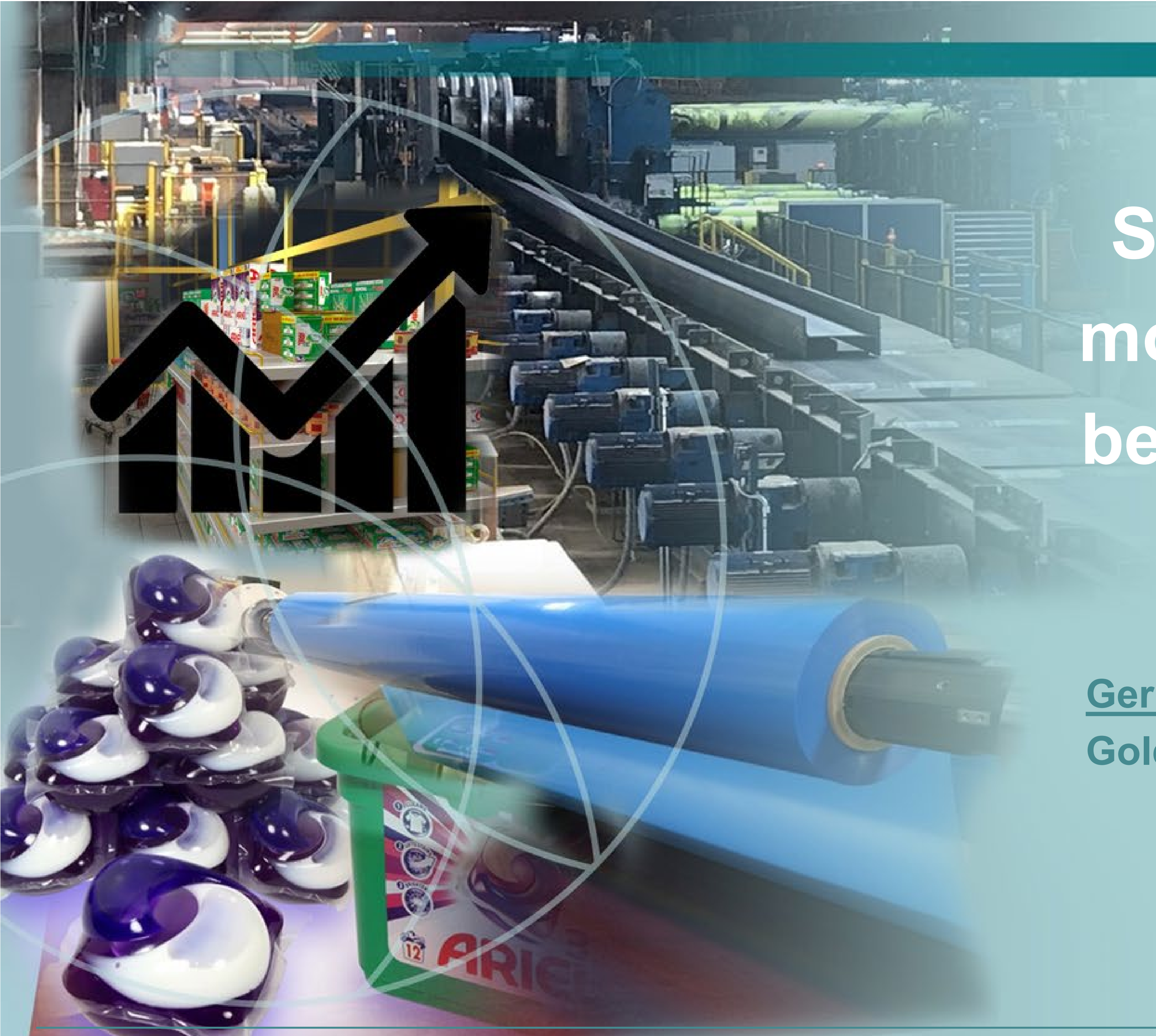
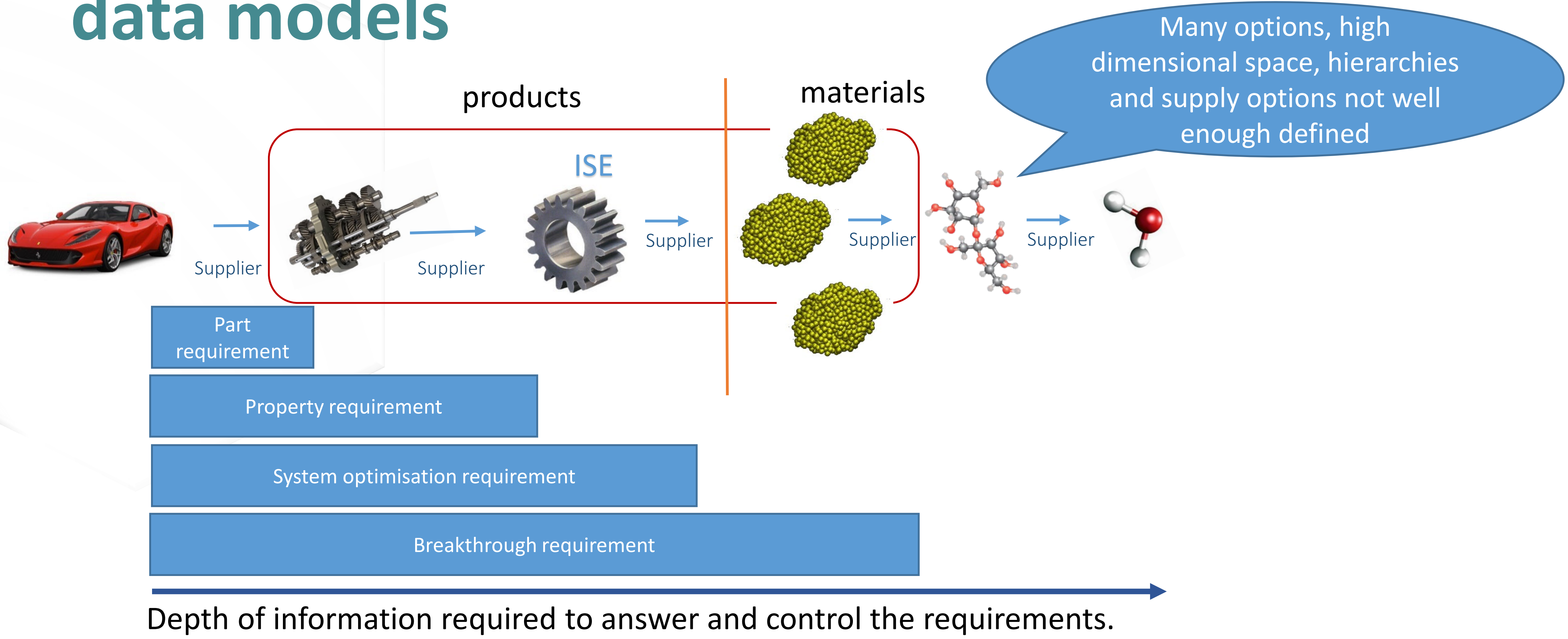


