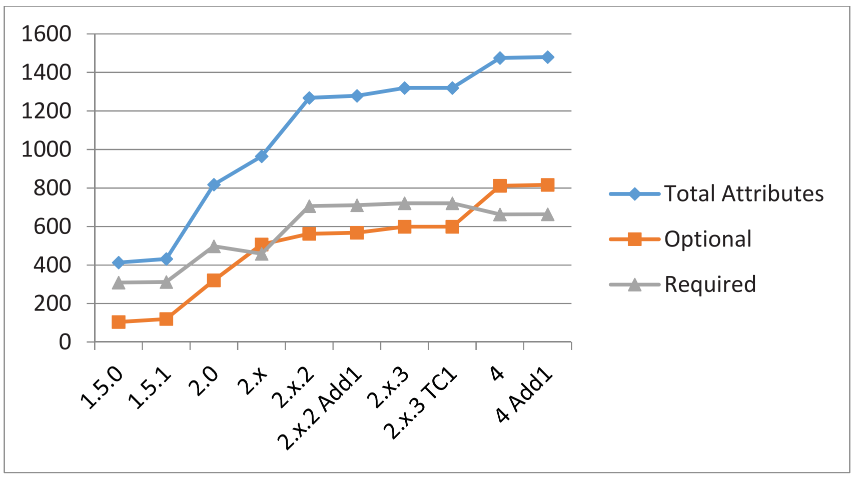

Figure 6 Optional and required attributes

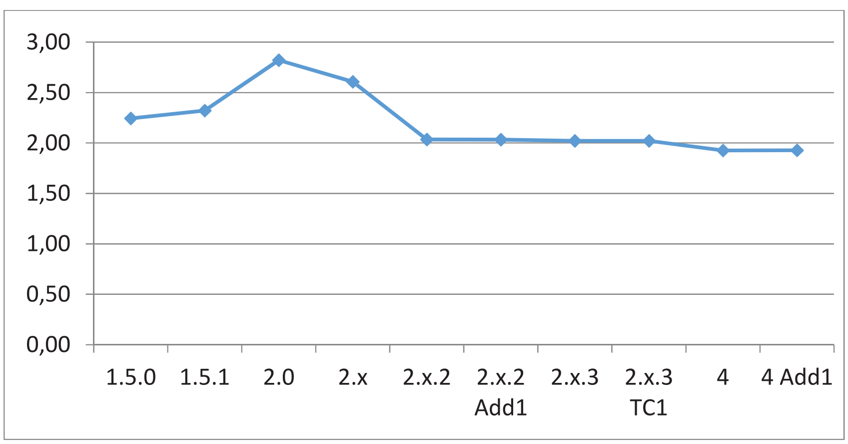

Figure 7 Average number of attributes per entity

There are two further implications of this approach in IFC. One is that the majority of international agreements around the information needed for an entity is not being recorded in the EXPRESS schema used to represent the IFC and as published by ISO, but are in side agreements documented on the buildingSMART website. The second implication is in the performance and efficiency of the SPF used to transport data about a building. There will be added complexity and size in the SPF as attributes of an entity instance are not recorded directly in the entity but have to be associated to it through the Property Set mechanism.

One further change evident in the attributes described in the IFC schema is the increase in optional attributes over time as seen in Figures 6 and 8. Of particular interest is the large change in IFC 4 where the percentage of attributes which are optional has increased significantly. This is indicative of an attempt to make the IFC schema more flexible for implementers and users. If fewer attributes of an entity are required to be filled before the instance is valid then an instance can be created earlier in the AEC-FM process when less information is known about it, or can be used under a greater number of conditions with less data transferred. The tradeoff here is over how much data is needed to make an instance valid for a particular purpose (e.g., thermal simulation). However, with the MVD structures being developed on top of the core IFC schema to specify exact requirements for particular domains and processes these greater requirements can, and should, be specified at this level.

