A decorative background consisting of numerous light blue circles of varying sizes, arranged in a pattern that suggests a molecular structure or a network.

Crystal Screening / Ranking and the DNA / EDNA projects

**Olof Svensson
EDNA Project Manager
ESRF Scientific Software Group**

Why screen and rank?

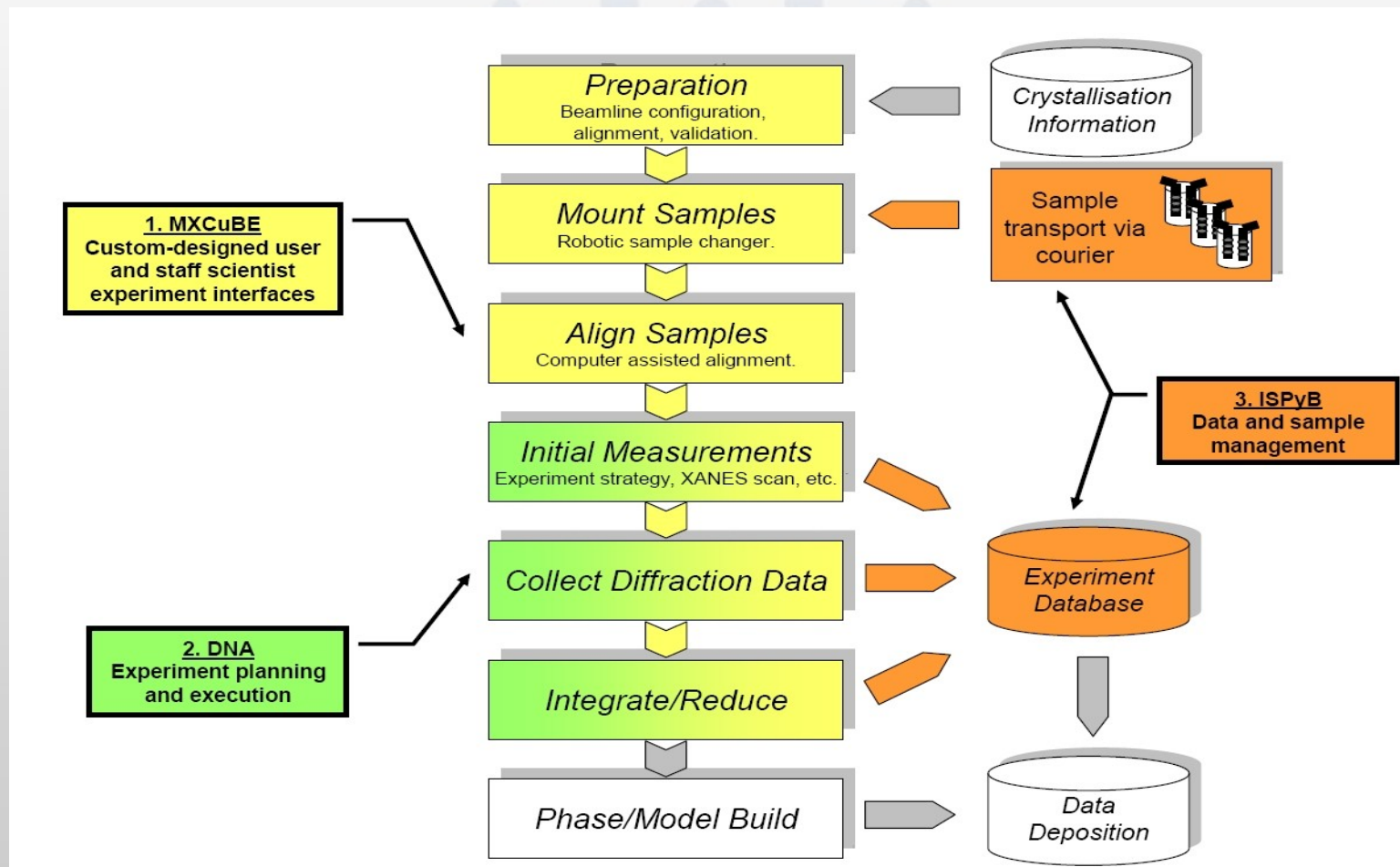
- On average at the ESRF : 24 data collections per PDB deposition
- Hence it's important to collect data on the best crystals
- Screening : collect reference images from a large set of samples of the same structure and calculate a score for each sample
- Ranking : choose the sample with the highest score
- Without automation : screening is time-consuming and tedious
- Thanks to automation implemented on ESRF MX beamlines, screening and ranking are fast and easy to use
- In this presentation I'll describe the current ESRF software for screening / ranking (DNA) and the future developments (EDNA)

