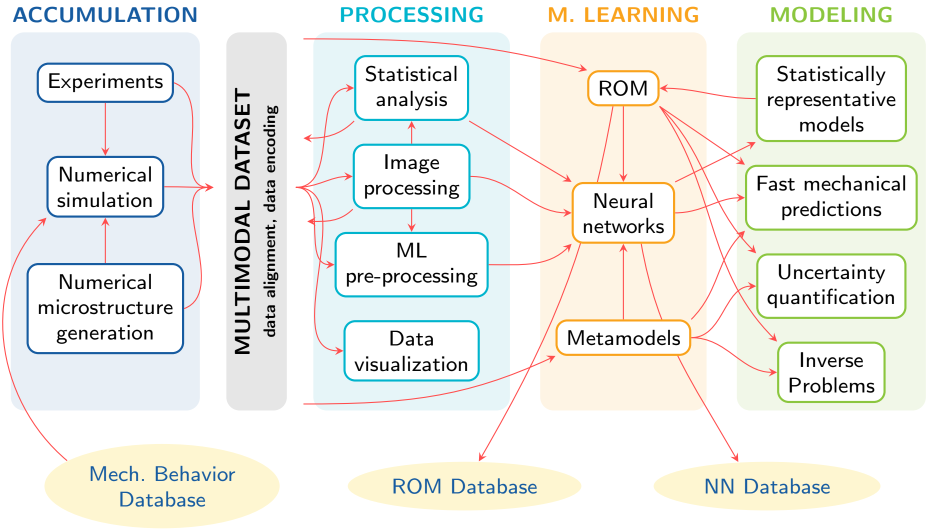
Orientations et Premiers Développements pour la Plate-forme de Données BIGMECA

A. Marano

COPIL des Travaux de la Chaire BIGMECA

8 octobre 2020

BIGMECA Chair Paradigm



Post-doc Goals

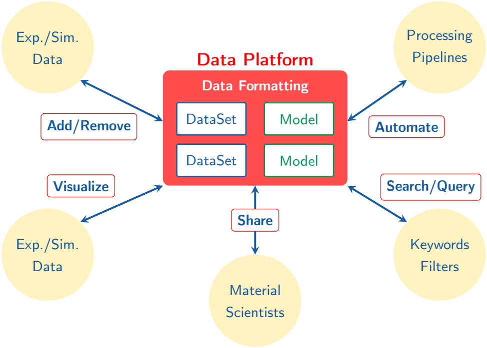
- Solve data management issues inherent to 4D material science workflows
- Implement a high level data platform package
- Digital twins numerical simulation (FFT methods, FEM, crystal plasticity)
- Support for C. Ribart and A. Aublet
- Applications :
  - microstructural uncertainty propagation
  - machine learning for hybrid modeling
  - slip localization ?

Table of Contents

BIGMECA data platform

Problems to solve

- Store and organize Large and Multimodal data
- Visualization
- Automation
- Open data (FAIR)
- High level interface
- Remote access



Review of existing platforms : synthesis

Literature review of scientific data platform in various fields (astronomy, heliophysics, geophysics, material science) available in **first technical report**.

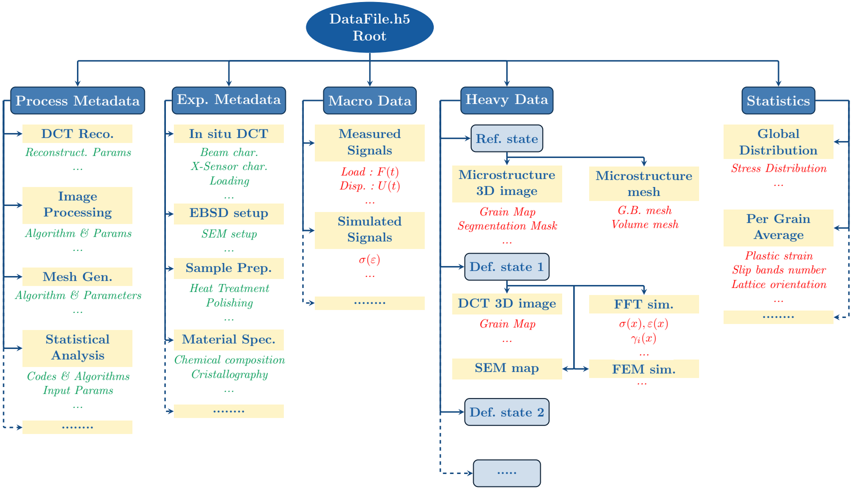
- Data standards set and used by the whole community : EOS, SeaDataNet, VOATable...
- Data production and processing tools based on data standards
- Flexible, scalable, hierarchical, self-descriptive file formats
- Service oriented architecture, high level interface
- Web applications (visualization, data processing), cloud computing/storage

