



# The European Materials Modelling Council

Data sharing across the NMBP  
programme via Industry Commons:  
from ontologies to marketplaces.

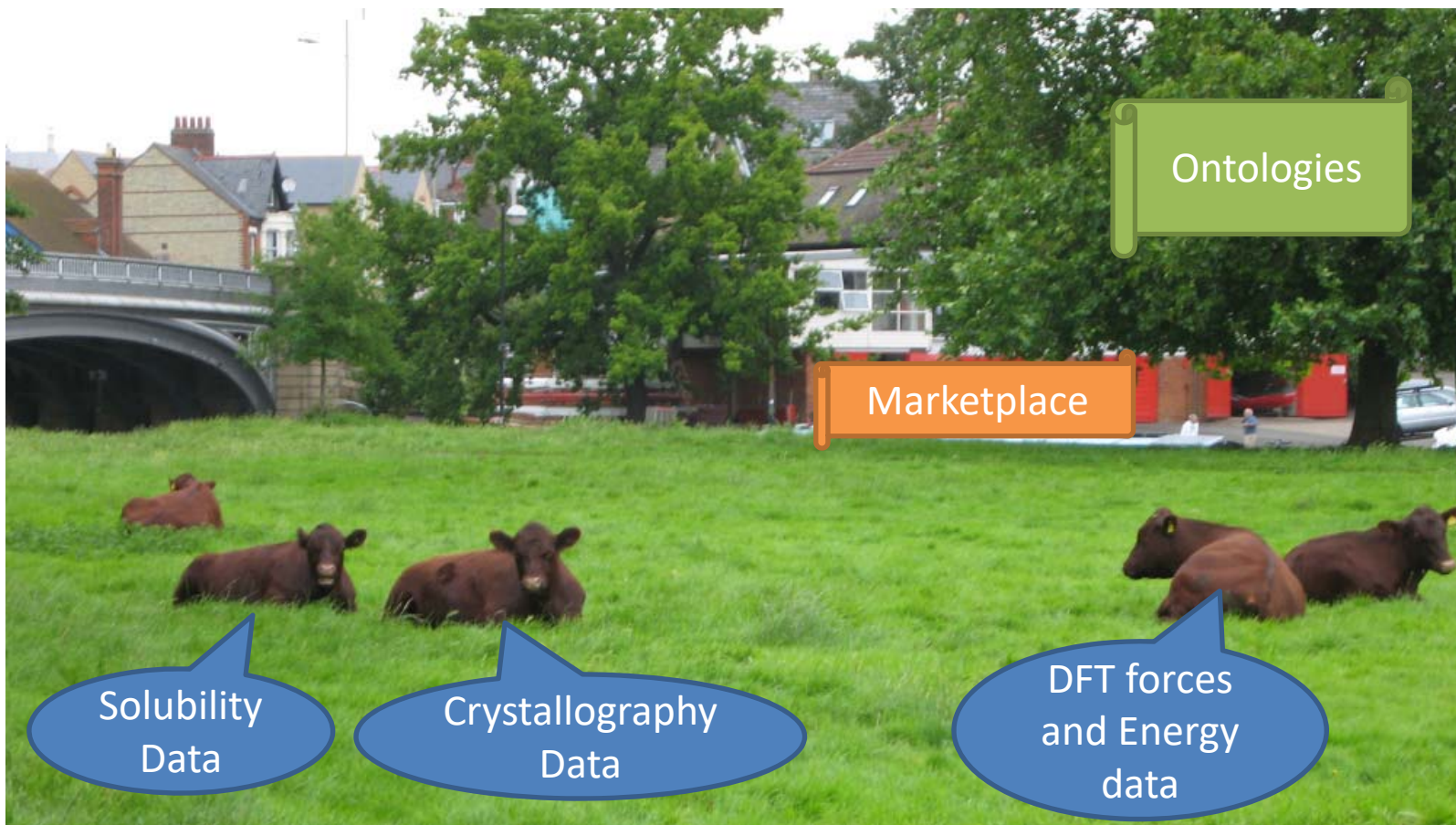
Alexandra Simperler and Gerhard Goldbeck



