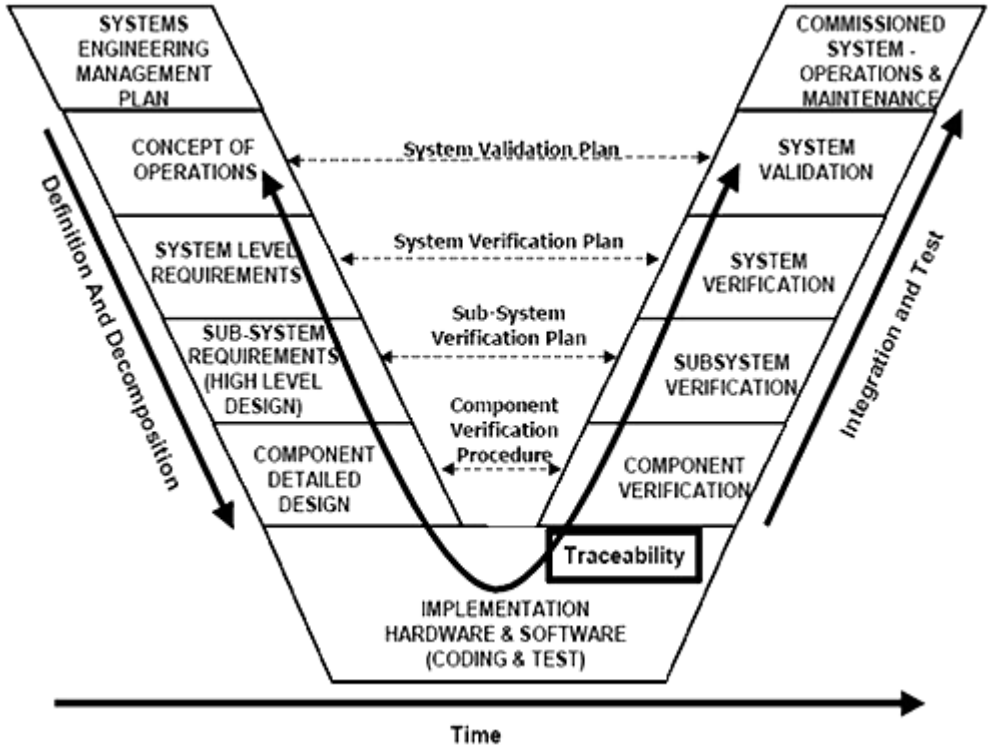
Information Integration using a Linked Data approach

OSLC Tool-chain architecture for data-intensive computations

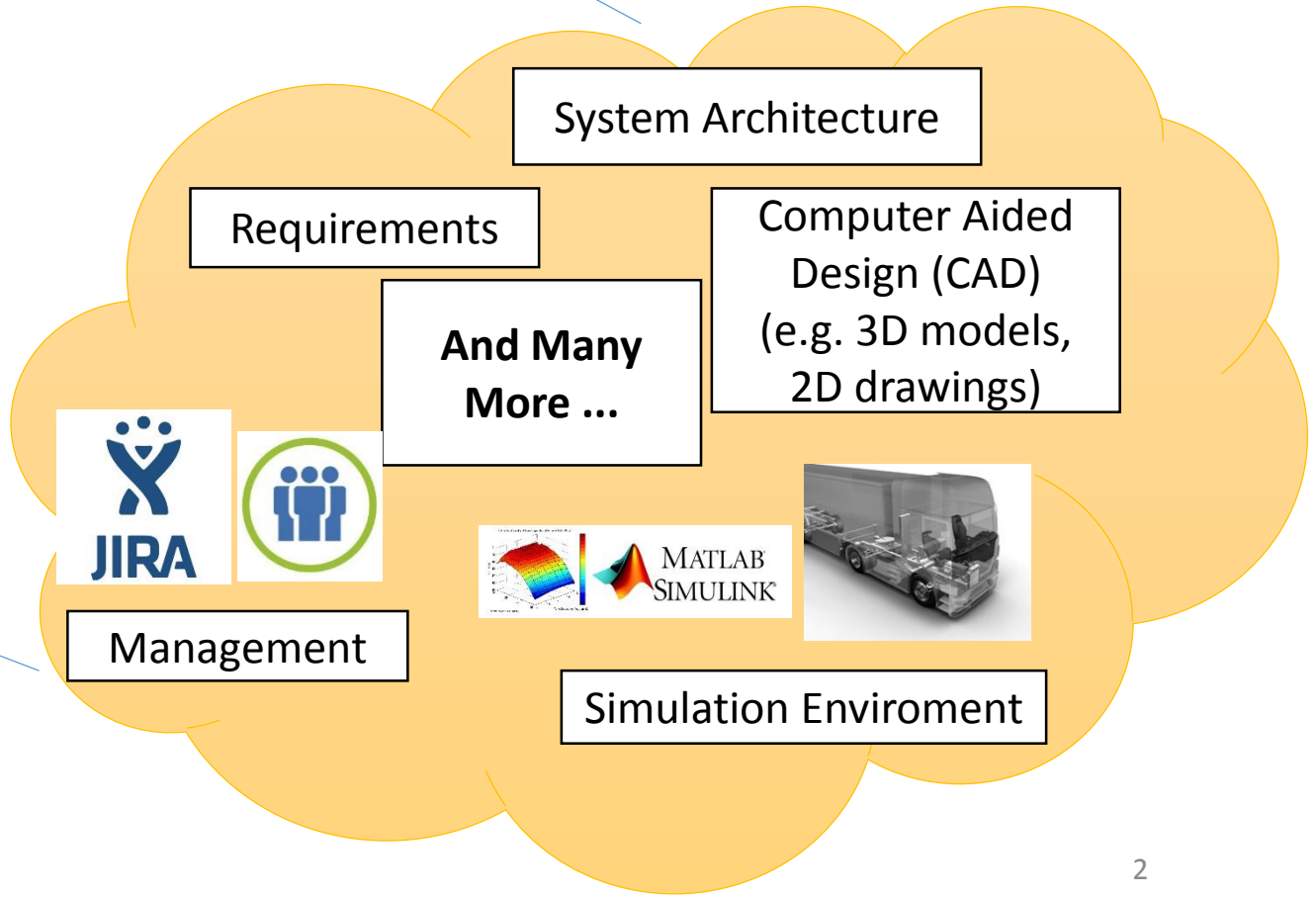
Jawad Munir
KTH, Royal Institute of Technology
Scania AB

Models-based Systems Engineering Development

ISO/IEC 15288 SE Process V-Model

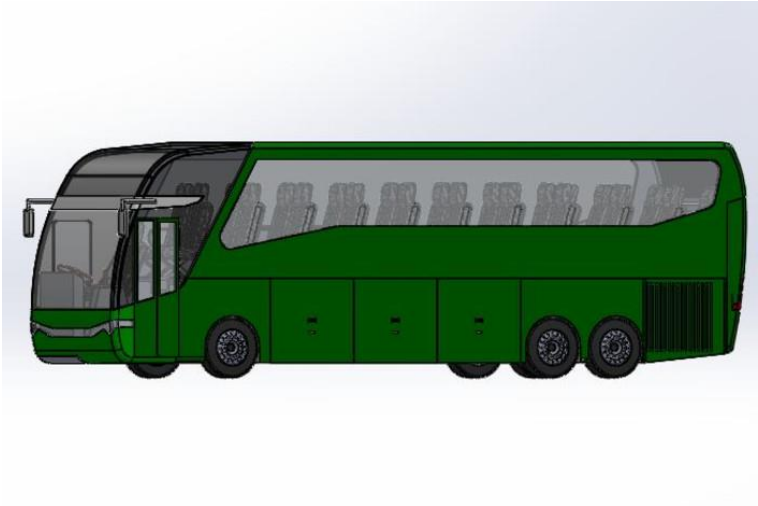


Many Relationships & Overlaps between these Models



Problem: Rollover Risk of Buses

- ECE R66.02 Valid from 2017, regulates the strength of the assembley in buses and the requirement of the roll-over test.



The strength of bus Superstructure

<<Requirement>>

Roof Crush Resistance

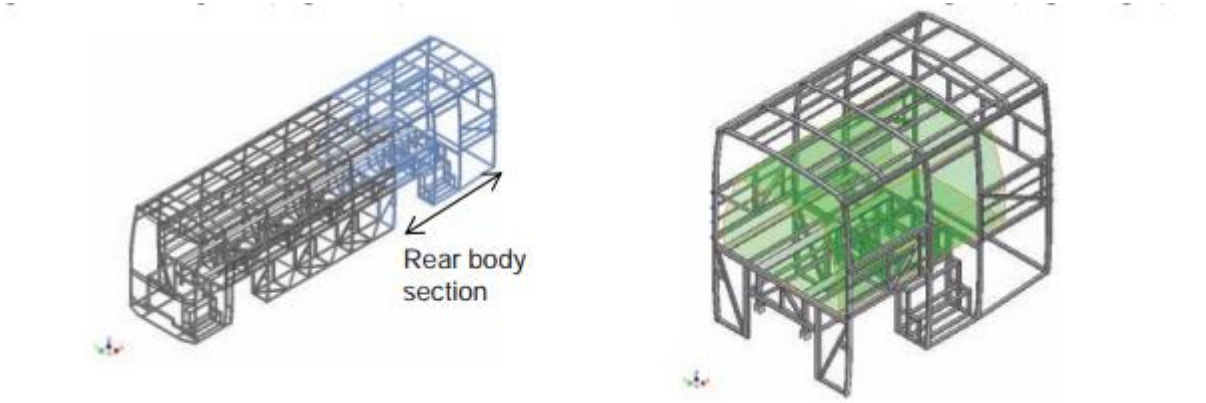
Bus superstructure section consist of four bays including rear entrance door, emergency door and rear end. Residual Space envelope requirement, in case of structural deformation in rollover test: **residual space need to be 38.0 mm**

System Egineer define the test

<<verify>>

<<Test Case>>

Accuracy test for bus superstructure model, quasi static load test



The bus superstructure (left) & CAD model of rear bus body section with 3D residual space envelop (right). The Models developed using CATIA software

Mechanical Engineer designs the bus geometrical model

Dynamic complete computer rollover simulation test for bus body

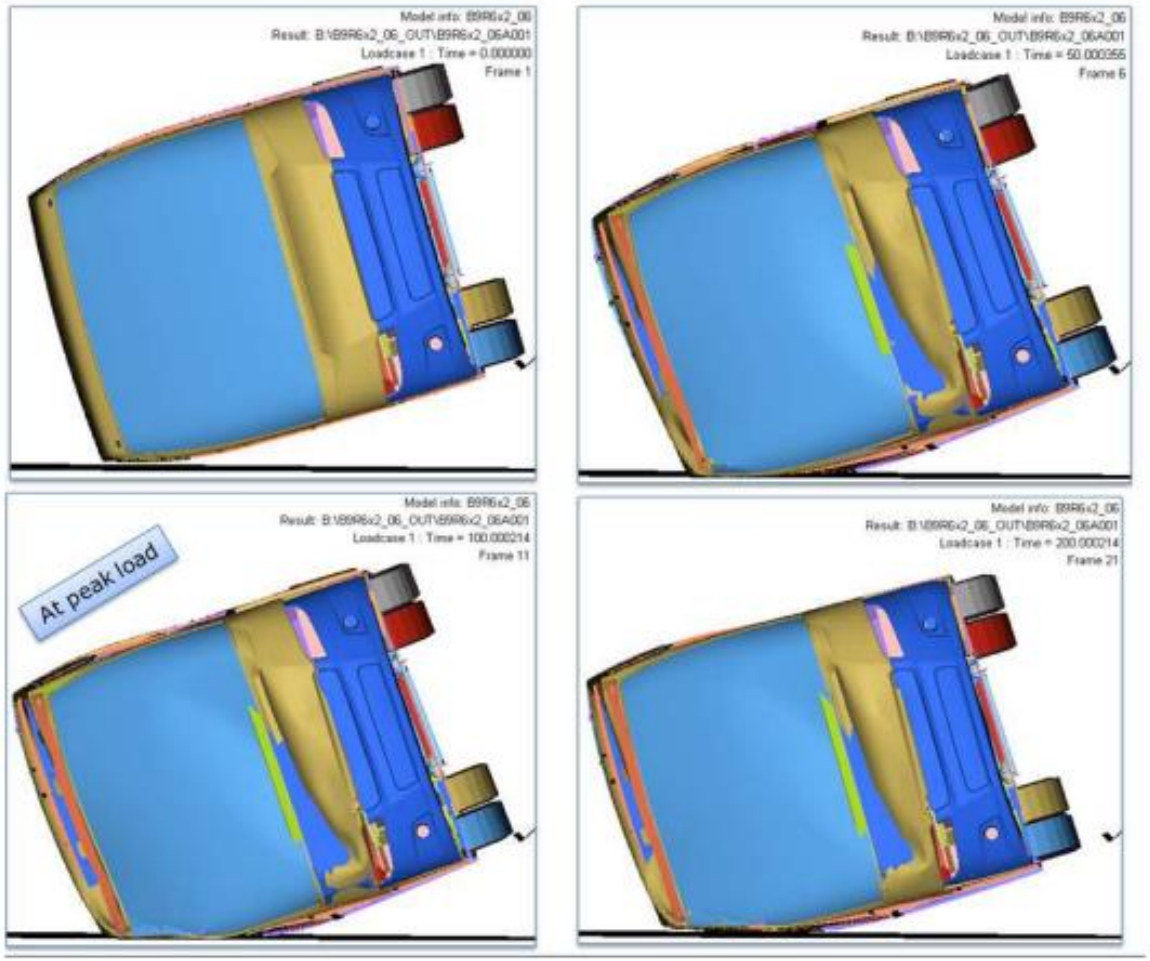
<<requirement>>
Vehicle in motion rollover test

