Lessons Learned from Designing and Using bcfOWL

Oliver Schulz, Jyrki Oraskari and Jakob Beetz

1Department of Design Computation, RWTH Aachen University, Aachen, Germany
2Department of Individualized Production, RWTH Aachen University, Aachen, Germany

Abstract
The bcfOWL ontology has been developed as part of the EU Horizon 2020 BIM4Ren project to enable communication between BIM Collaboration Format (BCF) Issues and Linked Building Data (LBD) concepts described on the Semantic Web. This paper evaluates the current approach in bcfOWL based on its use in the BIM4Ren project. The ontology serves as an interlanguage for component-based communication, providing a gateway to the systems of different domains. During its use in the project, new insights into the usability of the ontology were gained. We discuss these findings and provide suggestions to complement the original design principles of bcfOWL, targeting the LBD domain. Our work should guide future research in component-based communication in building-related projects and provide helpful considerations for future ontology designs. We also discuss potential areas of improvement for bcfOWL, including versioning, Ontology Design Patterns and validation. Overall, bcfOWL aims to improve querying capabilities and connectivity with Linked Building Data, making it a valuable tool for building-related projects.

Keywords
BCF, Issue Management, Linked Data, Linked Building Data, Ontology Design

1. Introduction
The bcfOWL ontology [1] was developed as part of the EU Horizon 2020 BIM4Ren project. The BIM4Ren project aimed to harness the potential of BIM to enable energy-efficient renovation of existing buildings across the construction value chain. The development of bcfOWL was explicitly aimed at incorporating the localisation of site images and the spatialisation of tasks created by a project manager and completed by an assigned person in the field. bcfOWL is based on buildingSMART’s BIM Collaboration Format (BCF) [2], an open standard for sharing and communicating issues related to building elements. The BCF standard was developed to facilitate the exchange of issue descriptions among various stakeholders in the context of the building information modelling methodology. The standard employs concepts for Topics, Comments, and Viewpoints to describe Issues and is available in two formats: file-based (BCF XML) and server-based (BCF API). While both formats share most concepts, buildingSMART
has not yet developed a common schema for BCF.

bcfOWL aims to enable access between issue communication-related information and Linked Building Data (LBD) concepts described on the Semantic Web. For instance, bcfOWL can connect BCF issues with Linked Building Data elements and make it searchable using open standards. Moreover, the ontology can be a shared schema to ensure compatibility with and conversion to BCF XML and the BCF API.

In this paper, we want to evaluate the current approach in bcfOWL based on its usage throughout the BIM4Ren project. Original design principles of bcfOWL were introduced in [1] and are still valid. The suggestions in this paper complement these and target the Linked Building Data domain. Therefore, this work should serve as a guide for future research in component-based communication in building-related projects. On the other hand, it shall provide helpful considerations for future ontology designs.

The paper is organised as follows: We begin by introducing the current state of bcfOWL and describe its design and usage throughout the BIM4Ren project. Next, we identify potential areas for improvement that became apparent during our work with bcfOWL and discuss them at the end of each subsection. Finally, we outline a path forward for the ontology and conclude the paper in the last section.

The namespaces in Listing 1 are referred to throughout the paper.

<table>
<thead>
<tr>
<th>prefix</th>
<th>URI</th>
</tr>
</thead>
<tbody>
<tr>
<td>bcfOWL</td>
<td><a href="http://lbd.arch.rwth-aachen.de/bcfOWL#">http://lbd.arch.rwth-aachen.de/bcfOWL#</a></td>
</tr>
<tr>
<td>bot</td>
<td><a href="https://w3id.org/bot#">https://w3id.org/bot#</a></td>
</tr>
<tr>
<td>cto</td>
<td><a href="https://w3id.org/cto#">https://w3id.org/cto#</a></td>
</tr>
<tr>
<td>ifcOWL</td>
<td><a href="https://standards.buildingsmart.org/IFC/DEV/IFC4/ADD2_TC1/OWL#">https://standards.buildingsmart.org/IFC/DEV/IFC4/ADD2_TC1/OWL#</a></td>
</tr>
<tr>
<td>prov-o</td>
<td><a href="http://www.w3.org/ns/prov#">http://www.w3.org/ns/prov#</a></td>
</tr>
<tr>
<td>dot</td>
<td><a href="https://w3id.org/cto#">https://w3id.org/cto#</a></td>
</tr>
<tr>
<td>opm</td>
<td><a href="https://w3id.org/opm#">https://w3id.org/opm#</a></td>
</tr>
<tr>
<td>omg</td>
<td><a href="http://w3id.org/omg#">http://w3id.org/omg#</a></td>
</tr>
</tbody>
</table>

