Sem4Tra Workshop

On the visualization of semantic-based mappings

Nicolò Pinciroli, Mario Sacaj, Mersedeh Sadeghi, Safia Kalwar, Andreas Vogelsang and Matteo Rossi

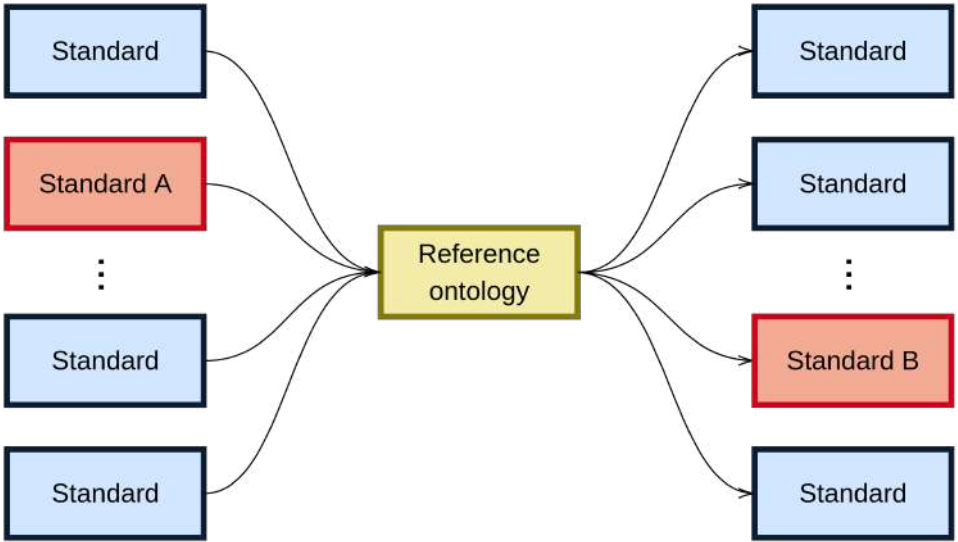
September 6, 2021

Introduction

Research question

How can coordinated views be employed by users to translate terms from a standard to an ontology?

Research question



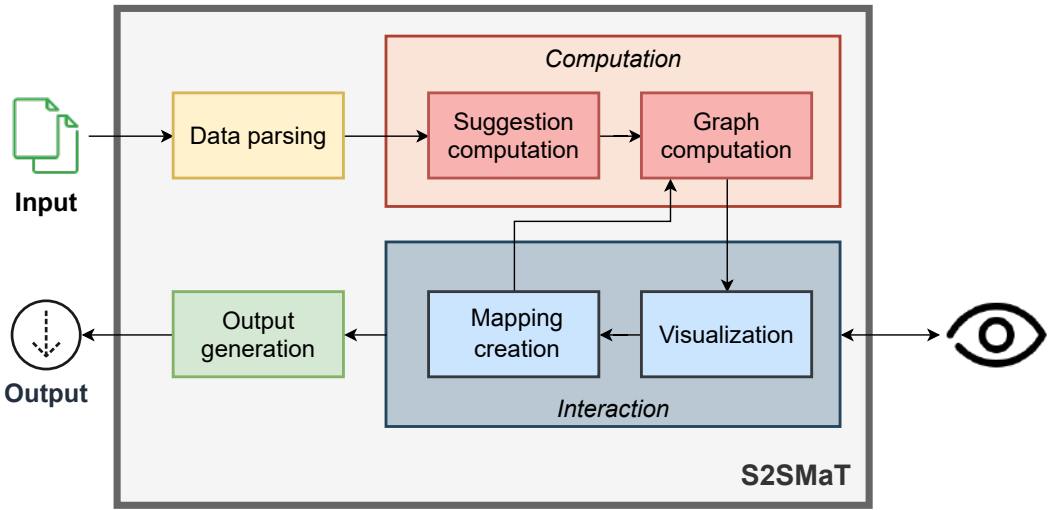
Motivations

- Railway systems concepts represented using **ontologies**, hardly **readable** by humans;
- **Inaccurate** automatic associations creation;
- Effectiveness of **coordinated views**.

Background

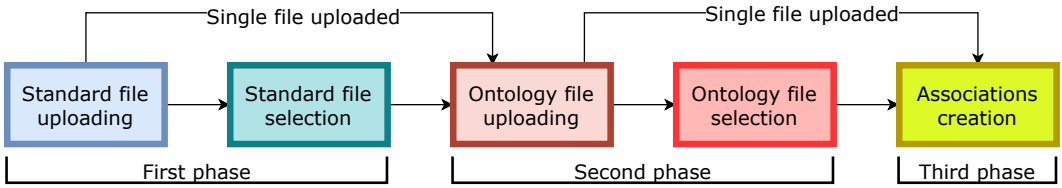
- **Ontologies** used only in specific sub-domains
  (e.g., railway safety [Hul+18] and the Dutch railways [BRS17]);
- Existing libraries (e.g., VOWL [Loh+16]) allow to represent **structured** data as graphs
- Existing tools allow to create **coordinated views**; [Vag+20], [Spu+20];
- Existing predictive **model** for suggesting associations.

Architecture



User Interface Design

The application workflow



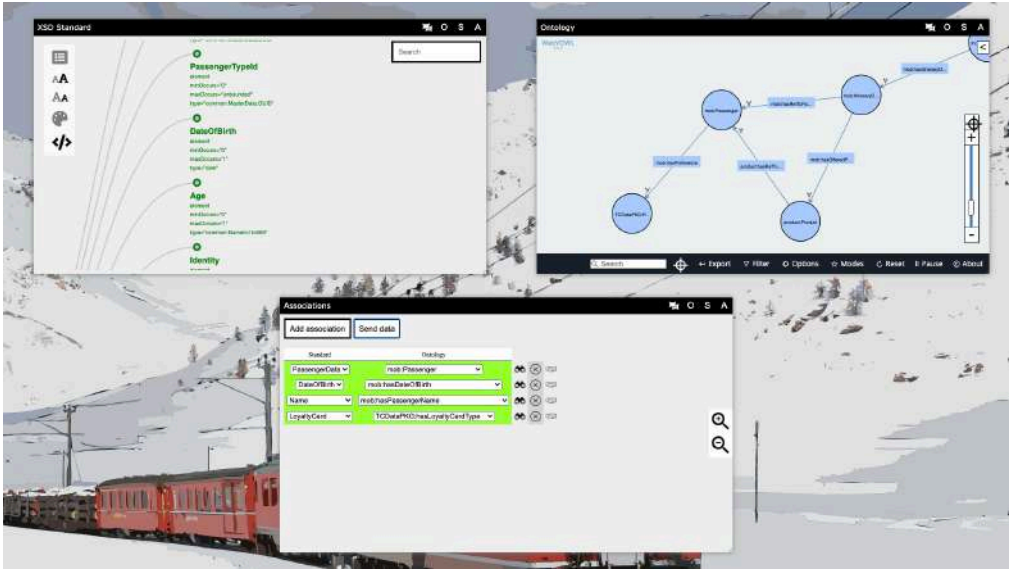
Binding representation

- Complex Type (XSD) ⟷ Class (Ontology)
- Element/Attribute (XSD) ⟷ Property (Ontology)