The superstructure of the vehicle shall be of sufficient Strong so that in the event of rollover it doesn't intrude inside the Residual space, for it, dynamic simulation should verify it at speed of 40km/h

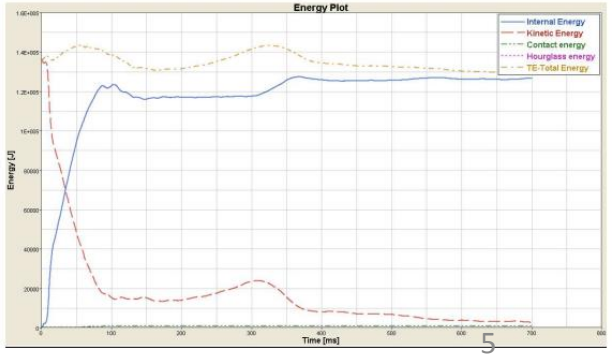
System engineer defines simulation testCase

<<verify>>

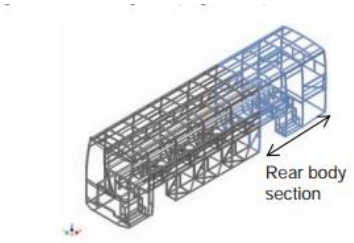
<<testCase>>
Vehicle in motion rollover simulation

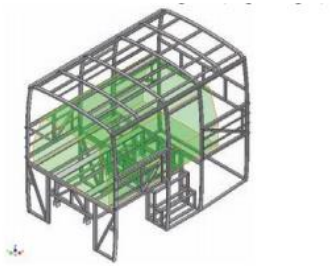


Mechanical Engineer performs simulation with dynamic system model



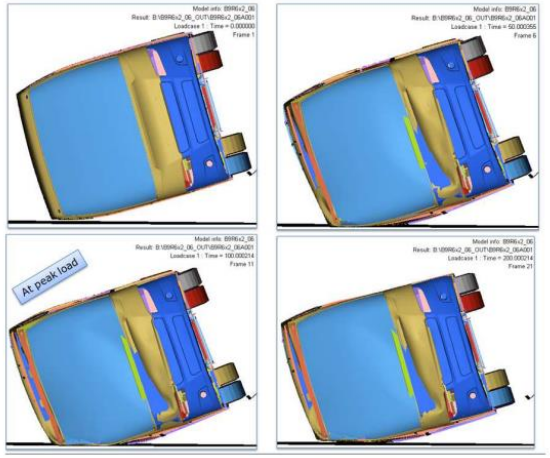
Link between Residual Space parameter of Geometric Model and Simulation Model