EuroNano2019 Bucharest, 12.6.2019



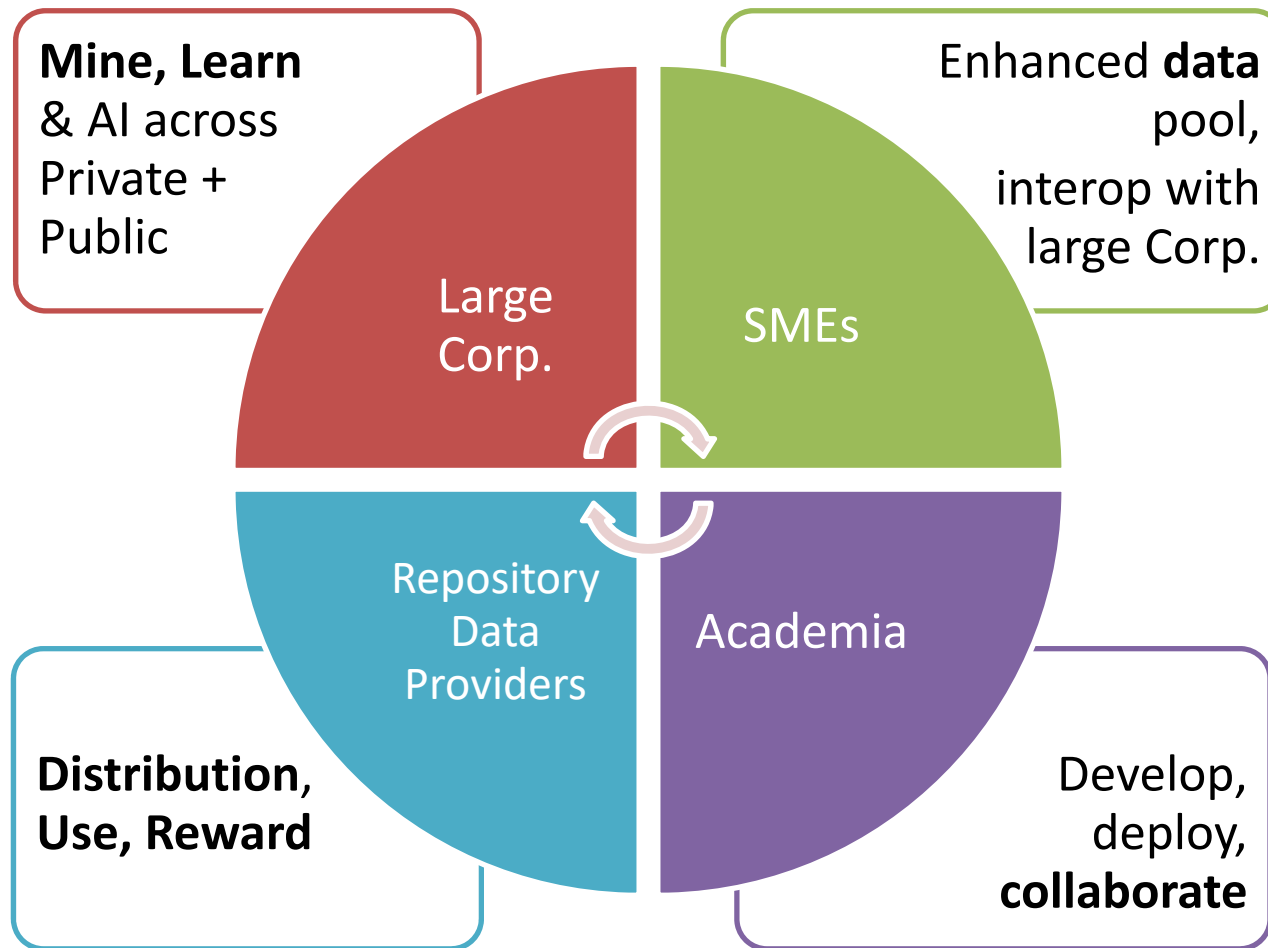
# Commons



Midsummer Common, Cambridge is registered pursuant to the provisions of the [Commons Registration Act 1965](#): registration of (often historical) rights to common land, town greens, and village greens in England and Wales.



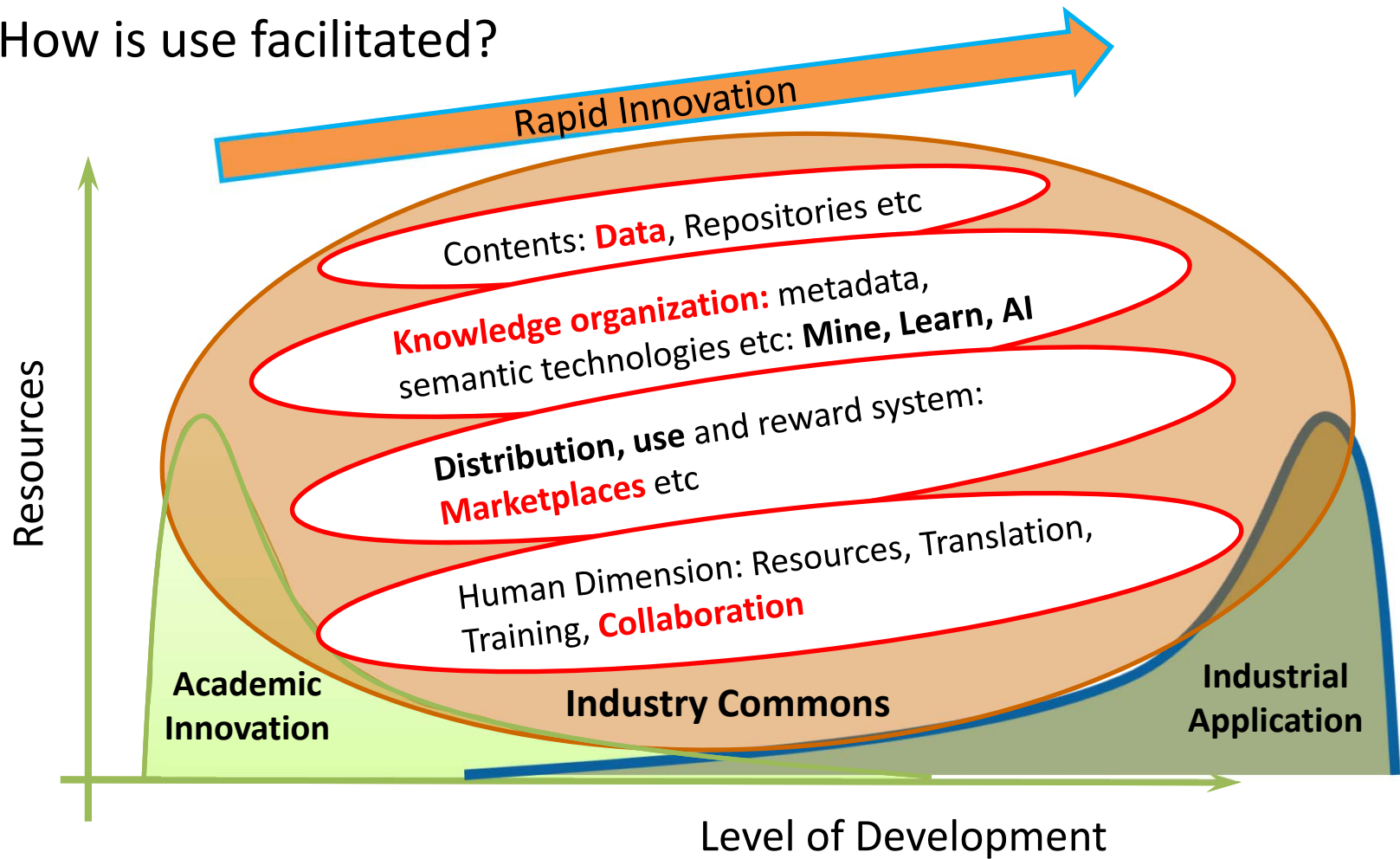
# Commons: value proposition



The Internet of Materials: An AI-Enabled Data Economy for Materials Discovery, E Pyzer-Knapp et al (IBM Research UK and STFC Hartree Centre)

# Industry Commons

- How are they ‘registered’?
- How is use facilitated?





