



Hands-on Workshop on



tools + services

September 29-30, 2016
National Technical University of Athens, Greece

Extracting knowledge from data using the JaqPot Modelling Tool

Philip Doganis , NTUA



Nanomaterials

- ❑ [Nanomaterials are defined](#) as:
 - ❑ *A natural, incidental or manufactured material containing particles, in an unbound state or as an aggregate or as an agglomerate and where, for 50 % or more of the particles in the number size distribution, one or more external dimensions is in the size range **1 nm - 100 nm**.*
- ❑ They show **different properties** that other materials that exhibit structure in larger scale because of their very small dimensionality and their physicochemical properties, such as **shape** and **surface** properties.
- ❑ They have been tested and show great potential for use in **Information Technology, Communications, Medicine, Pharmaceutical production and delivery**.
- ❑ Nanomaterials are not intrinsically hazardous per se but there may be a need to take into account specific considerations in their risk assessment.



Role of Modelling in Nano

- In 2006, the European Community established the **REACH** (Regulation concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals) and founded the European Chemicals Agency (**ECHA**). Companies have the responsibility of collecting information on the **properties and the uses of substances** that they manufacture or import at or above one tonne per year. They also have to make an **assessment of the hazards and potential risks** presented by the substance.
- Learning more about chemicals by testing them on animals is a last resort – registrants may only carry out new tests when they have exhausted all other relevant and available data sources.

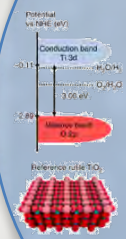
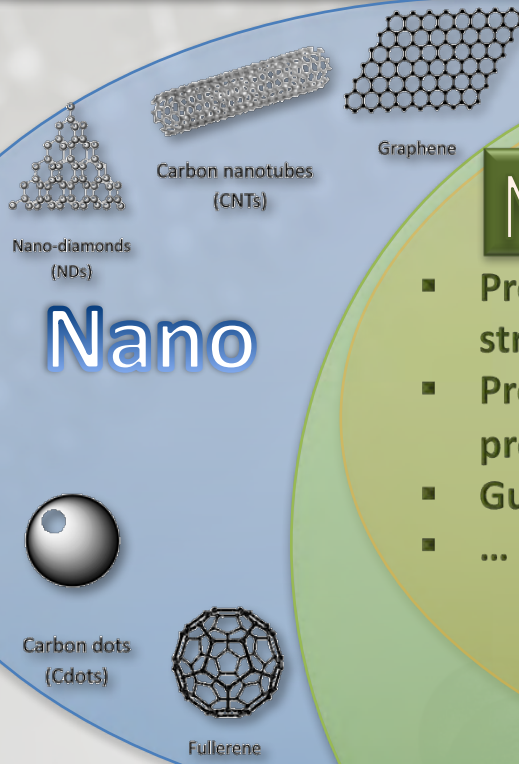
Source: <https://echa.europa.eu/>

$$\text{Activity, Toxicity, Properties} = f \left(\begin{array}{l} \text{electronic properties} \\ \text{physicochemical properties} \\ \text{topology} \end{array} \right)$$



NanoInformatics

Nano



Nanoinformatics

- Predict properties of existing structures
- Propose design that fits properties
- Guide experiments
- ...

Info

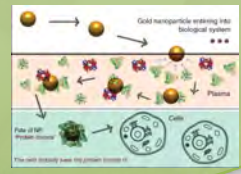


$$\frac{\partial}{\partial a} \ln f_{a,\sigma^2}(\xi_i) = \frac{(\xi_i - a)}{\sigma^2} f_{a,\sigma^2}(\xi_i) - \frac{1}{\sqrt{2\pi\sigma^2}}$$

$$\int \mathcal{T}(x) \cdot \frac{\partial}{\partial \theta} f(x, \theta) dx = M \left(\mathcal{T}(\xi) \cdot \frac{\partial}{\partial \theta} \ln l(\xi, \theta) \right)$$



Bio



Data mining - Computational Intelligence

- **Data mining** is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems.
- Computational Intelligence is an important approach of Artificial Intelligence with three basic techniques:
 - **Fuzzy Sets- Fuzzy Logic:** they lead to systems that operate and extract knowledge using “natural” language to encode human knowledge and intuition.
 - **Neural networks:** generalised mathematical structures to model nonlinear relations between input/output variables, without the need for a fundamental model.
 - **Evolutionary algorithms:** global optimization methods that perform a systematic random search based on the process of natural selection.

