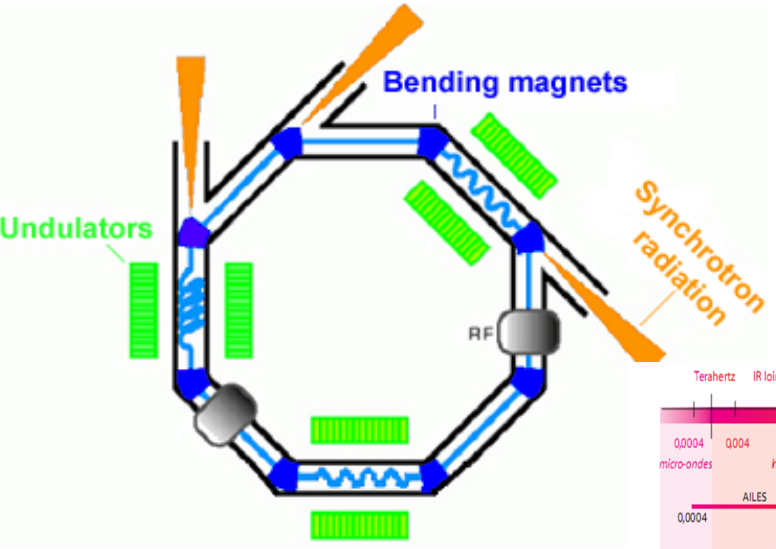
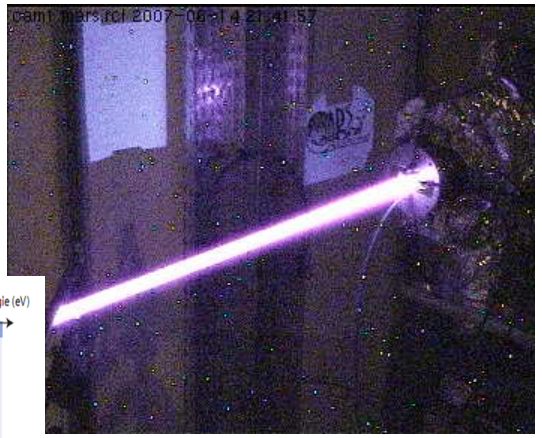
Challenges for Data Management and Treatment at a Large Scale Facility – Synchrotron – SOLEIL.

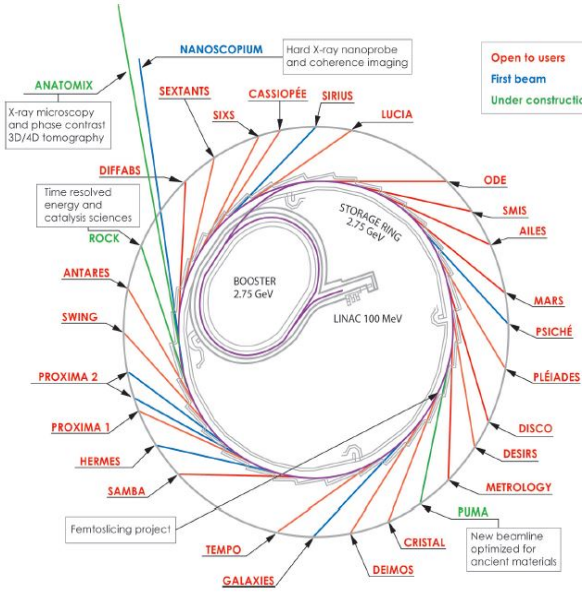
Andrew Thompson (Director for Life Sciences, SOLEIL) in collaboration with G. Kneller (CBM, Orleans).

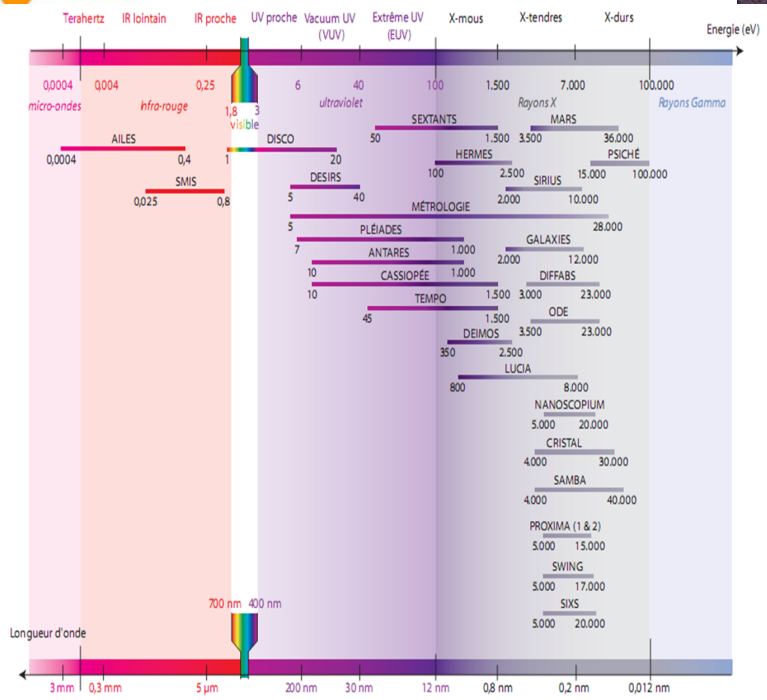
- Brief Introduction to SOLEIL
- Current trends
- Several examples of data collection and data type
- Problems posed by explosion of data volume

