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

Crystal Screening / Ranking and the DNA / EDNA projects

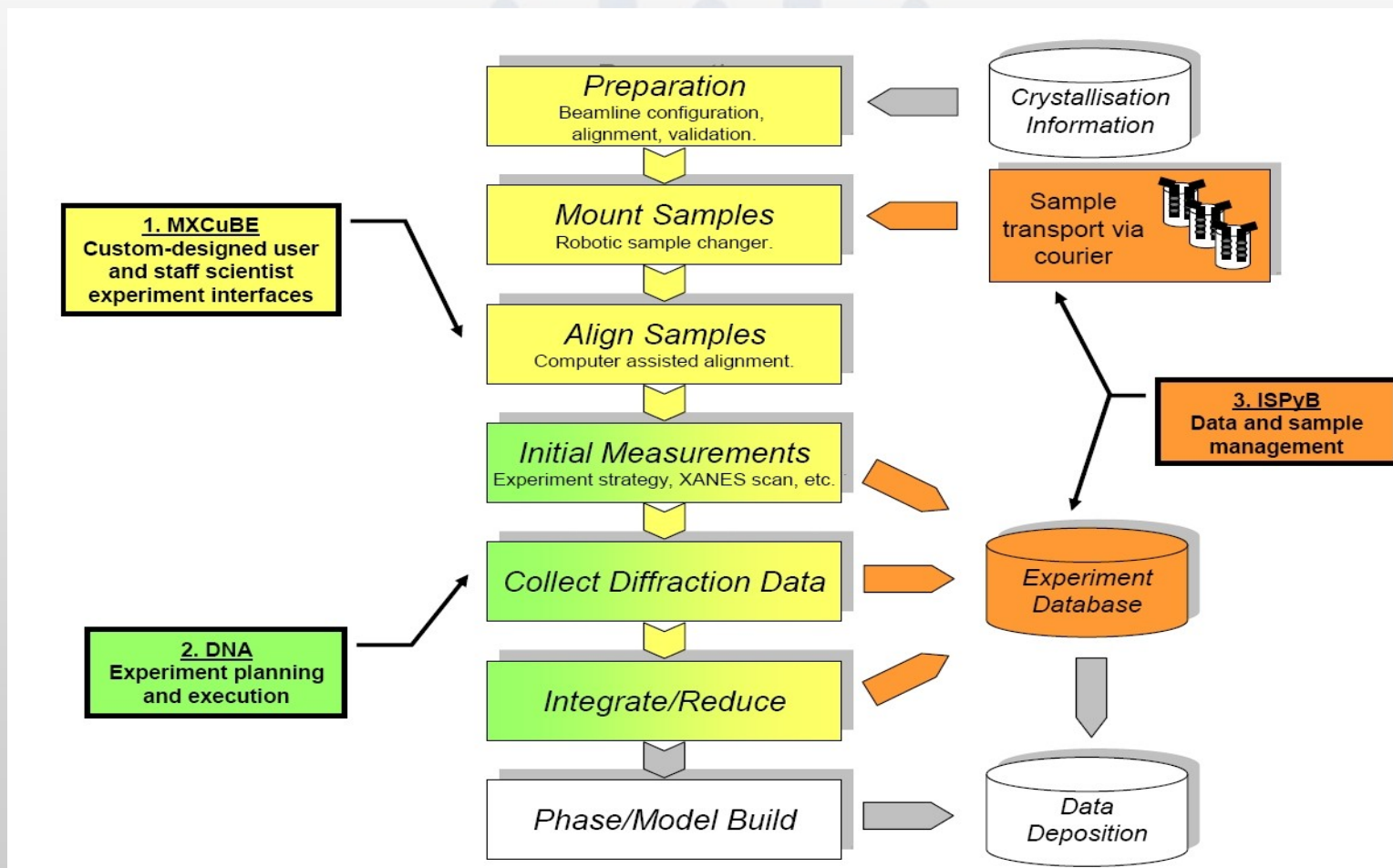
**Olof Svensson
Data Analysis Unit
Instrument Support and Development Division
ESRF**

Why screen and rank?