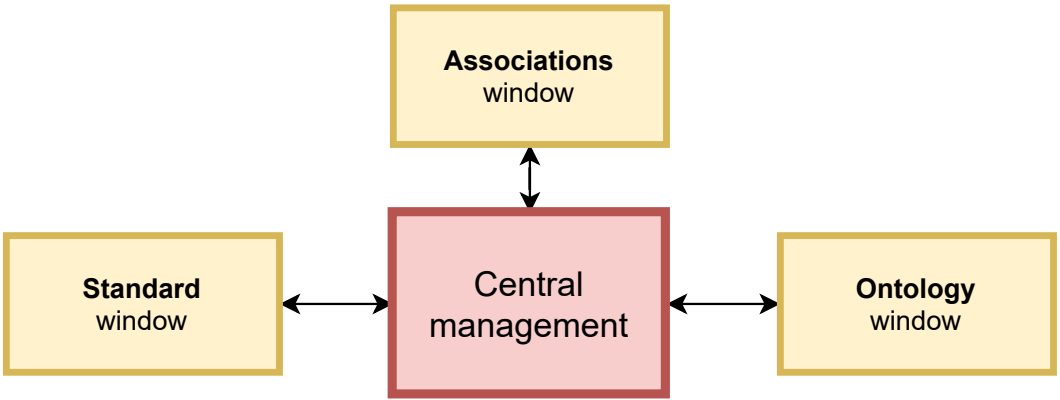
Suggestion computation

- Suggested one-to-one **mapping** between the concepts in **XSD** and the ones in the **ontology**
- Exploits a Word2vec-trained model (based on the Google News dataset)
- **Suggestions** → topmost similar terms
- Filtering based on binding representations
- Output → XSD-Ontology pairs

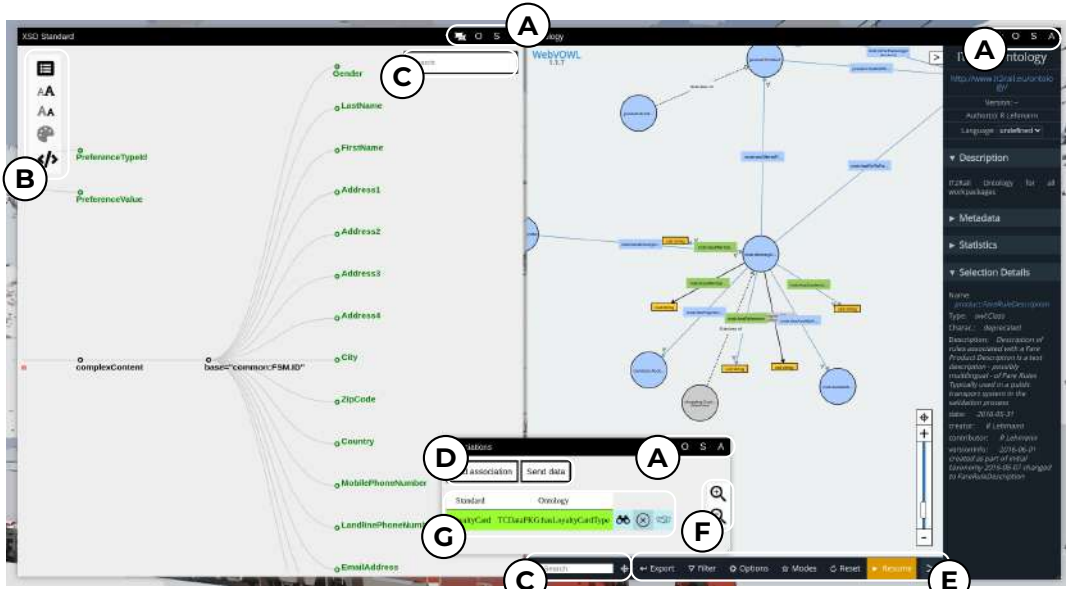
Associations creation overview



Associations creation overview

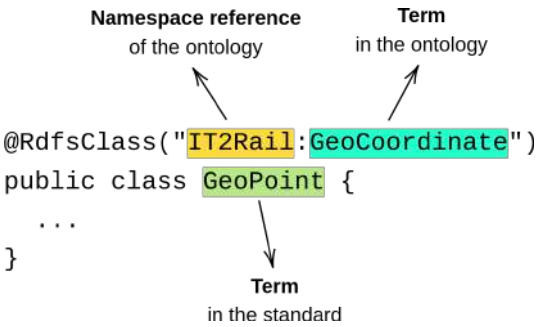


Associations creation interface and Gestalt principles



Output generation

1. XSD representation → Java constructs (mappings **materialization**)
   - Complex types → Java classes
   - Elements and attributes → Attributes and setter/getter methods
2. Annotation of Java constructs
3. Generation of a compressed file



Conclusions

- **Modular** application
- **Focus** on more intuitive data visualization and visualization principles
- **Integration** with the back-end
- Part of a **broader** project


Thank you for your attention

References I

[BOH11] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. "D^3 Data-Driven Documents". In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (Dec. 2011), pp. 2301–2309. ISSN: 1077-2626. DOI: 10.1109/TVCG.2011.185. URL: https://doi.org/10.1109/TVCG.2011.185.

[BRS17] Bas Bach, Mark von Rosing, and Henrik von Scheel. "Using Ontology and Modelling Concepts to Develop Smart Applications: Example Dutch Railway". In: *Int. J. Concept. Struct. Smart Appl.* 5.1 (2017), pp. 48–69. DOI: 10.4018/IJCSSA.2017010103. URL: https://doi.org/10.4018/IJCSSA.2017010103.

