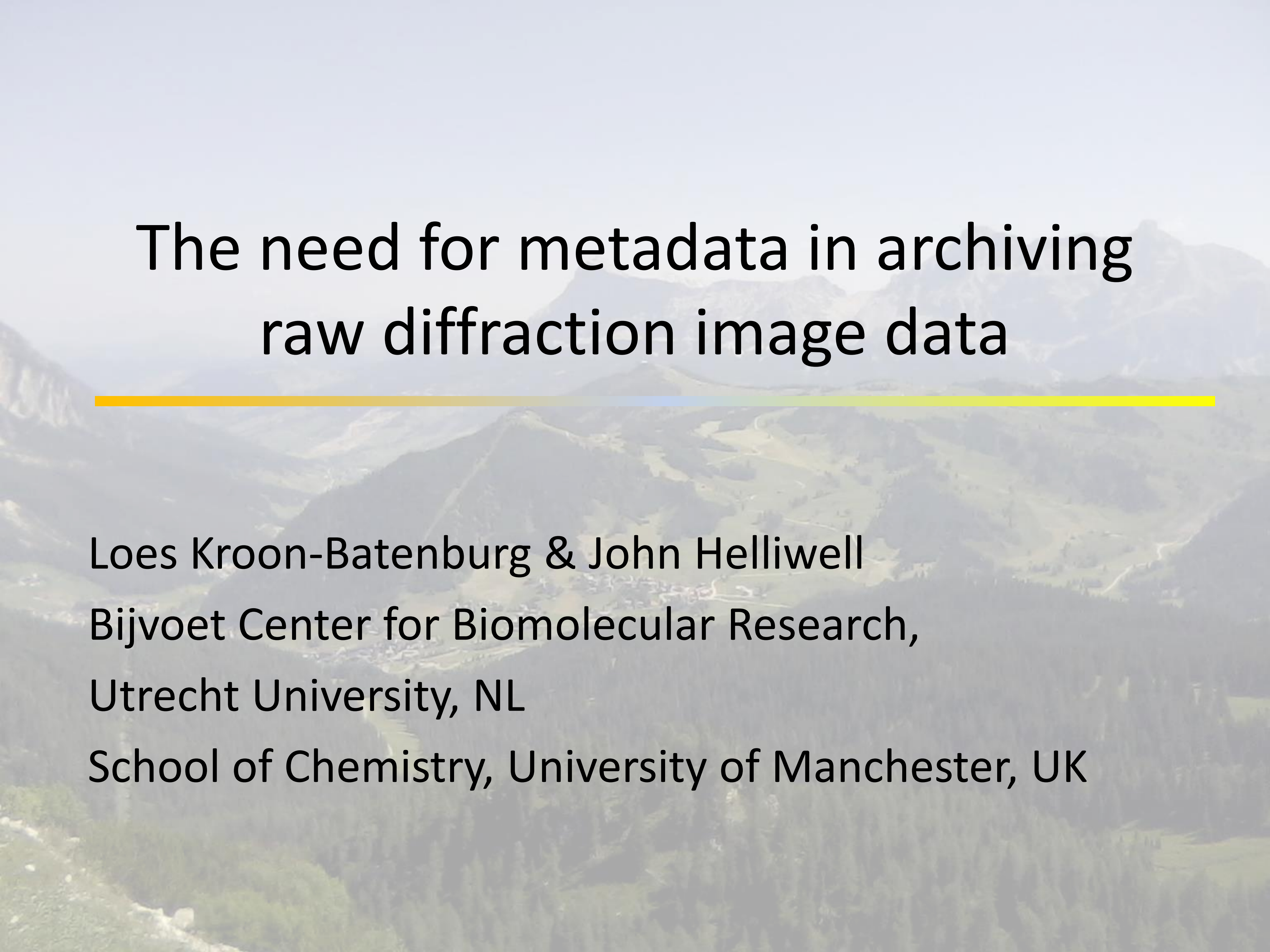


The need for metadata in archiving raw diffraction image data

Loes Kroon-Batenburg & John Helliwell
Bijvoet Center for Biomolecular Research,
Utrecht University, NL
School of Chemistry, University of Manchester, UK

Reasons for archiving raw data

- Allow reproducibility of scientific data
- Safeguarding against error and fraud
- Allow further research based on the experimental data and comparative studies
- Allow future analysis with improved techniques, changed standards or new science
- Provide example materials for teaching

Processed data

- Crystallographic structure determination is based on integration of Bragg spots
- Structure refinement against *I* or *F* (*i.e.* derived data) and prior knowledge
- PDB stores/requires:
  - Final refined coordinates
  - Derived data (structure factors)
  - Associated publication
  - Allows systematic re-analysis of all structures in the PDB, *e.g.* PDB-REDO

Raw data

- Raw images contain additional information:
  - Detector non-uniformity, spatial distortion, background scattering, detector gain: allows error analysis
  - Symmetry (unmerged data could be stored in PDB)
  - Shape of the spots (crystal size; mosaicity; lattice): improve spot integration
  - Composite patterns
  - Scattering between the spots:
    - diffuse scattering: streaks for packing defects; variational for protein dynamics
    - satellites due to incommensurately modulated structure
  - Development of validation criteria

Raw data

- Raw diffraction data that did not lead to structure solution:
  - no associated publication....
  - but very useful for methods development (see Michael Wall)
  - Data lost for ever?.....

What would be needed to run a project like PDB-REDO on raw data?

- Well documented data format, but not necessarily just one: CBF/imgCIF....
- Meta data: long term preservation

  sufficient meta data to allow

  - Discovery
  - Interpretation
  - Re-use

What is the current status?

DLS/ESRF and meta data

- Diamond Light Source and ESRF use data management system ISPyB
- Data analysis is automatically performed
- Scripts with meta data are provided for *Mosflm*, *XDS*..
- User is responsible for transfer and backup
- Long term preservation?

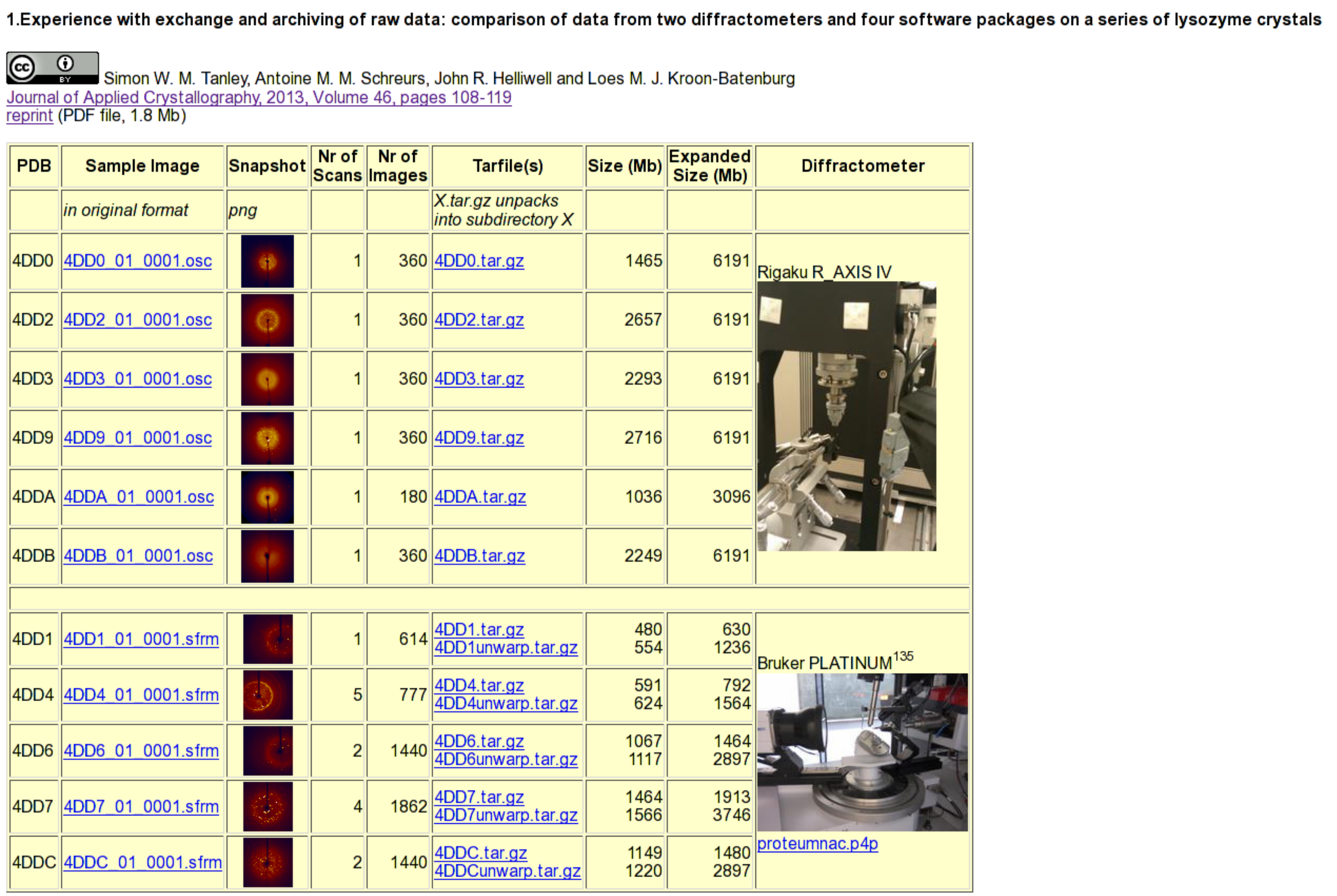
Experience with making diffraction image data available

- Initially data of 11 lysozyme crystals, on two different diffractometers, originally processed with the equipment’s built-in software, which we wanted to re-process with *EVAL* to rule out software effects
- Data were transferred from Manchester to Utrecht
- 35.3 Gb of uncompressed data. Transfer took 30 hours, spread over several days
- Data were compressed in Utrecht, using ncompress (lossless data compression with LZW algorithm) to 20 Gb, and can readily be read with *EVAL* software

Unwarping and data conversion

- Proteum Bruker images were in native u8-format (.sfrm)
- Unwarped in Manchester using .p4p file containing the distortion table
- Written to “*Mosflm*” format (2 byte) using *FrameUtility*
- Could not be read into *EVAL* in Utrecht because end-of-header at end of line

rawdata.chem.uu.nl



1.Experience with exchange and archiving of raw data: comparison of data from two diffractometers and four software packages on a series of lysozyme crystals

