



<http://www.3dem-noe.org/>

**Project no:** : LSHG-CT-2004-502828

**Project acronym:** 3DEM

**Project full name:** Three Dimensional Electron Microscopy

**Instrument type:** Network of Excellence

**Priority name:** "Life sciences, genomics and biotechnology for health,"

**Deliverable reference number:** De1.11

**WorkPackage:** WP1 Training Programme

**Title:** Course 3DEM Developers Workshop

**Due date of deliverable:** Month 14

**Actual submission date:** 12<sup>th</sup> October 2005

**Start date of project:** 1-March-2004

**Duration:** 5 years

**Organisation name of lead contractor for this deliverable:** EMBL-EBI  
European Molecular Biology Laboratory - European Bioinformatics Institute,  
Hinxton, Cambridge, UK.

**Revision:** Version 1.0

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
<b>PU</b>	Public	12-October 2005
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission Services)	
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	

## 3DEM Workshop 12-14\_oct05

PF6 Network of Excellence - 3DEM Workshop

[http://www.ebi.ac.uk/msd-srv/docs/3DEM\\_7.html](http://www.ebi.ac.uk/msd-srv/docs/3DEM_7.html)

schema / metadata / visualisation / conventions / standards / data-harvesting / software-standards / integration

"let's go for it and give it a try"  
Programme for the Meeting

### 12.10.2005

#### AFTERNOON SESSION

12.30-14.00	Registration & Lunch	
14.00-14.20	Welcome and General Introduction	Kim Henrick ( <a href="#">PDF</a> )
14.20-15.00	The 3dem dictionary in mmcif terms	Matt Baker ( <a href="#">PDF</a> )
15.00-15.40	RCSB software for handling the 3dem dictionary	John Westbrook ( <a href="#">PDF</a> )
15.40-16.00	Tea/Coffee break	
16.00-16:40	Integration of SPIDER with the Wadsworth Centre Project Archive covering data flow from EM acquisition to completed reconstruction	Bill Baxter ( <a href="#">PDF</a> )
16.40-17.20	Integration of eMAN with the Baylor object-oriented database and electronic notebook for Electron Cryomicroscopy	Steve Ludtke ( <a href="#">PDF</a> )
17.20-18.00	IMAGIC and data harvesting	Marin van Heel ( <a href="#">PDF</a> )
18.00	Dinner & bar	

### 13.10.2005

#### MORNING SESSION

8.50-9.30	IPLT software and data harvesting in XML for Electron Microscopy	Ansgar Philippsen ( <a href="#">PDF</a> )
9.30-10.00	Flat file database for documenting image processing micrographs using Bsoft and STAR format	Bernard Heymann ( <a href="#">PDF</a> )
10.00-10.30	IMAGIC header mapping between different formats with em2em	Michael Schatz ( <a href="#">PDF</a> )
10.30-10.50	Coffee/Tea break	
10.50-11.30	Harvesting in tomography through the TOM suite of software using XML formats for image data in relation to high-throughput electron microscopy	Christof Best ( <a href="#">PDF</a> )
11.30-11.50	COMET (tomography) software	Lars-Goran Ofverstedt ( <a href="#">PDF</a> )
11.50-12.20	MAP conventions and XML representation	David Belnap ( <a href="#">PDF</a> )
12.20-12.40	A hands-on example on software interoperability: model refinement interconnecting spiker and xmipp	Carlos Perez Roca ( <a href="#">PDF</a> )

12.40-13:00 Software interoperability: library level Carlos Oscar Sanchez Sorzano ([PDF](#))  
13.00-14.00 Lunch

#### AFTERNOON SESSION

14.00-18.00 The meeting will be held in an IT room with the opportunity for developers to bring along the current version of their software and to have hands-on sessions to begin the conversation of the software to using the new dictionary items as either mmCIF or XML. We can either work on a single software library for all software developers to use in a software jamboree or developers can work together to see how their packages can be converted and whether this requires additions to the dictionary.

