SEMANTiCS 2018 – 14th International Conference on Semantic Systems

- Vienna, Hilton Waterfront Conference Hotel, Meeting Room 5
- Session 2.4, 12 Sept 2018, 5 pm

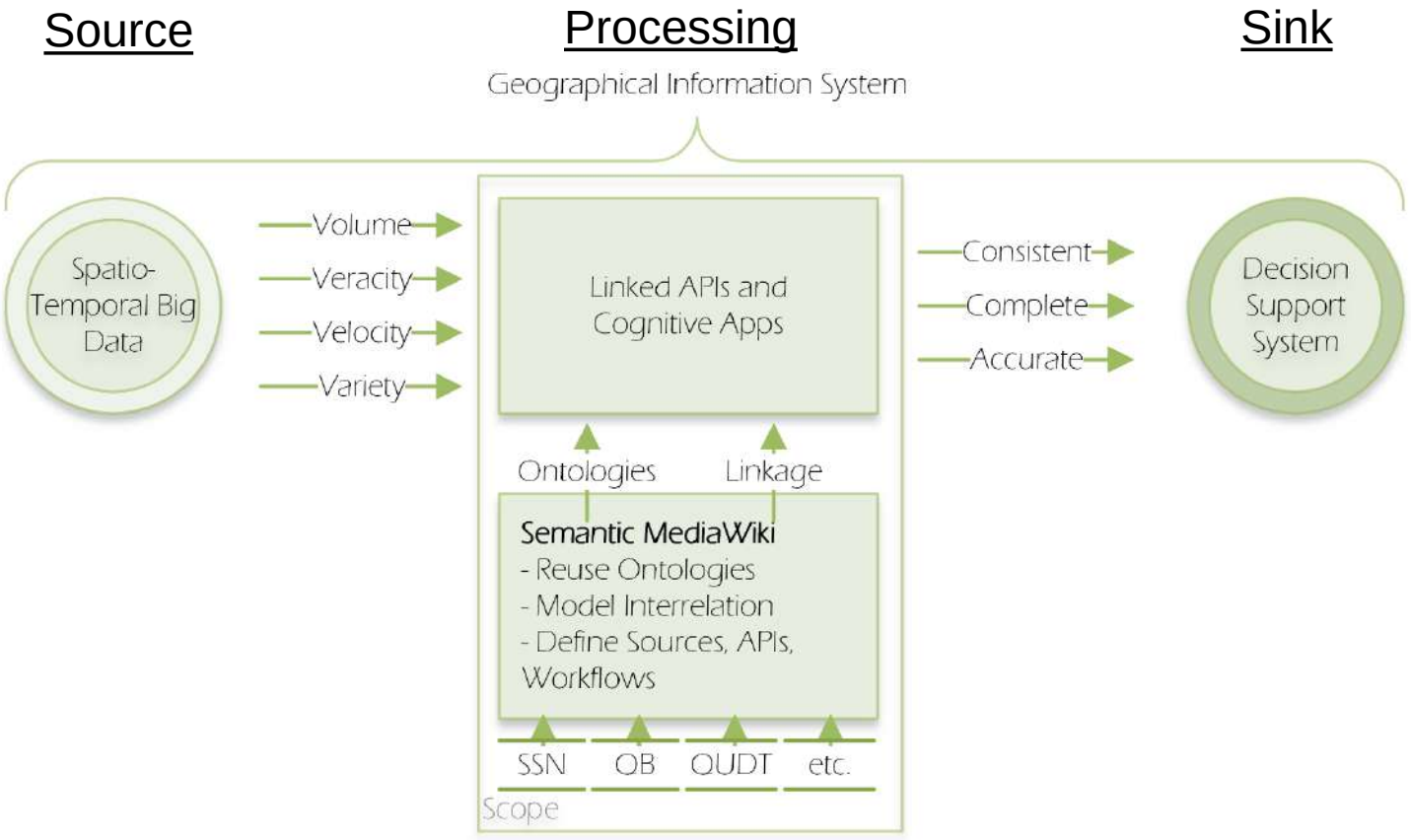
LSane: Collaborative Validation and Enrichment of Heterogeneous Observation Streams

Matthias T. Frank[a], **Sebastian Bader[b]**, Viliam Simko[a], Stefan Zander[c]

[a] FZI Research Center for Information Technology, Karlsruhe, Germany
[b] Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Sankt Augustin, Germany
[c] Hochschule Darmstadt, Darmstadt, Germany

