



Data and Modelling Integration

Hein Koelman, Don Pellegrino, Joey Storer

EMMC Workshop November 7-8 2017 – Cambridge UK

The Dow Chemical Company



Key Facts About Dow Chemical

2016 Sales of \$48.2 Billion



- Agricultural Sciences
- Consumer Solutions
- Infrastructure Solutions
- Performance Materials & Chemicals
- Performance Plastics

Dow Sites Around the World

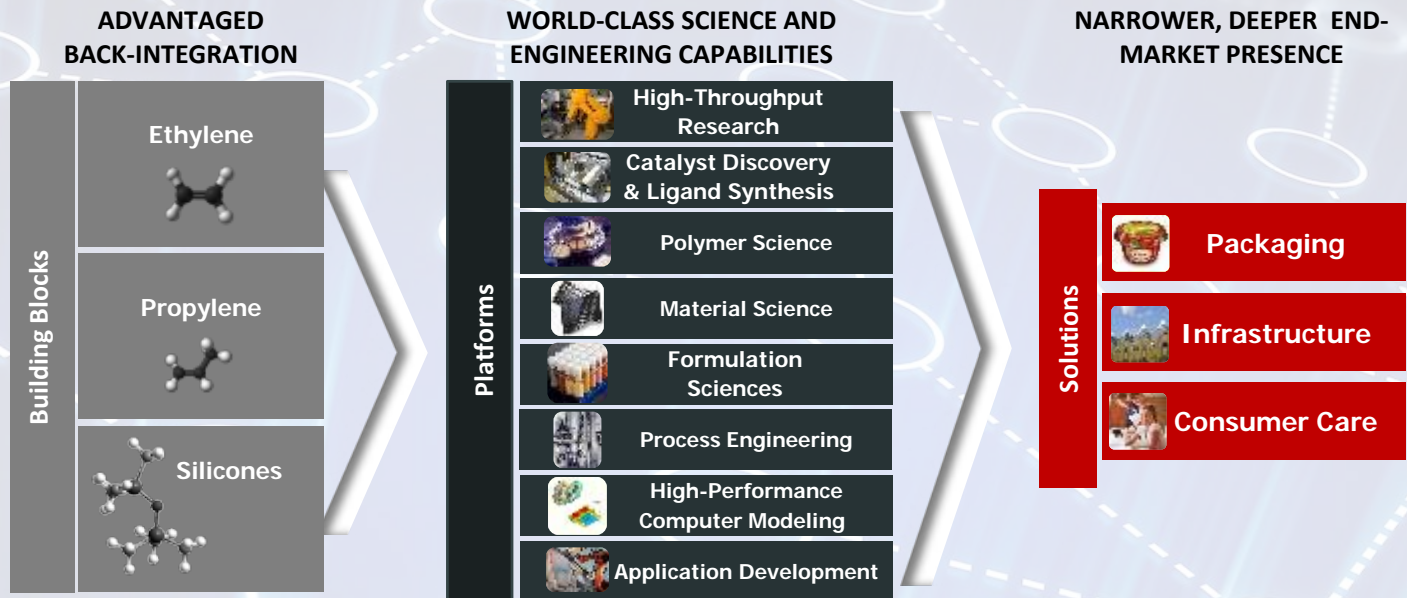


Corporate Stats

- 2016 Op. EBITDA¹ of \$9.8B
- 2016 Cash Flow from Operations of \$5.5B
- ~56,000 employees
- >7,000 product families
- Sales in 175 countries

1. Operating EBITDA is defined as EBITDA excluding the impact of "Certain Items," where "EBITDA" is defined as earnings (i.e., "Net Income") before interest, income taxes, depreciation and amortization.

Materials Science Division



Consumer-Driven Demand in Core End-Markets Propels Growth, Driven By Sustainable Urbanization Trends in Emerging Geographies

Towards a connected world in material science

Vision to have at the fingertips of the scientist the relevant (holistic) info, data, context, models, analytics and recommendations on what to use, study and develop as per the (customer) needs.