Calculation of a screening score

- Characterisation : indexing, integration and strategy calculation given one or several reference images
- After characterisation : if the images can be indexed and integrated, many parameters can be extracted, for example :
 - Theoretical resolution (BEST “ranking” resolution)
 - Total data collection exposure time
 - Mosaicity
 - Number of spots indexed
 - Number of images
 - etc.
- The score after screening a sample can either be one of, or a combination of, these parameters
- Different type of experiments will have different score definitions

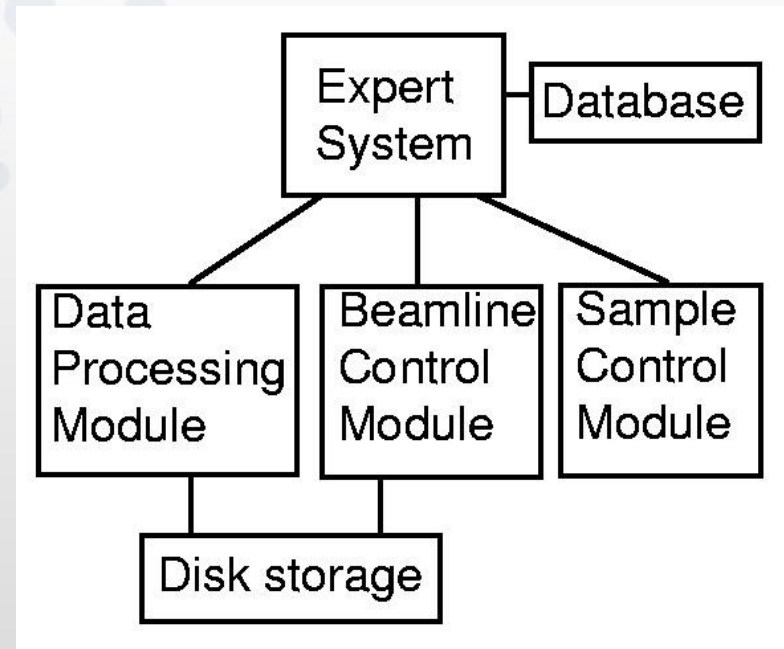
MX Automation

- Goal: Full automation from sample loading to reduced (integrated and scaled) data



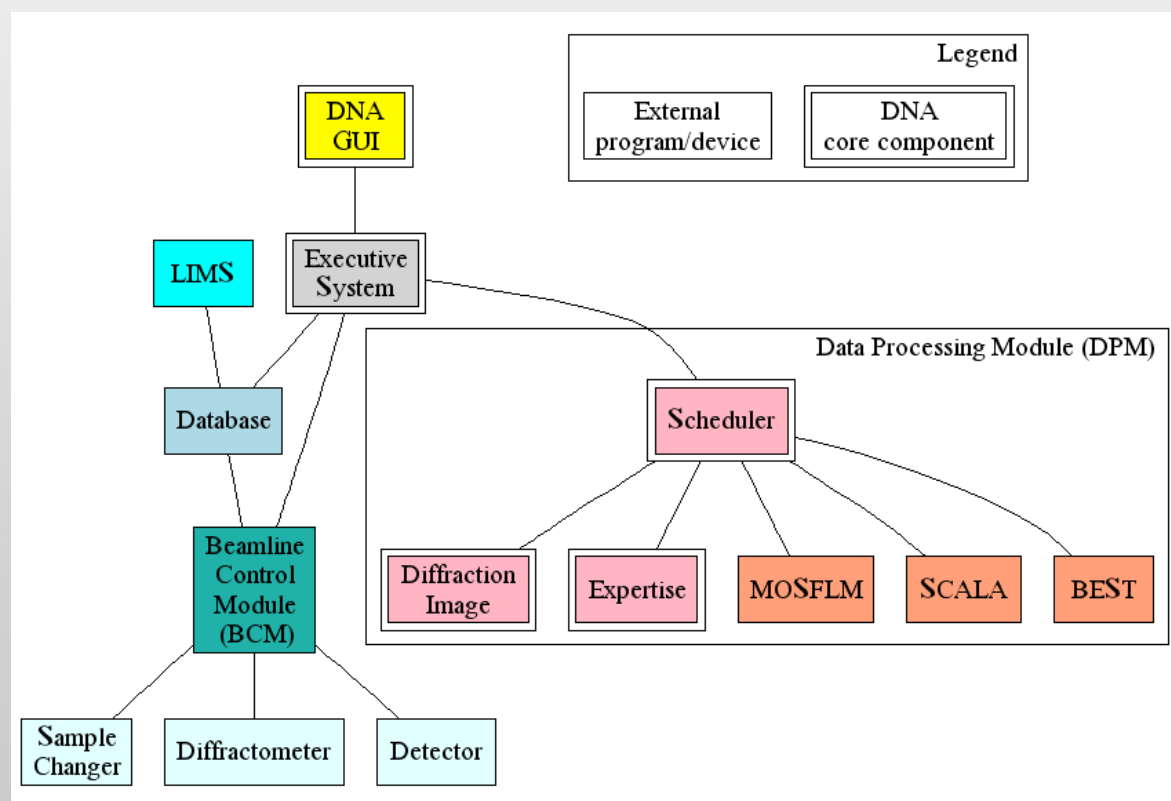
Creation of the DNA collaboration

- Kick-off meeting in 2001
- Initial collaborators :
 - ESRF
 - Daresbury SRS
 - MRC LMB Cambridge
- Initially no external funding
- Meaning of "DNA" :
 - automated collection of data