Platform core design

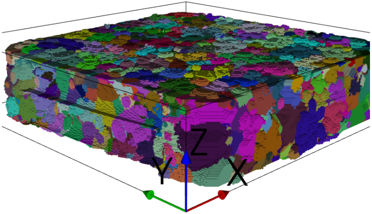
- Objective : Provide a simple Python API to add / remove / modify data in datasets
- File format : **HDF5** (*PyTables* Python packages)
  - Visualize file hierarchy and content : h5ls, h5dump, h5view, **ViTables**
- Systematically extended with a **XDMF** file (Paraview support)
  - XML Metadata model to describe arrays in HDF5 files
- Central class handling format, compression, data model, and HDF5/XDMF synchronization : **SampleData**
- Generic volume containers objects models for meshes and images
- Compatibility with VTK, Matlab, Zset outputs for now

Data Model

1 File = 1 physical asset = 1 Microstructure/Mechanical Sample

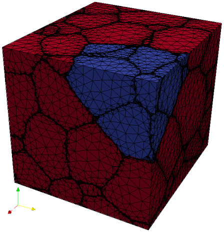


Use case 1 : DCT experiment

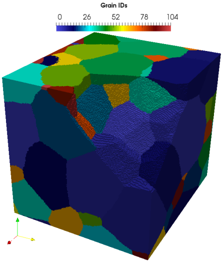


- Grain map X-DCT reconstruction $\sim 170^3$ voxels
- grain boundaries and volume mesh (F. Nguyen)
- FFT simulation

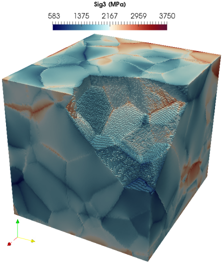
Grain Boundary Mesh



Grain Map



FFT simulation stress field



Use case 1 : DCT experiment

HDF5 file hierarchy (ViTables)

- AlLi_DCT_sample_visio.h5
  - Grain_data
    - Grains_subset
    - Grain_volumes
    - Grains_barycenter_coords
  - Volume_data
    - Image3D
      - Grain_map
      - Grain_subset
      - Sig3
    - Surface_mesh
      - Nodes
      - Elements
    - Volume_mesh
      - Nodes
      - Elements

Dataset creation

```
Sample = samples.SampleData(filename='AlLi_DCT_sample_visio',
                            sample_name= 'Test_sample_visio',
                            complib='zlib', bitshuffle=False)
```

3D Image container creation

```
Sample.add_image_from_file(imagefile= 'Mesh_test_input1_trans.mat',
                           imagename= 'Image3D',
                           location='/Volume_data/',
                           matlab_variables = ['/im3D'],
                           matlab_field_names =['Grain_map'])
```

Volume Mesh Attributes

Field_dim	3
element_topology	['Tetrahedron', 1]
mesh_description	3D volume mesh of linear tetrahedral elements
xdmf_geometry_path	Domain/Grid[3]/Geometry
xdmf_path	Domain/Grid[3]
xdmf_topology_path	Domain/Grid[3]/Topology

Data Compression

- Grain Map compression : naturally high compression ratios (few values)

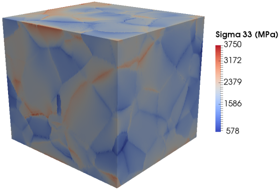
Compression Level	Library	File Size (Mo)	Ratio	Time (s)
0.00	none	10.76	1.00	0.12
1.00	zlib	0.40	**27.03**	0.14
1.00	lzo	0.70	15.39	0.14
1.00	blosc	2.99	3.60	0.17
5.00	zlib	0.28	**38.90**	0.21
5.00	lzo	0.70	15.39	0.16
5.00	blosc	0.68	15.71	0.17
9.00	zlib	0.21	**50.35**	0.46
9.00	lzo	0.70	15.39	0.14
9.00	blosc	0.68	15.71	0.16