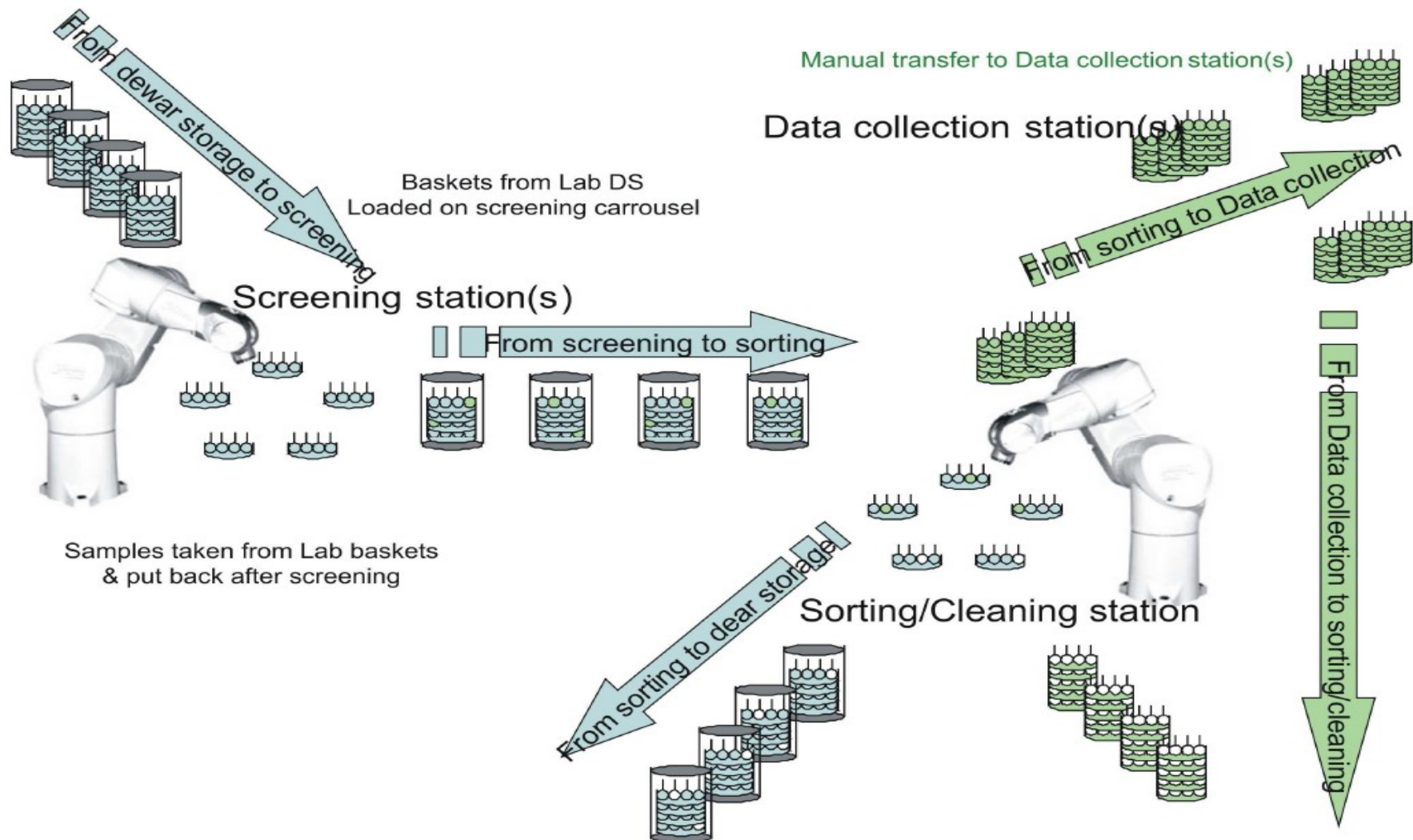
- On average at the ESRF : ca 20 data collections per PDB deposition
- Hence it's important to collect data on the best crystals
- Screening : collect reference images from a large set of samples of the same structure and calculate a score for each sample
- Ranking : choose the sample with the highest 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, EDNA and ISPyB)

MX Automation

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



Challenge for the ESRF Upgrade : Massively Automated Sample Selection Integrated Facility

