

Optimising the Diamond experience from a user's perspective

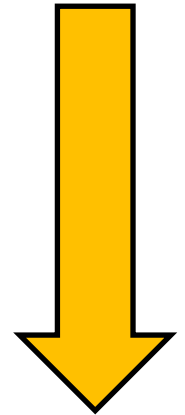
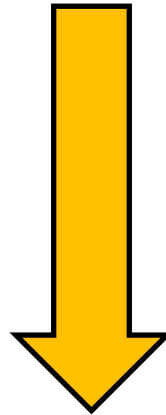
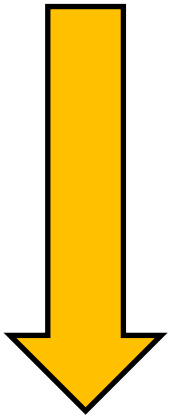
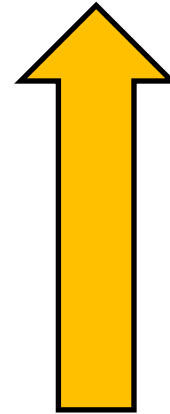
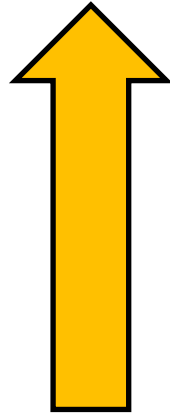
Dave Lawson

Department of Biological Chemistry

John Innes Centre, Norwich, UK



Data collection philosophy...



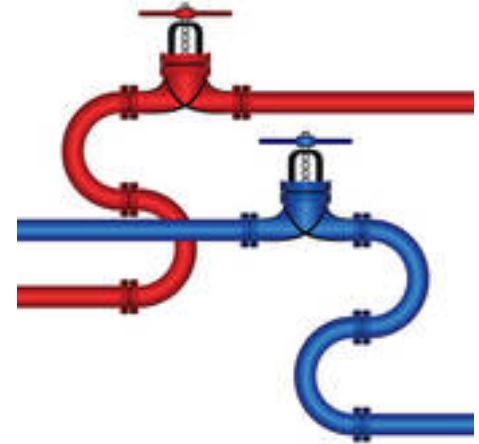
Strategy



Tools



Pipelines



Remote



- Maximizing efficiency
- Minimizing time commitment
- How to ~~collect~~ data

“Routine” data collection:

- on non-hazardous samples
- at cryogenic temperatures
- on pre-cooled crystals in pucks
- at conventional wavelengths

Disclaimer: the thoughts, views, and opinions expressed here are entirely my own and do not reflect the opinions of my employer, my wife, my kids, my...

Rules of thumb...

- Think!
- Be pessimistic about symmetry
 - if unsure - assume P1
- Be optimistic about resolution



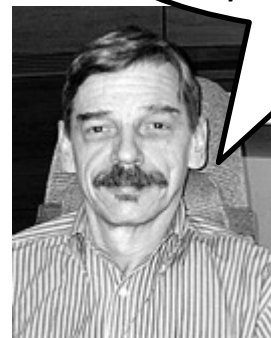
- Be prepared!



Primary goals (data quantity **AND** quality...)

- To collect as much data as possible...
- To collect the best possible data...
- To collect the data that will enable me to:
 - Solve my structure
 - Extend the resolution of my structure
 - Show that my ligand is bound

Secondary goal (get a life!)



Zbyszek
Dauter

"Data collection is
the last
experimental step!"

Ask yourself:

1. Do I spend a long time re-processing and analysing data at the home lab?
2. Does the next visit come along before I have analyzed all the data from the previous trip?

Primary goals (data quantity AND quality...)

- To collect as much data as possible...
- To collect the best possible data...
- To collect the data that will enable me to:
 - Solve my structure
 - Extend the resolution of my structure
 - Show that my ligand is bound



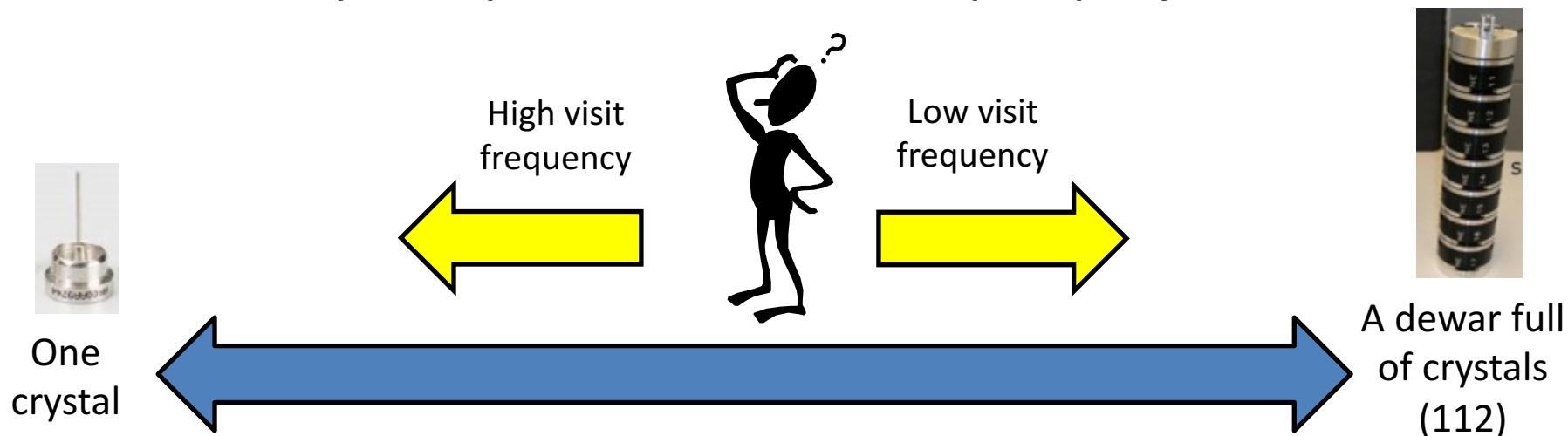
“Data collection is
the last
experimental step!”

Zbyszek
Dauter

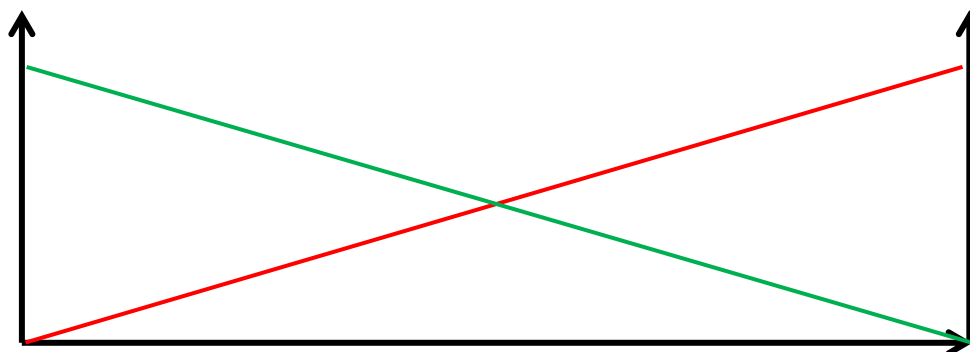
Secondary goal (get a life!)

- Try to answer these questions before end of visit:
 - Can I solve my structure?
 - Can I extend the resolution of my structure?
 - Is my ligand bound?
- Minimise the amount of follow-up work...

How many samples to take/send per project...

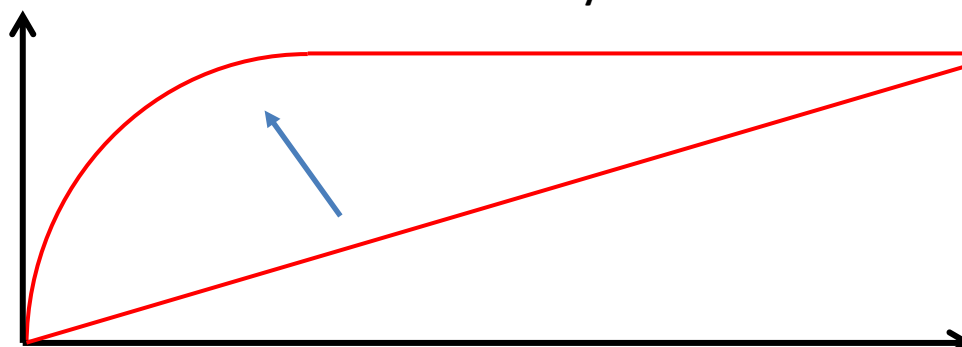


Effort before, during and after beamtime



Time available for other projects/users

Likelihood of 'success' within a single visit



Number of crystals

Data collection is always a compromise...

*How much
data do I
need?*

*What
resolution
do I need?*

*What are
the data
for?*

*How much
time do I
have?*



*How many
crystals do
I have?*



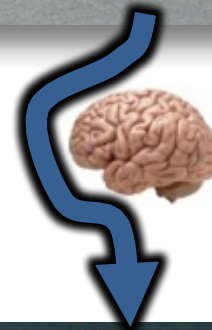
Be prepared!

Do as much of this thinking **before** your beamtime as possible!

MX data collection has become faster...

20th century data collection

21st century data collection



MX data collection has become faster...

The Pilatus detector – a real game changer!

- Typical dataset takes <3 min to collect
 - 10 - 20 GB raw data per data set!
 - >1.0 TB data possible in 24 hrs!
- In practice, 4 “useful” data sets per hour is good
 - includes screening for “best” crystal
 - ...and thinking!
- But - beamtime is still in high demand
 - users are generating crystals more rapidly
 - can get away with “marginal” samples
- So - each session is often split between several groups
 - therefore need to be efficient and organised...



Keeping pace with data collection...

data
analysis



data
collection



MX beamline tools

Timeline...