Semantic data and model management: benefits for Industry

Gerhard Goldbeck
Goldbeck Consulting Ltd

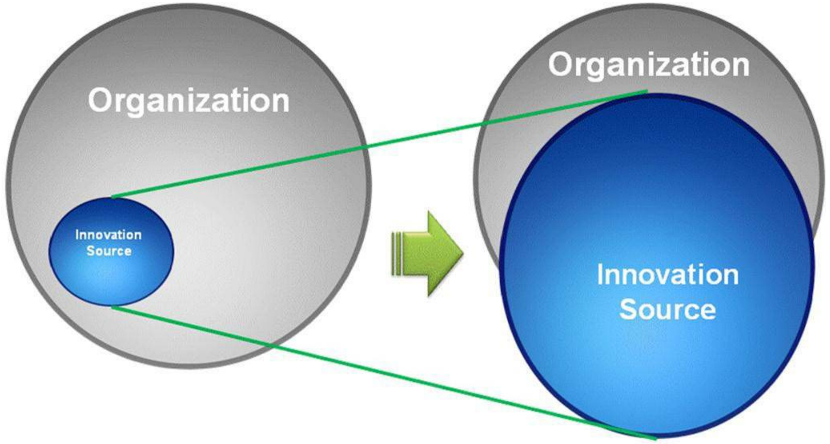
Complex requirements require deeper, semantic data models



…to support translation at enterprise level

From reliance in individual ‘heroes’ to systematic innovation management

- Reflect and maintain enterprise knowledge
- Combine **diverse** content in connected data source
- Find and utilise **relevant** information more quickly
- Provide more **transparency**
- Make better informed decisions
- Uncover **insights**



Innovation Vision by Malcolm Frank EVP – Strategy, Cognizant

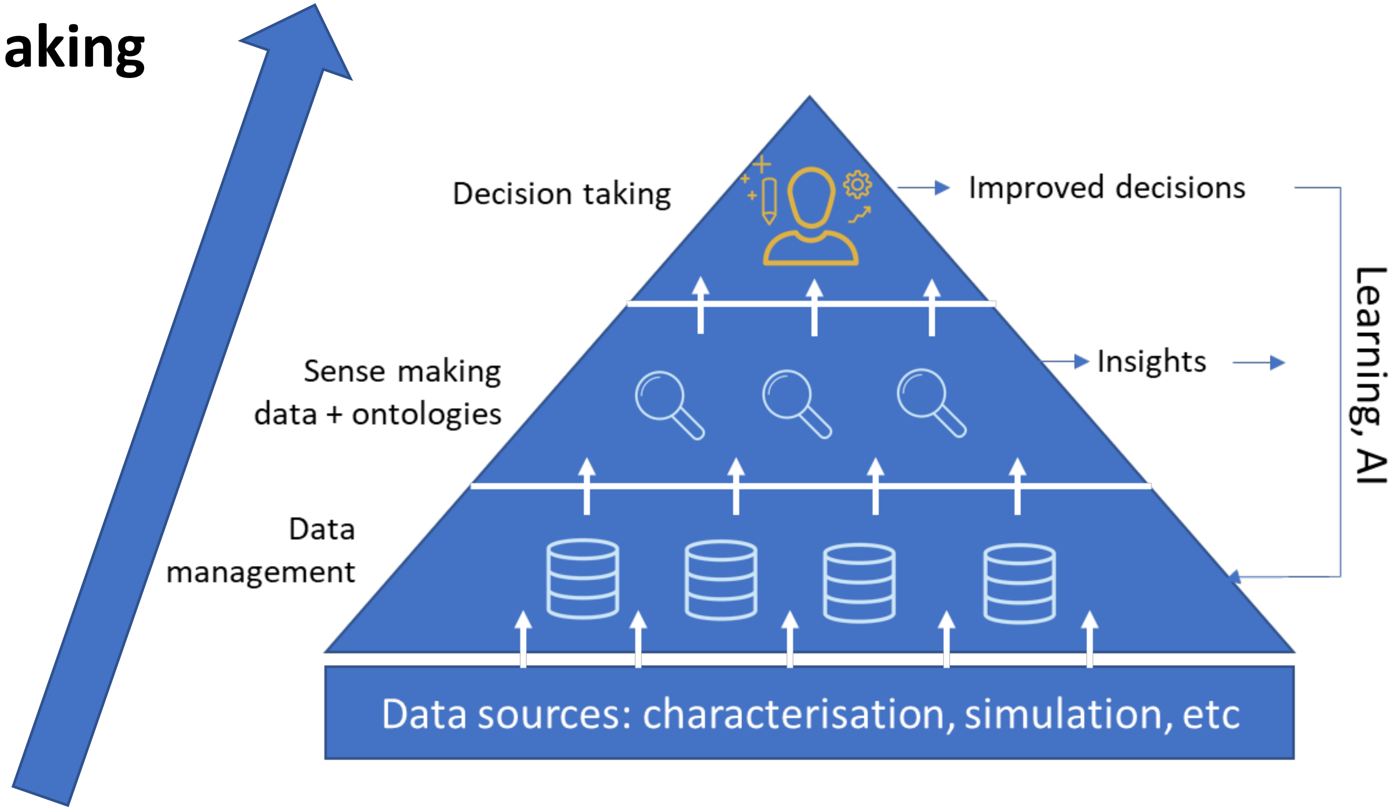
cf https://www.ibm.com/cloud/learn/knowledge-management

Open Translation Environment:

A Knowledge Management System for Research and Innovation

Towards better informed decisions

- **Knowledge exploration and decision making**
  - Exploration of complex information
  - Support multi-criteria decision making
- **Enhanced knowledge generation**
  - AI boosted by ontologies
- **Data documentation**
  - FAIR data, metadadata and ontologies
- **Data/knowledge generators**
  - Materials models (physics and data-based)
  - Characterisation, sensing