Residual Space

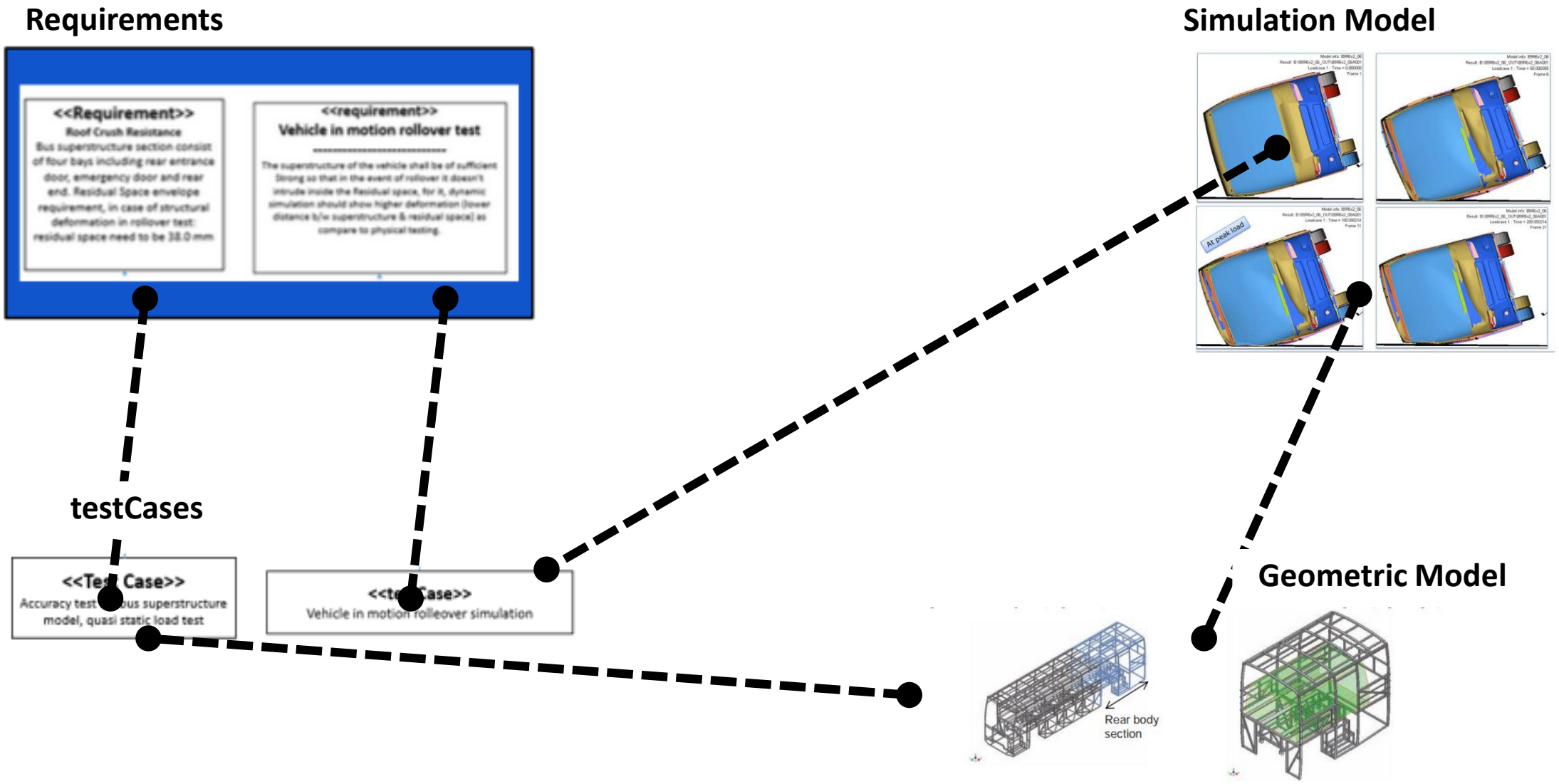
Computer simulation: inertia



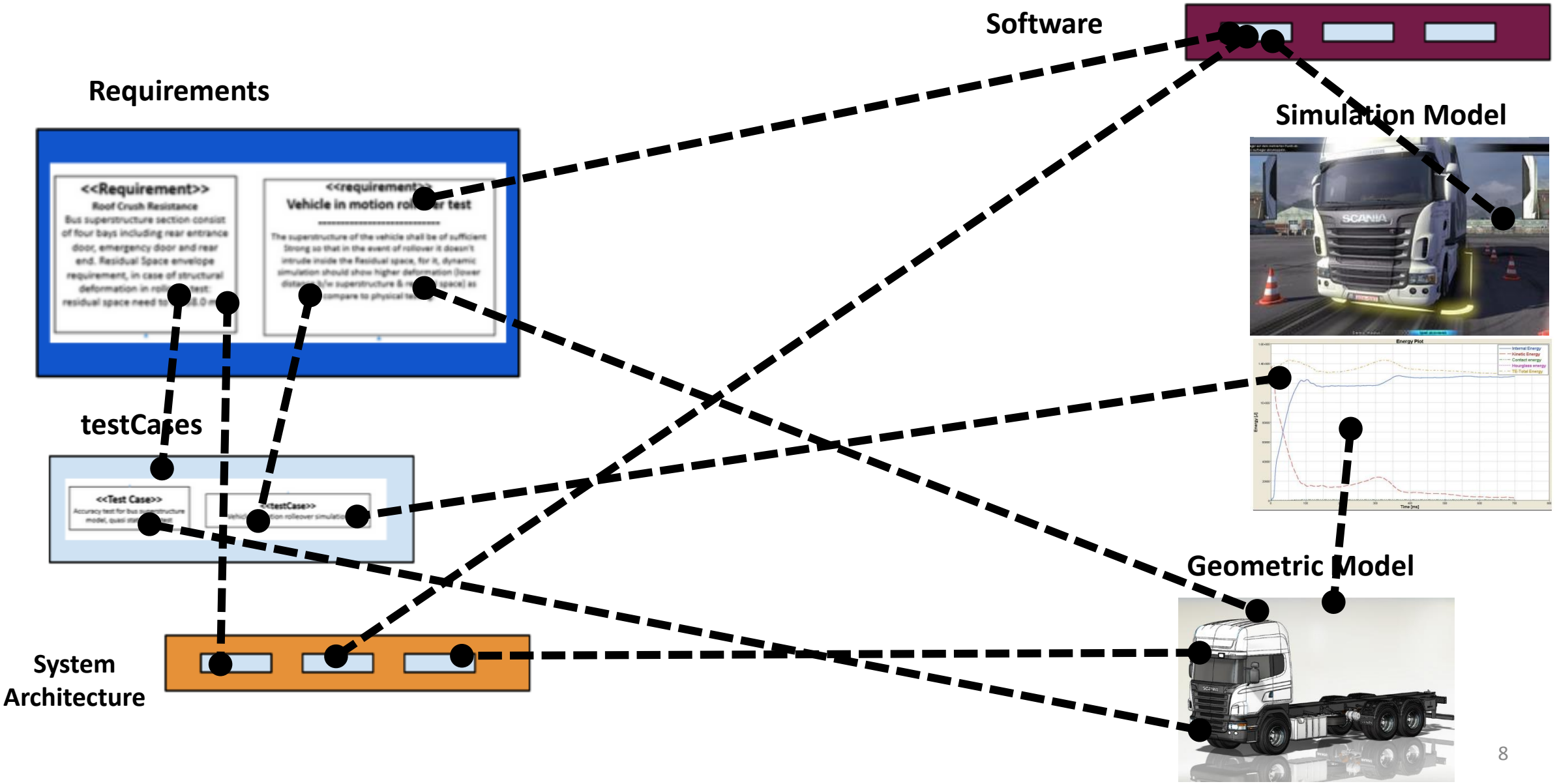
Residual Space in Geometric Model

Residual Space in simulation model

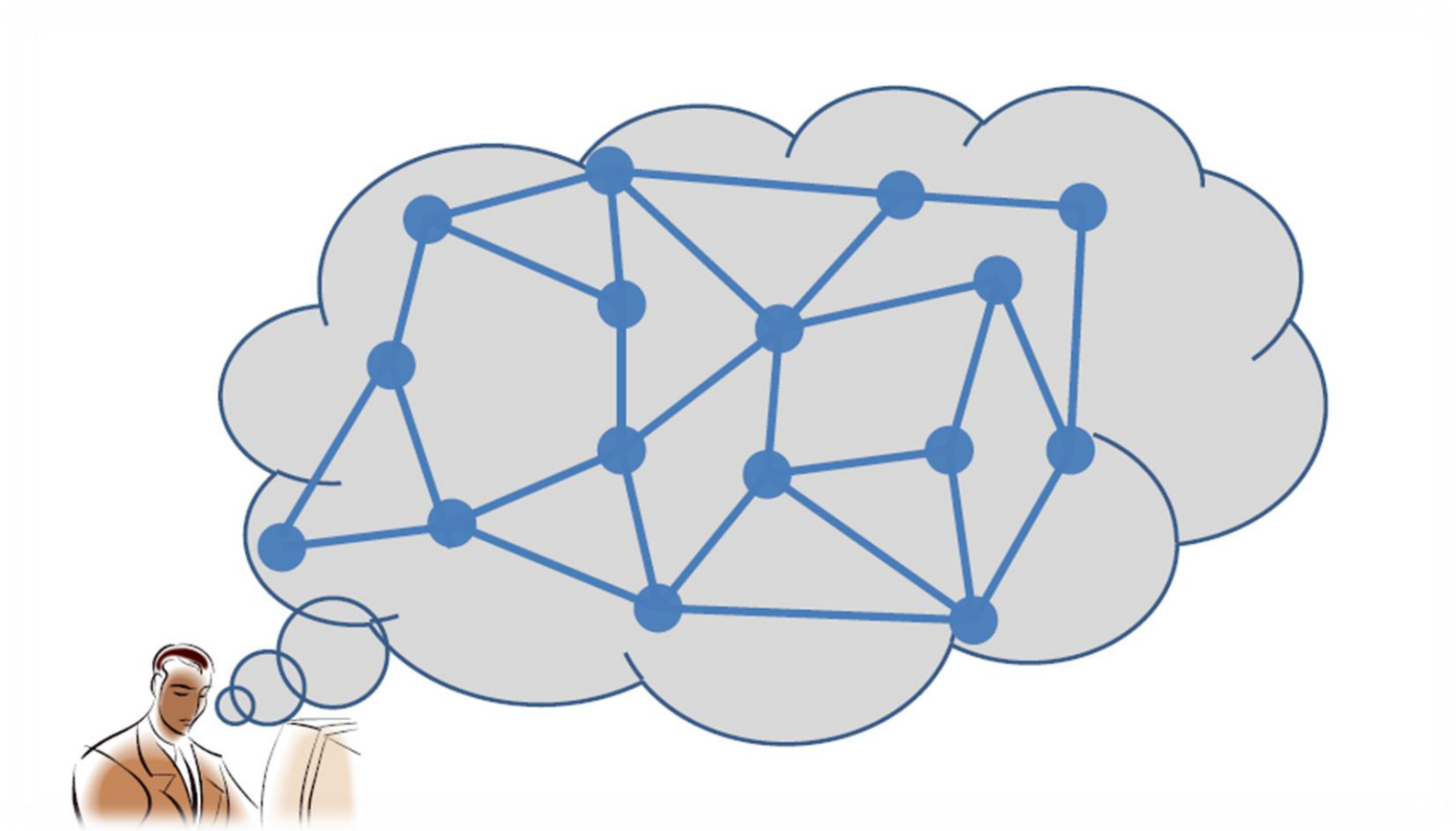
Relationships between Engineering Data



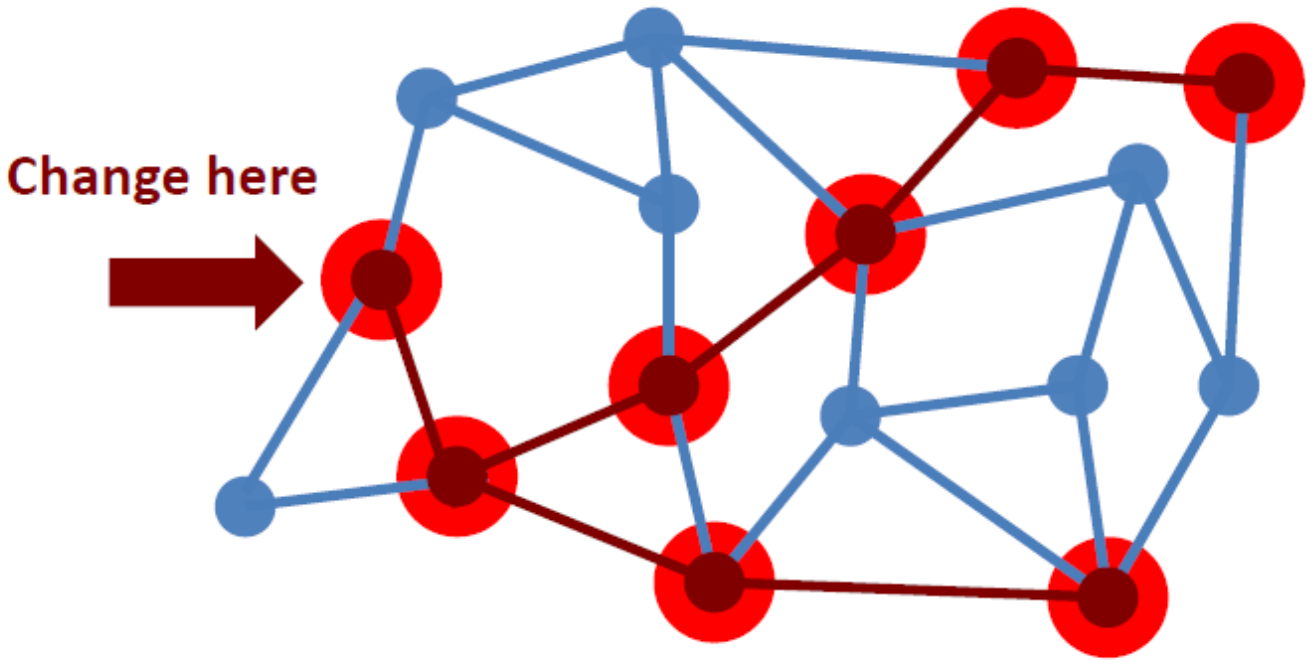
Reality in Complex System Design: (Too Many) Relationships between Engineering Data



Network of Relationships

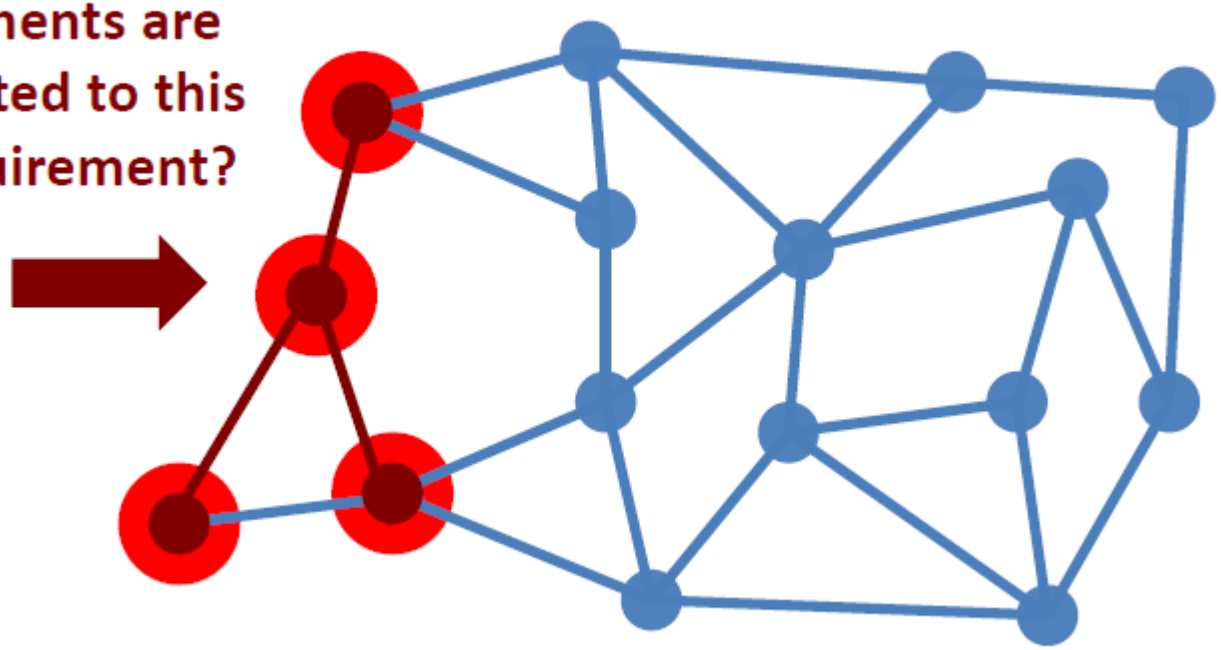


Impact Analysis

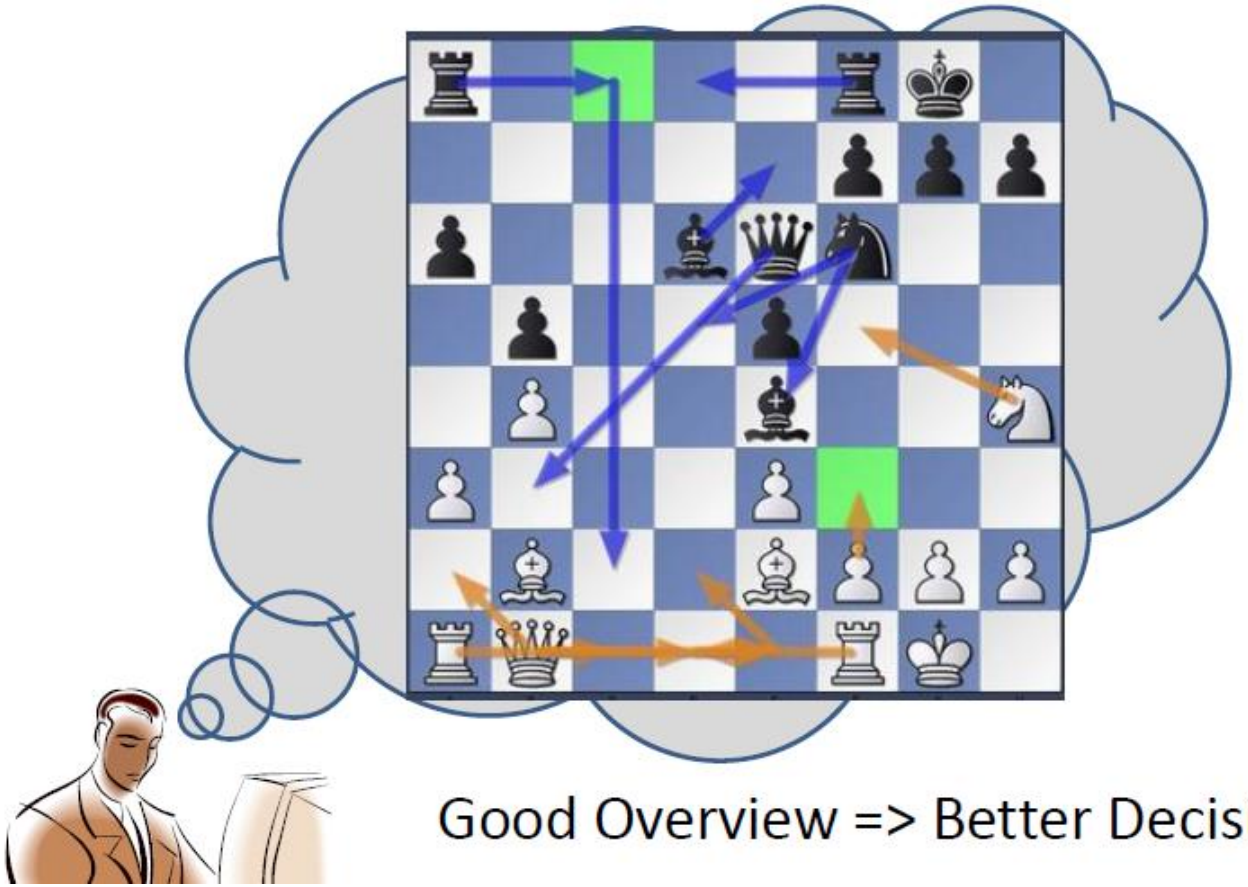


Queries

Query: Which elements are related to this requirement?