Data Compression

- Stress field, **lossy** compression to achieve high ratios :
  $D = \text{round}\left[\frac{D-\bar{D}}{\sigma(D)}, 10^{-N}\right]$

Significant digits	Level	Library	File Size (Mo)	Ratio
None	0.00	none	21.37	1.00
4	1	zlib	8.27	2.58
2	1	zlib	3.10	6.90
1	1	zlib	1.33	16.02
1	9	zlib	0.94	22.68

uncompressed



2 significant digits

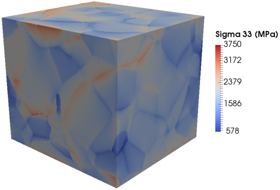
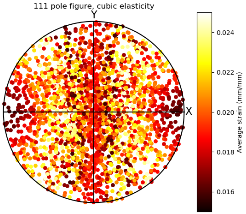
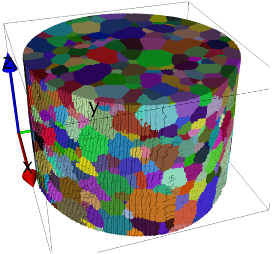


Table of Contents

Pymicro package (H. Proudhon)

- Open source Python package to manipulate and visualize 3D material data
- Designed for polycrystals
- Compatible with HDF5, DREAM3D datasets
- Crystallography tools
- Xray tomography, diffraction simulation tools
- FEM (Zset) tools

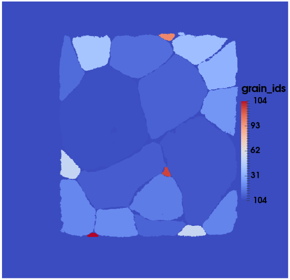




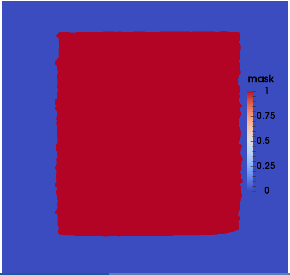
Integration (undergoing work)

- Introduce **SampleData** as Pymicro data management core
  - Integration between main classes of both codes
  - Construct datasets with organized data model from DCT files
  - easily transfer DCT datasets to HDF5 / XMDF files / compression
  - Pymicro class gain lazy structure access to data arrays
- Ensures maintenance, distribution and version control of the data platform code

DCT Grain map



Sample Mask



User friendly Data Model Implementation

- Access to data through keys : Index Dictionary
- Aliases system : user defined names for dataset features
- Aliases and Index stored in HDF5 file
- "Rigid" data model in memory but flexible interface

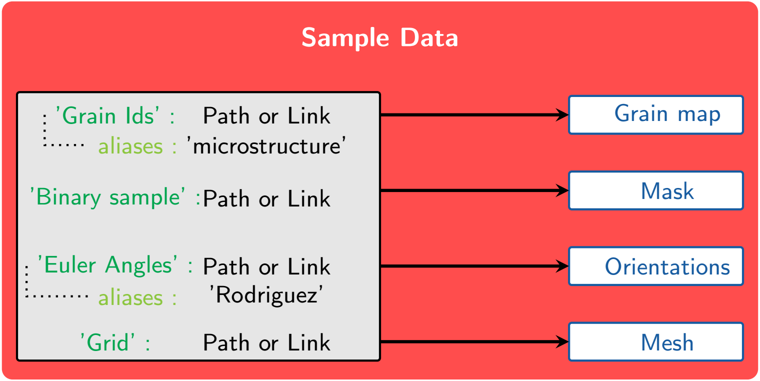
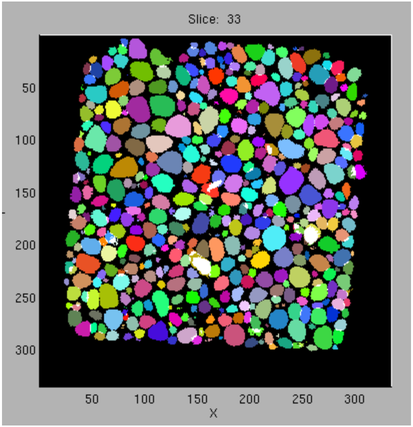