Before beamtime:

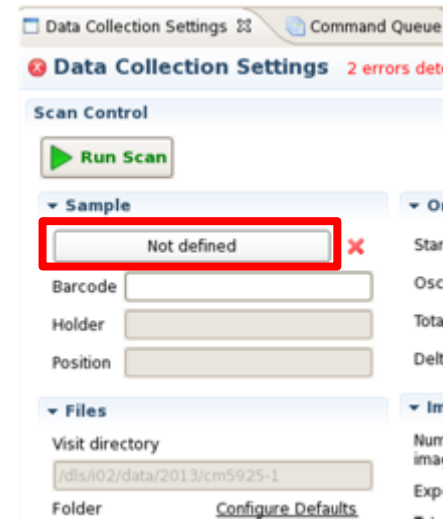
- Load crystals into pucks
- Enter sample information into ISPyB
- Upload PDBs for Dimple into ISPyB
- Upload sequences for MrBUMP into ISPyB
- Decide on running order for users/samples (may change later...)
- Share your plans with LC and EHC

Also – to avoid wasting time:

- Gather useful information:
 - Space groups and cell parameters of previous data sets
 - Phone numbers (beamline, LC, EHC, control room, other users...)
- Read beamline documentation – especially if new user
- Setup shortcuts/aliases to avoid too much typing:
 - e.g. `alias m67="cd /dls/i03/data/2016/mx12345-67"` (put in `.bashrc` file)

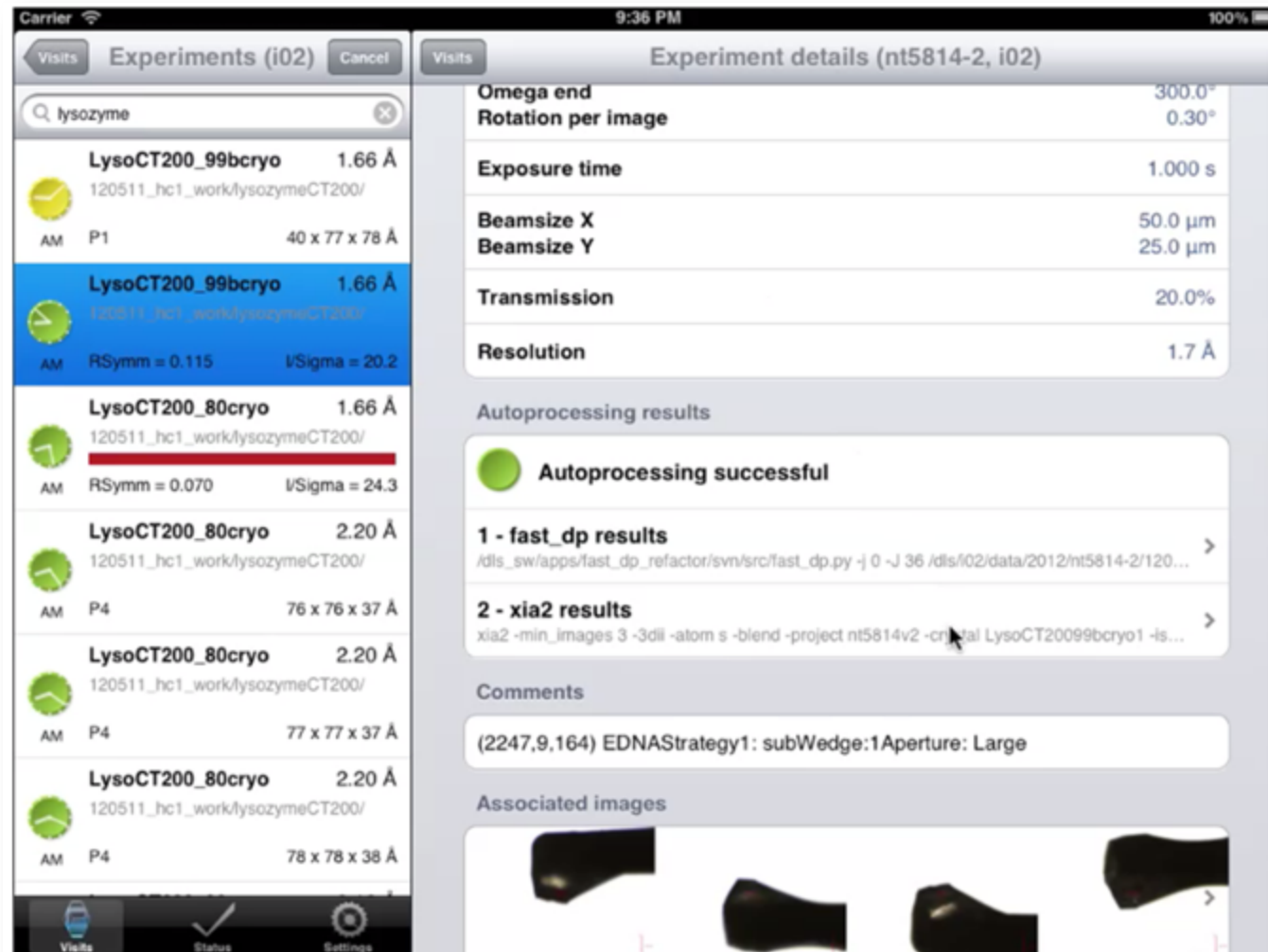
During beamtime:

- Load pucks and assign positions in ISPyB
- Select required sample from drop-down menu in GDA
 - no need to enter sample information or specify sample location
 - less likely to get the wrong sample
- Collect test images
- Based on appearance/EDNA/prior knowledge/requirements/(time left) decide to: collect/don't collect/revisit later
- Collect data set
- Analyse data as it is collected (ideally a second person...)
 - let the automated tools do the hard work (fast_dp, fast_ep, Dimple, MrBUMP...)
 - **BUT – remember to sense check the output!**
 - re-run jobs if necessary (e.g. to set space group, resolution, exclude images...)
 - (may revise sample running order/strategy based on processing results)
- Move onto next sample...



Keeping track of your data:

- Using SynchLink App (iPad/iPhone)

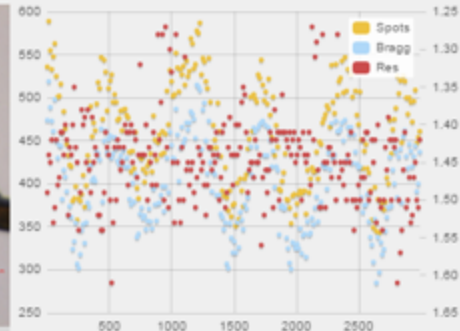
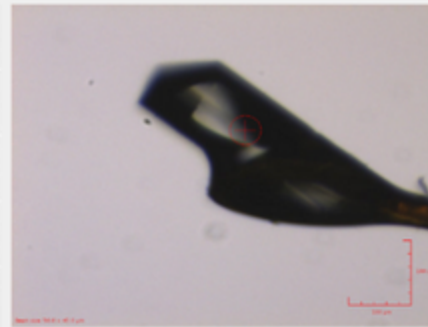
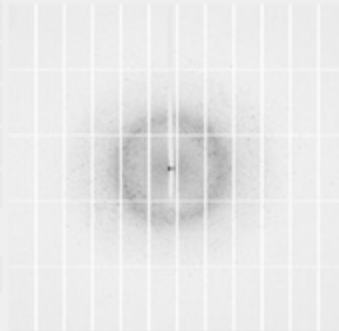


Keeping track of your data:

- Using SynchLink App (iPad/iPhone)
- Through ISPyB web interface (SynchWeb)

02-05-2014 12:37:21 - JIC/ADH4/ADH4-9/ADH4_M759_6_###.cbf

Sample: ADH4-9	Ω Start: 190.0°
Ω Osc: 0.15°	Ω Overlap: 0°
No. Images: 3000	Resolution: 1.50Å
Wavelength: 1.2820Å	Exposure: 0.040s
Transmission: 3.00%	Beamsize: 50x45µm
Type: Data Collection	
Comment: (-445,10,380)	



Auto Processing

Fast DP: Xia2:

Fast DP: [XIA2 3da](#) [XIA2 3dali](#) [XIA2 2da](#)

Space Group	A	B	C	α	β	γ
P 21 21 2	102.18	112.20	57.12	90.00	90.00	90.00

[MTZ file](#) [Log file](#) [Backup Cell](#)

Shell	Observations	Unique	Resolution	Rmeas	I/sig(I)	Completeness	Multiplicity	Anom Completeness	Anom Multiplicity
overall	2100048	148145	1.30 - 23.32	0.071	24.0	91.7	14.2	90.4	7.1
innerShell	30183	1988	5.81 - 23.32	0.061	59.0	98.4	15.2	99.1	8.8
outerShell	25217	5665	1.30 - 1.33	0.607	2.5	48.2	4.5	41.3	2.0

Downstream Processing

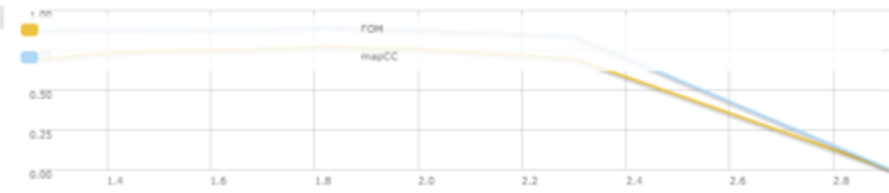
Fast EP: Dimple:

Fast EP

Figure of Merit: 0.726
Pseudo-free CC: 74.88

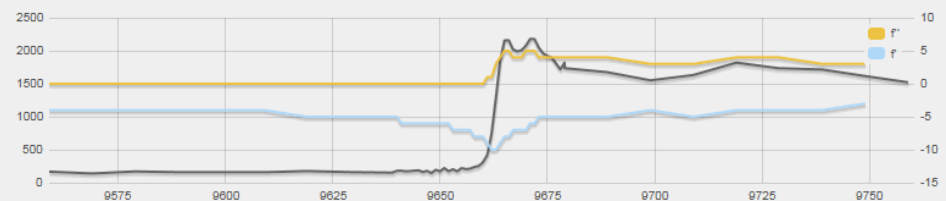
[Display / Model Viewer](#) [Download MTZ file](#) [Log file](#)

#	X	Y	Z	Occ
1	4.081	51.507	19.743	1.000
2	29.957	24.573	1.395	0.845
3	19.209	7.842	11.593	0.693
4	6.537	3.166	16.611	0.684
5	8.444	42.030	27.339	0.108



[mx9475-2] 02-05-2014 12:15:21 - Zn Edge Scan

Sample: ADH4-9	Scan File: ADH4_9_Zn_edgescan1.fluo
E(Peak): 9671eV (1.2820Å)	f': 5.23 / f'': -6.73e
E(Inf): 9663eV (1.2831Å)	f': 3.15 / f'': -10.45e
Exposure: 1s	Transmission: 0.40%
Comment: Click to edit	



xia2 Processing Report: mx9475v2

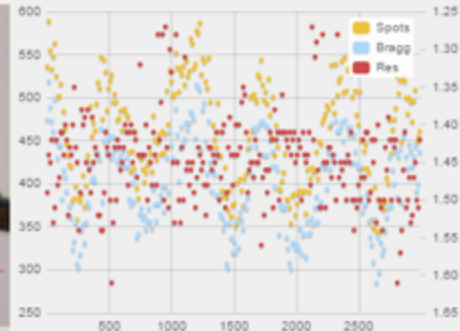
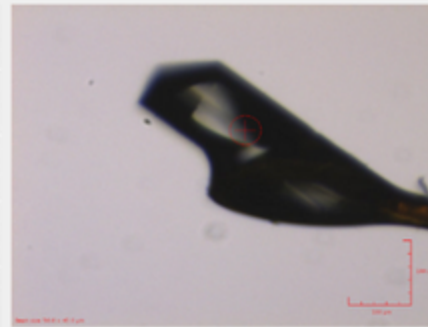
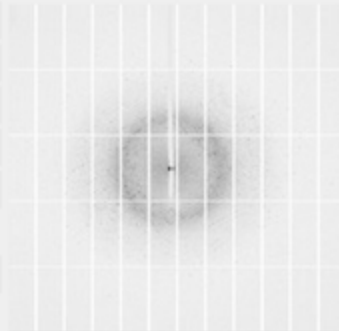
XIA2 version 0.3.6.3 completed with status 'normal termination'