Listing 1: Ontologies with their prefixes used throughout the paper

## 2. Using bcfOWL in BIM4Ren

The need for an ontology for BCF arose from work on the BIM4Ren project, in which images were spatially located based on plans and models by using BCF [3]. Even though the overall functionality worked with the BCF API, it lacked flexible query abilities. The API had to be extended with new API routes to cope with this drawback. Moreover, in [4], the BCF API was extended with a concept of *Originating Documents* for Viewpoints and *Spatial Representations* for Documents in order to make the camera view locations queryable over different representations of a building. Furthermore, a geometry concept was added to be able to save damage markups within the Viewpoints. However, BCF data had, and still has, very limited possibilities to connect to Linked Data concepts in the LBD domain, such as the Damage Topology Ontology (DOT) [5] and the Building Topology Ontology (BOT) [6].

While these proposed extensions partially addressed the drawbacks, adding to the BCF for each missing condition proved disadvantageous. Therefore, a transfer to an ontology seemed desirable, as it could ensure more flexible query ability and connectivity to other concepts in the LBD domain. Furthermore, the creation of the ontology does not exclude the further
exploitation of the original BCF formats. Still, it merely extends them with a common data format that can be compatible with both the XML and the API versions.

bcfOWL was first presented in [1] as a hand-crafted ontology to express the BCF API and BCF XML formats in the W3C Web Ontology Language (OWL). No new concepts were introduced or changed to ensure consistency between the different BCF representations. The ontology was used throughout the BIM4Ren project for various functions. Initially, it was used to locate images spatially. Then, to express tasks performed in a specific order by a particular person (see Figure 1) on the construction site. To achieve the latter, bcfOWL was integrated with the Construction Task Ontology² (CTO), as BCF cannot specify the order of task execution or provide shared task specifications. An example of this integration is shown in the Listing 2.

For communicating the BCF content with the different clients in the BIM4Ren Project, we provided a middleware that allowed us to query the content either by using the BCF API or SPARQL. The content itself was stored on an Apache Jena Fuseki server⁴ (see Figure 2).

---

Figure 1: Spatially located images and tasks in the BIM4Ren project. The BCF Viewpoints are located on the 2D floor plan, which is superimposed with the model.

---

inst:Intervention_Technical_pre_work_inspection

```
inst:Intervention_Technical_pre_work_inspection
    a bcfOWL:Topic, cto:Task;
    bcfOWL:hasAssignedTo
        inst:Oliver_Schulz;
    bcfOWL:hasCreator
        inst:Oliver_Schulz;
    bcfOWL:hasCreationDate
        "2022-05-24T00:00:00Z"^^xsd:dateTime;
```


Technical pre-work inspection. The quality of the facade needs to be checked, some measure be performed, and the accessibility needs to be ensures.

"2022-05-25T00:00:00Z"^^xsd:dateTime

"Technical pre-work inspection"

Listing 2: Example from connecting bcfOWL with CTO

In addition, the use of bcfOWL in container-based environments such as the Information Container for Linked Document Delivery (ICDD) [7] and the Linked Data Platform (LDP) [8] - in its manifestation of Solid [9], and ConSolid [10] - was investigated during BIM4Ren. Since bcfOWL was missing the ability to point to containers, the ontology was extended with the concepts of bcfOWL:TopicsContainer, bcfOWL:ViewpointsContainer, and bcfOWL:CommentsContainer. These can be used to point to containers that store BCF content about Topics, Viewpoints, and Comments. It is not necessary that the container content is serialised in RDF, but they could also be used to refer to, for example, JSON or XML content. This extension went hand in hand with testing federated Common Data Environments (CDEs), which was investigated in [11] using the CDE subset of BCF as an example.

3. Potential Areas of Improvements

In this section, we discuss the different potential areas for improvement for bcfOWL regarding future developments in the Linked Building Data domain.