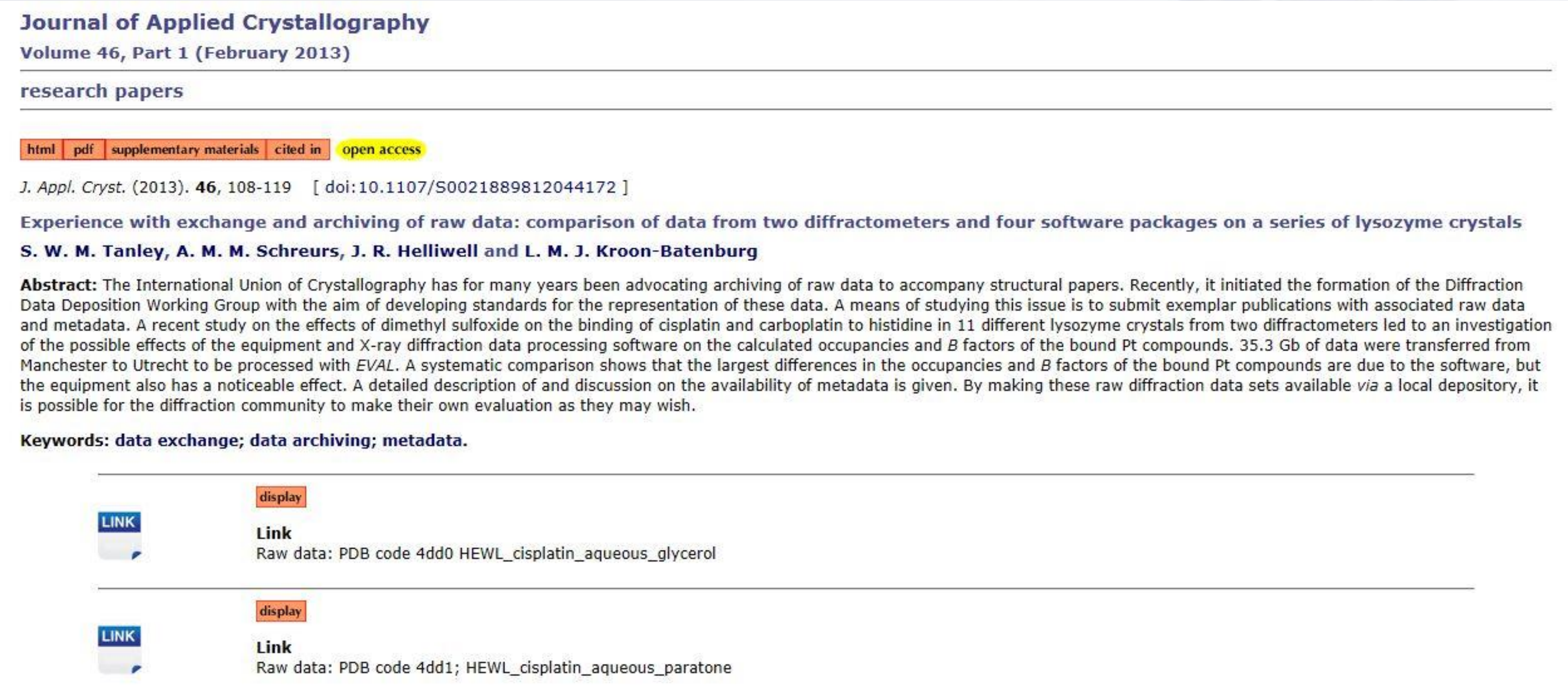
Simon W. M. Tanley, Antoine M. M. Schreurs, John R. Helliwell and Loes M. J. Kroon-Batenburg
Journal of Applied Crystallography, 2013, Volume 46, pages 108-119
reprint (PDF file, 1.8 Mb)

PDB	Sample Image	Snapshot	Nr of Scans	Nr of Images	Tarfile(s)	Size (Mb)	Expanded Size (Mb)	Diffractometer
	in original format	*png*			*X.tar.gz unpacks into subdirectory X*			
4DD0	4DD0_01_0001.osc		1	360	4DD0.tar.gz	1465	6191	Rigaku R_AXIS IV
4DD2	4DD2_01_0001.osc		1	360	4DD2.tar.gz	2657	6191	
4DD3	4DD3_01_0001.osc		1	360	4DD3.tar.gz	2293	6191	
4DD9	4DD9_01_0001.osc		1	360	4DD9.tar.gz	2716	6191	
4DDA	4DDA_01_0001.osc		1	180	4DDA.tar.gz	1036	3096	
4DDB	4DDB_01_0001.osc		1	360	4DDB.tar.gz	2249	6191	
4DD1	4DD1_01_0001.sfrm		1	614	4DD1.tar.gz 4DD1unwarp.tar.gz	480 554	630 1236	Bruker PLATINUM^{135} proteumnac.p4p
4DD4	4DD4_01_0001.sfrm		5	777	4DD4.tar.gz 4DD4unwarp.tar.gz	591 624	792 1564	
4DD6	4DD6_01_0001.sfrm		2	1440	4DD6.tar.gz 4DD6unwarp.tar.gz	1067 1117	1464 2897	
4DD7	4DD7_01_0001.sfrm		4	1862	4DD7.tar.gz 4DD7unwarp.tar.gz	1464 1566	1913 3746	
4DDC	4DDC_01_0001.sfrm		2	1440	4DDC.tar.gz 4DDCunwarp.tar.gz	1149 1220	1480 2897	

rawdata.chem.uu.nl

- Downloads by other groups: 150 GB was retrieved from rawdata.chem.uu.nl and some data were reprocessed
- Discussion with Kay Diederichs on Bruker Fixed CHI goniometer: see *Acta Cryst*. D**70** (2014) 2520 supplemental
- We welcome the interest in cisplatin binding to histidine studies by Shabalin *et al*., who reprocessed at least 1 data set

J. Appl. Cryst. link to raw data



Journal of Applied Crystallography
Volume 46, Part 1 (February 2013)

research papers

J. Appl. Cryst. (2013). **46**, 108-119 [doi:10.1107/S0021889812044172]

Experience with exchange and archiving of raw data: comparison of data from two diffractometers and four software packages on a series of lysozyme crystals

S. W. M. Tanley, A. M. M. Schreurs, J. R. Helliwell and L. M. J. Kroon-Batenburg

Abstract: The International Union of Crystallography has for many years been advocating archiving of raw data to accompany structural papers. Recently, it initiated the formation of the Diffraction Data Deposition Working Group with the aim of developing standards for the representation of these data. A means of studying this issue is to submit exemplar publications with associated raw data and metadata. A recent study on the effects of dimethyl sulfoxide on the binding of cisplatin and carboplatin to histidine in 11 different lysozyme crystals from two diffractometers led to an investigation of the possible effects of the equipment and X-ray diffraction data processing software on the calculated occupancies and *B* factors of the bound Pt compounds. 35.3 Gb of data were transferred from Manchester to Utrecht to be processed with *EVAL*. A systematic comparison shows that the largest differences in the occupancies and *B* factors of the bound Pt compounds are due to the software, but the equipment also has a noticeable effect. A detailed description of and discussion on the availability of metadata is given. By making these raw diffraction data sets available *via* a local depository, it is possible for the diffraction community to make their own evaluation as they may wish.

Keywords: data exchange; data archiving; metadata.

Link
Raw data: PDB code 4dd0 HEWL_cisplatin_aqueous_glycerol

Link
Raw data: PDB code 4dd1; HEWL_cisplatin_aqueous_paratone

HEWL_cisplatin_aq_glycerol: 4dd0

Tanley, Simon
[Research data] version 46. 2013. The University of Manchester.

Access to files

cisplatin_aq_glycerol_1-70.zip (zip)
cisplatin_aq_glycerol_71-140.zip (zip)
cisplatin_aq_glycerol_141-210.zip (zip)
cisplatin_aq_glycerol_211-280.zip (zip)
cisplatin_aq_glycerol_281-360.zip (zip)

Abstract

Abstract: The International Union of Crystallography has for many years been advocating archiving of raw data to accompany structural papers. Recently, it initiated the formation of the Diffraction Data Deposition Working Group with the aim of developing standards for the representation of these data. A means of studying this issue is to submit exemplar publications with associated raw data and metadata. A recent study on the effects of dimethyl sulfoxide on the binding of cisplatin and carboplatin to histidine in 11 different lysozyme crystals from two diffractometers led to an investigation of the possible effects of the equipment and X-ray diffraction data processing software on the calculated occupancies and B factors of the bound Pt compounds. 35.3 Gb of data were transferred from Manchester to Utrecht to be processed with EVAL. A systematic comparison shows that the largest differences in the occupancies and B factors of the bound Pt compounds are due to the software, but the equipment also has a noticeable effect. A detailed description of and discussion on the availability of metadata is given. By making these raw diffraction data sets available via a local depository, it is possible for the diffraction community to make their own evaluation as they may wish.

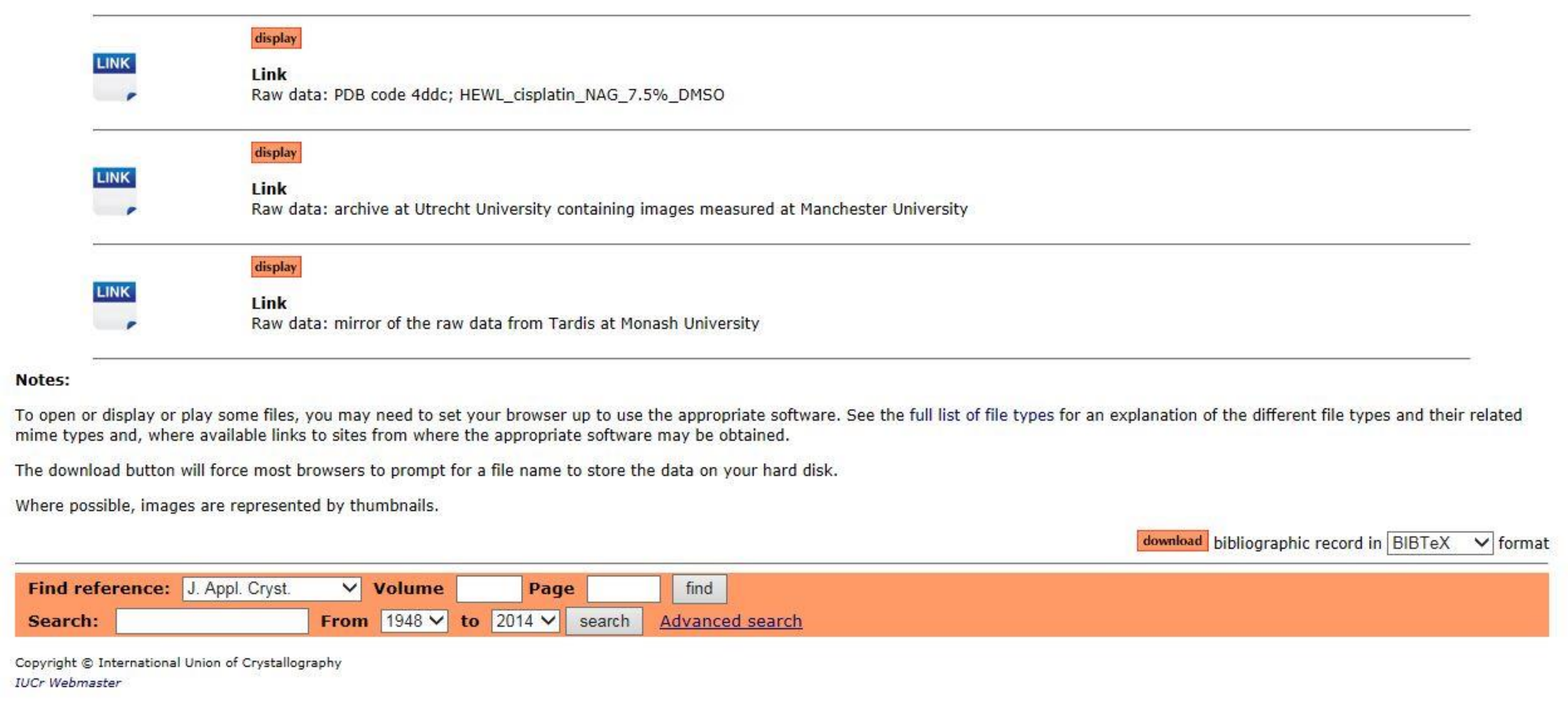
Related resources

Full-text held externally
DOI: 10.15127/1.219230
DOI: doi:10.1107/S0021889812044172

University researcher(s)

Academic department(s)
Faculty of Engineering and Physical Sciences' website
Faculty of Life Sciences' website
School of Chemistry 's website

J. Appl. Cryst. link to raw data



Link
Raw data: PDB code 4ddc; HEWL_cisplatin_NAG_7.5%_DMSO

Link
Raw data: archive at Utrecht University containing images measured at Manchester University

Link
Raw data: mirror of the raw data from Tardis at Monash University

Notes:

To open or display or play some files, you may need to set your browser up to use the appropriate software. See the full list of file types for an explanation of the different file types and their related mime types and, where available links to sites from where the appropriate software may be obtained.

