

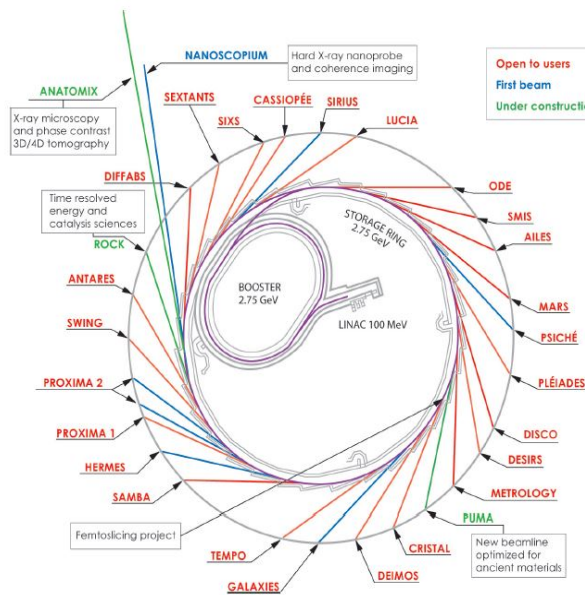
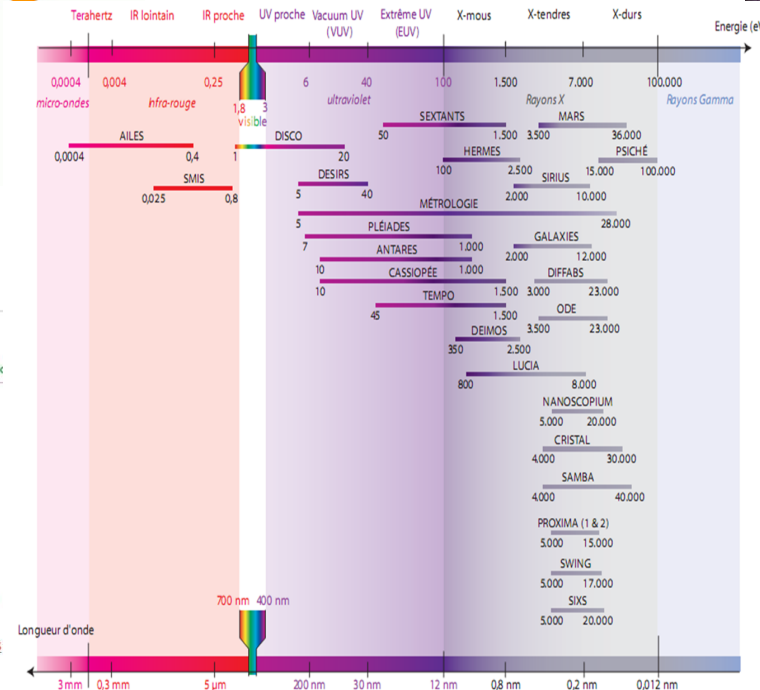
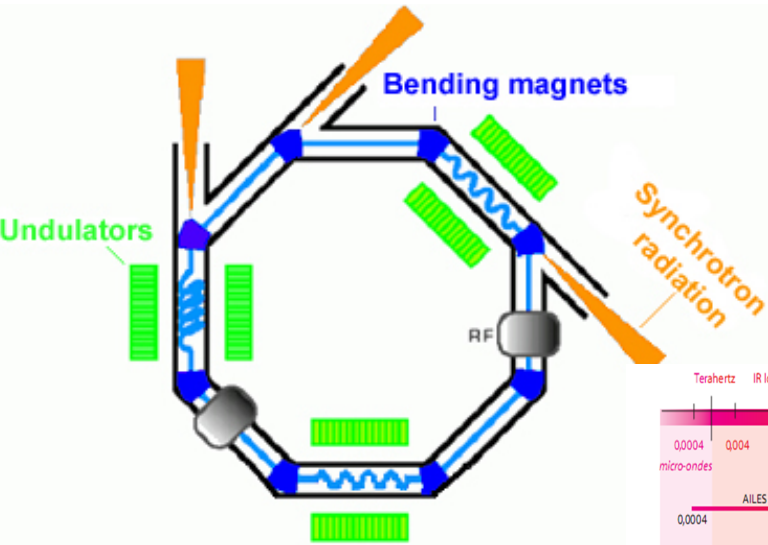
Challenges for Data Management and Treatment at a Large Scale Facility – Synchrotron – SOLEIL.

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

Spectroscopies

Time resolved (fs, ms)
very high resolution
in situ

Diffraction / Scattering

Automation, Kinetics
coherence

3D Imaging

High resolution phase
contrast multi-scale,
multi-modal,
coherence

Chemistry, Physical chemistry

Activity and reactivity of complex
system

Biology- Health

From molecule to
tissus

Complex Materials

Ancient, nanomaterials
extreme conditions

Physics

Fundamental properties of matter
from ideal to complex systems from
nm to macroscopic

Partnerships:

Medium and long-term projects, instrumentation and support, R&D

INRA-SOLEIL

IPANEMA

MICASOL

COSMETOMICS



5019 hours for the beamlines.