SOLEIL in a nutshell



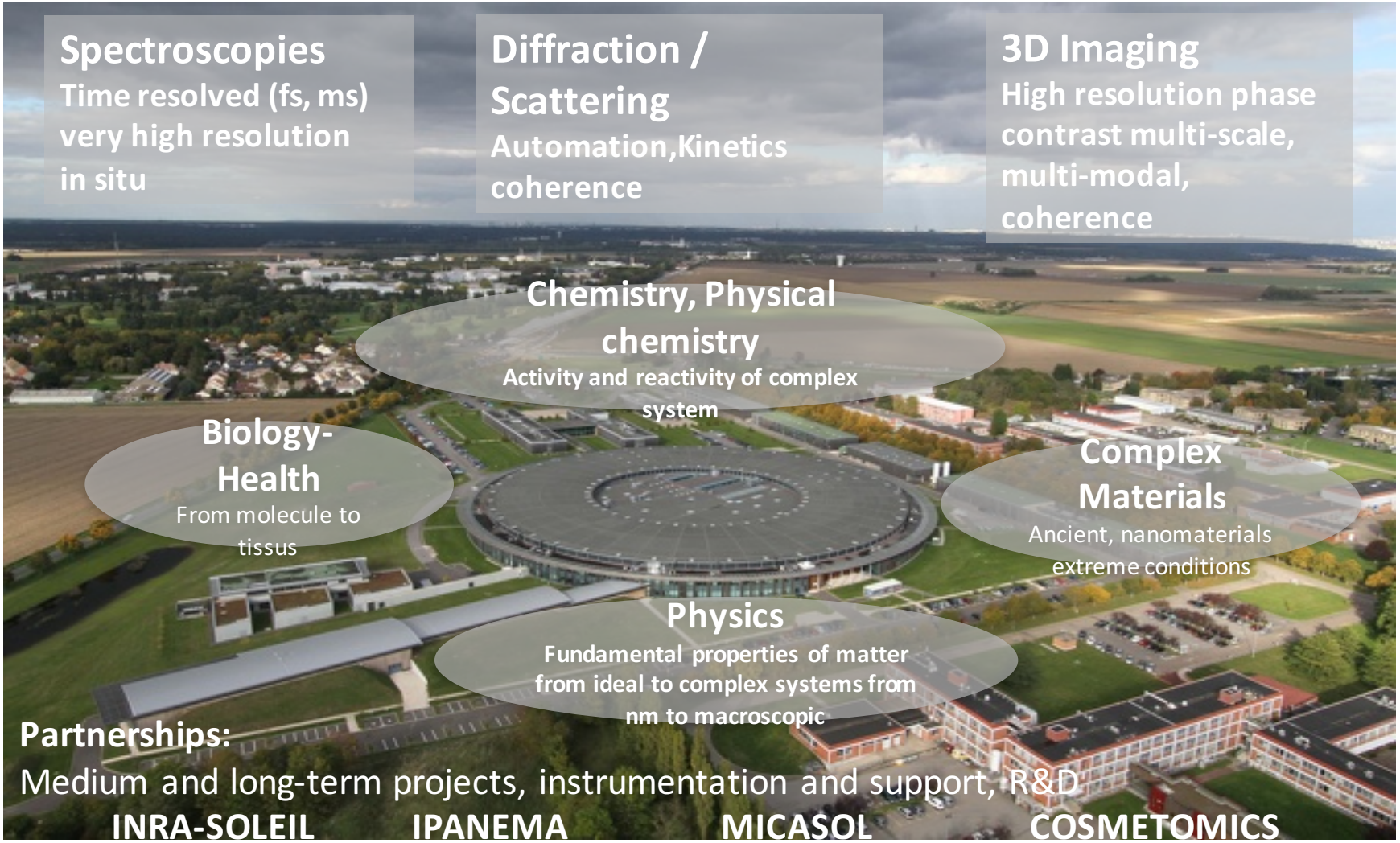






- 29 beamlines open to users.
- 3 beamlines on project funding (EquipEx,CPER).
- 21 on insertion devices; 8 on bending magnets.

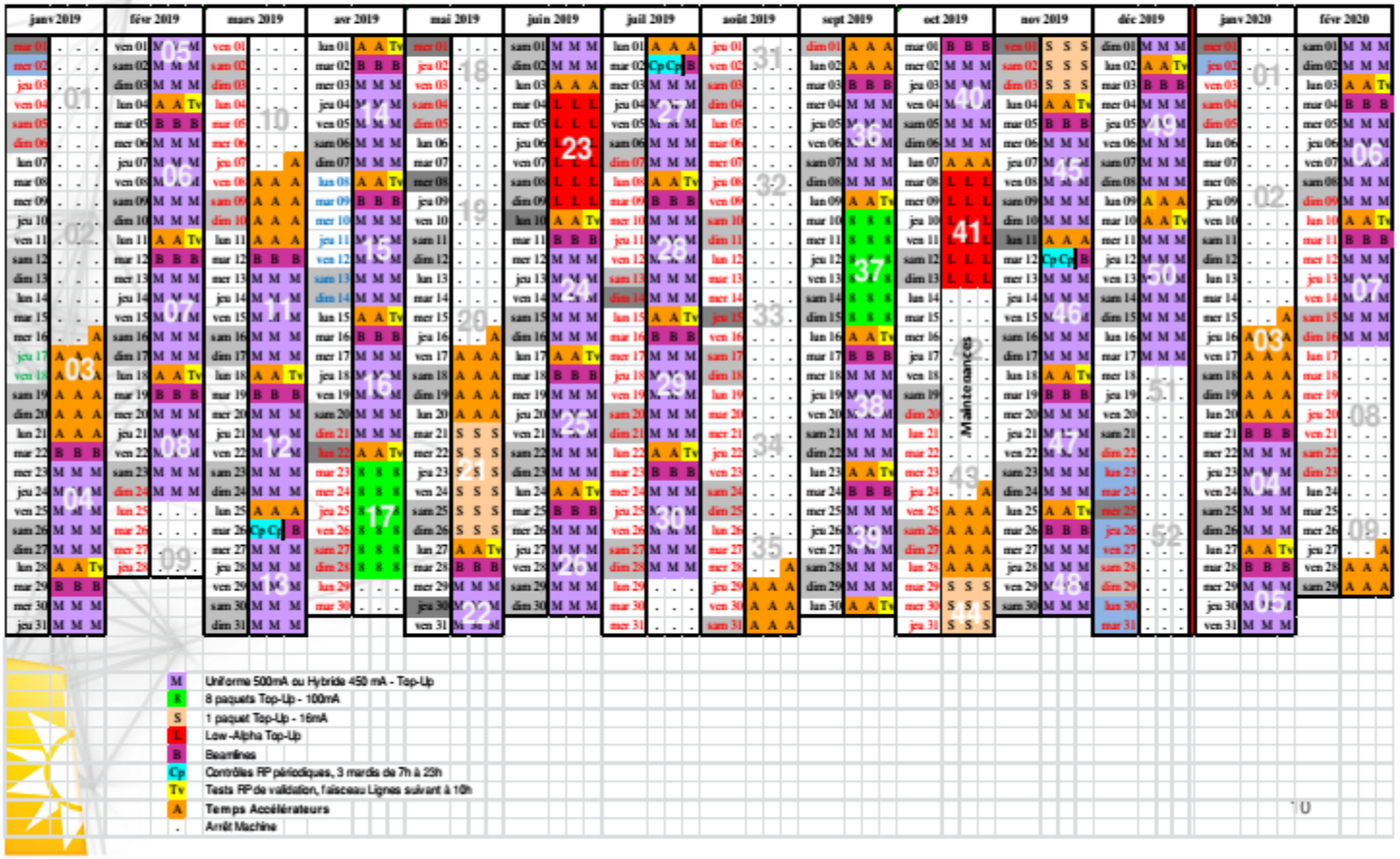
Scientific Strategy



Beam Time Schedule in 2019

SOLEIL SYNCHROTRON

5019 hours for the beamlines.



Approximately 200 days of beam supplied per year on 29 different experimental facilities. The highest producing beamlines collect ~ 1 terabyte today, but increasing …

Current trends in Synchrotron Facilities.

- Evolving synchrotron sources and optics. Developing tools to study heterogeneous objects at low (few microns) and high spatial resolution (5 – 10 nm).
- Increased multimodal measurements (hyperspectral imaging).
- Increased detector speeds and scanning speeds / need for statistically representative cohort of data.

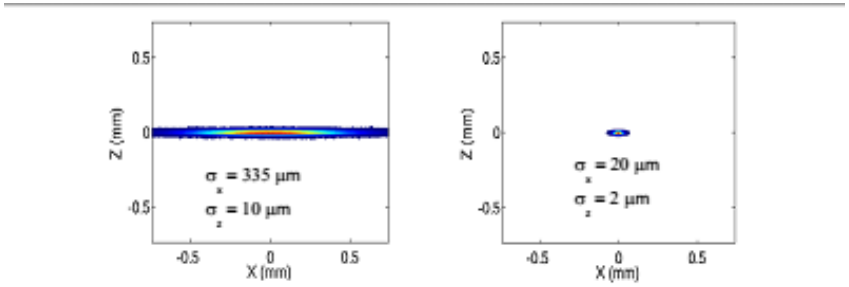
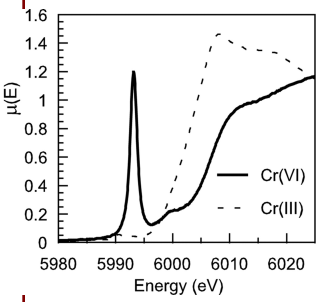


Figure 25: Comparison of the transverse beam profiles of SOLEIL and SOLEIL upgrade baseline lattice in a short straight section.