3.1. Versioning

As Issue Management is a process of exchanging information, the details of issues can change frequently. For example, the status of an issue may change from active when it is created to closed after it has been resolved. Although the BCF API can track these changes using an event system that allows queries to see what has changed, bcfOWL initially focused on the core aspects of BCF and did not consider the event system from the beginning.

Versioning for bcfOWL was first investigated in [12]. Different approaches were considered for describing an issue’s different versions in a typical Issue Management process. The core idea underlying the various approaches was to avoid deleting information from the graph and instead
add new information to it. As a result, the graph retains a comprehensive record of all past events, ensuring a complete history is preserved. The following approaches were considered as most promising to go forward:

1. A State’s approach, inspired by the LBD ontologies the Ontology for Property Management (OPM) [13] and The Ontology for Managing Geometry (OMG) [14]. These States can be used to externalise the information part of a subject to another triple that links to the original subject. Thereby newly created States never delete the old States and can be filtered by, for example, the parameter time.

2. A Statement approach for annotating triples using triples. The approach was first envisioned in [15] and culminated in developing RDF-star ⁵ which was used for this work.

As of now, a decision on what versioning system should be used is not yet made. The State’s approach is very much in line with current approaches in the LBD Community Group and is straightforward to implement with current Linked Data technologies. Statements with RDF-star showed a performance comparable to the State approach but still have open questions, for example, on how to express it in an ontology and how to use it for reasoning. In contrast, the statements approach could be used to only slightly modify the current structure of bcfOWL and add annotations to the triples. The States approach, on the other hand, would mean a significant restructuring of bcfOWL.

While this is not necessary for all processes, versioning should be considered from the beginning when creating new ontologies. For describing version content, PROV-O [16] is considered a starting point.

---

⁵W3C, RDF-star and SPARQL-star https://w3c.github.io/rdf-star/cg-spec/editors_draft.html accessed 27.02.2023
3.1.1. Signing

We considered the possibility of a user signing the changes on the graph when creating a new problem description or adding descriptions of how the problem was handled. The idea is to use a public key (PK) infrastructure so that only the user can create the signature, but anyone can check who has signed the part of the graph. The signatures can serve two purposes: to indicate ownership of the content, but more importantly, to provide non-repudiation of the claims, e.g. in a situation where someone declares the issue fixed. While such a safeguard is not necessary for all ontologies, it is a very important factor, especially in the construction industry - where many projects end up in court. Data protection should be ensured, particularly in areas where processes, responsibilities, tasks and the completion of tasks are recorded.

Carroll, Sayers et al., and Hogan have proposed algorithms to create a canonical serialisation of an RDF graph in [17], [18] and [19], respectively. The algorithms allow for checksum calculation for sub-graphs and, thus, when encrypted with the user’s private key, can be used to sign the content changes.

3.2. Validation

As seen in Figure 2, the client in the BIM4Ren workflow is able to communicate changes and additions in the BCF context via the middleware with the BCF API and SPARQL. While we identified the static structure of the API as a disadvantage in specific scenarios, its rigidity can also be an advantage, for example, when POST and PUT operations are performed. The precise structure of the API prevents incorrectly formed BCF content from being sent to the server. In these cases, the API gives direct feedback to the client and makes it aware of this via error messages. In the case of SPARQL Updates, however, this is not necessarily the case. Although a request is also returned as incorrect if it does not comply with the SPARQL syntax, it is not validated for completeness or content. While OWL allows us to reason, it is not necessarily helpful in validating the data content because of its open-world assumption [20]. The Shapes Constraint Language (SHACL) [21] could be a helpful addition to BCF on the semantic web by validating it with SHACL shapes before sending the content to the graph. An example shape for bcfOWL is provided in Listing 3.

```xml
bcfOWL:TopicShape
  a sh:NodeShape ;
  sh:targetClass bcfOWL:Topic ;
  sh:property [ sh:path bcfOWL:hasTitle ;
    sh:datatype xsd:string ;
    sh:minCount 1 ;
  ] ;
  [...],
  sh:property [ sh:path bcf:description ;
    sh:datatype xsd:string ;
  ] .
```

Listing 3: Example of how a BCF Shape could look like
However, a potential conflict may arise if versioning via RDF-star is used as described in Section 3.1, since SHACL cannot yet work with RDF-star. However, there are first discussions within the SHACL community to incorporate RDF-star. A potential workaround to this problem in the meantime could be to enable validation via regular SPARQL queries – as described in [20] – but this would only be a temporary fallback until the SHACL community resolves this matter.