Beam Time Schedule in 2019

janv 2019		févr 2019		mars 2019		avr 2019		mai 2019		juin 2019		juil 2019		août 2019		sept 2019		oct 2019		nov 2019		déc 2019		janv 2020		févr 2020			
mar 01		ven 01	M M M	ven 01		lan 01	A A Tv	mer 01		sam 01	M M M	lan 01	A A A	jeu 01		dim 01	A A A	mar 01	B B B	ven 01	S S S	dim 01	M M M	mer 01		sam 01	M M M	lan 01	M M M
mer 02		sam 02	M M M	sam 02		mar 02	B B B	jeu 02		dim 02	M M M	mar 02	Cp Cp B	ven 02		lan 02	A A A	mer 02	M M M	sam 02	S S S	lan 02	A A Tv	jeu 02		dim 02	M M M	lan 02	M M M
jeu 03		dim 03	M M M	dim 03		mer 03	M M M	ven 03		lan 03	A A A	mer 03	M M M	sam 03		mar 03	B B B	jeu 03	M M M	dim 03	S S S	mar 03	B B B	ven 03		lan 03	A A Tv	lan 03	M M M
ven 04		lan 04	A A Tv	lan 04		jeu 04	M M M	sam 04		mer 04	L L L	jeu 04	M M M	dim 04		ven 04	M M M	mer 04	M M M	ven 04	A A Tv	mer 04	M M M	sam 04		lan 04	A A Tv	lan 04	M M M
sam 05		mar 05	B B B	mar 05		ven 05	M M M	dim 05		mer 05	L L L	ven 05	M M M	lan 05		jeu 05	M M M	sam 05	M M M	mar 05	B B B	jeu 05	M M M	dim 05		mer 05	M M M	lan 05	M M M
dim 06		mer 06	M M M	mer 06		sam 06	M M M	lan 06		jeu 06	23	sam 06	M M M	mer 06		ven 06	M M M	dim 06	M M M	mer 06	M M M	ven 06	M M M	lan 06		jeu 06	M M M	lan 06	M M M
lan 07		jeu 07	M M M	jeu 07		dim 07	M M M	mar 07		ven 07	L L L	dim 07	M M M	mer 07		sam 07	M M M	lan 07	M M M	jeu 07	M M M	dim 07	M M M	mer 07		jeu 07	M M M	lan 07	M M M
mar 08		ven 08	M M M	ven 08		lan 08	A A Tv	mer 08		sam 08	L L L	lan 08	A A Tv	jeu 08		dim 08	M M M	mar 08	L L L	ven 08	M M M	dim 08	M M M	mer 08		jeu 08	M M M	lan 08	M M M
mer 09		sam 09	M M M	sam 09		mar 09	B B B	jeu 09		dim 09	L L L	mar 09	B B B	ven 09		lan 09	A A Tv	mer 09	M M M	dim 09	M M M	lan 09	A A A	jeu 09		dim 09	M M M	lan 09	M M M
jeu 10		dim 10	M M M	dim 10		mer 10	M M M	ven 10		lan 10	A A Tv	mer 10	M M M	sam 10		mar 10	M M M	jeu 10	M M M	dim 10	M M M	mar 10	A A Tv	ven 10		lan 10	A A Tv	lan 10	M M M
ven 11		lan 11	A A Tv	lan 11		jeu 11	M M M	sam 11		mar 11	B B B	jeu 11	M M M	dim 11		mer 11	M M M	ven 11	M M M	lan 11	A A A	jeu 11	M M M	sam 11		mar 11	M M M	lan 11	M M M
mar 12		mer 12	B B B	mer 12		ven 12	M M M	dim 12		lan 12	M M M	ven 12	M M M	lan 12		jeu 12	M M M	sam 12	M M M	jeu 12	M M M	dim 12	M M M	mer 12		sam 12	M M M	lan 12	M M M
dim 13		mer 13	M M M	mer 13		sam 13	M M M	lan 13		jeu 13	M M M	sam 13	M M M	mar 13		ven 13	M M M	dim 13	M M M	mer 13	M M M	ven 13	M M M	lan 13		jeu 13	M M M	lan 13	M M M
lan 14		mer 14	M M M	jeu 14		lan 14	M M M	mar 14		ven 14	M M M	dim 14	M M M	mer 14		sam 14	M M M	lan 14	M M M	jeu 14	M M M	sam 14	M M M	mar 14		jeu 14	M M M	lan 14	M M M
mar 15		ven 15	M M M	ven 15		lan 15	A A Tv	mer 15		sam 15	M M M	lan 15	A A Tv	jeu 15		dim 15	M M M	mar 15	M M M	ven 15	M M M	dim 15	M M M	mer 15		sam 15	M M M	lan 15	M M M
mer 16		sam 16	M M M	sam 16		mar 16	B B B	jeu 16		dim 16	M M M	mar 16	B B B	ven 16		lan 16	A A Tv	mer 16	M M M	sam 16	M M M	lan 16	M M M	jeu 16		dim 16	M M M	lan 16	M M M
jeu 17		dim 17	M M M	dim 17		mer 17	M M M	ven 17		lan 17	A A Tv	mer 17	M M M	sam 17		mar 17	M M M	jeu 17	M M M	dim 17	M M M	mar 17	M M M	ven 17		lan 17	M M M	lan 17	M M M
ven 18		lan 18	A A Tv	lan 18		jeu 18	M M M	sam 18		mar 18	B B B	jeu 18	M M M	dim 18		mer 18	M M M	ven 18	M M M	lan 18	A A Tv	mer 18		sam 18		mar 18	M M M	lan 18	M M M
sam 19		mar 19	B B B	mar 19		ven 19	M M M	dim 19		lan 19	A A A	mer 19	M M M	lan 19		jeu 19	M M M	sam 19	M M M	mar 19	B B B	jeu 19		dim 19	A A A	mar 19	M M M	lan 19	M M M
dim 20		mer 20	M M M	mer 20		sam 20	M M M	lan 20		jeu 20	M M M	sam 20	M M M	mar 20		ven 20	M M M	dim 20	M M M	mer 20	M M M	ven 20		lan 20	A A A	jeu 20		dim 20	M M M
lan 21		jeu 21	M M M	jeu 21		dim 21	M M M	mar 21		ven 21	M M M	dim 21	M M M	mer 21		sam 21	M M M	lan 21	M M M	jeu 21	M M M	sam 21		mar 21	B B B	ven 21		lan 21	M M M
mer 22		mar 22	B B B	mar 22		ven 22	M M M	mer 22		sam 22	M M M	lan 22	A A Tv	jeu 22		dim 22	M M M	mer 22	M M M	ven 22	M M M	dim 22	M M M	mer 22		sam 22	M M M	lan 22	M M M
mer 23		sam 23	M M M	sam 23		mer 23	M M M	jeu 23		dim 23	M M M	mar 23	B B B	ven 23		lan 23	A A Tv	mer 23	M M M	sam 23	M M M	lan 23		jeu 23	M M M	dim 23		lan 23	M M M
jeu 24		dim 24	M M M	dim 24		mer 24	M M M	ven 24		lan 24	A A Tv	mer 24	M M M	sam 24		mar 24	B B B	jeu 24	M M M	dim 24	M M M	mar 24		jeu 24	M M M	lan 24		lan 24	M M M
ven 25		lan 25	M M M	lan 25		jeu 25	M M M	sam 25		mar 25	B B B	jeu 25	M M M	dim 25		mer 25	M M M	ven 25	M M M	lan 25	A A Tv	mer 25		sam 25	M M M	mar 25		lan 25	M M M
sam 26		mar 26	M M M	mar 26		ven 26	M M M	dim 26		mer 26	M M M	ven 26	M M M	lan 26		jeu 26	M M M	sam 26	A A A	mar 26	B B B	jeu 26		dim 26	M M M	mer 26		lan 26	M M M
dim 27		mer 27	M M M	mer 27		sam 27	M M M	lan 27		jeu 27	M M M	sam 27	M M M	mar 27		ven 27	M M M	dim 27	M M M	mer 27	M M M	ven 27		lan 27	A A Tv	jeu 27		lan 27	M M M
lan 28		jeu 28	A A Tv	jeu 28		dim 28	M M M	mar 28		ven 28	M M M	dim 28	M M M	mer 28		sam 28	M M M	lan 28	A A A	mer 28	A A A	lan 28		jeu 28	M M M	dim 28		lan 28	M M M
mar 29		ven 29	M M M	ven 29		jeu 29	M M M	mer 29		sam 29	M M M	lan 29		jeu 29		dim 29	M M M	mar 29	S S S	jeu 29	M M M	dim 29		lan 29	M M M	mer 29		sam 29	M M M
mer 30		sam 30	M M M	sam 30		jeu 30	M M M	dim 30		mer 30	M M M	lan 30		jeu 30		dim 30	M M M	mer 30	A A A	lan 30	A A A	mer 30		sam 30	M M M	jeu 30		lan 30	M M M
jeu 31		dim 31	M M M	dim 31		ven 31	M M M	lan 31		jeu 31		sam 31		sam 31		lan 31	A A A	mer 31	S S S	jeu 31	S S S	dim 31		lan 31	M M M	mer 31		lan 31	M M M