FZI FORSCHUNGSZENTRUM INFORMATIK

Scope of Work



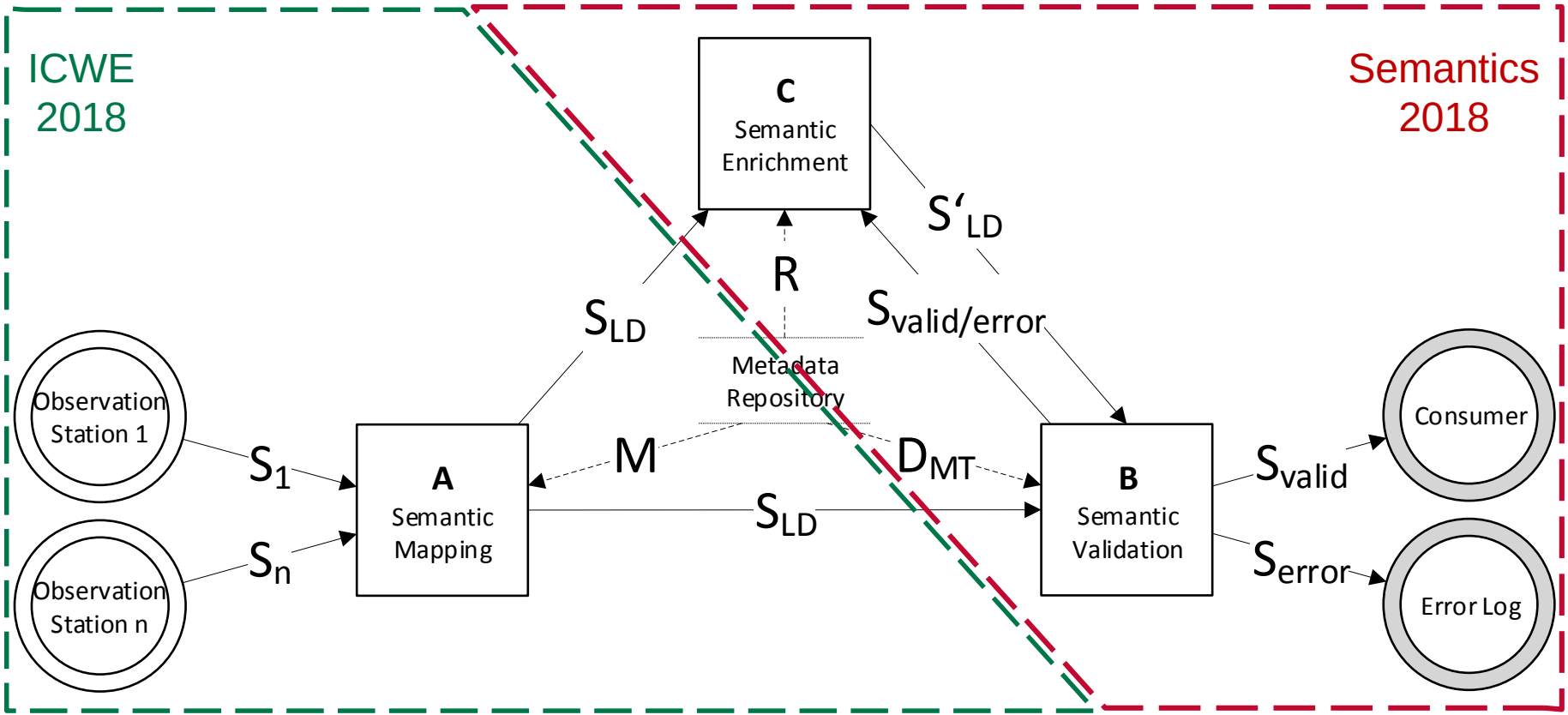
Research Questions

1. How can **heterogeneous** messages of environmental observations be **collaboratively** validated with uniform shapes?
2. How can we exploit these annotations to derive and apply rules for **semantic enrichment** to heterogeneous messages of environmental observations?

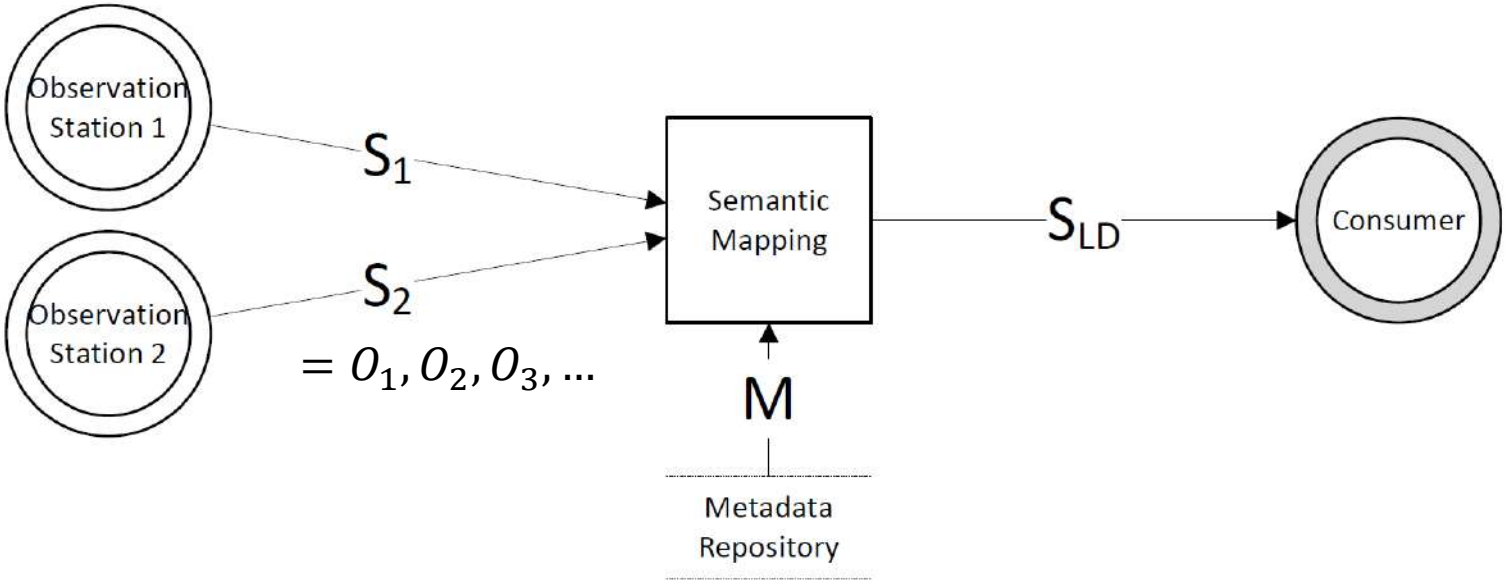
Concept Matrix: Semantic for IoT Streams

Article	Semantic Annotation	Semantic Stream Processing	Collaborative Annotations	Semantic Validation	Semantic Enrichment
Amiguet-Vercher et al (2010)	✓	✓			
Barnaghi et al (2013)	✓	✓			
Daniele et al (2015)	✓				
Das & Cyganiak (2012)	✓				(✓)
Dimou et al (2014)	✓				(✓)
Duy et al (2017)	✓	✓			
Frank & Simko (2018)	✓	✓	✓		
Kolozali et al (2014)	✓	✓			
Markovic & Edwards (2016)	✓				
Markovic et al (2016)	✓	✓			
Bermúdez-Edo et al (2017)	✓				
Wiener et al (2016)	✓			(✓)	(✓)

Approach (Overview)



Semantic Mapping (ICWE 2018)



$$SemMap(O, M) = \{(u, v) : (k, v) \in O, (k, u) \in M\}$$

Semantic Mapping (ICWE 2018)