Read output from /dls/i04/data/2014/mx9475-2/processed/JIC/ADH4/ADH4-9/ADH4_M7S9_6_/xia2/3da-run

SAD	
Wavelength (Å)	1.28202
High resolution limit	1.30 (5.81 - 1.30)
Low resolution limit	23.32 (23.32 - 1.33)
Completeness	91.7 (98.4 - 48.2)
Multiplicity	14.2 (15.2 - 4.5)
I/sigma	24.0 (59.0 - 2.5)
R _{merge}	0.060 (0.048 - 0.486)
Anomalous completeness	90.4 (99.1 - 41.3)
Anomalous multiplicity	7.1 (8.8 - 2.0)
See all statistics	
Unit cell dimensions: a (Å)	102.180
b (Å)	112.200
c (Å)	57.120
α (°)	90.000
β (°)	90.000
γ (°)	90.000
Spacegroup	P 2 ₁ 2 ₁ 2
Sfcheck twinning score	2.16
Your data do not appear to be twinned	
All crystallographic parameters..	

02-05-2014 12:37:21 - JIC/ADH4/ADH4-9/ADH4_M759_6_###.cbf

Sample: ADH4-9	Ω Start: 190.0°
Ω Osc: 0.15°	Ω Overlap: 0°
No. Images: 3000	Resolution: 1.50Å
Wavelength: 1.2820Å	Exposure: 0.040s
Transmission: 3.00%	Beamsize: 50x45µm
Type: Data Collection	
Comment: (-445,10,380)	



Auto Processing

Fast DP: ✓ Xia2:

Fast DP: ☒ Xia2 3dai ☒ Xia2 2dai ☒ Xia2 2dai

Space Group	A	B	C	α	β	γ
P 21 21 2	102.18	112.20	57.12	90.00	90.00	90.00

[Download MTZ file](#) [Download Log file](#) [Download Cell](#)

Shell	Observations	Unique	Resolution	Rmeas	I/sig(I)	Completeness	Multiplicity	Anom Completeness	Anom Multiplicity
overall	2100048	148145	1.30 - 23.32	0.071	24.0	91.7	14.2	90.4	7.1
innerShell	30183	1988	5.81 - 23.32	0.061	59.0	98.4	15.2	99.1	8.8
outerShell	25217	5665	1.30 - 1.33	0.607	2.5	48.2	4.5	41.3	2.0

Downstream Processing

Fast EP: ✓ Dimple:

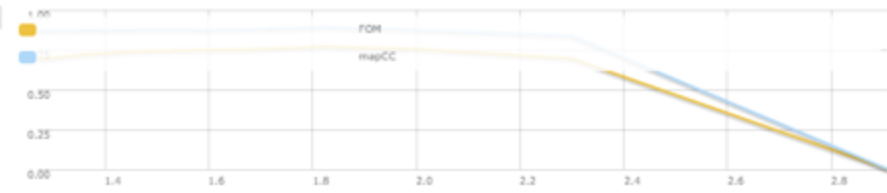
Fast EP

Figure of Merit: 0.726
Pseudo-free CC: 74.88

[Display / Model Viewer](#)

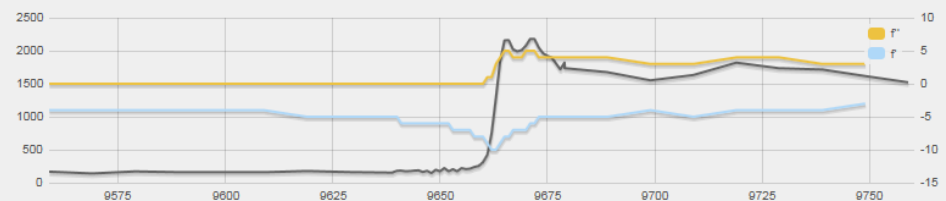
[Download MTZ file](#) [Download Log file](#)

#	X	Y	Z	Occ
1	4.081	51.507	19.743	1.000
2	29.957	24.573	1.395	0.845
3	19.209	7.842	11.593	0.693
4	6.537	3.166	16.611	0.684
5	8.444	42.030	27.339	0.108



[mx9475-2] 02-05-2014 12:15:21 - Zn Edge Scan

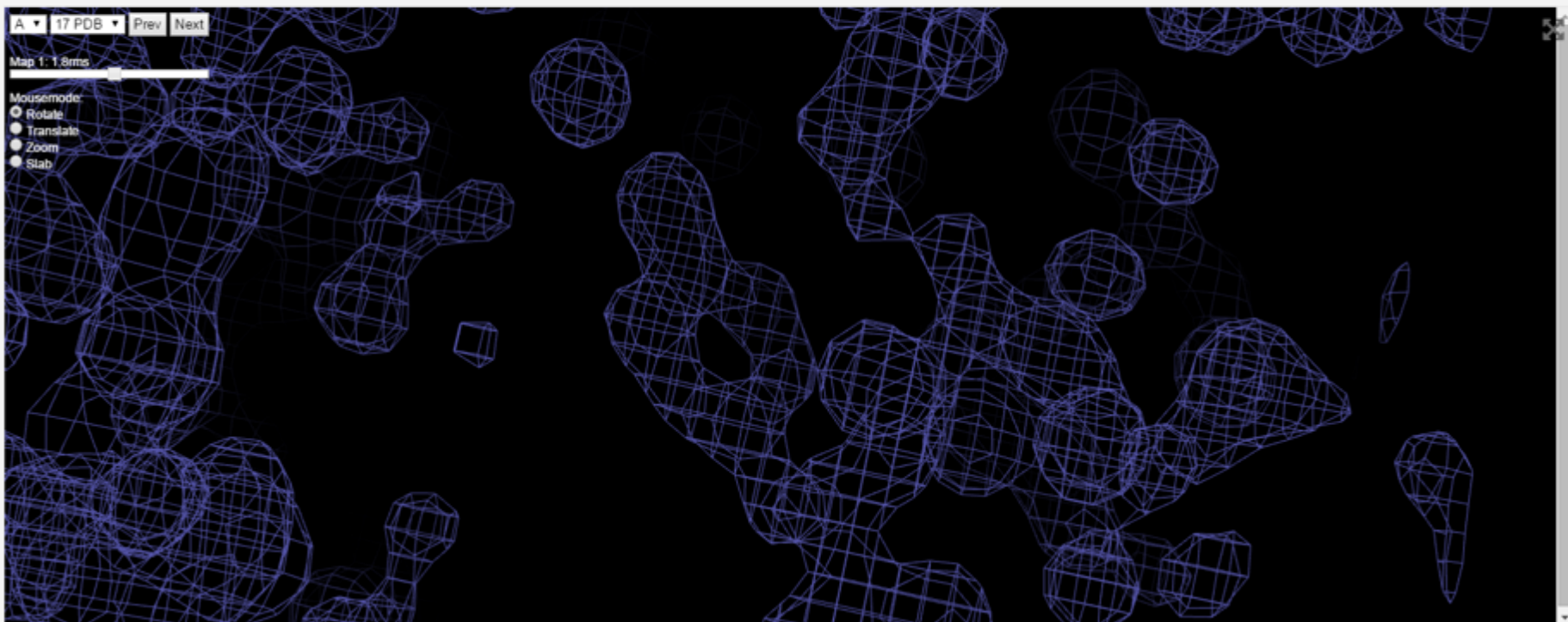
Sample: ADH4-9	Scan File: ADH4_9_Zn_edgescan1.fluo
E(Peak): 9671eV (1.2820Å)	f': 5.23 / f'': -6.73e
E(Inf): 9663eV (1.2831Å)	f': 3.15 / f'': -10.45e
Exposure: 1s	Transmission: 0.40%
Comment: Click to edit	



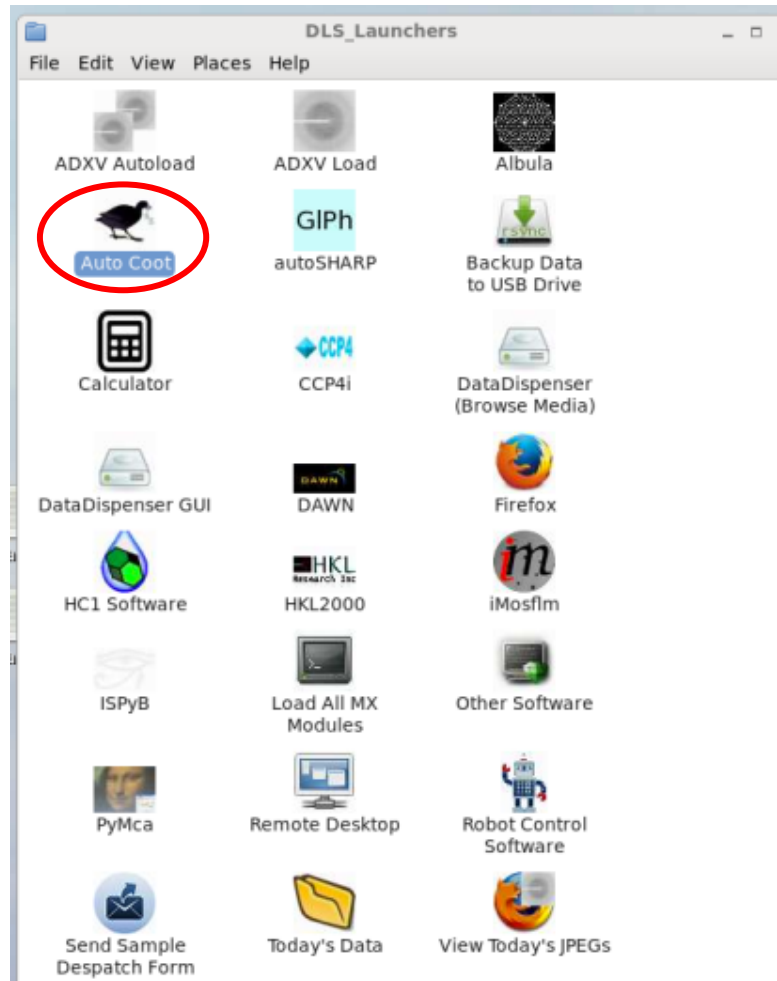
This is the new interface to ISPyB at Diamond. If you want to escape to the old version click [here](#)

mx9475-2: JIC/ADH4/ADH4-9/ADH4_M7S9_6

This page is an interactive map and model viewer. You can rotate the view using left mouse. Mousewheel zooms in and out. You can navigate through residues by pressing space or shift+space. Toggle fullscreen with the icon in the top right. CTRL + left click to translate