- M Uniforme 500mA ou Hybride 450 mA - Top-Up
- 8 paquets Top-Up - 100mA
- S 1 paquet Top-Up - 16mA
- Low-Alpha Top-Up
- B Beamlines
- Cp Contrôles FP périodiques, 3 mards de 7h à 23h
- Tv Tests FP de validation, faisceau Lignes suivant à 10h
- A Temps Accélérateurs
- Arrêt Machine

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.

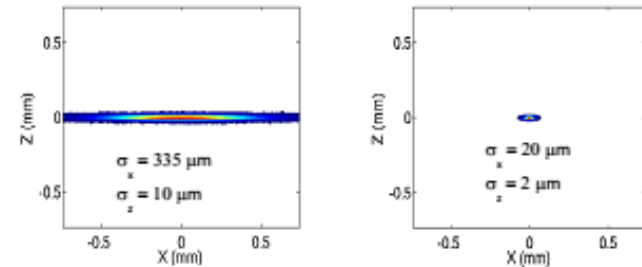


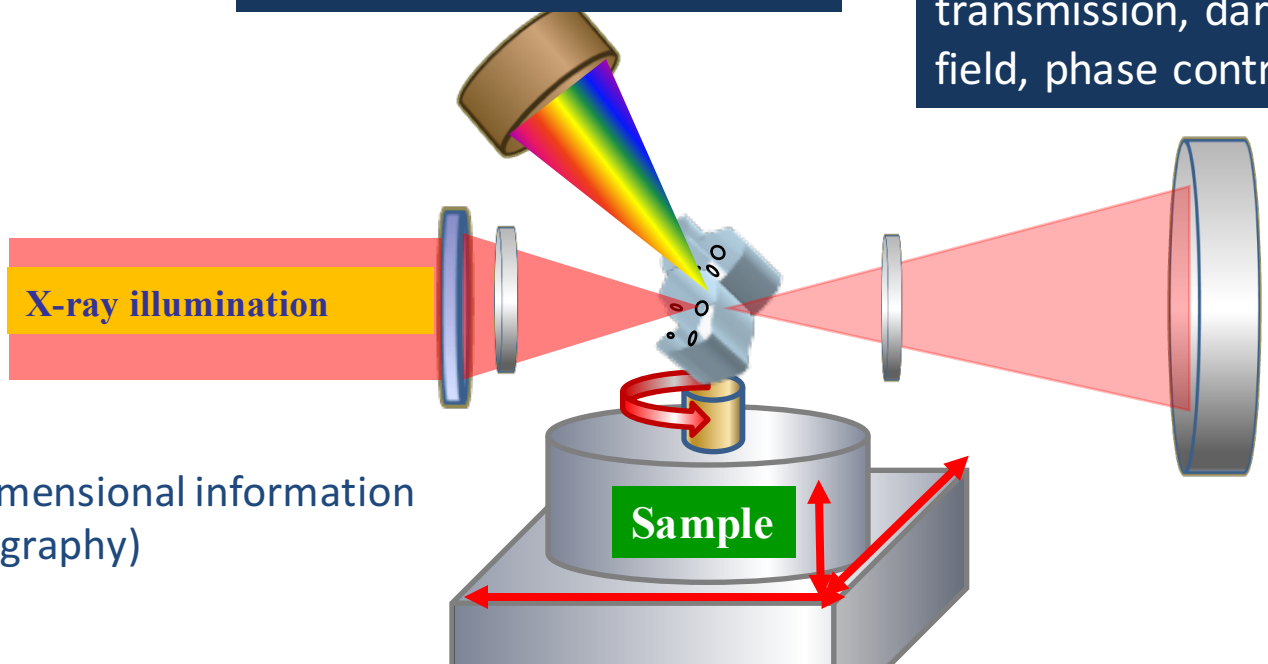
Figure 25: Comparison of the transverse beam profiles of SOLEIL and SOLEIL upgrade baseline lattice in a short straight section.

