

## **The Chemistry KnowBase**

### A New Form of Knowledge Source for the Digital Age

IOCD promotes the pursuit and application of the chemical sciences for sustainable, equitable human development and economic growth, especially in low and middle-income countries.

Major activities involve capacity building and education. In the last domain we have developed:

- (i) Organic chemistry course in Spanish (<http://organica1.org>)
- (ii) Training in medicinal chemistry (<http://ntpd.pharm.ku.edu/IOCD/>),
- (iii) Global microscience project (in collaboration with UNESCO & IUPAC) (<http://www.iocd.org/WhatWeDo/microscience.shtml>).

The General assembly of IOCD has decided to initiate the KnowBase project that involve to create “unprecedented resources” aimed to help teachers to prepare their original courses according to official curriculum, local context and their own view.

#### **The KnowBase Project: Objectives**

KnowBase will at first cover chemistry in the broad sense at two different levels (i) higher education (U-ChemKnowBase) and (ii) secondary education (S-ChemKnowBase). We also expect to create an original resource for public education (P-ChemKnowBase).

KnowBase will cover the theoretical as well as experimental field and will link them to serious games. It will be composed of texts, equations, references, videos and a table of content and a dictionary of unusual format, it will not be organized as lectures, but as lucid and clearly illustrated descriptions and explanation of the data. Modularity and interoperability are the most striking characteristics of the KnowBase project. It will be built specifically to foster informal learning formalized focused on the table of contents (GTC). It will provide, under editorial control, authoritative and organized sources of information references and knowledge offering to teachers and learners global standard of knowledge at a specified level.

KnowBase will be built collaboratively by several scientists working for the same chapter side by side, that will transfer their personal knowledge with a shared vision and the same objectives.

KnowBase should be freely accessible in several languages through internet by anyone but will be first drafted in English. DVDs of the GTC will be available as well as specific data on request. The GTC will be used by teachers to select the topics they want to include in their courses and prepare their own table of content. They will then select the "Modules of knowledge" or part of them that will constitute their courses (Text, equations, references). Those can be accessed by their students or kept as models for others on activation of the related resource.

We expect that ChemKnowBase will offer a place where experiments and games will meet theory, knowledge and skills.

We will first focus on the U-ChemKnowBase project that involves the highest level of knowledge involved from which other sources will be derived after appropriate adjustments.

It will be the first place to cite in the introduction of any research paper, a place to read before starting any novel research project, a PhD thesis or a postdoc. It will be fully useful in industry, and will allow their technicians to be aware of a subject or to find the most important theoretical and experimental informations in order to conduct any type of reactions.

### **U-ChemKnowBase Project: Comparison with some existing resources**

Chem-KnowBase will be different from any existing available resources. Although related to:

- (i) Manuals. Those are intended to be used by the learners and not by their teachers, who are expected to know much more,
- (ii) Wikipedia, is not relevant for teaching. It misses the required organization, many important topics and the editorial control. It is the encyclopedia that "anyone can contribute".
- (iii) Series of books. Comprehensive series from Elsevier and Science of Synthesis from Georg Thieme Verlag (<http://www.thieme-chemistry.com/en/formate/referenzwerke/science-of-synthesis.html>) are related to research and not to education. They are therefore difficult to handle because the basic notions are not explicitly described and their content is more encyclopedic. Moreover they are not freely accessible what is a handicap especially for those living in low and middle-income countries,
- (iv) EdX a technological platform from MIT designed to offer online versions of the courses of MIT, Harvard, Berkeley and Georgetown Universities, They do not help the teachers to learn and have a few courses in science. Please have a look at : <https://www.edx.org/> and enrol for one course in chemistry.
- (vi) Chemistry related websites such as ChemSpider (<http://www.chemspider.com/>), Encyclopedia of Reagent Organic Chemistry Portal ([www.organic-chemistry.org](http://www.organic-chemistry.org)); Portal Chemistry (<http://en.wikipedia.org/wiki/Portal:Chemistry>); Chemical Portal (<http://www.webqc.org/>) do not provide textual information.