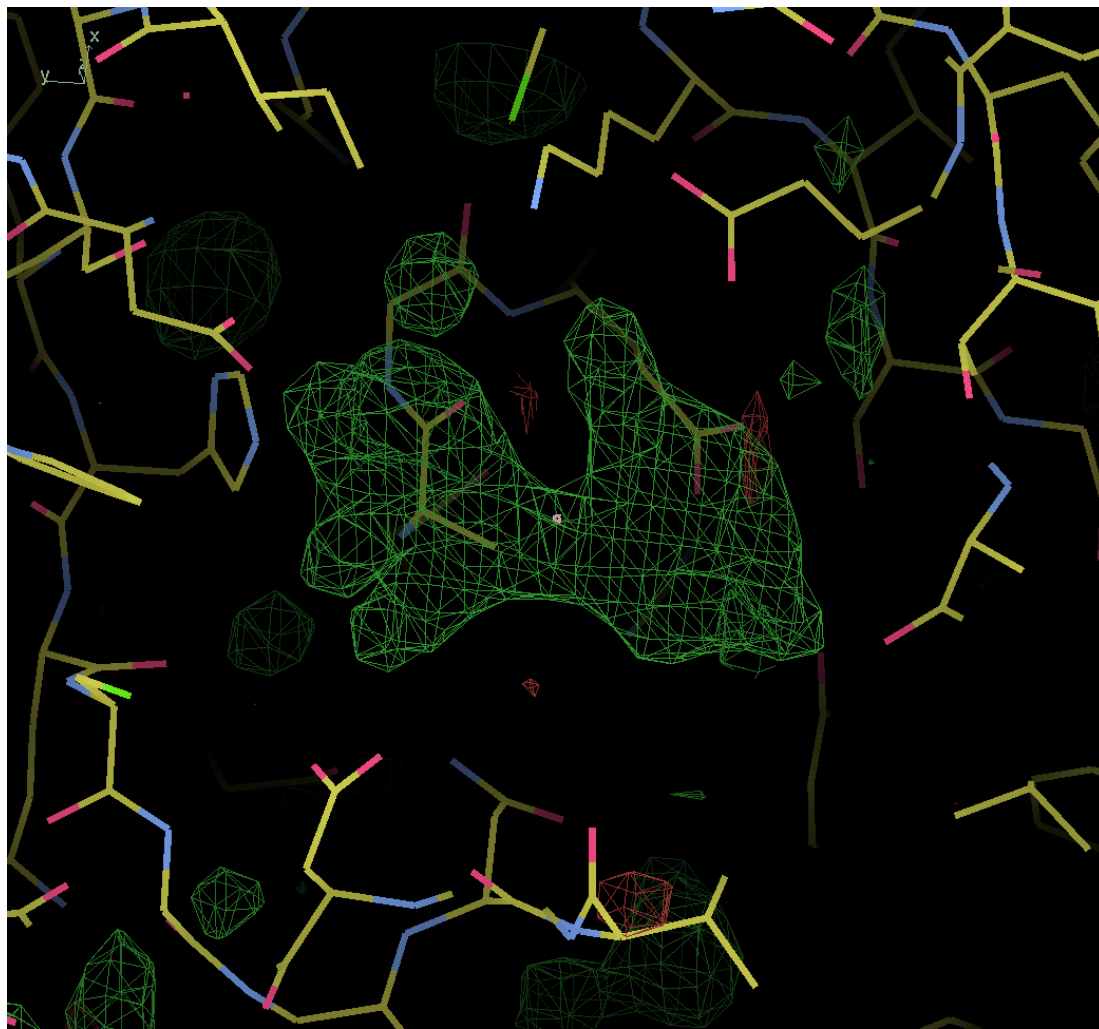


Looking at maps in Coot:



...or go to the `dimple/fast_ep` directory in “processed” and type:
`./coot.sh`

Dimple (difference map pipeline)



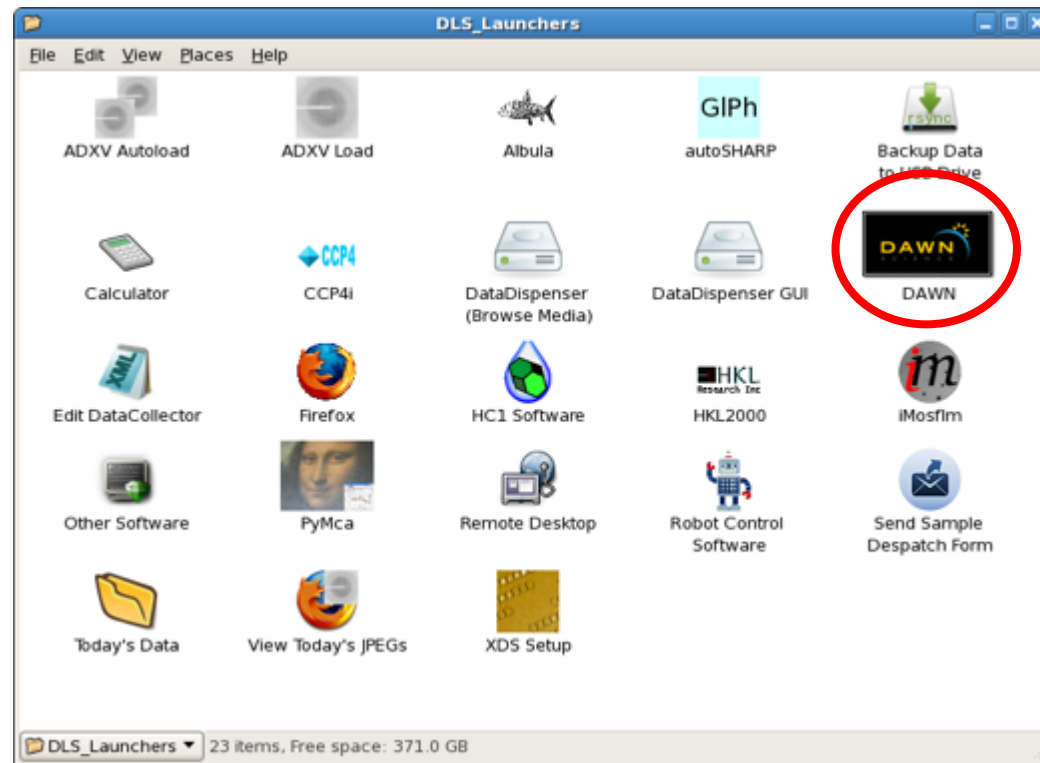
Step	Approx. time (mins)
Data collection	2
Fast_dp	2
Dimple	1
1 st image -> map	5

Get the “answer” whilst working on the next crystal...
...assuming no jobs fail!

NB. Sense check the result e.g. are R-values reasonable?

Re-running jobs:

- Most pipelines will run from the command line
- Fast_dp, Xia2 and DIMPLE will also run through the DAWN GUI:



- Coming soon: running DIALS through SynchWeb (ISPyB interface)...

By end of visit:

- Each data set should be characterised as:
 - useful/may be useful/not useful
- Some data sets may be re-processed to your satisfaction
- You may have interpretable experimental maps
- You may have preliminary structures
- You **will** have less follow-up work to do!

After your beamtime:

- Re-process “useful” and “may be useful” datasets if necessary
- Use Diamond computers through the NXclient:
 - using latest versions of software
 - no local installation of pipelines required
 - likely to be much faster than your workstation
- Copy results home
- With help of ISPyB database:
 - keep track of data collections
 - document your results and conclusions



This is the new interface to ISPyB at Diamond. If you want to escape to the old version

Visits for mx9475

This page lists the visits available to the

Show entries

Start Date

End Date

21:00 01-08-2014

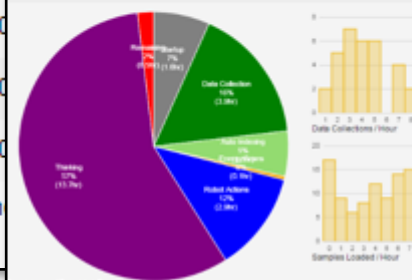
02:00 02-08-2014

21:00 01-08-2014

02:00 02-08-2014



Started: 02-05-2014 09:00
Data Collections: 313 (74 / 179)
Collections Skipped: 5
Energy Scans: 1
MCA Spectra: 1
Samples Loaded: 162



Sample	Images	Res	k	Q Osc	Spacegroup	Unit Cell	Process	Resolution
ADH4-9	1000	1.2	0.9795	0.10	P 21 2 2	101.85, 112.21, 57.09	57.09 - 1.26	57.09 - 1.26
ADH4-9	5	1.5	1.2820	0.50				
ADH4-9	3000	1.5	1.2820	0.15	P 21 2 2	102.18, 112.28, 57.12	102.18 - 1.3	102.18 - 5.81
ADH4-9	3000	1.5	1.2820	0.15	P 21 2 2	102.22, 112.25, 57.15	75.58 - 1.31	75.58 - 5.86
ADH4-9	2000	2.4	2.0000	0.15	P 21 2 2	102.28, 112.32, 57.20	112.32 - 2.01	112.32 - 8.99
KJ_NRC1_45	43	3.0	2.0000	0.00				
KJ_NRC1_45	12	3.0	2.0000	0.00				
KJ_NRC1_45	35	3.0	2.0000	0.00				
KJ_NRC1_45	18	3.0	2.0000	0.00				
KJ_NRC1_45	5	3.0	2.0000	0.50				
KJ_NRC1_45	3800	3.0	2.0000	0.20	P 6	111.06, 111.06, 54.21	27.76 - 6.39	0.000 94.5
KJ_NRC1_50	56	3.0	2.0000	0.00				
KJ_NRC1_50	15	3.0	2.0000	0.00				
KJ_NRC1_50	5	3.0	2.0000	0.50				
KJ_NRC1_50	3800	3.0	2.0000	0.20	P 62	110.66, 110.66, 53.97	53.97 - 6.3	0.227 99.7
KJ_NRC1_54	5	3.0	2.0000	0.50				
KJ_NRC1_53	24	3.0	2.0000	0.00				
KJ_NRC1_53	12	3.0	2.0000	0.00				

to edit

173

to edit

73

to edit

257

to edit

50

to edit

51

to edit

85

to edit

0

to edit

102

to edit

310

i03

Dr Graeme Winter

Click to edit

b21

Dr Katsuaki Inoue

Click to edit

i04-1

Dr Keith Henderson

Click to edit

i04

Dr Pierpaolo Romano

Click to edit

First Previous 1 2 3 4 Next Last

Summary - why use the automated tools?