While the open-world assumption is integral to the Linked Data philosophy, it is not always helpful when it comes to tightly defined schemas. This can be remedied by using SHACL to validate the data for conformance using the shapes even before the content is stored in the graph. It should be noted that OWL and SHACL are not mutually exclusive and can be used complementary to each other. Therefore, if an ontology maps to a fixed schema, the corresponding SHACL shapes should be created in addition to the description in OWL. However, it should be further investigated whether checking incoming issues that first have to be checked by means of SHACL has a negative effect on the performance and responsiveness of the server communication.

3.3. Generic Framework

bcfOWL is a translation of a specific Issue Management format to an ontology. Thereby, its concepts - although theoretically applicable in many different scenarios - are always tightly connected to BCF. This approach is comparable to translating the IFC EXPRESS Schema to ifcOWL [22]. The different components of BCF, the Topics, Viewpoints and Comments, are generally not unique to BCF, but on the opposite, very generic concepts that could be used for scenarios outside Issue Management. As an example:

The concept of the BCF Viewpoint is used throughout the BIM4Ren project to spatially locate images taken on-site in the digital building model, as seen in [3].

BCF is the only standardised and widely supported format that is currently allowing to achieve this in a software-independent manner. But in order to use this concept, the Viewpoint is always accompanied by the BCF concepts of Topics and Comments. Furthermore, the Viewpoint lacks vital information to be usable on its own. For example, it does not come with an author or a date. This information has to be saved to the Topic. However, this only partially solves the problem since a Topic can have multiple Viewpoints simultaneously. Therefore, from an ontological perspective, using the Viewpoint from BCF to describe camera views of real pictures on-site without intending to remain in the BCF context makes little sense.

Although bcfOWL and BCF use the general concepts, they do not describe them sufficiently to be used meaningfully in other scenarios. Therefore, more generic frameworks are needed to describe these concepts that can then be used to create BCF information.

A solution to this problem could be to base bcfOWL on the principles of Ontology Design Patterns (ODP) [23]. These principles require keeping ontologies small and modular since large ontologies tailored to specific schemas are usually challenging to reuse for other purposes. Suitable examples in the LBD community built on these principles are the Building Topology GitHub, Modifications to SHACL to incorporate RDF-star: https://github.com/w3c/rdf-star/issues/122 accessed: 24.02.2023

GitHub, RDF-star and SPARQL-star support for SHACL: https://github.com/w3c/shacl/issues/15 accessed: 24.02.2023
Figure 3: The main concepts of BCF are the Topic, the Comment and the Viewpoint with its many sub-components. The information in these concepts can be classified into two categories: component-based communication and spatial


Although many ontologies are created to bring an existing schema into the Semantic Web, it should be checked beforehand whether they can be built on top of existing higher-level ontologies. If these do not exist, these ontologies could be created first, potentially keeping the ODP principles in mind.

The core components of BCF and bcfOWL can be divided into two main concepts: component-based communication information and spatial information. Even though these components cannot always be clearly assigned to only one of the concepts, we will show the most important overlaps in the following two subsections.

3.3.1. Component-Based Communication Information

The concepts of Topics and Comments in BCF mainly contain semantic information for carrying messages and providing a context for these messages. An example structure for BCF can be seen in Figure 3. Most of the properties in these concepts - like the title, status, assignment etc. - can be used for communication in general. For example, the ISO 10303 [24] describes an approval process that contains substantial overlaps with the BCF information. The VDI/bs 2552 Part 11.2 [25] describes another component-based process with significant overlaps with the BCF standard. These examples show that many component-based communication processes are
founded on a common information root, but no connection exists between them. Furthermore, is the link in BCF to the targeted building elements in the communication relatively weak and placed in a sub-concept of the Viewpoints. This prevents the potential of BCF from being adequately exploited. By linking to the respective building elements, it should be possible to find out which problems selected building elements have or have had in the past. This would be an essential function in the digital planning of a building and its elements over its entire life cycle - for example, in the domain of facility management.

However, this link in BCF cannot be easily queried from the server without querying or downloading the entire database [3]. bcfOWL can circumvent this problem by querying via SPARQL, but this again has the disadvantage that this can only happen in a BCF context. If information were noted for building elements in a non-BCF context, this would lead to a theoretical misuse of bcfOWL. A higher-level framework for component-based communication could bypass this problem and maintain compatibility with different processes.