Files to be made available:  
Current 3DEM database entries mapped to new dictionary  
Mapping of Bsoft star format to new dictionary  
Mapping of TOM XML to new dictionary

19.00 Dinner & bar

**14.10.2005**

#### MORNING SESSION

9.00-9.40 An Instrument Manufacture view of data harvesting through to archive database Reinier Perquin ([PDF](#))  
9.40-10.00 Visualisation trials via the WWW for 3dem maps using the AstexViewer™ Tom Oldfield ([PDF](#))  
10.00-10.40 Discussion on practicality of using the dictionary and next steps  
10.40-11.00 Coffee/Tea break  
11.00-12.30 Programming session  
13.00 CLOSE & LUNCH

#### Summary of talks and discussions (cf ppt or pdf presentations):

##### **The 3dem dictionary in mmcif terms Matt Baker (Baylor)**

Described the further development of the EM dictionary set up initially at the EBI with the addition of new data items following a meeting at the RCSB. cf Cryo-EM development definition web page [http://mmcif.pdb.org/dictionaries/mmcif\\_emx.dic/Index/index.html](http://mmcif.pdb.org/dictionaries/mmcif_emx.dic/Index/index.html) ([http://iims.ebi.ac.uk/cif\\_iims.dic/Groups/em\\_group.html](http://iims.ebi.ac.uk/cif_iims.dic/Groups/em_group.html)). He further described a 3DEM ontology cf presentation.

##### **RCSB software for handling the 3dem dictionary John Westbrook (RCSB)**

Described how the new EM dictionary: ([http://mmcif.pdb.org/dictionaries/mmcif\\_emx.dic/Index/index.html](http://mmcif.pdb.org/dictionaries/mmcif_emx.dic/Index/index.html)) was

generated encoded in mmCIF Software tools are/will be available from RCSB to convert mmCIF to XML.

**Integration of SPIDER with the Wadsworth Centre Project Archive covering data flow from EM acquisition to completed reconstruction Bill Baxter (Albany)**

Described an automated microscopy for SPIDER:

Data-flow from EM acquisition to completed reconstruction.

Microscopy—Reconstruction—Archive DB

Mostly manual input

Described SPIRE: SPIRE is controlled by a graphical user interface to processing with SPIDER batch files, and contains a database for keeping track of reconstruction projects. It provides a simple interface that ties buttons to batch files and displays their outputs. SPIRE detects if a copy of Jweb is running, and if so, uses it to display SPIDER images.

SPIRE software package:

GUI

PYTHON wrapper

Eg

Reconstruction experiment-----SQL + JAVA-?-----Archive DB

Automatic particle picking

Log files generated contains much more information than is archived.

Export XML readily achieved.

**Integration of eMAN with the Baylor object-oriented database and electronic notebook for Electron Cryomicroscopy Steve Ludtke (Baylor)**

EMEN (Electron Microscopy Electronic Notebook) is an object-oriented database used to record information about all aspects of transmission electron microscopy experiments, from specimen purification through final 3D reconstruction. EMEN is built on top of the object-oriented database and web-based application server, ZOPE.

EMEN- total experiment recording

eMAN-

Some problems with the internal db :

- Poor data mining

- Search speed sluggish after 0.5 million records

- No schema mapping

**Flat file database for documenting image processing micrographs using Bsoft and STAR format Bernard Heymann (NIH)**

Bsoft is a collection of programs and a platform for development of software for image and molecular processing in structural biology. Problems in structural biology are approached with a highly modular design, allowing fast development of new algorithms without the burden of issues such as file I/O. It provides an

easily accessible interface, a resource that can be and has been used in other packages.

Bsoft software in STAR format.

Reads old flat files with advantages:

STAR file is version 'ignorant' so script changes not a problem

STAR micrograph file:

Multiple micrographs of same field at beginning of STAR file have same history in header

Data blocks grouped into one file or separated

STARfile encodes micrograph parameters and particle orientations and

May be expanded to include multiple conformations.

Experimental stage XML converter available.

Future developments:

Support for XML parameters file

Web-based workflow

Deposition button to EMDB

<b>IPLT software and data harvesting in XML for Electron Microscopy Ansgar Philippsen (Biozentrum)</b>
--

IPLT is a modular, collaborative, integrated, open-source architecture for image processing of electron microscopy images. It is designed around object oriented paradigms and implemented using the programming languages C++ and Python. In many aspects it deviates from classical image processing approaches.

IPLT uses C++ libs and STL standard template library

Recommended is the 'boost' C++ software developers site (

<http://www.boost.org/>)

HTF file format

Lively debate on whether effort to make useable wrapper worthwhile. At this stage not.

