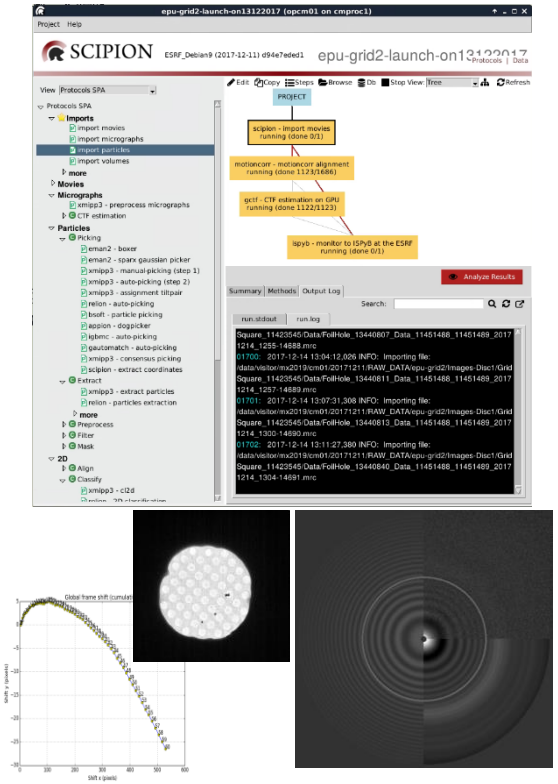
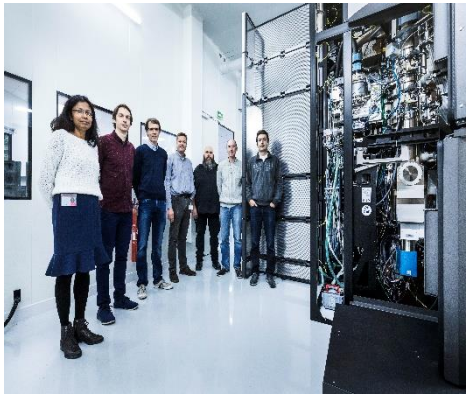


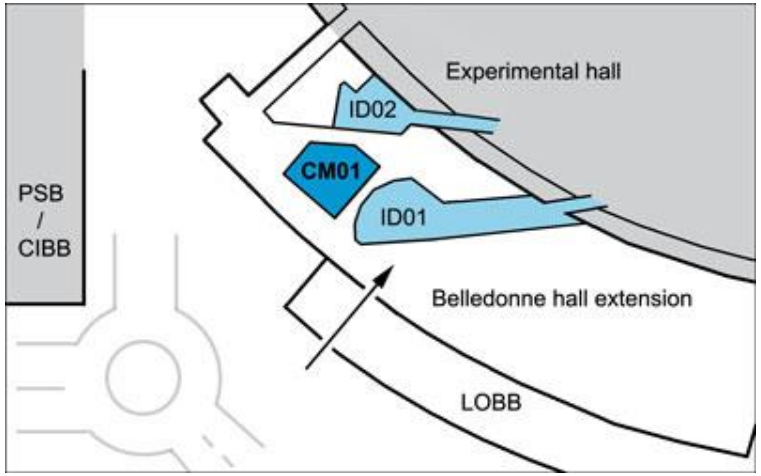
Connection ISPyB and Scipion: Current status at the ESRF and future plans