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

Detector 2:
transmission, dark field, phase contrast

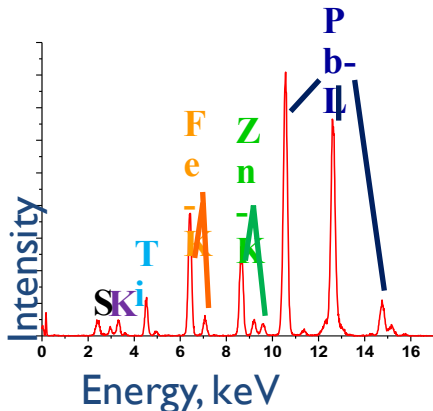
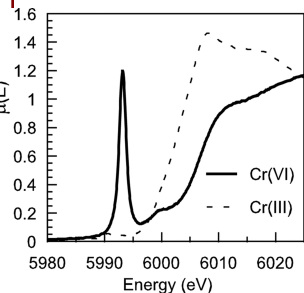


X-ray illumination

Sample

Resolution: ~30-400 nm

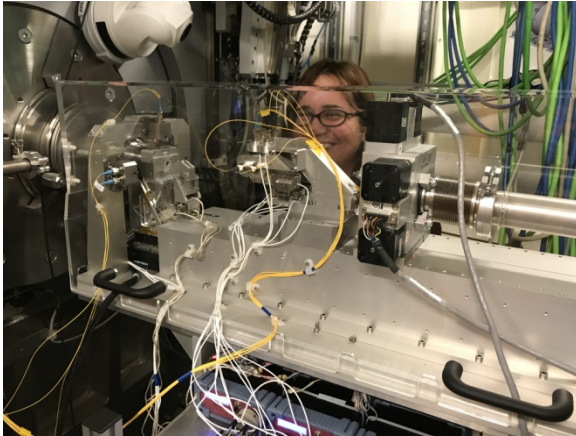
2/3 dimensional information (tomography)



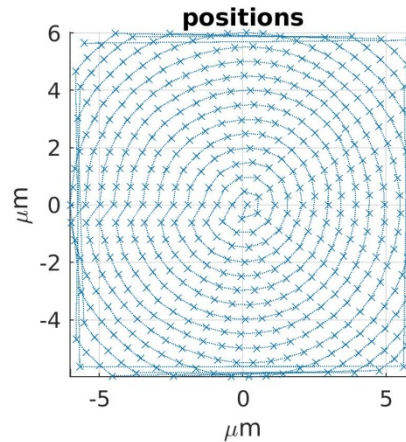
Cryo-cooling being developed.
Nanoscopium team (A. Somogyi, K. Medjoubi, G. Baranton)

2D-Ptychography at beamline Swing

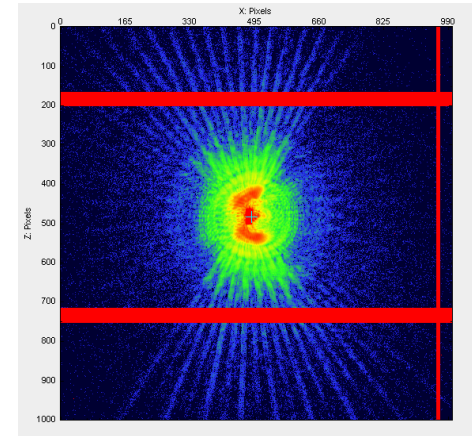
All ptychographic results generated via *cSaxs ptychography application (PSI)*
(Pierre Thibault, Manuel Guizar-Sicairos and coll., 2008-2018)



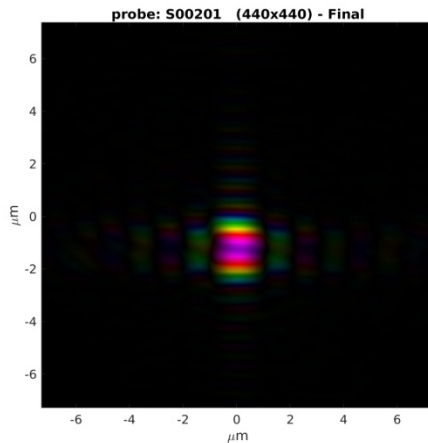
2D Nanoprobe set-up installed on Swing beamline (Kubsky, Somogyi, MAX IV)



Spiral scan generated via *Passerelle* application



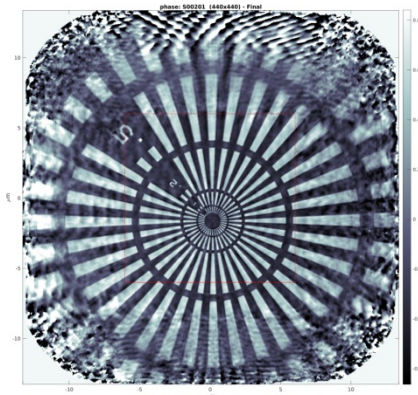
Diffraction pattern on Eiger4M detector at 6.2 m
All data collected using *Tango* devices.



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

Main involved people:

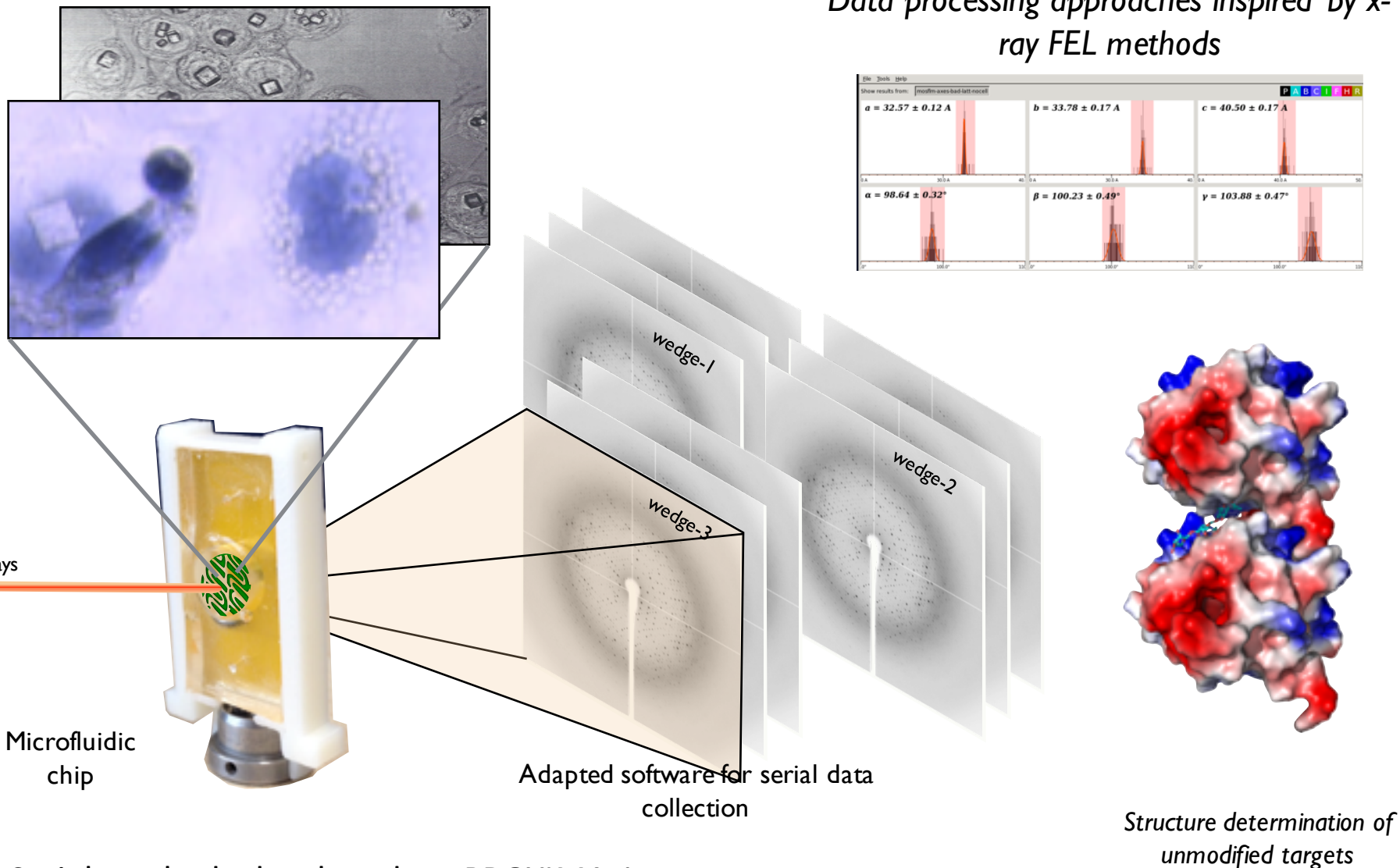
Ch. Engbloem (ECA)
F. Alves (CI)
A. Lestrade (ALI)
Y.M. Abiven (ECA)
F. Langlois (ICA)
P. Montaville (PX1)
F. Berenguer (CRISTAL)
J. Pérez (SWING)



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

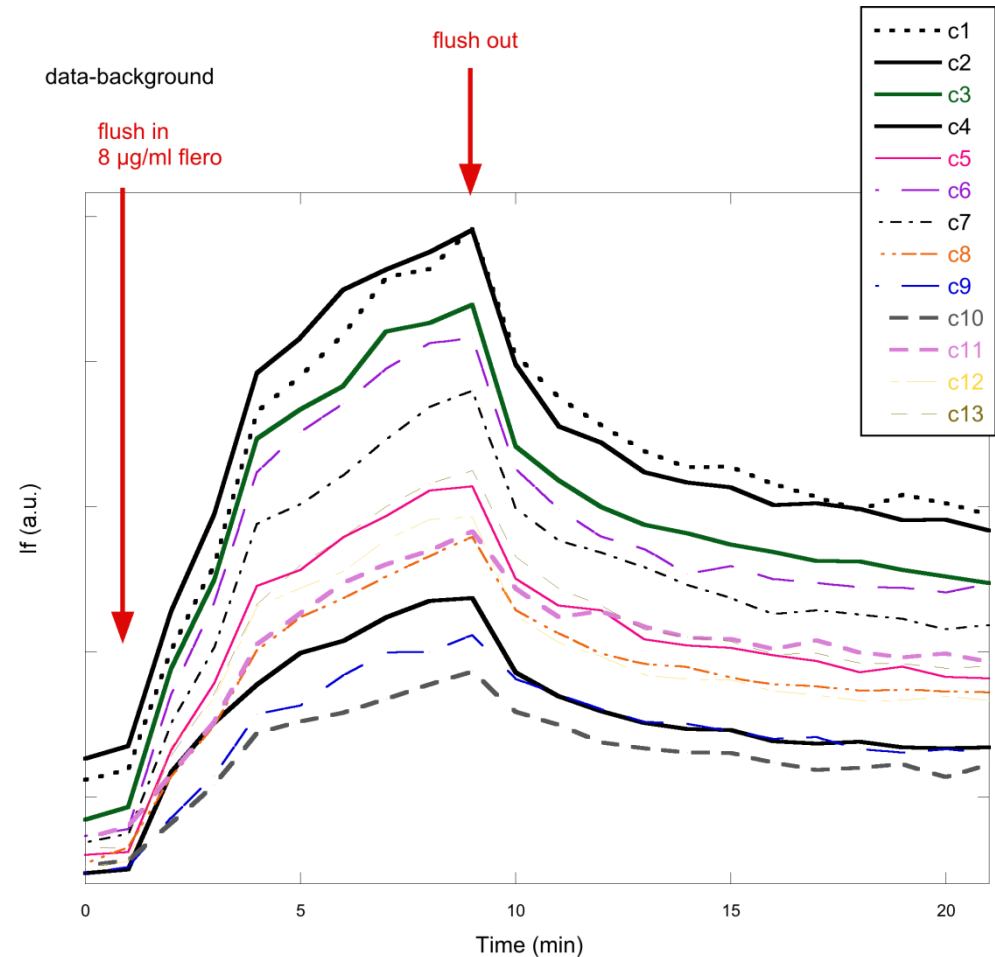
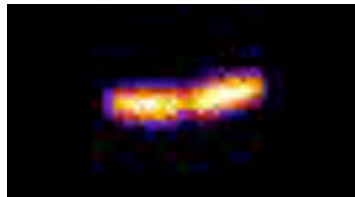
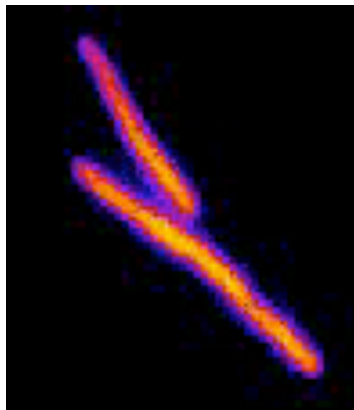
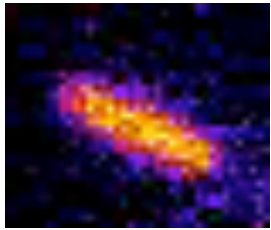
Application to *in vivo* grown crystals – satellite to SUM 2018

