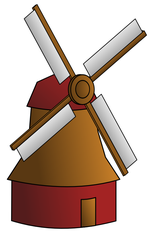


Quixote Project

The Quixote Project

http://quixote.wikispot.org

- an Open Source, Open Data, International collaboration to develop the infrastructure to organise, share, and query computational chemistry data
- no centralised structure, internet-based, and run entirely by motivated scientists
- create a useful infrastructure and consolidate the model around the tools; the "If you build it, they will come" approach.
- collaboration managed using skype conferences, wikis, etherpads and mailing lists
- have started to attract funding and collaborators

Openly Accessible Databases of Results

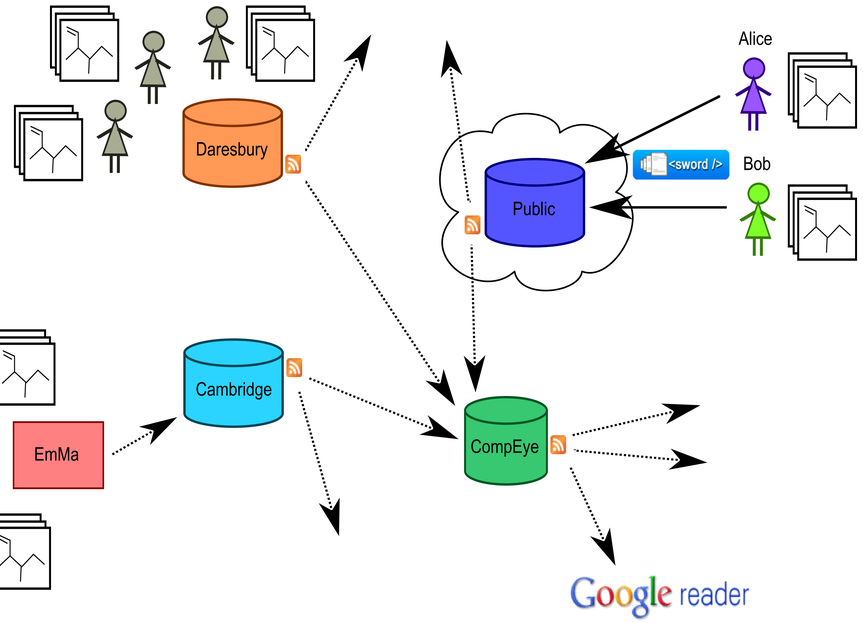
These would provide a valuable resource as:
- computational codes can be easily validated and benchmarked
- developers of new methods can check their results against similar calculations
- costly calculations only need to be run once as others can then access the data
- the wealth of data would be extremely useful for data mining
- standard repositories would provide an easy, automated way of generating and archiving supporting information for publications

Why Quioxte?

- there is currently no standard way to archive or search the data from computational chemistry calculations; as a result valuable data sits festering on hard drives, lost to everyone
- there isn't even a standard data format (despite the data being rigorously defined) so each computational chemistry code needs specialised tools to understand its output
- lots of people have tried to solve this (seemingly trivial) problem, but there is currently no accepted solution
- the name Quioxte is after Miguel Cervantes "Don Quixote de la Mancha" as the project was started following a CECAM meeting in Zaragoza, Spain http://neptuno.unizar.es/events/qcdatabases2010/

Test Cases/Collaborations

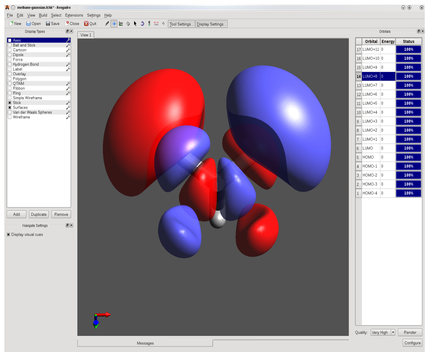
- we already have a number of groups who are testing the alpha implementation with their own data.
- we have funding available for some cloud storage for an initial public database
- a project has been started to use to the tools to conduct a scientific study on the conformations of cyclobutadiene entirely in the open



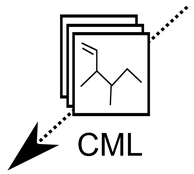
Current Architecture

- modular, open source tools - collaborate with existing projects where possible
- requirements: lightweight, easily installable, flexible, easily up-dateable, simple user interface and support the major Quantum Chemistry codes
- same infrastructure to manage a local datastore or a public data repository
- repositories expose atom feeds for aggregation/indexing/status updates.

GUI e.g. Avogadro



Input in CML



File and upload managment

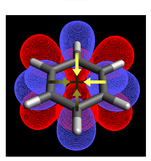
- lensfield2 monitors filesystem and manages file transformations
- can automate file uploads/downloads to local/remote repositories
- RESTful system for uploading and aggregation
- use of SWORD (http://swordapp.org) to facilitate authentication/ metadata management when uploading to repositories
- EMMA embargo system can control what is published from local to remote repositories

Chemical Markup Language (CML)

CML is a well-developed and established mark-up language for chemical data however it is yet to be adopted as the standard for all codes.

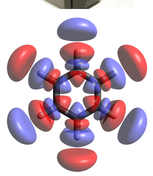
Such adoption would allow:
- different codes to interoperate to create complex workflows
- tools (e.g. GUIs such as Avogadro) to operate on the input or output of any code supporting the format
- data to be automatically validated, as a semantic model underlies the format

We are developing Quixote around CML, but this requires the development of converters from legacy formats to CML. Several projects (e.g. jumbo-converters, openbabel, cclib) already do much that is needed in this area.



NWChem runs initial calculation on computer A

CML Output used as input



Gaussian runs final calculation on computer B

Quixote People

The nature of the Quixote project means that there is no "membership" as such. However, a (necessarily incomplete) list of active participants includes:

Sam Adams: http://www-pmr.ch.cam.ac.uk/wiki/Sam_Adams
Tamás Beke: Chalmers Univesity, Sweden.
Pablo Echenique: http://www.pabloechenique.com
Jorge Estrada: BIFI, University of Zaragoza, Spain
Marcus D. Hanwell: http://www.kitware.com/company/team/hanwell.html
Peter Murray-Rust: http://www.ch.cam.ac.uk/staff/pm.html
Jens Thomas: STFC Daresbury Laboratory, U.K.
Joe Townsend: http://www-pmr.ch.cam.ac.uk/wiki/Joe_Townsend
Lance Westerhoff: QuantumBio Inc, USA.