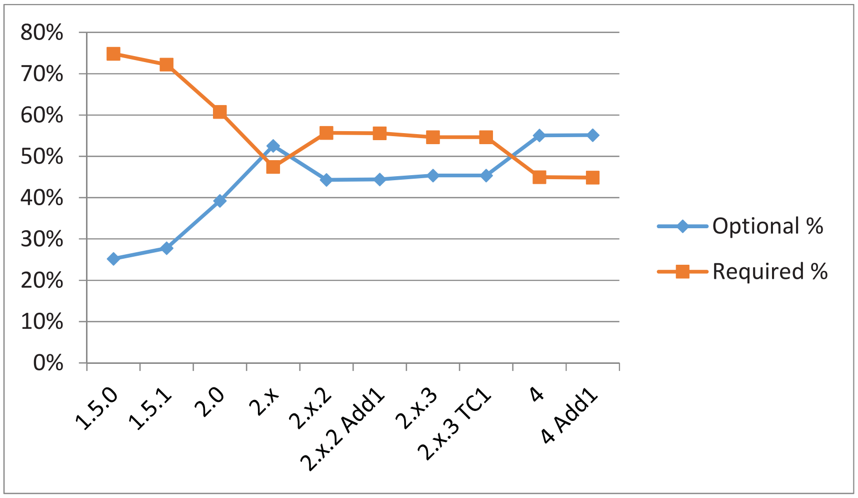

Figure 8 Percentage of optional and required attributes

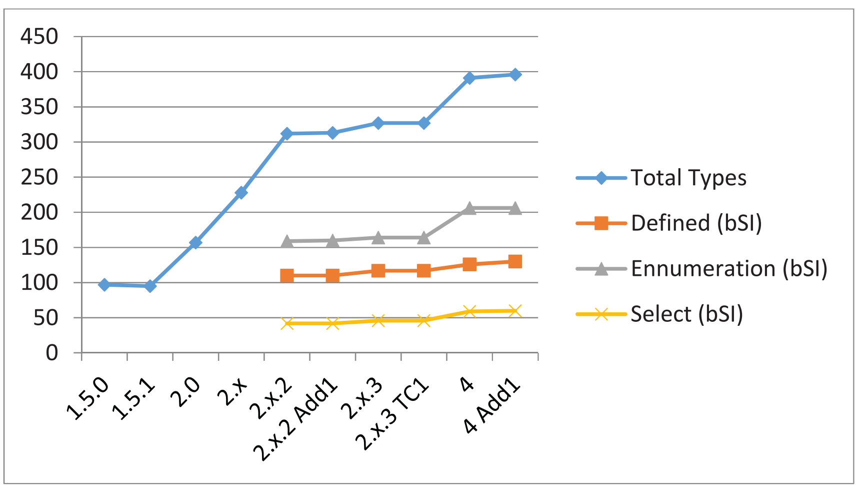

Figure 9 Types used in the IFC schema

2.1.3 Types

Figure 9 shows the growing number of types specified in the IFC schema (and a breakdown of defined, enumeration and select types as provided by buildingSMART) with growth mirroring the growth in entities from IFC 2.x onwards. This growth is unsurprising for this period as the IFC schema is covering more domains in AEC-FM and it would be expected that new types are required to reflect information requirements in these new domains. What does seem a little unusual is the quantity of types in relation to the number of attributes. With a ratio of around 4:1 for attributes to types it is clear that a number of these types are only specified for one or two attributes in the schema. However, as the type structures provide some semantics about the expectations of allowable data for an attribute it would seem worthwhile continuing to capture this within the schema.

2.1.4 Semantic codification

EXPRESS provides for the specification of some semantics to be attached to entities with functions and rules as well as with WHERE clauses inside entities. Figure 10 documents that this is a seldom used feature in the IFC schema (though over half of the entities do have WHERE clauses) and is not growing significantly over time. Given that there is significant semantics about the entities in each schema written up in the online documentation there could be greater effort put into the computer interpretable specification of this semantics. With automated translators of the EXPRESS semantic information into other programming languages this would give a way of automatically ensuring matching functionality in a wide range of software tools.

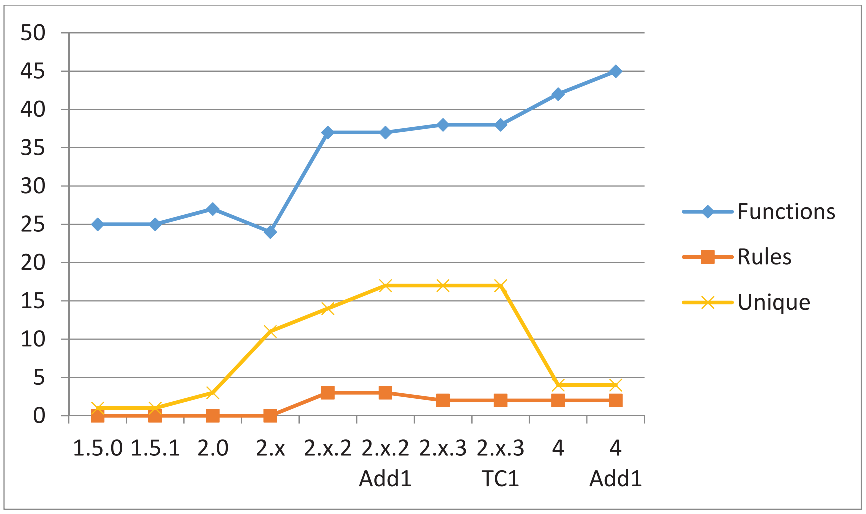

Figure 10 Semantic structures in IFC schema

Figure 10 also documents the use of the UNIQUE keyword on attributes in the schema. This gives a way of specifying that there can be no duplicates of the value of an attribute in a model. For example, the *GlobalId* in *IfcRoot* which is a very strict requirement as indicated below:

> *"All entities having semantic significance derive from IfcRoot, where instances are identifiable within a data set using a compressed globally unique identifier (IFC-GUID). This identifier must never change during the lifetime of an object, which allows data to be merged, versioned, or referenced from other locations." IFC (2015)*

It is unsurprising that the number of unique attributes increases over time, but what is surprising is the effort that has obviously been put into IFC 4 to reduce these unique attributes. The benefit of the reduction will be in consistency checking and maintenance of data inside a software tool, where less effort needs to go towards ensuring the unique property holds at all times. The potential loss is that items which should be unique are not enforced in that way. While IFC 4 maintains unique attributes for the entities *IfcApplication*, *IfcPropertyEnumeration*, and *IfcRoot* it discards previous unique constraints in entities such as *IfcActionRequest*, *IfcCostSchedule*, *IfcOrderAction*, *IfcPermit*, and *IfcProjectOrder* where it could be considered that there does need to be a unique identifier for information of this type. For example, where *IfcPermit* used to have an enforced unique *PermitId* there is now only an optional *Identification* provided through *IfcControl*.

2.1.5 Property Sets

The use of Property Sets has been a dominant feature of IFC schema since 2.x.2 with an increasing number of Property Sets being specified in each new version, though as seen in Figure 11 this is not mirrored in a consequent growth of unique properties across the Property Sets. Section 2.1.2 contrasts the ratio of Property Sets to attributes in regards to where information about an instance resides and the implications this can have on lowered efficiency of representation and increased processing time.

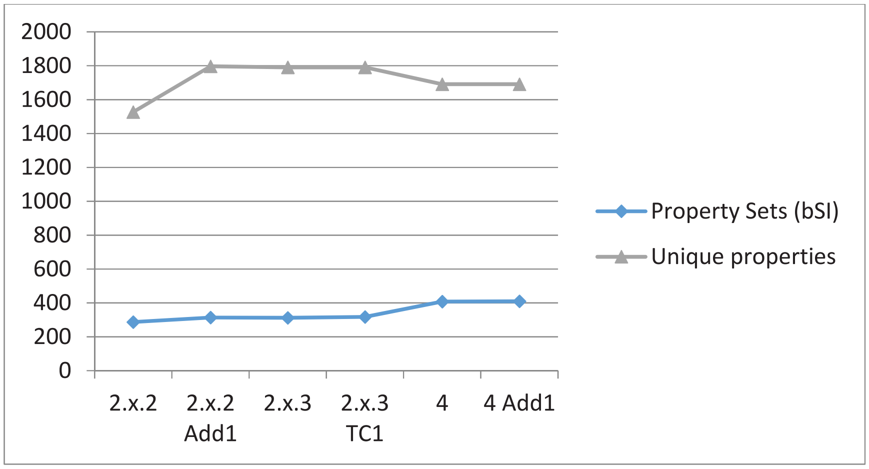

Figure 11 Property sets in the IFC schema

2.2 Schema Metrics

There are a large number of metrics that can be calculated on the graph representation of a schema. Each of these metrics providing some evidence of complexity and connectivity in the graph. In Figure 12 the average and maximum depth of the inheritance tree is documented. While the maximum depth has not changed over many versions of IFC it is clear that the average depth of inheritance structures is increasing. This is unsurprising as given the layered structure of IFC (as shown in Figure 2) and the fact that new versions of IFC are adding entities for domain specific coverage then most of these entities will be appearing towards the leaf end of the schema. The growing impact of this metric is that there is more dependence upon root entities in the schema and any changes near the top of the inheritance hierarchy is going to have a proportionally larger impact for developers (who have to propagate and verify that change to a greater number of entities) than would be the case if the schema had a flatter structure.