Evolution in knowledge systems

Example Knowledge Graphs – Siemens Technology drives innovation from world-class research to company-wide adoption

SIEMENS
Ingenuity for life

Milestones

- Google: Google Knowledge Graph announced
- SIEMENS: Breakthrough: ML on graphs
- Startups & big players entering KG market
- Gartner: Gartner add KGs to hype cycle
- SIEMENS: Mindsphere SDI adding graph support
- SIEMENS: > 30 KG projects, > 10 productive use > 6 CCTs
- SIEMENS: Shared ontologies

Technology innovation activities

Funded projects

LarKC
THESEUS
Optique

Cognitive Deep Learning

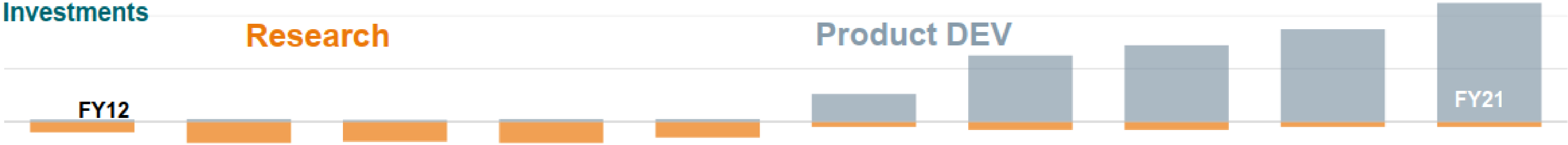
Business Buy-in

Data Strategy

Ontology libraries

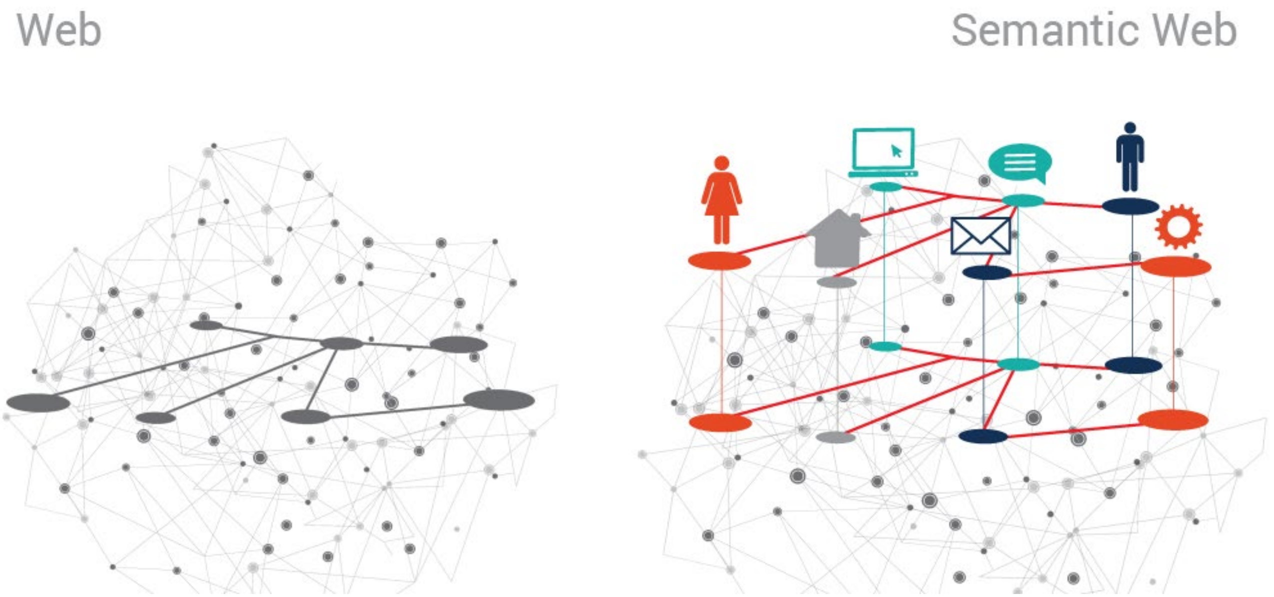
- emantic knowledge graphs & reasoning
- Knowledge Graphs as data layer on top of SQL Databases
- Machine Learning on graphs
- Proof-of-values and competence ramp-up
- Accelerate adoption & roll-out projects
- Shared & reusable ontologies

Technology Investments



Semantic data management: what and why?

- Applying semantic web technologies to the management of data
- Benefit is a **flexible, interconnected Data Model**
- **Saves time, cost, and improves maintainability**—over traditional technologies such as relational databases, due to
  - Open World Assumption (OWA): we don't know all the facts; no answer does not mean something is cannot be true; new facts can be easily included.



https://devopedia.org/semantic-web

https://cambridgesemantics.com/blog/semantic-university/semantic-technologies-applied/applying-semantic-web-two-camps/

When semantic data management makes sense

- Especially relevant for research where new facts/factors may emerge.
- In general when:
  - The data being managed changes constantly, e.g. different types of materials with different range of properties
  - The required views on those data (e.g. calculations, analytics, etc.) change constantly.
  - A lot of cross-organizational collaboration takes place
  - A lot of heterogeneous data sources need to be queried together: more agile and incremental than large data warehouses etc.
  - There is a lack of industry data standards.
  - Semantics provides Information classification, interoperability without syntactic standards, reasoning and retrieval

https://cambridgesemantics.com/blog/semantic-university/semantic-technologies-applied/example-semantic-web-applications/

Benefits

- Enables less technical users to quickly find, access, integrate and share data
- Allows subject matter experts in the business to become a part of the data modeling process
- Reduces the cycle time of accessing ready-to-use data#
- Reduce time for integration design by 30%, deployment by 30%, and maintenance by 70%.

- **the power emerging from the ability to reason on semantically integrated data** provides emergent insights that outweigh the large data integration benefits by more than 10:1. (for eccenca customers such as Siemens, Bosch, Nokia or AstraZeneca) (Chris Brockman, CEO of Eccenca on Linked-in)

Sources: Understand the Role of Data Fabric (Gartner), Gartner Top 10 Data and Analytics Trends for 2021

Example

Product Management System

- Enabling **comparability of specifications for 500,000 components**
- Enabling search for compatible spare parts for outdated machinery
- Integrating data of compatible products from competitors

- the different data sources of more than 500,000 products and their descriptions were integrated in knowledge graphs. Each knowledge graph links data of a logical domain from different sources that can be automatically updated at defined intervals. Additionally, expert knowledge and metadata on machines and spare parts were added to profit from and share best practices. To

98 % REDUCTION
Of Search Time

100 % SELF-SERVICE
Via Web Interface

100 % SCALABLE
For Future Projects

https://eccenca.com/success-story/product-data-management-system

Example: SpringerMaterials Accelerates Research With Semantic Search

STARDOG

Data from many different sources, hard to continue integration and to permit users to see the dynamic connections within and relationships between different materials, properties, and publications.