# Benefits vs Challenges

## Benefits

- **Sustainable** collaboration and integration platform
- Integration of knowledge for **accelerated and targeted innovation**
- **Stronger collaboration** and open innovation
- **Synergies** with A. I. , Machine Learning, Big-Data and Open Science Cloud initiatives

## Challenges

- Continuous **production of data** is **not synonymous** with **knowledge**
- **Isolated** communities, **siloed** data
- Terminology **confusion**
- **Lack of standards** and **curation** infrastructure
- **Lack of interoperability**



# Challenge: Data

- 1000s of projects
- Flood of data, Less than 1% of data collected is analysed for value contained within it
- Data in repositories but no common 'structure'
- No interoperability even if 'connected'
- Where is the Commons?
- **BUT: Data is key to speeding up materials development. Developing materials faster will require managing and using data more effectively**, which includes consolidating data into a single consistent searchable format, as well as **structuring**, storing, and using materials data to harness the power of artificial intelligence.

*Source:* The Internet of Materials: An AI-Enabled Data Economy for Materials Discovery, E Pyzer-Knapp et al (IBM Research UK and STFC Hartree Centre), Feb 2018.  
Courtesy of InnovateUK.



# Solutions: Data

<https://www.rd-alliance.org/international-materials-resource-registries-wg-report>

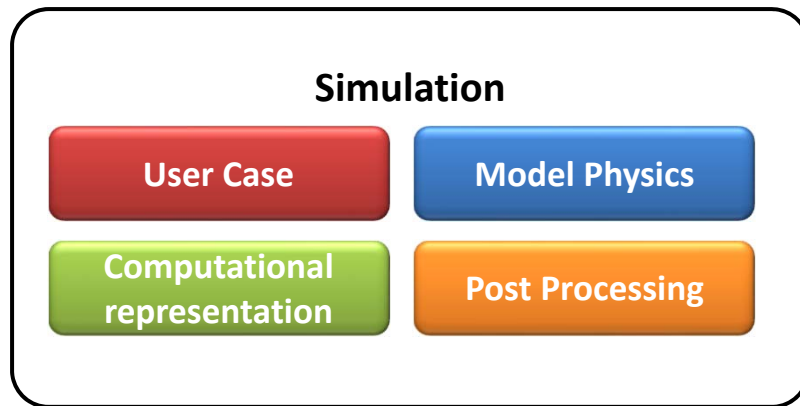
- RDA Working Group Report Mar-2018
  - establish a network of Materials Resource Registries
  - model for a federated registry framework that can support the distributed creation, sharing, and searching of data resource descriptions.
  - extendable resource metadata.
- In response, project SIDAP (including Bayer, Covestro, Evonik) works on a “scalable integration concept”
- The data integration architecture takes care of the **semantics of the different types of data** to achieve a conceptualised, integrated and future-proof data representation that enables interactive analysis.

Source: Chemanager: **“Big Data analysis fails mostly due to unstructured data”**  
<https://www.chemanager-online.com/themen/mess-automatisierungstechnik/sidap-datenintegration-und-datenanalyse>

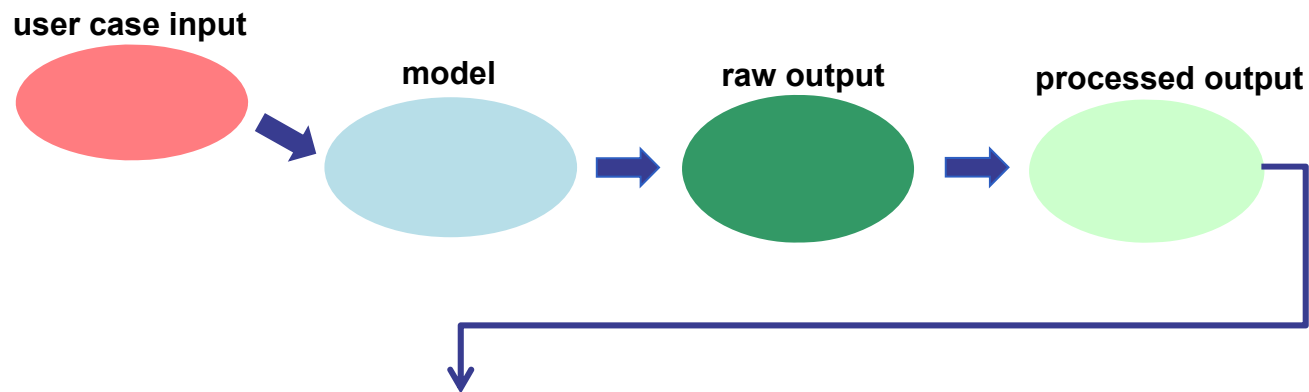


# Solutions: MODA

... an example of an agreed data documentation system and already in wide spread use



Online tool to document and register projects on EMMC.info



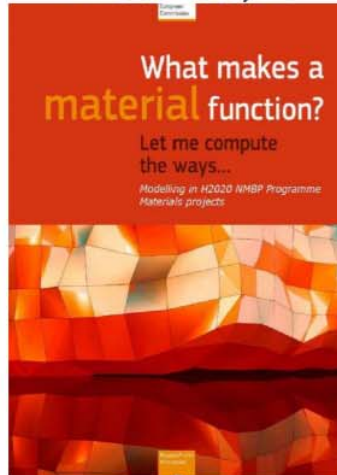




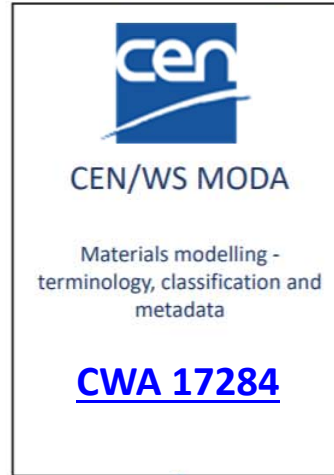
# Process of Standardisation

From Survey of the field (Review of Materials Modelling, RoMM)  
to standardised terminology and documentation (CWA)