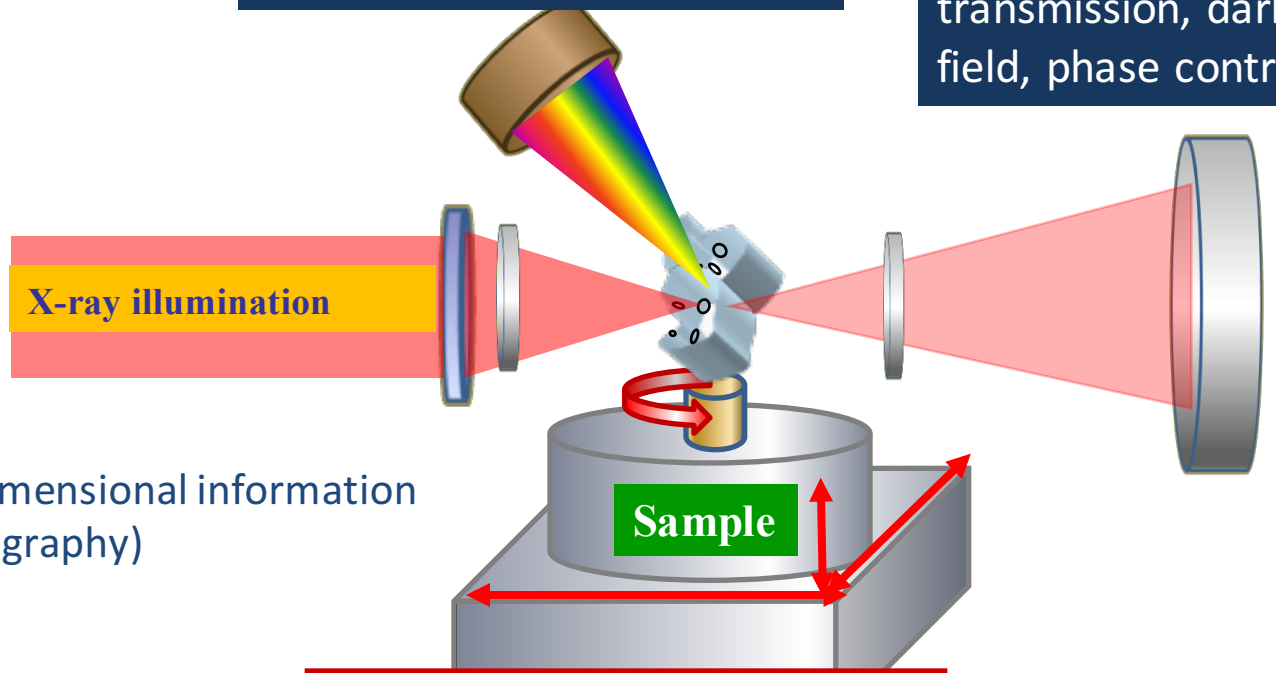
Scanning hard X-ray nano-imaging at Nanoscopium: a multi-technique tool for simultaneous trace metal and morphology studies

Image is collected step-by step

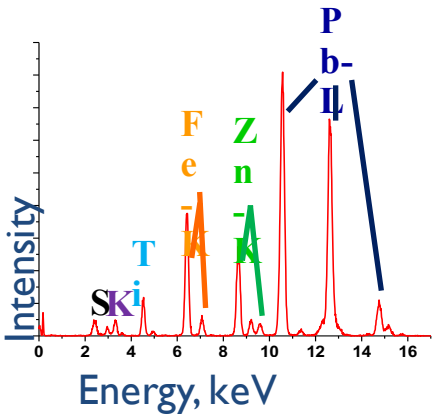


Detector 1:
Composition, chemical state

Detector2:
transmission, dark field, phase contrast



2/3 dimensional information (tomography)

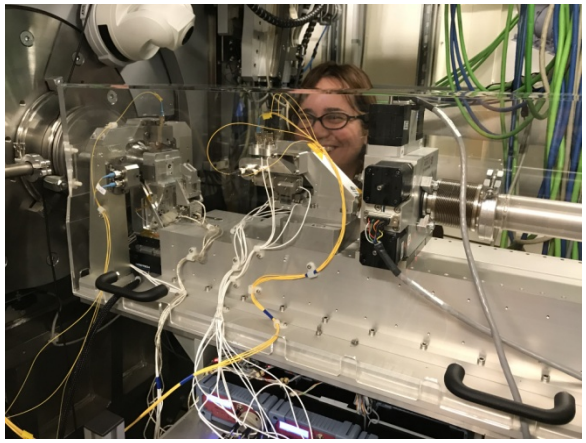


Resolution: ~30-400 nm

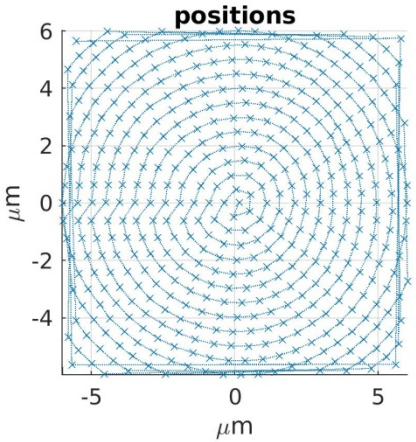
Cryo-cooling being developed.
Nanoscopium team (A. Somogyi, K. Medjoubi, G. Baranton)

2D-Ptychography at beamline Swing

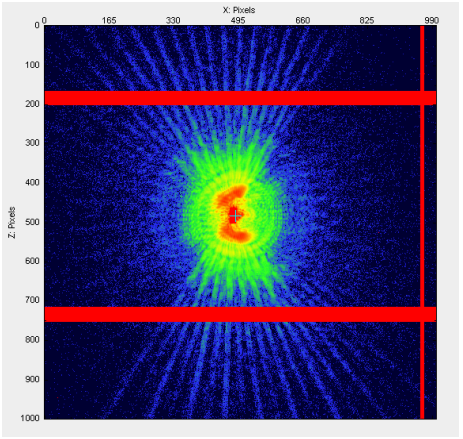
All ptychographic results generated via cSaxs ptychography application (PSI)
(Pierre Thibault, Manuel Guizar-Sicairos and coll., 2008-2018)



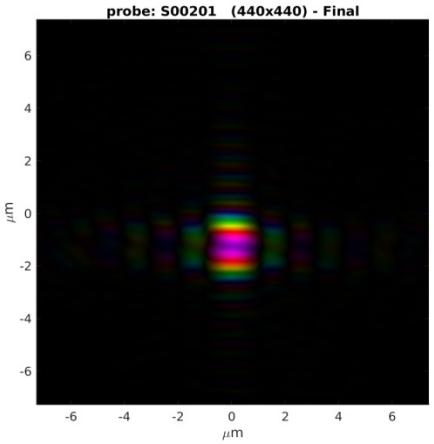
2D Nanoprobe set-up installed on Swing beamline (Kubsky, Somogyi, MAX IV)



Spiral scan generated via *Passerelle* application



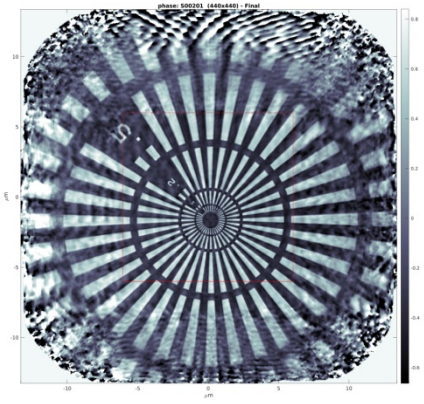
Diffraction pattern on Eiger4M detector at 6.2 m
All data collected using *Tango* devices.