Solution:

- Unifying data in a knowledge graph: a single declarative model to bring all the connections between the content into a unified view.

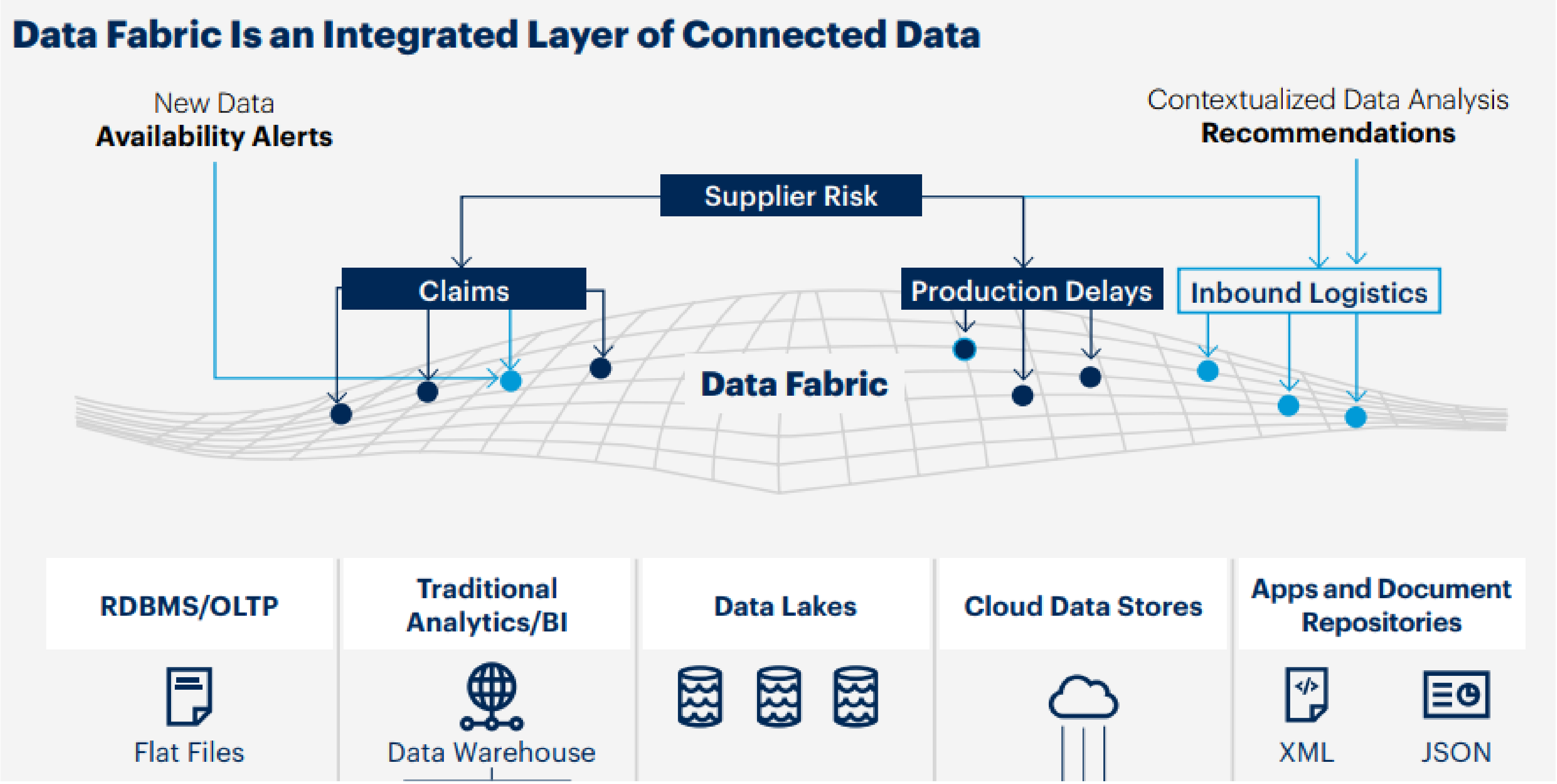
- SpringerMaterials knowledge graph contains 100 million triples spanning data sources like the Landolt-Börnstein series, Springer Handbooks, and relational databases that cover topics from corrosion to polymers.

- Uncover complex relationships from simple keywords.
- Reverse search to find substances based on property values
- Smart light-weight inference on substance and property identifiers
- etc

https://www.stardog.com/company/customers/springer-nature/

Semantic data management ->Data Fabric

- Data fabrics utilize semantic knowledge graphs, metadata management, and machine learning to unify data across various data types and endpoints. https://www.ibm.com/topics/data-fabric



https://www.gartner.com/en/publications/essential-guide-to-data-fabric

Under the hood

- Knowledge Graphs:
  - e.g. dbpedia https://www.dbpedia.org/about/ crowd-sourced community effort to extract structured content from the information created in various Wikimedia projects.
- Ontology that defines the semantics of the KG
  - Dbpedia ontology: consists of 320 classes which form a subsumption hierarchy and are described by 1,650 different properties.