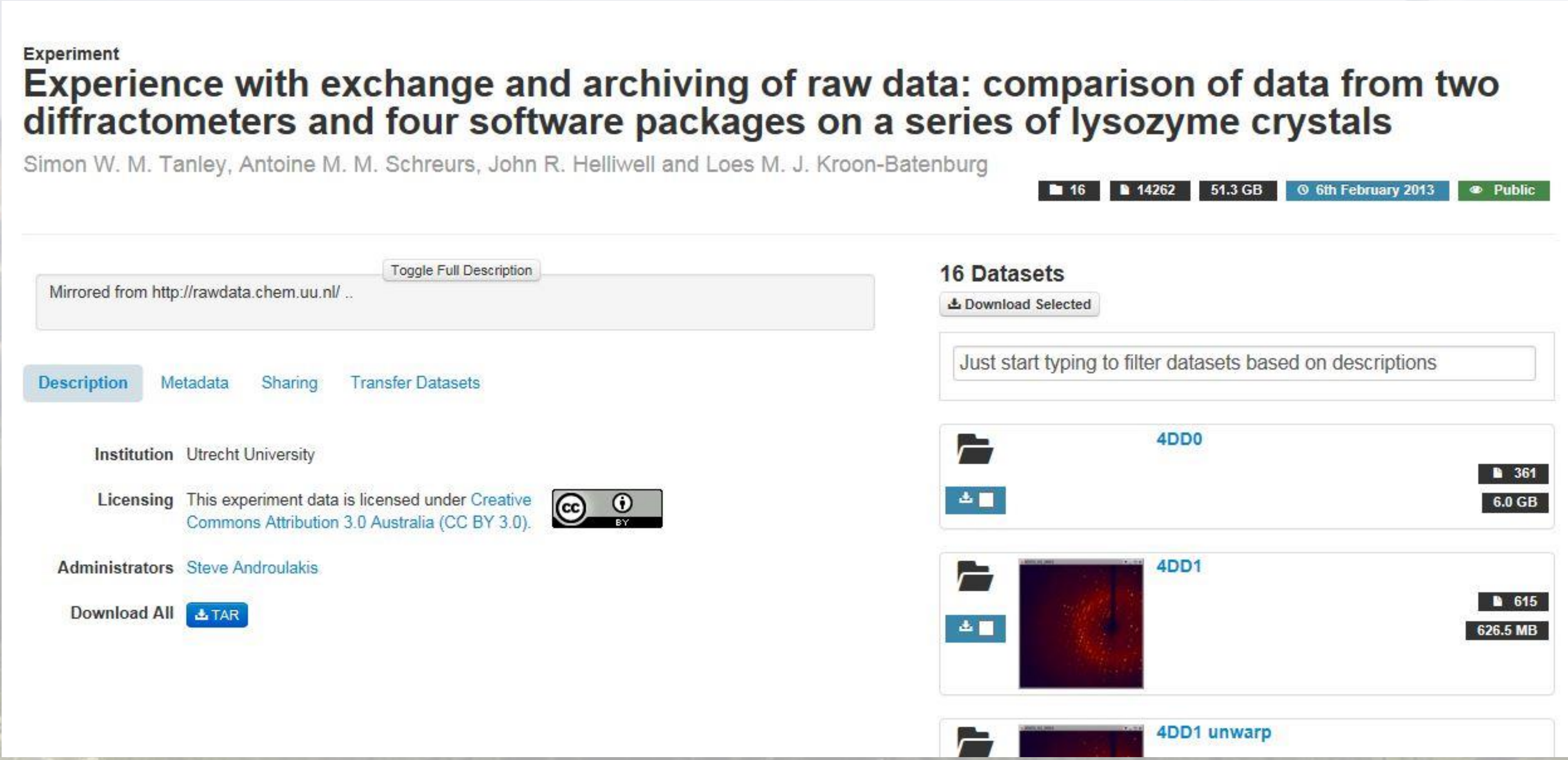
The download button will force most browsers to prompt for a file name to store the data on your hard disk.

Where possible, images are represented by thumbnails.

download bibliographic record in BIBTeX format

Find reference: J. Appl. Cryst. **Volume** **Page** find
Search: **From** 1948 **to** 2014 search Advanced search

Tardis



Experiment

Experience with exchange and archiving of raw data: comparison of data from two diffractometers and four software packages on a series of lysozyme crystals

Simon W. M. Tanley, Antoine M. M. Schreurs, John R. Helliwell and Loes M. J. Kroon-Batenburg

Mirrored from http://rawdata.chem.uu.nl/ ..

Institution	Utrecht University
Licensing	This experiment data is licensed under Creative Commons Attribution 3.0 Australia (CC BY 3.0).
Administrators	Steve Androulakis
Download All	TAR

16 Datasets

Dataset	Files	Size
4DD0	361	6.0 GB
4DD1	615	626.5 MB
4DD1 unwarp		

General purpose repositories

DRYAD

- Has integrated submission for a growing list of journals (hundreds of journals already)
- Data have their own DOI but have to be associated with some sort of publication
- Max. 10 GB (otherwise additional costs)
- Data publishing charges:
  - -sometimes covered by the journal
  - -80 $ per data package

figshare

- Cloud based
- Unlimited storage space for publicly available data
- Publish negative data
- All data has a DOI
- Papers under CC-By
- Data under CC0
- Figshare for publishers: host data for online papers

General purpose repositories

zenodo

- EU-funded OpenAIRE project:
- Uses the cloud infrastructure of CERN Large Hadron Collider
- Uploaded data get DOI
- All research outputs welcome!
- Files up to 3 GB
- Searching for “diffraction data” gave 15 hits of which 6 protein data sets

The Dataverse Network™ Project

- A repository for research data that: takes care of long term preservation
- No DOI’s
- Up to 1TB (max 2 GB per file)
- Works together with Dryad

Diffraction data repositories

- DIMER: University of Queensland Diffraction Image Experiment Repository (37)
- JCSG: Joint Center for Structural Genomics (1720)
- TARDIS/Store.Synchrotron (31 public data)

Meta data

- Usually limited amount of listed meta data, e.g. DIMER repository:
- Publication
- PDB entry

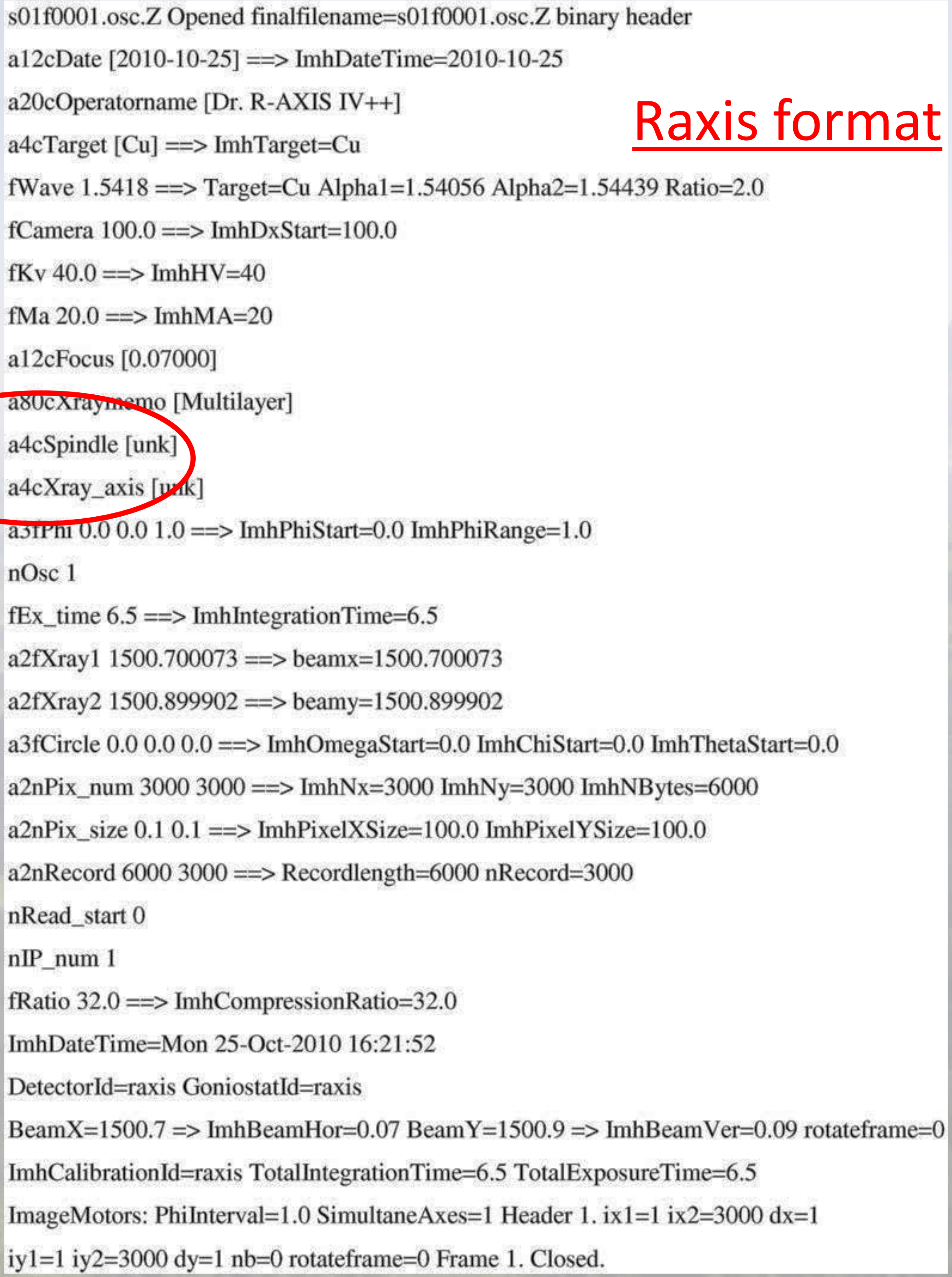
Metadata

Collection date	07 Nov 2012
Exposure time	1.000 s
Oscillation range	1.000° to 2.000°
Two theta	-0.000°
Detector distance	330.010 mm
Beam centre	158.453 mm, 158.030 mm
X-ray wavelength	0.979 Å
Image size	3072 px, 3072 px
Pixel size	0.103 mm, 0.103 mm
Image format	SMV