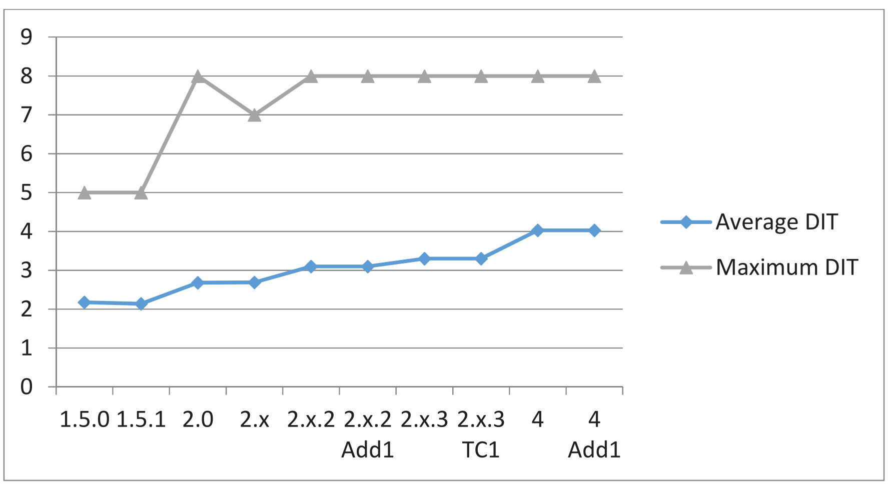

Figure 12 Changing average Depth of the Inheritance Tree (DIT)

2.3 MVD Impacts

While the analysis presented in Sections 2.1 and 2.2 looks at the IFC schema as a whole it is certain that no one should ever use or implement the whole of the IFC schema for a particular interoperability need. buildingSMART provide the Model View Definition (MVD) approach to identifying a subset of the IFC schema for any particular transfer, see below:

> *"A subset of the data schema and referenced data is referred to as a model view definition. A particular model view definition is defined to support one or many recognized workflows in the building construction and facility management industry sector. Each workflow identifies*

data exchange requirements for software applications. Conforming software applications need to identity the model view definition they conform to." IFC (2015)

