iDMT
Innovation Centre in Digital Molecular Technologies

Access to data in a fully digital chemical lab

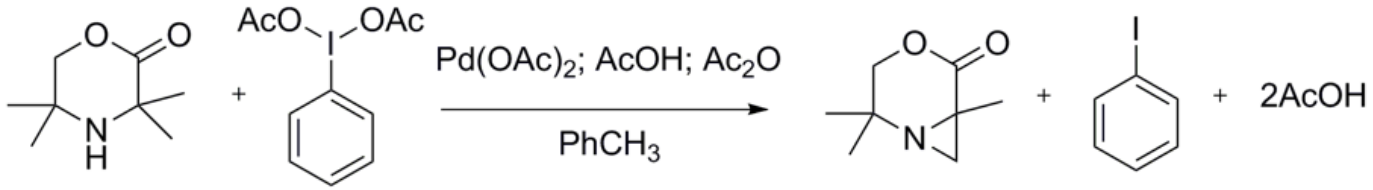
Alexei Lapkin
8/06/2021

EUROPEAN UNION
European Regional Development Fund

UNIVERSITY OF CAMBRIDGE
AstraZeneca
SHIONOGI

What is Chemical Data

- Human readable vs machine readable chemical information




```
RInChI=0.03.1S/C10H11IO4/c1-8(12)
14-11(15-9(2)13)10-6-4-3-5-7-10/h3-7H,
1-2H3!C8H15NO2/c1-7(2)5-11-6(10)8(3,4)9-7/h9H,
5H2,1-4H3<>C2H4O2/c1-2(3)4/h1H3,(H,3,4)!C6H5I/
c7-6-4-2-1-3-5-6/h1-5H!C8H13NO2/c1-7
(2)5-11-6(10)8(3)4-9(7)8/h4-5H2,1-3H3<>2C2H4O2.
Pd/c2*1-2(3)4;/h2*1H3,(H,3,4);/q;;+2/p-2!C2H4O2/
c1-2(3)4/h1H3,(H,3,4)!C4H6O3/c1-3(5)7-4(2)
6/h1-2H3!C7H8/c1-7-5-3-2-4-6-7/h2-6H,1H3/d+
```

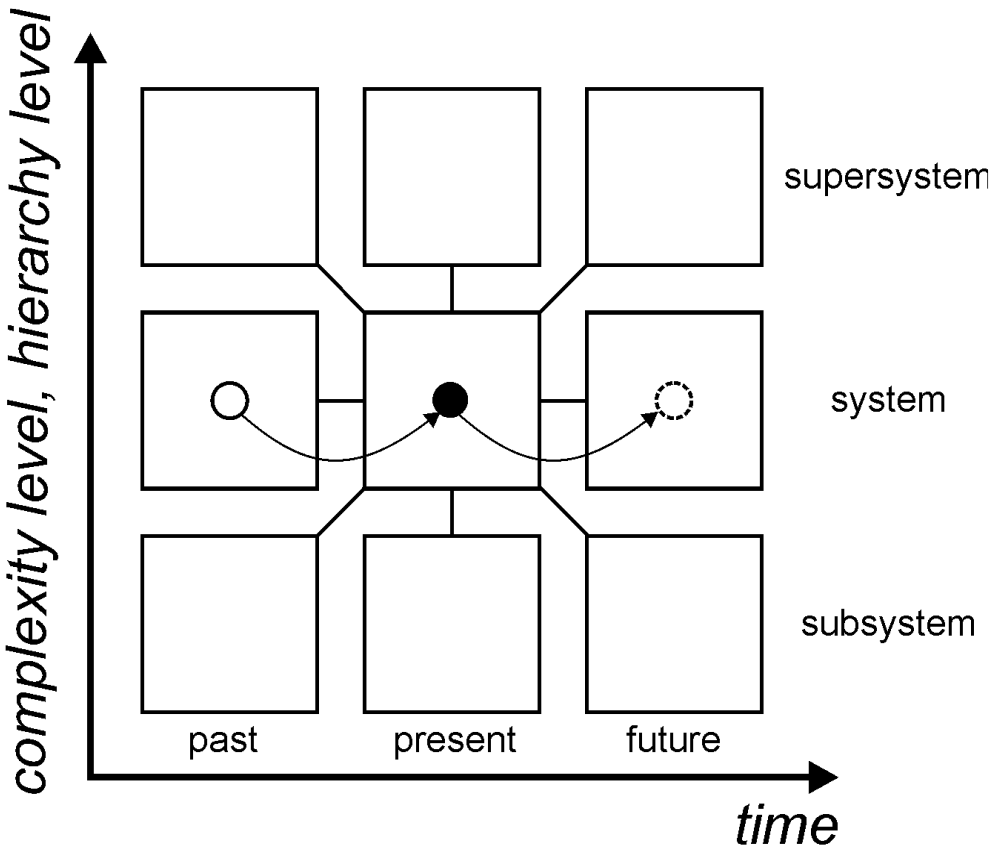
```
ProcAuxInfo=$$PAI0.01|1:2|1;1|2;1;
1|393|6E5|60:0.06;120:0.14;180:0.20;
240:0.32;300:0.40;360:0.52;420:0.70;
480:0.90;540:1.0;600:1.0|?;?;0.90|8.3E-7;
8.3E-7|0;0;0|4.2E-9:m:f;8.3E-6:m:f;
1.7E-6:m:f;1.5E-4:m:f|1E-5
```