[Cow01] Nelson Cowan. "The magical number 4 in short-term memory: A reconsideration of mental storage capacity". In: *Behavioral and Brain Sciences* 24.1 (2001), pp. 87–114. DOI: 10.1017/S0140525X01003922.

References II

[Hul+18] Bernhard Hulin et al. “Towards a Common Ontology of Safety Risk Concepts for Railway Vehicles and Signaling”. In: *Computer Safety, Reliability, and Security - 37th International Conference, SAFECOMP 2018, Västerås, Sweden, September 19-21, 2018, Proceedings*. Ed. by Barbara Gallina, Amund Skavhaug, and Friedemann Bitsch. Vol. 11093. Lecture Notes in Computer Science. Springer, 2018, pp. 297–310. DOI: 10.1007/978-3-319-99130-6_20. URL: https://doi.org/10.1007/978-3-319-99130-6%5C_20.

[Loh+16] Steffen Lohmann et al. “Visualizing Ontologies with VOWL”. In: *Semantic Web* 7.4 (2016), pp. 399–419. DOI: 10.3233/SW-150200. URL: http://dx.doi.org/10.3233/SW-150200.

[Mil94] George A Miller. “The magical number seven, plus or minus two: Some limits on our capacity for processing information.”. In: *Psychological review* 101.2 (1994), p. 343.

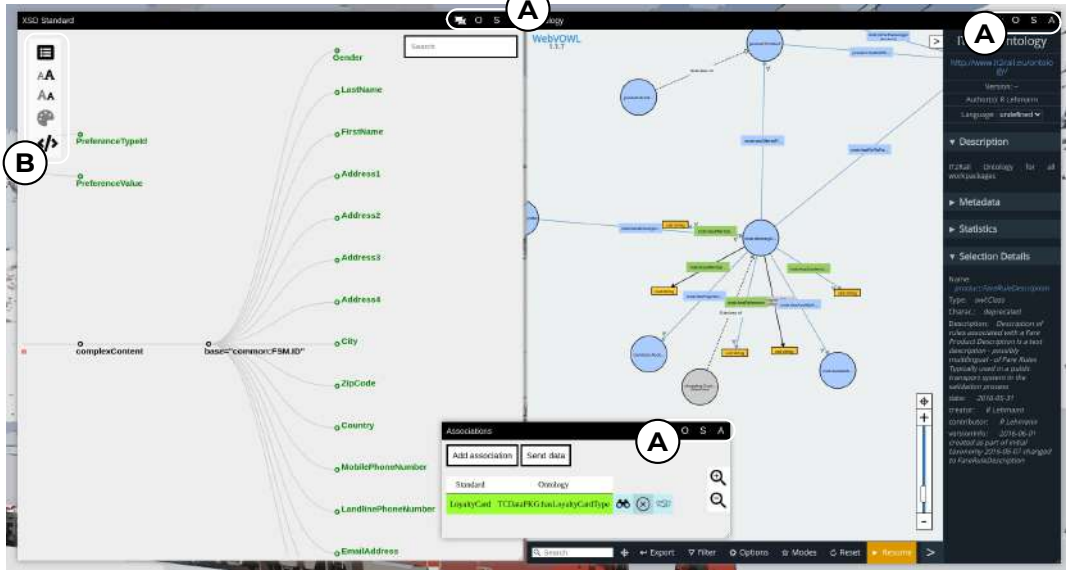
References III

[Spu+20] Maxim Spur et al. “Exploring Multiple and Coordinated Views for Multilayered Geospatial Data in Virtual Reality”. In: *Inf.* 11.9 (2020), p. 425. DOI: 10.3390/info11090425. URL: https://doi.org/10.3390/info11090425.

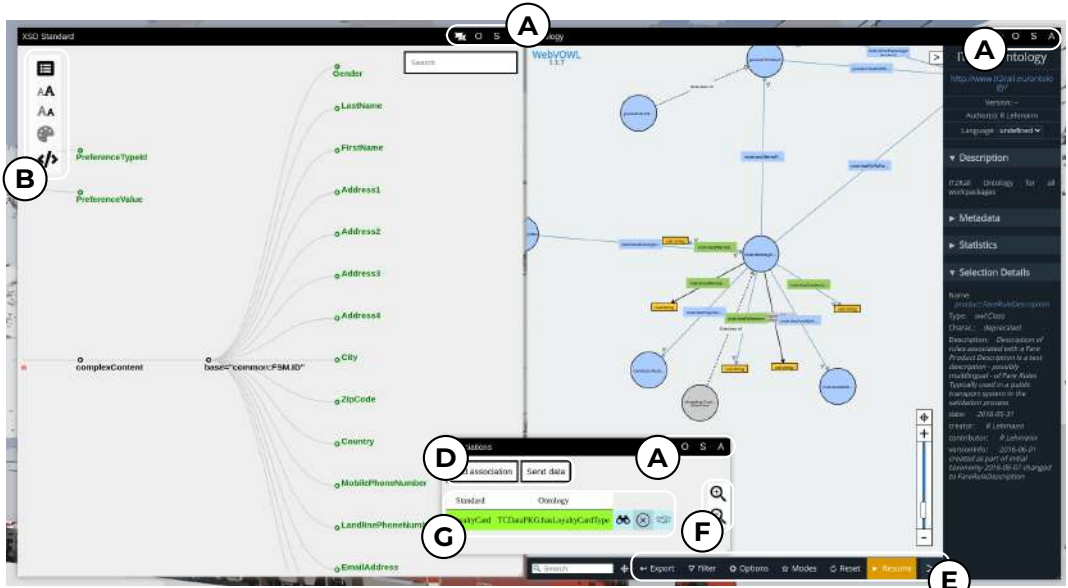
[Vag+20] Nicolo Oreste Pinciroli Vago et al. “INTEGRA: An Open Tool To Support Graph-Based Change Pattern Analyses In Simulated Football Matches”. In: *Proceedings of the 34th International ECMS Conference on Modelling and Simulation, ECMS 2020, Wildau, Germany, June 9-12, 2020*. Ed. by Mike Steglich et al. European Council for Modeling and Simulation, 2020, pp. 228–234. DOI: 10.7148/2020-0228. URL: https://doi.org/10.7148/2020-0228.