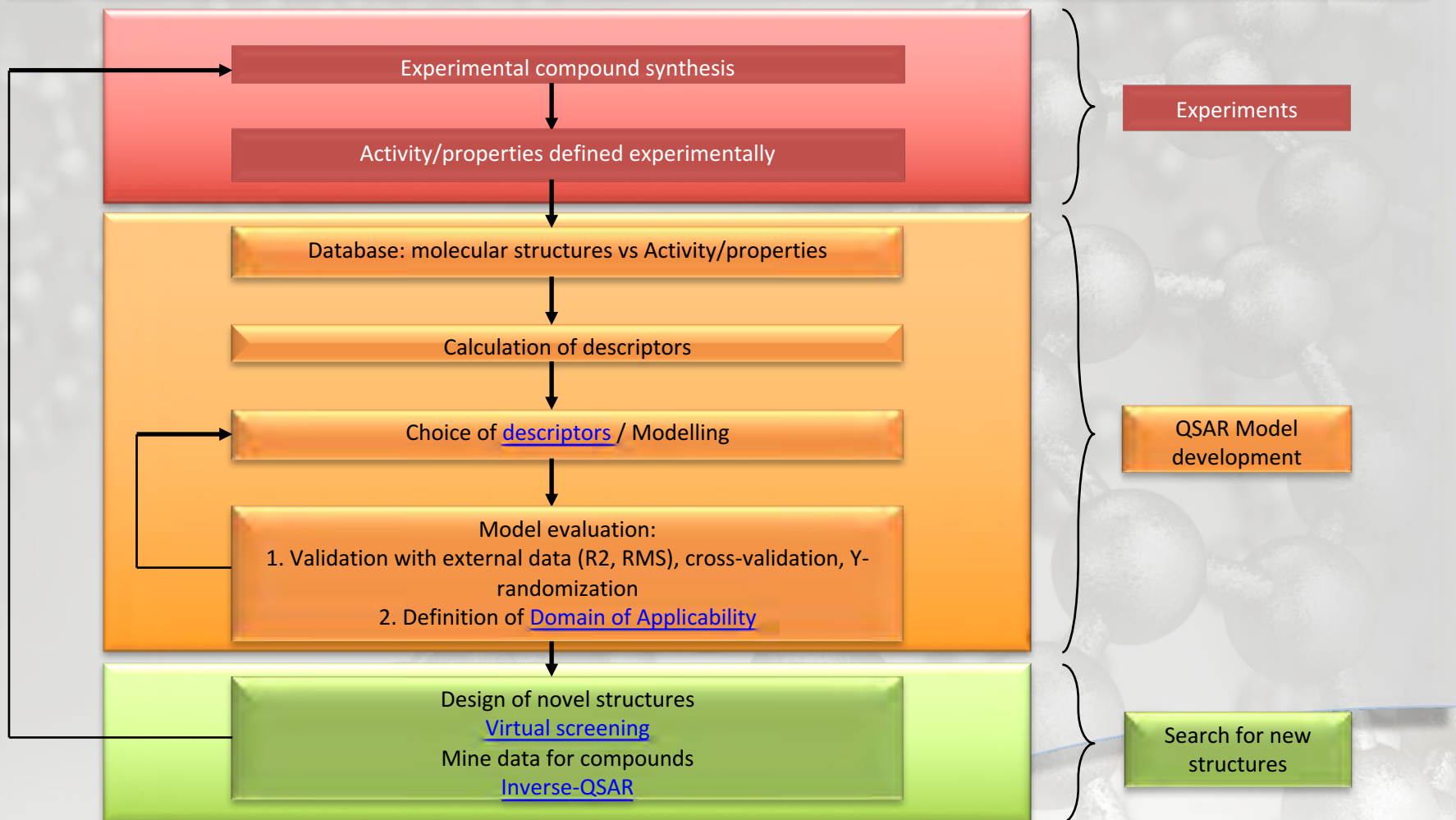


NanoInformatics - NanoQSAR

- ❑ An important research direction in the area of nanomaterials is **Safety by design**.
- ❑ Central to this premise is the development of **quantitative structure activity relationships (nanoQSAR)**, that identify relationships between nanomaterial properties (including toxicity) and their structural characteristics.
- ❑ We describe NMs using [Descriptors](#): numerical values, derived from sources that include **theory**, **Microscopy images** and **Omics data**, that are used to represent