Olof Svensson
Data Analysis Unit
ISDD / ESRF

ESRF CRYO-EM CM01

CM01



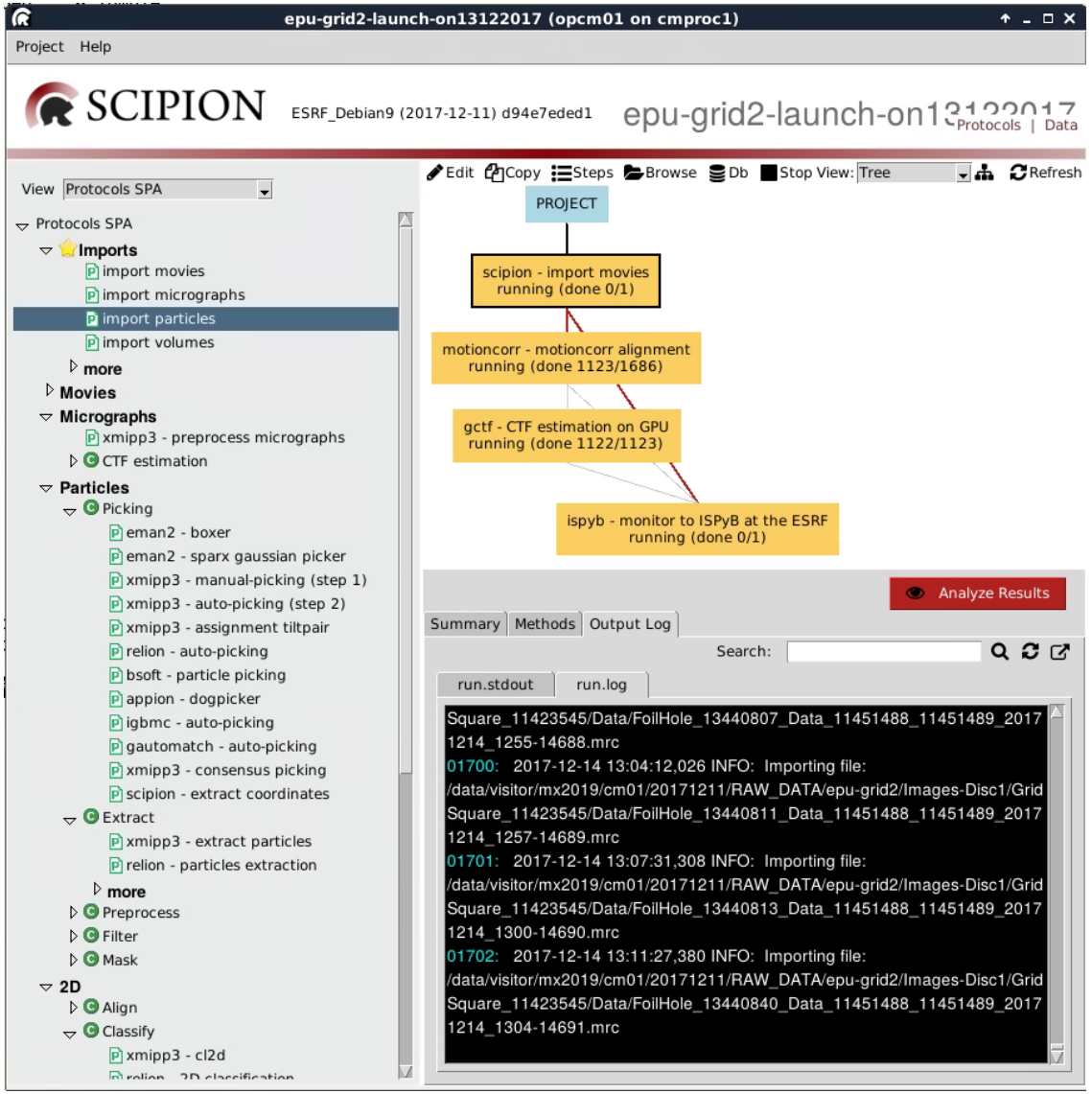


http://www.esrf.eu/home/news/general/content-news/general/inauguration-of-a-cryo-electron-microscope-platform-at-the-esrf.html

AUTOMATIC CRYOEM DATA PROCESSING AT THE ESRF

Computing ressorces:

- One dedicated server
- 28 cores
- 2 GPU (GTX 1080 Ti)



ESRF SCIPION DEVELOPMENTS

Customized scipion startup script:

- Creates a new Scipion project and executes it.
- Command line driven, started by the CM01 scientists (GUI to be developed).
- Uses Scipion un-modified actors for movie import, motioncorr2 and Gctf.
- ISPyB monitor developed using the one written by (Kevin Savage and Jose Miguel de la Rosa Trevin) as base.
- Runs very reliable! Very nice features like possibility of individual start and stop of actors.
- Uploads meta-data to ISPyB using three web service calls.
- In operation since November 2017, 7 user experiments so far, > 25000 movies uploaded to ISPyB.

ISPYB CRYOEM WEB SERVICES

https://github.com/antolinos/ispyb-client/wiki/EM-Workflow

addMovie :

- proposal: Proposal type and proposal number of the current experiment. For instance: mx415
- sampleAcronym: Acronym of the sample.
- movieDirectory: [DATACOLLECTION.imageDirectory] Directory that contains movies and will identify the grid square as a data collection.
- movieFullPath: [MOVIE.movieFullPath] The full path of the movie.
- movieNumber: [MOVIE.movieNumber] The unique number of a movie for a grid square.
- micrographFullPath: [MOVIE.micrographFullPath] Full path of a micrograph produced by instrument.
- micrographSnapshotFullPath: [MOVIE.micrographSnapshotFullPath] Full path of a snapshot of the micrograph.
- xmlMetaDataFullPath: [MOVIE.xmlMetaDataFullPath] Full path of a XML file containing movie metadata.
- voltage: [DATACOLLECTION.voltage] Microscope voltage in kV.
- sphericalAberration: [DATACOLLECTION.sphericalAberration]Spherical aberration in mm.
- amplitudeContrast: [DATACOLLECTION.amplitudeContrast]Amplitude contrast.
- magnification: [DATACOLLECTION.magnification]Magnification.
- pixelSizeOnImage: [DATACOLLECTION.pixelSizeOnImage]scanned pixel size in microns.
- numberOfImages: [DATACOLLECTION.numberOfImages] Number of frames in a movie.
- dosePerImage: [MOVIE.dosePerImage] Dose per image in e/Å^2.
- positionX: [MOVIE.positionX] X position of movie.
- positionY: [MOVIE.positionY] Y position of movie.
- beamlineName
- startTime SimpleDateFormat dt = new SimpleDateFormat("yyyyy-mm-dd hh:mm:ss");

