Demonstrating a pragmatic solution to context associations in RDF using Blank Node Graphs

Ruben Dedecker[0000−0002−3257−3394], Jos De Roo[0000−0001−8862−0666], Beatriz Esteves[0000−0003−0259−7560], and Pieter Colpaert[0000−0001−6917−2167]

IDLab, Department of Electronics and Information Systems,
ruben.dedecker@ugent.be
https://knows.idlab.ugent.be
Ghent University – imec, Technologiepark-Zwijnaarde 122, 9052 Ghent, Belgium

Abstract. Actors in Web ecosystems integrating data from multiple sources–such as in the Solid project or in thematic data spaces–must be able to express requirements over a combination of exchanged data and its associated context. Examples include assigning confidence to integrated data, tracking provenance, signing a specific set of statements, or the expression of policies for usage control. Despite earlier attempts, there is no general consensus on how to model such context associations, resulting in a lack of interoperability for exchanging combined context and data across actors. With this paper, we propose a pragmatic modeling approach for context associations, by referencing Blank Node Graphs, thereby enforcing a specific interpretation of the graph name and restricting its definition to a blank node to avoid accidental merging of graphs. We introduce the "RDF context associations" specification, which defines the data model and the interpretation. The demonstrator then showcases both the modeling of context information over given source data and the filtering of the resulting context associations using SPARQL.

Keywords: Linked Data · RDF · Context Associations · Blank Node Graphs · personal data

1 Introduction

In a 2003 presentation [1] of envisioned challenges for the Semantic Web ecosystem, Sir Tim Berners-Lee proposed the concept of a Web of Trust, stating that all statements on the Web occur in some context, and applications need this context to evaluate the trustworthiness of these statements, doing so with a small set of verifying tools. Recent initiatives aimed at including personal data in Web ecosystems, such as Solid [4] or thematic data spaces[1], benefit from the inclusion of context information in data exchanges for the modeling of signatures and usage restrictions that support their requirements for access and usage control.

[1] https://digital-strategy.ec.europa.eu/en/policies/data-governance-act

In section 2, we propose a pragmatic way to model the association of context information over exact sets of target statements through Blank Node Graphs. In section 3, we then provide a live demonstrator showing this approach in republishing RDF data with associated context that can be filtered using SPARQL.

2 Context associations using Blank Node Graphs

The use of named graphs to model context information is not new [2], but their practical integration in data exchange ecosystems requires imposing both a concrete semantic interpretation of the graph name, discussed in section 2.1, and a syntactic constraint described in section 2.2. We define a Blank Node Graph as a named graph that conforms to both these requirements.

The specification document for the modeling and interpretation of context associations using Blank Node Graphs can be found at `https://w3id.org/context-associations/specification`. An example context association that expresses provenance and a usage policy over data is shown in listing 1.1.

```
@prefix ca: <https://w3id.org/context-associations#>.
@prefix prov: <https://example.org/provenance#>.
@prefix pol: <https://example.org/policy#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
_:dataGraph {
   <http://people.org/ruben> foaf:name "Ruben";
      foaf:age 28.
}
_:dataGraph a ca:GraphIdentifier;
   prov:origin <http://dataspace.org/ruben/resource1>.
_:policy a pol:Policy;
   pol:target _:dataGraph;
   pol:issuer <http://people.org/ruben>;
   pol:permission [
      pol:action pol:Use;
      pol:legalBasis pol:InformedConsent;
      pol:purpose pol:AgeVerification;
   ].
```

Listing 1.1. An association of provenance and policy information to a target Blank Node Graph.

2.1 Defining name interpretation for named graphs

The introduction of named graphs in the RDF syntax[2], represented as a pair consisting of a graph name and an RDF graph, deliberately left the nature of the relationship between the two unspecified. However, their use for modeling context over the RDF graph relies on the interpretation of the graph name as a reference to this RDF graph.

With the context association specification, we impose this interpretation using an HTTP profile `https://w3id.org/context-associations/profile`

[2] `https://www.w3.org/TR/rdf11-concepts/`

expressed over a data source, or by explicitly typing the graph name as a `https://w3id.org/context-associations#GraphIdentifier`. A compliant client can now process the graph name as referencing the connected RDF graph.