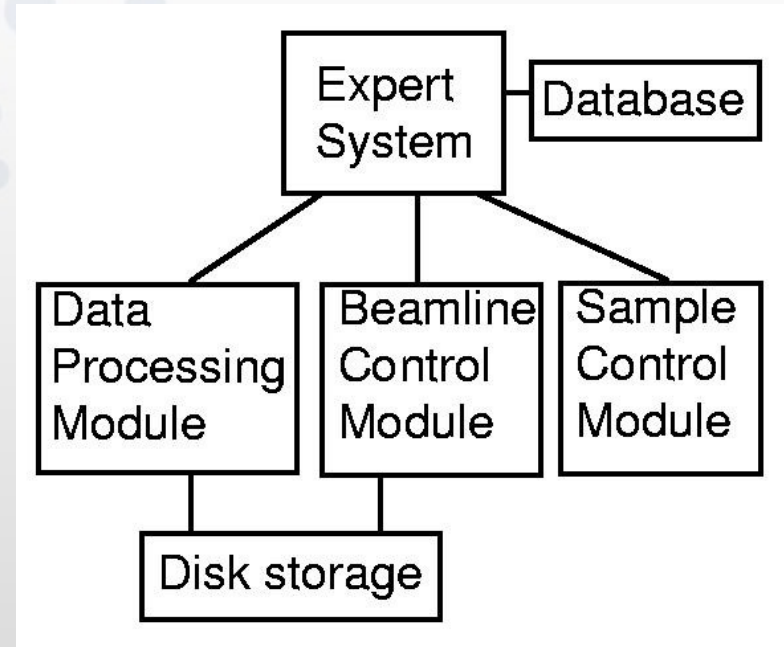


Calculation of a screening score

- 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

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
- 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, DLS