RoMM  
Review of Materials Modelling VI  
*Anne de Baas, EC*



CEN Workshop Agreement  
*Endorsed by >15 EU organisation*



EMMC - MODA template

MODA for <user>-case>  
Simulated in project <acronym>

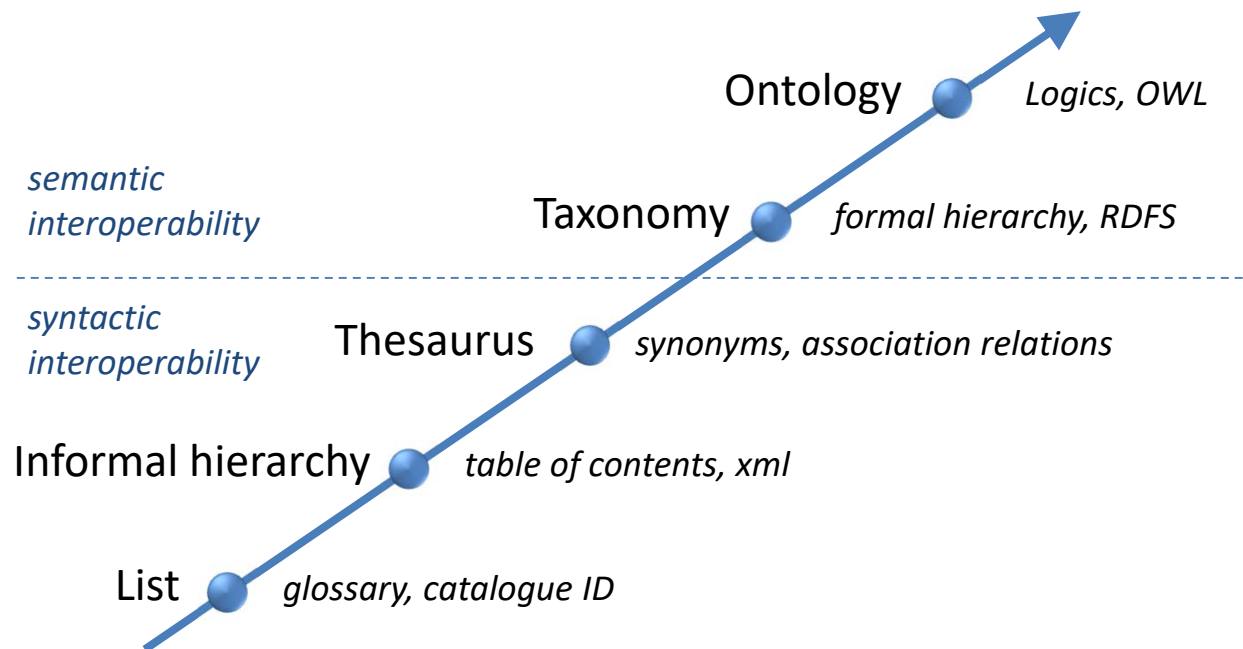
OVERVIEW of the SIMULATION	
1	<b>USER CASE</b> Please give the properties and behaviour of the particular material, manufacturing process and/or on-service behaviour to be simulated. No information on the modelling should appear here. The idea is that this simulation can also be simulated by others with other models and that the results can then be compared.
2	<b>CHAIN OF MODELS</b> Please identify the first model. Note these are assumed to be physics-based models unless it is specified differently. Most modelling projects consist of a chain of models, (workflows). Here only the Physics Equations should be given and only names appearing in the content list of the Review of Materials Modelling VI should be entered. This review is available on <a href="http://www.cen.eu/cen/publications/technical_technological_library.cfm">http://www.cen.eu/cen/publications/technical_technological_library.cfm</a> . All models should be identified as electronic, atomic, mesoscopic or continuum. Model 2 Please identify the second model. Model 3 Please identify the second model. <b>DATA-BASED MODELS</b> If data-based models are used, please specify.
3	<b>PUBLICATION REFERENCES</b> Please give the publication which documents the data of this CWA simulation. This article should ensure the quality of this data set (and not only the quality of the models).
4	<b>ACCESS CONDITIONS</b> Please list whether the model and/or data are free, commercial or open source. Please list the owner and the name of the software or database (include a web link if available).
5	<b>WORKFLOW AND ITS RATIONALE</b> Please give a textual rationale of why you as a modeller have chosen these models and this workflow, knowing other modellers would simulate the same and-user case differently. This should include the reason why a particular aspect of the user case is to be simulated with a particular model.

**EMMO**  
EUROPEAN MATERIALS MODELLING ONTOLOGY



# Challenge: Knowledge Organisation

Semantics and metadata allow a resource to be understood by both humans and machines → promote interoperability.



INTEROPERABILITY !!!

**Machine can interpret information and reason**

Machine can process information due to compatible syntax.

*Ability of computer systems or software to exchange and make use of information*



Existing solutions:

- Target a **wide range of different objectives**
- Lack a **common basis** in the **semantics** of materials
- **Lack generality**
- **Dependent** on adhering to **particular formats**
- Require **input by experts** to go across domain boundaries:  
**laborious and error prone**



# Solutions: Knowledge Organisation

## 1. Terminology and classification

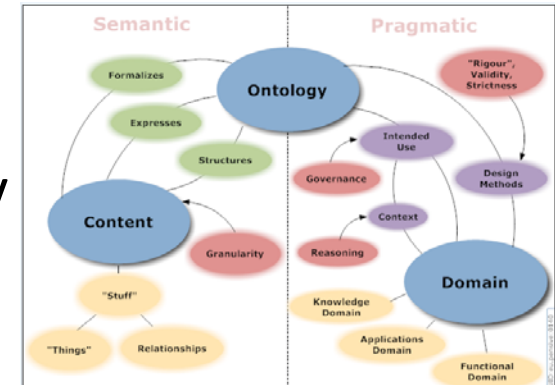
CEN Workshop Agreement



## 2. Taxonomy and Ontology

European Materials & Modelling Ontology

EMMO



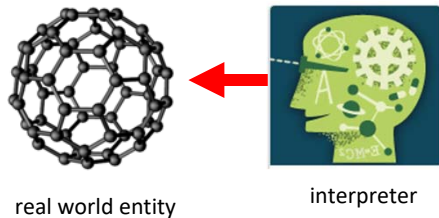
## 3. Open standards for interoperability