### **U-ChemKnowBase Project: The Models**

As far as the textual content is concerned, we have selected as basic model the textbook "Modern Organic reactions" written by H. House, Professor at MIT at the time, that appeared in the mid sixties (1965, 309 pages) and has been updated once on the same topics six years later (1971, 856 pages). It covers only few relevant topics (3) divided each in about four chapters of about fifty pages including text, schemes and references. The scientific message is nicely delivered; the examples are adequately selected as well as the references. Another textbook from J. March covered much more topics and much more references but the text does not transfer a message of similar quality and the equations and references not properly selected.

### **U-ChemKnowBase Project: Challenges**

**Chemistry.** The knowBase project should cover much more fields and although it will at first involve organic chemistry it should take its root from physical chemistry use as much structural data such as Xray crystallography and other data explaining how they have been correlated. It should cover different disciplines ranging from synthesis, supramolecular chemistry, medicinal chemistry and spectroscopy, analytical chemistry and many others.

Its conception should allow expanding the number of its topics, to insert new chapters and new paragraphs to the existing ones and of course to allow regular updating of the content of existing ones. Furthermore we want the same "Module of knowledge" to be meaningfully included at several locations of the GTC to avoid redundancies.

It will offer access to a series of typical experiments and to the definition of chemical terms through "ChemDic". ChemDic will play a key role in KnowBase that encompasses the characteristic of all types of dictionaries being a glossary of chemical terms and acronyms, providing short definitions, grammatical analysis possessing etymology and ontology entries and being as well an encyclopedic dictionary as an original entry to ChemKnowBase and a spellchecker. It will include for each entry a series of sentences using the term in all the contexts it is used and a series of related leading references. It will also allow access to information about equipment and its use and the name of the chemists who have play a major role.

Such sort of certified dictionary does not exist and the GoldBook from IUPAC (International Union of Pure and Applied Chemistry) although certified is only a chemical glossary missing references.

ChemBase will be built collaboratively and although the authors should be free to deliver their message, it should be delivered in an integrated mode. Credit and scientific rewards should be given to the authors that will provide the content costless. We therefore expect to get worldwide the support of national science foundations.

**ICT.** Information Technology is a key component of ChemKnowBase that should rely on the way the software is organized and built. It should allow to connect all the components, text, equation, reference and video databases, and provide interconnected data (to how many chapters a reference or an equation is related for example and its specific location). It should allow easy selection, from many places simultaneously, of data from the GCT to allow each one to build its own table of content and the related chapters. It should also allow original collaborative tools and protocols to favor the discussions between authors, teachers, learners and crosslinked interaction to adjust the system to the real needs. It should be advisable to built a tool based on freely available references and abstracts to automatically extract the topics, the related review article, and senior authors to organize the content of ChemBase, to select the most valuable authors and to alert for updating. In a second phase it is envisaged to built our own search engines and perception tools and to take advantage of ChemDic to favor the building of specific ontologies collaboratively.

## **U-ChemKnowBase Project: Implementation**

We have selected at first Prof. Erick Carreira (ETH, Zurich, <http://www.carreira.ethz.ch/people/emc>) and Gary Molander (Penn State U; <http://www.chem.upenn.edu/chem/research/faculty.php?id=28>) to be the partners of IOCD for initiating the process but many other scientists will join us once the preliminary modeling will be achieved.

ChemDic is being build by Atomicka, an Indian dedicated Company, and will be available to be filled within six months, and the original tolls to manage the references are already built. ChemAxon tolls (Budapest, Hungary) could be freely accessible to manage chemical equations (<http://www.chemaxon.com/>) on request.

We have agreed to collaborate with the Palais de la decouverte in Paris.

IOCD is involved to promote Microscience project (<http://www.iocd.org/WhatWeDo/microscience.shtml>) with African Academy of Sciences (<http://www.aasciences.org/>) and UNESCO ([http://portal.unesco.org/science/en/ev.php-URL\\_ID=5052&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/science/en/ev.php-URL_ID=5052&URL_DO=DO_TOPIC&URL_SECTION=201.html)) Those experiments will be linked to our SC-ChemKnowBase project.

Prof. Stefano Cerri (U Montpellier, <http://www.meetup.com/webscience-montpellier/members/13281991/>) will advice about building the collaborative tools being a specialist of WebScience an initiative that also includes MIT and the University of Southampton ([http://en.wikipedia.org/wiki/Web\\_Science\\_Trust](http://en.wikipedia.org/wiki/Web_Science_Trust)).