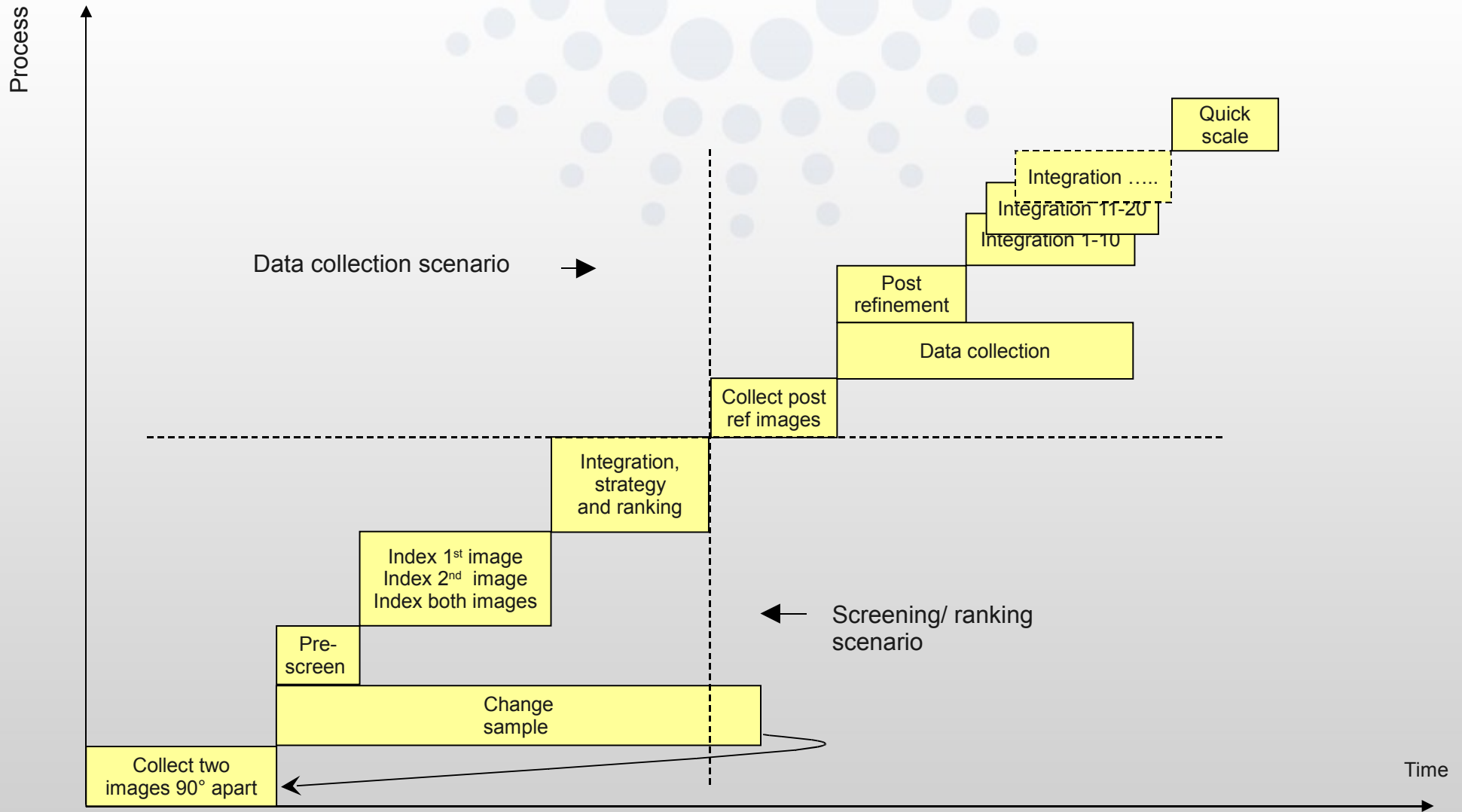


Evolution of the DNA collaboration

- Main development period 2001 – 2005
- More collaborators and more developers entered the project, mainly thanks to external fundings : BioXHIT and e-HTPX.
- Since 2005 an integral part of the ESRF "data collection pipeline"
- Installed and used at :
 - ESRF
 - Diamond
 - Recently used at NSLS (Brookhaven), now using EDNA