2.2 Preventing graph merges through Blank Node Graphs

The RDF specification defines that merging RDF sources requires the merging of named graphs with identical graph names. The use of URIs as graph names in context associations therefore risks accidentally merging graphs when combining data from different sources. To ensure immutability, the use of blank nodes as graph names prevents such accidental merges, as RDF requires the relabeling of blank nodes when combining data from different sources [3].

The modeling of a Blank Node Graph as a named graph with a blank node as graph name that is typed according to section 2.1, provides a graph that can be referenced by context information while remaining unambiguous under storage, exchange, and processing in RDF Web ecosystems. In cases where external references are required, graph names can be defined as skolem identifiers to ensure their unique generation while supporting external references [3].

2.3 Querying context associations with SPARQL

The context association model defined above is native to RDF, and hence can be stored in RDF-compatible graph stores and queried using SPARQL. An example evaluation of policy and provenance requirements over the context association defined in listing 1.1 using SPARQL can be found in listing 1.2. This example can be run on the Comunica Web interface.

```
PREFIX po: <http://purl.org/ontology/po/>
PREFIX ca: <https://w3id.org/context-associations#>
PREFIX prov: <https://example.org/provenance#>
PREFIX pol: <https://example.org/policy#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
CONSTRUCT {
   ?person foaf:name ?name.
} WHERE {
   ?targetDataGraph a ca:GraphIdentifier;
      prov:origin <http://dataspace.org/ruben/resource1>.
   ?policy a pol:Policy;
      pol:target ?targetDataGraph;
      pol:issuer <http://people.org/ruben>;
      pol:permission ?perm.
   ?perm pol:action pol:Use.
   ?perm pol:purpose pol:AgeVerification.
   GRAPH ?targetDataGraph {
      ?person foaf:name ?name.
   }
}
```

Listing 1.2. A SPARQL Construct query asserting statements based on combined requirements for target statements and associated policy and provenance information.

3 Demonstrator

To demonstrate the use of Blank Node Graphs to define and integrate associated context information to target data, we provide a Web demonstrator[3]. The demonstrator consists of two interfaces. The first interface allows a visitor to provide an RDF resource and a selection of provenance, policy, and signature information to attach to this data. The resulting context associations are then displayed and re-published on the Web. The second interface then demonstrates how these context associations can be processed, allowing the visitor to provide a set of resources and to apply context filters that result in the generation of a SPARQL query. This query is then evaluated over the provided resources, resulting in only the data that conforms to the requested context filters.

4 Conclusion and Perspective

With this paper, we show that any RDF data can be annotated with contextual information today, the result of which can be stored, exchanged, and processed in a Web ecosystem as RDF data, and queried using SPARQL.

With the perspective of the upcoming RDF 1.2 standard, our position is that the inclusion of Triple Terms in the RDF standard does not obsolete the use of Blank Node Graphs to model context associations, as per-triple annotations similarly require additional semantic definitions to model closed sets of statements.

Acknowledgments. Supported by SolidLab Vlaanderen (Flemish Government, EWI and RRF project V023/10)

References

1. Berners-Lee, T.: The semantic web and challenges. W3C Web site Slideshow (2003)
2. Carroll, J.J., Bizer, C., Hayes, P., Stickler, P.: Named graphs, provenance and trust. In: Proceedings of the 14th international conference on World Wide Web. pp. 613–622. WWW '05, Association for Computing Machinery, New York, NY, USA (2005). https://doi.org/10.1145/1060745.1060835, https://dl.acm.org/doi/10.1145/1060745.1060835
3. Hogan, A., Arenas, M., Mallea, A., Polleres, A.: Everything you always wanted to know about blank nodes. Journal of Web Semantics **27-28**, 42–69 (Aug 2014). https://doi.org/10.1016/j.websem.2014.06.004, https://linkinghub.elsevier.com/retrieve/pii/S1570826814000481
4. Sambra, A.V., Mansour, E., Hawke, S., Zereba, M., Greco, N., Ghanem, A., Zagidulin, D., Aboulnaga, A., Berners-Lee, T.: Solid: a platform for decentralized social applications based on linked data. MIT CSAIL & Qatar Computing Research Institute, Tech. Rep. **2016** (2016)

[3] The demonstrator is located at https://github.com/KNowledgeOnWebScale/rdf-context-associations-demo