Experience

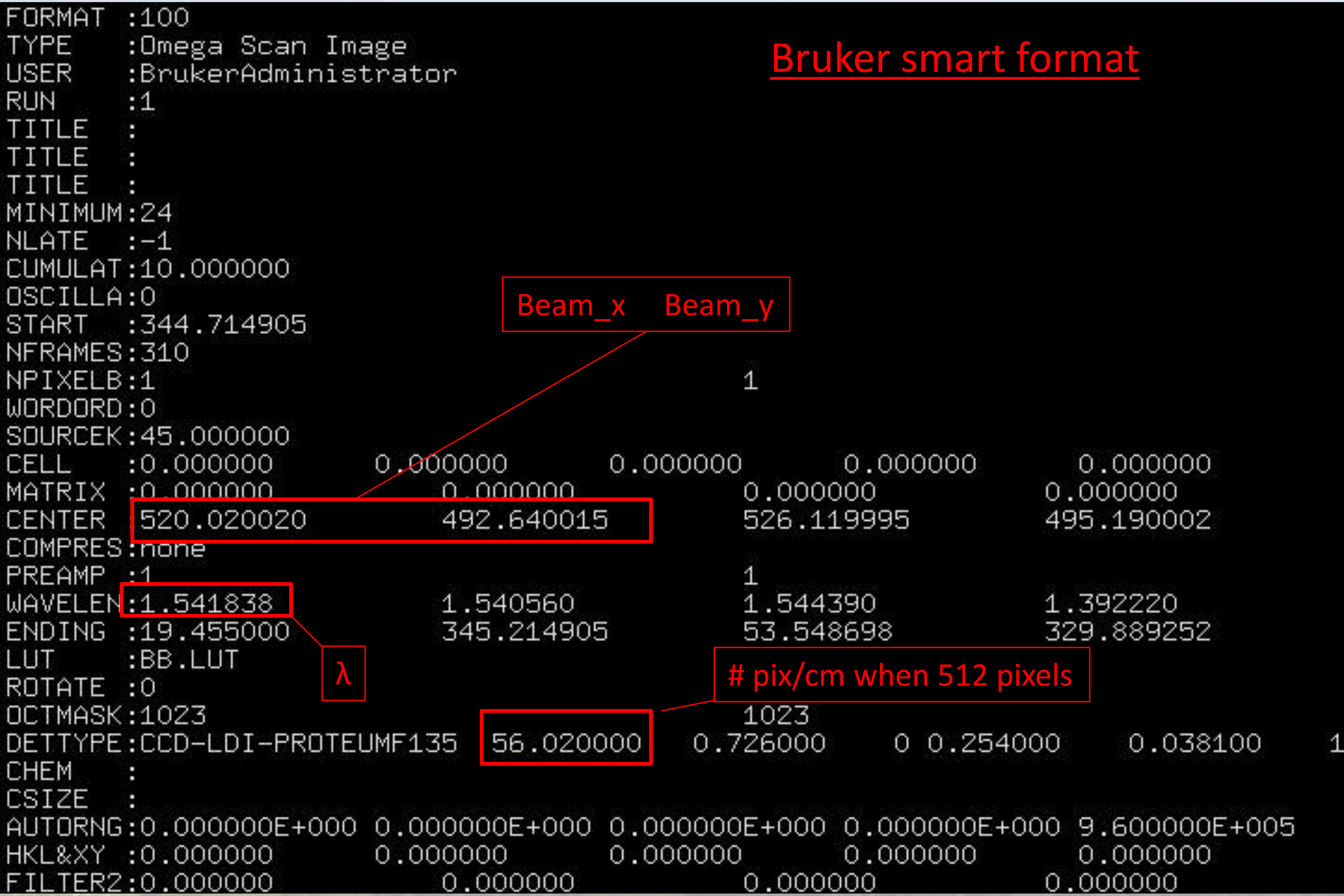
- What is our experience with the level of meta data ?

Raxis format



```
s01f0001.osc.Z Opened finalfilename=s01f0001.osc.Z binary header
a12cDate [2010-10-25] ==> ImhDateTime=2010-10-25
a20cOperatorname [Dr. R-AXIS IV++]
a4cTarget [Cu] ==> ImhTarget=Cu
fWave 1.5418 ==> Target=Cu Alpha1=1.54056 Alpha2=1.54439 Ratio=2.0
fCamera 100.0 ==> ImhDxStart=100.0
fKv 40.0 ==> ImhHV=40
fMa 20.0 ==> ImhMA=20
a12cFocus [0.07000]
a80cXraymemo [Multilayer]
a4cSpindle [unk]
a4cXray_axis [unk]
a3fPhi 0.0 0.0 1.0 ==> ImhPhiStart=0.0 ImhPhiRange=1.0
nOsc 1
fEx_time 6.5 ==> ImhIntegrationTime=6.5
a2fXray1 1500.700073 ==> beamx=1500.700073
a2fXray2 1500.899902 ==> beamy=1500.899902
a3fCircle 0.0 0.0 0.0 ==> ImhOmegaStart=0.0 ImhChiStart=0.0 ImhThetaStart=0.0
a2nPix_num 3000 3000 ==> ImhNx=3000 ImhNy=3000 ImhNBytes=6000
a2nPix_size 0.1 0.1 ==> ImhPixelXSize=100.0 ImhPixelYSize=100.0
a2nRecord 6000 3000 ==> Recordlength=6000 nRecord=3000
nRead_start 0
nIP_num 1
fRatio 32.0 ==> ImhCompressionRatio=32.0
ImhDateTime=Mon 25-Oct-2010 16:21:52
DetectorId=raxis GoniostatId=raxis
BeamX=1500.7 => ImhBeamHor=0.07 BeamY=1500.9 => ImhBeamVer=0.09 rotateframe=0
ImhCalibrationId=raxis TotalIntegrationTime=6.5 TotalExposureTime=6.5
ImageMotors: PhiInterval=1.0 SimultaneAxes=1 Header 1. ix1=1 ix2=3000 dx=1
iy1=1 iy2=3000 dy=1 nb=0 rotateframe=0 Frame 1. Closed.
```

Mostly manufacturers provide detailed description of the image format



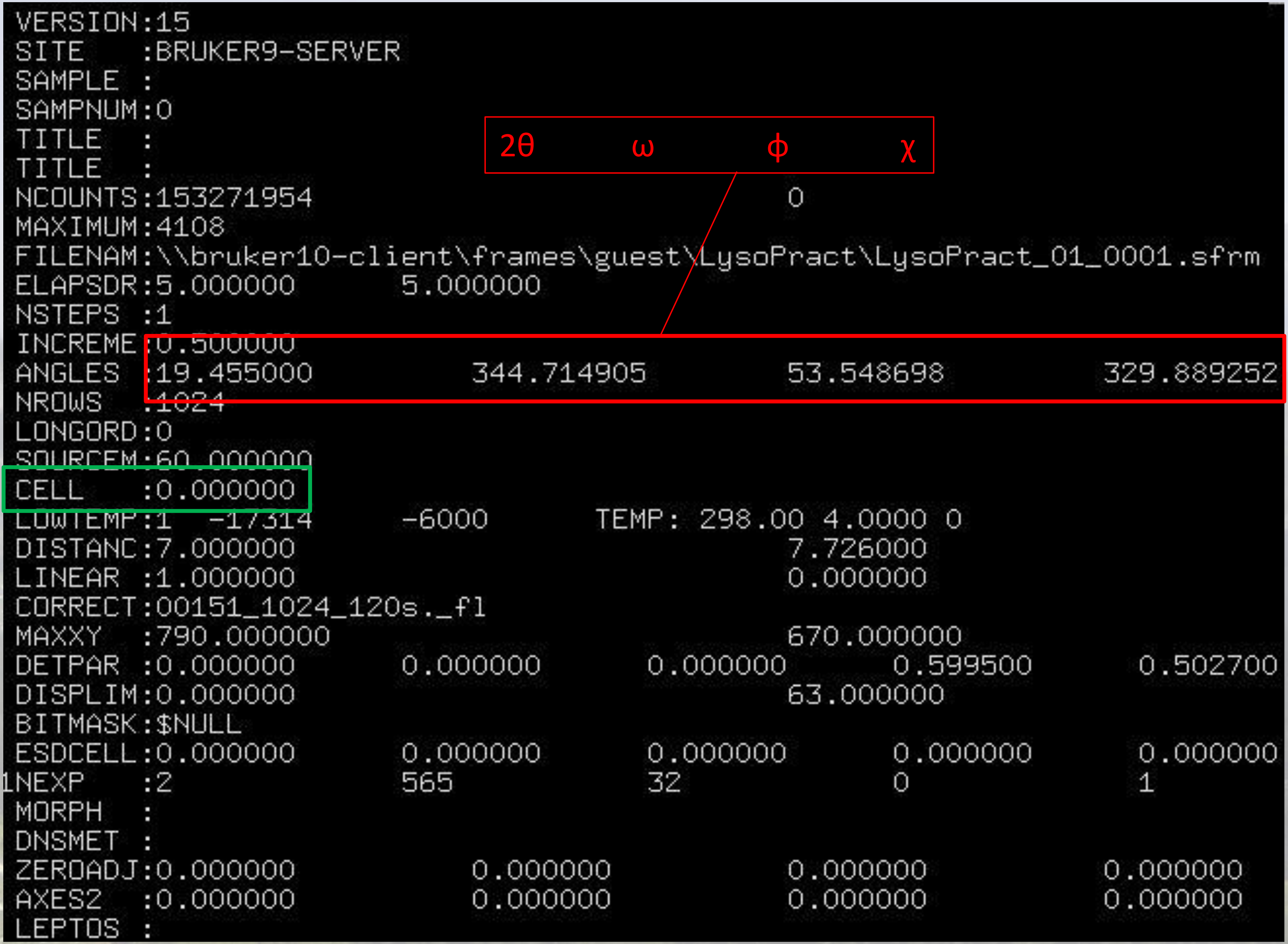
Bruker smart format

```
FORMAT  :100
TYPE    :Omega Scan Image
USER    :BrukerAdministrator
RUN     :1
TITLE   :
TITLE   :
TITLE   :
MINIMUM:24
NLATE   :-1
CUMULAT:10.000000
OSCILLA:0
START   :344.714905
NFRAMES:310
NPIXELB:1                                   1
WORDORD:0
SOURCEK:45.000000
CELL    :0.000000      0.000000      0.000000      0.000000      0.000000
MATRIX  :0.000000          0.000000          0.000000          0.000000
CENTER  :520.020020        492.640015        526.119995        495.190002
COMPRES:none
PREAMP  :1                                  1
WAVELEN:1.541838          1.540560          1.544390          1.392220
ENDING  :19.455000         345.214905        53.548698         329.889252
LUT     :BB.LUT
ROTATE  :0
OCTMASK:1023                                1023
DETTYPE:CCD-LDI-PROTEUMF135  56.020000   0.726000    0 0.254000    0.038100    1
CHEM    :
CSIZE   :
AUTORNG:0.000000E+000 0.000000E+000 0.000000E+000 0.000000E+000 9.600000E+005
HKL&XY  :0.000000      0.000000      0.000000      0.000000      0.000000
FILTER2:0.000000          0.000000          0.000000          0.000000
```