DNA screening / data collection scenarios



DNA GUI : Screening Input

DNA Expert System

File Screening

DNA - automated collection of data

Sample Screening Sample Ranking Collect Reference Images Auto Index Strategy Results

Collect parameters

Directory = /data/id14eh2/inhouse/mx415/280806/run1

PUBLIC copy of ref. images = ☒ Enabled

Exposure Time = 0.2

Oscillation = 1.0

Resolution = 2.0

Index and strategy parameters

Bravais Lattice = unknown

Minimum Desired Oscillation = 0.1

Multiplicity = 2.0

Anomalous = ☐ Enabled

Screening Strategy

Screen	Prefix	Cell	Symmetry	Resolution	Bar Code	Location
<input type="checkbox"/> tryp-C32		34.000 34.000 34.000 0.0 0.0 0.0		2.0	HA00AA1356	1 2
<input type="checkbox"/> TLN-TLN10		0.000 0.000 0.000 90.0 90.0 120.0	P6122	2.0	HA00AK1582	2 7
<input type="checkbox"/> TLN-TLN4		0.000 0.000 0.000 90.0 90.0 120.0	P6122	2.0	HA00AK1600	2 4
<input type="checkbox"/> tryp-D51		34.000 34.000 34.000 0.0 0.0 0.0		2.0	HA00AK1594	1 3
<input type="checkbox"/> tryp-D52		34.000 34.000 34.000 0.0 0.0 0.0		2.0	HA00AK1587	1 4
<input type="checkbox"/> TLN-TLN8		0.000 0.000 0.000 90.0 90.0 120.0	P6122	2.0	HA00AK1583	2 9
<input type="checkbox"/> TLN-TLN7		0.000 0.000 0.000 90.0 90.0 120.0	P6122	2.0	HA00AK1592	2 8
<input type="checkbox"/> TLN-TLN2		0.000 0.000 0.000 90.0 90.0 120.0	P6122	2.0	HA00AK1588	2 2
<input type="checkbox"/> TLN-TLN1		0.000 0.000 0.000 90.0 90.0 120.0	P6122	2.0	HA00AK1574	2 1
<input type="checkbox"/> TLN-TLN9		0.000 0.000 0.000 90.0 90.0 120.0	P6122	2.0	HA00AK1590	2 10
<input type="checkbox"/> tryp-C23		34.000 34.000 34.000 0.0 0.0 0.0		2.0	HA00AB4220	1 8
<input type="checkbox"/> TLN-TLN6		0.000 0.000 0.000 90.0 90.0 120.0	P6122	2.0	HA00AK1589	2 6
<input type="checkbox"/> tryp-C21		34.000 34.000 34.000 0.0 0.0 0.0		2.0	HA00AK1579	1 6
<input type="checkbox"/> tryp-C24		34.000 34.000 34.000 0.0 0.0 0.0		2.0	HA00AA3904	1 9
<input type="checkbox"/> tryp-C22		34.000 34.000 34.000 0.0 0.0 0.0		2.0	HA00AC8667	1 7
<input type="checkbox"/> tryp-B11		34.000 34.000 34.000 0.0 0.0 0.0		2.0	HA00AJ9885	1 10
<input type="checkbox"/> TLN-TLN3		0.000 0.000 0.000 90.0 90.0 120.0	P6122	2.0	HA00AK1585	2 3

Get Sample Changer Contents Select All Samples Deselect All Samples Screen

Feedback for PROPOSAL: mx415

Collection status: Ready

Processing status: Ready

Control

Help Submit Feedback Abort

Executive Output MOSFLM Output

```

060906 11:13:39 : Exposure time
060906 11:13:39 : Resolution (requested) : 2.00
060906 11:13:39 : Resolution (edge of detector): 2.00
060906 11:13:39 : Collecting 2 images, first image ref-TLN-TLN4_1_001.img, in directory /data/id14eh2/inhouse/mx415/280806/run1
060906 11:13:39 : Warning message from the BCM - skipping data processing for this sample.
060906 11:13:39 : 
060906 11:13:39 : Waiting for collection of reference images for sample TLN-TLN4 to finish.
060906 11:14:57 : 
060906 11:14:57 : Data collection finished for sample ref-TLN-TLN4_1
060906 11:14:57 : Resolution (edge of detector): 2.00
060906 11:14:57 : 
060906 11:14:57 : Warning message from the BCM - skipping data processing for this sample.
060906 11:14:57 : Waiting for new command
060906 11:22:01 : Getting proposal information for mx415, please wait...
060906 11:22:04 : Proposal mx415 information received: 113 sessions
060906 11:22:04 : Getting a list of loaded samples - please wait...
060906 11:22:04 : Sending a sample_references request to the BCM...
060906 11:26:45 : 20 sample references provided by the BCM.
060906 11:26:48 : Information about 19 crystals received
060906 11:26:48 : Waiting for new command