Visualization principles

Consistency



Proximity and similarity



Short memory holds a limited amount of items [Mil94; Cow01]

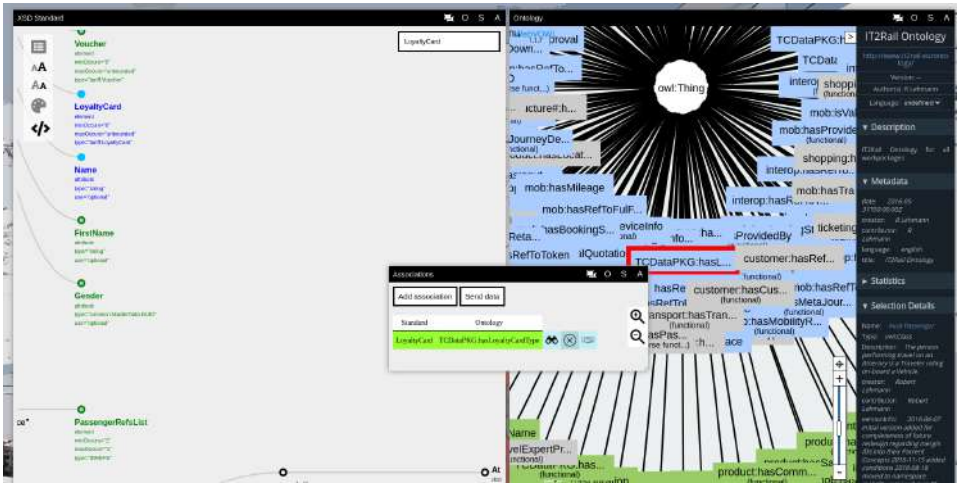
1. Shallow menus
2. Visual feedback after selection
3. Standard and consistent icons

Interaction

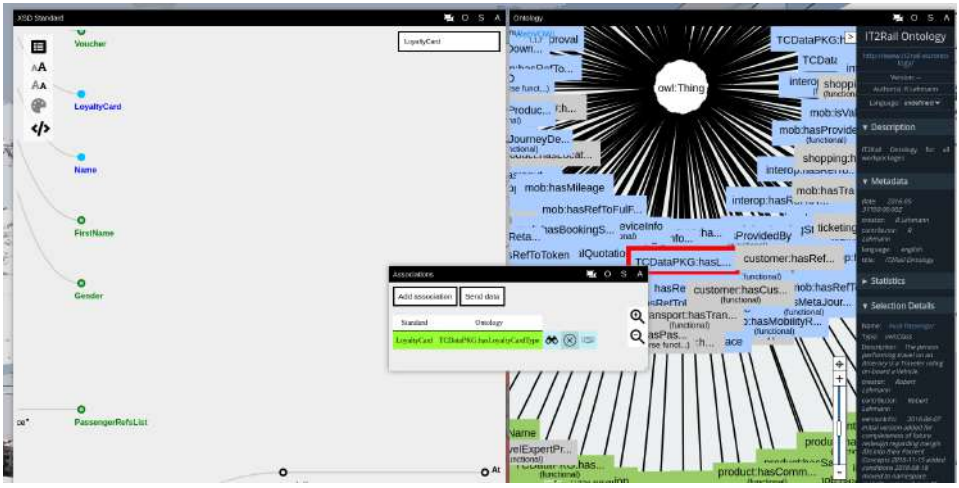
Thinking is interactive

1. Windows can be moved, resized and set always on top
2. Graphs can be dragged, zoomed, coloured, ...
3. Variable amount of details

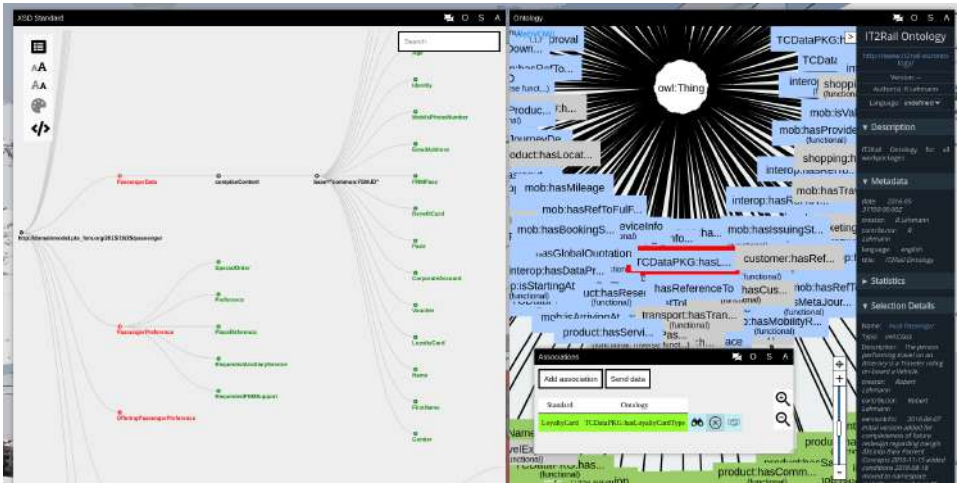
Details


XSD Standard
Voucher
LoyaltyCard
Name
FirstName
Gender
PassengerRefsList
Ontology
owl:Thing
IT2Rail Ontology
Description
Metadata
Statistics
Selection Details
Associations
Add association
Send data
Standard
Ontology
mob:hasMileage
mob:hasRefToFullF...
TCDataPKG:hasL...
customer:hasRef...
customer:hasCus...
product:hasComm...

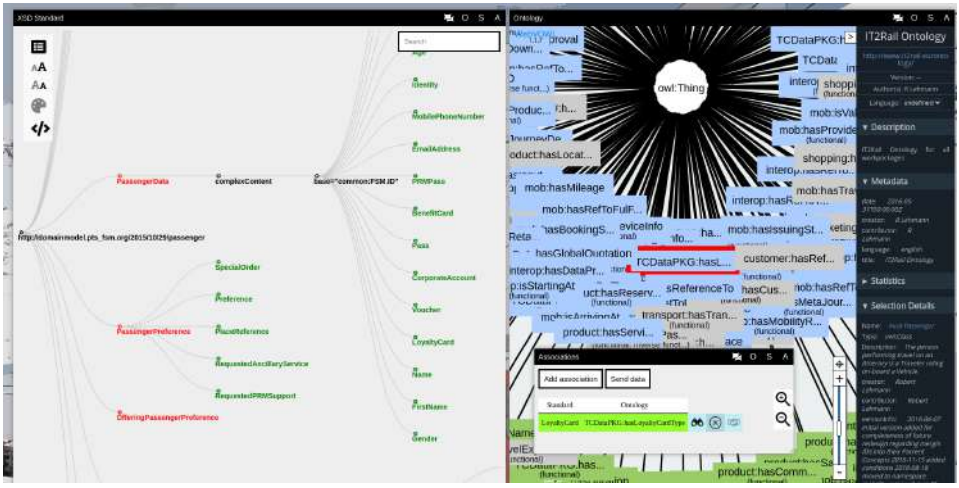
Details


XSD Standard
Voucher
LoyaltyCard
Name
FirstName
Gender
PassengerRefsList
Ontology
owl:Thing
IT2Rail Ontology
Description
Metadata
Statistics
Selection Details
Associations
Add association
Send data
Standard
Ontology
TCDataPKG:hasL...
customer:hasRef...
mob:hasMileage
mob:hasRefToFullF...
product:hasComm...