Overview of Development Process



Good Overview => Better Decisions

Scania – A case study

- 40000 employee
- Company in VM Group (Heavy Vehicles)

- In-house development
- Number of tools (in-house + commercial): 1000+

- The product: a truck in many versions and configurations
- Continous Development and integration

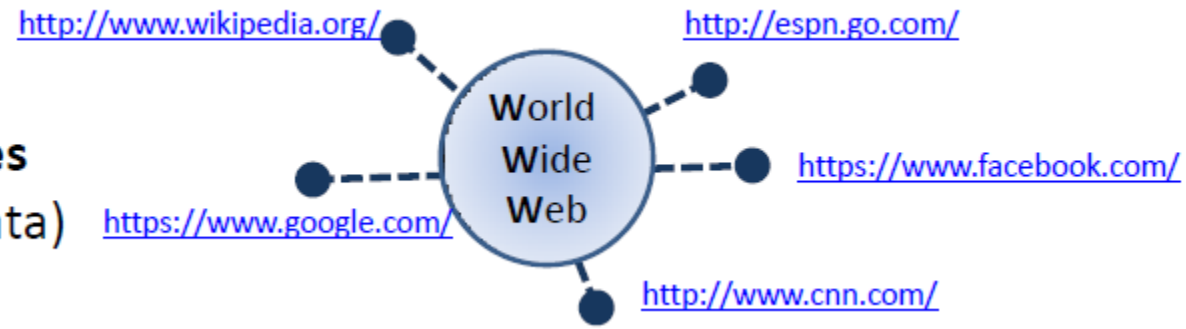
Tradional Tool and Data Integration

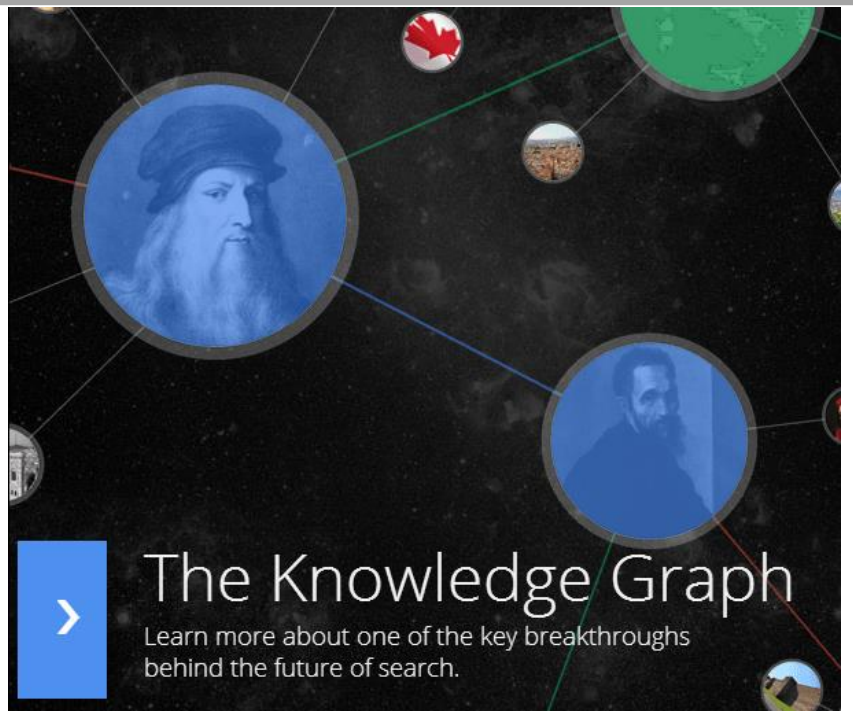
- Point-to-Point connections
  - Don’t scale
  - limited coverage
- Tight Coupling
  - Dependence on internal structures and meanings
  - Inability to identify crucial entities (components, locations etc.)
- Monocultures
  - Vendor lock-in
  - Processes and methods needs to follow the tool
  - Incomplete w.r.t domains
  - What to do with in-house tools? – results in Data duplication

- **There is no (can not be a) single Enterprise Information Model**

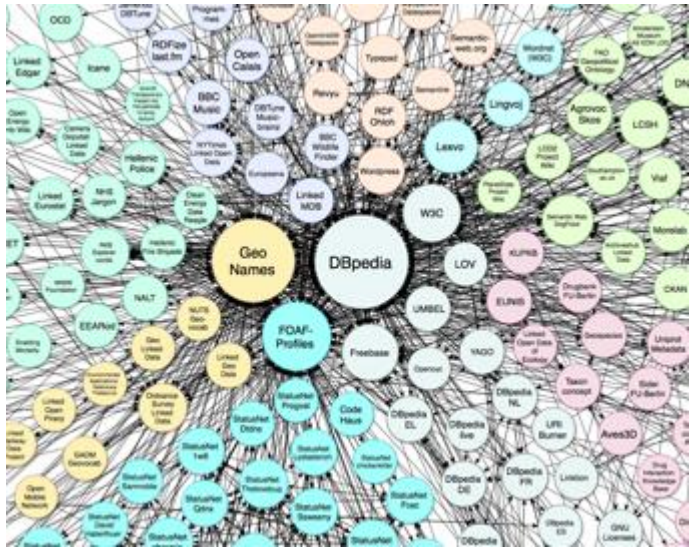
Relationships on the Web

Linked Web Pages
(Unstructured Data)

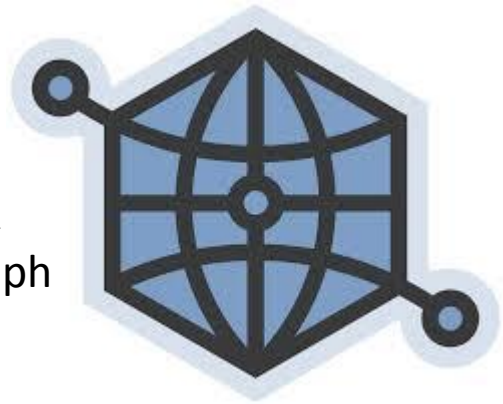




Linked Data
(Structured Data)



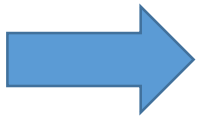
Facebook
Open Graph

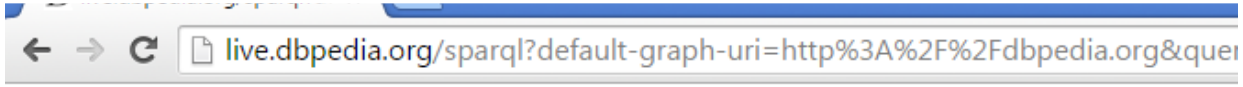


Example Linked Data: Dbpedia Query

Dbpedia Query: Give me name and date of establishment of universities in sweden?

```
SELECT DISTINCT ?name ?established
WHERE {
  [] dbpprop:country dbpedia:Sweden;
     rdf:type dbpedia-owl:University;
     dbpprop:nativeName ?name;
     dbpprop:established ?established.
}
ORDER BY ?established
LIMIT 50
```





name	established
"Kungliga Tekniska högskolan"@en	1827
"Chalmers tekniska högskola"@en	1829
"Konstfack"@en	1844
"Göteborgs universitet"@en	1954
"Lunds Tekniska Högskola"@en	1961
"Ekonomihögskolan i Lund"@en	1961
"Ekonomihögskolan i Lund"^^<http://www.w3.org/2001/XMLSchema#string>	1961
"Linköpings universitet"^^<http://www.w3.org/2001/XMLSchema#string>	1969
"Linköpings universitet"@en	1969
"Linköpings Tekniska Högskola"^^<http://www.w3.org/2001/XMLSchema#string>	1970
"Tekniska högskolan vid Linköpings universitet"@en	1970
"Linköpings Tekniska Högskola"@en	1970
"Högskolan i Jönköping"^^<http://www.w3.org/2001/XMLSchema#string>	1977
"Högskolan i Borås"^^<http://www.w3.org/2001/XMLSchema#string>	1977
"Högskolan i Jönköping"@en	1977
"Högskolan i Borås"@en	1977
"Blekinge Tekniska Högskola"@en	1989
"Tekniska Högskolan i Jönköping"^^<http://www.w3.org/2001/XMLSchema#string>	1994
"Tekniska Högskolan i Jönköping"@en	1994
"Centrum för Mellanösternstudier vid Lunds Universitet"@en	2007
"Linnéuniversitetet"@en	2010
"Luleå tekniska universitet"^^<http://www.w3.org/2001/XMLSchema#string>	19711997
"Luleå tekniska universitet"@en	19711997

Tim Berners-Lee, Linked Data principles

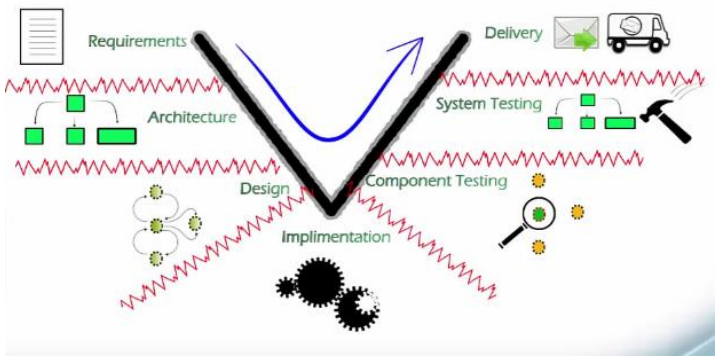
- **Tim berners-lee, Linked Data Principles**

1. Use **URIs** to identify the “things” in your data
2. Use **http:// URIs** so people (and machines) can look them up on the web
3. When a URI is **looked up**, **return** a **description** of the thing (in RDF format)
4. Include **links to related things**

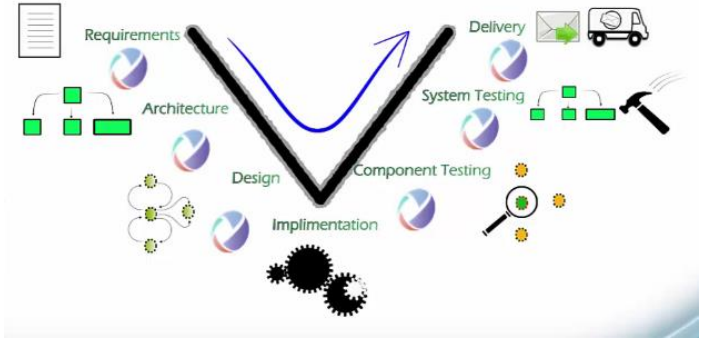
Open Services for Lifecycle Collaboration

- OSLC = Reusing the Web for tool integration
- Based on Web standards **(Linked Data and RESTful Web Services)**
- Initiated by IBM, It is **OPEN-initiative**
- Adopted by many tool vendors like Ericsson, IBM, Airbus etc ...
- Many tools have OSLC support like Jira, Jenkins, IBM rational tools etc ...
- Managed by OASIS