addMotionCorrection

- proposal: Proposal type and proposal number of the current experiment. For instance: mx415
- movieDirectory: Directory that contains movies and will identify the grid square as a data collection.
- micrographFullPath: Full path of a MRC file containing the micrograph.
- correctedDoseMicrographFullPath: Full path of a MRC file containing the micrograph corrected for dose.
- micrographSnapshotFullPath: Full path of a PNG file containing snapshot of micrograph.
- driftPlotFullPath: Full path of a PNG file containing plots of global shifts.
- logFileFullPath: Full path to the log file of motion correction.

addCTF

- proposal: Proposal type and proposal number of the current experiment. For instance: mx415
- movieFullPath: [MOVIE.movieFullPath] The full path of the movie.
- spectraImageSnapshotFullPath: Full path of a JPEG snapshot of the spectra image.
- spectraImageFullPath: Full path to the MRC file of the spectra image.
- defocusU: Defocus U in Å.
- defocusV: Defocus V in Å.
- angle: angle in degrees.
- crossCorrelationCoefficient: The Cross Correlation Coefficient (CCC).
- resolutionLimit: Resolution limit in Å.
- estimatedBfactor: Estimated B factor in Å^2.
- logFilePath: full path to the log file of CTF estimation

PRESENTATION OF RESULTS IN EXI

EM 23-11-2017 16:37:53
/data/visitor/mx415/cm01/20171123/RAW_DATA/testsecretin-grid1/Images-Disc1/GridSquare_23722826/Data

Sample	Sample6
Grid Squares	1
Voltage	300000
Spherical Aberration	2.7
# Frames	40
Amplitude Contrast	10%
Scanned Pixel Size	

2017-11-23 16:37:52

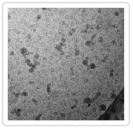
Movies	7
Motion Correction	100.0%
CTF	100.0%

Voltage	300000
Magnification	130000
Directory	/data/visitor/mx415/cm01/20171123/RAW_DATA/testsecretin-grid1/Images-Disc1/GridSquare_23722826/Data

Movie #3385 30-11-2017 10:55:23

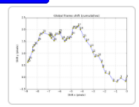
Movie

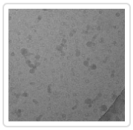
Number	3385
X, Y	-0.000, 0.000
Dose	0.1300