Jacob et al. J Cheminform (2017) 9:23

- Reactants (nature and relevant properties)
- Products (nature and relevant properties)
- Reagents (solvent, catalyst, ligand, magic dust…)
- Reaction conditions (T, P, residence time, etc)
- Yield of main product as a function of time

iDMT Innovation Centre in Digital Molecular Technologies

Chemical Reaction as a System



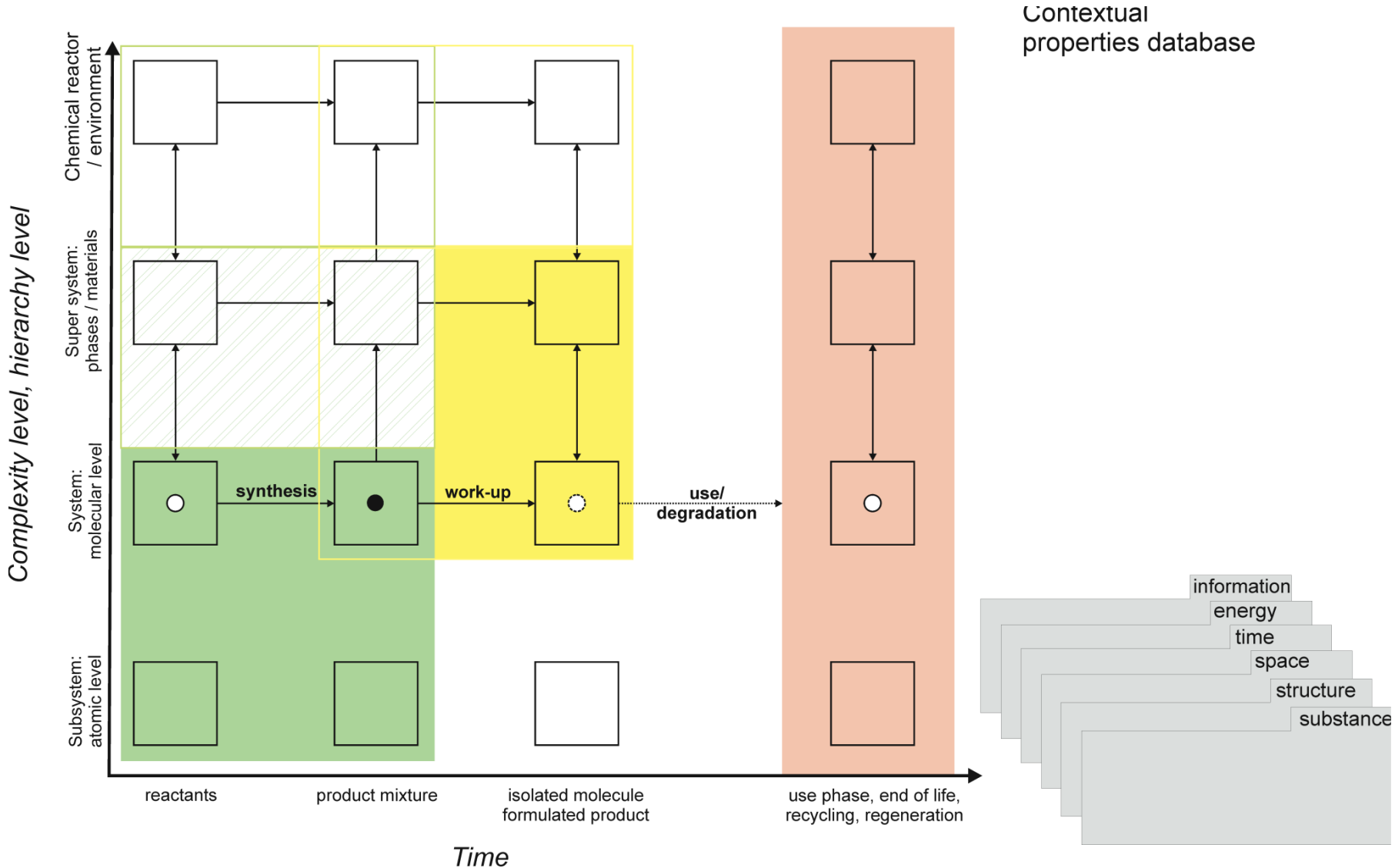
Description of interactions within complex hierarchical systems requires understanding of their intrinsic and emerging properties and their relation to scale and time.