```


DNA GUI : Ranking Result

NX - svensson@tet.esrf.fr:1037 - bet
 DNA Expert System

File Screening
 DNA - automateD collection of data

Sample Screening Sample Ranking Collect Reference Images Auto Index Strategy Results

Sample ranking

Redo Rank View Rank Result Import Rank Project Export Rank Project

Ranking Results

Collect	Rank	Prefix	SpaceGroup	Cell	Information
<input type="checkbox"/>	1	Dps_P23_-Dps1	P222	54 58 67 90 90 90	Rank: Total exposure time to collect the data = 4.9 [s]
<input type="checkbox"/>	2	Dps_P23_-Dps4	P222	54 57 66 90 90 90	Rank: Total exposure time to collect the data = 5.6 [s]
<input type="checkbox"/>	3	Dps_P23_-Dps2	P222	54 57 66 90 90 90	Rank: Total exposure time to collect the data = 5.7 [s]
<input type="checkbox"/>	4	Dps_P23_-Dps3	P222	54 57 66 90 90 90	Rank: Total exposure time to collect the data = 5.7 [s]
<input type="checkbox"/>	5	Dps_P23_-Dps6	P222	54 57 66 90 90 90	Rank: Total exposure time to collect the data = 5.7 [s]
<input type="checkbox"/>	6	Dps_P23_-Dps7	P222	54 57 66 90 90 90	Rank: Total exposure time to collect the data = 5.7 [s]
<input type="checkbox"/>	7	Dps_P23_-Dps5	P222	54 57 66 90 90 90	Rank: Total exposure time to collect the data = 5.7 [s]
<input type="checkbox"/>	8	Dps_P23_-Dps9	P222	54 57 66 90 90 90	Rank: Total exposure time to collect the data = 5.7 [s]
<input type="checkbox"/>	9	Dps_P23_-Dps8	P222	54 57 66 90 90 90	Rank: Total exposure time to collect the data = 5.7 [s]

Ranking criterion: Total exposure time

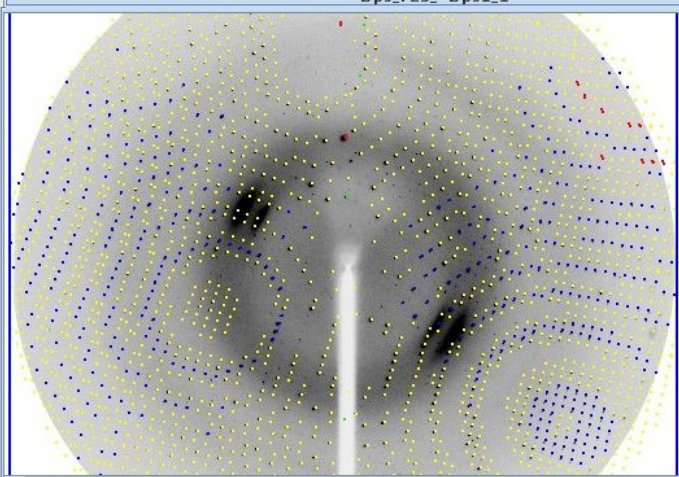
Feedback for PROPOSAL : mx415
 Collection status: Ready
 Processing status: Ready

Executive Output MOSFLM Output

```

20070322 14:13:11 : 
20070322 14:13:11 : =====
20070322 14:13:11 : The program 88ST
20070322 14:13:11 : This is the read
20070322 14:13:11 : 
20070322 14:13:11 : 88ST has calcula
20070322 14:13:11 : If you want to c
20070322 14:13:11 : Ranking engine u
20070322 14:13:11 : Ranking project
20070322 14:13:12 : 
20070322 14:13:12 : Screening done
20070322 14:13:12 : 
20070322 14:13:12 : 
20070322 14:13:12 : Waiting for new
  
```

Dps_P23_-Dps7_1 Dps_P23_-Dps5_1 Dps_P23_-Dps9_1 Dps_P23_-Dps8_1
 Dps_P23_-Dps4_2 Dps_P23_-Dps2_2 Dps_P23_-Dps3_2 Dps_P23_-Dps6_1
 Dps_P23_-Dps1_1



DNA Collaborators (in 2007)