Motion Correction

Total Motion	11.3
Avg. Motion/frame	0.3
Frame Range	1 - 40
Dose	1.0000

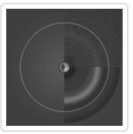


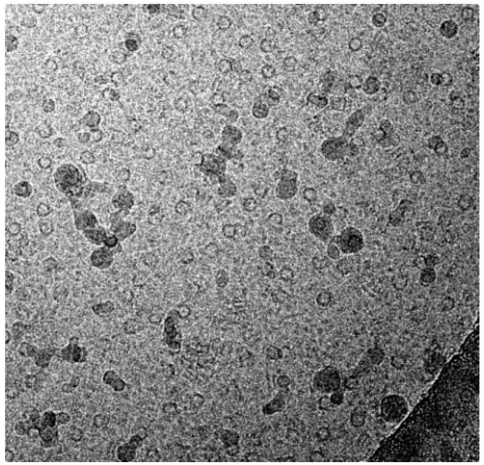


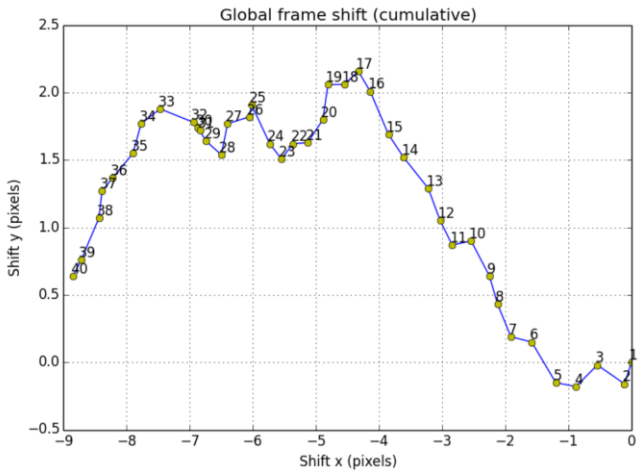
CTF

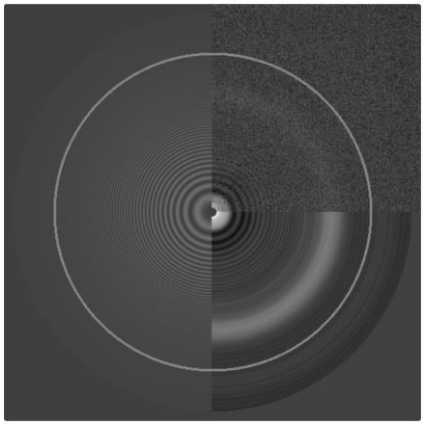
Resolution Limit	2.902
Correlation	0.068865
Defocus U	22280.72
Defocus V	21719.82