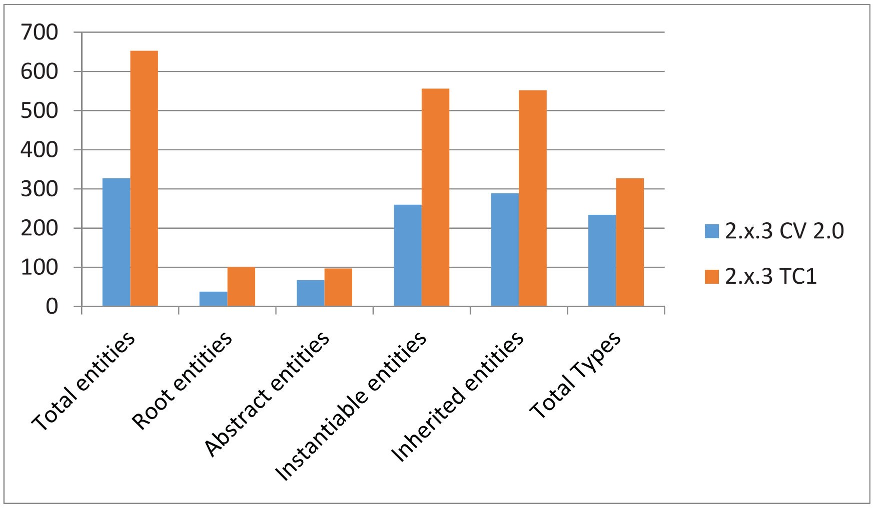

Figure 13 Entity comparison between IFC schema and the CV MVD

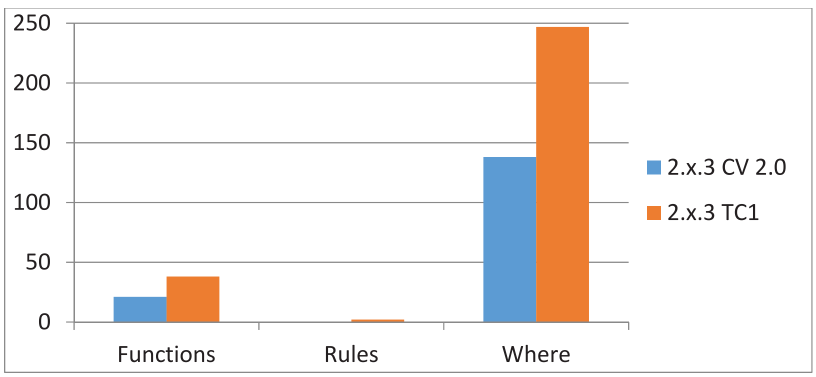

Figure 14 Semantics comparison between IFC schema and the CV MVD

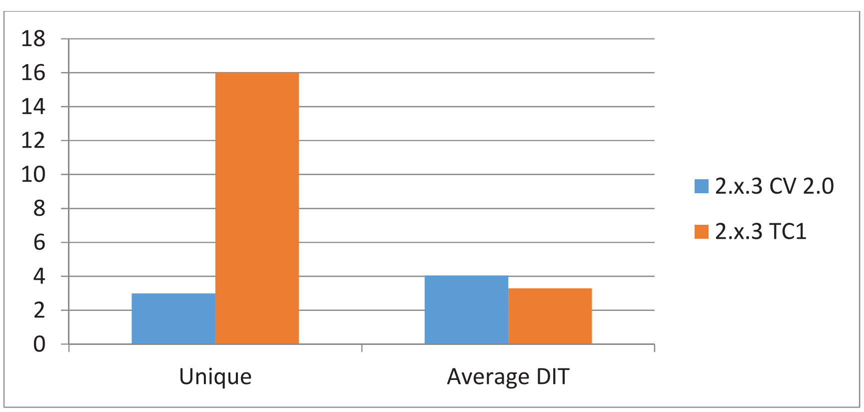

Figure 15 Metrics comparison between IFC schema and the CV MVD

While MVD are not typically published in EXPRESS longform we have performed an evaluation between the IFC 2.x.3 Coordination View version 2.0 (CV 2.0) and IFC 2.x.3 TC1. Figures 13 and 14 identify that CV 2.0 is approximately half the size of the full IFC schema, so provides a significant reduction in implementation effort required for a software company to develop and test. Intriguingly in Figure 15 one can see that CV 2.0 has very few of the UNIQUE constraints in 2.x.3 TC1, reducing the implementation requirements for this MVD but a little surprising the DIT of CV

2.0 is greater than for the base schema, indicating that more entities relating to domain requirements are incorporated into this MVD.

3 Conclusions

The IFC schema has grown significantly since the 1990s with another big jump in coverage introduced with IFC 4 in 2013. The analysis of the evolving IFC schema shows a fairly typical increase in size across the board of measured metrics for almost all versions. What is evident however is that in IFC 4 a significant effort was made to reduce some aspects of complexity in the schema. This is seen in the: decrease in root entities; increase in optional attributes; and decrease in unique attributes. These are all significant changes to the structure of the IFC schema which will reduce the development and checking effort required by software vendors looking to implement the latest version of the IFC schema. There are however semantic implications from these changes which have not been fully described by buildingSMART (e.g., why attributes that needed to be unique for the last decade are now optional).

While the changes in IFC 4 are likely to be welcome by software developers this analysis makes it clear that most of the metrics which describe increasing complexity, the impact on the ability to ensure correctness, and the impact on efficiency of the transferred data are increasing. Though there are no easy approaches to restructuring a schema to reduce these metrics it is clear that buildingSMART will consider restructures to reduce complexity as evidenced in IFC 4. It is also clear that very little of the semantics of the IFC schema (as published for readers to interpret on the website) is captured in computer interpretable specifications such as functions, rules and WHERE clauses. Increasing this specification of semantics would help ensure that all implementers of IFC code to the same standard.

Acknowledgements

Generous help in checking the reported metrics and rectifying calculation anomalies was provided by Johannes Dimyadi (University of Auckland) and Thomas Krijnen (Technical University of Eindhoven).

References

Amor, R., Jiang, Y. & Chen, X. (2007) BIM in 2007 - are we there yet?, *Proc. of CIB W78 conference on Bringing ITC knowledge to work*, Maribor, Slovenia, 26-29 June, pp. 159-162.

Genero, M., Oiattini, M. & Calero, C. (2000). Early Measures for UML class diagrams, *L'Objet*, 6 (4), pp. 1-28.

IFC (2015) Industry Foundation Classes, buildingSMART International, web site last accessed 29/5/2015. http://www.buildingsmart-tech.org/specifications

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

ISO-STEP (1994) ISO 10303-1:1994 Industrial automation systems and integration -- Product data representation and exchange -- Part 1: Overview and fundamental principles, ISO Standard.

Ma, H., Ha, K.M.E., Chung, C.K.J. & Amor, R. (2006). Testing Semantic Interoperability, *Proc. of the Joint International Conference on Computing and Decision Making in Civil and Building Engineering*. Montreal, Canada, 14-16 June.

Piattini, M., Genero, M. and Calero, C. (2002). Data Model Metrics, *Handbook of Software Engineering and Knowledge Engineering*, 2, ISBN: 981-02-4974-8.