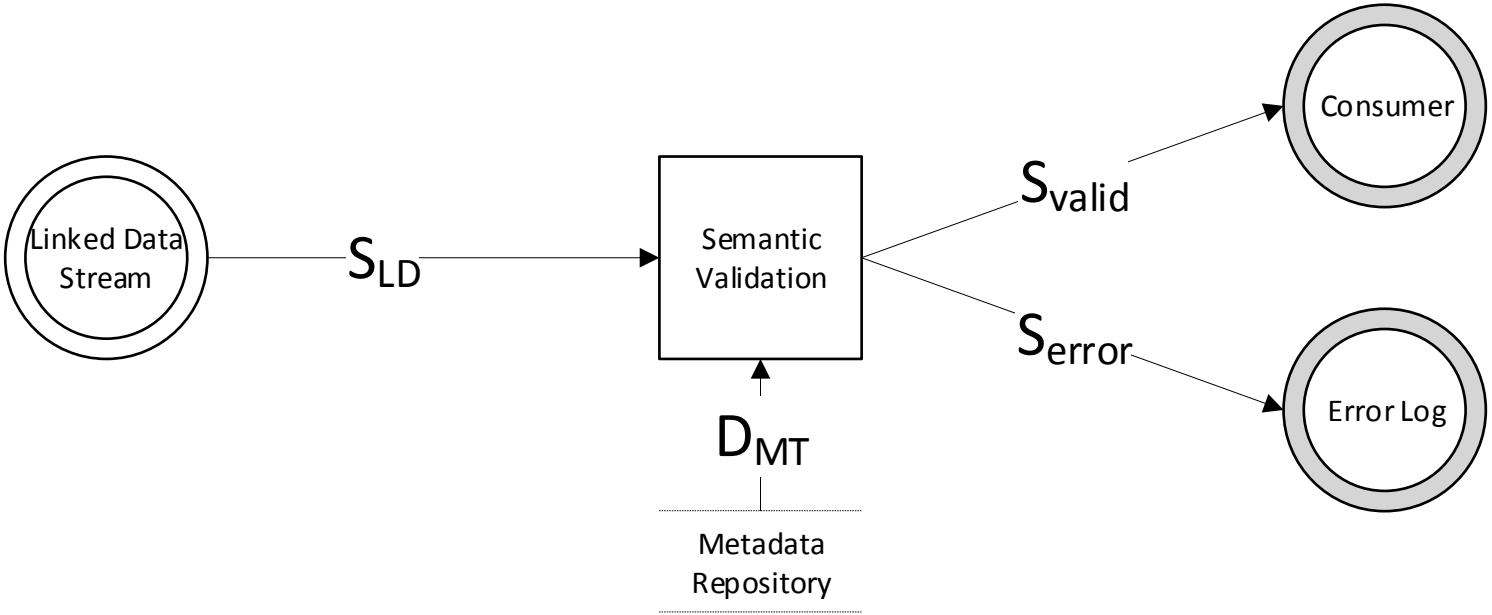
Key-value pairs

Key	Value
temp	28
no2	34
coordinates[0]	8.700066
coordinates[1]	49.30075

Triples

Subject	Predicate	Object
ex:stream1#temp	unit:DegreeCelsius	„28“
ex:stream1#no2	eea:g-m3	„34“
ex:stream1#lat	geo:lat	„8.700066“
ex:stream1#long	geo:long	„49.30075“

Semantic Validation



$$SemVal(O, D_{MT}) = \begin{cases} O_{valid} = O \cup R_{valid} & \text{if } D_{MT} \subseteq O \\ O_{error} = O \cup R_{error} & \text{if } D_{MT} \supset O \end{cases}$$

Semantic Validation – invalid

$$D_{MT} \supset O$$

D_{MT} (definiton of message type)

Subject	Predicate	Object
ex:shape1	rdf:type	sh:property
ex:shape1	sh:path	quantity: Thermodynamic Temperature
ex:shape1	sh:minCount	„1“
ex:shape1	sh:maxCount	„1“

O (observation message)

Subject	Predicate	Object
ex:stream1#temp	unit:DegreeCelsius	„28“
ex:stream1#no2	eea:g-m3	„34“
ex:stream1#lat	geo:lat	„8.700066“
ex:stream1#long	geo:long	„49.30075“

Semantic Validation – valid

$$D_{MT} \subseteq O$$

D_{MT} (definiton of message type)

Subject	Predicate	Object
ex:shape1	rdf:type	sh:property
ex:shape1	sh:path	quantity: Thermodynamic Temperature
ex:shape1	sh:minCount	„1“
ex:shape1	sh:maxCount	„1“

O (observation message)

Subject	Predicate	Object
ex:stream1#temp	unit:DegreeCelsius	„28“
ex:stream1#no2	eea:g-m3	„34“
ex:stream1#lat	geo:lat	„8.700066“
ex:stream1#long	geo:long	„49.30075“

schema knowledge

Subject	Predicate	Object
unit:DegreeCelsius	qudt:quantityKind	quantity:ThermodynamicTemperature

Semantic Validation – valid

$$D_{MT} \subseteq O$$

D_{MT} (definiton of message type)

Subject	Predicate	Object
ex:shape1	rdf:type	sh:property
ex:shape1	sh:path	quantity: Thermodynamic Temperature
ex:shape1	sh:minCount	„1“
ex:shape1	sh:maxCount	„1“

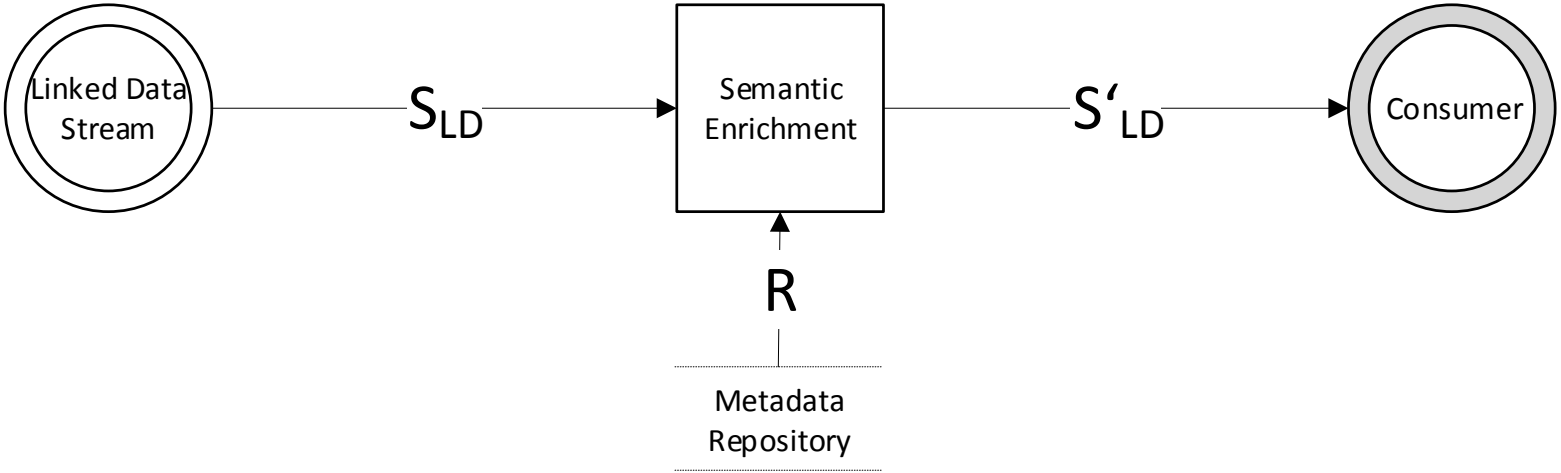
O (observation message)