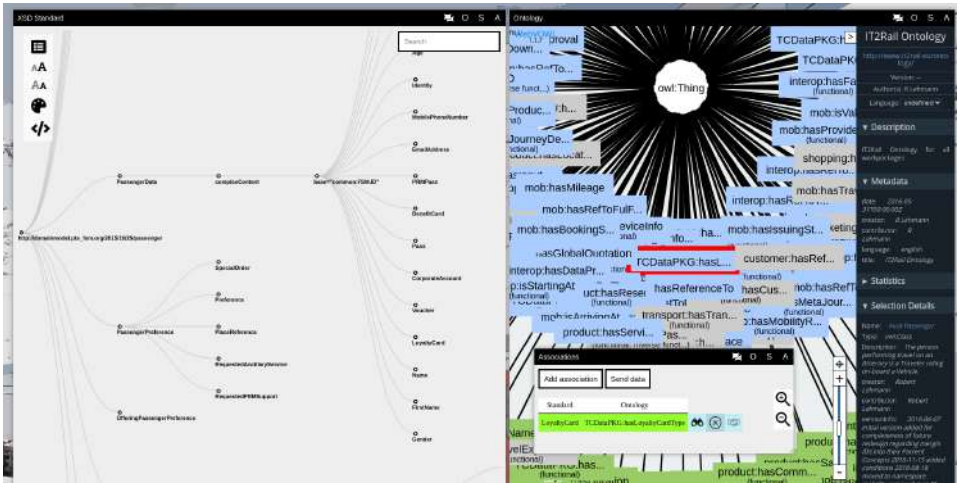
Text size


XSD Standard
Ontology
owl:Thing
IT2Rail Ontology
Associations
Add association
Send data
Description
Metadata
Statistics
Selection Details

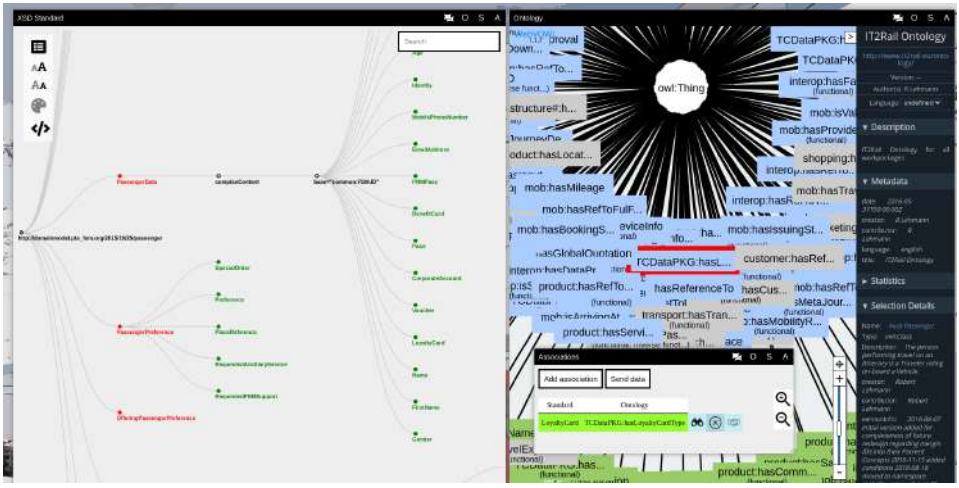
Text size


XSD Standard
PassengerData
complexContent
http://domainmodel.pts_fsm.org/2015102S/passenger
SpecialOrder
Preference
PassengerPreference
PaxIdReference
RequestedAuxiliaryService
RequestedPRMSupport
OfferingPassengerPreference
Identity
MobilePhoneNumber
EmailAddress
PRMPass
BenefitCard
Pass
CorporateAccount
Voucher
LoyaltyCard
Name
FirstName
Gender
Ontology
owl:Thing
IT2Rail Ontology
Description
Metadata
Statistics
Selection Details
Associations
Add association
Send data
mob:hasMileage
mob:hasRefToFullF...
TCDataPKG:hasL...
customer:hasRef...
product:hasServi...
transport:hasTran...
product:hasComm...

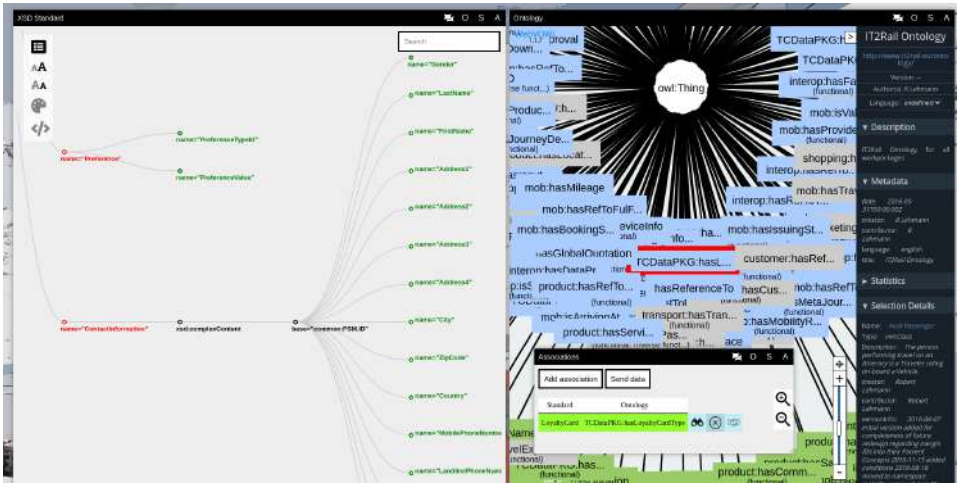
Color


XSD Standard
Ontology
owl:Thing
IT2Rail Ontology
Description
Metadata
Statistics
Selection Details
Associations
Add association
Send data
mob:hasMileage
mob:hasRefToFullF...
mob:hasBookingS...
customer:hasRef...
transport:hasTran...
product:hasServi...
product:hasComm...
TCDataPKG:hasL...
interop:hasDataPr...
p:isStartingAt
hasReferenceTo
mob:hasMobilityR...
mob:hasProvide
shopping:h
TCDataPK
Identity
MobilePhoneNumber
EmailAddress
PrintPass
DebitCard
Pass
CorporateAccount
Voucher
LoyaltyCard
Name
FirstName
Gender
PassengerData
completeContext
PassengerPreference
PassReference
Preference
SpecialOrder
OfferingPassengerPreference

