mmCIF and Dictionary Driven Software with the MSD Database Production Pipeline

Kim Henrick, Sameer Valenkar, Harry Boutselakis, Ayzaz Hussain, John Ionides, Adamandia Kapopoulou, Peter Keller, Richard Newman, Jorge Pineda, Antonio Suarez, Jawahar Swaminathan, John Tate, European Bioinformatics Institute, EMBL Outstation Hinxton, Wellcome Trust Genome Campus Hinxton Cambridge UK. E-mail: henrick@ebi.ac.uk

The Macromolecular Structure Database (MSD) group, http://www.ebi.ac.uk/msd, based at the European Bioinformatics Institute (EBI) (an outstation of the European Molecular Biology Laboratory EMBL) is a member of the wwPDB (http://www.wwpdb.org) and provides one of the PDB deposition sites. The MSD uses AutoDep4 (http://www.ebi.ac.uk/msdsrv/autodep4/) for PDB depositions and carries out not only the annotation tasks required to produce PDB entries it also provides a stable and clean repository of macromolecular structure data services that allow users to access, search and retrieve structural data. In addition the MSD handles the deposition and archive for Maps from cryo-electron microscopy through the separate deposition interface EMDep, http://www.ebi.ac.uk/msd-srv/emdep/. mmCIF and XML are used throughout the data processing pipeline including the mapping of PDB entries to an in-house extended mmCIF for loading intio the Oracle databases and in the numerous processes that are run on the deposition data to enrich the PDB data with extensive derive information. The MSD export the PDB entries to the RCSB in the standard pdbx dictionary format. This talk will outline part of the processing pipeline used.

Keywords: mmCIF, relational databases, processing pipeline