- Faster sample changing (ISPyB)
 - essential for remote...
- Informs the decision making process
 - make decisions sooner
 - revise strategy on the fly
 - (e.g. recollect dataset x... no more data required for project y...)
 - make better overall use of beamtime
- Easier to cope with a small team
- Reduces amount of post-beamtime follow-up work
- Simple to keep track of your samples and data (ISPyB)

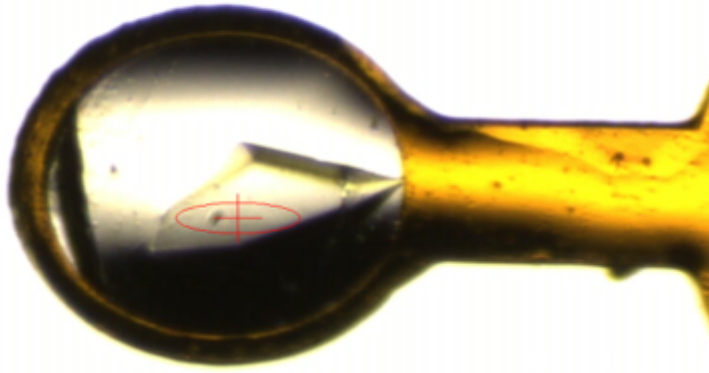
Remote collection



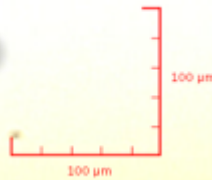
...from here



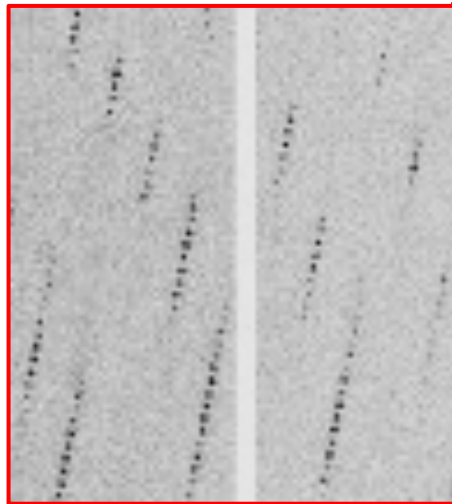
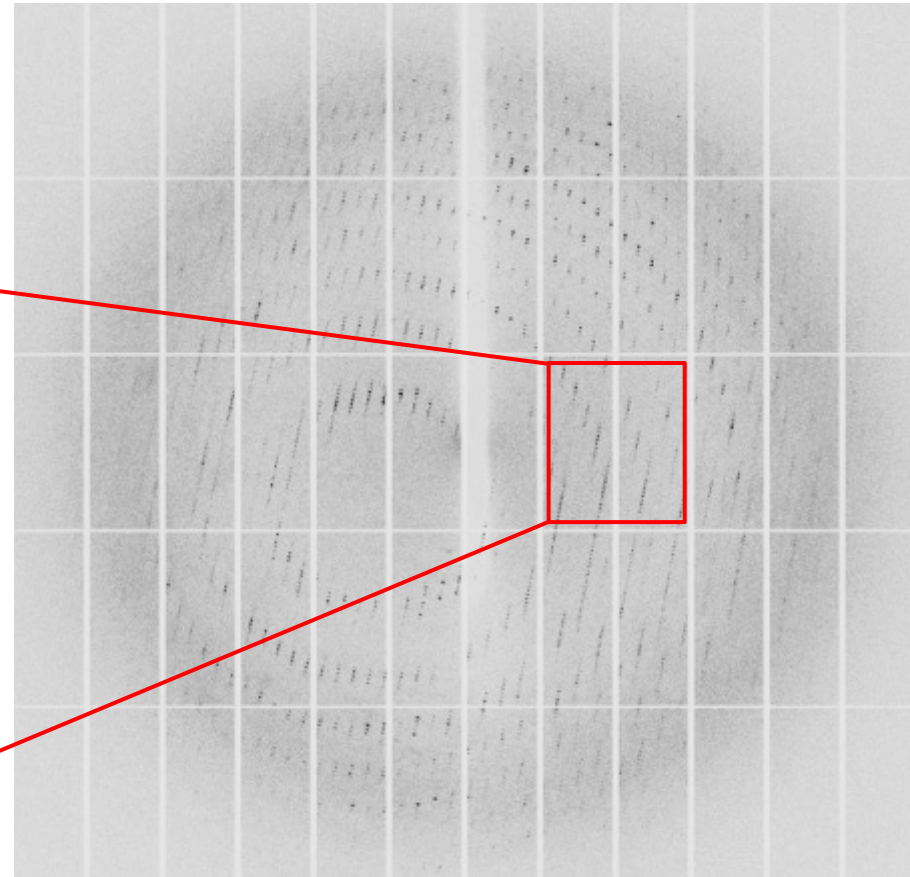
Some real data...



Crystal cooled in loop



Diffraction pattern recorded from crystal at Diamond

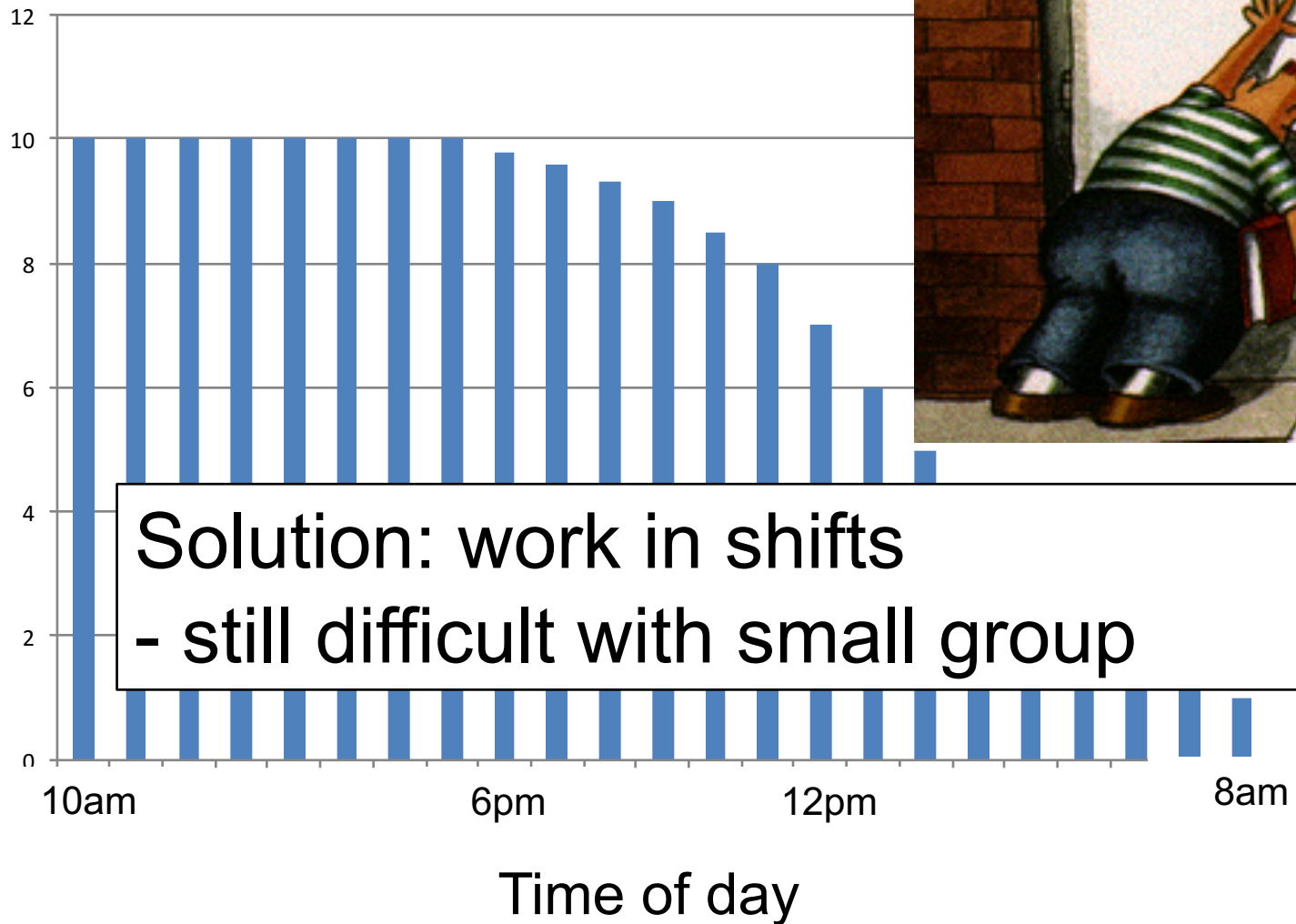


Beam size: 82.0 x 21.0 μm

Some fictitious data...

A day at the beamline...

Data quality



Who should collect the data?

- Someone who knows the project
 - what data are required?
 - what space group/cell parameters/resolution to expect
- Person who grew and mounted the crystals
- Someone who is not tired!

Ideal scenario:

- everyone collects data from their own crystals
- work in short shifts



Solutions:

1. everyone travels to Diamond (not practical ... and expensive)
2. small team travels to Diamond – everyone else collects data remotely (mixed access)
3. everyone collects data remotely (fully remote access)

Who travels to Diamond?

Without remote users:

- 3 people minimum – at least 2 are experienced users

With remote users (mixed access):

- 2-3 people – only 1 needs to be an experienced user
 - good way to introduce new users to DLS



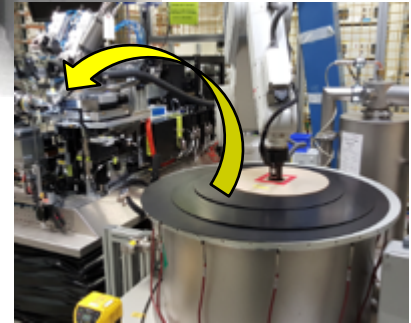
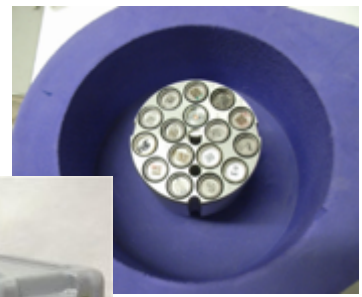
Who collects data remotely?