The value of this is huge: Accelerate Innovation

Typical Data Landscape within an R&D Organization

Excel files

Local / Files Shares



Test equipment
disc drives



Documents/ Reports
Local/File Shares

Additional Issues

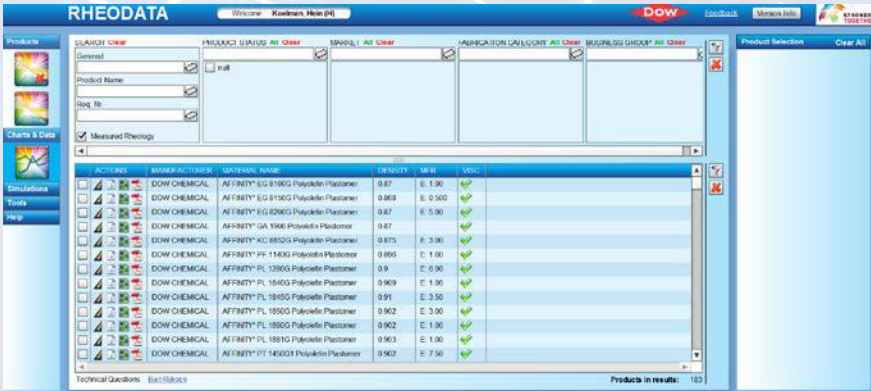
- Context Missing / Quality?
- Lots of vendor systems
- Each business own approaches
- Units / Test protocols
- “System of record”



Internal
Databases

Typical Model Landscape

Point solutions: specific models to solve/use for very specific needs. Models are on Islands, not connected, not visible when needed. Some dedicated widespread programs, e.g. Dow RheoData

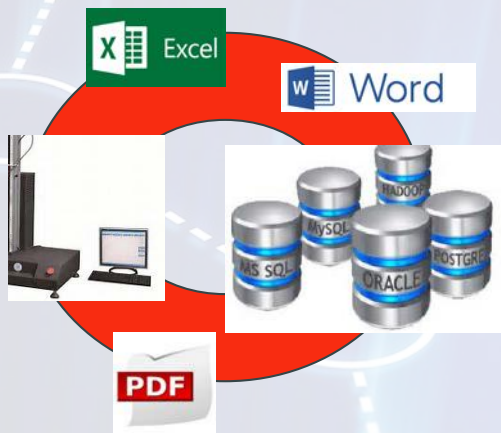


The screenshot displays the RHEODATA software interface. The main window shows a table of polymer products. The table has columns for 'ACTIVITY', 'MATERIAL NUMBER', 'MATERIAL NAME', 'DENSITY', 'M.W.', and 'VISC'. The 'MATERIAL NAME' column lists various Polyolefin Plastomers. The 'DENSITY' column shows values ranging from 0.87 to 0.962. The 'M.W.' column shows values ranging from 1.90 to 7.50. The 'VISC' column shows values ranging from 0.500 to 3.00. The interface also includes a search bar, a sidebar with navigation options, and a 'Product Selection' panel on the right.

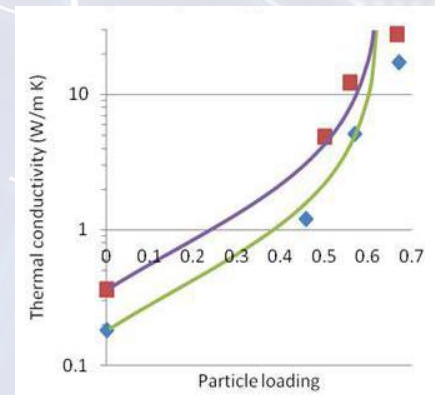
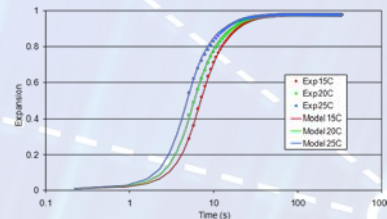
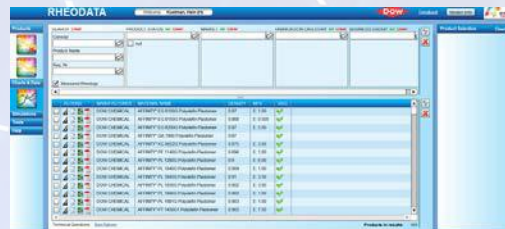
ACTIVITY	MATERIAL NUMBER	MATERIAL NAME	DENSITY	M.W.	VISC
		DOW CHEMICAL AF-FINITY-EG 8150G Polyolefin Plastomer	0.87	E: 1.90	✓
		DOW CHEMICAL AF-FINITY-EG 8150G Polyolefin Plastomer	0.868	E: 0.500	✓
		DOW CHEMICAL AF-FINITY-EG 8200G Polyolefin Plastomer	0.87	E: 5.90	✓
		DOW CHEMICAL AF-FINITY-GA 1900 Polyolefin Plastomer	0.87		✓
		DOW CHEMICAL AF-FINITY-KD 8820G Polyolefin Plastomer	0.875	E: 3.00	✓
		DOW CHEMICAL AF-FINITY-PP 1140G Polyolefin Plastomer	0.866	E: 1.90	✓
		DOW CHEMICAL AF-FINITY-PL 1260G Polyolefin Plastomer	0.9	E: 6.90	✓
		DOW CHEMICAL AF-FINITY-PL 1600G Polyolefin Plastomer	0.909	E: 1.90	✓
		DOW CHEMICAL AF-FINITY-PL 1840G Polyolefin Plastomer	0.91	E: 3.50	✓
		DOW CHEMICAL AF-FINITY-PL 1900G Polyolefin Plastomer	0.902	E: 3.00	✓
		DOW CHEMICAL AF-FINITY-PL 1900G Polyolefin Plastomer	0.902	E: 1.90	✓
		DOW CHEMICAL AF-FINITY-PL 1981G Polyolefin Plastomer	0.903	E: 1.90	✓
		DOW CHEMICAL AF-FINITY-TT 5400G Polyolefin Plastomer	0.902	E: 7.50	✓