The same description can be applied to a chemical reaction:

- Invariant properties of molecules
- Context-dependent properties of molecules
- Properties & phenomena dependent on time
- Properties & phenomena dependent on scale

iDMT Innovation Centre in Digital Molecular Technologies

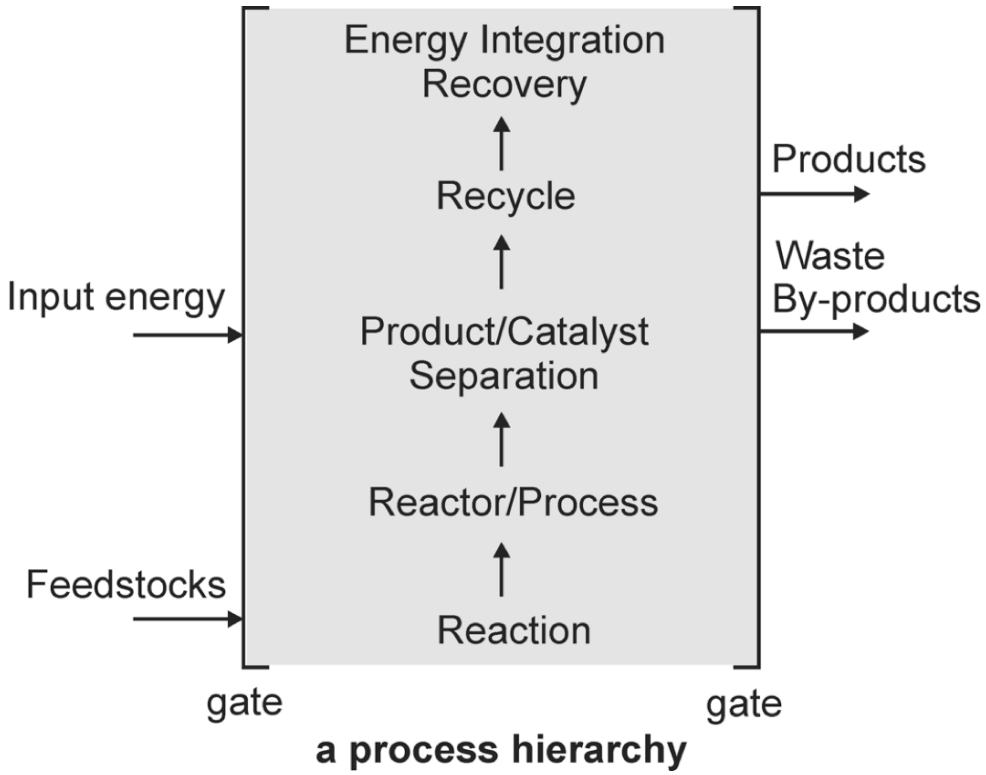