- Users at multiple sites; at each of these:
 - 1 experienced user can cope, but better with 2 (second person analyses data)
 - any number of new users (+observers)



For routine data collection at 100 K:

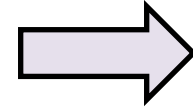
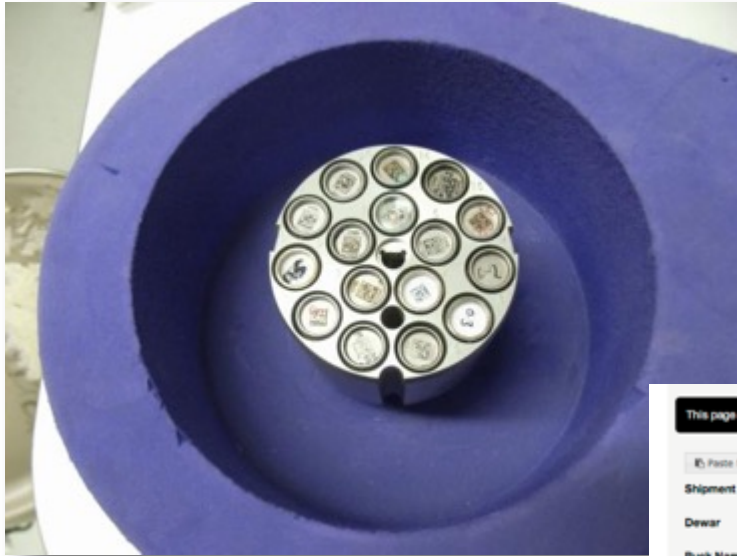
- Samples prepared and cryo-cooled in home lab
- Transported to DLS in dry shipping dewars
- Samples mounted robotically
- Data collection controlled through GDA GUI
- Only manual operation at DLS: loading/unloading pucks
 - Up to 480 or 592 samples can be loaded at once (depending on beamline)
 - may only need to enter hutch at beginning and end of session
- Everything else computer driven...



Therefore you don't need to be there!

ISPyB database & shipping

- Create shipment in ISPyB
- Enter sample info. directly into ISPyB database

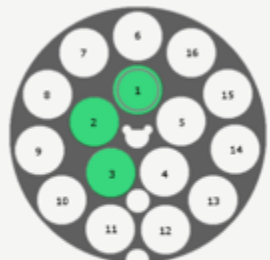


This page allows you to add containers to the selected dewar and shipment. If the protein you want to use isn't listed type in a new name and press tab. This will create a new protein

Shipment:

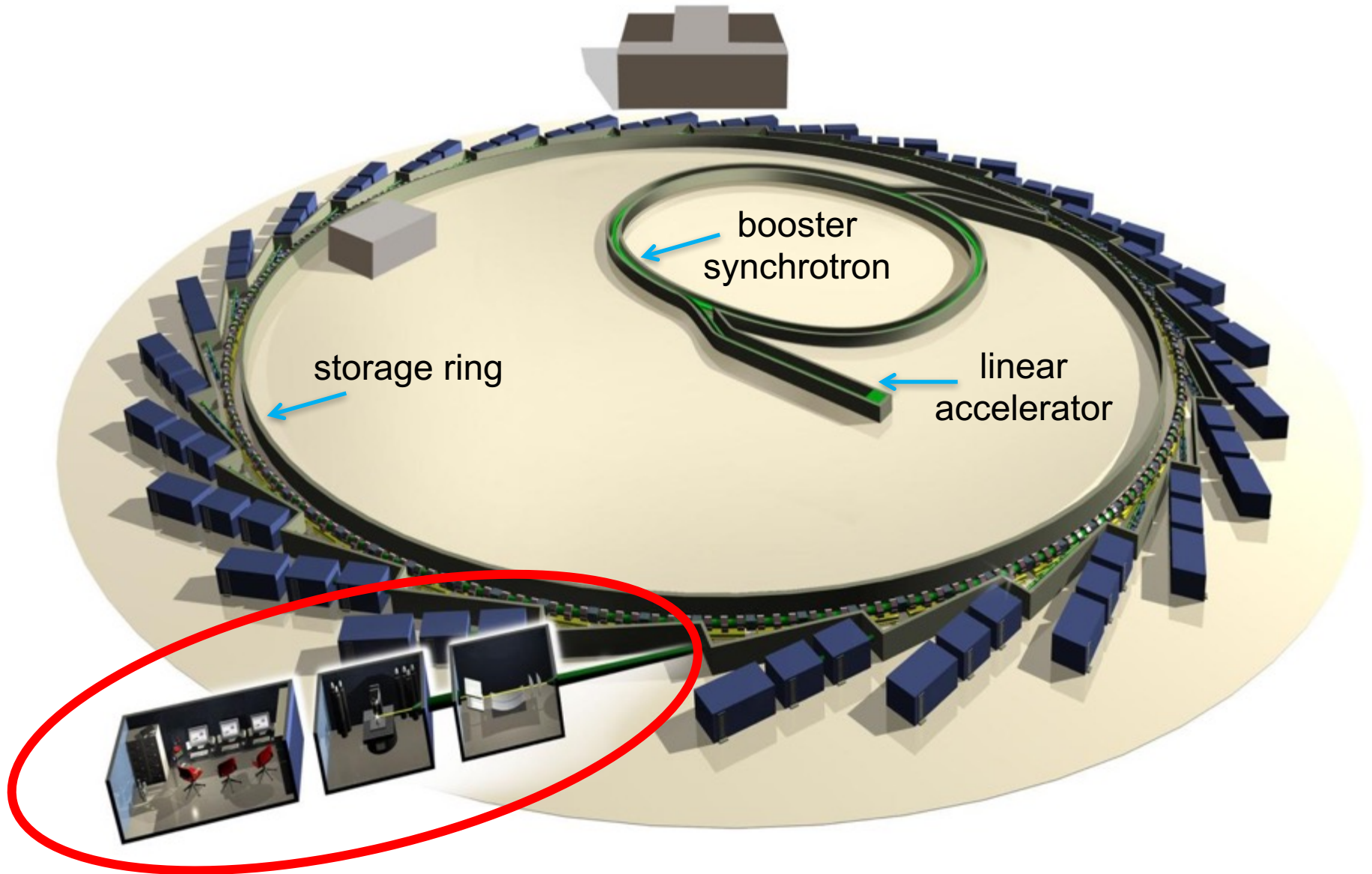
Dewar:

Puck Name:



Location	Protein Acronym	Sample Name	Spacegroup	Barcode	Comment	
1	lys	lys1	P			O X
2	lys	lys2	P			O X
3	lys	lys3	P			O X

What is remote data collection?





In This Section

Manual

[Preparing for Beamtime](#)
[Experiment DB - ISPyB](#)
[Shipping Samples](#)
[Sample Holders](#)
[Sample Changer](#)
[Remote Access](#)

Connection to Diamond

[Known Remote Access Issues](#)
[Speed Test](#)
[Webcams](#)
[Short Visits](#)
[Data Collection Software](#)
[Mini Kappa](#)
[Data Analysis](#)
[Data Backup](#)
[Webcams](#)
[Acknowledging MX beamlines](#)
[Quick Guide to MX Beamtime at](#)
[Diamond Light Source](#)
[Science Highlights](#)
[Beamtime shift Patterns](#)
[Calc](#)

Connection to Diamond

Remote Access

New remote access is available at Diamond to login on MX beamlines or NX cluster. Remote access is available from Linux, Mac OSX and Windows computers.

Please see the video below:

Diamond Remote Access



How to remotely connect to a beamline or Diamond cluster



If you are experiencing issues you can try alternative ways to connect to Diamond:

- [From Linux or Mac OS X](#)

iMac with extra monitor:
(i) data collection
(ii) data evaluation

View
today's
jpgs

2nd iMac:
(i) data evaluation
(ii) data (re)processing
(iii) Skype

Useful
phone
numbers

GDA
interface

webcams

ISPyB
interface

Coffee!

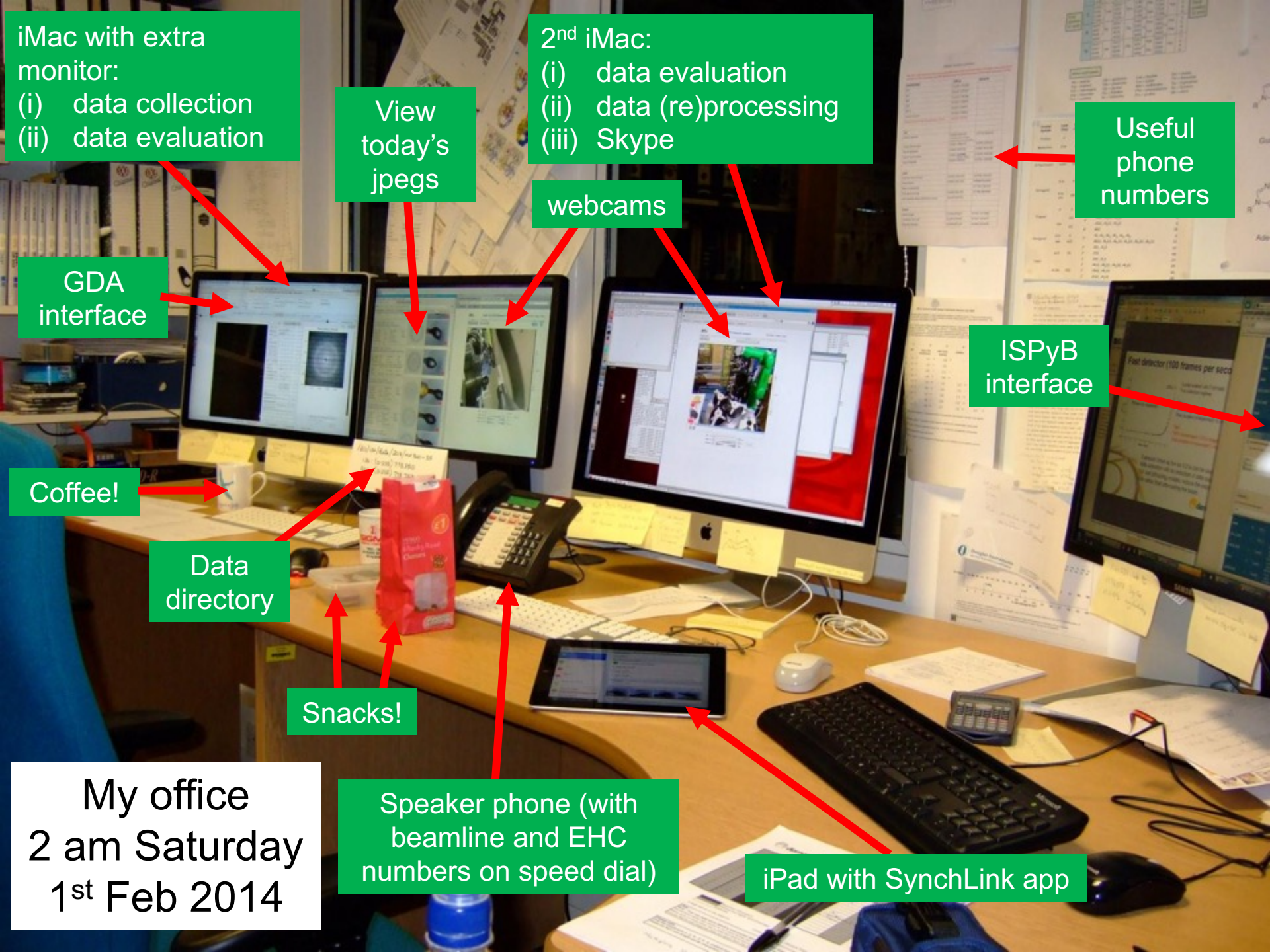
Data
directory

Snacks!