Calculated beam shape at sample position
FZP (f=164mm) + aperture (80 x 20 μm²)

Main involved people:

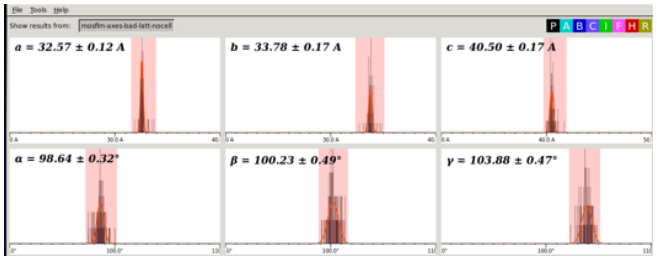
Ch. Engbloem (ECA)
F. Alves (CI)
A. Lestrade (ALI)
Y.M. Abiven (ECA)
F. Langlois (ICA)
P. Montaville (PX1)
F. Berenguer (CRISTAL)
J. Pérez (SWING)

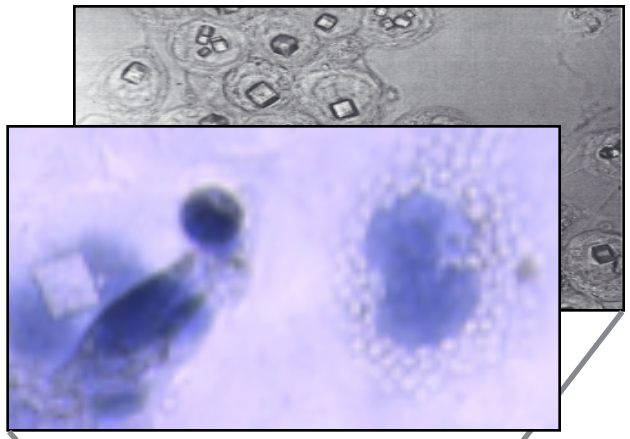


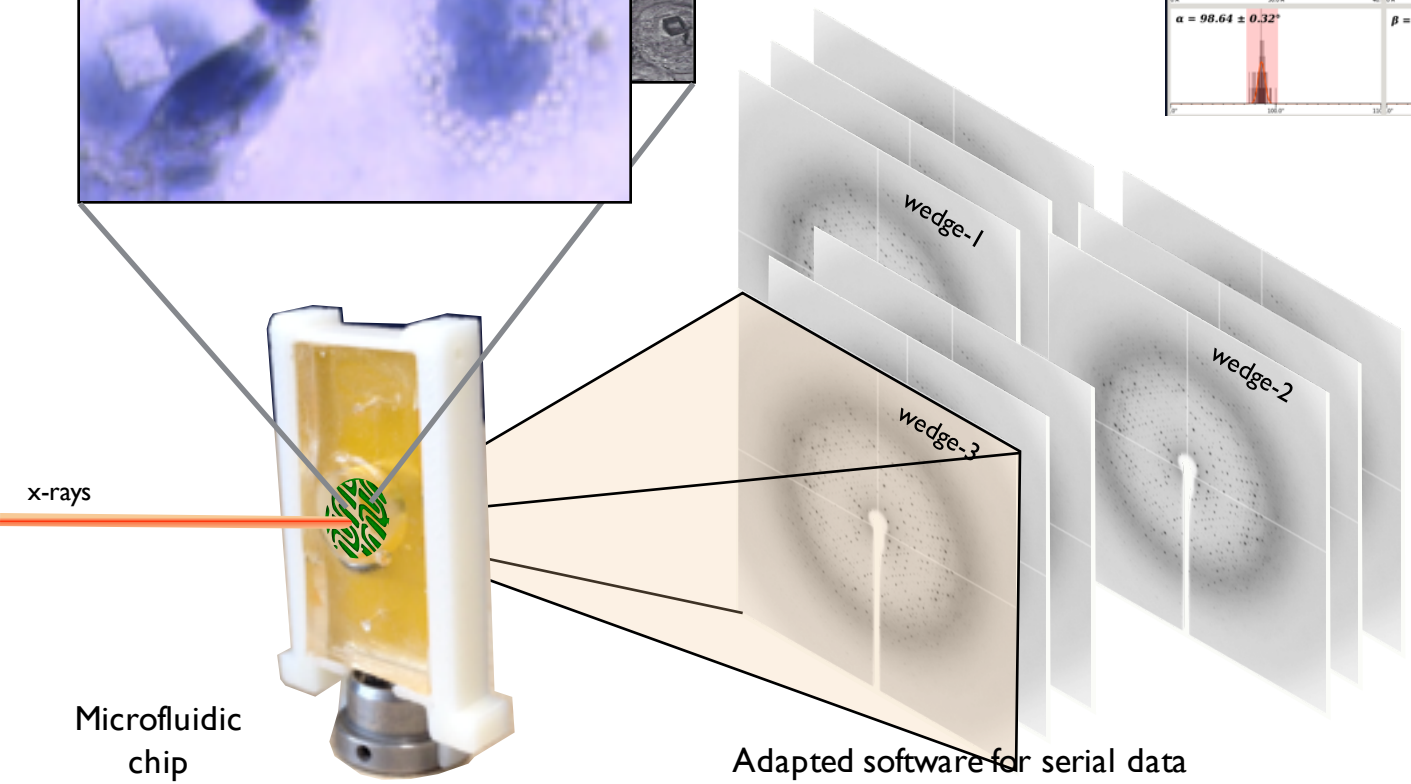
Phase map obtained after 400 iterations. Inner spacings : 50 nm