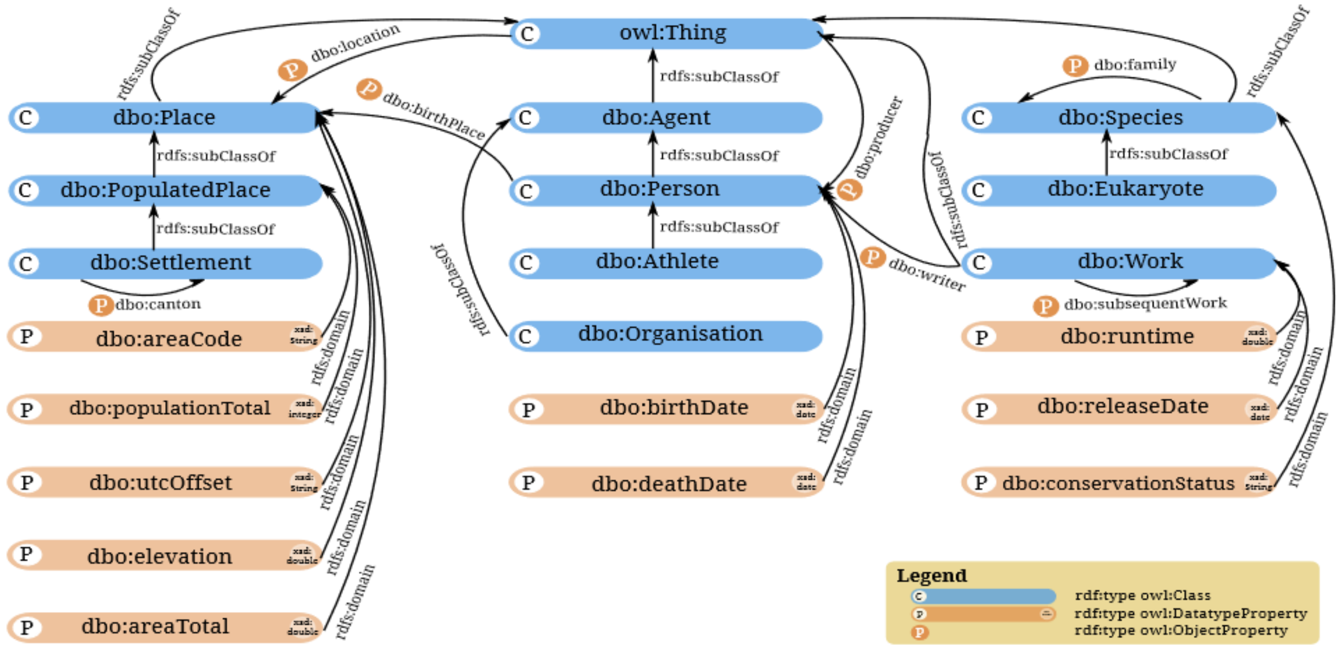


Fig. 3. Snapshot of a part of the DBpedia ontology.

Note: it has Chemical Substance class (similar to Schema.org/Google) but nothing about materials

Why Ontology based Materials KM

- Integration of disparate data sources into one queryable system: answer questions across data silos more efficiently
- Expressing scientific and organisational knowledge by means of curated entities and their relations: an ontology and knowledge graph
- Build a ‘corporate memory’ (see eccenca)
- Integrate external data more easily with internal data
- Support data extraction from literature and its integration with internal data by having well defined and organisationally meaningful concepts to map to.
- Manage and curate information exchange, e.g. between departments, along value chain, e.g. done by Catena-X for Battery passport: based on shared semantic model.
- Build new applications faster and in a more lightweight manner than by heavy data analytics: ability to play with and utilise new concepts and relations

How to construct a Materials KG?

What foundations can I build upon?

Examples from other fields:

- Chemistry has the IUPAC Goldbook: terminology with definitions. Ontologies have been built utilising these definitions (NB: there are still typically inconsistencies in concepts that need to be overcome for machine processable (logically consistent) ontologies

- WWW: e.g Schema.org

- Life Sciences: Gene Ontology, CHEBI ontology, are effectively large, curated catalogues.

- Materials Sciences ????

Challenges for semantic KM of Materials

Heterogenous terminology is prevalent.

EOSC Interoperability Framework presents the following challenges:
https://eosc-portal.eu/eosc-interoperability-framework

- **Lack of common explicit definitions about the terms that are used by user communities**
- Lacking common semantic artefacts across communities (e.g., general ontologies that can be shared)

Domain terminologies and ontologies SOTA

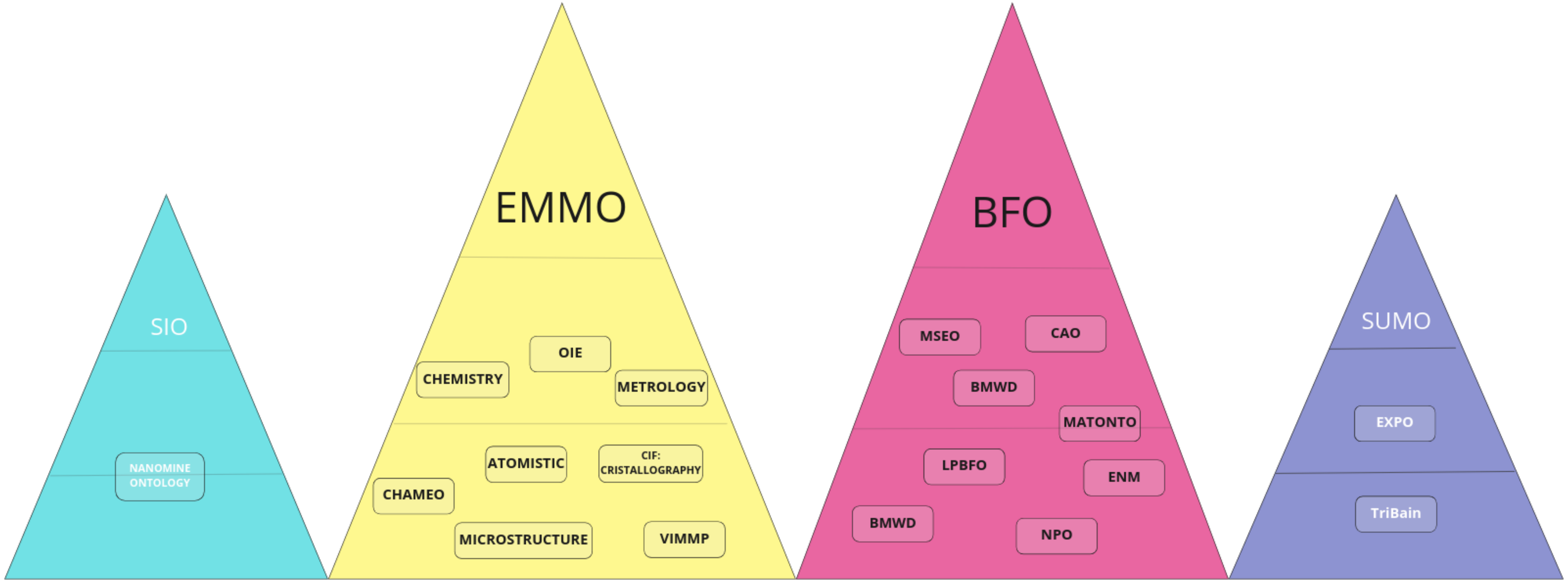
Review and Alignment of Domain-Level Ontologies for Materials Science

Anne De Baas, Pierluigi Del Nostro, Jesper Friis, Emanuele Ghedini, Gerhard Goldbeck, Ilaria Maria Paponetti, Andrea Pozzi, Lan Yang, Arkopaul Sarkar, Francesco Antonio Zaccarini and Daniele Toti

Forthcoming paper (to be submitted to IEEE Access)

Analysis of materials and chemistry domain Ontologies

- About 50 (accessible, machine readable) ontologies
  - distributed across SIO, EMMO, BFO, SUMO top level ontologies



Coverage of Materials Science terminology

Glossary of terms from

- some 700 terms harvested from Textbook by Callister “Modern materials science and engineering”
- Materials modelling CWA (based on the RoMM)
- Materials characterisation CWA
- ONTOCOMMONS demonstrators

Classification of 800 terms into seven subjects.