My office
2 am Saturday
1st Feb 2014

Speaker phone (with
beamline and EHC
numbers on speed dial)

iPad with SynchLink app

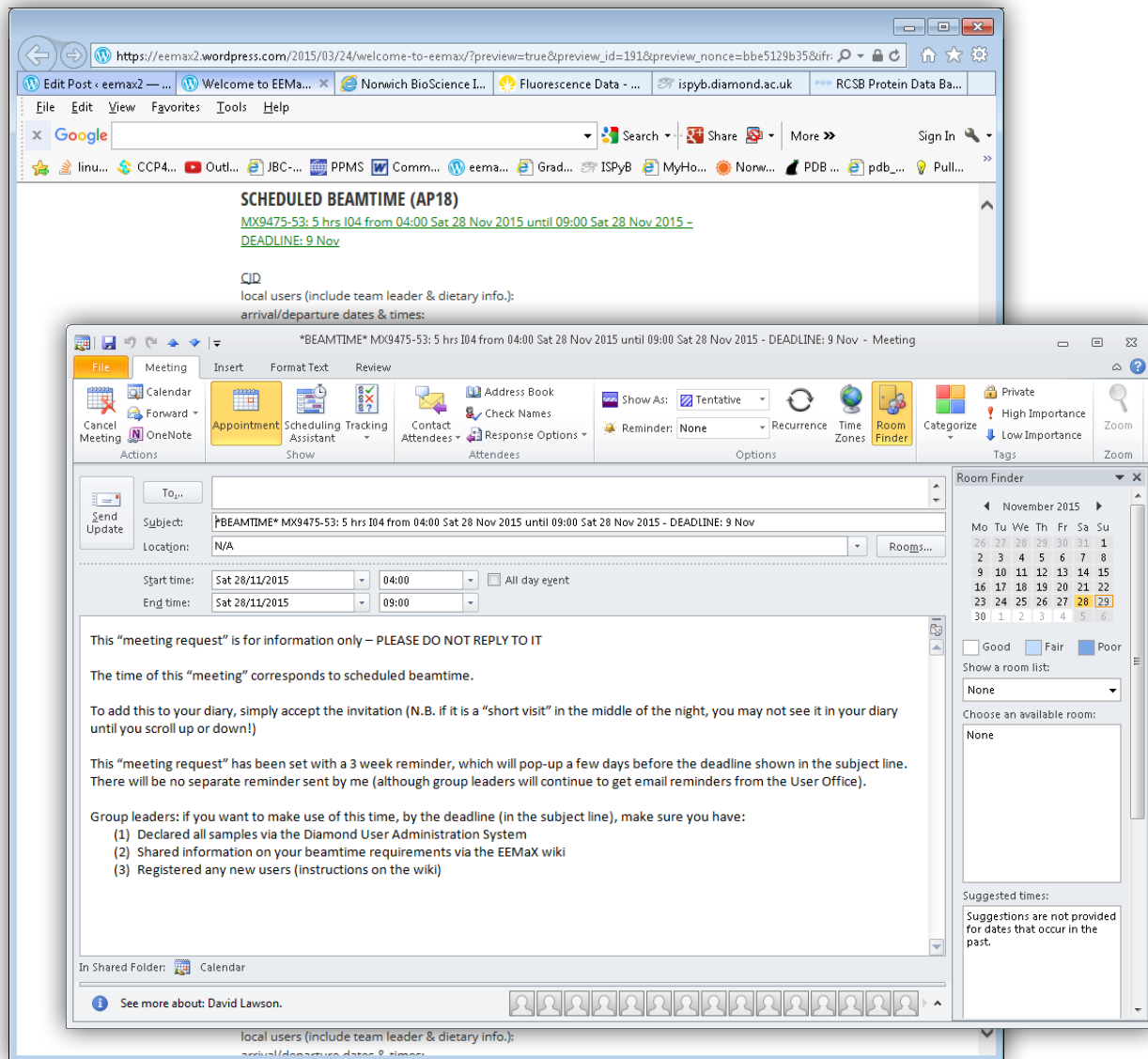


How do we manage the beamtime?

Several groups may participate in one session, so we need to be organised..

Beamtime requirements are entered into a BAG wiki

Scheduled beamtime and associated deadlines (with reminders) sent as meeting requests to Outlook diaries



How do we manage the beamtime?

- Decide on running order and approximate changeover times beforehand
 - need to be flexible e.g. if there is downtime
- For mixed access:
 - remote users generally get the night shift
 - remote users occasionally get daytime access (e.g. while on-site team analyse data or go for meals!)

Timeline for fully remote data collection

2 - 3 weeks before:

- decide who is sending samples (wiki)
- complete ERA and experiment notification form (UAS)

Week before:

- decide on running order and rough division of beamtime (via email)
- prepare samples and enter sample information into ISPyB database

2 - 4 days before:

- ship samples (N.B. must arrive during normal working hours)
- upload files for autoprocessing into ISPyB database (PDBs, sequences)

Timeline for fully remote data collection

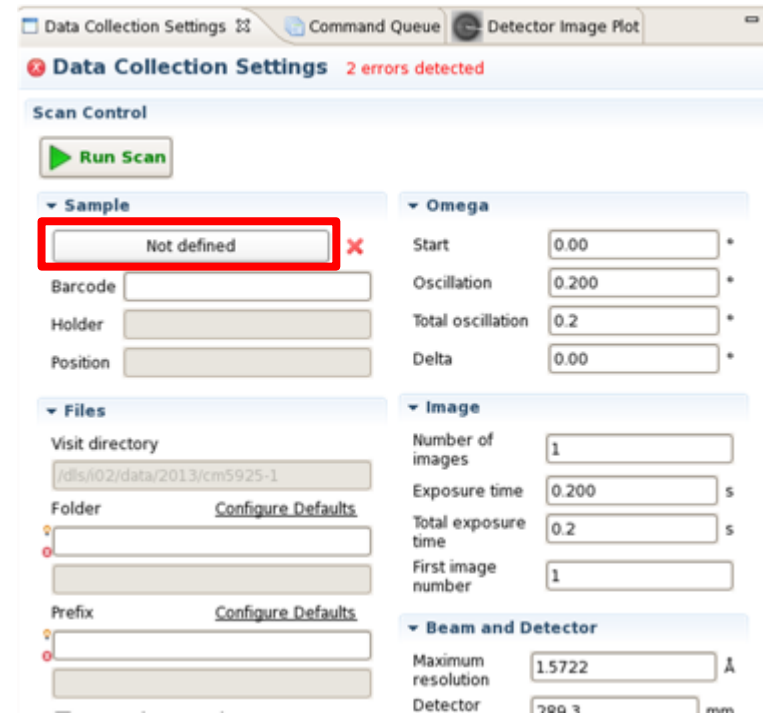
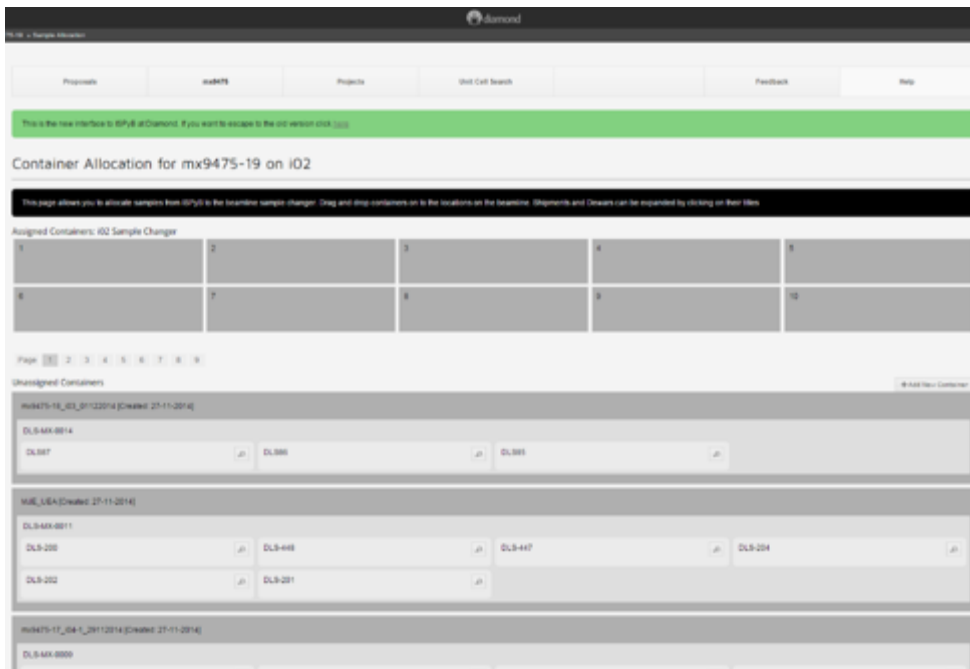
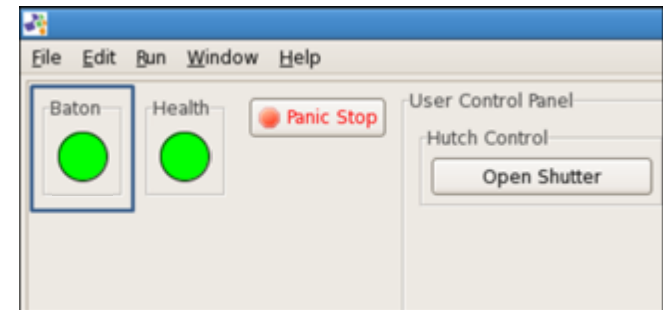
Sometime prior to shift:

- LC/EHC loads pucks into beamline dewar
- ~1 hr before start of shift (or slot in running order):
 - connect to beamline computer using NX client
 - start GDA
 - enter puck positions into ISPyB database (LC/EHC)



Zero hour:

- take the baton, select sample from menu, collect data!