Application to
in vivo grown crystals –
satellite to SUM 2018

Data processing approaches inspired by x-ray FEL methods

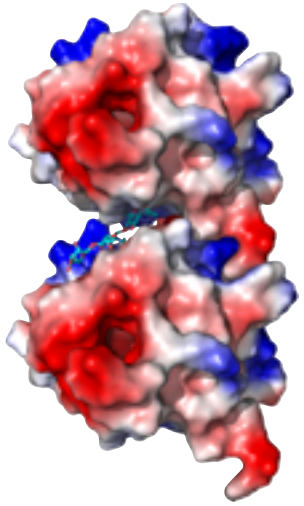






Microfluidic
chip

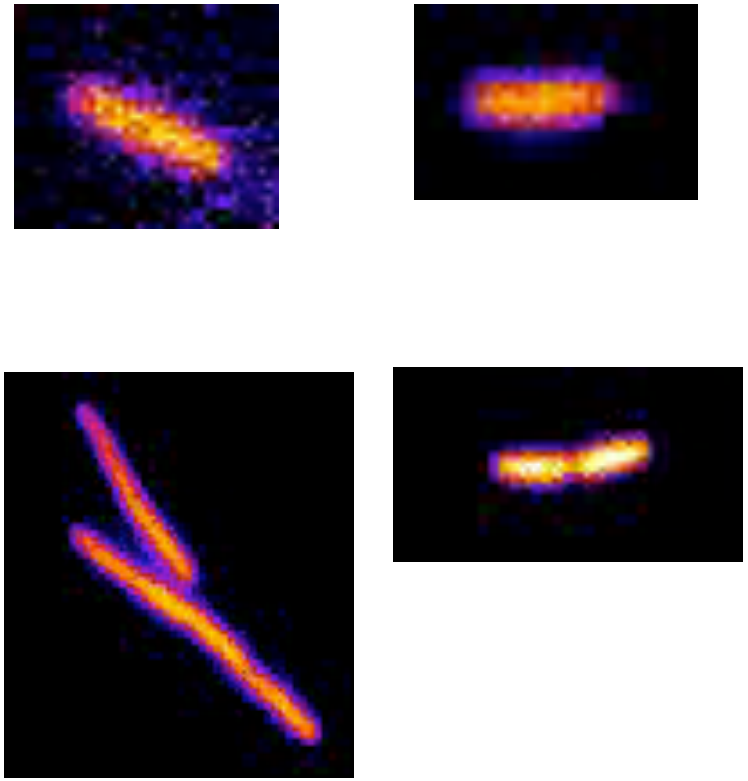
Adapted software for serial data
collection

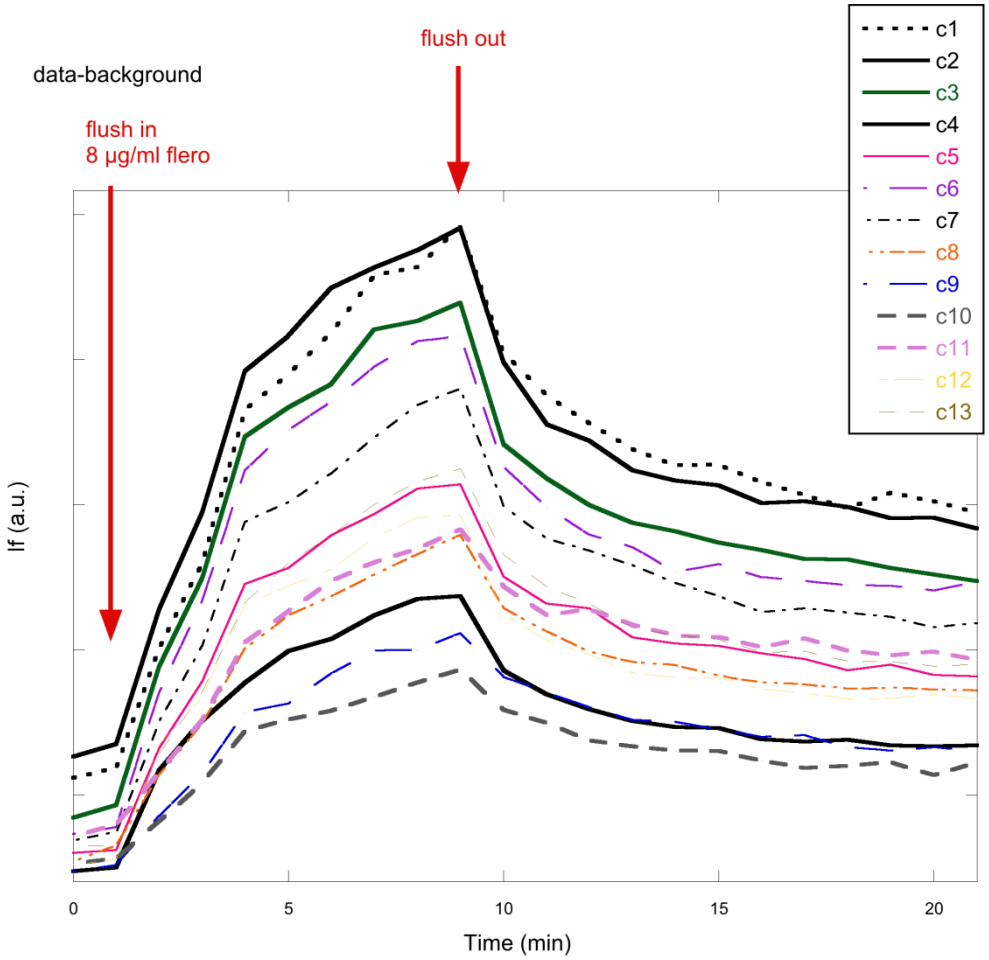


Structure determination of unmodified targets

Serial methods developed on PROXIMA 1
(L. Chavas) and PROXIMA 2a (W. Shepard)