Angle	62.99
Estimated B factor	86.87









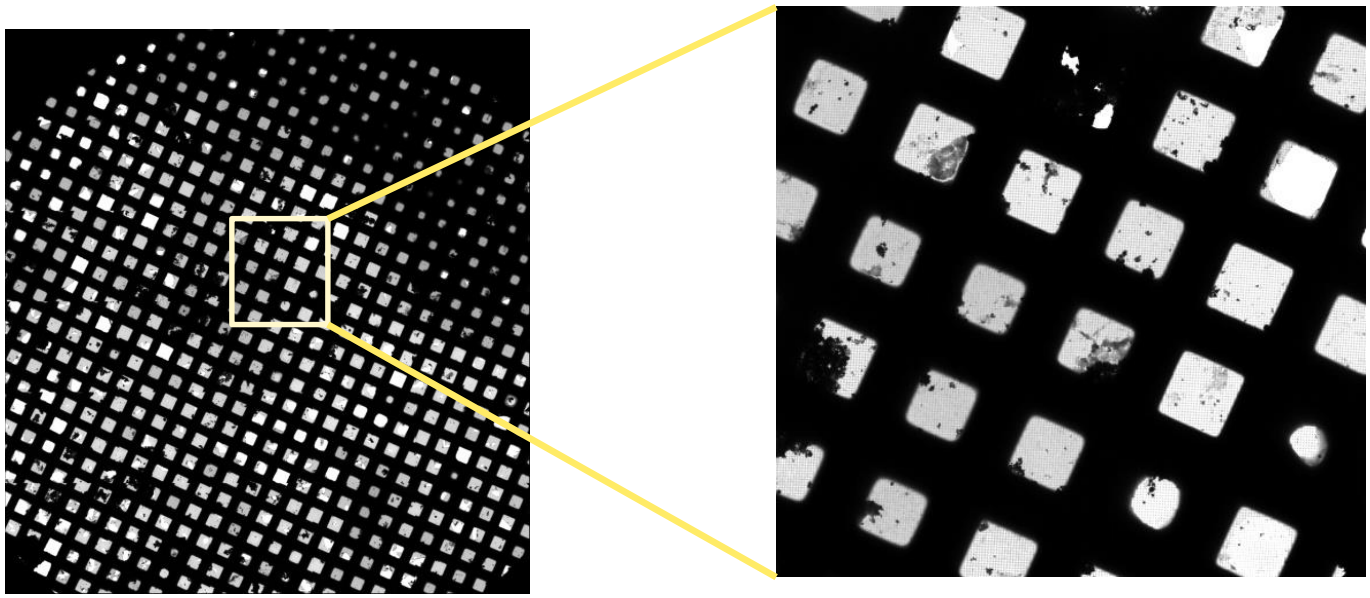
CRYOEM FUTURE DATA ANALYSIS DEVELOPMENTS

Extend automatic processing

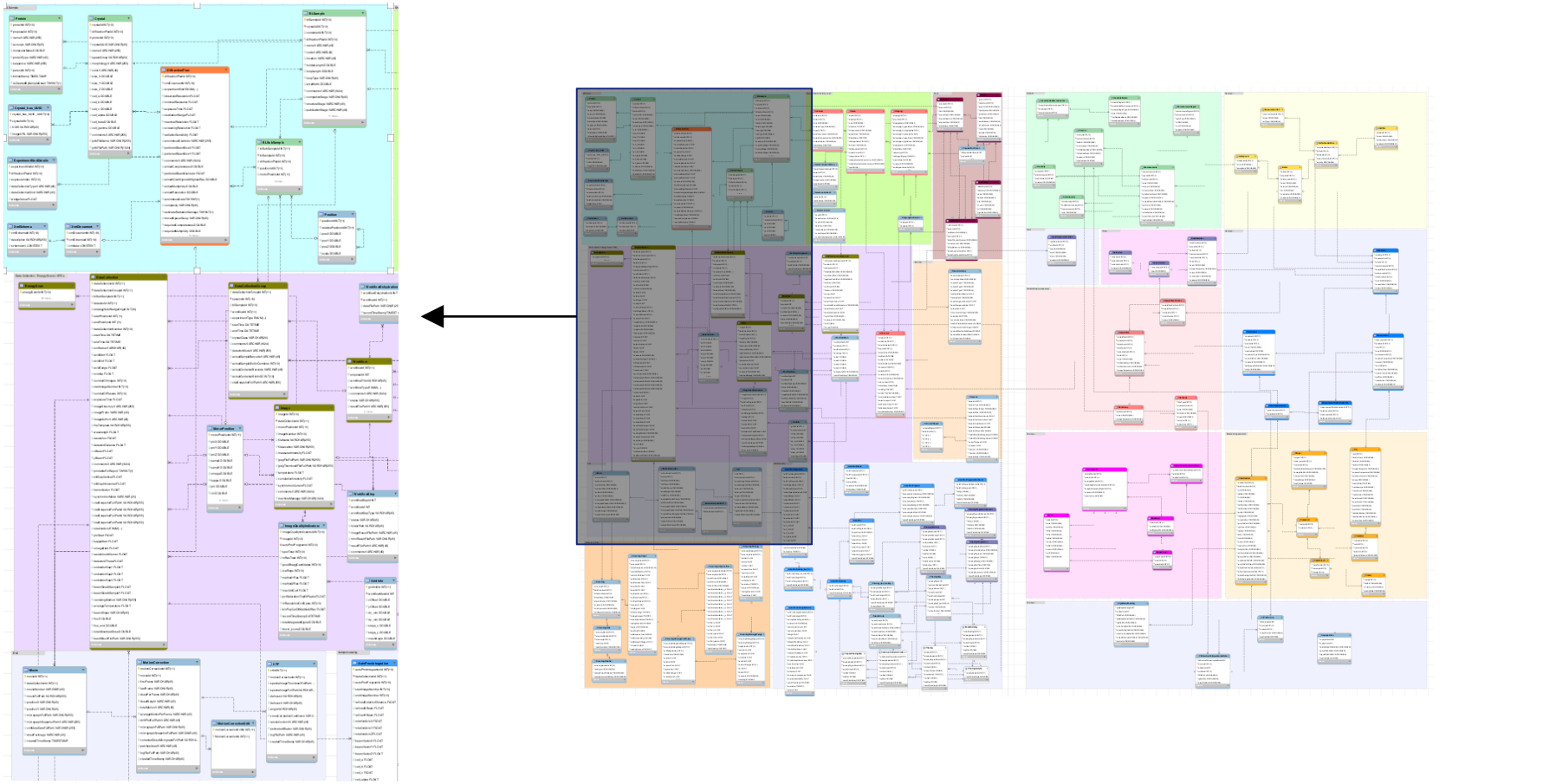
- Particle picking
- Particle extract
- 2D classification

Mail-in operation

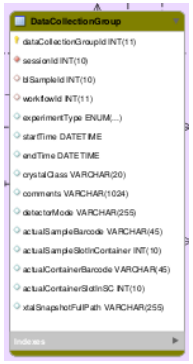
- Not possible to give users access to Cryo-EM data acquisition software like for MXCuBE
- Users will choose remotely which grid squares to use for data collection