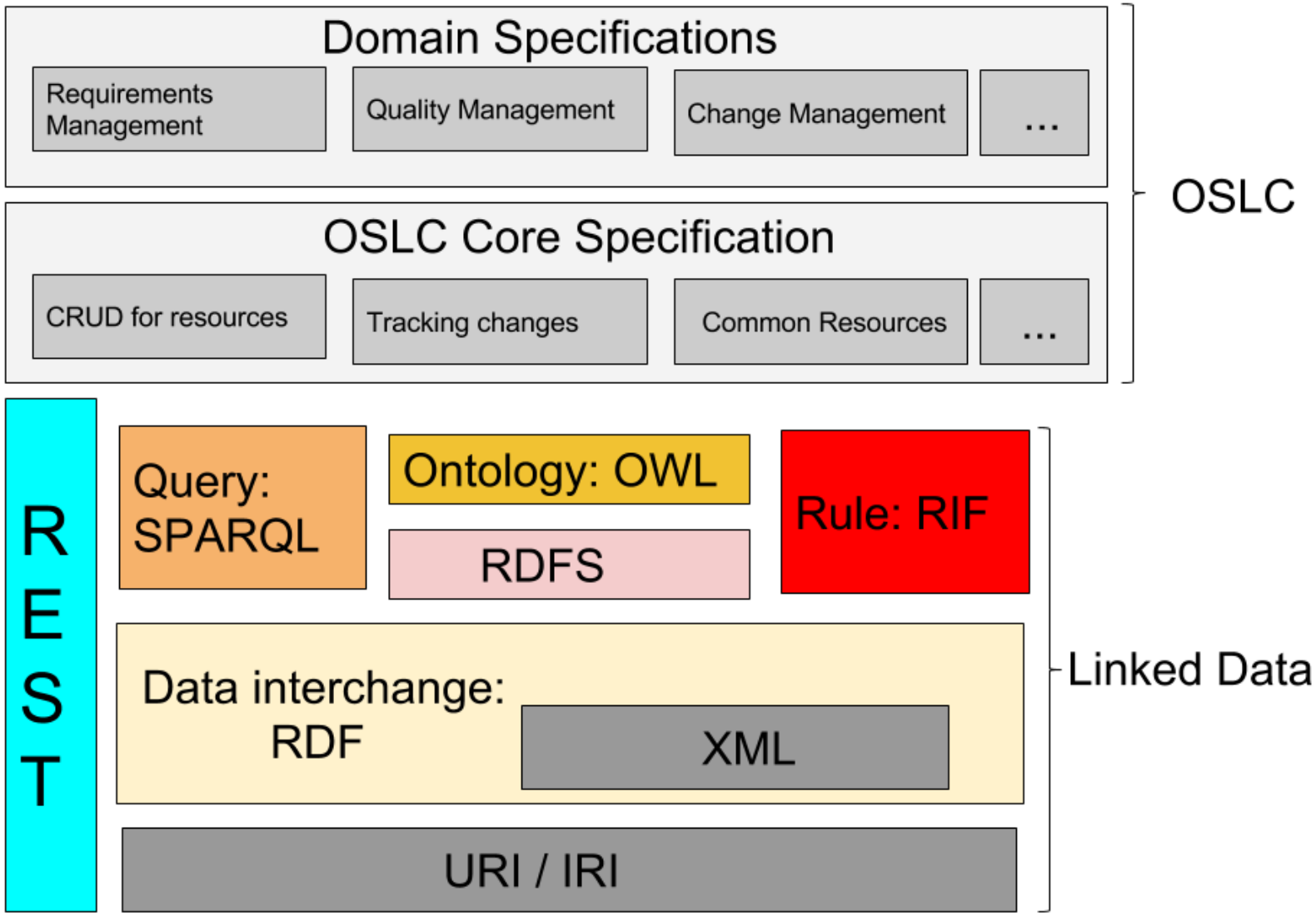
ISO SE V-Model



OSLC Tool Integration



The OSLC Technology Stack

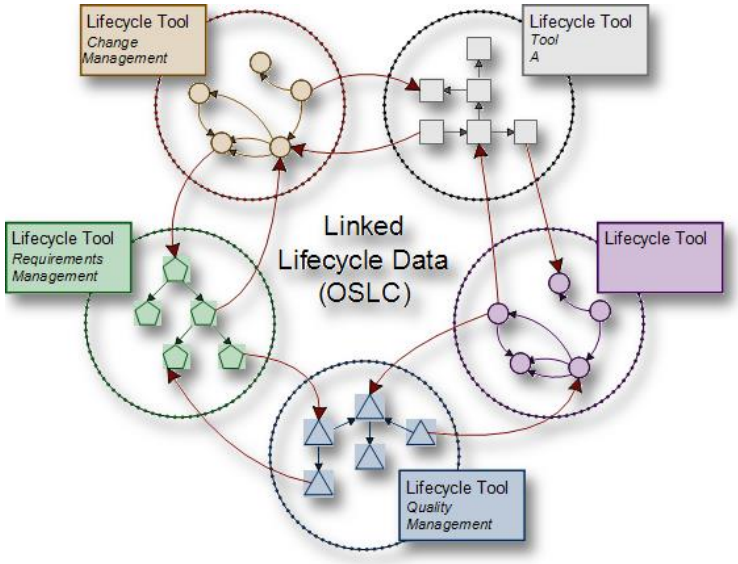


How OSLC works?

- OSLC defines set of Specifications and best practises for using linked data approach for tool integration and interoperability.

Operations on OSLC Resources

- **Publishing** OSLC resources
- **Retrieving** OSLC resources
- **Linking** OSLC resources across tools
- **Tracking changes** to OSLC resources



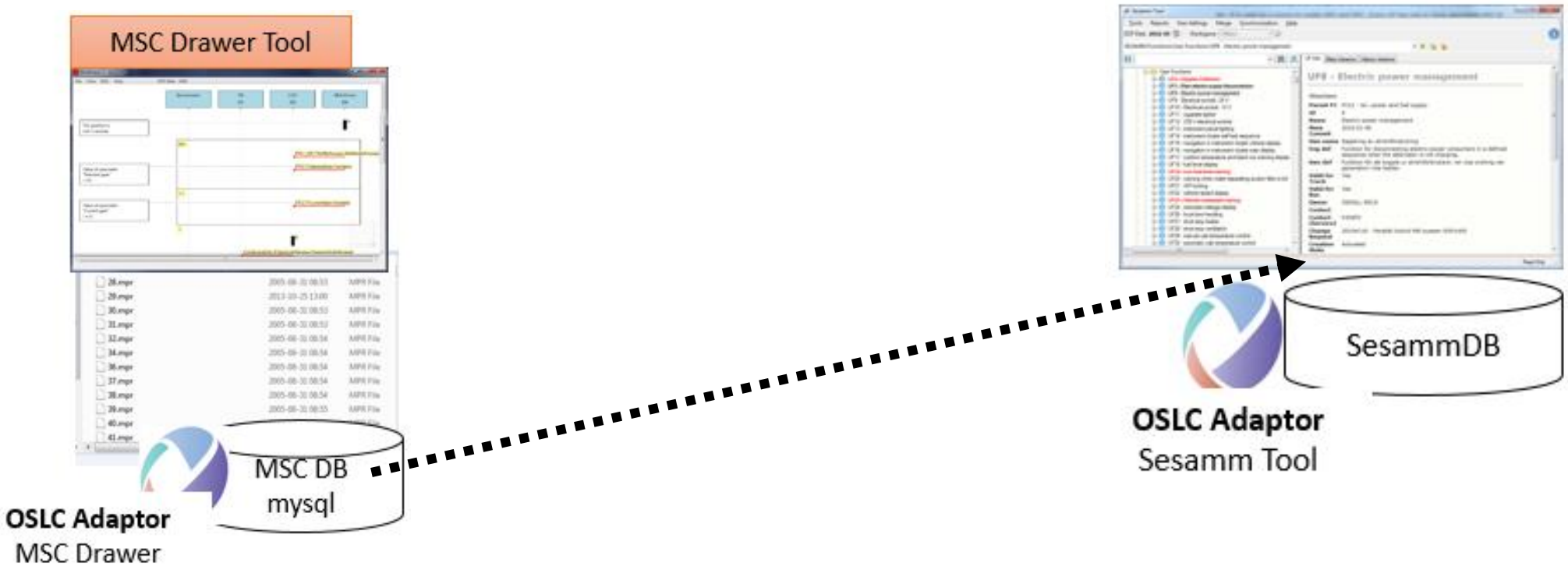
Thesis Project

- Assumption – **Building an OSLC Tool-chain data integration architecture for performing complicated cross domain queries.**

- Goal
  - Identify Integration Scenario for tool integration
  - Develop a prototype to support that Integration Scenario (Proof of Concept)
  - Test the prototype with cross domain queries

Integration Scenario – Scope + Tools involved

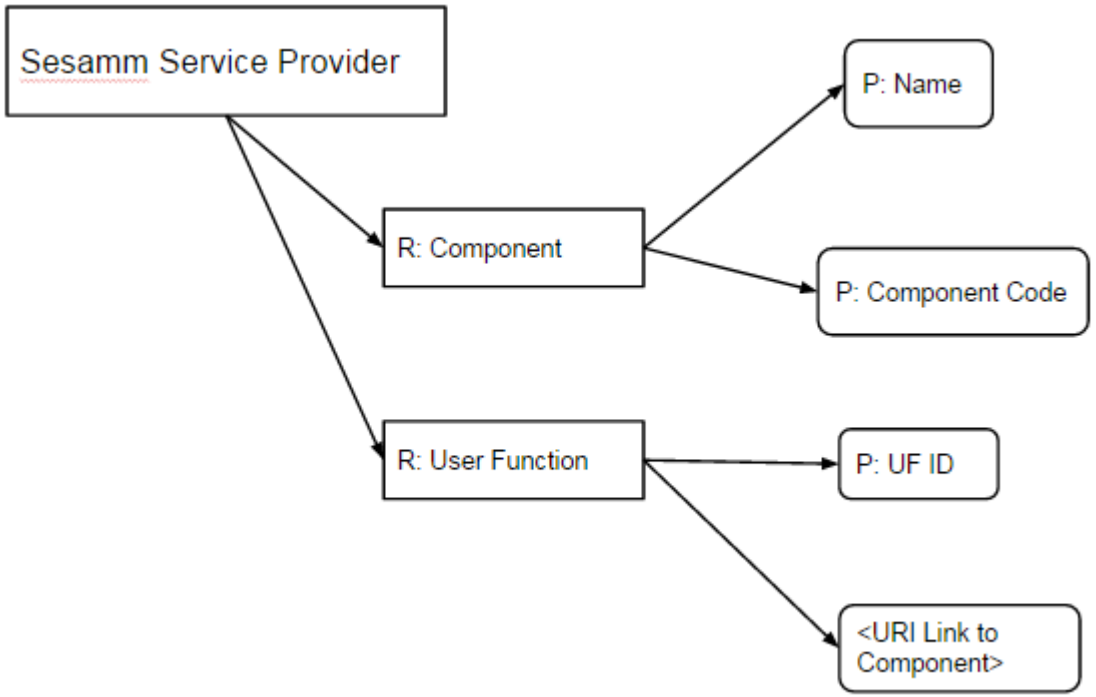
- We want to link **MSC Drawer** with **Sesamm Tool**
- Both tools have been developed by RESA