Chemical Reaction Context



Chemical Engineering & Processing. Process intensification, 50 (2011) 1027-1034

iDMT
Innovation Centre in Digital Molecular Technologies

Process as a Hierarchical System



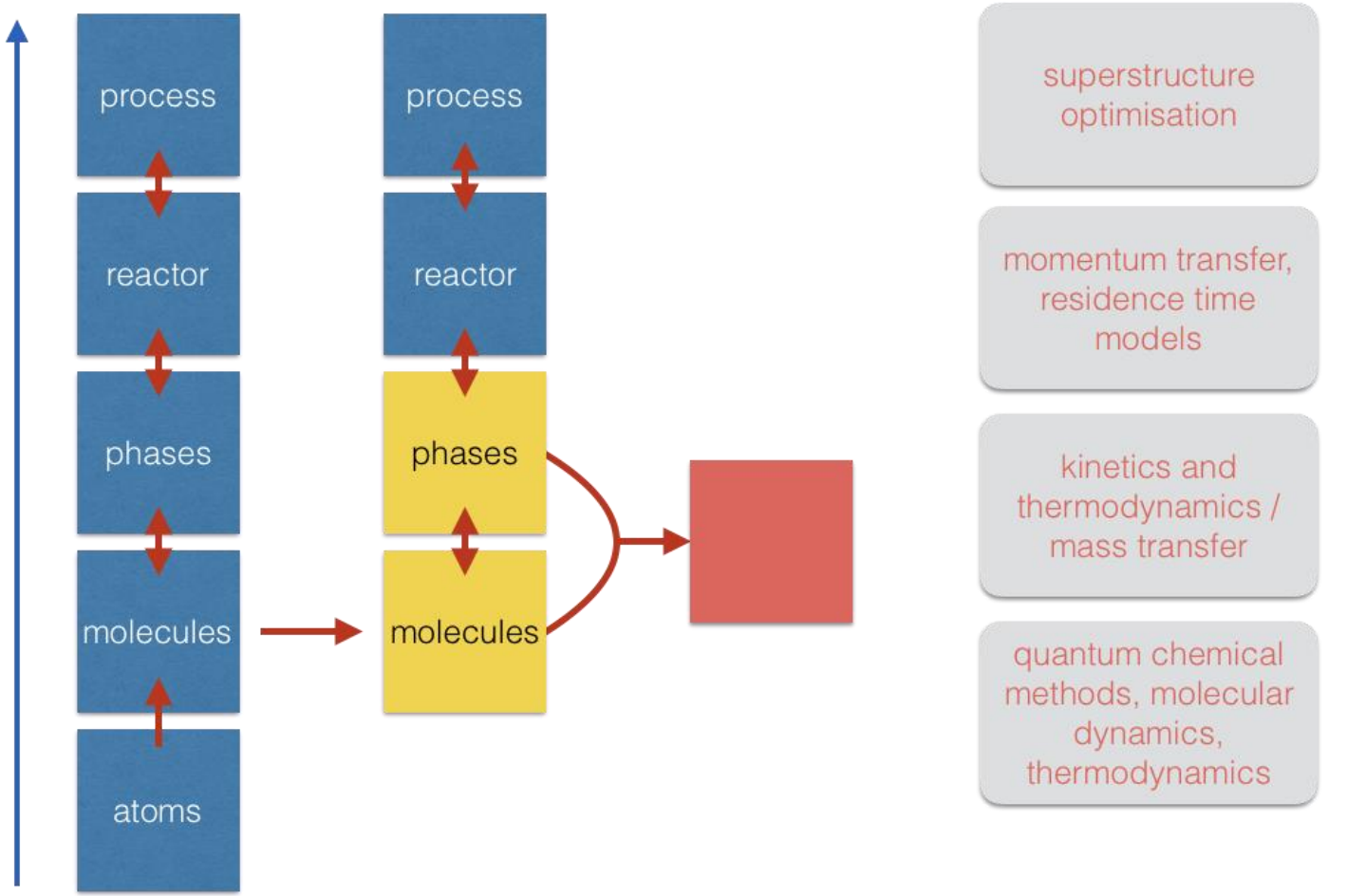
a process hierarchy

Data requirements for describing a process:

- All chemical species and their physico-chemical properties (under conditions in storage, feed, all process stages and final storage – multiple instances of specific 'context').
- All reactions (pathways [mechanisms], kinetics).
- Mass transfer (phases, rates of mass transfer in specific equipment).
- Heat transfer (phases, properties of fluids, rates of heat transfer in specific equipment).
- Impact of hydrodynamics on any properties.
- Neighbouring systems (for mass and energy integration).

iDMT Innovation Centre in Digital Molecular Technologies

Process Context = Reaction + Equipment + Infrastructure



superstructure optimisation

momentum transfer, residence time models

kinetics and thermodynamics / mass transfer

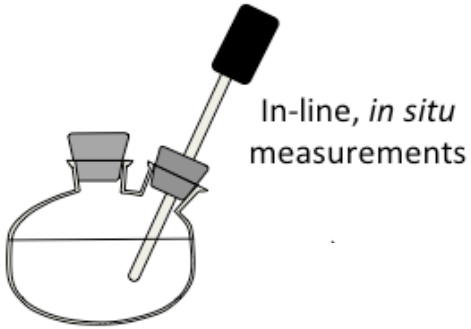
quantum chemical methods, molecular dynamics, thermodynamics

Reactants ***Products*** ***Use phase*** → *End of Life*

iDMT Innovation Centre in Digital Molecular Technologies

How to Manage Data in a Fully Digital Lab

Experiment Coding
Translation of DoE into automated experiment formulation
Equipment ontology



Automated experimental system

Data Ontology
Classification of experimental observations

Chemical Data:
Reaxys
ELNs
QC results
Reaction mechanisms
Thermodynamic properties
Molecular descriptors
Environmental impact factors
Toxicity data
….

Chemical Ontology
Relationships between data elements

Chemical Knowledge:
Kinetic model hypothesis
Formulation properties hypothesis
Bio-activity model
Reactivity pattern hypothesis
….

iDMT Innovation Centre in Digital Molecular Technologies

What Should iDMT Build?

Innovation Centre in Digital Molecular Technologies (iDMT) is a project co-funded by the European Regional Development Fund (ERDF), AstraZeneca, Shionogi and the University of Cambridge.

iDMT is an open innovation platform for collaborative R&D projects in the areas of:

- Artificial Intelligence in Molecular Technologies
- Robotic Equipment for Chemical Synthesis
- Algorithms and Tools for Digital Process Development

- High-throughput experiments will generate data (reaction outcomes)
- Formulation robot will generate data (formulation properties)
- Flow synthesis robots will generate data (optimal reaction conditions)
- QC methods and MD methods will generate data
- ML / AI algorithms will use the data
- What is the architecture for storing and maintaining the data?