Towards a Linked Data System

Data World



Model (Virtual) World



E.g. Rheological data -> recommended /available rheology models
Rheology model result -> Nearby real world data

Objectives in developing a linked data system

Objective 1 - Don't repeat an experiment

Objective 2 - Don't duplicate reference data

Objective 3 - Capture provenance

Objective 4 - Make information discoverable and expose context.

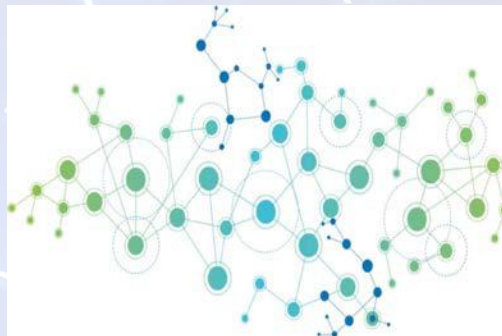
Objective 5 - Support modeling

Objective 6 - Support internal data creation, publication, and reuse

Objective 7- further develop / serve as a knowledge system

Ontology “ the one to connect them all “

















































- BFO
- CHEMINF
- OBO Foundry/Ontobee
-
- PubChem RDF



... and graph database



LifeSciences vs Materials Ontologies

No.	Ontology Prefix	Ontology Full Name	OBO	List of Terms
1	AEO	Anatomical Entity Ontology	L	 
2	AERO	Adverse Event Reporting Ontology	L	 
3	AGRO	Agronomy Ontology	L	 
4	APD	Ascomycete phenotype ontology	L	 
5	APOLLO_SV	Apollo Structured Vocabulary	N	 
6	BCGO	Beta Cell Genomics Ontology	L	 
7	BCO	Biological Collections Ontology	L	 
8	BFO	Basic Formal Ontology	F	 
9	BFO11	Basic Formal Ontology (BFO) 1.1	L	 
10	BILA	Bilateria anatomy	N	 
11	BSPQ	Biological Spatial Ontology	L	 
12	BTO	BRENDA tissue / enzyme source	L	 
13	CARO	Common Anatomy Reference Ontology	L	 
14	CDAO	Comparative Data Analysis Ontology	L	 
15	CEPH	Cephalopod Ontology	L	 
16	CHEBI	Chemical Entities of Biological Interest	F	 
17	CHEMINF	Chemical Information Ontology	L	 
18	CHMO	Chemical Methods Ontology	L	 
19	CIO	Confidence Information Ontology	L	 
20	CL	Cell Ontology	L	 
21	CLO	Cell Line Ontology	L	 
22	CMO	Clinical measurement ontology	L	 
23	CTCAE-OAEView	OAE CTCAE view	N	 
24	CTENO	Ctenophore Ontology	L	 

..... <http://www.ontobee.org/>

Hundreds of ontologies in LifeScience space versus 1.... In Materials Space
PubChemRDF (Substances)