Individualising the analysis of each bacterium



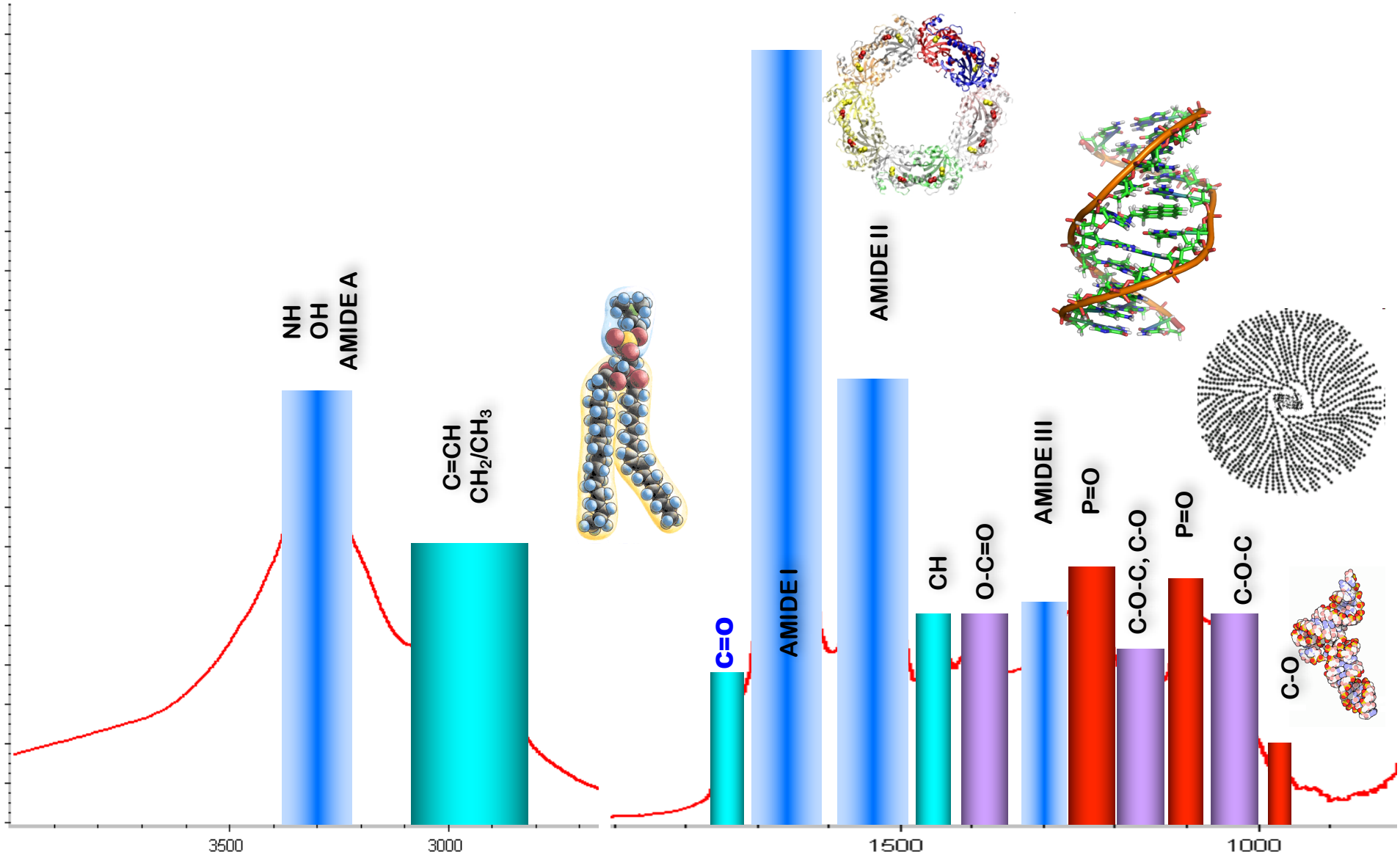


DISCO beamline : M. Refregiers et al.)

raw data

By analysing each individual curve, singular behaviour (mutations, reactions, …) may be isolated.

SYNCHROTRON INFRARED MICROSCOPY:
MOLECULAR FINGERPRINTS



Imaging micrometeorites

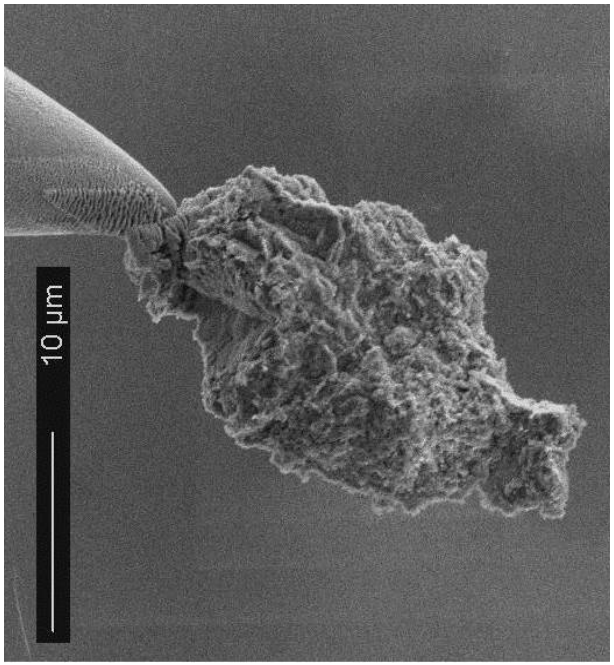
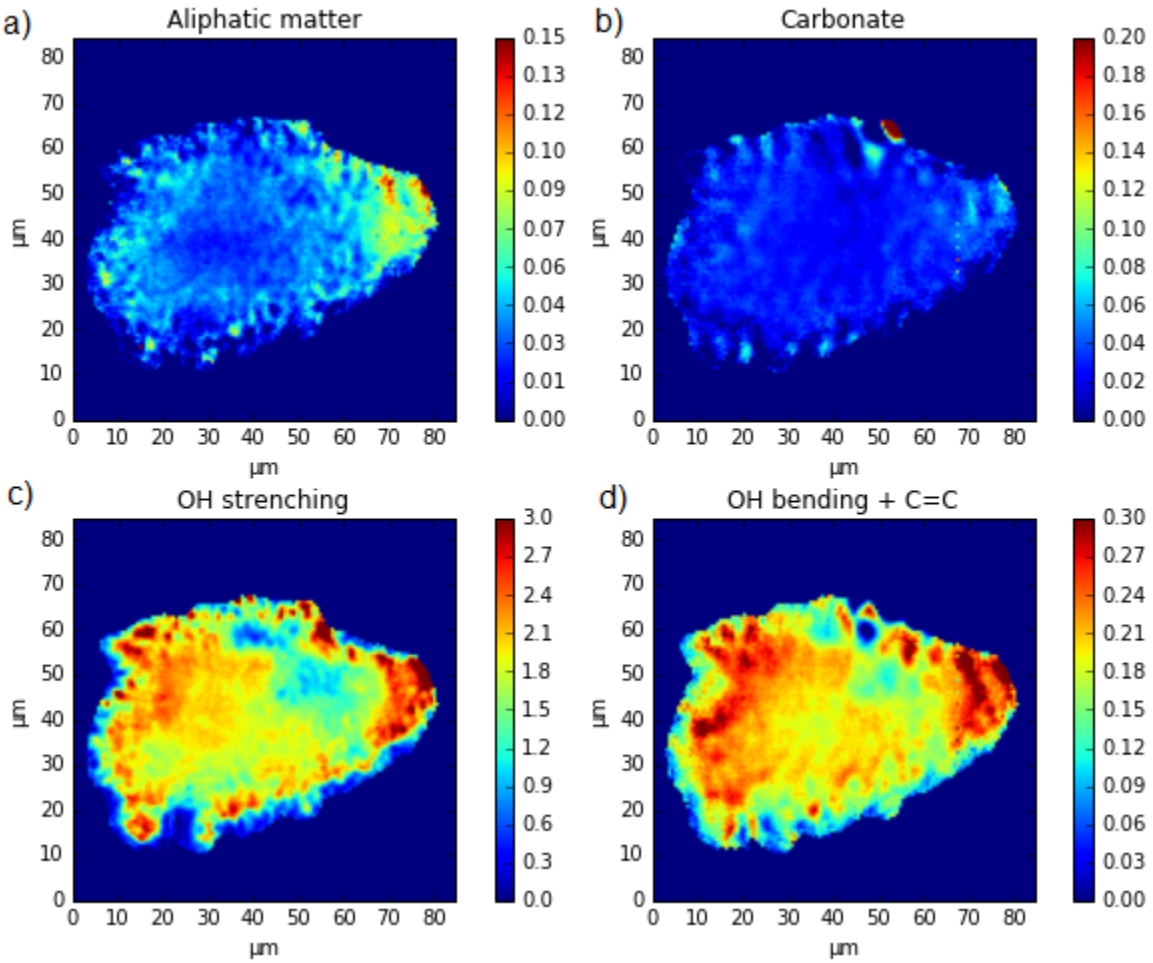
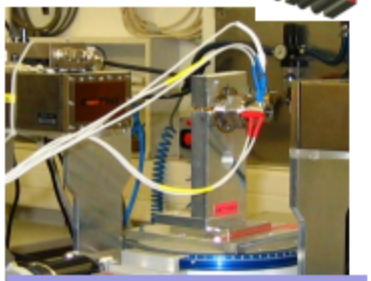


Image from FEM



SMIS Beamline – F. Borondics et al.