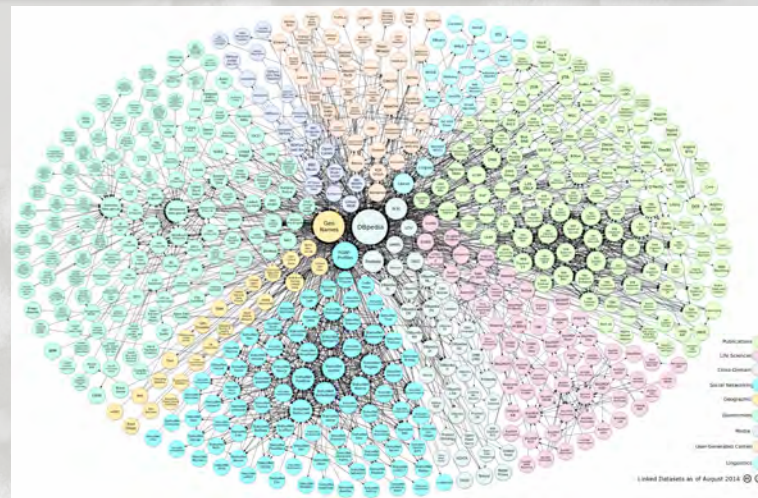


QSAR workflow



Structured Data – Linked data

- **Linked Data** is a method of publishing **structured data** so that it can be interlinked and become more useful through [semantic queries](#)
- It builds upon standard Web technologies such as [HTTP](#), [RDF](#) and [URIs](#)
- Share information in a way that can be read **automatically** by computers
- This enables data from different sources to be connected and queried

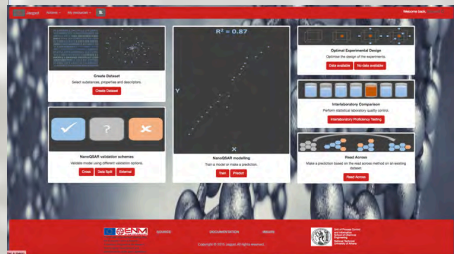


In Jaqpot we use URIs for:

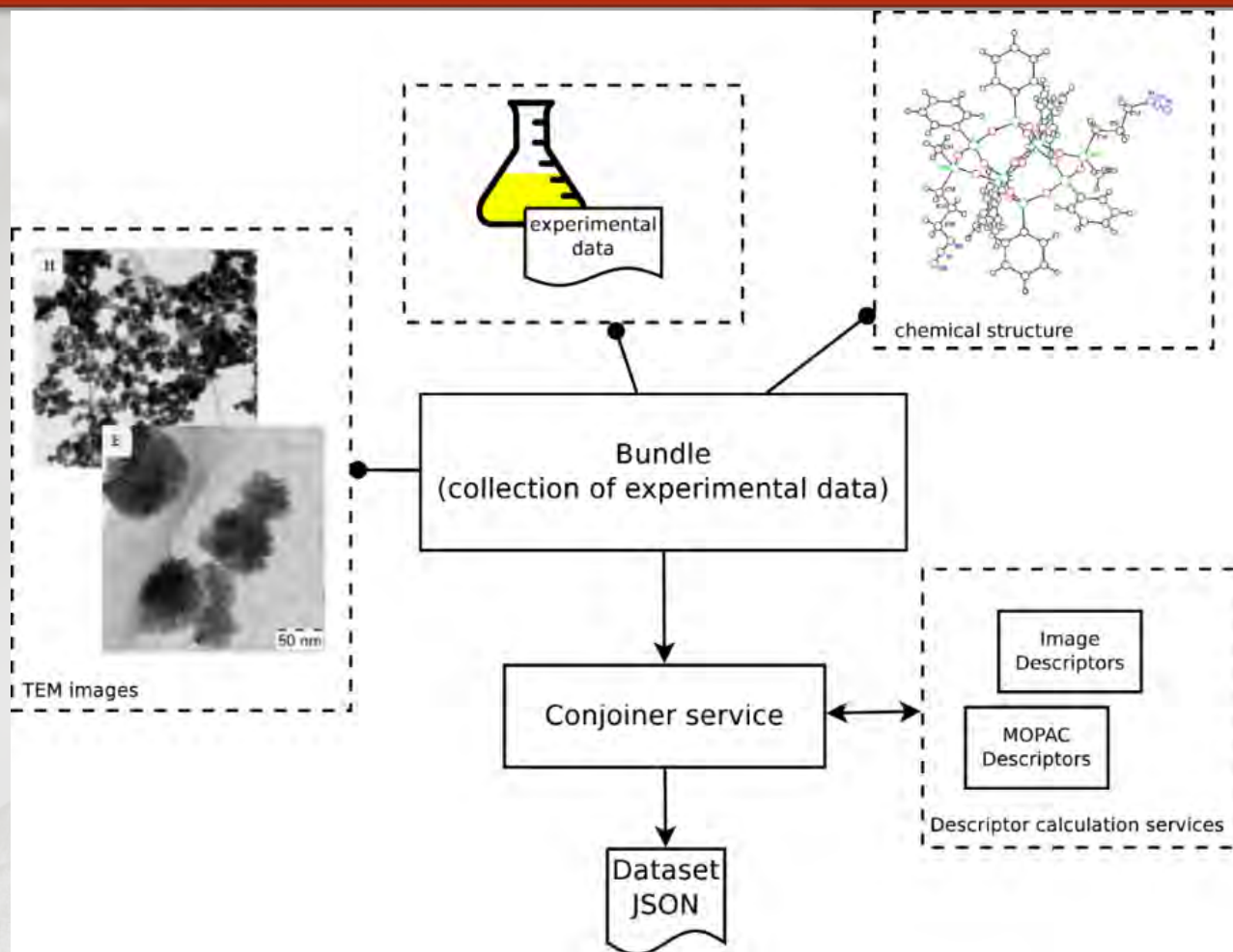
- Datasets
 - Models
 - Services
 - Properties
 - Results
- } Everything!



Jaqpot



<http://jaqpot.org>



Jaqpot - Functionality

The screenshot displays the Jaqpot web application interface. At the top, there is a navigation bar with 'Jaqpot', 'Home', 'My resources', and 'Welcome back, [user]'. The main content area is divided into several panels:

- Create Dataset:** Select substances, properties and descriptors. Includes a 'Create Dataset' button.
- NanoQSAR validation schemes:** Validate model using different validation options. Includes buttons for 'Cross', 'Data Split', and 'External'.
- NanoQSAR modelling:** Train a model or make a prediction. Shows a scatter plot with $R^2 = 0.87$. Includes 'Train' and 'Predict' buttons.
- Optimal Experimental Design:** Optimize the design of the experiments. Includes 'Data available' and 'No data available' buttons.
- Interlaboratory Comparison:** Perform statistical laboratory quality control. Includes 'Interlaboratory Proficiency Testing' button.
- Read Across:** Make a prediction based on the read across method on an existing dataset. Includes 'Read Across' button.