OSLC Adaptor for Sesamm Tool

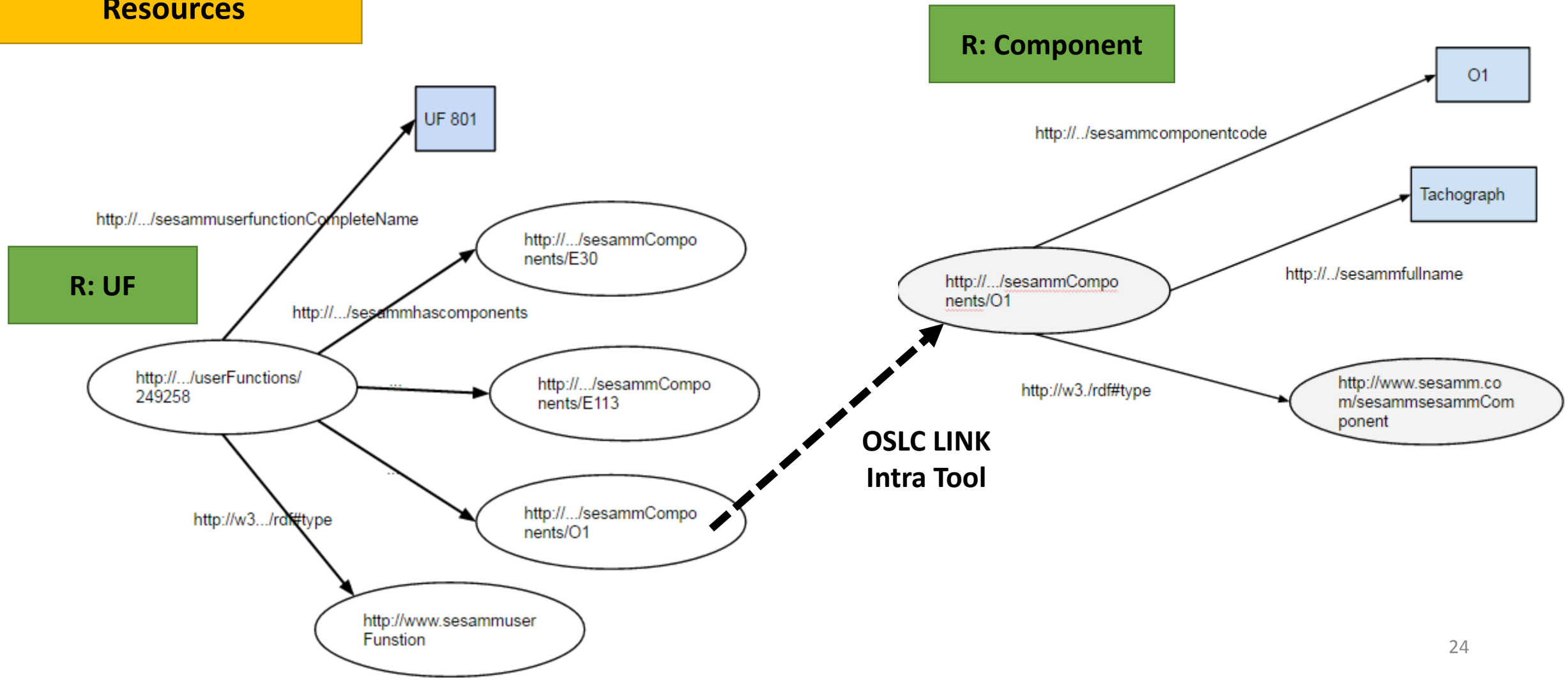
- Sesamm OSLC adaptor publish two resources to the web
  - Components – ECU´s
  - UF

Sesamm OSLC Provider adaptor Model



Sesamm Tool OSLC Adaptor

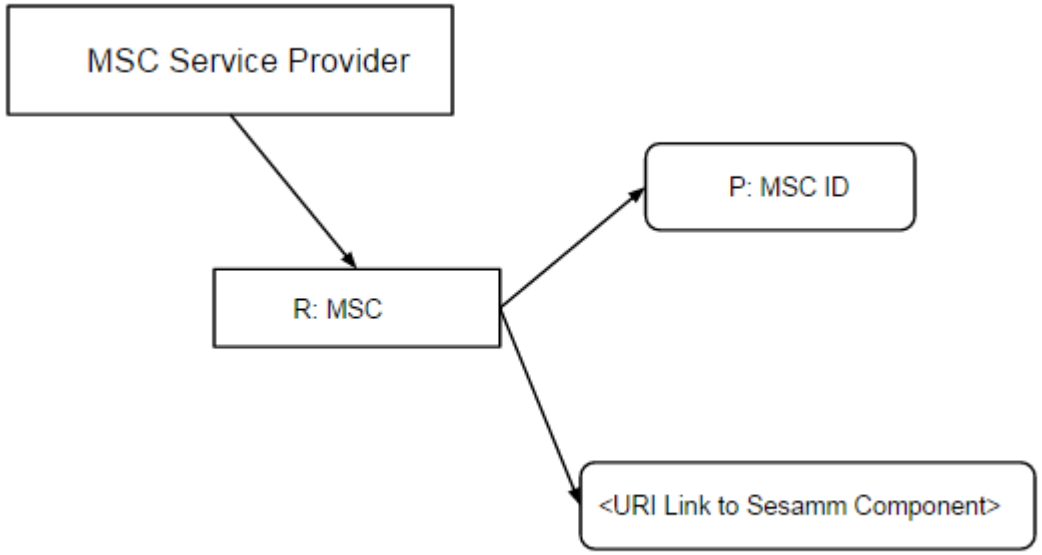
Example: Showing Sesamm Tool Adaptor Resources



OSLC Adaptor for MSC Drawer

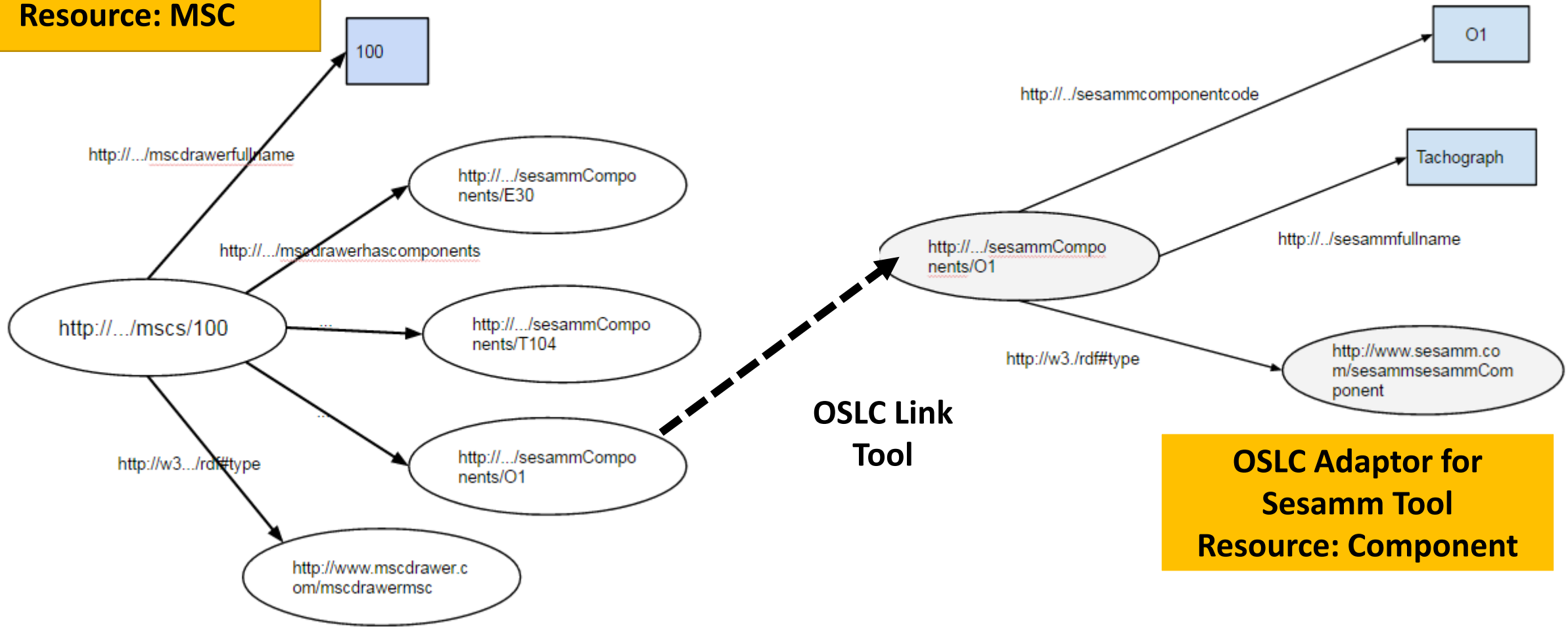
- MSC OSLC adaptor publish one resource to web
  - MSC – SCN of an UF

MSC Drawer OSLC Provider adaptor Model



Linking Resources between MSC Drawer and Sesamm Tool – Detail

OSLC Adaptor for MSC-Drawer
Resource: MSC



OSLC Adaptor for Sesamm Tool
Resource: Component

Tracking changes to OSLC Resources

- Tracked Resource Set (TRS) Sepecification is used to track changes.
- Part of OSLC core Specifcation.
- The specification allows a OSLC server to expose an exact set of resources (TRS), track additions to and removals from the set, and track changes to the resources in the set.

TrackedResourceSet Properties

Prefixed Name	Occurs	Read-only	Value-type	Representation	Range	Description
trs:base	exactly-one	True	Resource	Reference	ldp:DirectContainer	An enumeration of the Resources in the Resource Set.
trs:changeLog	exactly-one	True	Resource	Either	trs:ChangeLog	A Change Log providing a time series of incremental adjustments to the Resource Set.

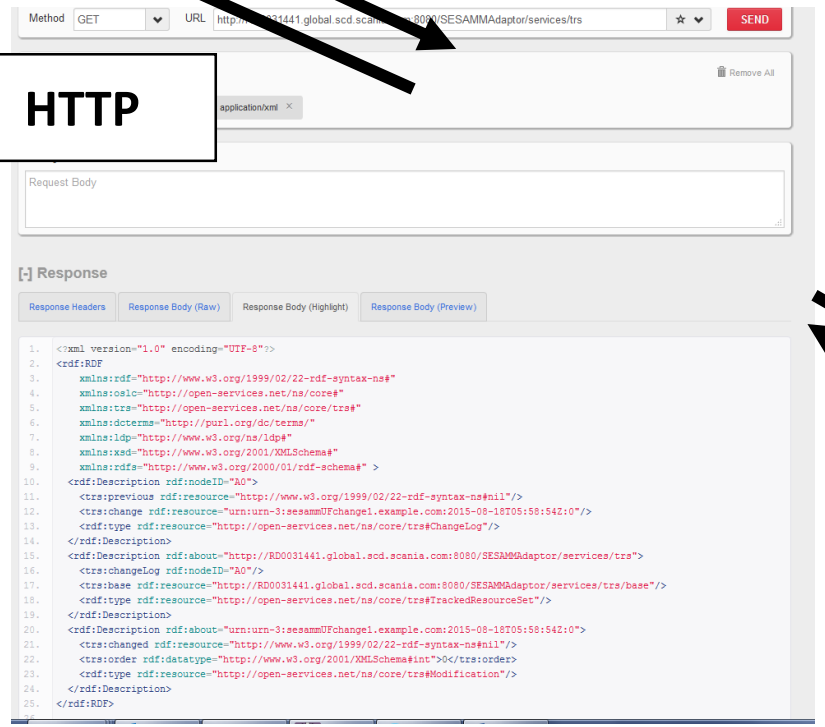
Retrieving Base and ChangeLog

IBM Rational Engineering Lifecycle Manager (RELM)