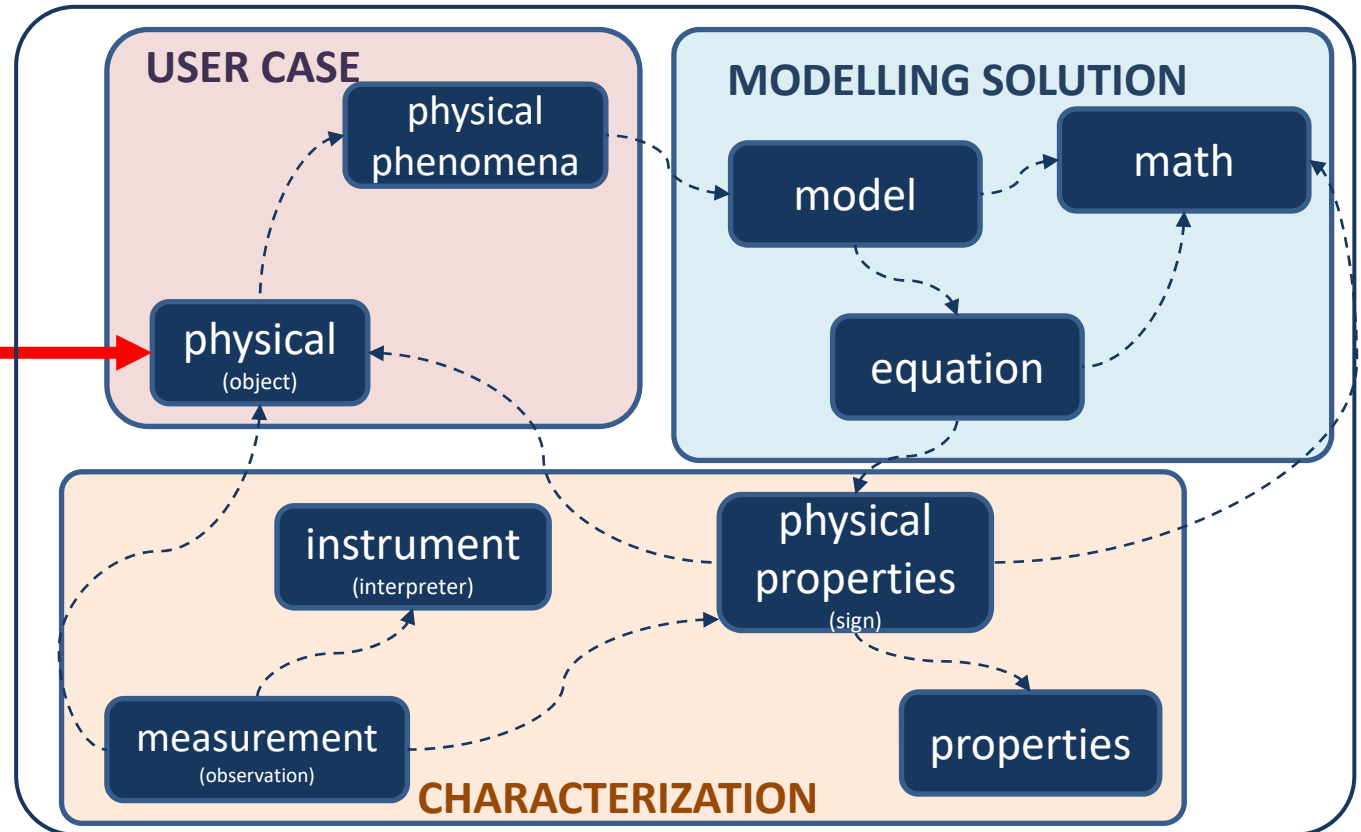


# EMMO Interoperability

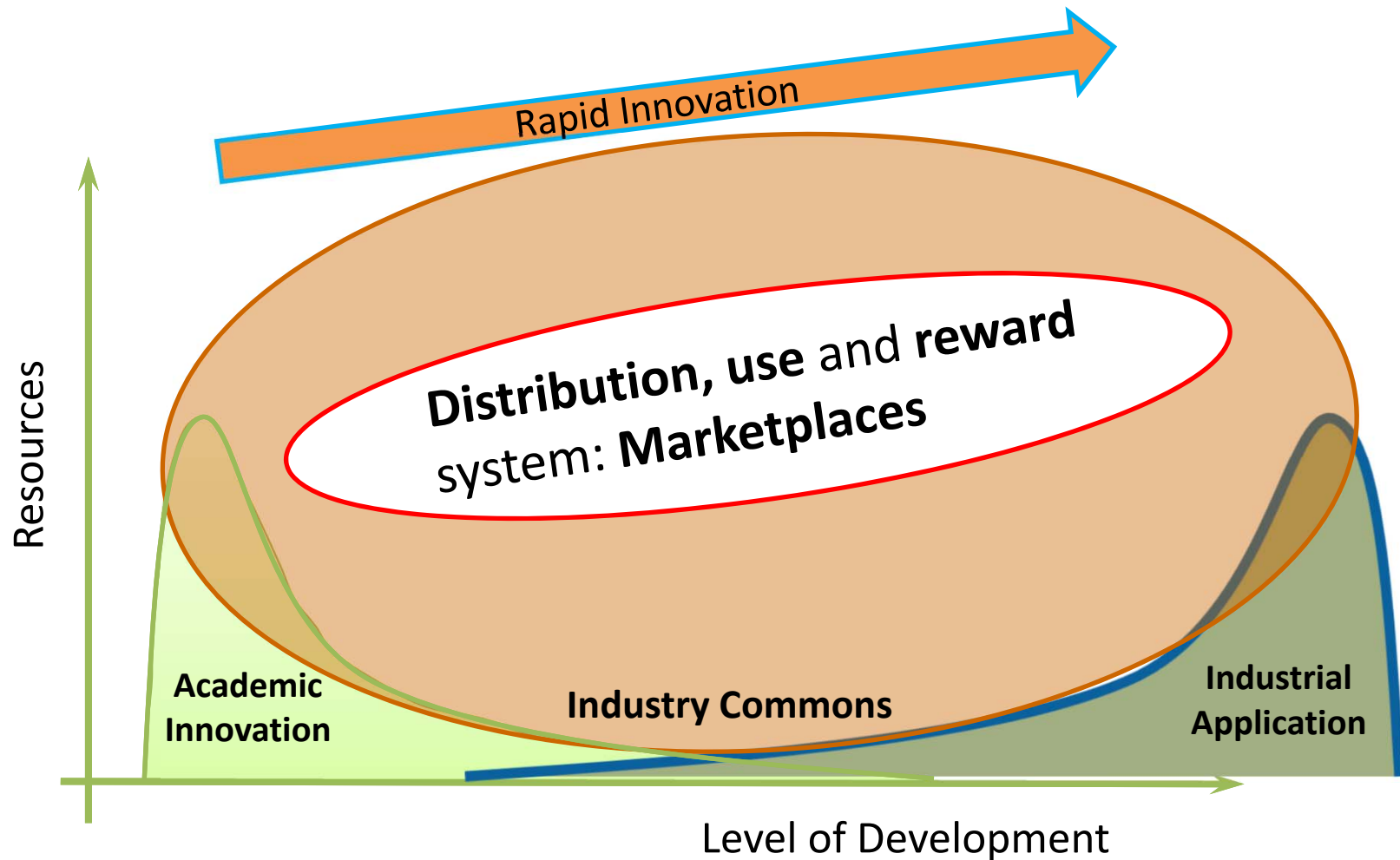
**Horizontal interoperability:**  
one user case,  
multiple modelling  
solutions.