Data processing approaches inspired by x-ray FEL methods



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

Individualising the analysis of each bacterium

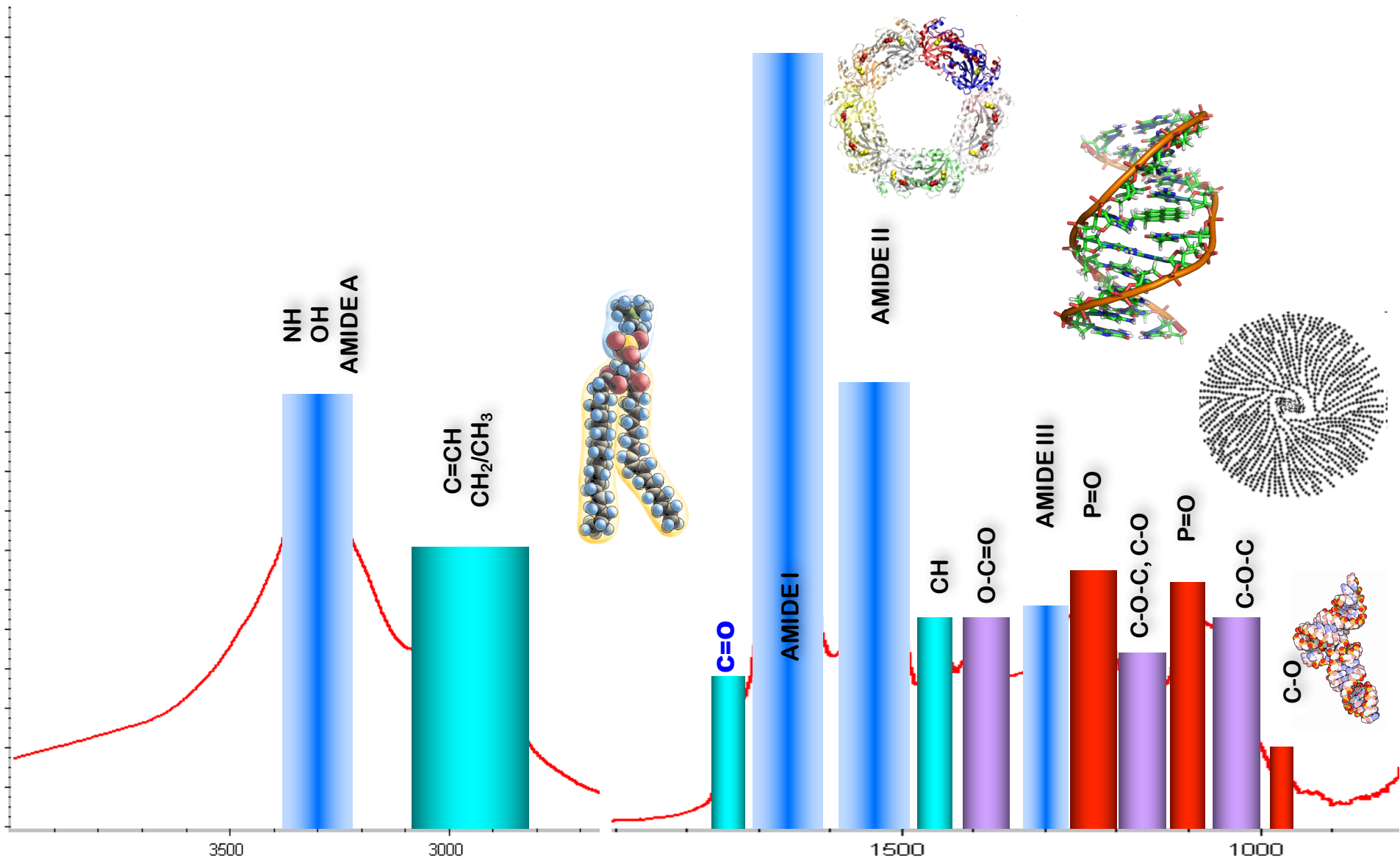


DISCO beamline :
M. Refregiers et
al.)

By analysing each individual curve, singular behaviour (mutations, reactions, ...)