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

MX - 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

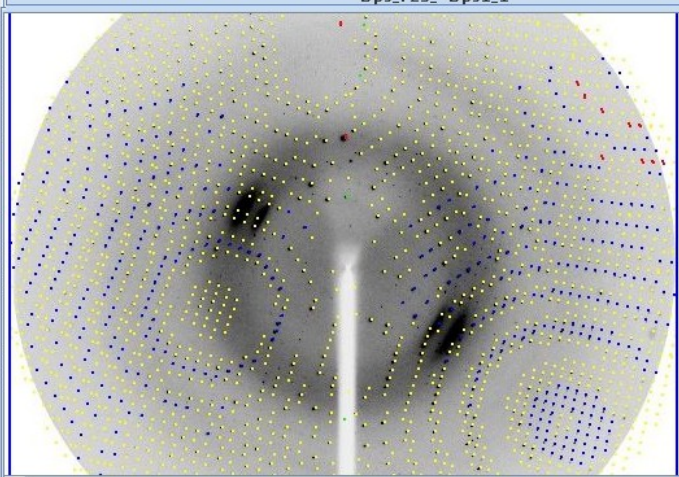
```

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

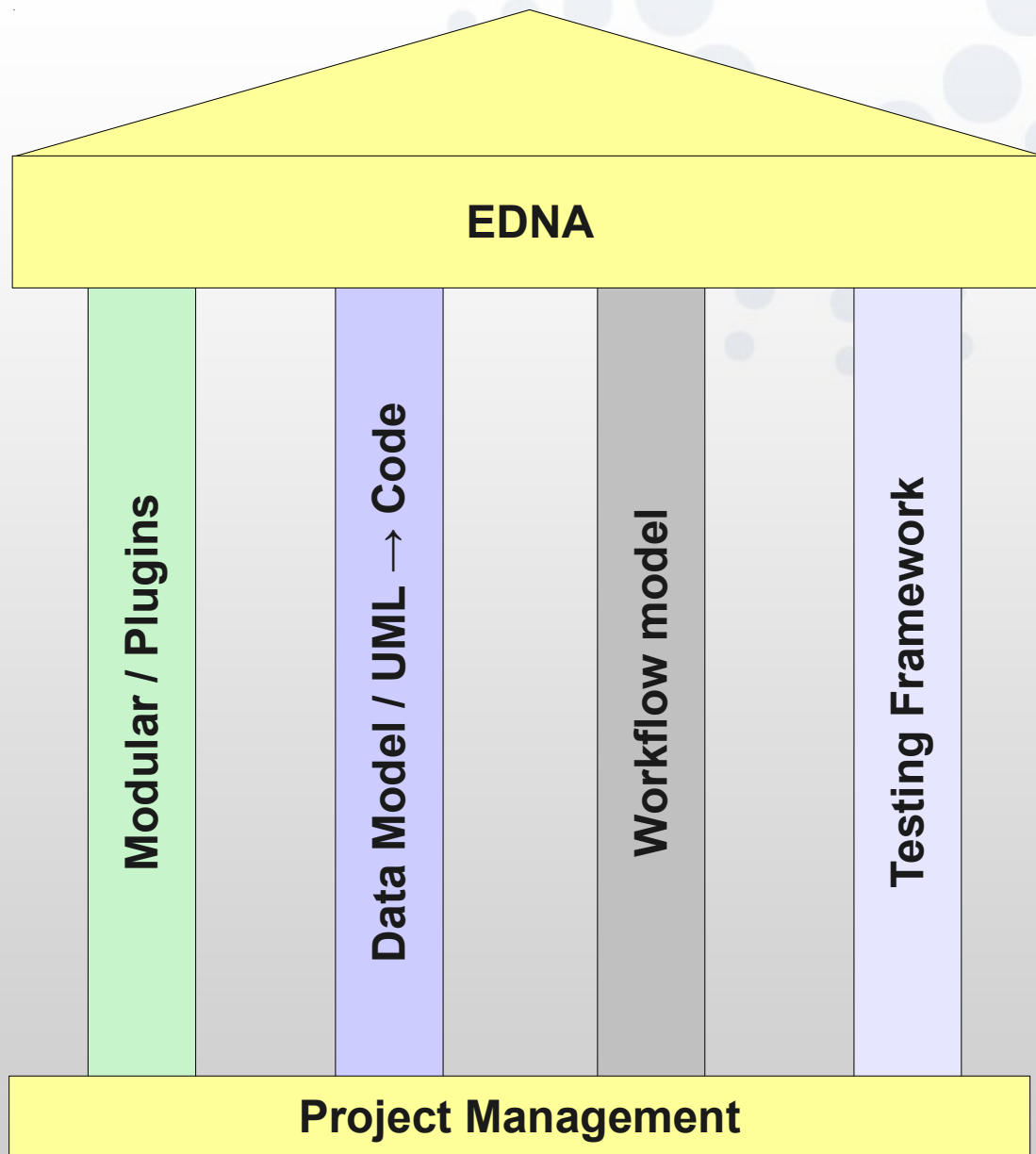
Dps_P23_-Dps1_1



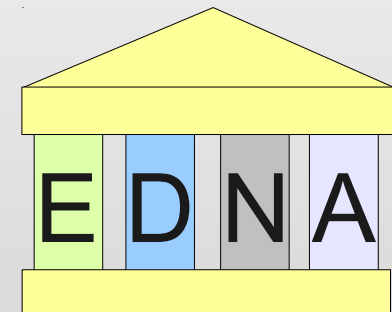
The EDNA project

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
 - EDNA applications for non-MX ODA are being developed at the ESRF
- The EDNA MXv1 characterisation application is currently being deployed on ESRF beamlines



Logo (suggestion):



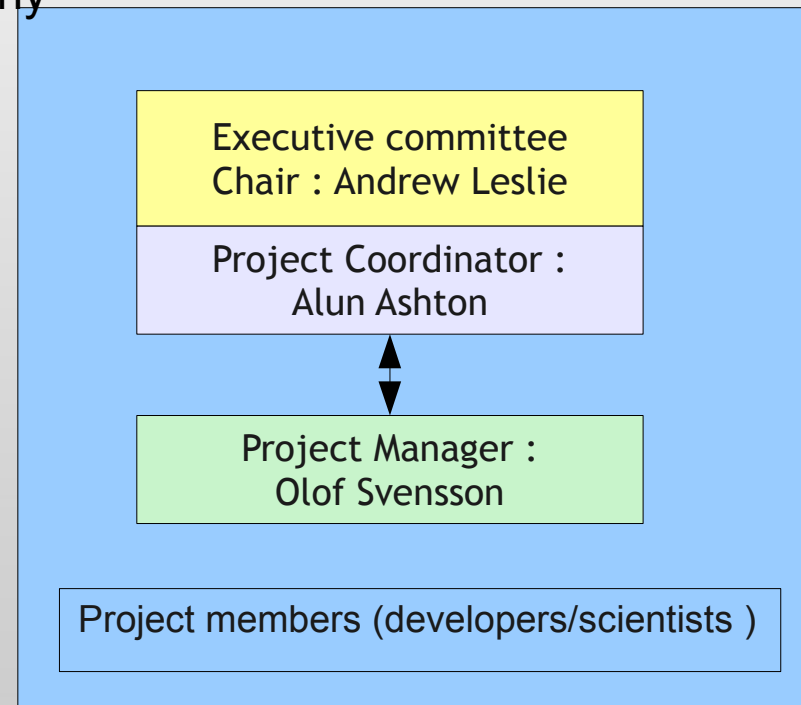
EDNA Project Management (1)

• Executive Committee:

- Alun Ashton, DLS, UK
- Gérard Bricogne, Global Phasing, UK
- Andrew Leslie, MRC LMB, Cambridge, UK
- Andrew McCarthy, EMBL-Grenoble, France
- Sean McSweeney, ESRF, Grenoble, France
- Thomas Schneider, EMBL-Hamburg, Germany
- Andrew Thompson, Synchrotron Soleil, France

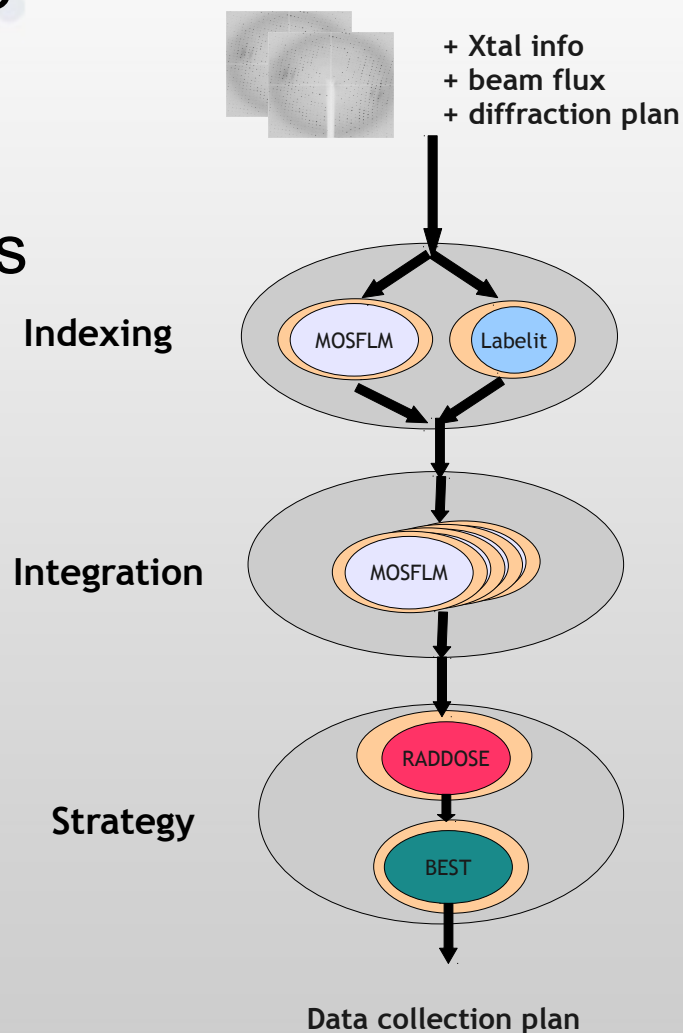
• Other members from:

- BESSY, Berlin, Germany
- MAX LAB, Lund, Sweden
- NSLS, Brookhaven, U.S.
- SLS, Villigen, Switzerland
- University of Sydney, Australia
- University of York, UK



EDNA MXv1 Characterisation