<b>A hands-on example on software interoperability: model refinement interconnecting spider and xmipp Carlos Perez Roca (CSIC) Interoperability between different packages in a workflow system. and Carlos Perez Roca (CSIC-Spain) Carlos Interoperability in practice.</b>
--

Comments again reflected problems with workflow approach in terms of recognition for separately developed packages in terms of getting funding for such standalones as when considered as part of workflow. Fraught with difficulty eg because of possible changes within individual packages not being caught so simple change can cause major crash somewhere in pipeline-difficult to trace. A workflow system using standard components was suggested using webservice for raw data transfer. There is a recognition that there is heterogeneous software

and standards for protocols and data formats was required. Decision making success/failure software conventions.

**IMAGIC header mapping between different formats with em2em Michael Schatz (Image Science)**

Described the non-trivial header conversion processes for EM maps, using em2em. (<http://www.imagescience.de/em2em/em2em.htm>). Particular problems when packages change their header information as this is usually not published or made known at the time. Has supplied information on header conversion information to Spanish web-site (include URL). Exemplified requirement for format encompassing new conventions and dictionary with facility for enhanced header information.

**Harvesting in tomography through the TOM suite of software using XML formats for image data in relation to high-throughput electron microscopy Christof Best (MPI)**

Discussion of various softwares with examples of a method to use a matrix XML facility.

TOM software can export XML standardised for the new dictionary terms.

**COMET (tomography) software Lars-Goran Ofverstedt (Karolinska Inst.)**

COMET allows an efficient use of low-dose cryo TEM images of unstained specimen. In addition, an automated 3-D correlation and averaging technique to reach intermediate resolutions on single vitrified objects is being developed. COMET:Automated tomography at very low dose (20 e/A<sup>2</sup>) for all 120 images. Fortran and C programming.Flat files with numbers for image parameters for CTF correction etc.Has reconstruction XML in C++ environment.Has unique parameters so should use STANDARD parameters from new dictionary for interchange. Will be able to export XML in interchange/archive format.

**MAP conventions and XML representation David Belnap**

Presented the published 3DEM conventions describing the handling of maps. Suggested that the standards keeper would be EMDDB and a committee be setup to review further additions. EMDEP should enforce conventions at deposition. Developers were asked to implement the conventions and it was suggested that at the next Gordon conference on EM that the community should be encouraged to endorse these conventions.

**An Instrument Manufacture view of data harvesting through to archive database Reinier Perquin (FEI)**

Reviewed DICOM standards existing for medical science as an example for setting up agreed standards between hardware and associated software manufacturers. However not happening yet between EM suppliers. Described F.E.I. extension to the standard MRC (ccp4) map format and agreed with comments that the FEI 'MRC' image format should have a different extension to distinguish it from standard MRC formats.

**Visualisation trials via the WWW for 3dem maps using the AstexViewer™ Tom Oldfield (EBI)**

Demonstrated Astexviewer with EM map.  
Request for rendering visualisation  
Tom [cf 'bob' package for example of rendering-'cloud' type visualisation  
(<http://www.ahpcrc.org/software/bob/>)]

**Summarised Discussions.**

**(1) Discussion on how to obtain export XML from different packages, and general philosophy:**

What will benefit deposition of data in the future? and implementation of a search and retrieval system to supply standard data.

**Agreed conclusion:** Final reconstruction program dumps out the deposition XML file, Say using a 'deposition' button which will give a local XML mapping to the conventions XML. No requirement to archive raw data. This to be implemented by early 2006.

**(2) Discussion on conventions and image formats.**

Conventions as set out in DB and BH paper agreed upon. Including image (map) origin, units for data items etc (cf URL to paper) N.B. Marin pointed out published rotation matrices incorrect? Test sets of images to be made available by DB for circulation and testing to establish practical test methods for eg existing EMDB maps.

**Agreed conclusion:** New map format to be established based on agreed conventions. Format to include XML header with range of information as prescribed. Implementation during 2006-2007.

**(3) Discussion of new dictionary.**

John Westbrook presented newly updated dictionary containing micrograph and particle data-items generated during the meeting.

**Agreed conclusion:** Similarly other data items regarding Tomography, 2d- crystallisation and filament processing would be made available from BH and other users to JW. Implementation early 2006