Check the webcams!

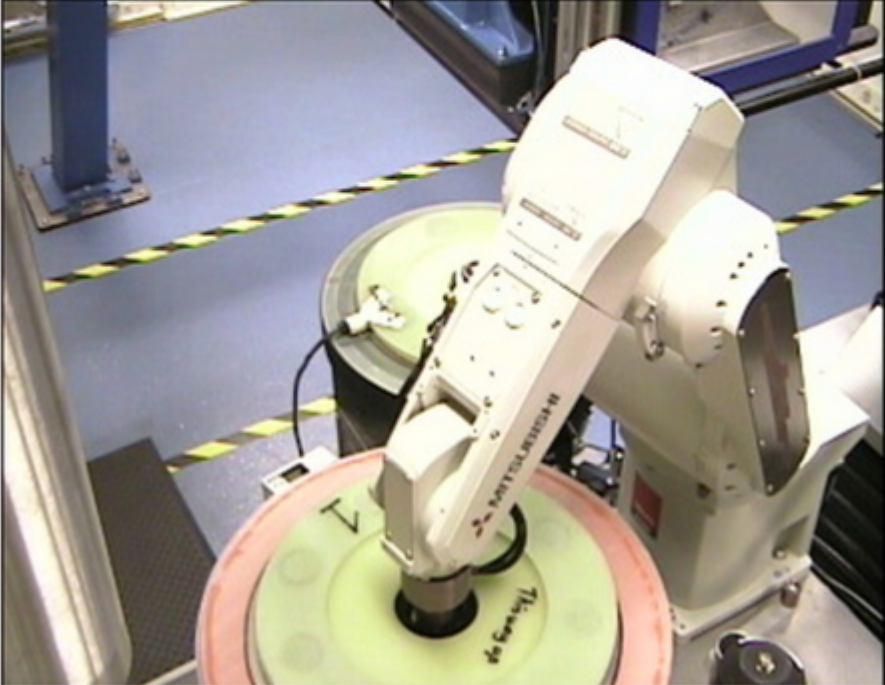
[Proposals](#)[mx9476](#)[Projects](#)[Unit Cell Search](#)[Feedback](#)[Help](#)

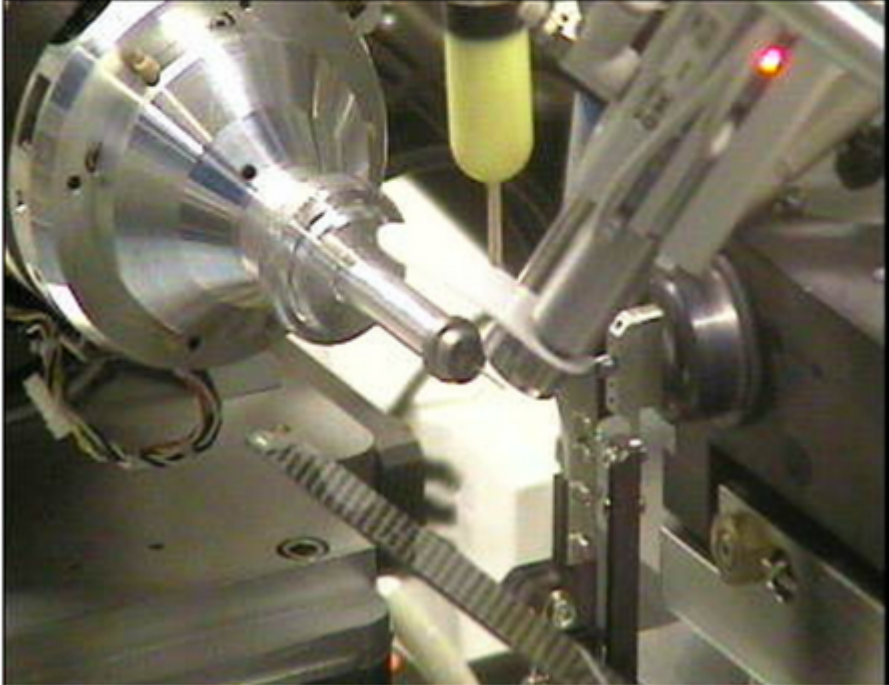
This is the new interface to ISPyB at Diamond. If you want to escape to the old version click [here](#)

This page shows all data collections for the selected visit. If the visit is ongoing the page will automatically update as new data is collected. Auto processing results will be displayed

i03 Webcams & Beamline Status

Ring Current 232.281	Refill 130.903	Hutch Locked	Port Shutter Open	Expt Shutter Open	Fast Shutter Closed	Wavelength 0.976254	Transmission 100
-------------------------	-------------------	-----------------	----------------------	----------------------	------------------------	------------------------	---------------------

i03 Experimental Hutch, Mon Dec 1 2014 10:07:50

i03 Sample Position, Mon Dec 1 2014 10:07:50

[Data Collections](#)[Full Data](#)[Auto Integrated](#)[Screening](#)[Edge Scans](#)[MCA Spectra](#)[Robot Actions](#)[Sample Actions](#)[Favourite](#)

[Summary](#)[Visit Stats](#)

People can be set in their ways...



Let go Luke...

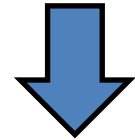


...use remote!

I know what
you're getting
for Christmas.

How?

Because I felt
your
presents!

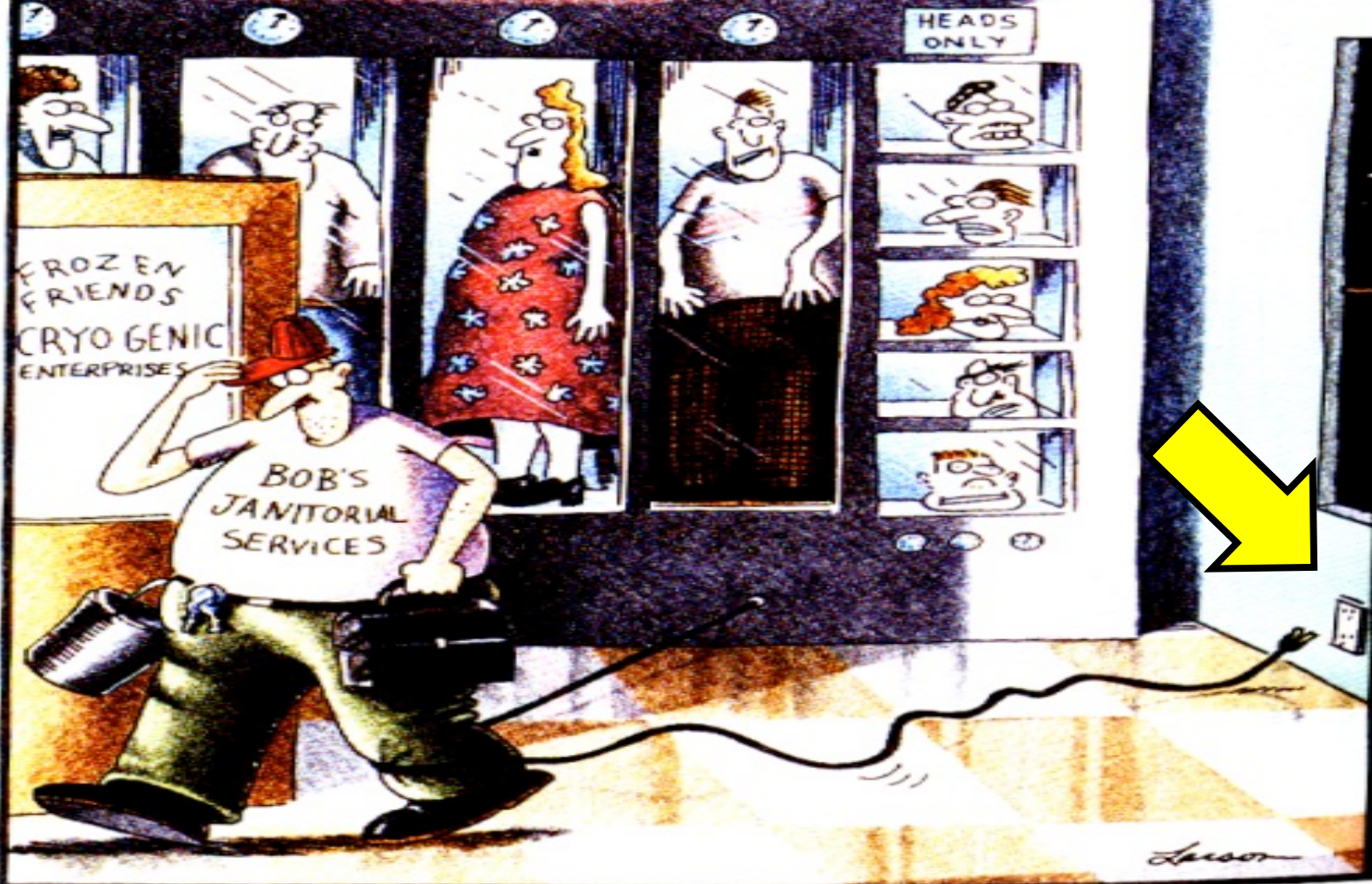


=



When things go wrong...

Warning: maintain at 100K



When things go wrong...

Can you fix things?

- probably not!

On-site user:

- During normal working hours – call local contact
- Out of hours – call EHC

Remote user:

- During normal working hours – call local contact
- Out of hours – call EHC

Therefore you don't need to be there!

Getting your data home...

- send hard drive with dewar
or...
- FTP data home



+



Both methods
are slow!

Alternative:

- Use autoproccessing output or...
- (re)process data remotely on DLS computers and FTP output only
- archive raw data later

Advantages of remote data collection:

- Beamtime is fully used
- Users collect data on their own crystals
- Most/all of the users stay at home labs
 - time commitment is low for remote users
 - stakeholders can observe data collection
 - useful for training non-experts
- Time can be used flexibly
- Less stressful
- Everyone gets some sleep!
- Difficult to “break” the beamline



Disadvantages of remote data collection:

- Risk of losing touch with beamline staff and latest developments (do BAG training, go to User meeting)
- Dewars/hard drives occasionally get misplaced...
- Don't get a cooked breakfast!



Take home messages

- If you are not using remote data collection – why not?
- If you are not using ISPyB – why not?
- If you are not using the software tools/pipelines – why not?
- If you don't have a life – you know why....



Acknowledgements

- Access to MX beamlines at Diamond
- Excellent support of beamline staff/EHCs
- Software developers
- BBSRC