Subject	Predicate	Object
ex:stream1#temp	unit:DegreeFahrenheit	„82“
ex:stream1#no2	eea:g-m3	„34“
ex:stream1#lat	geo:lat	„8.700066“
ex:stream1#long	geo:long	„49.30075“

schema knowledge

Subject	Predicate	Object
unit:DegreeFahrenheit	qudt:quantityKind	quantity:ThermodynamicTemperature

Semantic Enrichment



$$SemEnr(O, R) = O'$$

Semantic Enrichment

Observation

ID	Subject	Predicate	Object
$statement_1$	$observation_1$	hasValue	28.76324 *DegreeCelsius*
$statement_2$	$statement_1$	hasProvenance	$sensor_1$

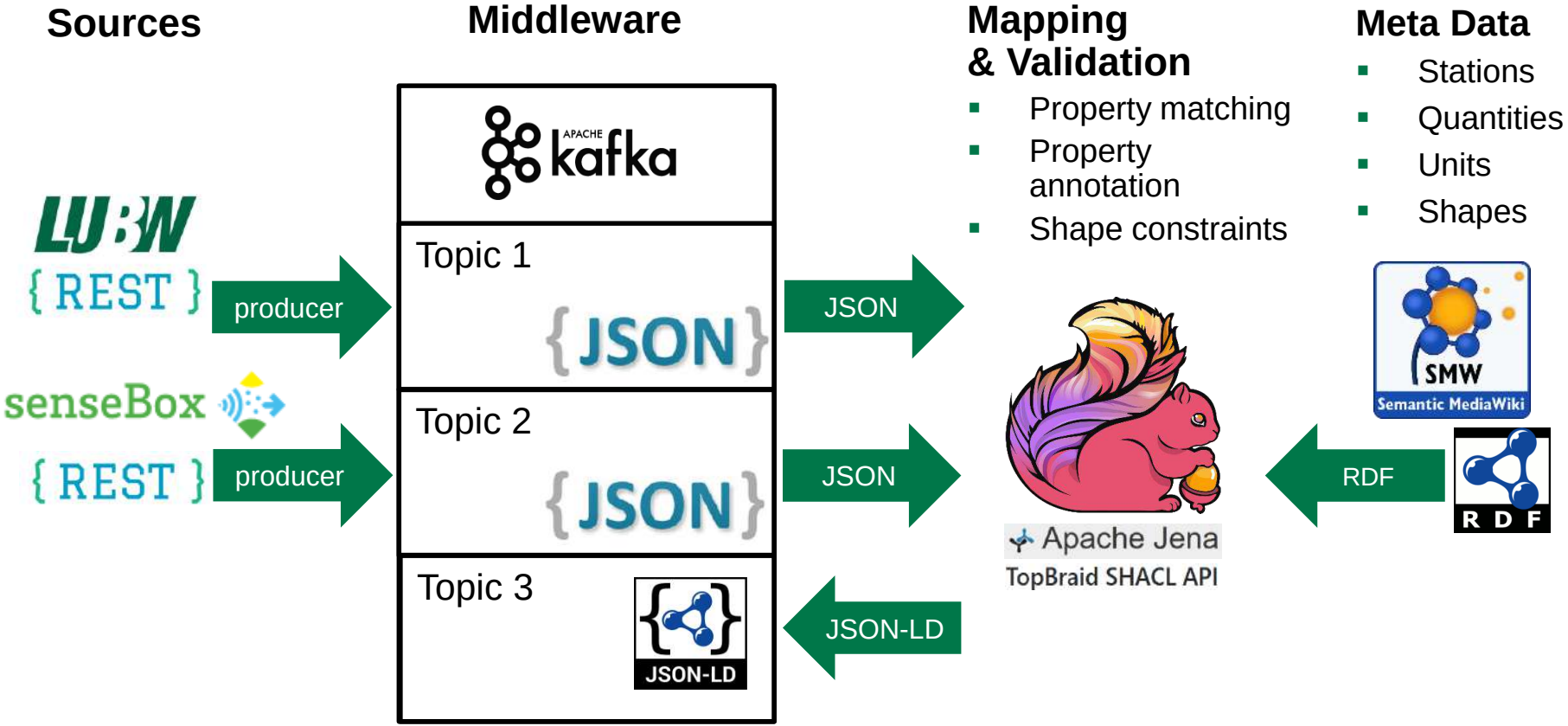
Rule

ID	Subject	Predicate	Object
$statement_3$	DegreeFahrenheit	isEqualTo	$DegreeCelsius * 1.80000 + 32.00000$

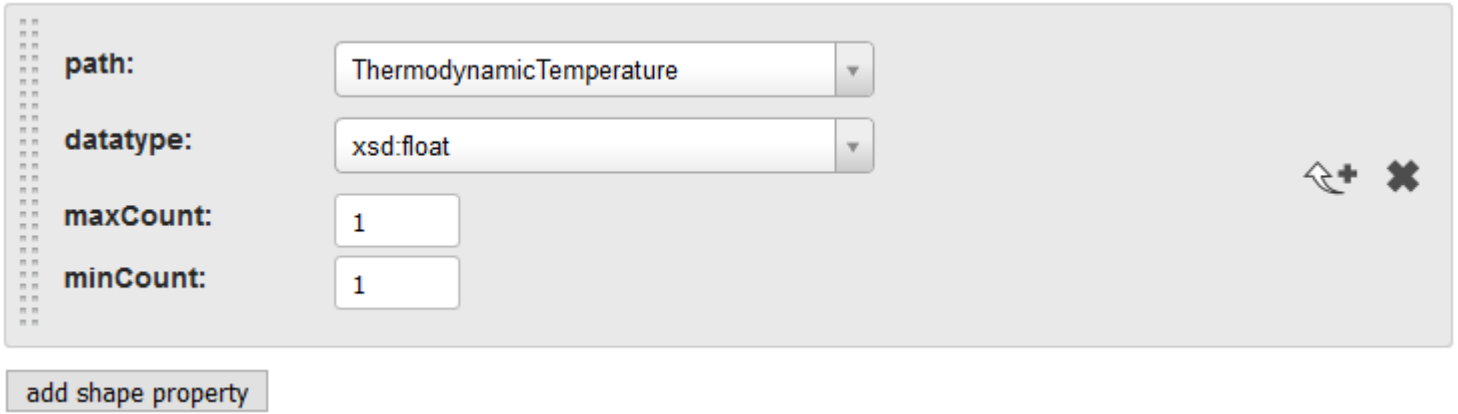
Derived

ID	Subject	Predicate	Object
$statement_1$	$observation_1$	hasValue	28.76324*DegreeCelsius*
$statement_4$	$observation_1$	hasValue	83.77383*DegreeFahrenheit*
$statement_2$	$statement_1$	hasProvenance	$sensor_1$
$statement_5$	$statement_4$	hasProvenance	$conclusion_1$
$statement_6$	$conclusion_1$	hasInput	$statement_1$
$statement_7$	$conclusion_1$	hasRule	$statement_3$