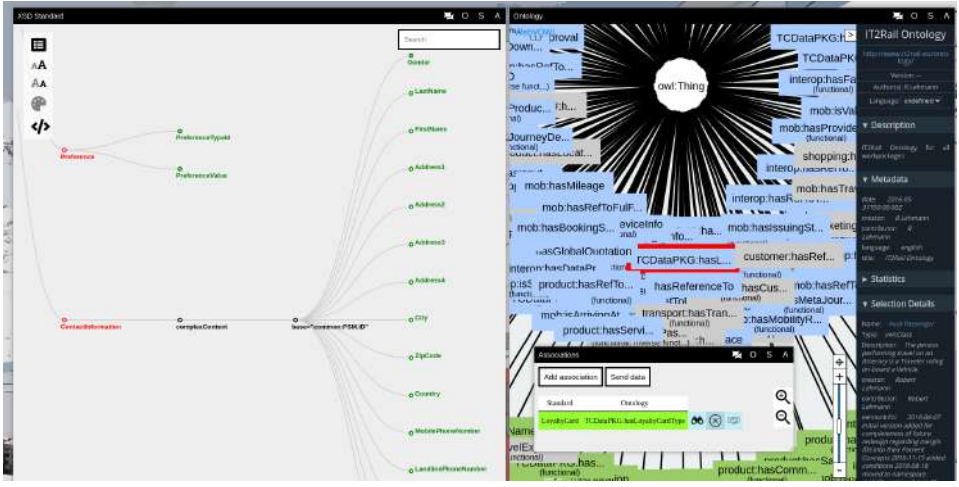
Color


XSD Standard
Ontology
owl:Thing
IT2Rail Ontology
Description
Metadata
Statistics
Selection Details
Associations
Add association
Send data
Standard
Ontology

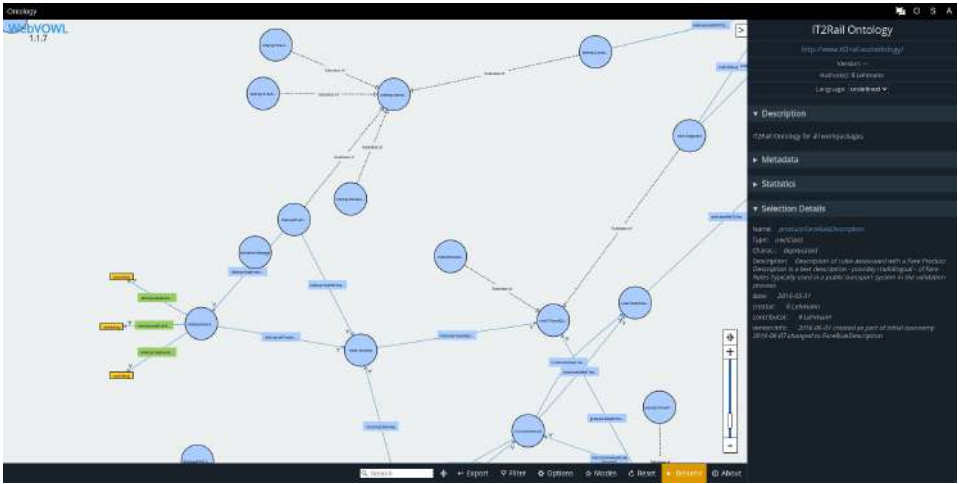
Verbose


XSD Standard
Ontology
owl:Thing
IT2Rail Ontology
Description
Metadata
Statistics
Selection Details
Associations
Add association
Send data
mob:hasMileage
mob:hasRefToFullF...
mob:hasBookingS...
mob:hasIssuingSt...
customer:hasRef...
hasReferenceTo
transport:hasTran...
product:hasServi...
product:hasComm...
interop:hasR...
TCDataPKG:hasL...
TCDataPKG:h...
TCDataPK...
mob:hasProvide...
shopping:h...
mob:hasTra...
mob:isVa...
interop:hasFa...
product:hasRefTo...
hasGlobalOrientation
hasCus...
hasMobilityR...
LoyaltyCard
TCDataPKG:hasLoyaltyCardType

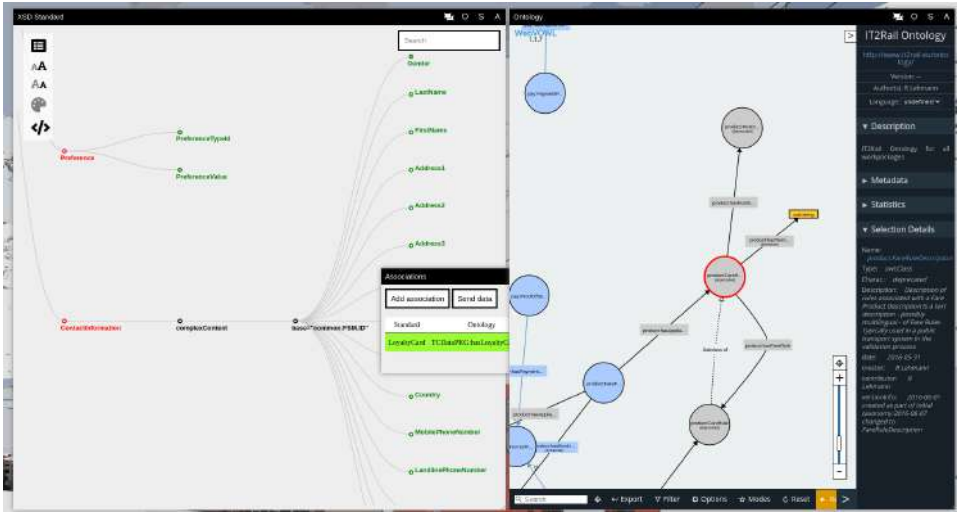
Verbose


XSD Standard
Ontology
owl:Thing
IT2Rail Ontology
Description
Metadata
Statistics
Selection Details
Associations
Add association
Send data
Standard
Ontology
LoyaltyCard
TCDataPKG:hasLoyaltyCardType