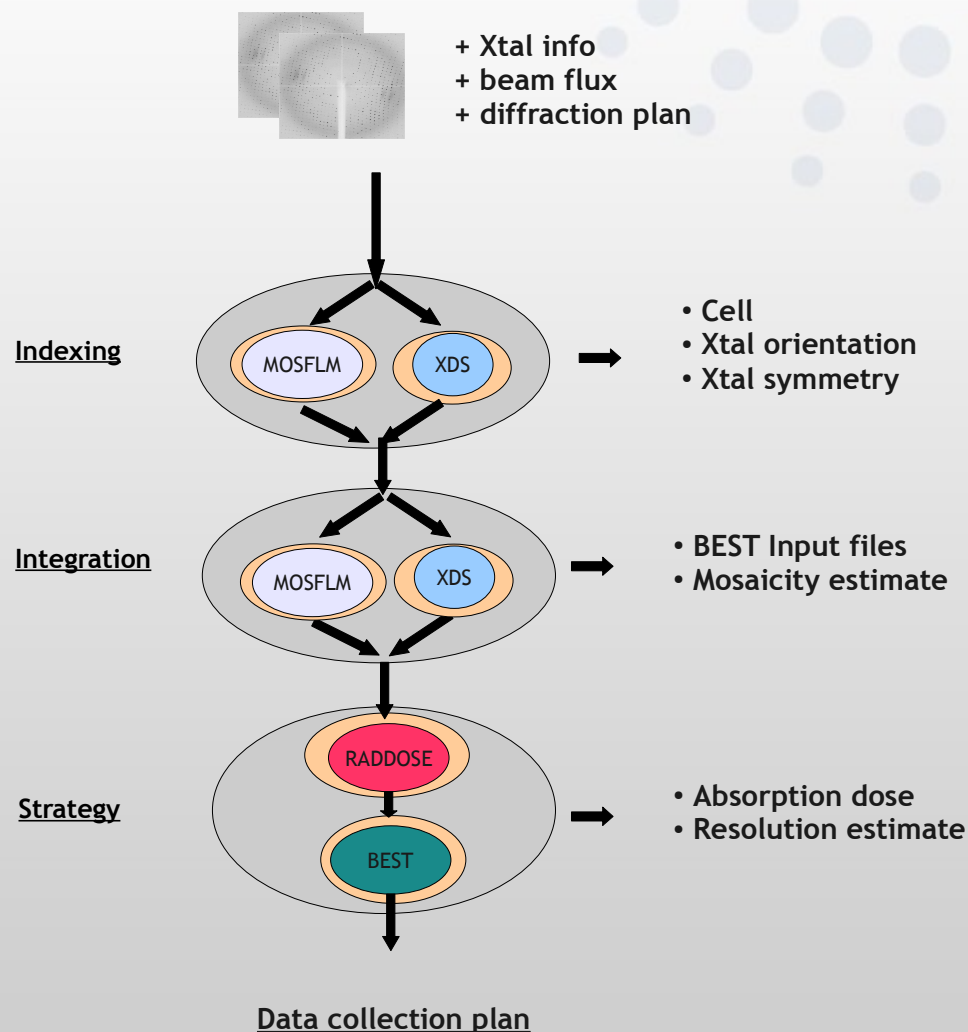
Home institute	Name and link to email	DNA related work funded by	DNA tasks
Brookhaven Nat'l Lab.	Alex Soares	BNL	DNA co-ordinator at BNL
	Bob Sweet		
	John Skinner		
Diamond Light Source	Alun Ashton	DLS	Acting DNA project co-ordinator
	Colin Nave		
	Elizabeth Duke		DNA co-ordinator at DLS
	Karl Levik		DNA developer
	Katherine McAuley		Testing and implementation of DNA at Diamond Light Source
EMBL Grenoble	Raimond Ravelli	EMBL Grenoble	DNA co-ordinator at the EMBL Grenoble
	Sandor Brockhauser	BioXHIT	Working on introducing kappa geometry strategy into DNA
EMBL Hamburg	Alexander Popov	BioXHIT	DNA co-ordinator at EMBL Hamburg
	Gleb Bourenkov	DESY	
	Venkataraman Parthasarathy	SPINE	Integration of the BEST strategy software
ESRF	Sean McSweeney	ESRF	DNA co-ordinator at the ESRF
	Darren Spruce		Responsible for the ESRF BCM (ProDC) and the DNA - LIMS connection
	Olof Svensson		Responsible for the DNA Executive System and DNA 2.0 Project Manager
	Marie Françoise Incardona	BioXHIT	Working on DNA 2.0
	Romeu Pieritz		Responsible for developing the ranking module and working on DNA 2.0
Global Phasing	Gérard Bricogne	Global Phasing	DNA co-ordinator at Global Phasing
	Peter Keller	BioXHIT	Working on DNA 2.0
MRC LMB Cambridge	Andrew Leslie	MRC	DNA co-ordinator at Cambridge
	Harry Powell	CCP4	Responsible for the DNA DPM based on MOSFLM
MRC France - BM14 at the ESRF	Ludovic L auner	e-htpx	Working together with Darren on the DNA - beamline database connection
STFC Daresbury	Graeme Winter	e-htpx	Responsible for the DNA Scheduler
SLS - PSI	Takashi Tomizaki	SLS - PSI	DNA co-ordinator at SLS
Synchrotron Soleil	Andrew Thompson	Synchrotron Soleil	DNA co-ordinator at Soleil
	Lucile Roussier		Working on DNA - database connection for Soleil
	Eric Girard		Working on offline tests of DNA
	Pierre Legrand		Working on integrating XDS as a DPM in the DNA system

The EDNA project

What is EDNA?