Implementation: Architecture



Implementation: Forms for Data Shapes



Shape	Component	Message
Property quantity:ThermodynamicTemperature: datatype=xsd:integer, minCount=1, maxCount=1, name=temparature	sh:minCount	Property needs to have at least 1 values, but found 0

LSAnE Consumer

- Start
- LUBW
- SenseBox
- Enriched Stream
- Valid Messages
- Invalid Messages

Enriched Stream


```
    {
      "@id": "https://amazonas.fzi.de/biggis/index.php/My.weather.lubw#pm10",
      "@type": "_:b0",
      "value": "0"
    },
    {
      "@id": "https://amazonas.fzi.de/biggis/index.php/My.weather.lubw#so2",
      "@type": "https://amazonas.fzi.de/biggis/index.php/G-m3",
      "value": "0"
    },
    {
      "@id": "https://amazonas.fzi.de/biggis/index.php/My.weather.lubw#station",
      "@type": "https://amazonas.fzi.de/biggis/index.php/Station",
      "value": "\"DEBY189\""
    },
    {
      "@id": "https://amazonas.fzi.de/biggis/index.php/My.weather.lubw#timestamp",
      "@type": "https://amazonas.fzi.de/biggis/index.php/UnixTimestamp",
      "value": "1525858195421"
    }
  ],
  "@context": {
    "value": {
      "@id": "http://www.w3.org/1999/02/22-rdf-syntax-ns#value"
    },
    "Has_subobject": {
      "@id": "https://amazonas.fzi.de/biggis/Property:Has_subobject",
      "@type": "@id"
```

Contribution

1. How can **heterogeneous** messages of environmental observations be **collaboratively** validated with uniform shapes?
   - Metadata management with Semantic MediaWiki
   - Semantic mappings on the fly
   - Schema inferencing and additional reasoning
2. How can we exploit these annotations to derive and apply rules for **semantic enrichment** to heterogeneous messages of environmental observations?
   - Rule-based creation of explicit data
   - Stream enrichment using SPARQL Rules

Conclusion

- LSane for collaborative definitions of
  - Semantic shapes of observation messages
  - Enrichment rules for observation messages
- Data stream functions for
  - Semantic Validation *SemVal(O, D_{MT})*
  - Semantic Enrichment *SemEnr(O, R)*
- Further research:
  - Applicability and scalability of a large amount of observation streams
  - Usability for domain experts in different domains

FZI

FZI FORSCHUNGSZENTRUM INFORMATIK

Feedback & Discussion

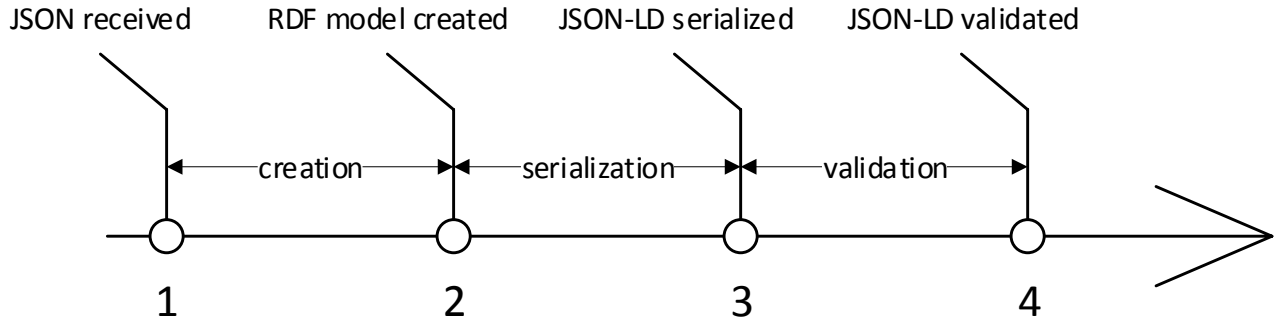
THANK YOU

FZI

FZI FORSCHUNGSZENTRUM INFORMATIK

BACKUP

Evaluation Results (ICWE 2018)



processing time in milliseconds

$f \approx 150 \frac{messages}{second}$	Arithmetic Mean	Median Value	Standard Deviation	Min Value	Max Value
Creation	1.460	0.341	3.699	0.196	47.750
Serialization	0.342	0.270	1.372	0.135	130.886
Validation	3.020	0.195	5.684	0.062	129.788
Total	4.822	0.979	6.813	0.468	276.622

Evaluation Data

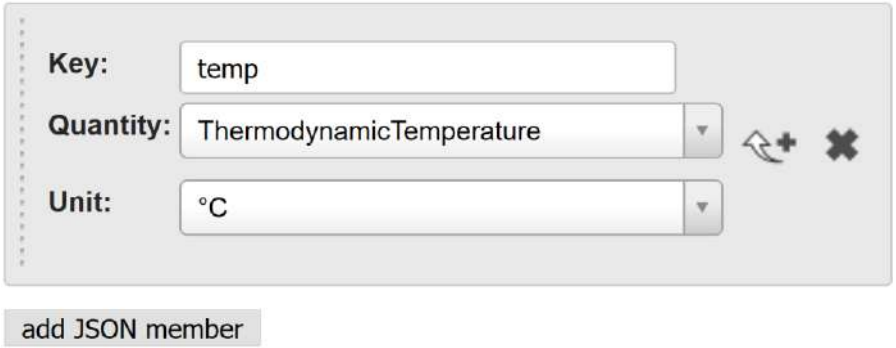
LUBW

http://lupo-messwerte.appspot.com/generic?table=bw_luft_stammdaten&limit=999

SenseBox

https://api.opensensemap.org/boxes/59ec966d49f6f80011c12395/

Implementation: Forms for Meta Data



My.weather.lubw

JSON property	Quantity	Unit
My.weather.lubw#longitude	Longitude	
My.weather.lubw#ozn	Ozn	G-m3
My.weather.lubw#height	Altitude	
My.weather.lubw#so2	So2	G-m3
My.weather.lubw#station	Station	
My.weather.lubw#luqx		
My.weather.lubw#pm10	Pm10	
My.weather.lubw#getRequestTimestamp		
My.weather.lubw#timestamp	Timestamp	
My.weather.lubw#latitude	Latitude	
My.weather.lubw#no2	No2	G-m3