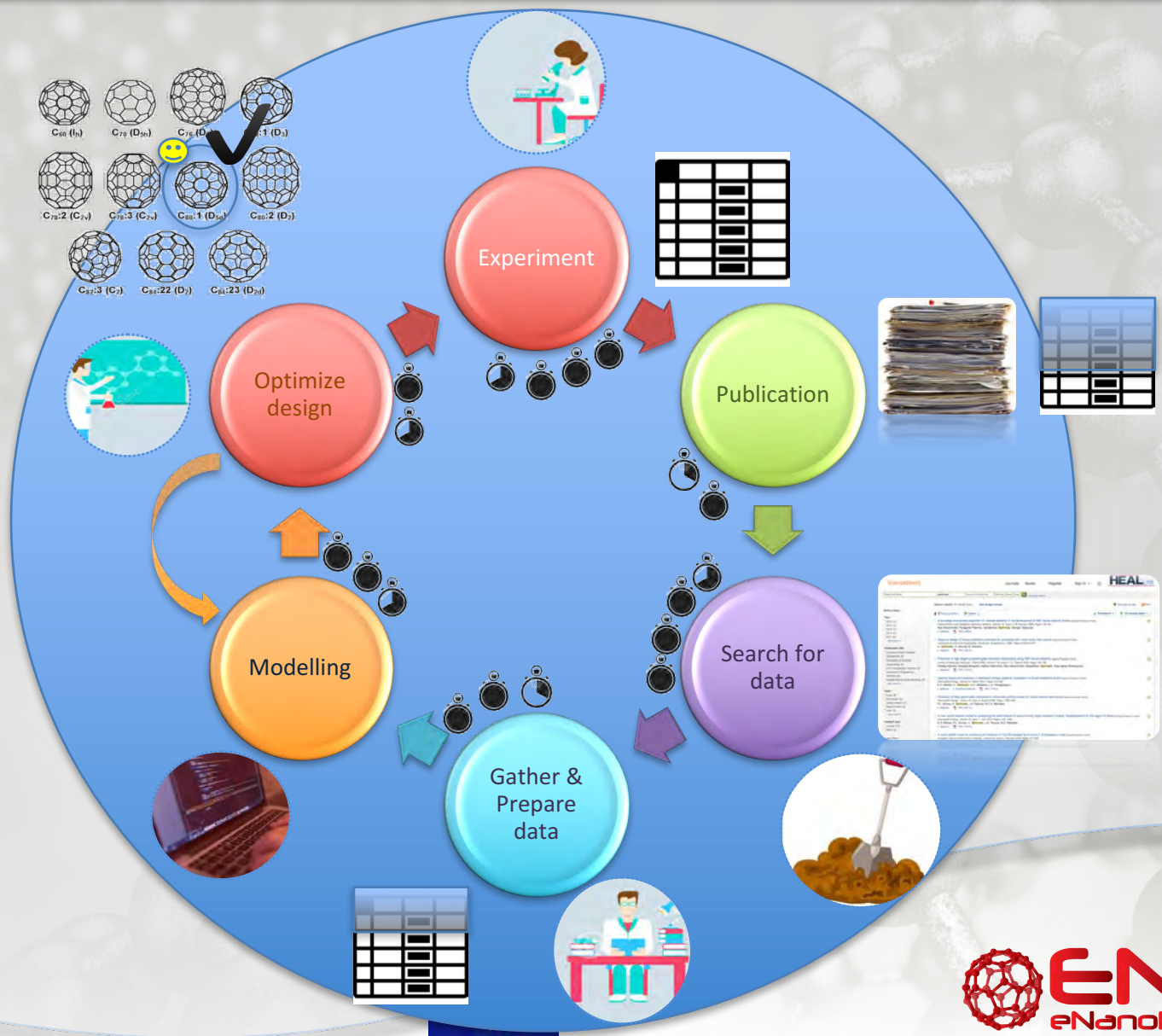
The footer contains logos for ENM (eNanoMapper), INM, and the University of Twente, along with copyright information: 'Copyright © 2015 Jaqpot. All rights reserved.'