**Linking** between  
properties  
database, models  
and user cases to  
facilitate  
**validation** and  
**data collection.**



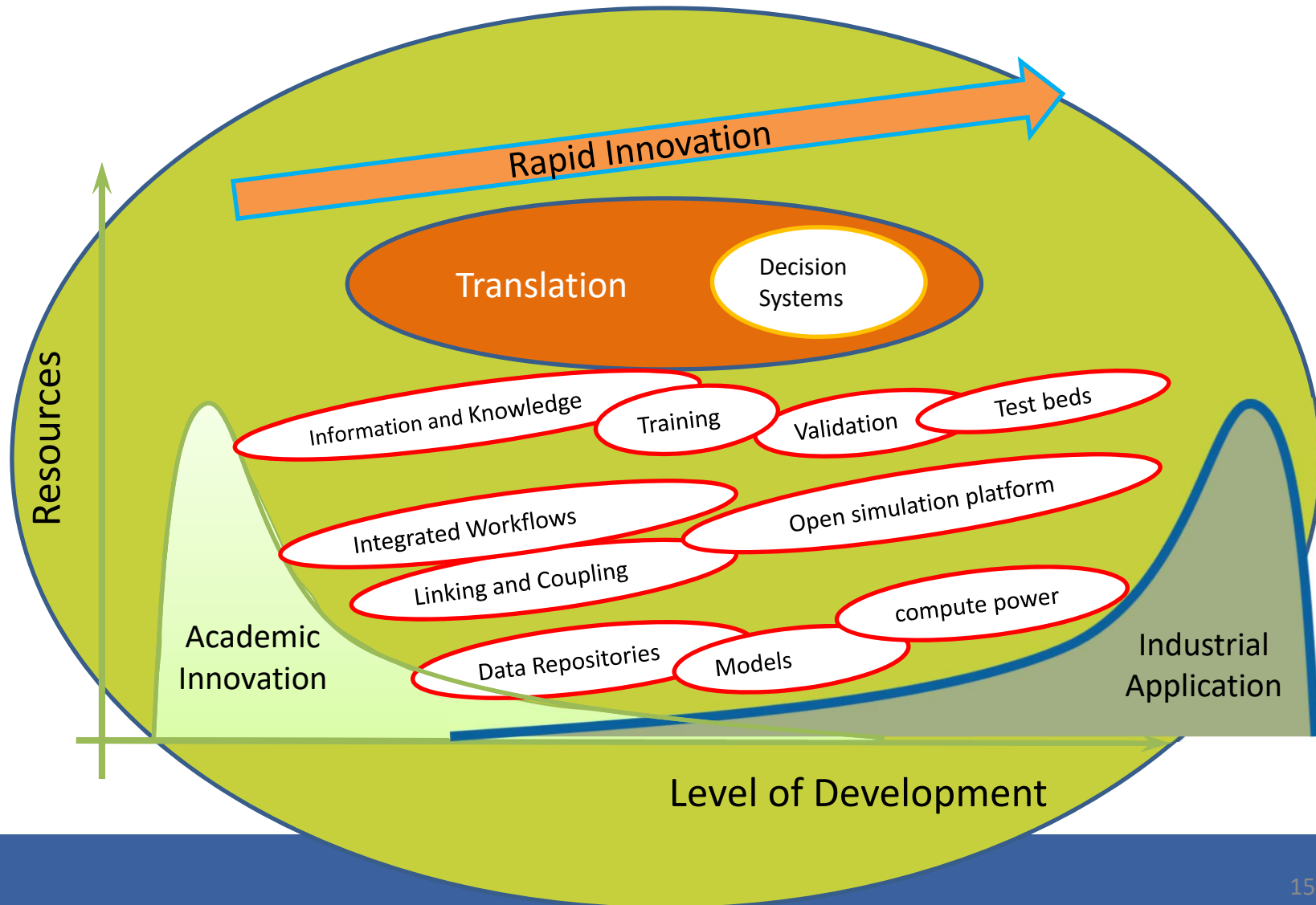
# Challenge and Solution: Marketplaces





# What is available at the materials modelling market place?

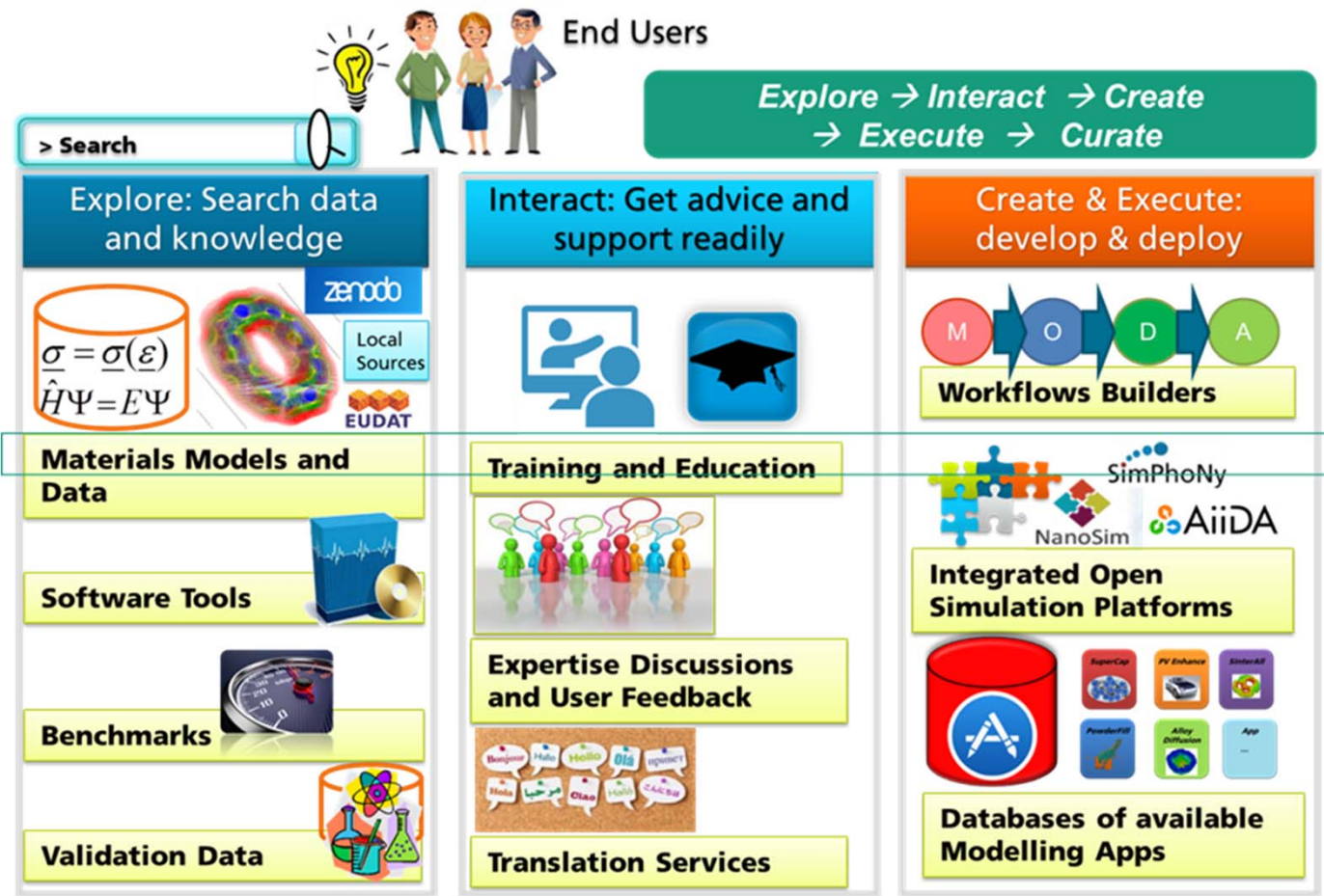