Kategorie: Topic

Evaluation Meta Data

My.weather.lubw

JSON property	Quantity	Unit
My.weather.lubw#longitude	Longitude	
My.weather.lubw#ozn	Ozn	G-m3
My.weather.lubw#height	Altitude	
My.weather.lubw#so2	So2	G-m3
My.weather.lubw#station	Station	
My.weather.lubw#luqx		
My.weather.lubw#pm10	Pm10	
My.weather.lubw#getRequestTimestamp		
My.weather.lubw#timestamp	Timestamp	
My.weather.lubw#latitude	Latitude	
My.weather.lubw#no2	No2	G-m3

Kategorie: Topic

```
<swivt:Subject rdf:about="http://amazonas.fzi.de/biggis/index.php/Spezial:URI-Aufl%C3%B6ser/My.weather.lubw%23no2">
  <rdf:type rdf:resource="http://amazonas.fzi.de/biggis/index.php/Spezial:URI-Aufl%C3%B6ser/Category-3AJsonproperty"/>
  <swivt:masterPage rdf:resource="http://amazonas.fzi.de/biggis/index.php/Spezial:URI-Aufl%C3%B6ser/My.weather.lubw"/>
  <swivt:wikiNamespace rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">0</swivt:wikiNamespace>
  <property:Quant rdf:resource="http://amazonas.fzi.de/biggis/index.php/Spezial:URI-Aufl%C3%B6ser/No2"/>
  <property:Unit rdf:resource="http://amazonas.fzi.de/biggis/index.php/Spezial:URI-Aufl%C3%B6ser/G-2Dm3"/>
  <swivt:wikiPageSortKey rdf:datatype="http://www.w3.org/2001/XMLSchema#string">My.weather.lubw</swivt:wikiPageSortKey>
</swivt:Subject>
```

G-m3

- Titel (de): Mikrogramm pro Kubikmeter
- Einheitenzeichen: µg/m³
- Link: https://www.eea.europa.eu/themes/air/air-quality/resources/glossary/g-m3

Fakten zu „G-m3“

Einheitenzeichen	µg/m³
Titel-de	Mikrogramm pro Kubikmeter
URL	https://www.eea.europa.eu/themes/air/air-quality/resources/glossary/g-m3

LSAnE Consumer

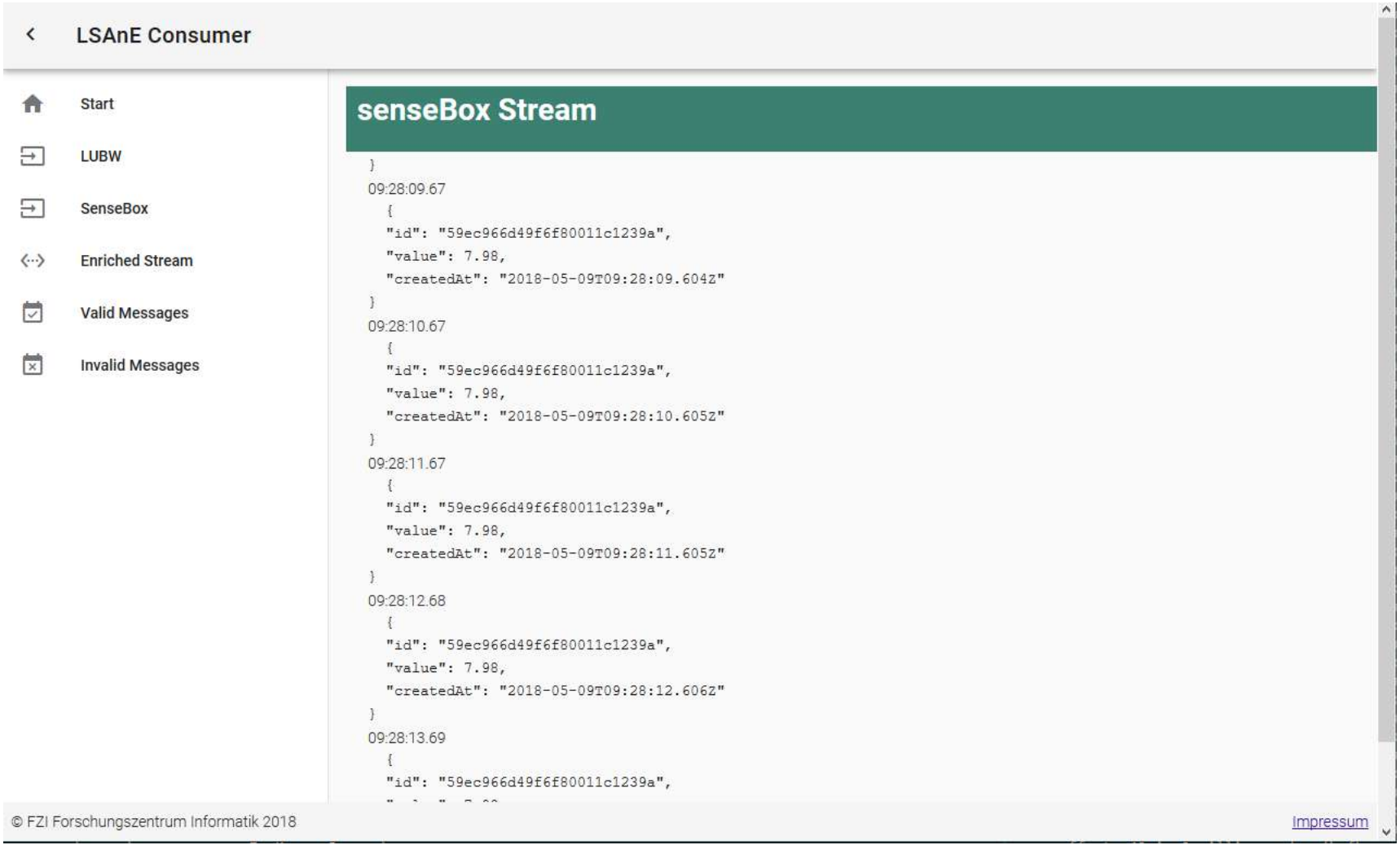
- Start
- LUBW
- SenseBox
- Enriched Stream
- Valid Messages
- Invalid Messages