- A collaborative project between several international laboratories and synchrotron facilities. Developed on the foundation of the DNA project
- A framework for Online Data Analysis of X-ray experiments
- A first application has been developed for the automation of the collection of X-ray diffraction data from macromolecular crystals. The first goal was to produce an appropriate ready-to-use data collection strategy from a few initial diffraction images taking into account radiation damage.

The EDNA prototype



The main scientific goal:

From a few diffraction images, calculate a data collection plan with minimum radiation damage.

Launch sequentially the crystal characterization and the data collection strategy steps by executing one or several parallel external programs according to the user configuration :

Indexing (MOSFLM and/or XDS and/or other)
Integration (MOSFLM and/or XDS)
Radiation damage estimation (RADDose)
Data collection strategy (BEST)

A prototype has been released to our partners.
GUIs are currently under development:
CBASS at NSLS, CCP4i, MXCube (soon)

EDNA Collaborators

Alexander Popov^(d)
 Alun Ashton^(e)
Andrew Leslie^(h)
Andrew McCarthy^(b)
Andrew Thompson^(k)
 Clemens Schulze^(j)
 Clemens Vornrhein^(f)
 Darren Spruce^(d)
 Elspeth Gordon^(d)
 Ezequiel Panepucci^(j)
Gérard Bricogne^(f)
 Gerrit Langer^(b)
Gleb Bourenkov^(b)
 Gordon Leonard^(d)
 Harry Powell^(h)
 Johan Turkenburg^(m)
Johan Unge^(g)
John Skinner⁽ⁱ⁾

Karl Levik^(e)
 Katherine McAuley^(e)
 Lucile Roussier^(k)
Marie-Françoise Incardona^(d)
Mark Basham^(e)
 Meitian Wang^(j)
 Michael Hellmig^(a)
 Olga Roudenko^(k)
Peter Keller^(f)
 Peter Turner^(l)
Pierre Legrand^(k)
 Robert Sweet⁽ⁱ⁾
Romeu Pieritz^(d)
Sandor Brockhauser^(b)
Sean McSweeney^(d)
 Takashi Tomizaki^(j)
Thomas Schneider^(b)
 Uwe Mueller^(a)

(a) BESSY, Berlin, Germany
 (b) EMBL, Grenoble, France
 (c) EMBL, Hamburg, Germany
 (d) ESRF, Grenoble, France
 (e) Diamond Light Source, UK
 (f) Global Phasing, Cambridge, UK
 (g) MAX LAB, Lund, Sweden
 (h) MRC LMB, Cambridge, UK
 (i) NSLS, Brookhaven, U.S.
 (j) SLS, Villigen, Switzerland
 (k) Synchrotron Soleil, France
 (l) University of Sydney, Australia
 (m) University of York, UK