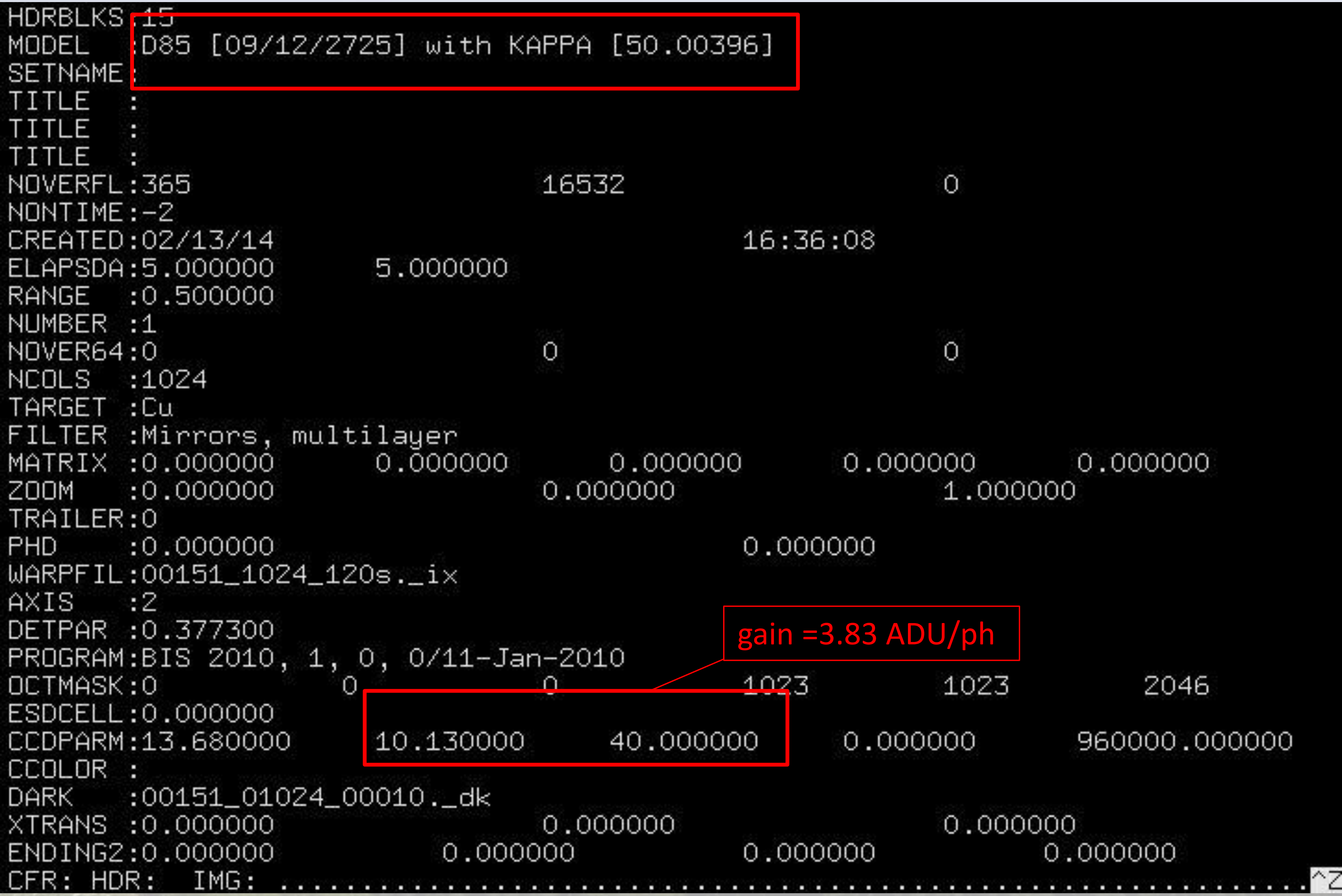
Beam_x Beam_y

λ

pix/cm when 512 pixels



```
 VERSION:15
 SITE   :BRUKER9-SERVER
 SAMPLE :
 SAMPNUM:0
 TITLE  :
 TITLE  :
 NCOUNTS:153271954                           0
 MAXIMUM:4108
 FILENAM:\\bruker10-client\frames\guest\LysoPract\LysoPract_01_0001.sfrm
 ELAPSDR:5.000000      5.000000
 NSTEPS :1
 INCREME:0.500000
 ANGLES :19.455000         344.714905        53.548698         329.889252
 NROWS  :1024
 LONGORD:0
 SOURCEM:60.000000
 CELL   :0.000000
 LOWTEMP:1  -17314     -6000      TEMP: 298.00 4.0000 0
 DISTANC:7.000000                            7.726000
 LINEAR :1.000000                            0.000000
 CORRECT:00151_1024_120s._fl
 MAXXY  :790.000000                          670.000000
 DETPAR :0.000000      0.000000      0.000000      0.599500      0.502700
 DISPLIM:0.000000                            63.000000
 BITMASK:$NULL
 ESDCELL:0.000000      0.000000      0.000000      0.000000      0.000000
1NEXP   :2             565           32            0             1
 MORPH  :
 DNSMET :
 ZEROADJ:0.000000          0.000000          0.000000          0.000000
 AXES2  :0.000000          0.000000          0.000000          0.000000
 LEPTOS :
```



```
HDRBLKS:15
MODEL  :D85 [09/12/2725] with KAPPA [50.00396]
SETNAME:
TITLE  :
TITLE  :
TITLE  :
NOVERFL:365                     16532                   0
NONTIME:-2
CREATED:02/13/14                            16:36:08
ELAPSDA:5.000000      5.000000
RANGE  :0.500000
NUMBER :1
NOVER64:0                       0                       0
NCOLS  :1024
TARGET :Cu
FILTER :Mirrors, multilayer
MATRIX :0.000000      0.000000      0.000000      0.000000      0.000000
ZOOM   :0.000000                0.000000                1.000000
TRAILER:0
PHD    :0.000000                            0.000000
WARPFIL:00151_1024_120s._ix
AXIS   :2
DETPAR :0.377300
PROGRAM:BIS 2010, 1, 0, 0/11-Jan-2010
OCTMASK:0           0           0           1023        1023        2046
ESDCELL:0.000000
CCDPARM:13.680000     10.130000     40.000000     0.000000      960000.000000
CCOLOR :
DARK   :00151_01024_00010._dk
XTRANS :0.000000                0.000000                0.000000
ENDING2:0.000000          0.000000          0.000000          0.000000
CFR: HDR:  IMG: ..............................................................^Z
```

gain =3.83 ADU/ph

Standard deviations of integrated reflection intensities

- Realistic standard deviations:

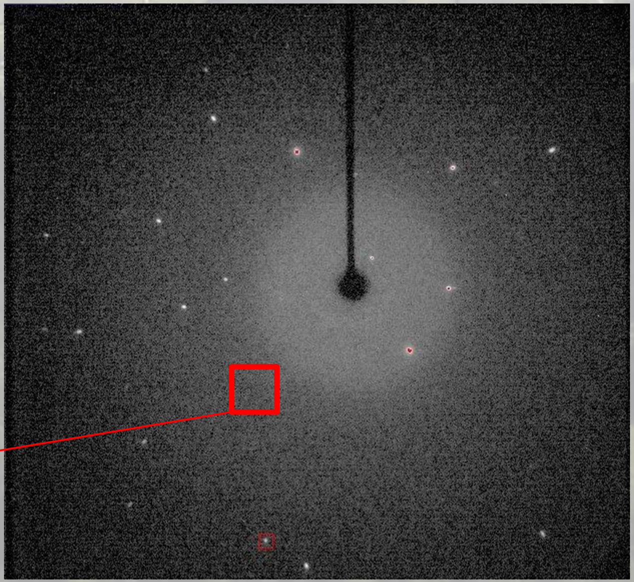
$\sigma_c = K\,[\sigma_I^2+(g\langle I\rangle)^2]^{1/2}$ $\sigma_I = \sqrt{I}$ Poisson

K and g determined from χ^2

typically: $K \approx 1$ and $g \approx 0.02\text{-}0.04$ if gain is correct

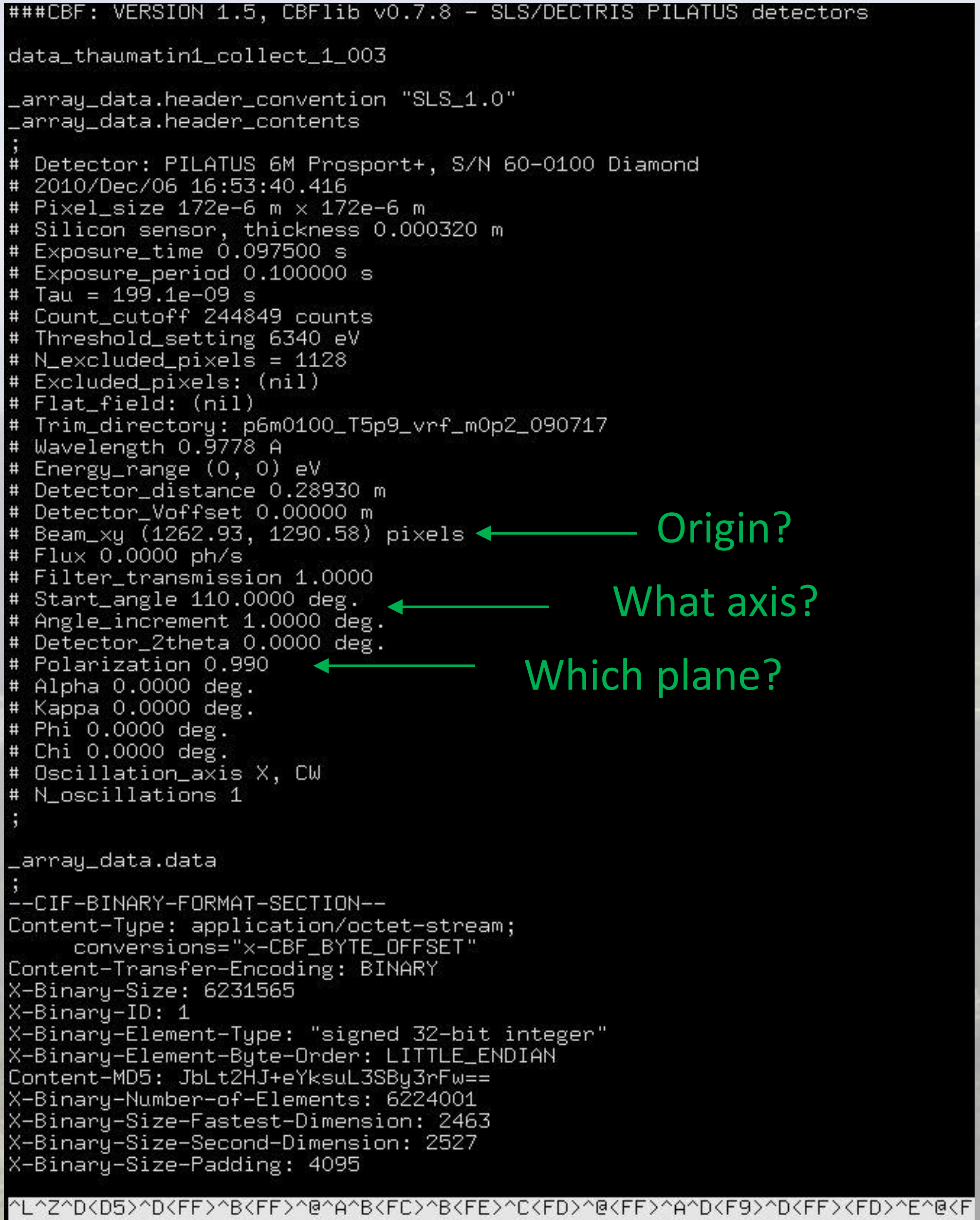
- Alternatively from:

$\sigma_{int}=[\Sigma_i(I_i-\langle I\rangle)^2/(N-1)]^{1/2}$



$\sigma^2 / \langle I \rangle$

Pilatus 6M header



```
###CBF: VERSION 1.5, CBFlib v0.7.8 - SLS/DECTRIS PILATUS detectors

data_thaumatin1_collect_1_003

_array_data.header_convention "SLS_1.0"
_array_data.header_contents
;
# Detector: PILATUS 6M Prosport+, S/N 60-0100 Diamond
# 2010/Dec/06 16:53:40.416
# Pixel_size 172e-6 m x 172e-6 m
# Silicon sensor, thickness 0.000320 m
# Exposure_time 0.097500 s
# Exposure_period 0.100000 s
# Tau = 199.1e-09 s
# Count_cutoff 244849 counts
# Threshold_setting 6340 eV
# N_excluded_pixels = 1128
# Excluded_pixels: (nil)
# Flat_field: (nil)
# Trim_directory: p6m0100_T5p9_vrf_m0p2_090717
# Wavelength 0.9778 A
# Energy_range (0, 0) eV
# Detector_distance 0.28930 m
# Detector_Voffset 0.00000 m
# Beam_xy (1262.93, 1290.58) pixels
# Flux 0.0000 ph/s
# Filter_transmission 1.0000
# Start_angle 110.0000 deg.
# Angle_increment 1.0000 deg.
# Detector_2theta 0.0000 deg.
# Polarization 0.990
# Alpha 0.0000 deg.
# Kappa 0.0000 deg.
# Phi 0.0000 deg.
# Chi 0.0000 deg.
# Oscillation_axis X, CW
# N_oscillations 1
;

_array_data.data
;
--CIF-BINARY-FORMAT-SECTION--
Content-Type: application/octet-stream;
     conversions="x-CBF_BYTE_OFFSET"
Content-Transfer-Encoding: BINARY
X-Binary-Size: 6231565
X-Binary-ID: 1
X-Binary-Element-Type: "signed 32-bit integer"
X-Binary-Element-Byte-Order: LITTLE_ENDIAN
Content-MD5: JbLt2HJ+eYksuL3SBy3rFw==
X-Binary-Number-of-Elements: 6224001
X-Binary-Size-Fastest-Dimension: 2463
X-Binary-Size-Second-Dimension: 2527
X-Binary-Size-Padding: 4095

^L^Z^D<D5>^D<FF>^B<FF>^@^A^B<FC>^B<FE>^C<FD>^@<FF>^A^D<F9>^D<FF><FD>^E^@<F
```

Origin?

What axis?

Which plane?

Knowledge is required about experimental set-up

EVAL: horax goniometer

Meta data for raw diffraction data

- Header contains usually limited information: only in the context of the equipment/beam line
- Goniometer set-up: axes orientations and rotation directions; zero positions; type of scan (continuous rotation, still, helical scan); oscillation range

kappa axes: omega=z kappa=k phi=z swing=z
rotationdirection -1 -1 -1 -1
values: omega kappa phi swing dist alpha

euler axes: omega=z chi=x phi=z swing=z
rotationdirection 1 1 1 1
values: omega chi phi swing dist

horax axes: omega=y chi=x phi=z swing=y
rotationdirection 1 1 1 1
values: omega chi phi swing dist

dtb axes: omega=z chi=-x phi=z swing=y
rotationdirection -1 -1 -1 1
values: omega chi phi swing dist

x8 axes: omega=z chi=x phi=z swing=z
direction 1 -1 -1 1
values: omega+180 chi phi+90 swing

x8c axes: omega=z chi=x phi=z swing=z
direction 1 -1 -1 1
values: omega+180 chi phi+90 swing

raxis axes: omega=z chi=x phi=z swing=z
direction -1 1 -1 1
values: omega chi phi swing

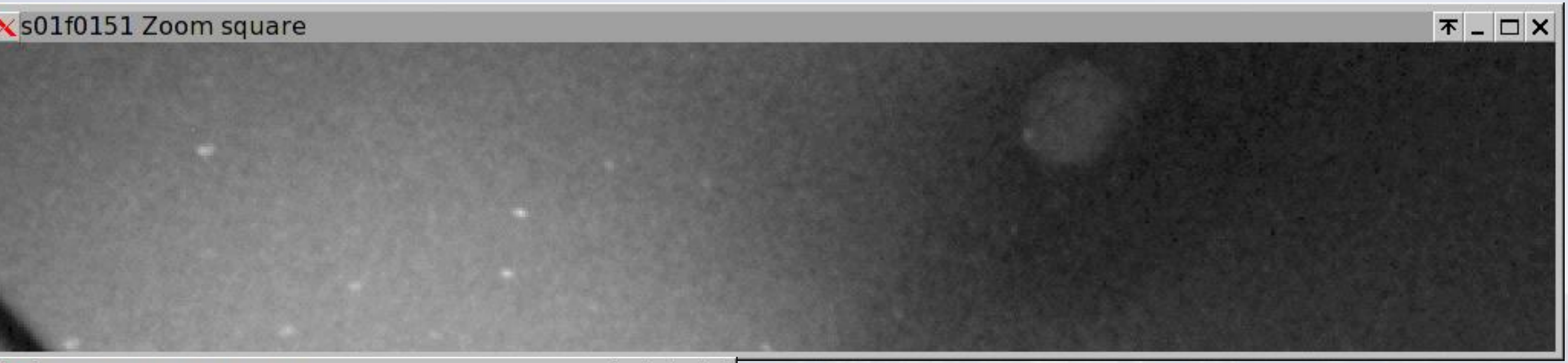
kappa180 axes: omega=z kappa=k phi=z swing=z
direction: -1 -1 -1 -1
values: omega+180 kappa phi swing

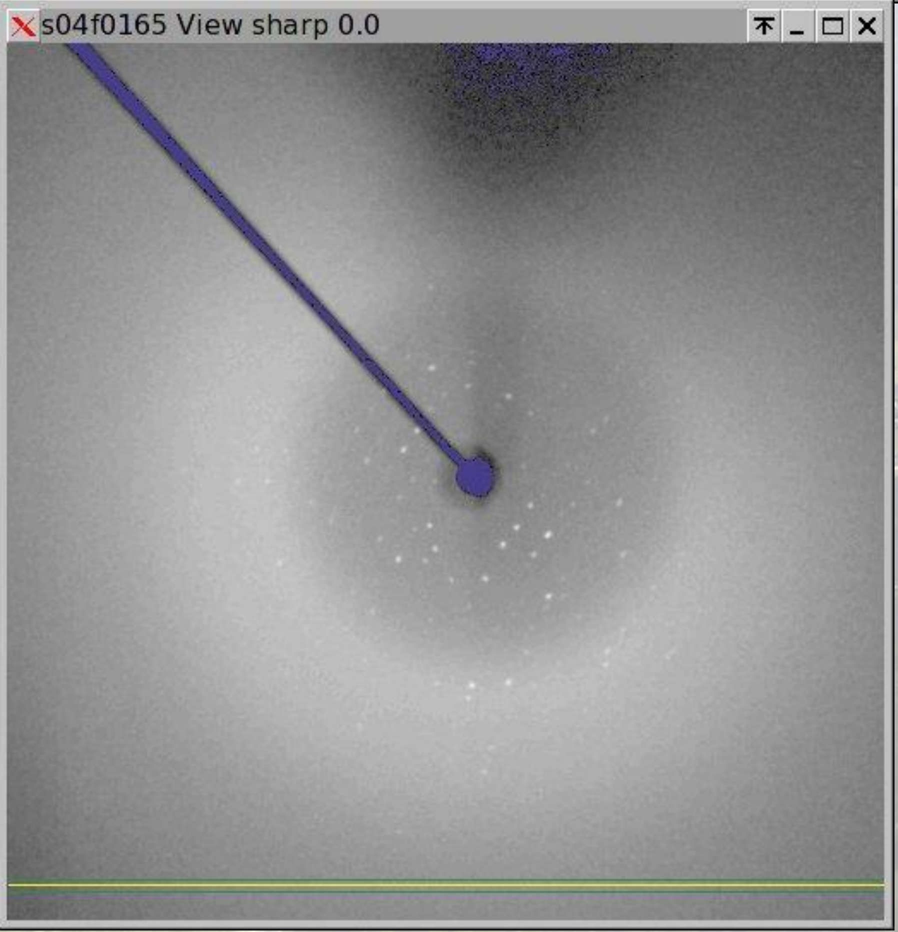
..continued..

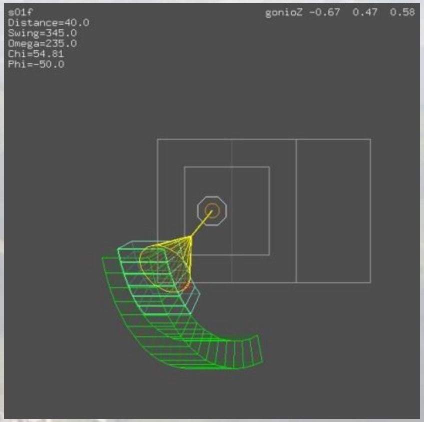
- Detector:
  - – Type and serial number
  - – Data format: byte-storage architecture
  - – Overflow handling; base line offset; gain
  - – Swing axis; Distance
  - – Fast, slow running pixel coordinates
  - – Origin: LL, LR, UR, UL
  - – Number of pixels, pixel size, binning
  - – Corrections applied:
    - Dark curent
    - Distortion table: usually manufacturer applied
    - Non-uniformity: flood field

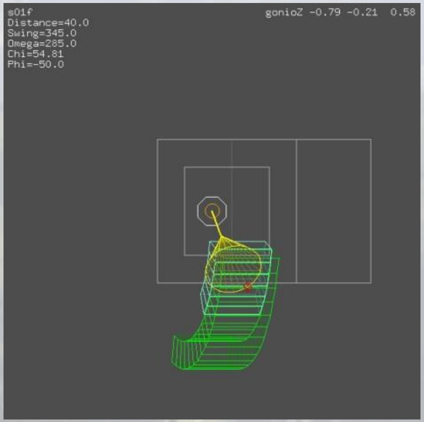
..continued..