LUBW Stream


```
  "timestamp": 1525857986963,
  "longitude": 11.46445
}
09:26:27.98
  {
  "no2": 61,
  "ozn": 10,
  "luqx": 0,
  "getRequestTimestamp": 1525857795155,
  "latitude": 48.18169,
  "heigth": 510,
  "so2": 0,
  "station": "DEBY189",
  "pm10": 0,
  "timestamp": 1525857987963,
  "longitude": 11.46445
}
09:26:28.98
  {
  "no2": 61,
  "ozn": 10,
  "luqx": 0,
  "getRequestTimestamp": 1525857795155,
  "latitude": 48.18169,
  "heigth": 510,
  "so2": 0,
  "station": "DEBY189",
  "pm10": 0,
```

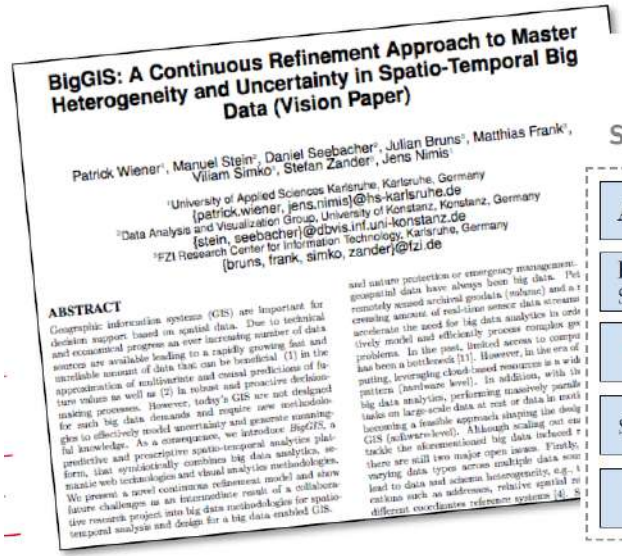

Impressum

FZI


LSAnE Consumer
Start
LUBW
SenseBox
Enriched Stream
Valid Messages
Invalid Messages
senseBox Stream
}
09:28:09.67
{
"id": "59ec966d49f6f80011c1239a",
"value": 7.98,
"createdAt": "2018-05-09T09:28:09.604Z"
}
09:28:10.67
{
"id": "59ec966d49f6f80011c1239a",
"value": 7.98,
"createdAt": "2018-05-09T09:28:10.605Z"
}
09:28:11.67
{
"id": "59ec966d49f6f80011c1239a",
"value": 7.98,
"createdAt": "2018-05-09T09:28:11.605Z"
}
09:28:12.68
{
"id": "59ec966d49f6f80011c1239a",
"value": 7.98,
"createdAt": "2018-05-09T09:28:12.606Z"
}
09:28:13.69
{
"id": "59ec966d49f6f80011c1239a",
© FZI Forschungszentrum Informatik 2018
Impressum

Preliminary Work: Paper ACM SIGSPATIAL 2016

Continuous Refinement Model



BigGIS: A Continuous Refinement Approach to Master Heterogeneity and Uncertainty in Spatio-Temporal Big Data (Vision Paper)

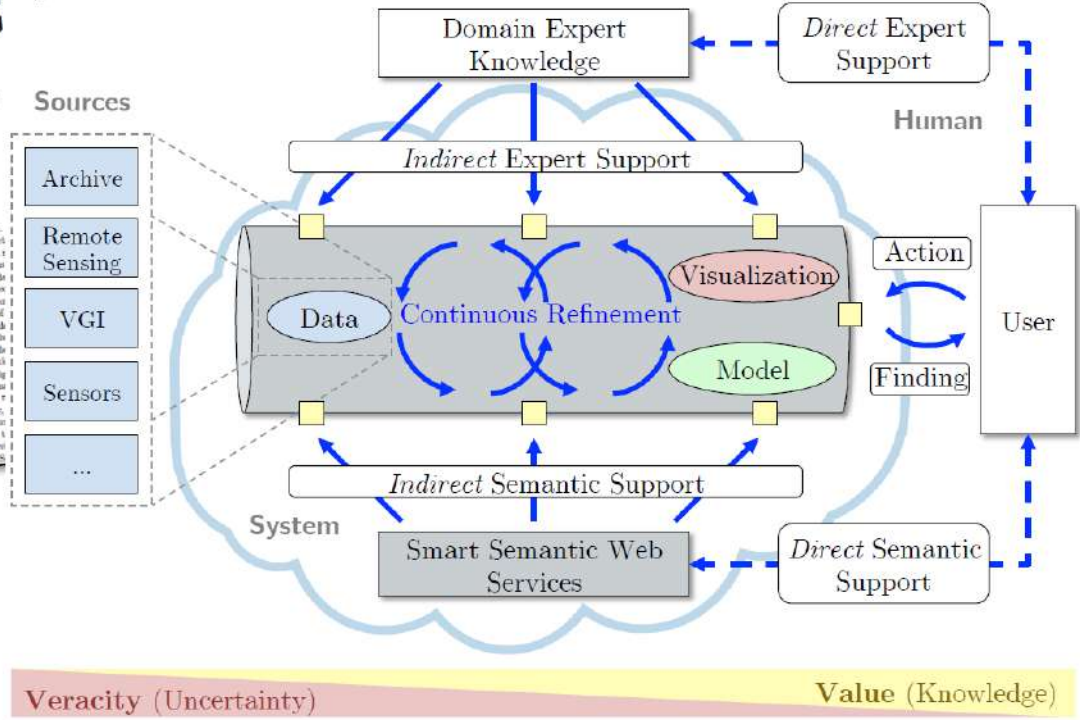
Patrick Wiener[1], Manuel Stein[2], Daniel Seebacher[2], Julian Bruns[3], Matthias Frank[3], Viliam Simko[3], Stefan Zander[1], Jens Nimis[1]

[1]University of Applied Sciences Karlsruhe, Karlsruhe, Germany
{patrick.wiener, jens.nimis}@hs-karlsruhe.de
[2]Data Analysis and Visualization Group, University of Konstanz, Konstanz, Germany
{stein, seebacher}@dbvis.inf.uni-konstanz.de
[3]FZI Research Center for Information Technology, Karlsruhe, Germany
{bruns, frank, simko, zander}@fzi.de