raw data

SYNCHROTRON INFRARED MICROSCOPY: MOLECULAR FINGERPRINTS





Imaging micrometeorites

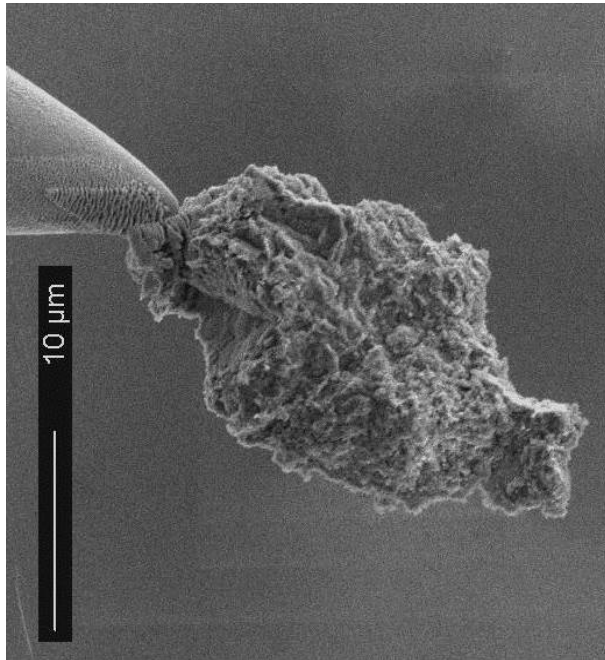
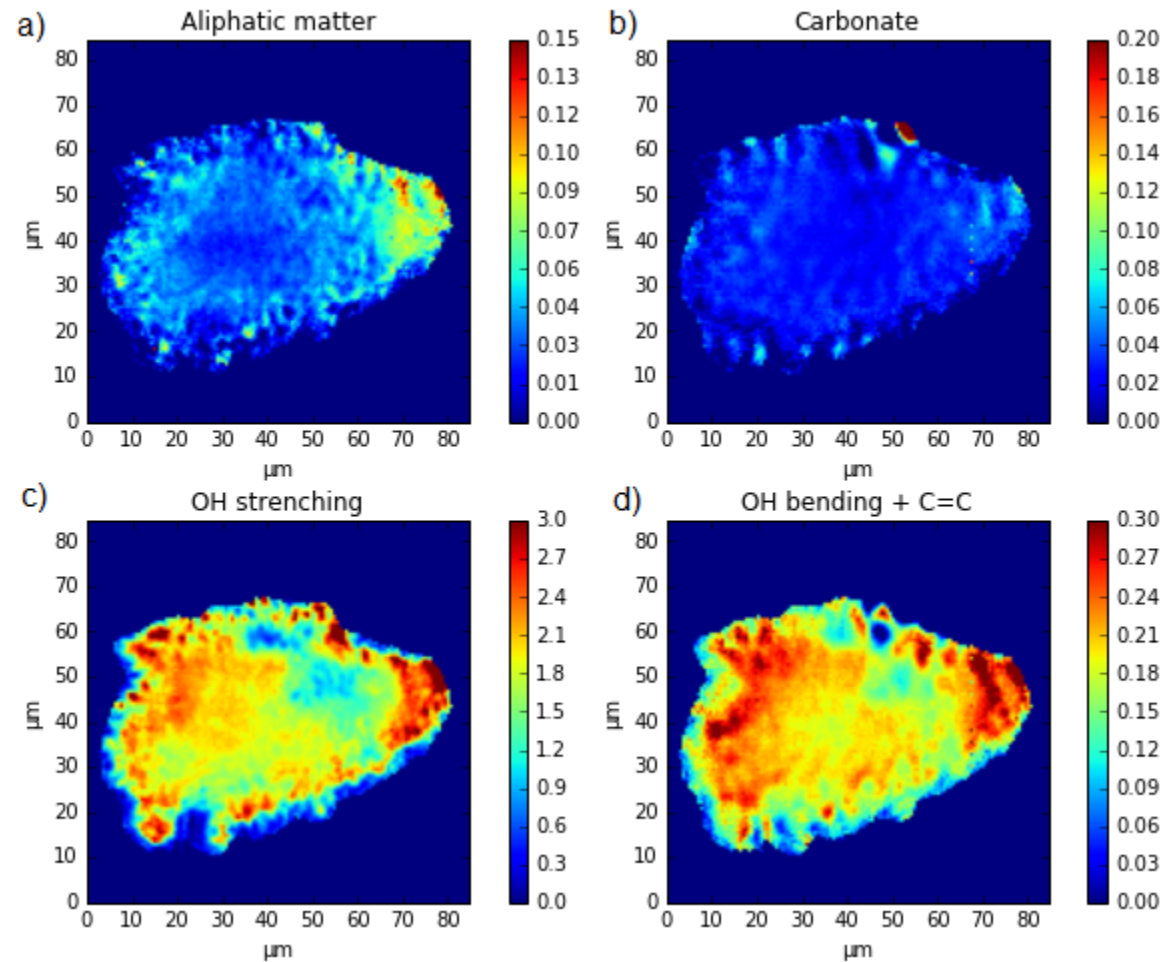
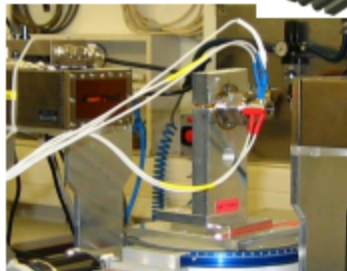


Image from FEM



SMIS Beamline – F.
Borondics et al.

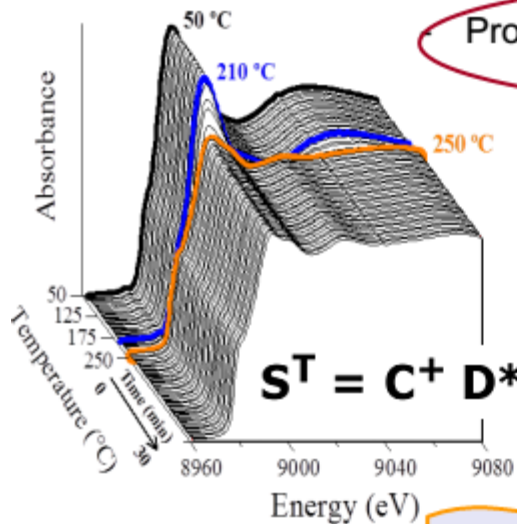


Electrochemical cells

- 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)

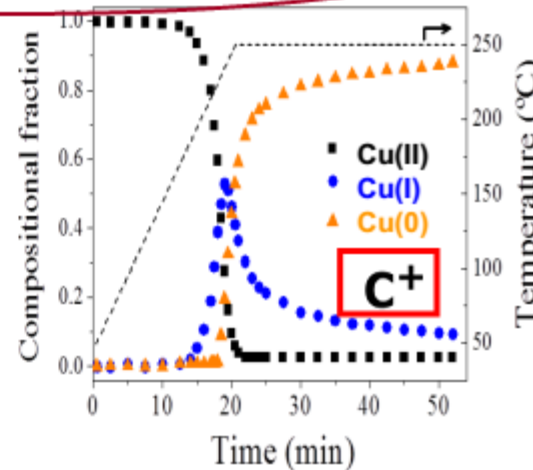
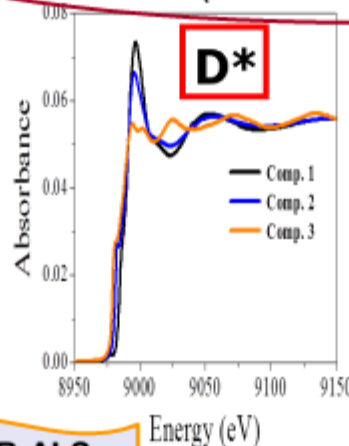
V.Briois,
SAC 2013