Client

Only ChangeLog

HTTP



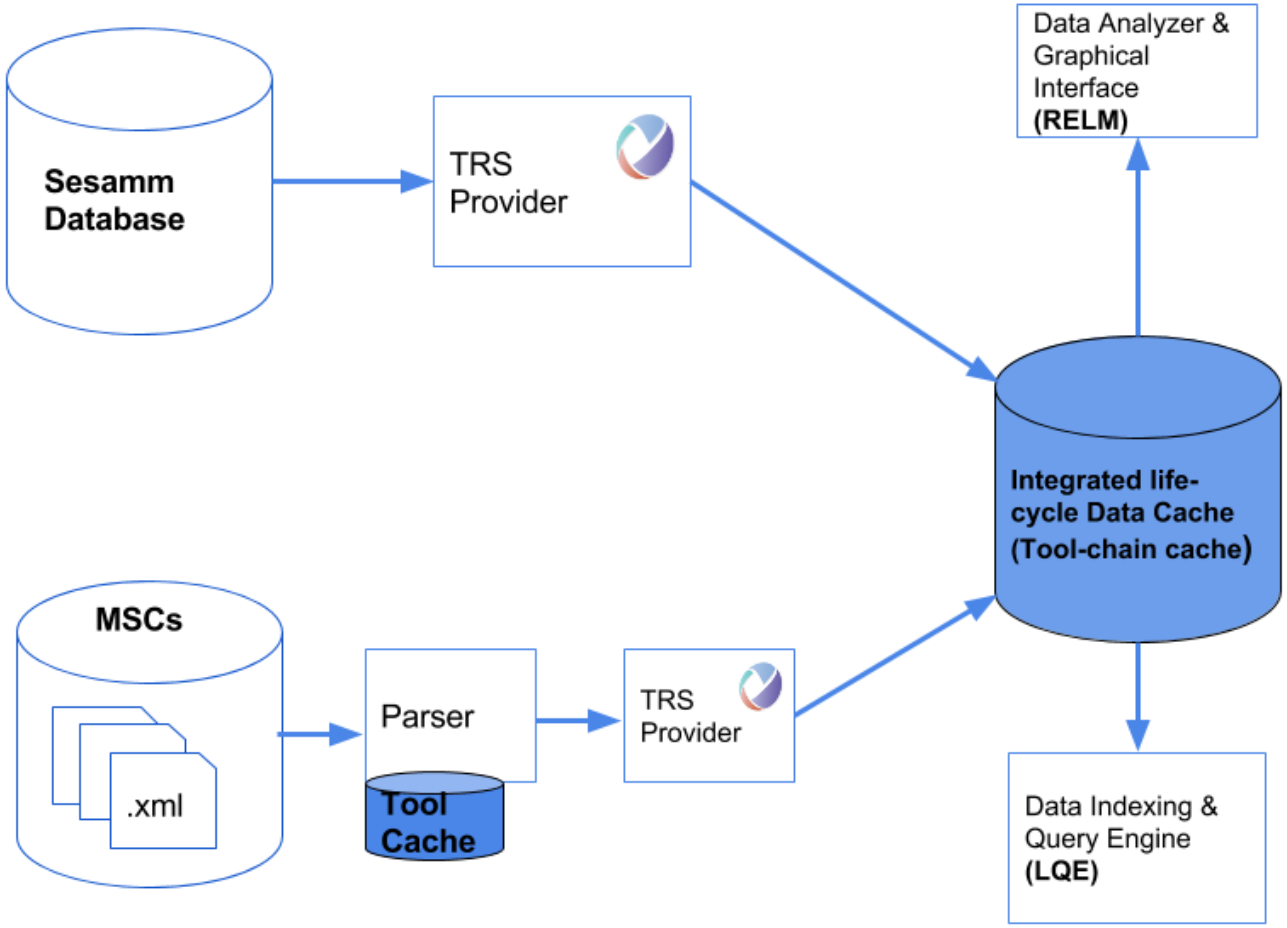
Sesamm Tool
TRS Adaptor

Support for tracked resource set protocol

SesammDB

OSLC Adaptor
Sesamm Tool

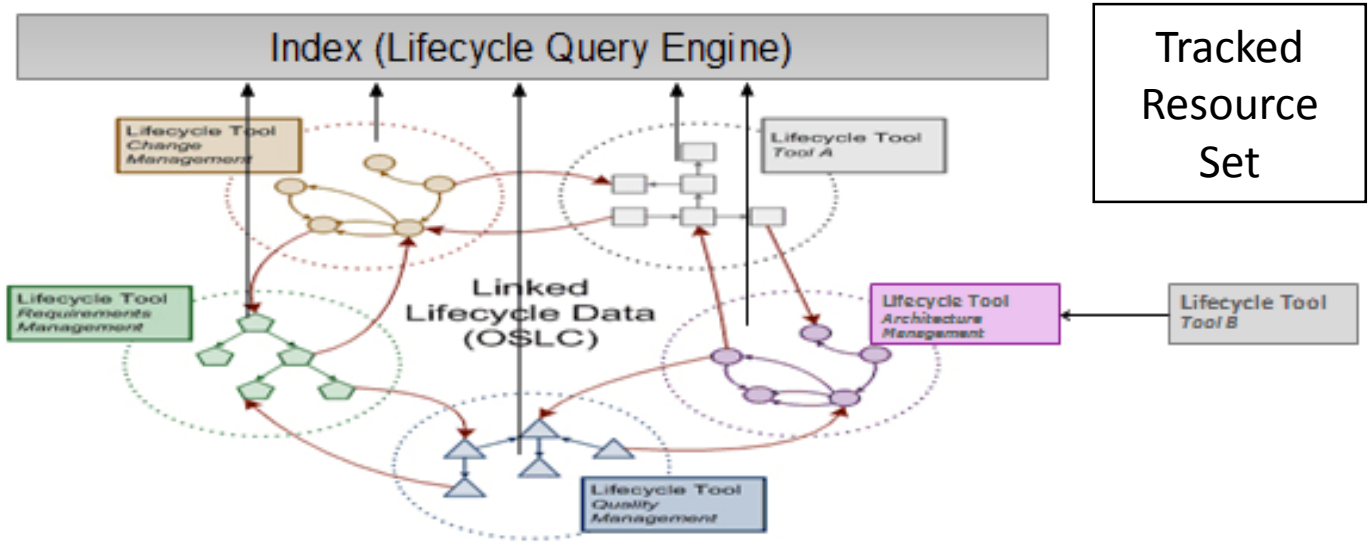
OSLC Tool-Chain Architecture for data integration



IBM Rational Engineering LifeCycle Manager (RELM)

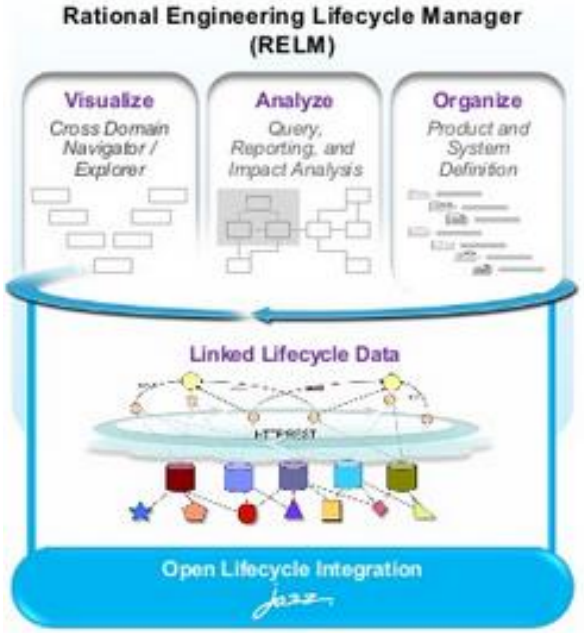
RELM is analysis tool, which help engineering teams to visualize, analyze and organize engineering data and their relationships.

Client Application LQE indexes TRS



Tracked Resource Set

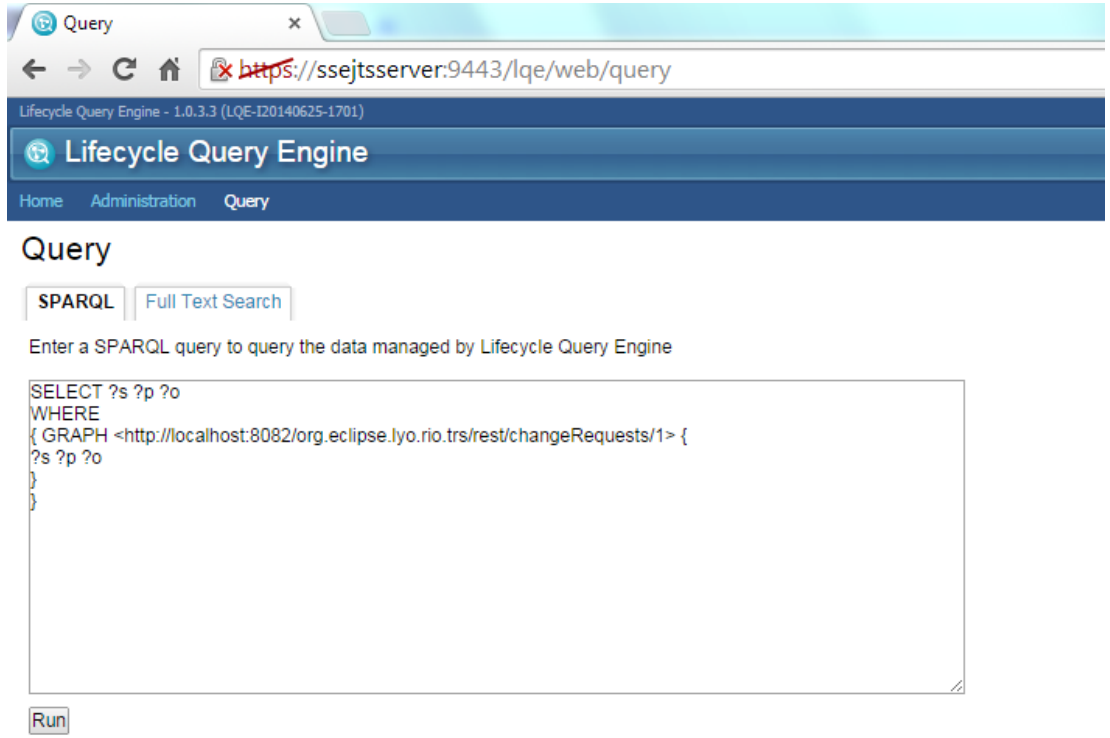
Analysis Tool



LifeCycle Query Engine

- Indexes reources from TRS providers (TRS Client)
- Indexed resources are represented as RDF
- Exposes an ´**Endpoint**´ for running queries on that data
  - **Endpoint:** A server that exposes its data via the SPARQL protocol

LifeCycle Query Engine SPARQL Endpoint



Test Queries: Cross Domain

Query 1:

“for all the components in one UF, give me all the MSCs where the components exists?”

Query 2:

“for all components in one MSC, give me all the UFs where the components exists?”

Query3:

*“for all components in UFs starting with **UF1**** *give me all the MSCs where the components exists?”*

DEMO – OSLC Tool-chain prototype

Results

Sr No.	Queries	Time of Execution (In seconds)
1	*"for all the components in one UF, give me all the MSCs where the components exists?"*	<1
2	*"for all components in one MSC give me all the UFs where the components exists?"*	<1
3	*"for all components in UFs starting with **UF1**** give me all the MSCs where the components exists?"*	<10

Table: Query Processing Time for each cross domain query

Results

Sr No.	Data Source	Indexing time (hh:mm:ss)
1	TRS Provider for Sesamm Tool	2:30
2	TRS Provider for MSC Drawer	1:10

Table: Data indexing time for each tool

Conclusions

- Data Consistency and Tracebility across the tool-chain.
- Capability of computation intensive, cross domain queries.
- Caching at individual Tool level and at Tool-Chain level.
- A federated approach to SPARQL query processing: Federated over multiple single repositories.
- LQE is dead-end of the Tool-Chain
- FO, SA or ST were not interested in RELM as an analysis tool.