ABSTRACT

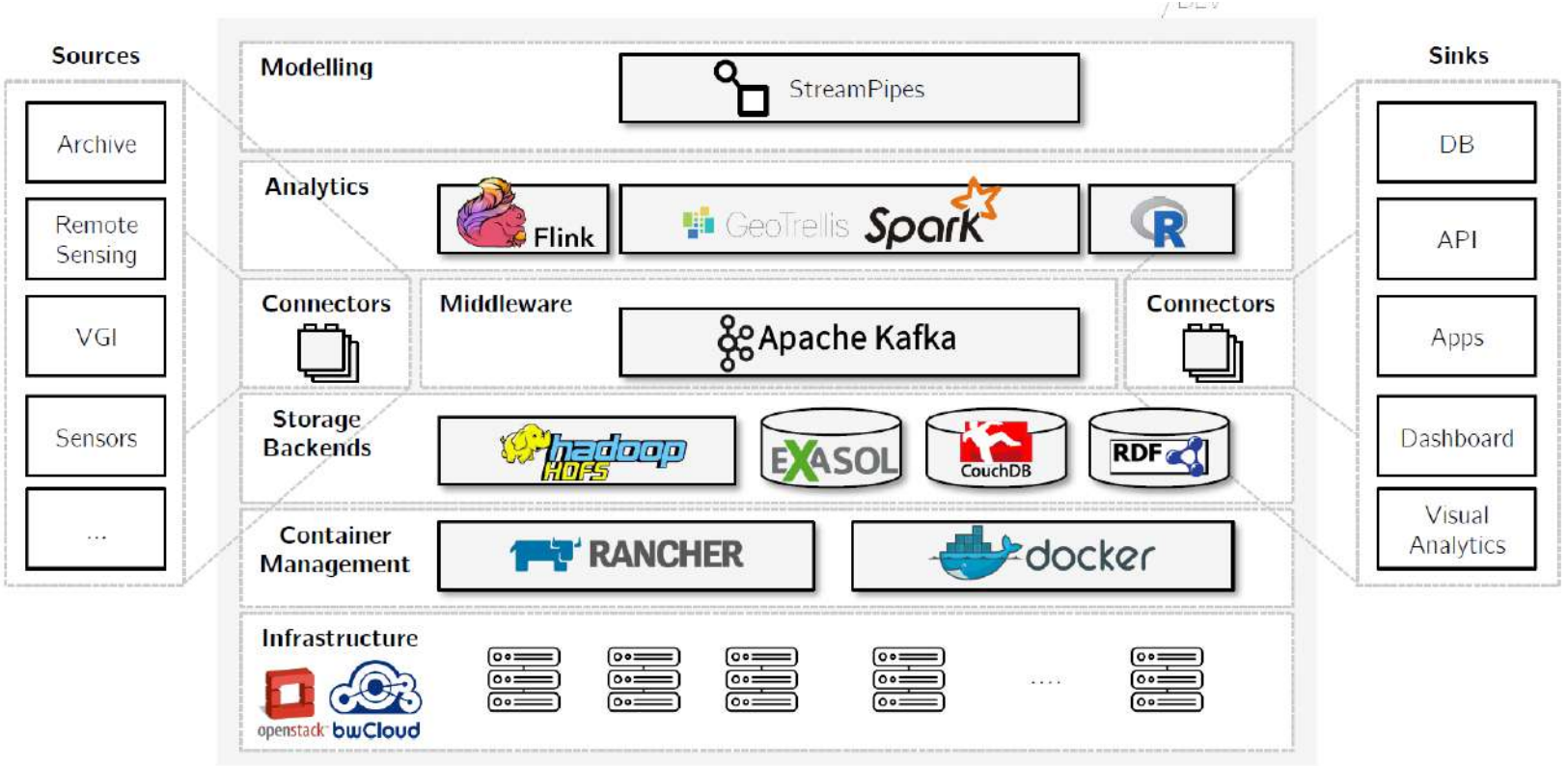
Geographic information systems (GIS) are important for decision support based on spatial data. Due to technical and economical progress an ever increasing number of data sources are available leading to a rapidly growing fast and unreliable amount of multivariate and causal predictions of future values as well as (2) in robust and proactive decision-making processes. However, today's GIS are not designed for such big data demands and require new methodologies to effectively model uncertainty and generate meaningful knowledge. As a consequence, we introduce *BigGIS*, a predictive and prescriptive spatio-temporal analytics platform, that symbiotically combines big data analytics, semantic web technologies and visual analytics methodologies. We present a novel continuous refinement model and show future challenges as an intermediate result of a collaborative research project into big data methodologies for spatio-temporal analysis and design for a big data enabled GIS.

① Integrated Analytical Pipeline
② Smart Semantic Web Services
③ Domain Expert Knowledge Extraction and Generation
④ Modelling Uncertainty



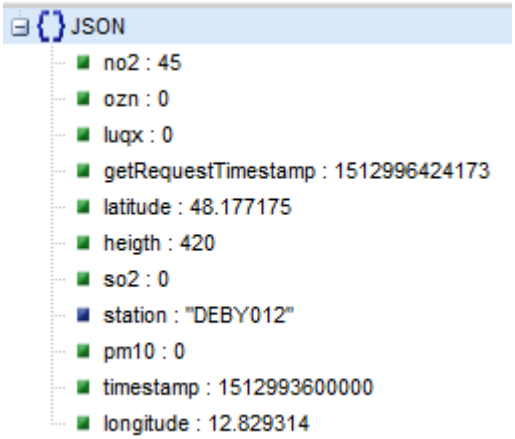
Preliminary Work: BigGIS Architecture

Spatio-Temporal Analytics at Scale



Input

Raw JSON Data



Meta Data (SMW)

My.weather.lubw

JSON property	Quantity	Unit
My.weather.lubw#longitude	Longitude	
My.weather.lubw#ozn	Ozn	G-m3
My.weather.lubw#height	Altitude	
My.weather.lubw#so2	So2	G-m3
My.weather.lubw#station	Station	
My.weather.lubw#luqx		
My.weather.lubw#pm10	Pm10	
My.weather.lubw#getRequestTimestamp		
My.weather.lubw#timestamp	Timestamp	
My.weather.lubw#latitude	Latitude	
My.weather.lubw#no2	No2	G-m3

Kategorie: Topic

Future Work:

- Suggest meta data of properties by analyzing JSON data
- User study to justify user experience
- Use cases for the produced JSON-LD stream