- Beam properties: beam_x,y; divergence_x,y; polarization
- Photon flux, exposure time; number of repeats;dose; wavelength; date
- crystal shape/size; chemical composition; heavy-atom derivatives; multiple wavelength experiments on one crystal;
- Sample mounting
- Sample name/characterization
- Type of diffraction: single crystal, twin, diffuse...






```
s01f                                   gonioZ -0.67  0.47  0.58
Distance=40.0
Swing=345.0
Omega=235.0
Chi=54.81
Phi=-50.0
```


```
s01f                                   gonioZ -0.79 -0.21  0.58
Distance=40.0
Swing=345.0
Omega=285.0
Chi=54.81
Phi=-50.0
```

Crystal mounted in open capillary
CuKα radiation
Shadow and reflections sneaking through

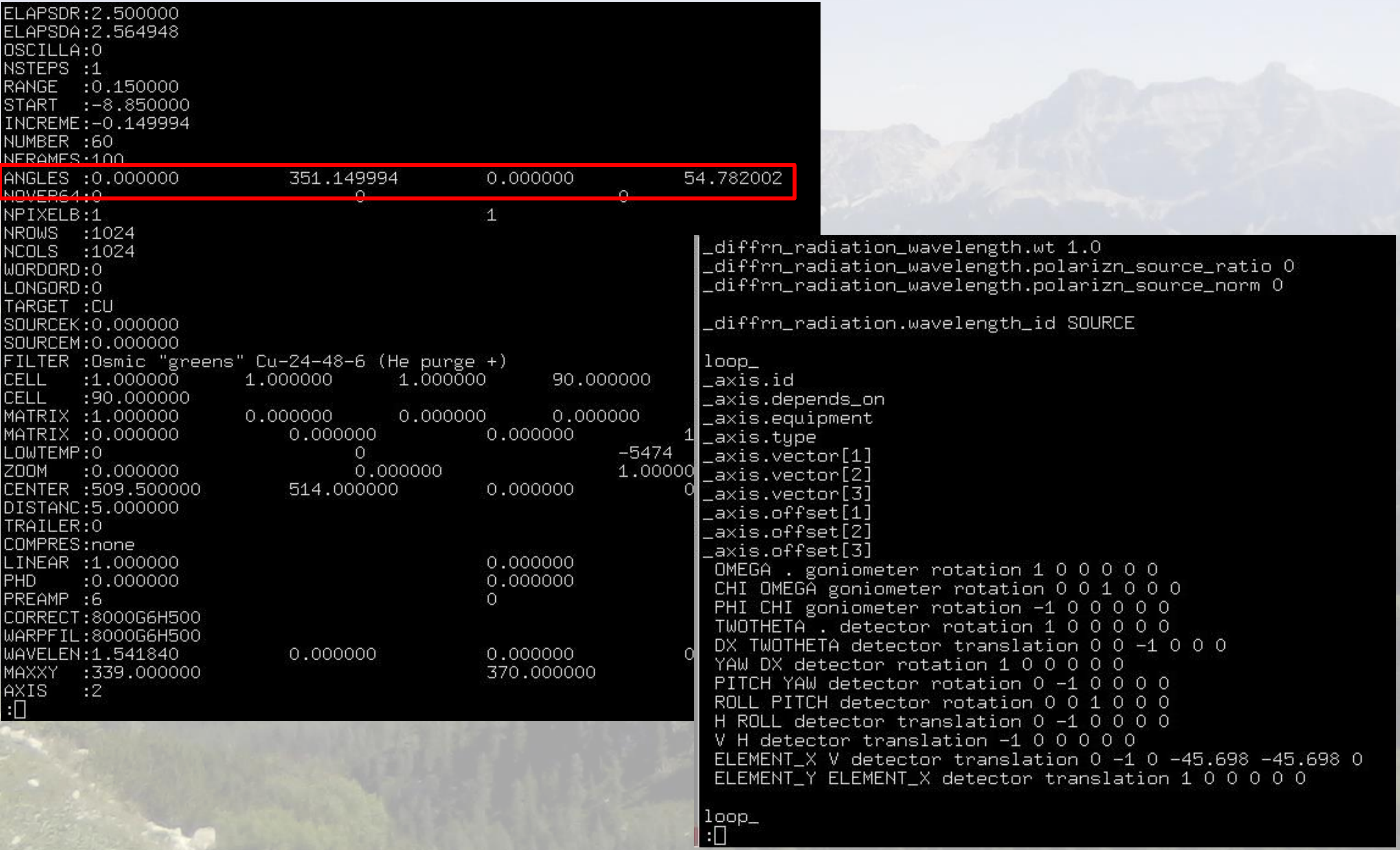
Photo of set-up is useful

Tanley, Schreurs, Kroon-Batenburg, Helliwell, *Acta. Cryst*. F**68** (2012), 1300
Link to video

Meta data

- Machine and human readable
- In imgCIF format?
- In HDF5-Nexus trees?
- Photo of set-up is not useful for high-throughput

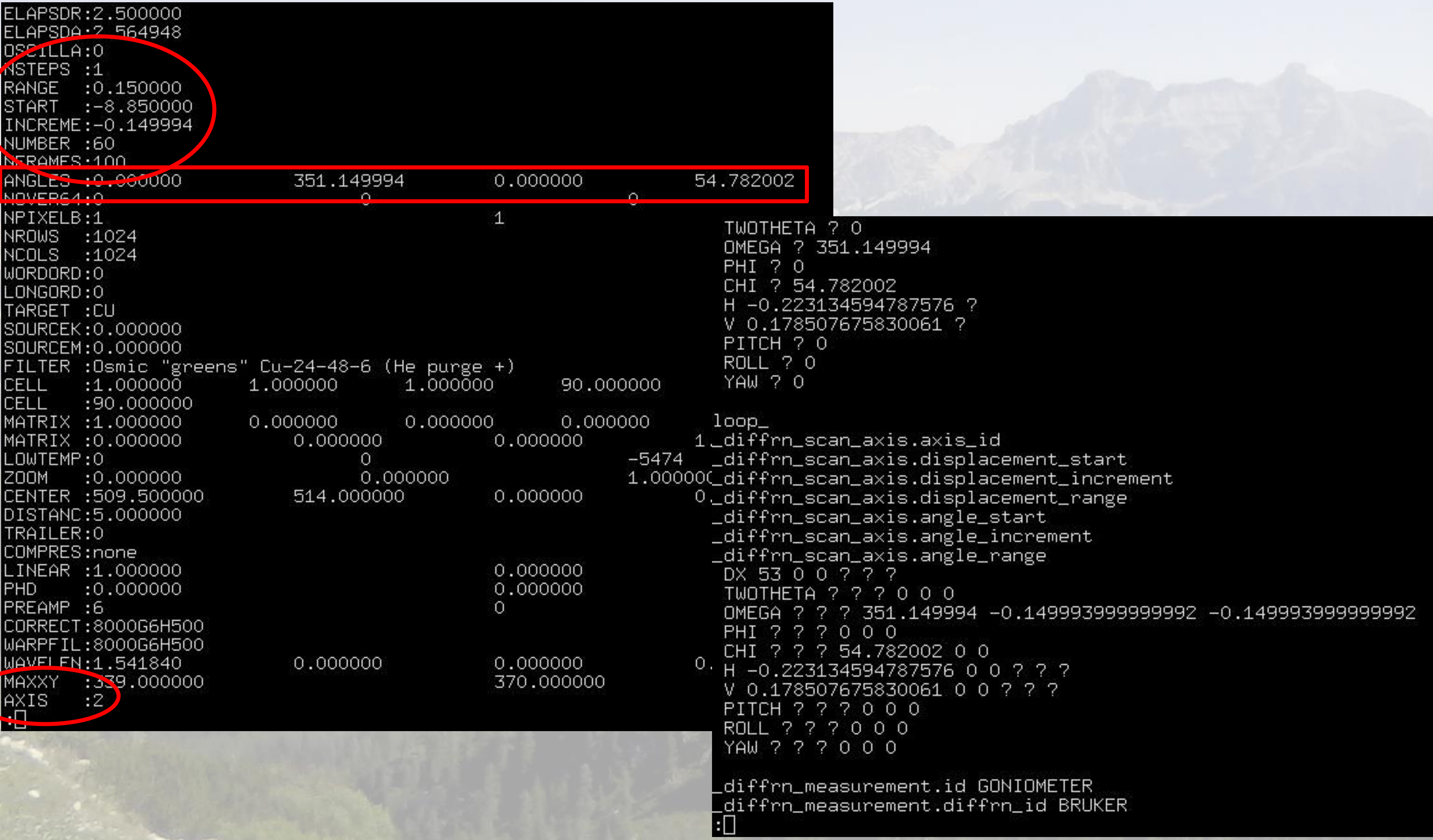
Comparison of .sfrm and .cbf



```
ELAPSDR:2.500000
ELAPSDA:2.564948
OSCILLA:0
NSTEPS  :1
RANGE   :0.150000
START   :-8.850000
INCREME:-0.149994
NUMBER  :60
NFRAMES:100
ANGLES  :0.000000          351.149994        0.000000          54.782002
NOVER64:0                       0                       0
NPIXELB:1                                   1
NROWS   :1024
NCOLS   :1024
WORDORD:0
LONGORD:0
TARGET  :CU
SOURCEK:0.000000
SOURCEM:0.000000
FILTER  :Osmic "greens" Cu-24-48-6 (He purge +)
CELL    :1.000000     1.000000      1.000000      90.000000
CELL    :90.000000
MATRIX  :1.000000     0.000000      0.000000      0.000000
MATRIX  :0.000000         0.000000          0.000000          1
LOWTEMP:0                       0                       -5474
ZOOM    :0.000000               0.000000                1.00000
CENTER  :509.500000       514.000000        0.000000          0
DISTANC:5.000000
TRAILER:0
COMPRES:none
LINEAR  :1.000000                           0.000000
PHD     :0.000000                           0.000000
PREAMP  :6                                  0
CORRECT:8000G6H500
WARPFIL:8000G6H500
WAVELEN:1.541840          0.000000          0.000000          0
MAXXY   :339.000000                         370.000000
AXIS    :2
:
```

```
_diffrn_radiation_wavelength.wt 1.0
_diffrn_radiation_wavelength.polarizn_source_ratio 0
_diffrn_radiation_wavelength.polarizn_source_norm 0

_diffrn_radiation.wavelength_id SOURCE

loop_
_axis.id
_axis.depends_on
_axis.equipment
_axis.type
_axis.vector[1]
_axis.vector[2]
_axis.vector[3]
_axis.offset[1]
_axis.offset[2]
_axis.offset[3]
 OMEGA . goniometer rotation 1 0 0 0 0 0
 CHI OMEGA goniometer rotation 0 0 1 0 0 0
 PHI CHI goniometer rotation -1 0 0 0 0 0
 TWOTHETA . detector rotation 1 0 0 0 0 0
 DX TWOTHETA detector translation 0 0 -1 0 0 0
 YAW DX detector rotation 1 0 0 0 0 0
 PITCH YAW detector rotation 0 -1 0 0 0 0
 ROLL PITCH detector rotation 0 0 1 0 0 0
 H ROLL detector translation 0 -1 0 0 0 0
 V H detector translation -1 0 0 0 0 0
 ELEMENT_X V detector translation 0 -1 0 -45.698 -45.698 0
 ELEMENT_Y ELEMENT_X detector translation 1 0 0 0 0 0

loop_
:
```