- MX sample characterisation taking into account radiation damage
- Indexing using MOSFLM or Labelit
- Parallel integration of reference images
- If flux + beamsizes + chemical composition:
 - RADDOSE for estimating radiation damage
- BEST strategy calculation
 - taking into account radiation damage
 - multi-subwedge data collection strategies



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
- Ranking via the ISPyB interface
 - Ranking can be made independently of data collection
 - Same interface for DNA / EDNA
 - No connection (yet) of ranking results to the data collection software (mxCuBE)

Ranking via ISPyB

- Data collection / Session: Select Samples & Rank

Data Collections		Energy Scans											
Image Prefix	Run No	Start Time	# images	Wavelength	Transm.	Ex. Time	Phi start	Phi range	Detector Resolution	Status	Sample Ranking	Skip	
mfe1fl9b	3	20:39:28	9	0,933	100	0,79	120	0,1	2,8		<input type="checkbox"/>	<input type="checkbox"/>	
postref-mfe1fl9b	3	20:38:44	3	0,933	100	0,79	210	0,1	2,8		<input type="checkbox"/>	<input type="checkbox"/>	
ref-mfe1fl9b	2	20:35:55	2	0,933	100	2	0	0,5	2,5		<input checked="" type="checkbox"/>	<input type="checkbox"/>	
