Ontokin: Building a chemistry knowledge base and services for bespoke chemical models in the process industry

Dr Daniel Nurkowski and Dr Amit Bhave

CMCL Innovations

Materials Ontology Workshop organised by the EMMC

Brussels, 29 June 2018

CMCL Innovations

CMCL: Computational Modelling Cambridge Ltd.

Mission

Delivering digital engineering software and solutions to industry and academia

Business model

Software | Consulting | Training

Market segments

Powertrains & fuels | Energy & chemicals

Simulation and design software supplier to industry and academia

Over a decade in innovative R&D and advanced engineering services

SME with an organically growing experienced team

www.cmclinnovations.com

Context:

- We specialise in digital engineering workflows that combine physico-chemical simulation with advanced statistical algorithms
- Ontology engineering forms the basic foundation (“the zeroth step”) in order to harness the data semantics, model interoperability, knowledge base and decision making
- **Use case:** Leveraging Ontokin, a chemical kinetics ontology developed and represented in OWL to offer on-demand chemical models and simulation services

Why Ontokin?

- Problems with the current approach
  - Fragmentation/heterogeneity of data sources and tools in industry
  - Inconsistencies in data, loss of information and logical errors
  - An example: Inconsistencies in chemical models
    - Discrepancies (up to 100 kJ/mol) in enthalpies of formation of same species in different models
    - Many orders of magnitude discrepancy in the rates of the same chemical reactions in different models

---

- Ontokin: Formal, explicit specification of a shared conceptualisation
- Ontology representation: Ontokin, formal reasoner-ready using one of the following syntax: OWL/XML, RDF/XML, TTL
- Ontokin→ Chemical knowledge base → Effective decision making

In collaboration with

CoMo GROUP

UNIVERSITY OF CAMBRIDGE

CAM.CREATE C4T Cambridge Centre for Carbon Reduction in Chemical Technology

The Power of Ontology

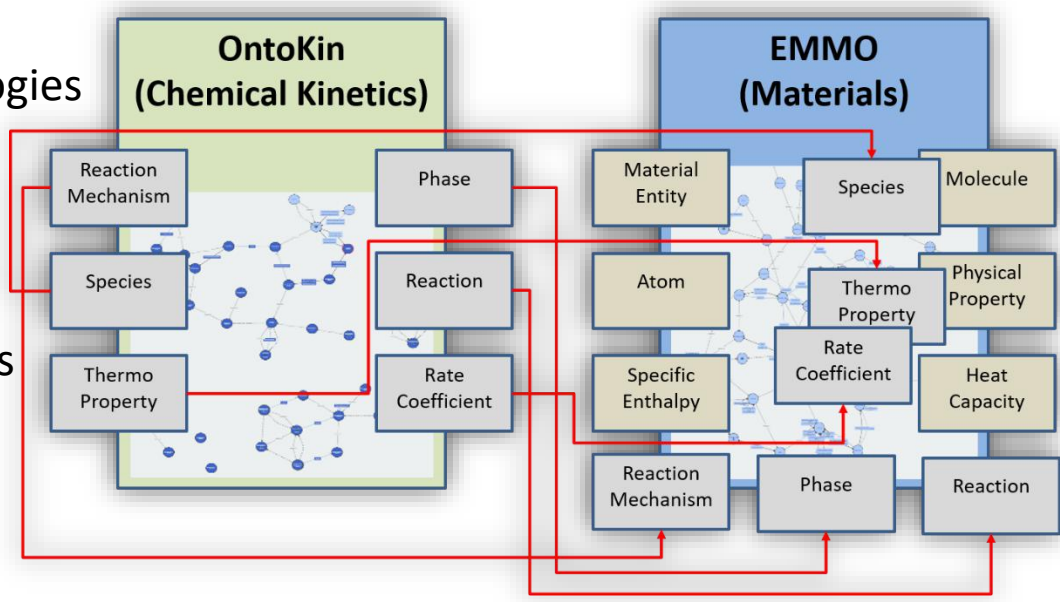
- Ontology sits on top of data to ensure a common meaning
  - Integrate and query heterogeneous resources within and across domains
  - Natural language processing of unstructured literature
  - Data sources enriched with meta-data

- Simplifies software engineering
  - Easy development of services and re-use of exiting resources
  - Advanced query and automation of tasks