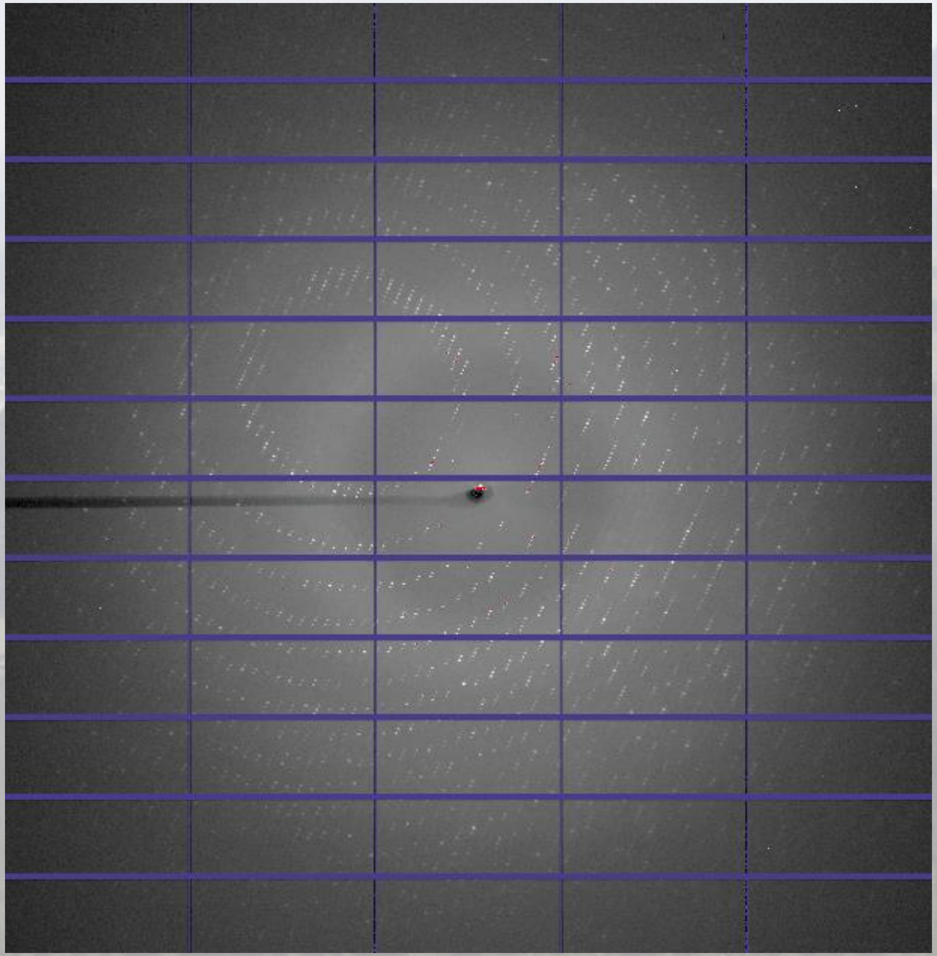
Comparison of .sfrm and .cbf



```
ELAPSDR:2.500000
ELAPSDA:2.564948
OSCILLA:0
NSTEPS :1
RANGE  :0.150000
START  :-8.850000
INCREME:-0.149994
NUMBER :60
NFRAMES:100
ANGLES :0.000000          351.149994         0.000000          54.782002
NOVER64:0                       0                       0
NPIXELB:1                                    1
NROWS  :1024
NCOLS  :1024
WORDORD:0
LONGORD:0
TARGET :CU
SOURCEK:0.000000
SOURCEM:0.000000
FILTER :Osmic "greens" Cu-24-48-6 (He purge +)
CELL   :1.000000      1.000000      1.000000      90.000000
CELL   :90.000000
MATRIX :1.000000      0.000000      0.000000      0.000000
MATRIX :0.000000          0.000000          0.000000          1.
LOWTEMP:0                     0                           -5474
ZOOM   :0.000000              0.000000                    1.00000
CENTER :509.500000        514.000000        0.000000          0.
DISTANC:5.000000
TRAILER:0
COMPRES:none
LINEAR :1.000000                            0.000000
PHD    :0.000000                            0.000000
PREAMP :6                                   0
CORRECT:8000G6H500
WARPFIL:8000G6H500
WAVELEN:1.541840          0.000000          0.000000          0.
MAXXY  :339.000000                          370.000000
AXIS   :2
:
```

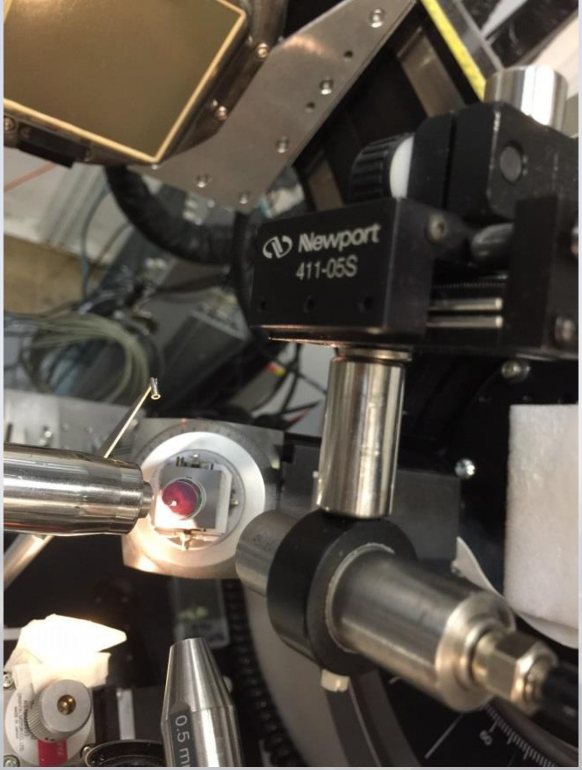
```
  TWOTHETA ? 0
  OMEGA ? 351.149994
  PHI ? 0
  CHI ? 54.782002
  H -0.223134594787576 ?
  V 0.178507675830061 ?
  PITCH ? 0
  ROLL ? 0
  YAW ? 0

 loop_
 _diffrn_scan_axis.axis_id
 _diffrn_scan_axis.displacement_start
 _diffrn_scan_axis.displacement_increment
 _diffrn_scan_axis.displacement_range
 _diffrn_scan_axis.angle_start
 _diffrn_scan_axis.angle_increment
 _diffrn_scan_axis.angle_range
  DX 53 0 0 ? ? ?
  TWOTHETA ? ? ? 0 0 0
  OMEGA ? ? ? 351.149994 -0.149993999999992 -0.149993999999992
  PHI ? ? ? 0 0 0
  CHI ? ? ? 54.782002 0 0
  H -0.223134594787576 0 0 ? ? ?
  V 0.178507675830061 0 0 ? ? ?
  PITCH ? ? ? 0 0 0
  ROLL ? ? ? 0 0 0
  YAW ? ? ? 0 0 0

_diffrn_measurement.id GONIOMETER
_diffrn_measurement.diffrn_id BRUKER
:
```

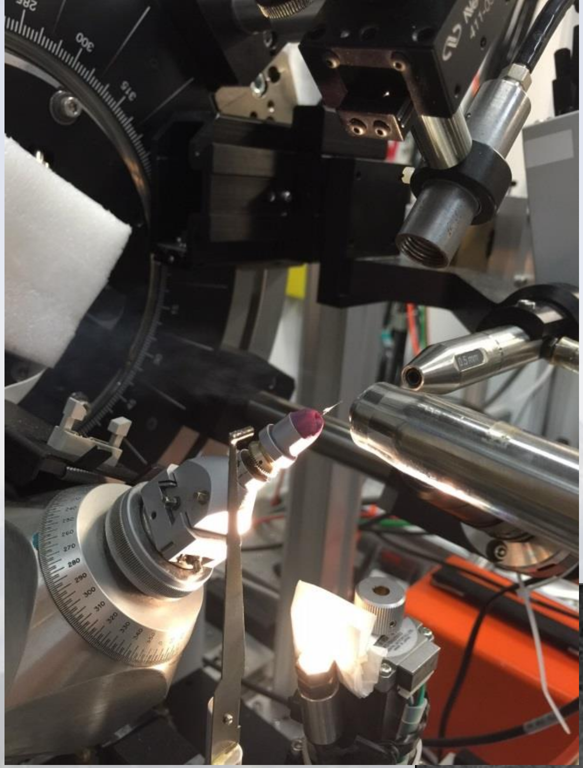
Pilatus image

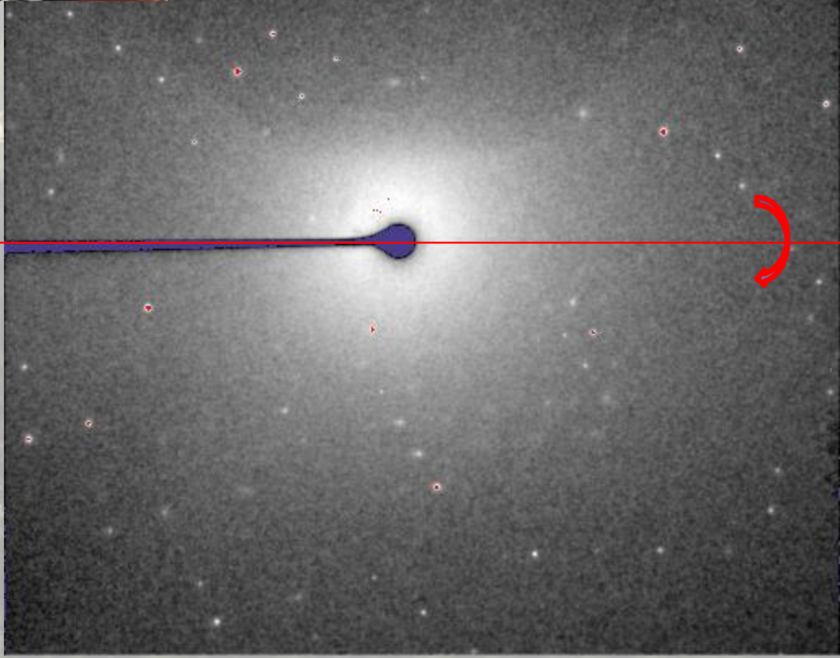


-1=strip

ADSC => 0







Argonne National Laboratory
Beamine 15ID-B

Synchrotron polarization:
weak component in horizontal plane

Images were converted to standard Bruker
Fixed CHI goniometer orientation

$$p = 1 - [I_{0w} \cos^2 \gamma \sin^2 2\theta + I_{0s} \sin^2 \gamma \sin^2 2\theta]$$

Data: Jozef Kozisek

Conclusions

- Long term data preservation and future use requires detailed and strict meta data policies
- Machine and human readable meta data format
- Meta data should be discoverable/searchable