ROCK : Road Map

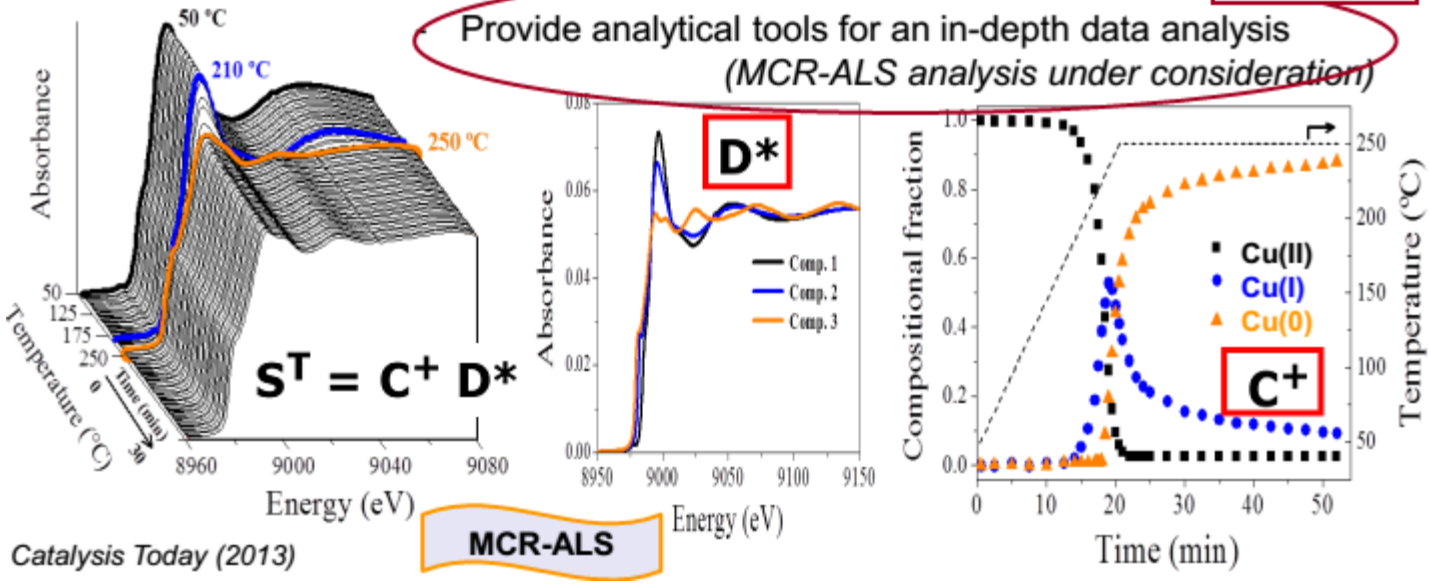


Electrochemical cells

- Provide equipments for the study of materials in working conditions : Some already existing, new under consideration *(Cells for reactions between liquid/solid; cryo-oven…)*
- Offer combination of techniques (Raman, UV-V… and propose new ones in routine (XRD)

V.Briois, SAC 2013

- Provide analytical tools for an in-depth data analysis *(MCR-ALS analysis under consideration)*

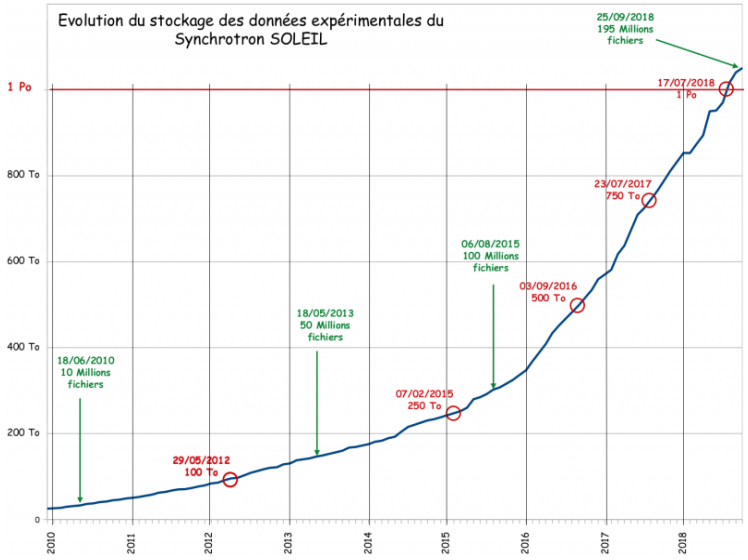


Catalysis Today (2013)

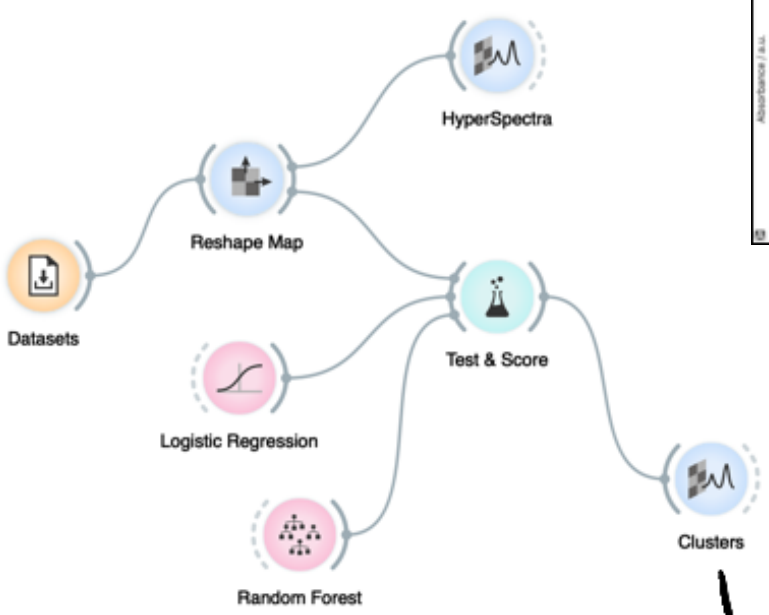
Time resolved spectroscopic measurements.
Extension towards 2D hyperspectral images with high spatial resulution.
ROCK beamline – V. Briois et al.

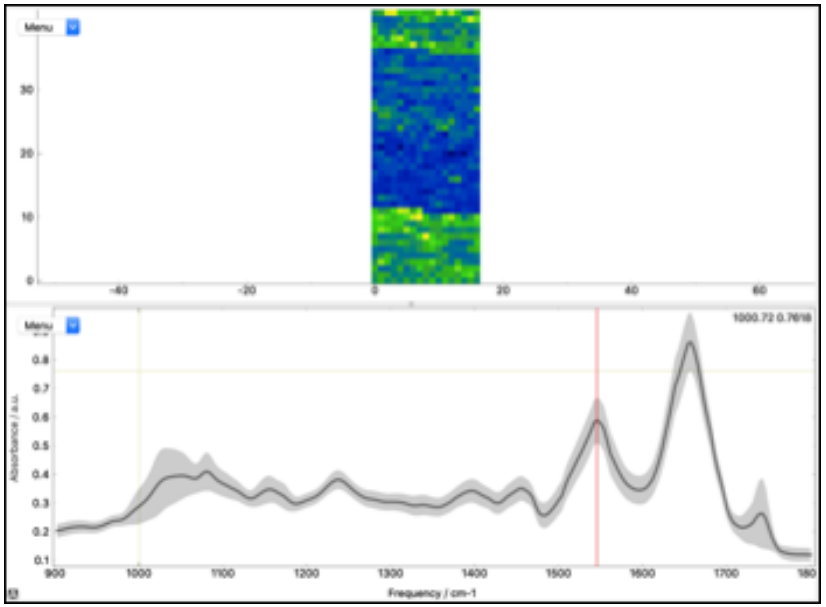
- **Towards an explosion of experimental data production :**
  - Fast 2D detectors produce very large data volumes
  - Flyscan data acquisition techniques for fast, simultaneous multi-detector scans @ high spatial resolution
  - Correlative imaging,
  - Time resolved spectroscopy,