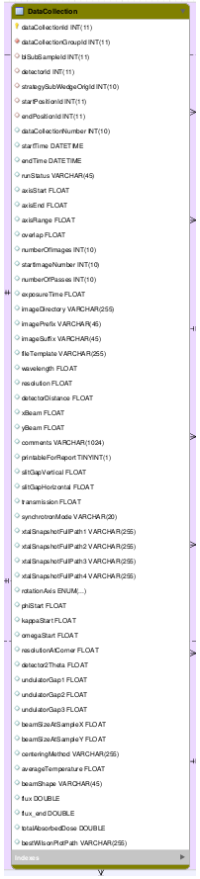


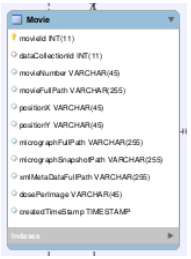
DATA MODEL

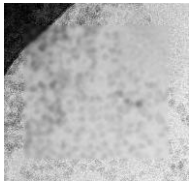


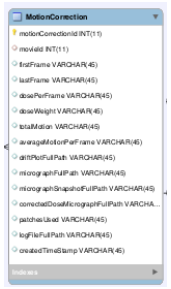
DATA MODEL

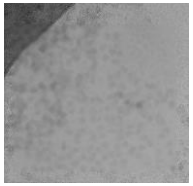

DataCollectionGroup
dataCollectionGroupId INT(11)
sessionId INT(10)
blSampleId INT(10)
workflowId INT(11)
experimentType ENUM(...)
startTime DATETIME
endTime DATETIME
crystalClass VARCHAR(20)
comments VARCHAR(1024)
detectorMode VARCHAR(255)
actualSampleBarcode VARCHAR(45)
actualSampleSlotInContainer INT(10)
actualContainerBarcode VARCHAR(45)
actualContainerSlotInSC INT(10)
xtalSnapshotFullPath VARCHAR(255)

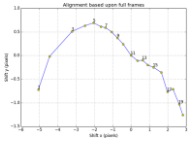

DataCollection
dataCollectionId INT(11)
dataCollectionGroupId INT(11)
blSubSampleId INT(11)
detectorId INT(11)
strategySubWedgeOrigId INT(10)
startPositionId INT(11)
endPositionId INT(11)
dataCollectionNumber INT(10)
startTime DATETIME
endTime DATETIME
runStatus VARCHAR(45)
axisStart FLOAT
axisEnd FLOAT
axisRange FLOAT
overlap FLOAT
numberOfImages INT(10)
startImageNumber INT(10)
numberOfPasses INT(10)
exposureTime FLOAT
imageDirectory VARCHAR(255)
imagePrefix VARCHAR(45)
imageSuffix VARCHAR(45)
fileTemplate VARCHAR(255)
wavelength FLOAT
resolution FLOAT
detectorDistance FLOAT
xBeam FLOAT
yBeam FLOAT
comments VARCHAR(1024)
printableForReport TINYINT(1)
slitGapVertical FLOAT
slitGapHorizontal FLOAT
transmission FLOAT
synchrotronMode VARCHAR(20)
xtalSnapshotFullPath1 VARCHAR(255)
xtalSnapshotFullPath2 VARCHAR(255)
xtalSnapshotFullPath3 VARCHAR(255)
xtalSnapshotFullPath4 VARCHAR(255)
rotationAxis ENUM(...)
phiStart FLOAT
kappaStart FLOAT
omegaStart FLOAT
resolutionAtCorner FLOAT
detector2Theta FLOAT
undulatorGap1 FLOAT
undulatorGap2 FLOAT
undulatorGap3 FLOAT
beamSizeAtSampleX FLOAT
beamSizeAtSampleY FLOAT
centeringMethod VARCHAR(255)
averageTemperature FLOAT
beamShape VARCHAR(45)
flux DOUBLE
flux_end DOUBLE
totalAbsorbedDose DOUBLE
bestWilsonPlotPath VARCHAR(255)

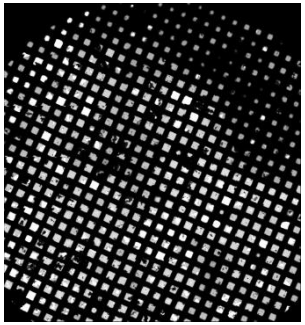

Movie
movieId INT(11)
dataCollectionId INT(11)
movieNumber VARCHAR(45)
movieFullPath VARCHAR(255)
positionX VARCHAR(45)
positionY VARCHAR(45)
micrographFullPath VARCHAR(255)
micrographSnapshotPath VARCHAR(255)
xmlMetaDataFullPath VARCHAR(255)
dosePerImage VARCHAR(45)
createdTimeStamp TIMESTAMP

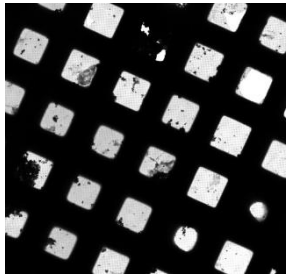


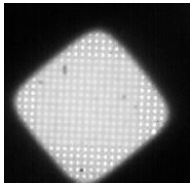

MotionCorrection
motionCorrectionId INT(11)
movieId INT(11)
firstFrame VARCHAR(45)
lastFrame VARCHAR(45)
dosePerFrame VARCHAR(45)
doseWeight VARCHAR(45)
totalMotion VARCHAR(45)
averageMotionPerFrame VARCHAR(45)
driftPlotFullPath VARCHAR(45)
micrographFullPath VARCHAR(45)
micrographSnapshotFullPath VARCHAR(45)
correctedDoseMicrographFullPath VARCHA...
patchesUsed VARCHAR(45)
logFileFullPath VARCHAR(45)
createdTimeStamp VARCHAR(45)

