UNIVERSITY OF CAMBRIDGE

Unilever Cambridge Centre for Molecular Informatics

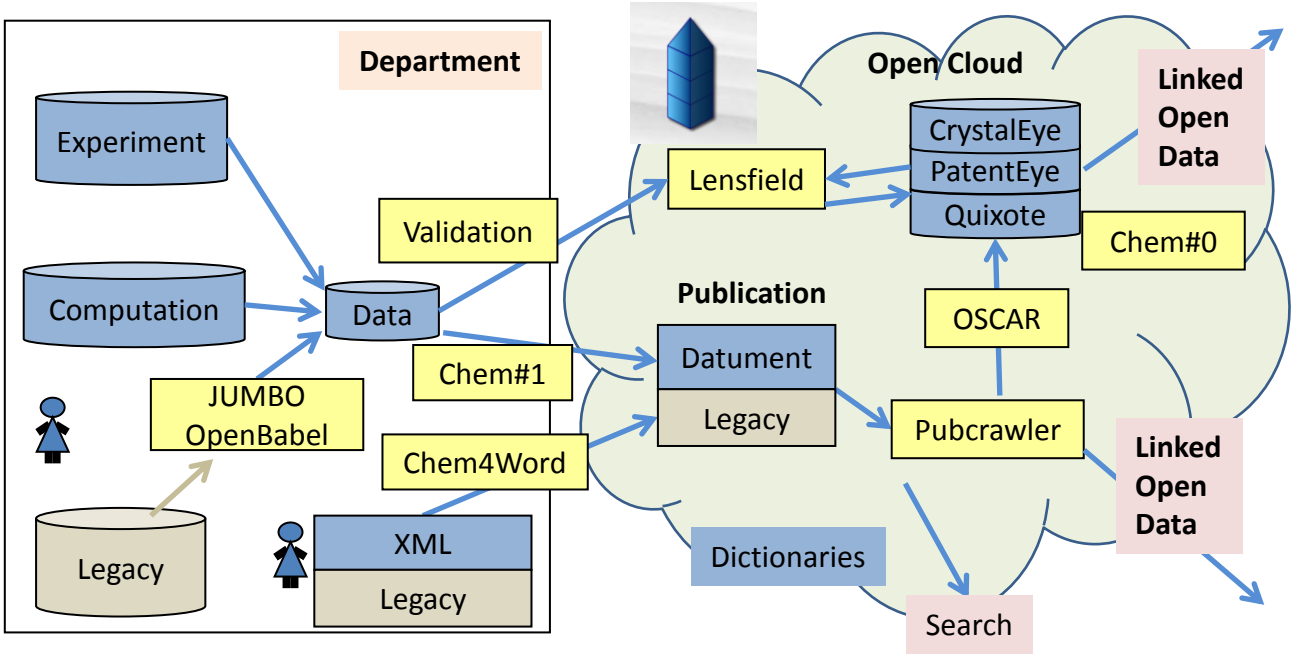
WWMM/CML Framework

The extended community

(PM-R group, Henry Rzepa, Blue Obelisk, Quixote, eMinerals, ChEng)

The **WorldWideMolecularMatrix** (2001-) is a design for capture and re-use of chemical information using Open semantic tools. There is no centre; scientists publish to the Matrix and re-use data from it. All data is Open. It interacts with the **Linked Open Data Cloud**

The diagram below shows ideas and components developed by our group and collaborators over 10 years, including **Authoring, semantic output, legacy conversion, text-mining, aggregation, crawling, and dictionaries.**



XML-CML/RDF framework → Software (Described on other posters)

The WWMM is fully interoperable with emerging Open Data activities such as linked Open Data and Blue Obelisk projects. Most data in chemistry is still closed, but CrystalEye, PatentEye and Quixote projects show how you can generate or extract data and publish it for Open re-use. The infrastructure is designed for early-adopter chemists, departments, publishers and repositories to capture, manage and publish their content.

The WWMM is also designed for citizen science, where no specialist knowledge of chemistry is required. The Green Chain Reaction project (September 2010) showed how a small group could deploy software and text-mine 100,000 reactions from patents in a month.
http://www-pmr.ch.cam.ac.uk/wiki/wwmm