ref-mfe1fl9b	1	20:34:00	2	0,933	100	2	0	0,5	3		<input checked="" type="checkbox"/>	<input type="checkbox"/>	

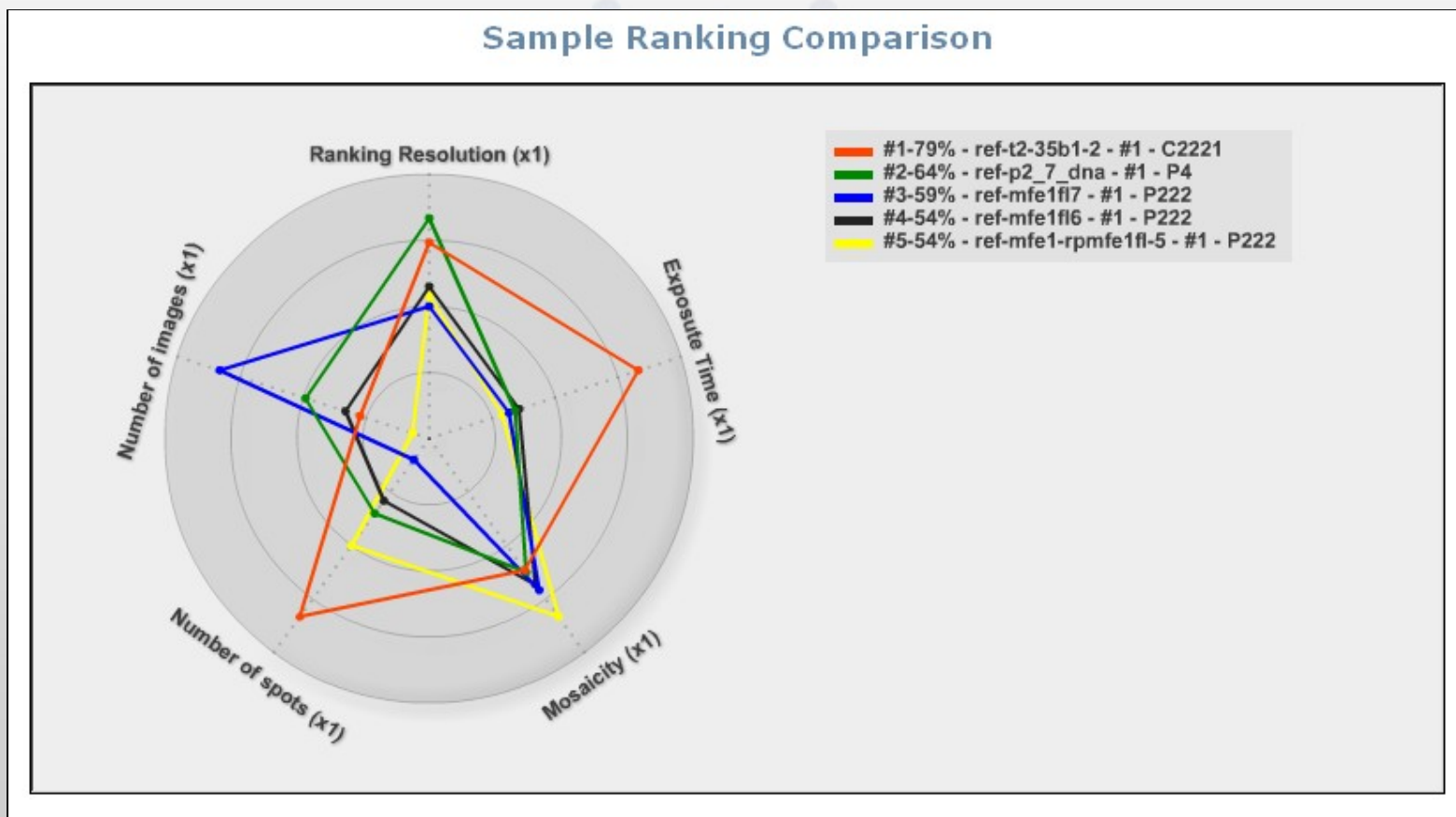
Sample Ranking

- Change weights & Rank again if necessary

Select Sample	Image Prefix	Run No	Start time	Space Group	UC a	UC b	UC c	Ranking resol. Å	Exposure time s	Mosaicity °	Number of spots	Number of images	Total
								Wt: 1	Wt: 1	Wt: 1	Wt: 1	Wt: 1	Rank
<input checked="" type="checkbox"/>	ref-t2-35b1-2	1	10:15:50	C2221	70	206	107	#1	223,0	#9 0,54	#1 2606	#13 107	#1 79 %
<input checked="" type="checkbox"/>	ref-p2_7_dna	1	15:31:14	P4	66	66	101	#2	547,5	#8 0,53	#9 1107	#4 61	#2 64 %
<input checked="" type="checkbox"/>	ref-mfe1fl7	1	17:37:39	P222	65	125	230	#3	593,6	#3 0,47	#25 318	#1 36	#3 59 %
<input checked="" type="checkbox"/>	ref-mfe1fl6	1	18:20:36	P222	65	126	227	#4	515,1	#5 0,49	#12 929	#9 89	#4 54 %
<input checked="" type="checkbox"/>	ref-mfe1-rpmfe1fl-5	1	11:52:09	P222	66	126	227	#5	628,7	#1 0,40	#4 1574	#19 430	#5 54 %
<input type="checkbox"/>	ref-mfe1fl6b	3	19:37:40	P222	66	126	227	#6	555,5	#2 0,46	#14 820	#8 85	#6 53 %
<input type="checkbox"/>	ref-mfe1-rpmfe1fl-17	1	16:38:46	P222	65	126	230	#7	3,40	#10 0,58	#22 498	#2 46	#7 52 %