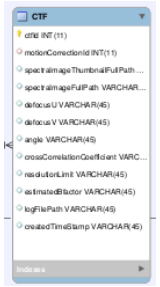


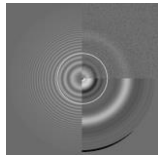

Alignment based upon full frames



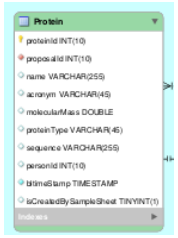


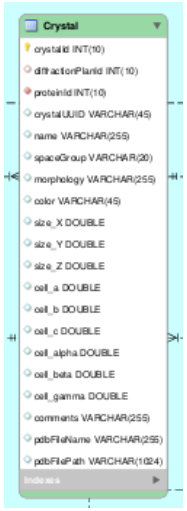


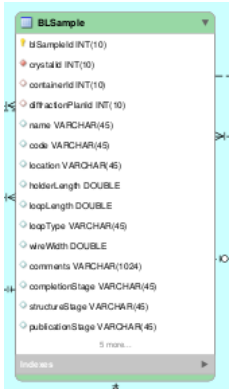

CTF
ctfId INT(11)
motionCorrectionId INT(11)
spectralImageThumbnailFullPath ...
spectralImageFullPath VARCHAR...
defocusU VARCHAR(45)
defocusV VARCHAR(45)
angle VARCHAR(45)
crossCorrelationCoefficient VARC...
resolutionLimit VARCHAR(45)
estimatedBfactor VARCHAR(45)
logFilePath VARCHAR(45)
createdTimeStamp VARCHAR(45)

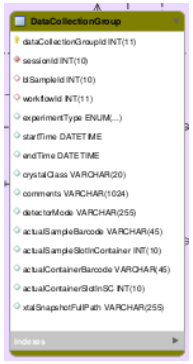


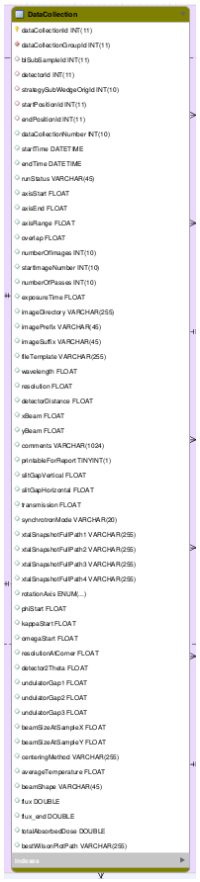
DATA MODEL

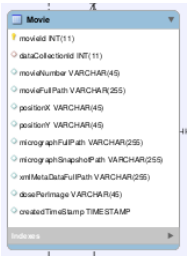

Protein
proteinId INT(10)
proposalId INT(10)
name VARCHAR(255)
acronym VARCHAR(45)
molecularMass DOUBLE
proteinType VARCHAR(45)
sequence VARCHAR(255)
personId INT(10)
bltimeStamp TIMESTAMP
isCreatedBySampleSheet TINYINT(1)
Indexes

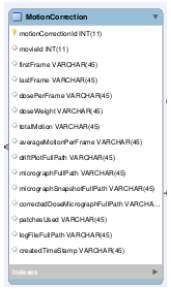

Crystal
crystalId INT(10)
diffractionPlanId INT(10)
proteinId INT(10)
crystalUUID VARCHAR(45)
name VARCHAR(255)
spaceGroup VARCHAR(20)
morphology VARCHAR(255)
color VARCHAR(45)
size_X DOUBLE
size_Y DOUBLE
size_Z DOUBLE
cell_a DOUBLE
cell_b DOUBLE
cell_c DOUBLE
cell_alpha DOUBLE
cell_beta DOUBLE
cell_gamma DOUBLE
comments VARCHAR(255)
pdbFileName VARCHAR(255)
pdbFilePath VARCHAR(1024)
Indexes

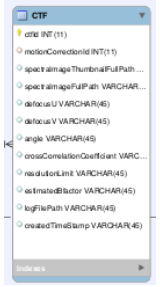

BLSample
blSampleId INT(10)
crystalId INT(10)
containerId INT(10)
diffractionPlanId INT(10)
name VARCHAR(45)
code VARCHAR(45)
location VARCHAR(45)
holderLength DOUBLE
loopLength DOUBLE
loopType VARCHAR(45)
wireWidth DOUBLE
comments VARCHAR(1024)
completionStage VARCHAR(45)
structureStage VARCHAR(45)
publicationStage VARCHAR(45)
5 more...
Indexes