1. Materials classes
2. Materials structure
3. Materials properties
4. Materials behaviour
5. Materials technologies
6. Materials theories
7. Products/devices

An allocation to a subject means this is a term that is generically used in this field and thus expected to figure in a related ontology

Terminology coverage in Ontologies

Ontology/Percentages of matching with domain terms	Materials classes	Materials structure	Materials properties	Materials behaviour	Materials Technologies	Theories for Materials
Chemical Methods Ontology	12,088	20,526	**35,354**	26,230	29,434	31,868
eNanomapper	18,681	23,158	31,313	24,590	26,792	27,473
MaterialsMine	9,890	15,789	23,232	24,590	13,962	10,989
NanoParticleOntology	9,890	16,842	22,222	13,115	16,604	24,176
NanoMine	6,593	13,158	21,212	21,311	10,566	9,890
DEB	6,593	10,000	14,141	8,197	9,811	3,297
EMMO-Mechanical Testing	12,088	10,000	13,131	8,197	14,340	2,198
MatOnto	8,791	6,842	12,121	14,754	3,774	4,396
MSEO	6,593	10,000	10,101	19,672	9,811	2,198
BUILDMAT	2,198	3,158	8,081	8,197	3,019	3,297
CIF Ontology	4,396	11,053	8,081	4,918	12,075	5,495
EMMO Metrology	0,000	3,158	6,061	3,279	6,792	3,297
OIE manufacturing	12,088	5,789	6,061	3,279	8,679	1,099
VIMMP marketplace-level domain ontologies	6,593	7,895	6,061	8,197	18,491	6,593
Characterisation Methodologies Ontology	1,099	4,737	5,051	4,918	20,000	3,297
Chemical information ontology	4,396	6,842	5,051	3,279	4,528	7,692
Chemical Analysis Ontology	6,593	7,368	4,040	3,279	9,811	3,297
MOL_TENSILE	1,099	1,053	4,040	6,557	2,264	0,000
Reaction ontologies	2,198	7,368	4,040	1,639	6,792	2,198
EMMO Materials	2,198	5,263	3,030	1,639	3,774	3,297
Material properties ontology	3,297	2,632	3,030	0,000	3,019	0,000
MESO	2,198	4,211	3,030	1,639	7,170	2,198
EMMO Domain Photovoltaics	0,000	2,632	2,020	1,639	1,887	2,198
EMMO Models	0,000	1,579	2,020	0,000	7,547	0,000
Precipitation Model	1,099	1,579	2,020	1,639	1,509	0,000
EMMO Manufacturing	0,000	1,579	1,010	0,000	2,642	0,000
EMMO-Atomistic	0,000	0,526	1,010	0,000	1,887	0,000
EMMO-Crystallography	0,000	4,211	1,010	0,000	1,887	0,000
EMMO-Microstructure	3,297	10,526	1,010	1,639	4,906	1,099
LPBFO	0,000	4,737	1,010	4,918	6,415	0,000
Materials Design Ontology	0,000	0,000	1,010	0,000	0,377	0,000
MDO-FULL	0,000	0,000	1,010	0,000	0,377	0,000
Metal-alloy	3,297	3,158	1,010	0,000	1,887	0,000
MOCO	0,000	0,526	1,010	0,000	1,132	0,000
Ontology for Simulation, Modelling, and Optimization	1,099	4,737	1,010	1,639	10,189	0,000
PMD Core Ontology	0,000	5,263	1,010	1,639	5,660	0,000
CHEBI (reduced)	3,297	4,211	0,000	0,000	2,264	0,000
OIE software	0,000	1,579	0,000	0,000	3,019	0,000
Semantic Types Ontology	1,099	1,053	0,000	0,000	3,019	0,000

Automatic analysis of 47 domain ontologies for assessing their coverage of the considered subject areas

Coverage maximum 35% of terms in an ontology

Ontologies were scanned for the **presence** within their elements of the relevant terms for each domain

Most frequent terms in DLOs

Most frequent terms derived from the automatic analysis of DLOs.

Terms	Occurrences	Percentage (across ontologies) [%]
material	42	89
atom	33	70
structure	33	70
component	32	68
property	29	62
phase	26	55
quantity	25	53
composition	24	51
measurement	23	49
particle	23	49
experiment	22	47
molecule	22	47