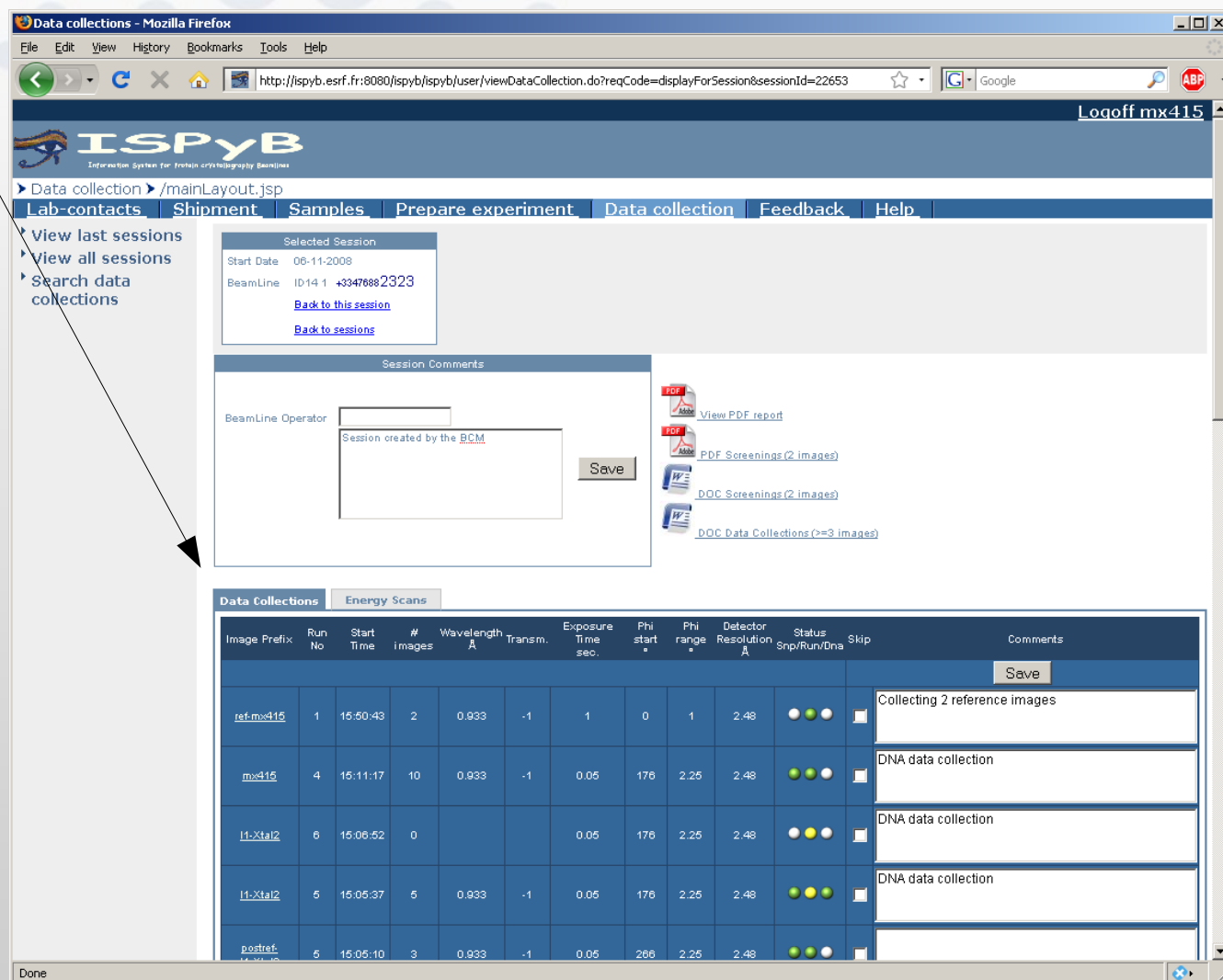
EDNA developers
Executive committee
 (*) Project manager

Screening and Ranking with EDNA

- Screening of an individual sample corresponds to a single DNA / EDNA characterisation
- Different philosophy of EDNA compared with DNA :
 - EDNA has no control of data collection, hence no GUI for screening
 - Screening will be handled by mxCuBE and other beamline control GUIs
- No implementation of ranking yet with EDNA
- Current idea : Ranking via the ISPyB interface
 - Advantages :
 - Ranking can be made independently of data collection
 - Same interface for DNA / EDNA
 - Disadvantage :
 - No connection (yet) of ranking results to the data collection software (mxCuBE)

Ranking via ISPyB

- Data collection page in the ISPyB GUI :
- If implemented :
 - A score for each screened crystal will appear in this page
 - Several different screenings should be selectable and form the input for ranking



Data Collections - Mozilla Firefox

http://ispyb.esrf.fr:8080/ispyb/ispyb/user/viewDataCollection.do?reqCode=displayForSession&sessionId=22653

Logoff mx415

ISPyB
Information System for Protein Crystallography Beamlines

Data collection > /mainLayout.jsp

Lab-contacts | Shipment | Samples | Prepare experiment | **Data collection** | Feedback | Help

View last sessions
View all sessions
Search data collections

Selected Session

Start Date: 06-11-2009
BeamLine: ID14 1 +33476882323
[Back to this session](#)
[Back to sessions](#)

Session Comments

BeamLine Operator:
Session created by the BCM

[View PDF report](#)
[PDF Screenings \(2 images\)](#)
[DOC Screenings \(2 images\)](#)
[DOC Data Collections \(>=3 images\)](#)

Data Collections | Energy Scans

Image Prefix	Run No	Start Time	# images	Wavelength Å	Transm.	Exposure Time sec.	Phi start °	Phi range °	Detector Resolution Å	Status Snp/Run/Dna	Skip	Comments
ref.mx415	1	15:50:43	2	0.933	-1	1	0	1	2.48	●●●●	<input type="checkbox"/>	Collecting 2 reference images
mx415	4	15:11:17	10	0.933	-1	0.05	176	2.25	2.48	●●●●	<input type="checkbox"/>	DNA data collection
11-Xtal2	6	15:06:52	0			0.05	176	2.25	2.48	●●●●	<input type="checkbox"/>	DNA data collection
11-Xtal2	5	15:05:37	5	0.933	-1	0.05	176	2.25	2.48	●●●●	<input type="checkbox"/>	DNA data collection
portref	5	15:05:10	3	0.933	-1	0.05	266	2.25	2.48	●●●●	<input type="checkbox"/>	

Done

Acknowledgements

- The DNA team
- The EDNA team
- The ISPyB team
- a big thank to Marie-Françoise who let me re-use many of her slides
- ...and thank you for your attention!