DataCollectionGroup
dataCollectionGroupId INT(11)
sessionId INT(10)
blSampleId INT(10)
workflowId INT(11)
experimentType ENUM(...)
startTime DATETIME
endTime DATETIME
crystalClass VARCHAR(20)
comments VARCHAR(1024)
detectorMode VARCHAR(255)
actualSampleBarcode VARCHAR(45)
actualSampleSlotInContainer INT(10)
actualContainerBarcode VARCHAR(45)
actualContainerSlotInSC INT(10)
xtalSnapshotFullPath VARCHAR(255)
Indexes


DataCollection
dataCollectionId INT(11)
dataCollectionGroupId INT(11)
blSubSampleId INT(11)
detectorId INT(11)
strategySubWedgeOrigId INT(10)
startPositionId INT(11)
endPositionId INT(11)
dataCollectionNumber INT(10)
startTime DATETIME
endTime DATETIME
runStatus VARCHAR(45)
axisStart FLOAT
axisEnd FLOAT
axisRange FLOAT
overlap FLOAT
numberOfImages INT(10)
startImageNumber INT(10)
numberOfPasses INT(10)
exposureTime FLOAT
imageDirectory VARCHAR(255)
imagePrefix VARCHAR(45)
imageSuffix VARCHAR(45)
fileTemplate VARCHAR(255)
wavelength FLOAT
resolution FLOAT
detectorDistance FLOAT
xBeam FLOAT
yBeam FLOAT
comments VARCHAR(1024)
printableForReport TINYINT(1)
slitGapVertical FLOAT
slitGapHorizontal FLOAT
transmission FLOAT
synchrotronMode VARCHAR(20)
xtalSnapshotFullPath1 VARCHAR(255)
xtalSnapshotFullPath2 VARCHAR(255)
xtalSnapshotFullPath3 VARCHAR(255)
xtalSnapshotFullPath4 VARCHAR(255)
rotationAxis ENUM(...)
phiStart FLOAT
kappaStart FLOAT
omegaStart FLOAT
resolutionAtCorner FLOAT
detector2Theta FLOAT
undulatorGap1 FLOAT
undulatorGap2 FLOAT
undulatorGap3 FLOAT
beamSizeAtSampleX FLOAT
beamSizeAtSampleY FLOAT
centeringMethod VARCHAR(255)
averageTemperature FLOAT
beamShape VARCHAR(45)
flux DOUBLE
flux_end DOUBLE
totalAbsorbedDose DOUBLE
bestWilsonPlotPath VARCHAR(255)
Indexes


Movie
movieId INT(11)
dataCollectionId INT(11)
movieNumber VARCHAR(45)
movieFullPath VARCHAR(255)
positionX VARCHAR(45)
positionY VARCHAR(45)
micrographFullPath VARCHAR(255)
micrographSnapshotPath VARCHAR(255)
xmlMetaDataFullPath VARCHAR(255)
dosePerImage VARCHAR(45)
createdTimeStamp TIMESTAMP
Indexes


MotionCorrection
motionCorrectionId INT(11)
movieId INT(11)
firstFrame VARCHAR(45)
lastFrame VARCHAR(45)
dosePerFrame VARCHAR(45)
doseWeight VARCHAR(45)
totalMotion VARCHAR(45)
averageMotionPerFrame VARCHAR(45)
driftPlotFullPath VARCHAR(45)
micrographFullPath VARCHAR(45)
micrographSnapshotFullPath VARCHAR(45)
createdTimeStamp VARCHAR(45)
Indexes


CTF
ctfId INT(11)
motionCorrectionId INT(11)
spectralImageFullPath VARCHAR...
defocusU VARCHAR(45)
defocusV VARCHAR(45)
angle VARCHAR(45)
resolutionLimit VARCHAR(45)
estimatedBfactor VARCHAR(45)
logFilePath VARCHAR(45)
createdTimeStamp VARCHAR(45)
Indexes

ACKNOWLEDGEMENTS

- Isai Kandiah, Alejandro de Maria, Solange Delagenière, Christoph Mueller-Dieckmann and Gordon Leonard (ESRF)
- Michael Hons (EMBL Grenoble)
- Gregory Effantin (IBS)
- Kevin Savage, Karl Levik, Neil Smith, Stuart Fischer and Alun Ashton (DLS)
- Pablo Conesa, Jose Miguel de la Rosa Trevin and Yaiza Rancel Gil (CNB)
- The ESRF Data Analysis Unit, Beamline Control Unit and Structural Biology group