Many ontologies, processes and schemas represent different subsets of component-based communication. Still, a higher-level schema is missing so far, which would allow for establishing compatibility and synergy between them. The higher-level ontology can then be used to describe the content of, e.g., BCF without clearly labelling it as bcfOWL. In fact, it could be argued that bcfOWL itself does not need to be specifically defined but that certain SHACL shapes, which check for BCF compliance and use the higher-level ontology, define whether the incoming data is BCF compatible or not.

### 3.3.2. Spatial Information

BCF provides users with a standardised and software-independent way to locate camera positions in a digital building model. While primarily targeted towards BIM, it can also be used in CAD or non-BIM models. While this is a significant achievement, it again comes with the disadvantage that this is always connected to the BCF context. Therefore, the core parts of the Viewpoint responsible for spatially locating an issue could be described in an upper-level ontology for locating and aligning geometry and documents like plans, models and images or, in summary: spatial representations of the building. These spatial representations are supposed to be structured in such a way that they can be placed in a spatial context to each other. This would allow heterogeneous data to be spatially linked to each other, which is a missing feature, especially for existing buildings, where often no coordination process of the digital data has taken place. Nevertheless, this process is also relevant for new buildings to be planned, too, since the coordination process is not always carried out in a satisfactory manner, and the possibility of spatial linking means that these errors can be rectified more quickly. An initial approach to the concept of spatial representations in BCF can be seen in [4], where BCF has been extended to include the functionality to locate damage in a digital building which could potentially be used to locate DOT damages.

In addition to the semantic component of communication, a further investigation of the spatial context becomes evident. Although we can already locate and link photos and geometries to each other, many open questions remain about how spatial representations should be linked to each other in a spatiotemporal context. A framework that allows us to link heterogeneous data spatially in a machine-readable way is not yet clearly available for the AEC domain, neither by
legacy means nor in the Linked Data context. Although this problem is usually circumvented by methods such as coordination processes, this does not always prevent errors. Without a spatial structure, correcting these errors can be very time-consuming. A spatial framework based on ODPs could address this issue as well.

4. Conclusion

The bcfOWL ontology, developed as part of the EU Horizon 2020 BIM4REN project, provides a promising approach to facilitating the communication of issue-related data in the context of BIM. By providing a means of expressing issue data on the Semantic Web, bcfOWL enhances the query ability and connectivity in the context of Linked Building Data, serving as an interlanguage for communication between different systems and domains. However, the use of bcfOWL in the BIM4Ren project has provided new insights into its usability, which this paper discusses.

It was discovered during the process that the ontology is insufficient for the functionality needed for use-cases in other component-based communication processes. While these processes share many similarities with the BCF concepts, they still differ in ways that prevent proper deployment. To address these shortages, this paper proposes decomposing the BCF concepts into parts to form BCF from several generic multi-purpose ontologies based on the Ontology Design Patterns, enabling other similar processes. Furthermore, with future developments, aspects such as versioning and validation should be strengthened in order to ensure the best possible applicability.

The consideration made in this paper should complement the original design of bcfOWL. While the suggestions made in [1] should ensure a sound basis when creating an ontology, this paper focuses on potential improvements and considerations that should help to make the ontologies concepts more reusable.

It is imperative to mention that a basic generalisation of our observation is impossible. However, it remains to be said that the translation of a schema from legacy formats to Linked Data is a frequently chosen approach, in which our presented points should at least be checked for applicability. This should benefit the overall goal of making ontologies more reusable, thus avoiding too much duplication of content between different ontologies.

Therefore, future work will further pursue component-based communication and spatial contextualisation. The goal is to establish higher-level ontologies that can be used to represent general functionalities and processes within buildings and CDEs and from which, in turn, the core components of, e.g. BCF can be derived. A comprehensive investigation is required to estimate how serving legacy formats, such as BCF, using the Semantic Web technology stack impacts the performance and robustness of client-server communication. The prototype servers, which were created within the framework of BIM4Ren to link BCF communication with Linked Data, cannot be sufficiently compared with servers from commercial providers. Nevertheless, by conducting a thorough evaluation, we can determine the optimal level of feature integration that maximizes both user value and server performance.
Acknowledgments

This research has been funded by the European Union’s Horizon 2020 project BIM4Ren.

References