<input type="checkbox"/>	ref-mfe1-rpmfe1fl-16	1	16:35:08	P222	65	126	227	#8	3,08	#7 0,52	#7 1143	#10 93	#8 52 %
<input type="checkbox"/>	ref-mfe1fl1b	1	20:16:36	P222	66	126	229	#9	2,85	#3 0,53	#16 755	#12 99	#9 52 %
<input type="checkbox"/>	ref-t2-35d6-1	4	11:25:50	P2	74	107	102	#10	3,34	#4 0,48	#15 775	#7 80	#10 51 %
<input type="checkbox"/>	ref-mfe1-rpmfe1fl-4	1	11:47:43	P222	66	126	227	#11	3,13	#15 0,66	#2 2229	#21 740	#11 51 %
<input type="checkbox"/>	ref-t2-35d6-1	1	10:19:09	P2	74	107	102	#12	2,99	#11 0,74	#10 1100	#5 70	#12 50 %
<input type="checkbox"/>	ref-dna-HA00AQ0066	1	16:09:47	P222	65	126	230	#13	3,08	#17 0,72	#21 510	#3 48	#13 50 %

Sample Ranking

- Sample Ranking Comparison



Future Developments

- EDNA :
 - DNA → EDNA on ESRF MX beamlines
 - Estimation of sensitivity to radiation damage by sacrificing a crystal
- MxCube :
 - Screening
 - Display selected samples in MxCube for Data Collect
- ISPyB :
 - More tables / columns for EDNA characterisation
 - Store Sample scoring in ISPyB database

Acknowledgements

- The DNA team
- The EDNA team
- The ISPyB team
- The MxCube team

- ...and thank you for your attention!