Data, models, tools, translation, training, support etc





# Materials Modelling Marketplaces

## *Integrate the whole ecosystem*



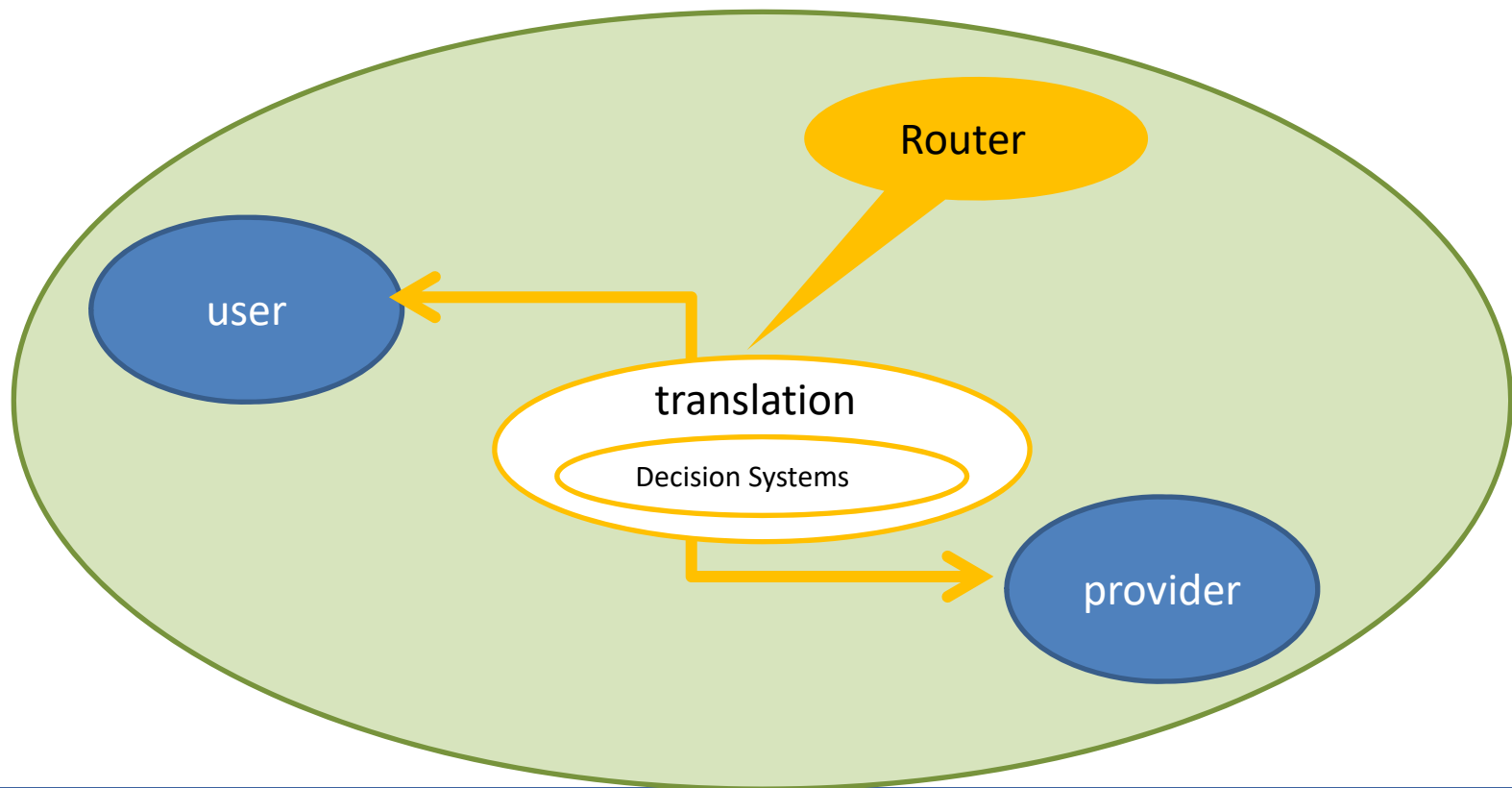
Materials Modelling Marketplaces are true **innovation hubs**