<http://jaqpot.org>

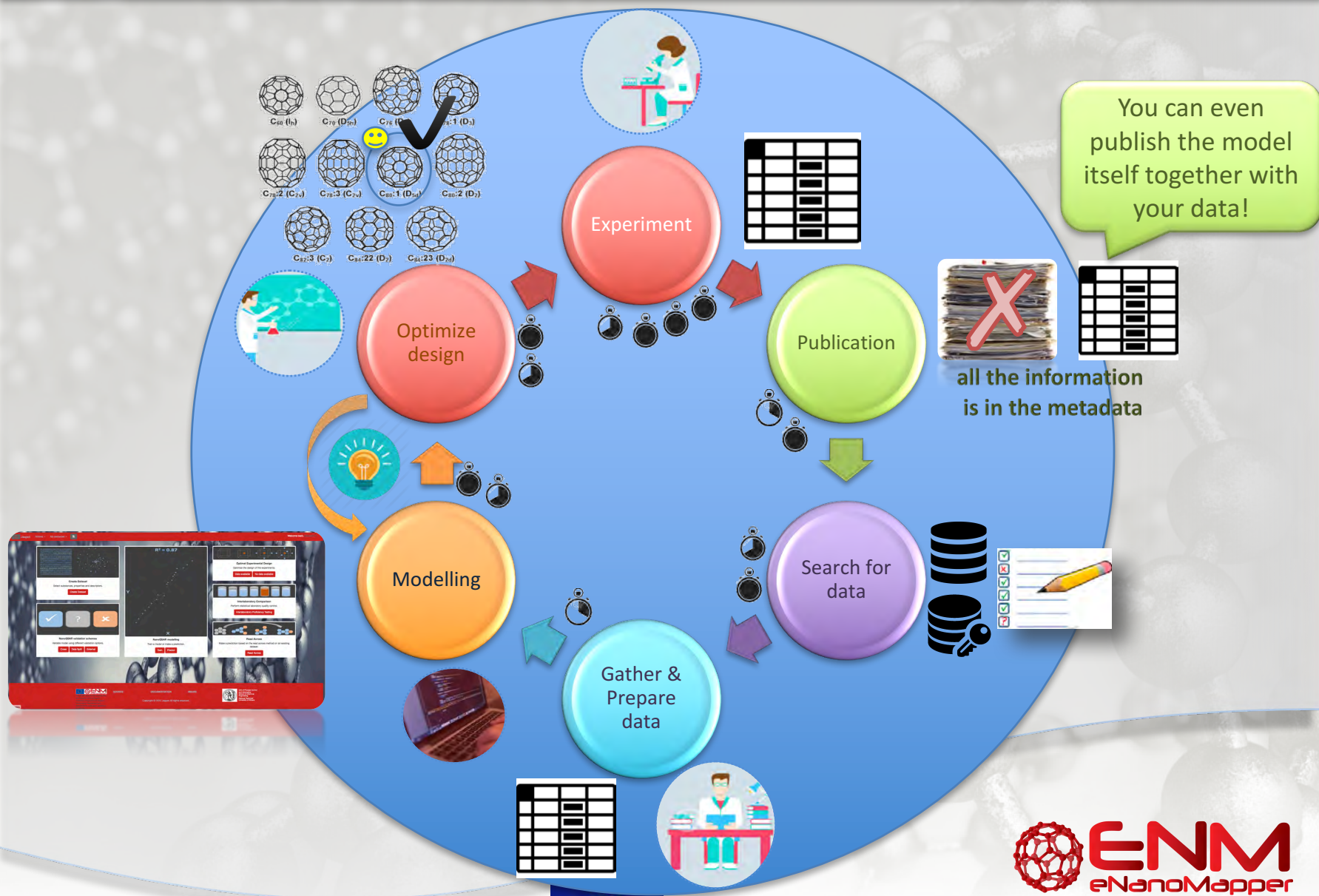
- nanoQSAR
 - Modelling
 - Validation
 - Image analysis
 - App
 - Web service
 - Quantum Mechanics calculations using MOPAC
- Part of v2.0, due soon:
- Optimal Experimental Design
 - Interlaboratory Comparison
 - Read Across



Doing research – old style



Role of Linked data & Jaqpot Modelling platform



eNanoMapper Objectives

- Improving the **utilisation of data** through the implementation of a modular infrastructure for **data storage, searching and sharing**, based on *open standards* and *semantic web* technologies, *minimum information standards* and established security solutions
- Accelerating knowledge *exchange* and *reuse* through the development of **ontologies** for the *categorisation* and *characterisation* of ENMs (pristine and in situ) in collaboration with other projects, including those launched following the NMP.2013-1.3-3 call (*Development of a systematic framework for naming and assessing safety of the next generations of nanomaterials being developed for industrial applications*).
- Enabling the creation of new **computational models** in nanomaterials safety through the implementation of interfaces for toxicity modelling and prediction algorithms which may process all data made available through eNanoMapper (statistical/data mining software)