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



Catalysis Today (2013)

MCR-ALS

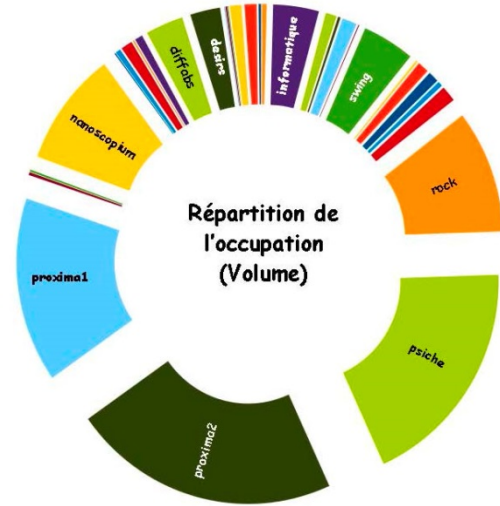
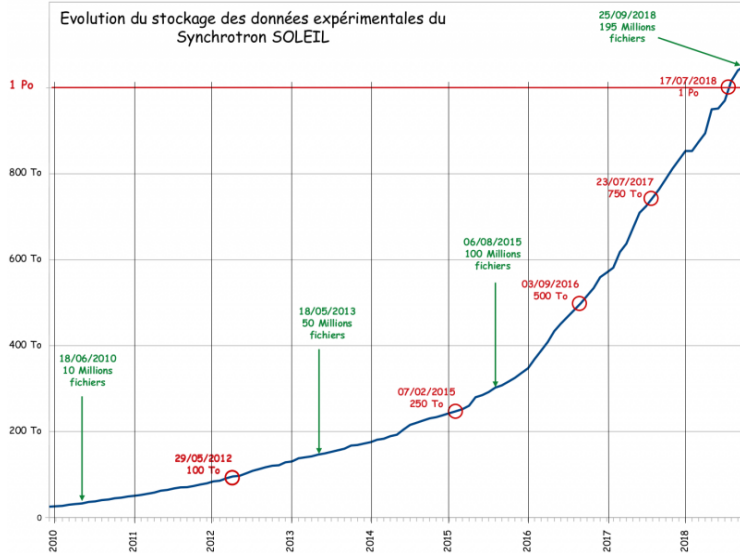


Time resolved spectroscopic measurements.
Extension towards 2D hyperspectral images with high spatial resolution.
ROCK beamline – V. Briois et al.

■ **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.

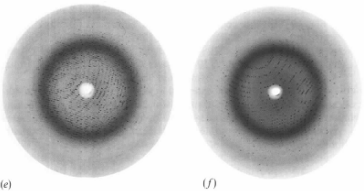
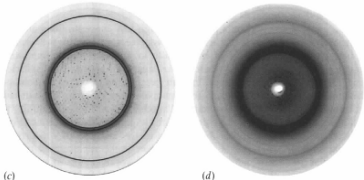
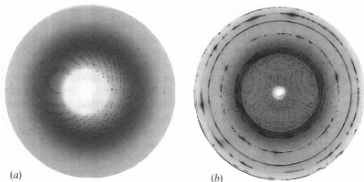


Conclusions

- **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.



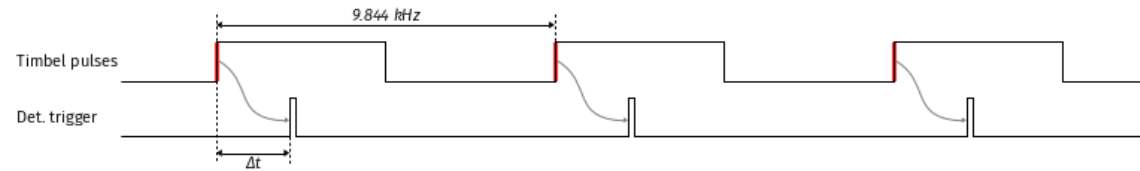
Sometimes it is visually easy to say what you should keep and what is rubbish, but not always.

It is impossible to annotate data by hand with collection speeds of « a few » to « several hundreds » of acquisitions per second.....

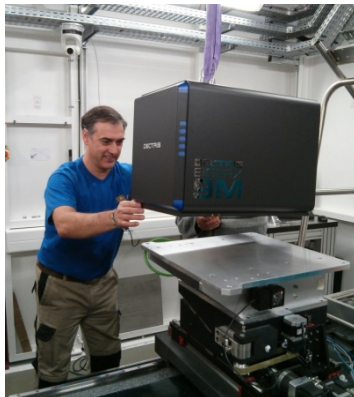
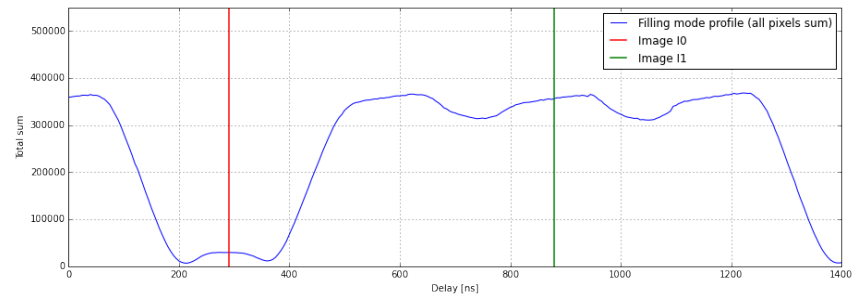
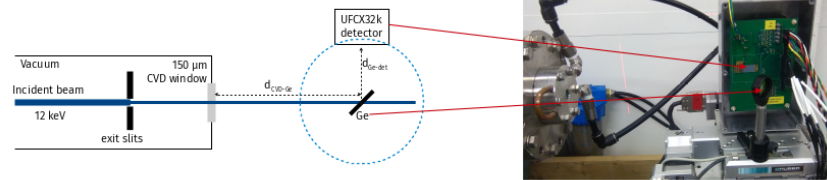
Need for « decision making software », tailored to different experimental methods. Machine learning with intelligence and mechanism adapted to different data types and experimental protocols?

Contact :

Andrew.thompson@synchrotron-soleil.fr



UFXC32k, collaboration with AGH in Cracow - 300 μ m Si, 75 μ m 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)