Need coherent representation of fundamental materials science concepts

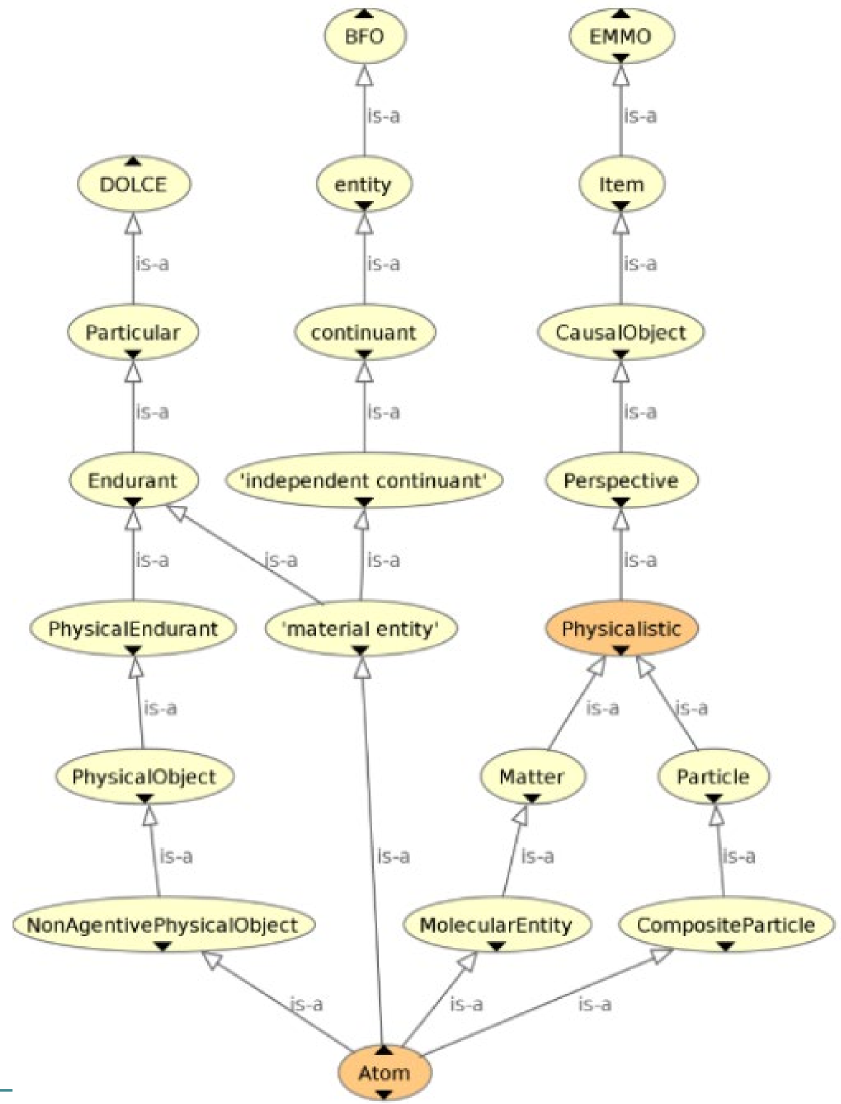
Key concepts include:

- Matter
- Material
- Atom
- Molecule
- Substance

- Physical Quantity
- Materials Property
- Materials Processing

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

OntoCommons:
Bridge Concepts development that provide detailed semantics and alignments.
Developed Bridge Concepts for the above as well as Materials Component and Materials Process



The Need for a Materials Ontology

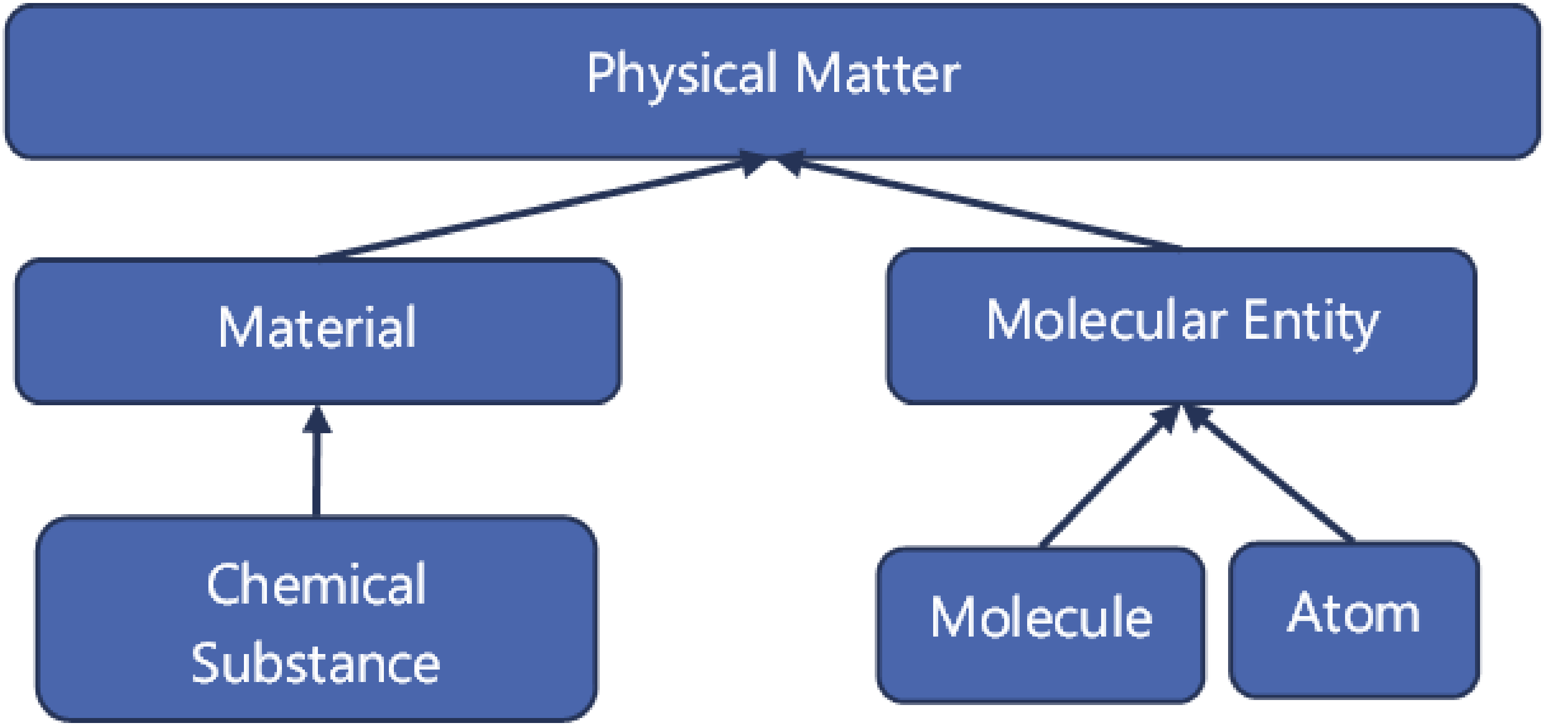
In 2018 several European practitioners in Materials Science under the **governance** of the **EMMC** expressed the need to develop a **knowledge framework** consistent with **scientific principles and methodologies** to complement the **existing physical-mathematical approach.**

https://emmc.eu/)