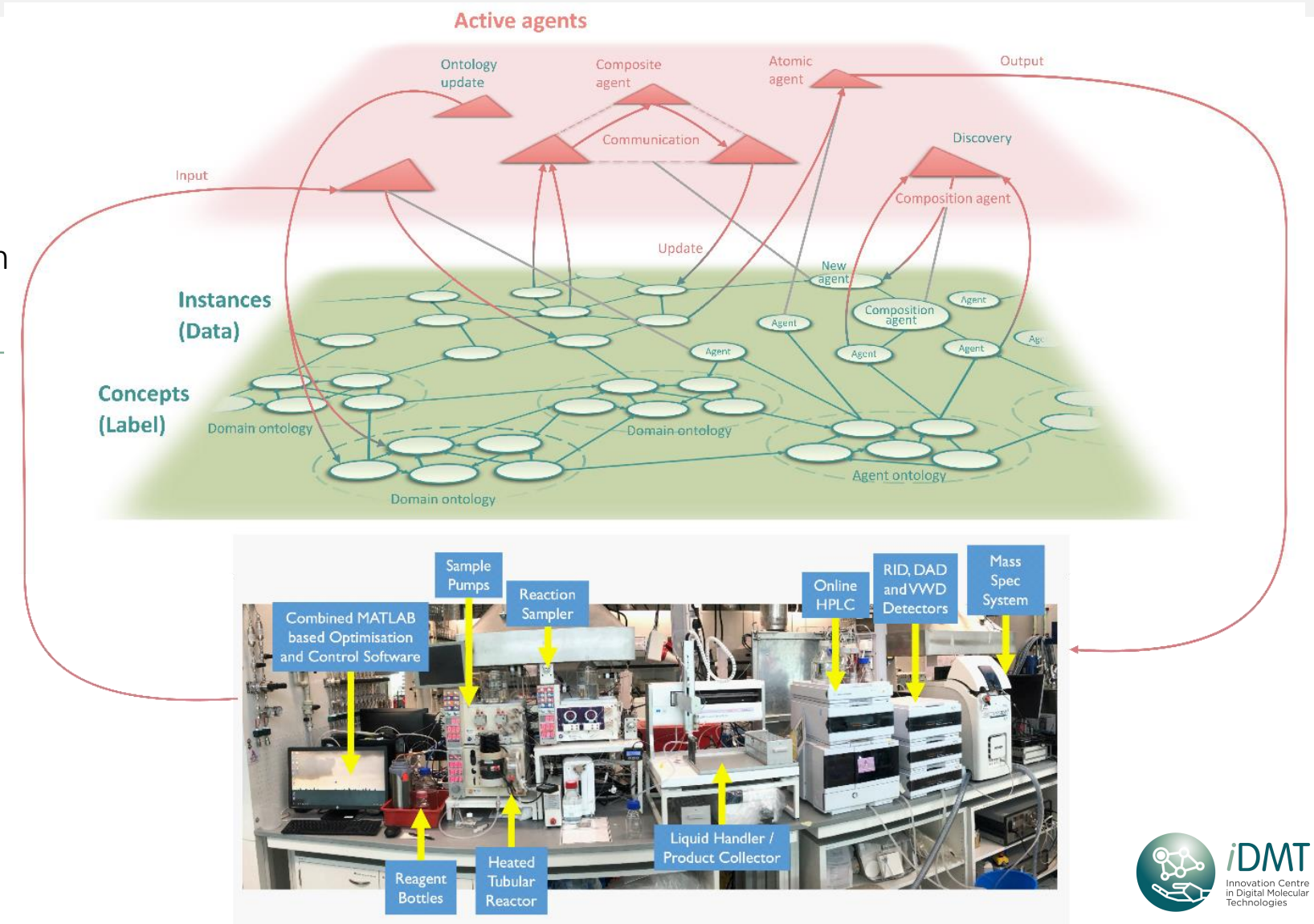
iDMT Innovation Centre in Digital Molecular Technologies

COMO/CARES/PIPS Knowledge Graph Technology

Who has the knowledge to develop such architecture

Partnership:

- Prof. Markus Kraft / COMO Group
- PIPS Data2Knowledge Project

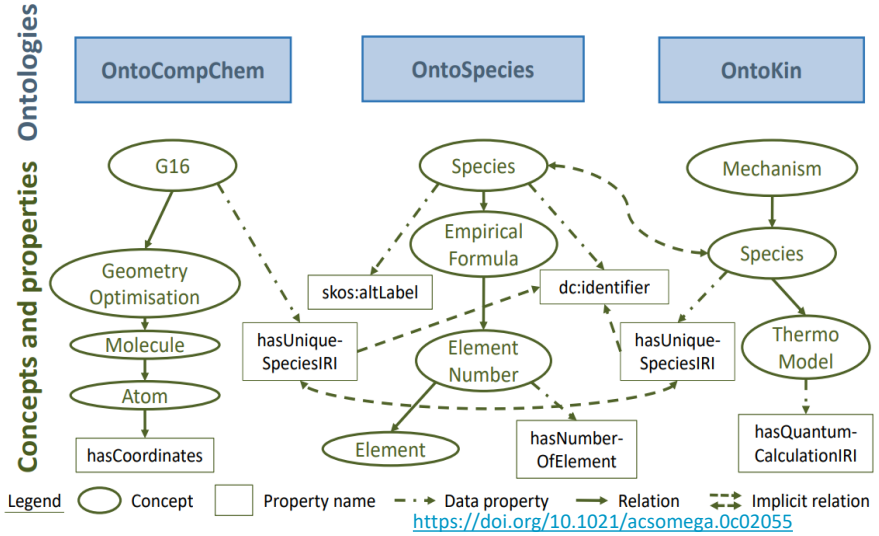


iDMT Innovation Centre in Digital Molecular Technologies

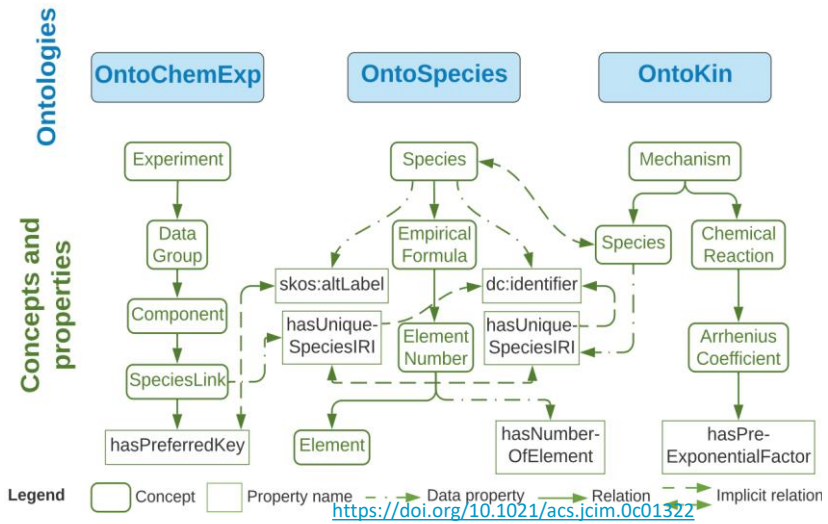
COMO/CARES Knowledge Graph Technology

- OntoSpecies – chemical species identification https://doi.org/10.1016/j.compchemeng.2020.106813
- OntoKin – chemical kinetic mechanisms https://doi.org/10.1021/acs.jcim.9b00960
- OntoCompChem – quantum chemistry calculation results https://doi.org/10.1021/acs.jcim.9b00227
- OntoChemExp – chemistry experiment data https://doi.org/10.1021/acs.jcim.0c01322

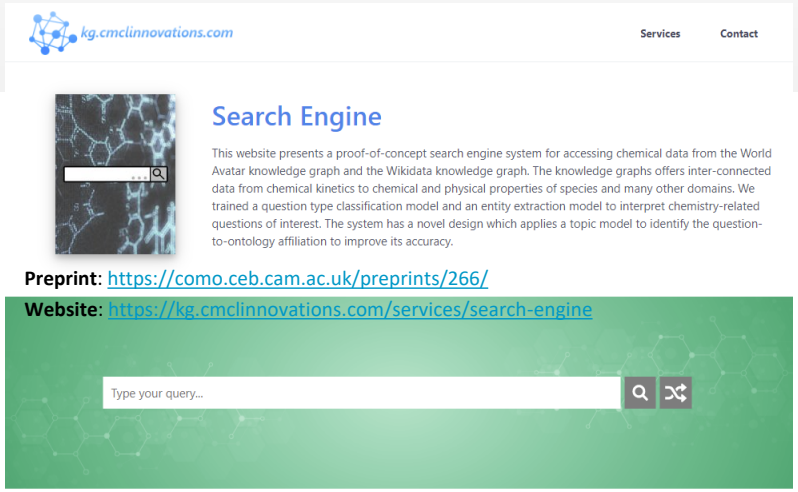
- All above ontologies are linked via OntoSpecies for unique chemical species identification



https://doi.org/10.1021/acsomega.0c02055



https://doi.org/10.1021/acs.jcim.0c01322



Preprint: https://como.ceb.cam.ac.uk/preprints/266/

Website: https://kg.cmclinnovations.com/services/search-engine

Example Questions

- General Computational Quantum Chemistry
- Kinetics & Thermodynamic
- Reactions & Mechanisms
- Molecule Classes
- Conditional Queries
- Queries by SMILES
- Molecule Properties

This content was developed, tested, and documented by the following commercial and academic partners. For more information on each institution, please follow the links listed below.

CMCL Innovations
cmclinnovations.com

CoMo GROUP
Computational Modelling Group
como.ceb.cam.ac.uk

CAMBRIDGE CARES
Cambridge CARES
cares.cam.ac.uk

iDMT
Innovation Centre in Digital Molecular Technologies

iDMT
Innovation Centre in Digital Molecular Technologies

Project “Innovation Centre in Digital Molecular Technologies” is part-funded by European Regional Development Fund

EUROPEAN UNION
European Regional Development Fund