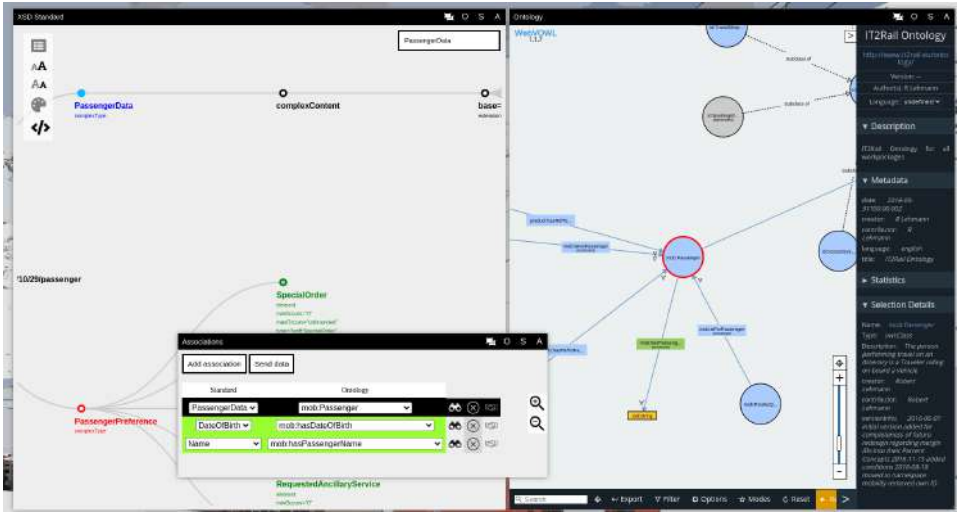
Maximization



Bring on top

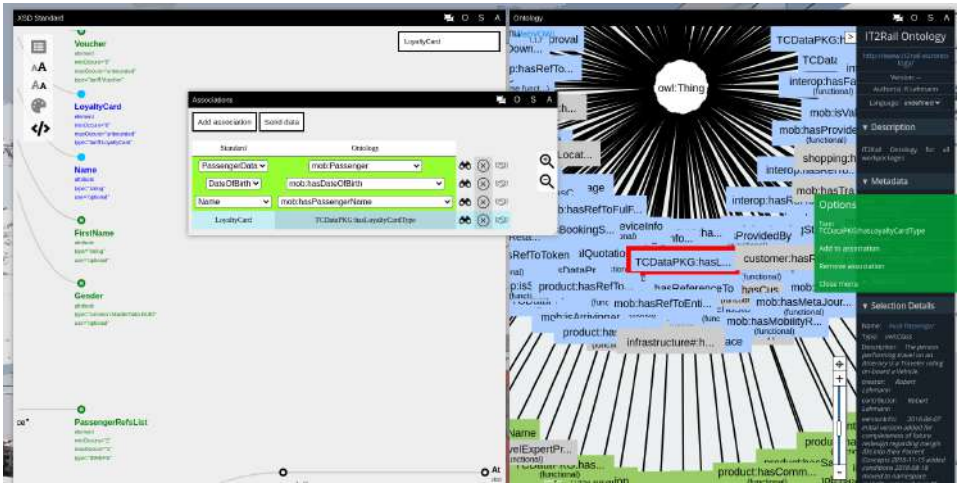

XSD Standard
Ontology
IT2Rail Ontology
Associations
Add association
Send data
Standard
Ontology
Description
Metadata
Statistics
Selection Details
Export
Filter
Options
Modes
Reset

Association retrieval



Association creation

Association creation



Implementation

Main contributions

- Integration and adaptation of existing visualization libraries and tools [Loh+16; BOH11; Vag+20]
- Implementation of a modular web-based application
- Integration with the back-end system

Strategy

- Each module creates independent visualizations
- One module corresponds to one window
- User interactions are managed on the client side

Asynchronous information retrieval

- Data are requested using JavaScript requests
- Asynchronous requests don’t block the interface while waiting
- Each module performs independent requests


Input and output examples

```
<xsd:complexType name="Identity">
    <xsd:complexContent>
        <xsd:extension base="common:FSM.ID">
            <xsd:attribute name="IdentityCard" type="xsd:string" use="required">
            </xsd:attribute>
            ...
        </xsd:extension>
    </xsd:complexContent>
</xsd:complexType>
```

```
<xsd:complexType name="PassengerPreference">
    <xsd:sequence minOccurs="0" maxOccurs="1">
        <xsd:element name="SpecialOrder" type="tariff:SpecialOrder" minOccurs
            ="0" maxOccurs="unbounded">
        </xsd:element>
        <xsd:element name="Preference" type="passenger:Preference" minOccurs="0"
             maxOccurs="unbounded">
        </xsd:element>
        …
    </xsd:sequence>
</xsd:complexType>
```

```
<owl:Class rdf:about="http://www.it2rail.eu/ontology/InteroperabilityFramework#
    PrivateStation">
    <rdfs:subClassOf rdf:resource="http://www.it2rail.eu/ontology/
        InteroperabilityFramework#StopPlace"/>
    <dc:contributor rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        Robert Lehmann</dc:contributor>
    <dc:creator rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Robert
         Lehmann</dc:creator>
    <dc:date rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime
        ">2016-07-12T10:57:53Z</dc:date>
    <dc:description rdf:datatype="http://www.w3.org/2001/XMLSchema#string"/>
    <owl:versionInfo rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    2016-07-12 initial version 2017-08-30 moved to namespace
        InteroperabilityFramework
    </owl:versionInfo>
</owl:Class>
```