Image courtesy of the MARKETPLACE IA project, GA. No. 760173. © MARKETPLACE



# Marketplace is a collaborative system

- Open to all providers, all users, all applications
- Includes translation and decision systems: Routers

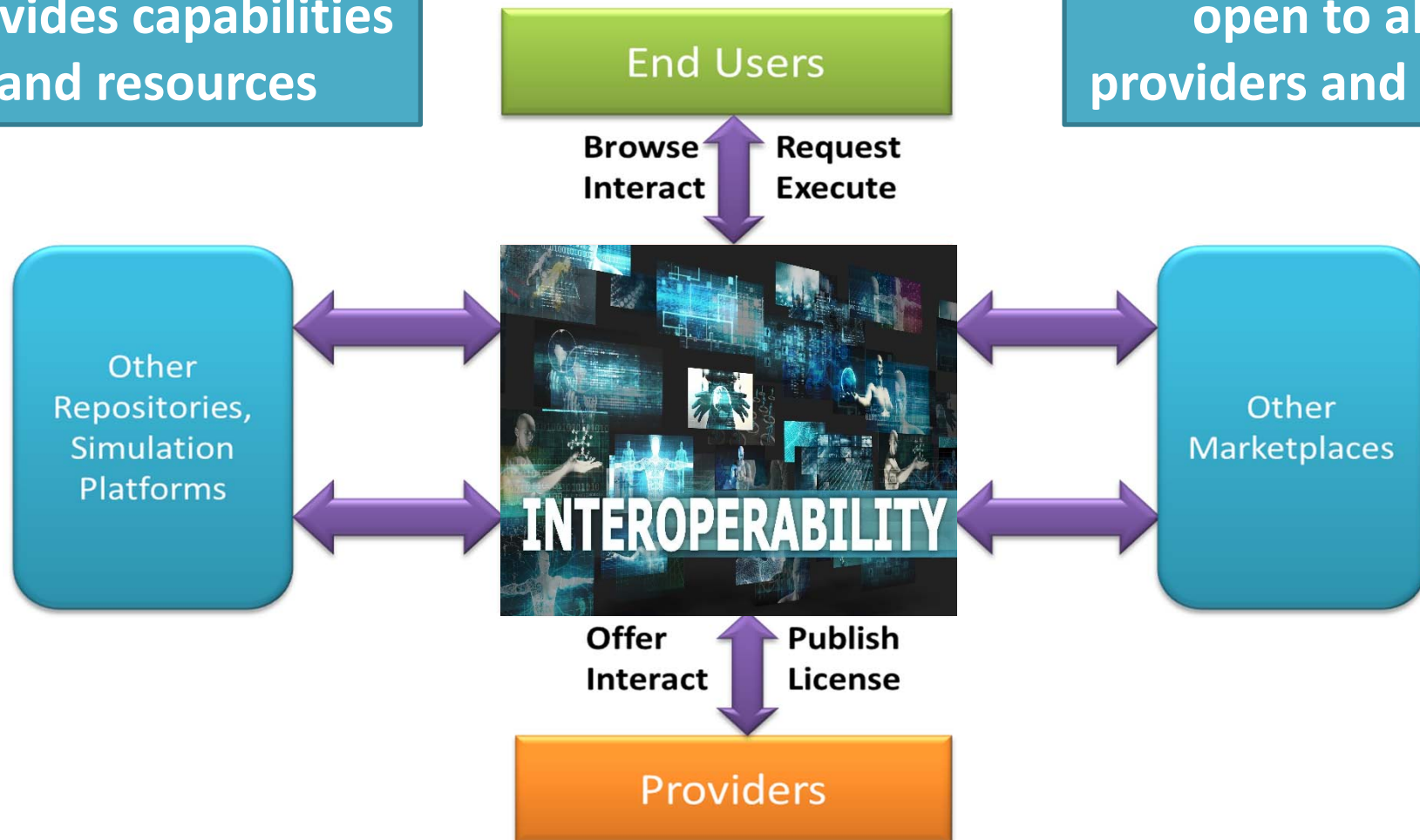




# Ultimate Industry Commons Vision: Digital Marketplaces

provides capabilities  
and resources

open to all  
providers and users





# Industry Commons “shopping list”

## FAIR Data

Annotation and linking,  
based on taxonomies

Interoperable, integrated  
repositories. Tools!

Start with open structured  
data, then closed (via  
open metadata)

Add unstructured (using  
semantics tools to  
structure)

## Semantics, AI

Concerted knowledge  
organisation effort:  
ontologies

Knowledge graphs,  
Automated data  
extraction

Analysis tools

AI based applications

## Marketplaces

Security, ownership

Human interactions,  
collaboration systems

Reward systems

Sustainability



# Actions in WP2020

- **DT-NMBP-39-2020: Towards Standardised Documentation of Data through taxonomies and ontologies (CSA)**
  - Standardised data documentation with metadata based on an agreed ontology across NMBP domains is critical for the widest use of data and, ultimately, reliable end-user products.
- **DT-NMBP-40-2020: Creating an open market place for industrial data (RIA)**
  - Make data FAIR through an effective common information system; data sharing in particular business-to-business
  - A marketplace open to all providers and users of data to maximise the spill over of knowledge across all economic sectors.