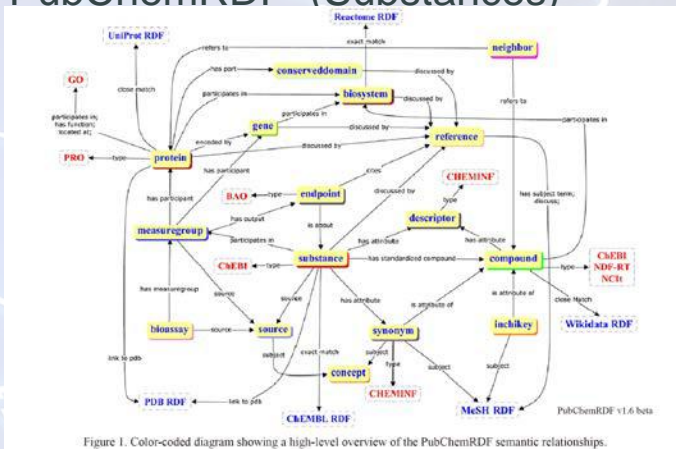


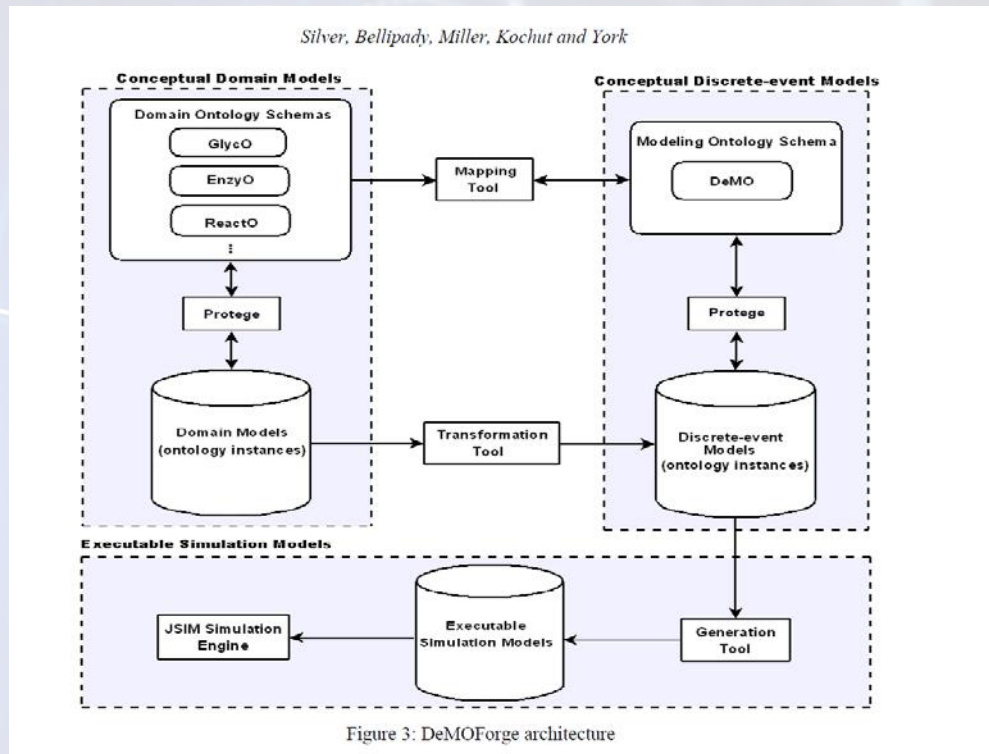
Figure 1. Color-coded diagram showing a high-level overview of the PubChemRDF semantic relationships.