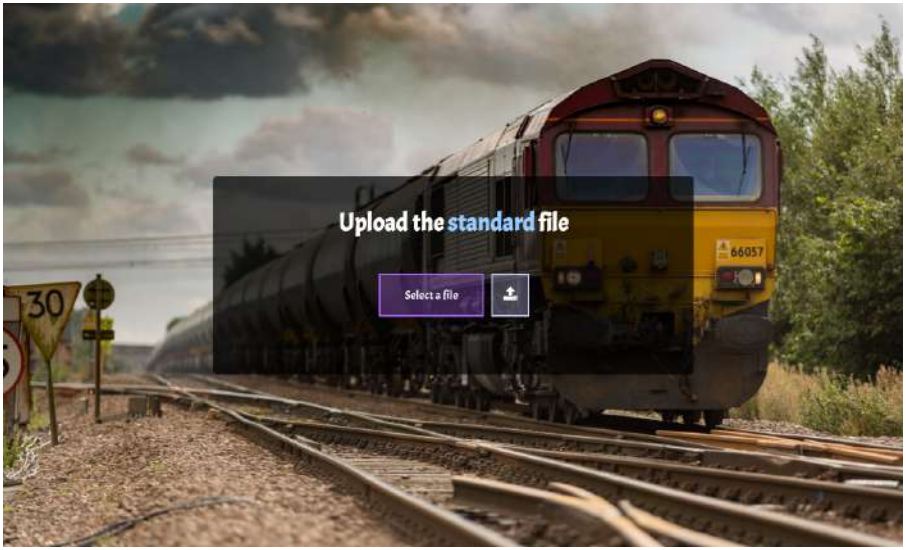
```
<owl:ObjectProperty rdf:about="http://www.it2rail.eu/ontology/transport#
    hasTransportServiceProviderID">
    <rdfs:subPropertyOf rdf:resource="http://www.it2rail.eu/ontology/hasID"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
    ...
    <dc:description rdf:datatype="http://www.w3.org/2001/XMLSchema#string"/>
    <an:i2rumlDomain rdf:datatype="http://www.w3.org/2001/XMLSchema#string
        ">TransportServiceProvider</an:i2rumlDomain>
    <an:i2rumlRange rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        TransportServiceProviderID</an:i2rumlRange>
    <owl:deprecated rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean">
        true</owl:deprecated>
    <owl:versionInfo rdf:datatype="http://www.w3.org/2001/XMLSchema#string
        ">2017-02-08 initial version</owl:versionInfo>
</owl:ObjectProperty>
```

Generated Java code

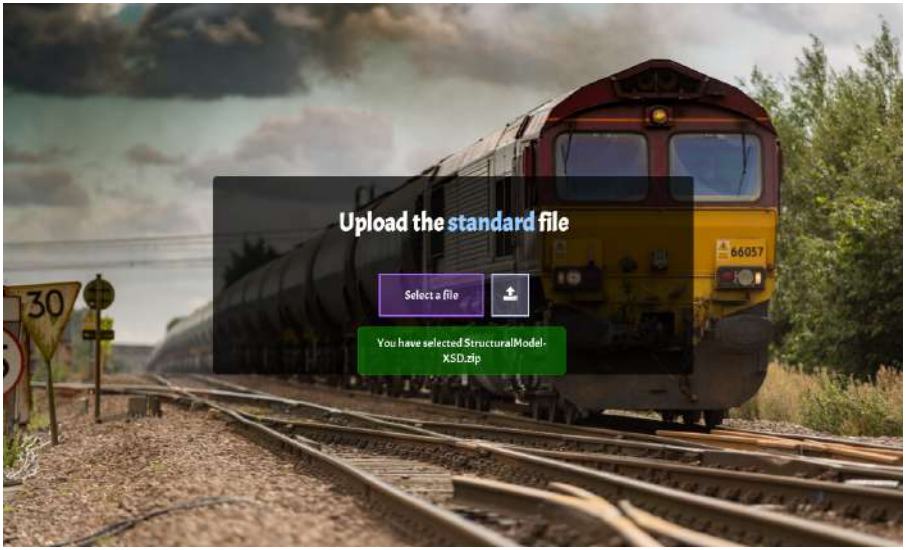

```
@RdfsClass("mob:Passenger")
@NameSpaces(...)
public class PassengerData extends FSMID
{
    @XmlElement(name = "DateOfBirth")
    @XmlSchemaType(name = "date")
    @RdfProperty(propertyName = "mob:hasDateOfBirth")
    protected XMLGregorianCalendar dateOfBirth;
    @XmlElement(name = "Voucher")
    protected List<Voucher> voucher;
    ...
    public List<String> getIdentityTypeId() {
        if (identityTypeId == null) {
            identityTypeId = new ArrayList<String>();
        }
        return this.identityTypeId;
    }
}
```

Workflow details

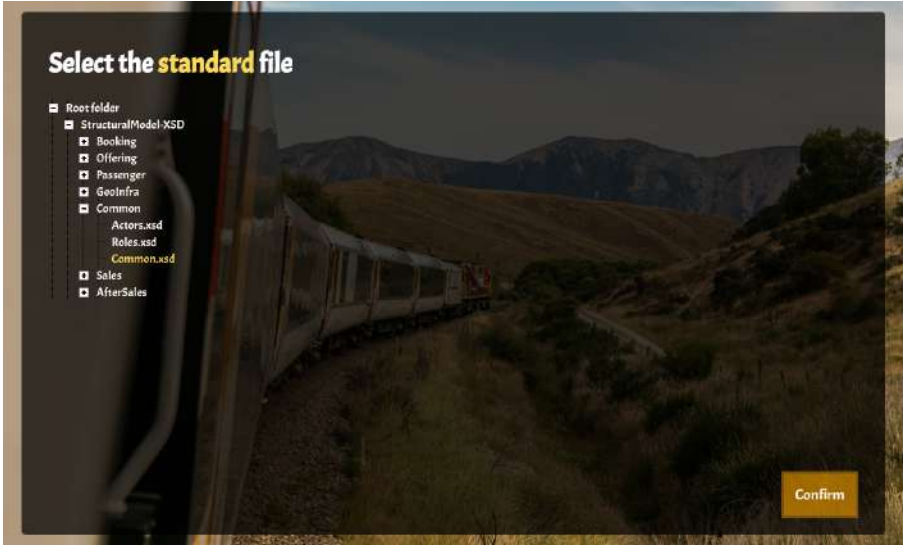
File selection



File selection



File selection



Data parsing

- Syntactic validity
- File parsing
- Pre-processing
- Cleaning

Suggestion computation

- Suggested one-to-one **mapping** between the concepts in **XSD** and the ones in the **ontology**
- Exploits a Word2vec-trained model (based on the Google News dataset)
  - Word → 300-dimensional feature vector
  - Relative distances between vectors ⟷ semantic similarity
- **Suggestions** → topmost similar terms
- Filtering based on binding representations
- Output → XSD-Ontology pairs