Conclusions

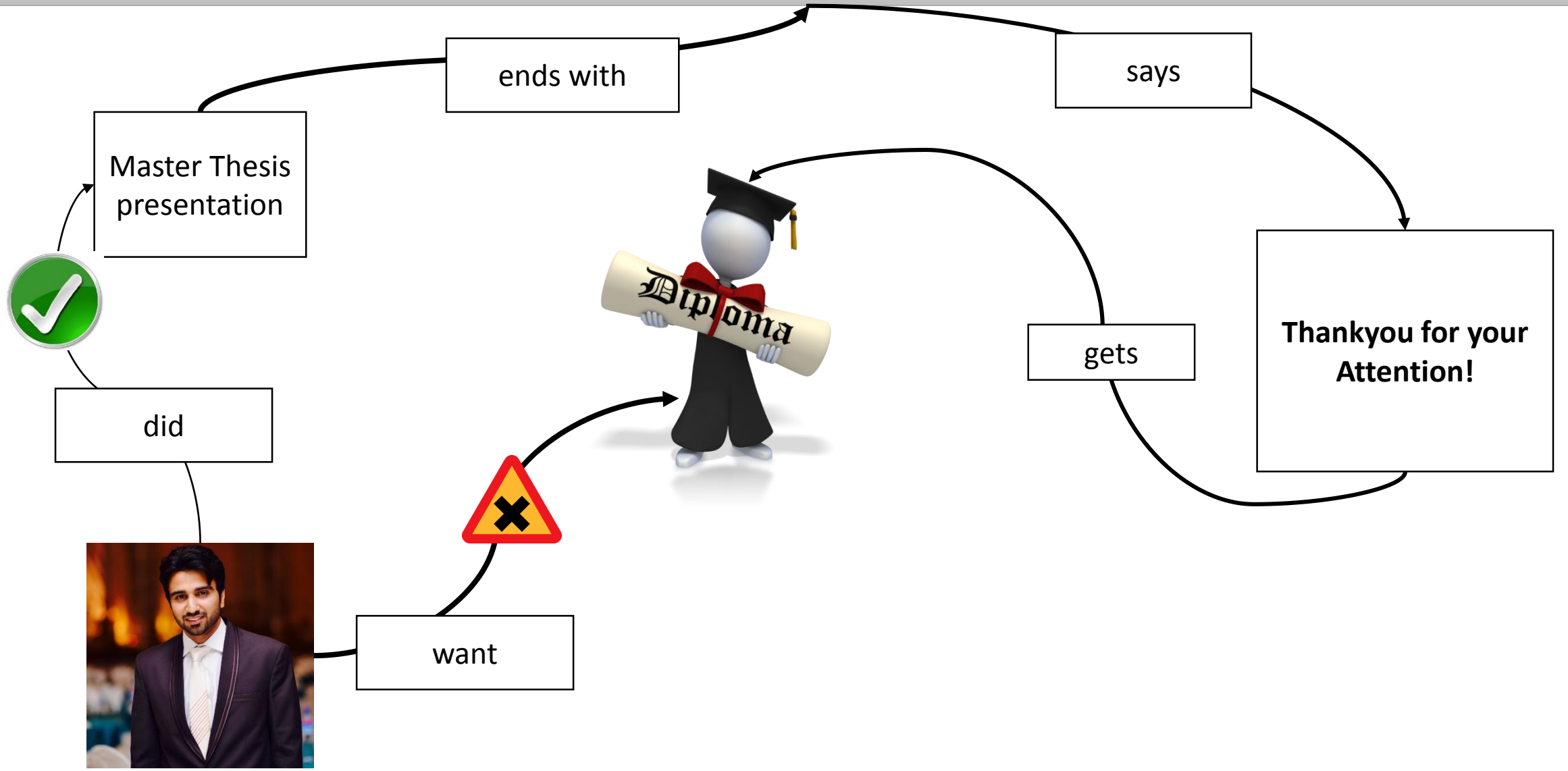
- **Key advantages**
  - Open Technologies (Using Open Web Standards XML, RDF, SPARQL etc...)
  - Dramatically reduce data integration costs, increase enterprise flexibility.
  - Scalability

- **Cons**
  - No open implementation for TRS Client exists.
  - Patent for TRS Client specification.
  - Installation of IBM RELM can be frustrating
  - No Interface for machines (APIs etc.)
  - SPARQL is still an expensive skill

Future Work

- HTTP PUSH capability in TRS Protocol
  - Tool data don’t change so often
  - Safety critical systems – instant update necesary
- Implentation of TRS Client
  - Interface for humans
  - Interface for machines (APIs etc.)
  - Analysis tool (Graphical tool) – Customizable for Specific needs.
- Further Development of prototype: further develop existing Adaptors + Including other tools

Q & A


ends with
says
Master Thesis presentation
Diploma
gets
Thankyou for your Attention!
did
want

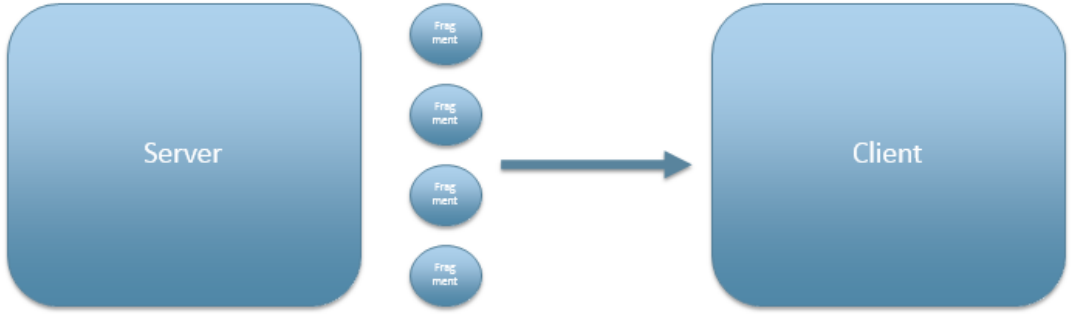
Appendix

Alternatives

- **SDshare** – A Specification used for publishing and consuming resource descriptions
- A protocol for tracking changes in semantic data store
- Based of Linked Data principles
- Defines how an RDF system can maintain a synchronized local copy of master data contained in another RDF system
- Not an active community behind the standard
- **Mature Product:**
  - Hafslund Sesam

Sesam Approach – SDshare implementation

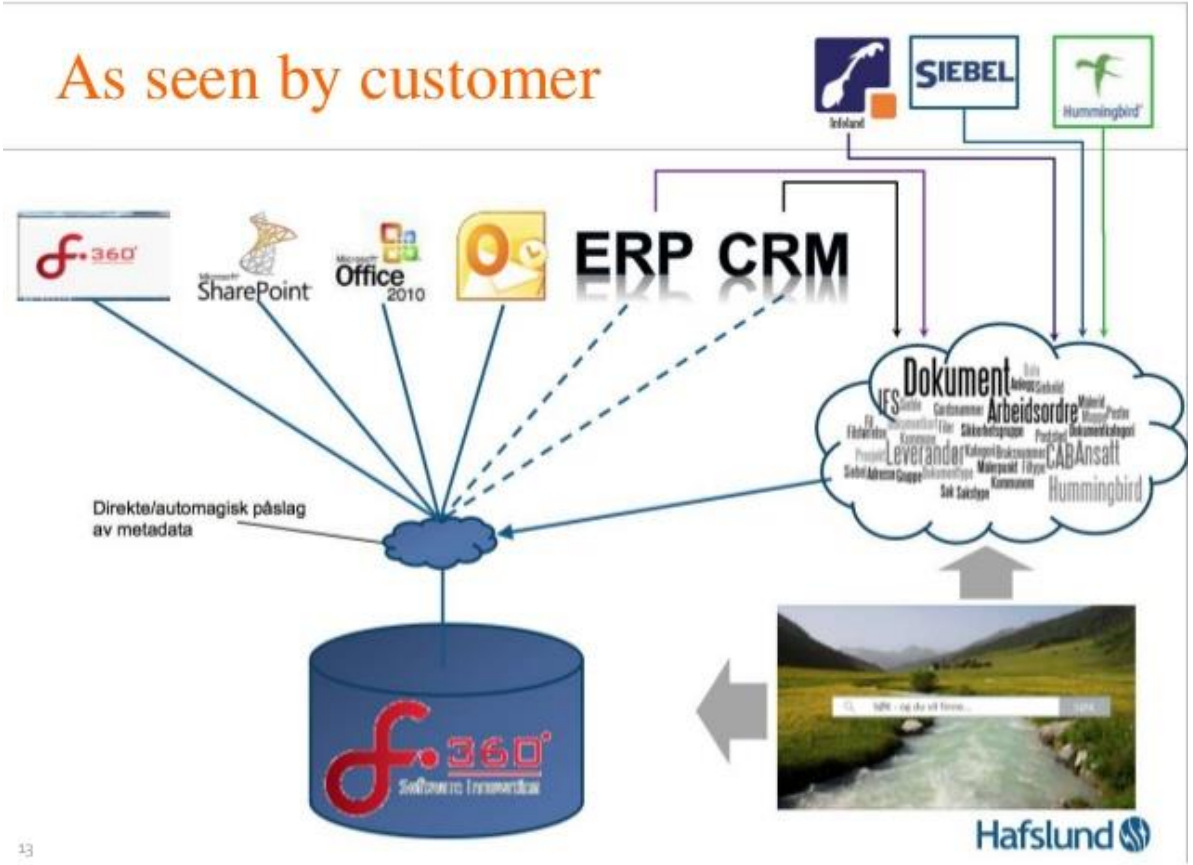
SESAM is an architectural approach and set of components for integrating, querying, reusing and repurposing enterprise data. SESAM enables the data from many sources to be seamlessly merged together in a schema-less, and scalable data store, that we call the DataHub.



Server publishes fragments representing changes in datastore

Client pulls these in, updates local copy of dataset

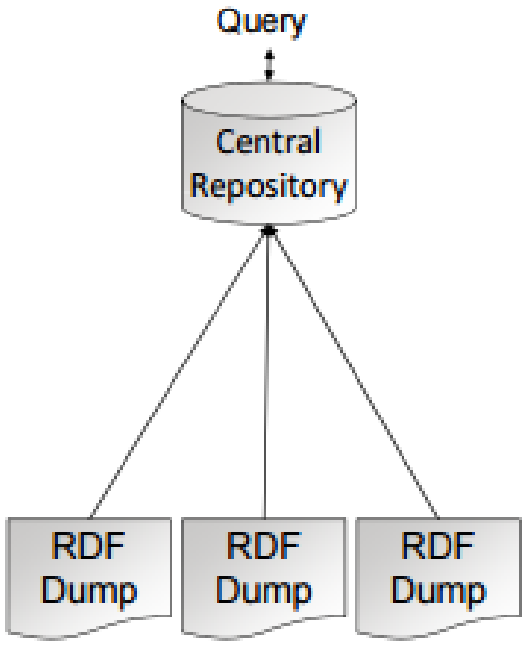
There is, however, more to it than just this



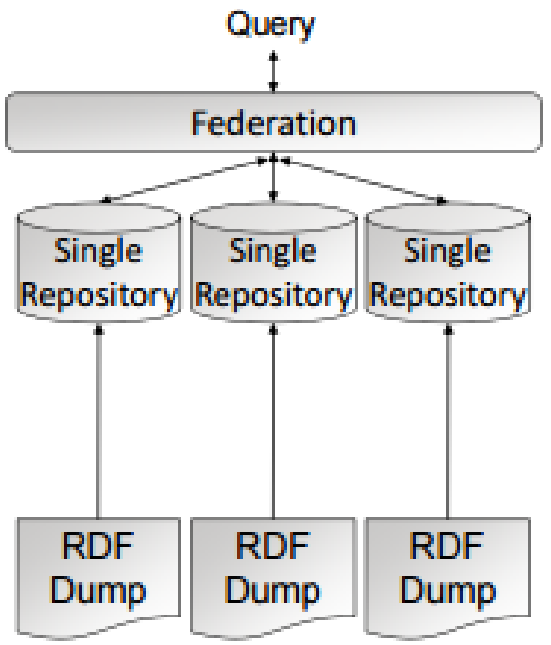
Sdshare – More Details

- SDshare spec
  - http://www.egovpt.org/fg/CWA_Part_1b
- SDshare issue tracker
  - http://projects.topicmapslab.de/projects/sdshare
- SDshare use cases
  - http://www.garshol.priv.no/blog/215.html

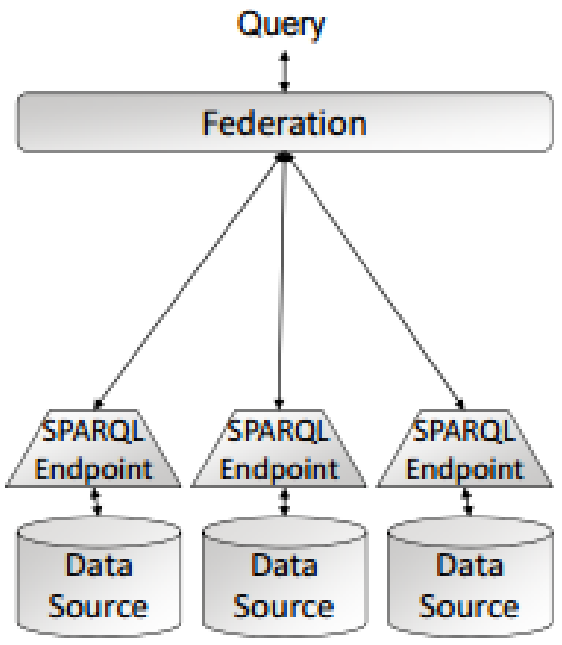
Approaches to Federation



a) Integration in a central repository



b) Federation over multiple single repositories



c) Federation over multiple SPARQL endpoints