<https://pubchem.ncbi.nlm.nih.gov/rdf/>

No open-source available material science ontology / publications

Connecting Modeling Ontologies to Domain Ontologies

Example in Biochemical domain



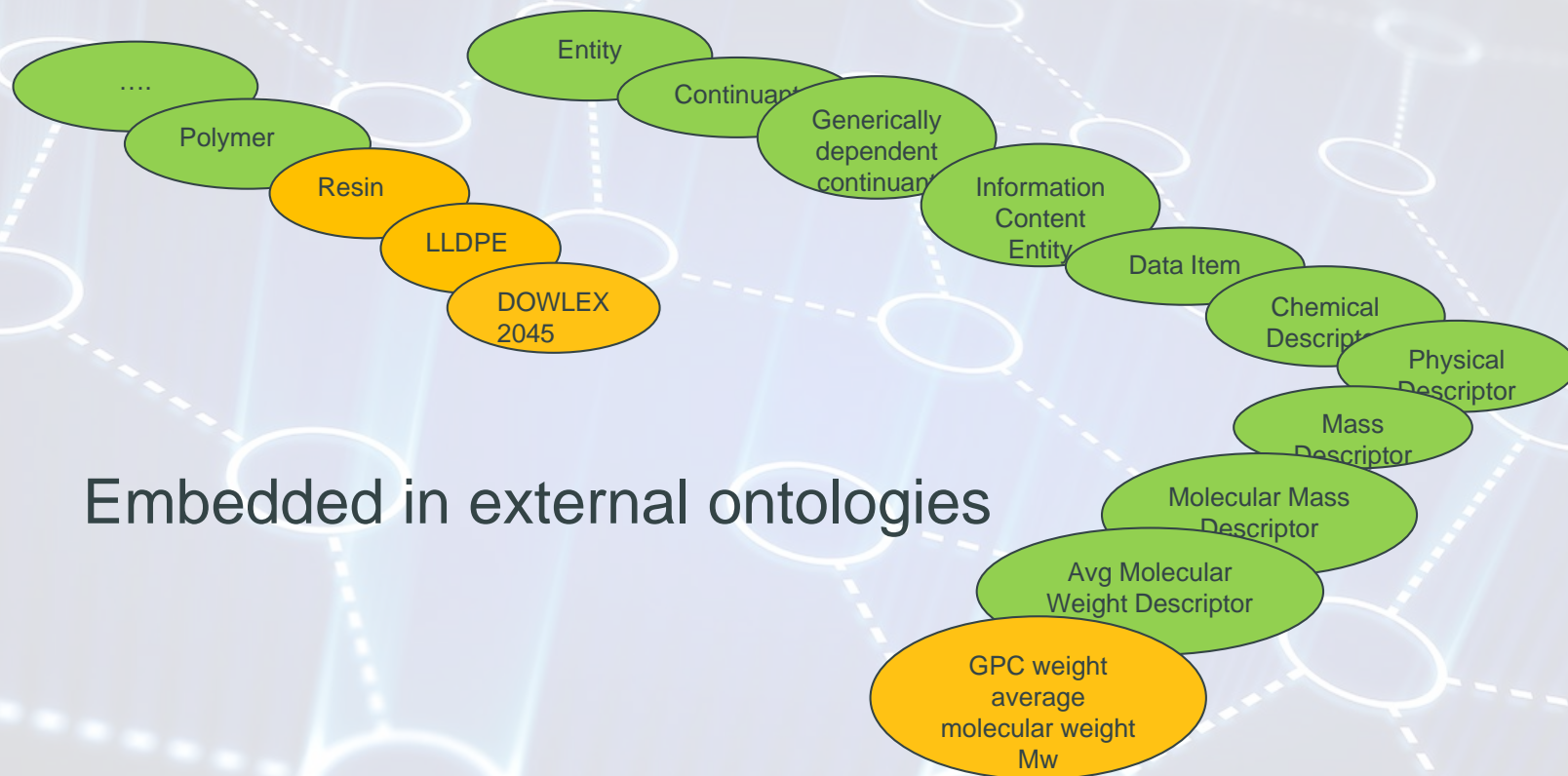
SUPPORTING INTEROPERABILITY USING THE DISCRETE-EVENT MODELING ONTOLOGY (DeMO) - 2009 IEEE - Gregory A. Silver e.a

Methodology

Design Dilemmas and choices made with justifications.

- PubChem RDF as a baseline due to the availability of data rather than any particular preference for an encoding.
- BFO as a baseline due to use by PubChem RDF, adoption by OBO Foundry
- Strict adherence to W3C standards to prevent vendor lock-in and maintain public/private interoperability. Avoidance of property stores Vendor-specific features.

Domain Specific Ontology



Embedded in external ontologies

Inference & Recommenders

Inference: Inferred classes of customer relevant applications based on the asserted material performance attributes.

Mereotopological Models: need to include into ontologies

-> Models: Solve performance requirements in the context of article structure (e.g. morphology and adjacency of materials)

Recommender: Recommend Materials/Structures/Morphologies to meet performance requirements

Gap

Public Material Science Ontology, including the Modeling Ontology:

- No place to add customer application performance indicators
- Difficult to achieve innovation by inferencing
- Recommenders are difficult to deliver
- Difficult to make full use of external knowledge / data