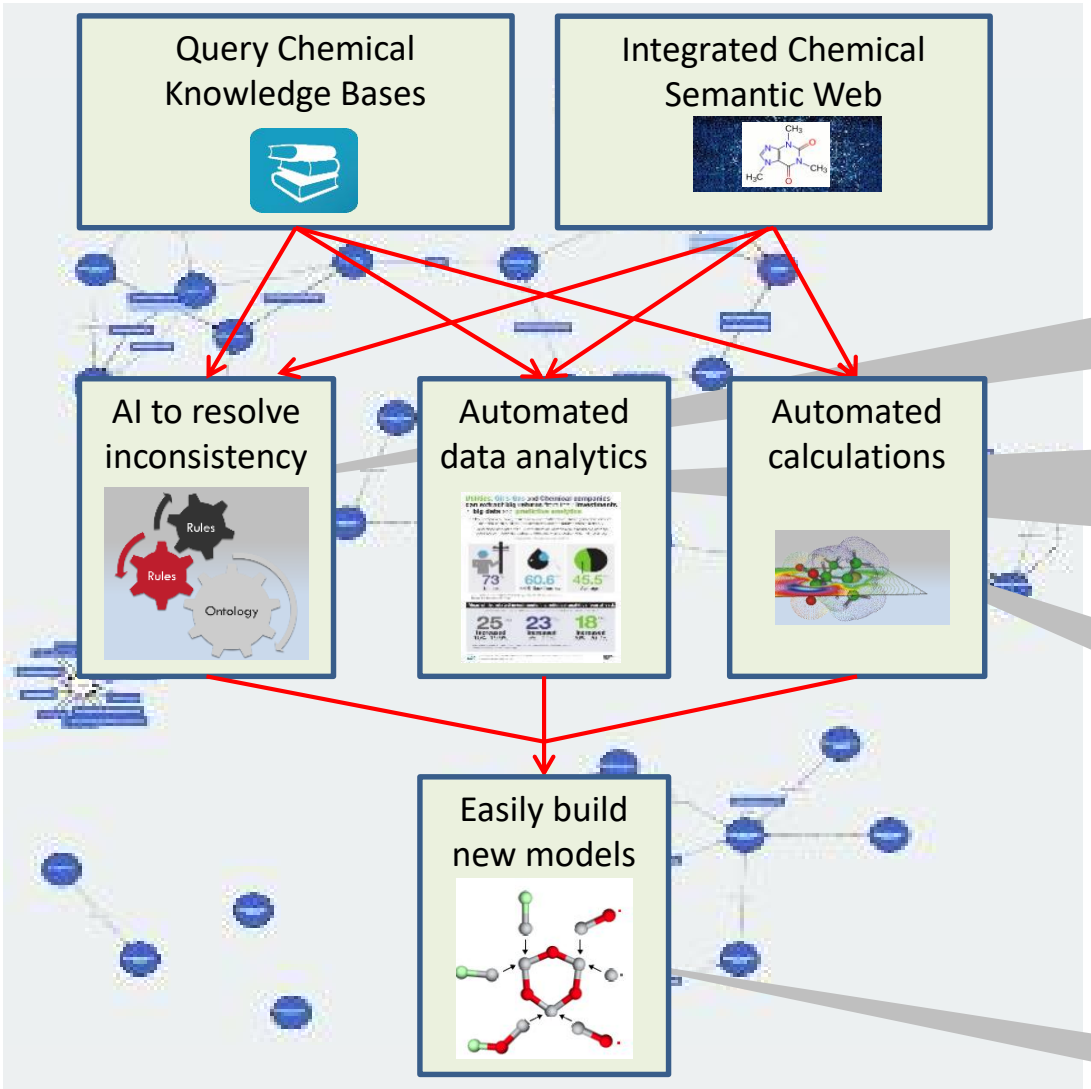
- Extensible
  - Re-use and extend existing ontologies
  - Standardisation via W3C

- Ontokin
  - Data: Computational chemistry calculations, reaction mechanisms

- Utilise machines to automate chemical models – *k*inetics, Cantera, Chemkin, etc.



Purpose & methodology: use-cases for the Process Industry



Ontology to sit on top of data

Services to automate tasks

Curation of knowledge

- Resolve inconsistencies
  - Species data
  - Reaction pathways
- Assessment of new data
  - Millions of compounds published every year
  - New properties added to existing data
- Automated calculation
  - Thermochemical data
  - Reaction rates

Automated bespoke models

- Development and verification of models for industry