Table of Contents

Microstructural uncertainty with DCT reconstruction

- DCT 3D algorithm : unassigned areas
- DCT 6D : reconstruction confidence index field
- Grains dilatation and scans fusion
- Grain boundary exact position not known with certainty



Propagation in numerical simulations : Methodology I

1. Full microstructure :
   Grain map obtained with DCT reconstruction + grain dilatation + Grain Orientations

2. High resolution simulation ($\sim 5.10^8$ voxels) / highly parallel FFT solver / crystal plasticity

3. Extract and mesh subset of interest Ω. Identify grain boundary nodes positions $\boldsymbol{X}_0^{GB}$ and elements attached, forming Ω^{GB}

4. Stochastic representation of grain boundary positions :
   $\boldsymbol{X}^{GB}(\boldsymbol{\theta}) = \boldsymbol{X}_0^{GB} + \boldsymbol{u}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a random variable (vector)

5. Construct a parametric representation of $\boldsymbol{u}(\boldsymbol{\theta})$:
   Eigenmodes problem resolution on Ω^{GB} with $\boldsymbol{u}(\boldsymbol{x}) = 0$ on $\partial\Omega^{GB}$
   $\Longrightarrow \boldsymbol{u}(\boldsymbol{\theta}) = \sum_k \alpha_k(\theta_k)\,\psi_k$

Propagation in numerical simulations : Methodology II

6. Consider the $\alpha_k\,(\theta_k)$ as random centered gaussian variables and identify their variance to well represent the position uncertainty range (in a sense to be defined..)

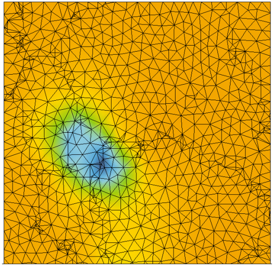
7. Use Monte-Carlo methods to generate hundreds of microstructures, and compute a FEM simulation (structural Zoom, Zset). Compute the distribution of interest quantities.

8. Machine Learning : during the Monte-Carlo process, mechanically cluster generated microstructures, and apply hyper-reduction to each cluster to accelerate future calculations

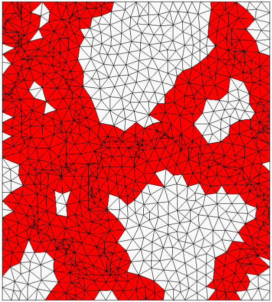
Grain boundary mesh deformation example

- One eigenmode ψ_2 of the parametric representation of $\boldsymbol{u}$.
- Elements in red are grain boundary neighbourhood elements in Ω^{GB}
- $\boldsymbol{u} = \alpha\psi_2$ applied to the original grain boundary mesh

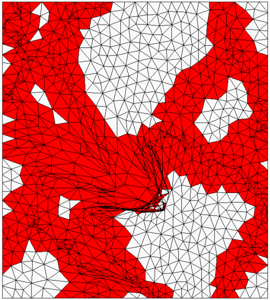
Eigen Mode (ψ_{2y})



Initial Mesh



Deformed Mesh



Challenges and Development

- Complex mesh handling with Zset and the platform $\Longrightarrow$ BasicTools ? (interface with BasicTools mesh objects)
- Mesh quality after microstructure generation ?
- Probabilistic indentification ?
- Possibility : voxelize deformed mesh, then FFT calculation again. FFT structural zoom techniques are emerging
- Automation
- Computational Resources ?

High Performance Computing at GENCI

- Resources allocation request addressed to GENCI
- 500.000 h requested on Jean Zay cluster at the moment
- Extensible allocation during the winter if the resources are granted
- Alternatives ?

Table of Contents

Conclusion and future works

- The data platform core is functional (data format, high level API, volume data storage, visualization, index dictionary)
- First release of Pymicro+SampleData coming soon !
- Many numerical simulations (FFT and FEM) in collaboration with C. Ribart
- Main focus in next months : uncertainty propagation chain specification and implementation
- Hopefully, applied on a HPC infrastructure