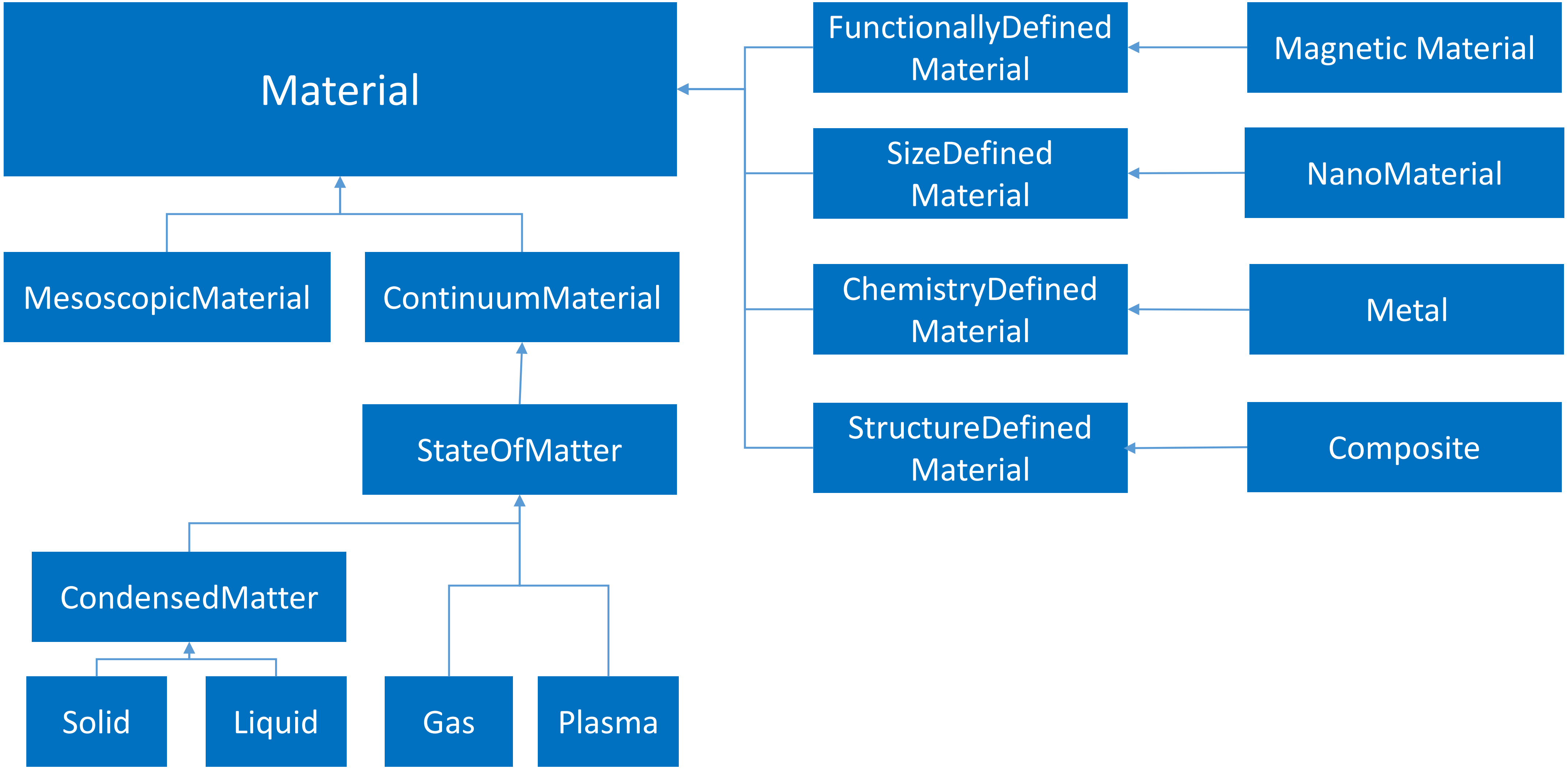
The **Elementary Multiperspective Material Ontology (EMMO)** is an ontology developed to represent such knowledge framework.

https://github.com/emmo-repo/EMMO

Materials and Chemistry Classes in EMMO



Materials classes in EMMO



OntoTrans OTE:

Knowledge Management with a difference

- Includes Models : “Live data”: connection to simulations, ML/AI
- Guided and curated knowledge capture in an ontology, EMMO based
- Reasoning, recommendation and knowledge graphs support exploration, insights and decisions
- Demonstrated in wide range of industries: highly flexible and adaptable system

Conclusions

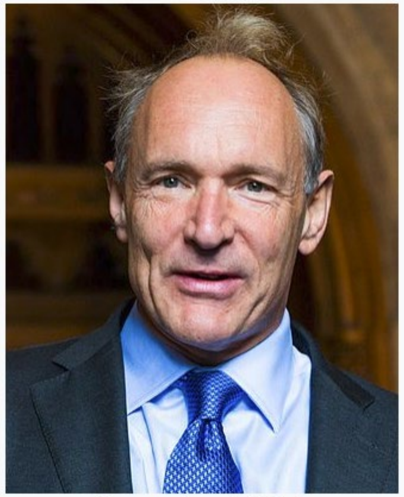
- Semantic Data Management brings a wide range of benefits that are very relevant to research environments.
- A wide range of examples demonstrate the benefits.
- Underlying Knowledge Graphs need to be based on a well develop, fit-for-purpose ontology.
- There are many KG and ontologies in different disciplines, but a severe lack in materials, e.g. lacking from dbpedia, no agreed terminologies, in fact often conflicting definitions etc.
- EMMO provides a solid ontology basis
- OntoTrans OTE is KM for materials with added ‘live’ data (from simulation)
- **We need Semantic Materials to realise the Semantic Web vision for materials innovation.**

Berners-Lee, Semantic Web vision,1999

I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers.

A "Semantic Web", which makes this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines.

https://en.wikipedia.org/wiki/Semantic_Web



Berners-Lee in 2014

Born	Timothy John Berners-Lee 8 June 1955 (age 68) London, England
Other names	TimBL TBL
Education	The Queen's College, Oxford (BA)
Known for	Invention of the World Wide Web
Spouses	Nancy Carlson (m. 1990; div. 2011) Rosemary Leith (m. 2014)
Children	2 children; 3 step-children

Acknowledgements

- Anne De Baas, Daniele Toti, Alex Simperler (Goldbeck Consulting)
- Emanuele Ghedini, Ilaria Paponetti, Francesco Zaccarini (Univ Bologna)
- Jesper Friis (SINTEF)

This work has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreements no. 958371 (OntoCommons) and no. 862136 (OntoTrans)

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

This work has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreements no. 958371 (OntoCommons) and no. 862136 (OntoTrans)

Thank you!

The OntoTrans project has received funding from the European Union‘s Horizon 2020 research and innovation programme under Grant Agreement No 862136.