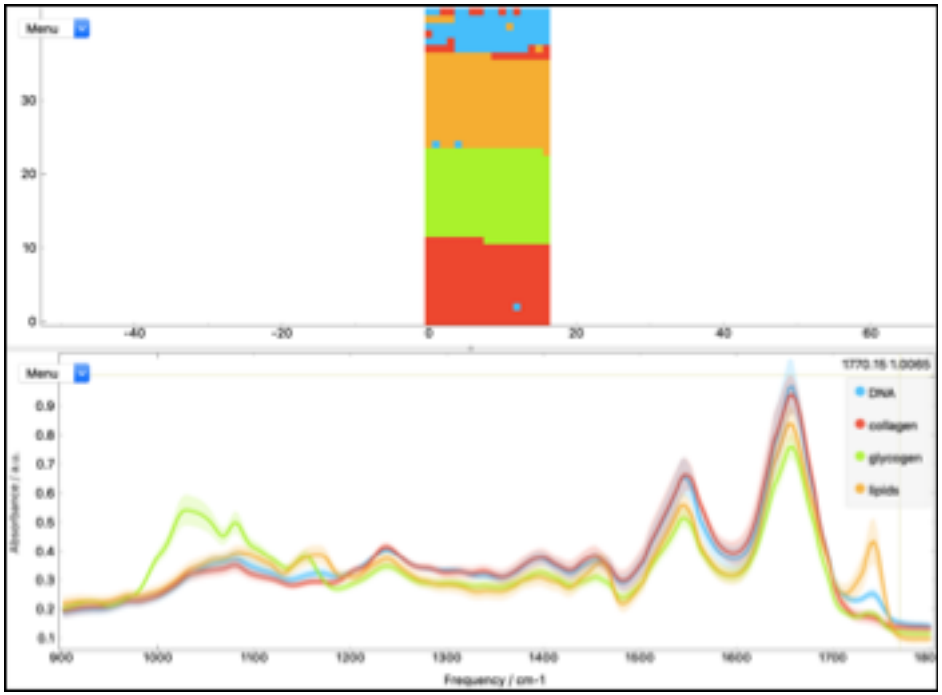
Based on projection of detector installation for the next 2 or 3 years, we expect more than 2 petabytes per year ! Two existing beamlines already collect > 1 terabyte per day.



quasar







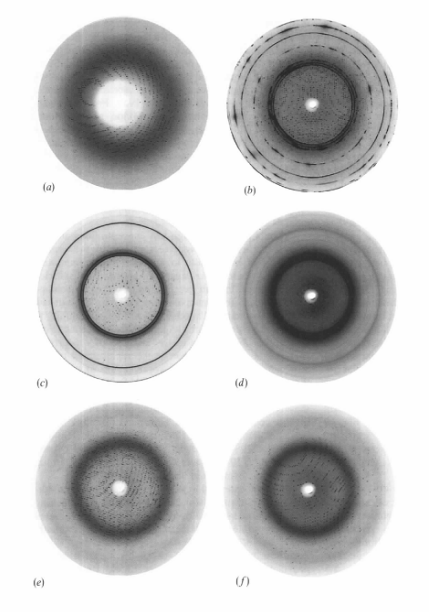
https://quasar.codes

https://github.com/quasars

Check online tutorial videos!

Conclusions

- **Soon facing the big data challenges : to manage and process large amounts of data**
  - Data handling and processing increasingly complex
  - Data too big to transfer
  - Online data reduction and analysis mandatory to decide what to do next
  - Need to avoid collecting and storing lots of unusable data
  - Online and offline data reduction and analysis requiring big compute resources and top level tools
  - …

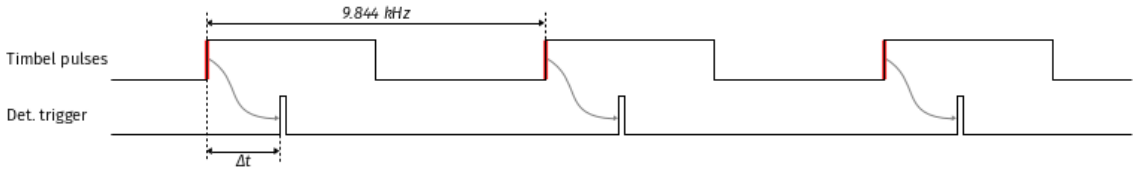


Need online evaluation tools in all cases, so as to make experimental choices during beamtime.

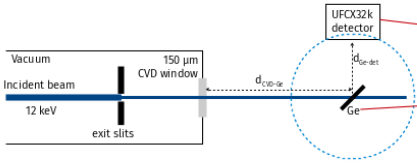
Sometimes it is visually easy to say what you should keep and what is rubbish, but not always.
It is impossible to annotate data by hand with collection speeds of « a few » to « several hundreds » of acquisitions per second…..

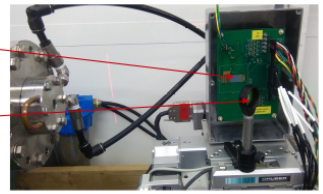
Need for « decision making software », tailored to different experimental methods. Machine learning with intelligence and mechanism adapted to different data types and experimental protocols?

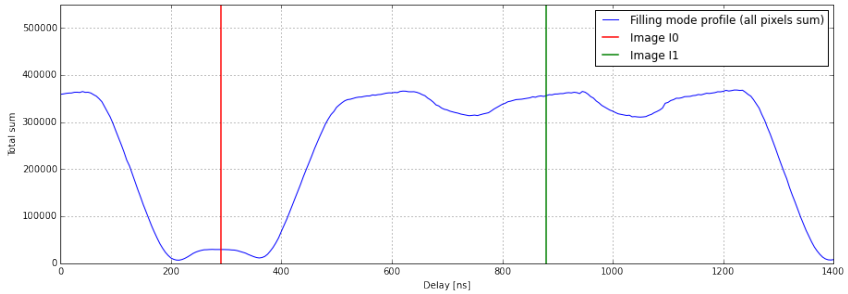
Contact :
Andrew.thompson@synchrotron-soleil.fr

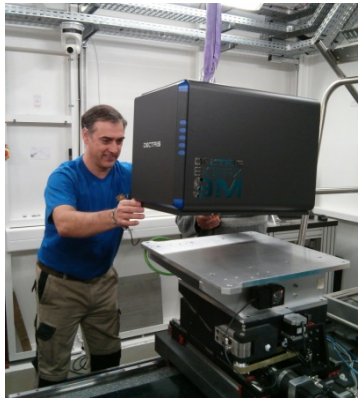


UFXC32k, collaboration with AGH in Cracow - 300µm Si, 75µm pixels, 2 module detector under construction (19.2 x 19.2 mm, 66 K pixels), 16 module detector will follow. Time resolution (in burst mode) is ~ 100 ns









18 mega pixels 32 bits @ 133 frames per second
Faster detectors under development (for example Jungfrau detector at PSI, 1.1 kHz frame rate, large area)