1 Introduction

This document describes the use cases and API's for the program Mosflm. Mosflm is an unusual case, because there are number of significant functions it performs - each of these will have to be a well identified use case.

1.1 Functionality

The program mosflm has the following key pieces of functionality which need to be exposed *via* the API:

- Autoindexing of diffraction patterns.
- Estimation of mosaic spread.
- Refinement of unit cell.
- Integration of diffraction patterns.
- Calculation of data collection strategies (deprecated.)

1.2 Input

Each of these tasks requires a certain amount of common input - it can only make sense to standardise on the API for this. Of particular importance are the content of the image header and the frames to use in processing.

2 Rules

2.1 Introduction

There are a couple of rules which are useful when running mosflm, and these should be encoded in the *wrapper* rather than the rest of the expert system because they are specific to this program.

These rules pertain to:

- Indexing errors, in particular lattice assignment.

2.2 Indexing Errors & Lattice Assignment

Prototype data set: 1VR9 native (12847) from the JCSG. This is correctly C2 symmetry, but is pseudo-I222.

Autoindexing with mosflm, even with the "refine" keyword assigned, gives a lattice of I222. However, when this is set the mosaic spread estimation & refinement fails, which is a pointer that there is something wrong. The estimation fails saying "you'll have to guess" or something, and is then refined to a negative value in cell refinement. This should be interpreted as a pointer that the lattice has too high a symmetry. Asserting a spacegroup of C2 corrects this, so clearly this is something important to build into the wrapper.

3 Making Sure Nothing is Lost

FIXED 16/AUG/06 the distortion & raster parameters decided on in the cell refinement stages need to be recycled to use in integration. This is demonstrated by running an interactive (through the GUI) mosflm autoindex and refine job, then dumping the runit script. The important information is in the following records:

```
 Final optimised raster parameters:   15   17   12    5    6
   => RASTER keyword
 Separation parameters updated to   0.71mm in X and  0.71mm in Y
   => SEPARATION keyword
   XCEN    YCEN  XTOFRA   XTOFD  YSCALE  TILT TWIST
 108.97  105.31  0.9980  149.71  0.9984   -13   -46
   => BEAM, DISTANCE, DISTORTION keywords (note that the numbers
      are on the next line here)
```

This should make the resulting integration more effective. The idea for this implementation is that the numbers end up in the "integrate set parameter" dictionary and are therefore recycled, in the same way that the GAIN currently works. Note well - some of this conflits with the definition of the FrameProcessor, in which case the relevant information should be passed back to that interface.

Have implemented an application xia2process, which should test out this wrapper. However, there seem to be some issues with the implementation when running with scripts, or Labelit as the indexer implementation, which I will need to fix.

This is reasonably effective... below are some results illustrating the changes for the 1VPJ/12287 dataset. Note that this is not quite a fair comparison, because the original set had the bad areas of the detector masked out. Still, it's interesting. The order is INFL orig, INFL new, LREM orig, LREM new.

```
Summary data for Project: demo Crystal: 12287 Dataset: infl

                                           Overall  InnerShell OuterShell

  Low resolution limit                       52.56     52.56      1.74
  High resolution limit                       1.65      5.22      1.65

  Rmerge                                     0.056     0.034     0.348
```

Rmeas (within I+/I-)	0.073	0.044	0.465
Rmeas (all I+ & I-)	0.096	0.074	0.522
Rpim (within I+/I-)	0.045	0.027	0.305
Rpim (all I+ & I-)	0.045	0.036	0.283
Fractional partial bias	-0.010	-0.032	0.008
Total number of observations	110206	3541	9776
Total number unique	26070	970	3281
Mean((I)/sd(I))	12.8	24.9	2.4
Completeness	97.9	99.1	87.8
Multiplicity	4.2	3.7	3.0
Anomalous completeness	96.8	81.2	
Anomalous multiplicity	2.3	1.7	
DelAnom correlation between half-sets	0.583	0.122	
Mid-Slope of Anom Normal Probability	1.573		

Summary data for Project: demo Crystal: 12287 Dataset: infl

	Overall	InnerShell	OuterShell
Low resolution limit	52.56	52.56	1.74
High resolution limit	1.65	5.22	1.65
Rmerge	0.055	0.032	0.347
Rmeas (within I+/I-)	0.071	0.041	0.463
Rmeas (all I+ & I-)	0.094	0.072	0.519
Rpim (within I+/I-)	0.044	0.026	0.304
Rpim (all I+ & I-)	0.044	0.035	0.281
Fractional partial bias	-0.010	-0.030	-0.038
Total number of observations	111491	3732	9910
Total number unique	26064	979	3304
Mean((I)/sd(I))	13.2	26.5	2.5
Completeness	97.9	99.3	88.1
Multiplicity	4.3	3.8	3.0
Anomalous completeness	96.9	81.5	
Anomalous multiplicity	2.3	1.7	
DelAnom correlation between half-sets	0.597	0.166	
Mid-Slope of Anom Normal Probability	1.611		

Summary data for Project: demo Crystal: 12287 Dataset: lrem

	Overall	InnerShell	OuterShell
Low resolution limit	52.56	52.56	1.74
High resolution limit	1.65	5.22	1.65
Rmerge	0.052	0.031	0.313
Rmeas (within I+/I-)	0.067	0.040	0.422

```
  Rmeas (all I+ & I-)                        0.065     0.038     0.457
  Rpim (within I+/I-)                        0.042     0.025     0.281
  Rpim (all I+ & I-)                         0.031     0.020     0.258
  Fractional partial bias                   -0.006    -0.017    -0.071
  Total number of observations              106594      3567      8233
  Total number unique                        25692       975      2969
  Mean((I)/sd(I))                             13.3      25.9       2.5
  Completeness                                96.4      99.1      79.4
  Multiplicity                                 4.1       3.7       2.8

  Anomalous completeness                      95.1      74.3
  Anomalous multiplicity                       2.3       1.5
  DelAnom correlation between half-sets     -0.192    -0.033
  Mid-Slope of Anom Normal Probability       0.868

Summary data for Project: demo Crystal: 12287 Dataset: lrem

                                           Overall  InnerShell OuterShell

  Low resolution limit                       52.56     52.56      1.74
  High resolution limit                       1.65      5.22      1.65

  Rmerge                                     0.052     0.031     0.313
  Rmeas (within I+/I-)                       0.067     0.040     0.422
  Rmeas (all I+ & I-)                        0.065     0.039     0.458
  Rpim (within I+/I-)                        0.042     0.025     0.281
  Rpim (all I+ & I-)                         0.031     0.020     0.259
  Fractional partial bias                   -0.007    -0.020    -0.068
  Total number of observations              107779      3737      8294
  Total number unique                        25678       982      2976
  Mean((I)/sd(I))                             13.6      27.1       2.6
  Completeness                                96.4      99.4      79.6
  Multiplicity                                 4.2       3.8       2.8

  Anomalous completeness                      95.1      74.5
  Anomalous multiplicity                       2.3       1.5
  DelAnom correlation between half-sets     -0.215    -0.020
  Mid-Slope of Anom Normal Probability       0.862
```

Note the very slight improvements in R, I/σ etc.

4 Thoughts

Could get some useful information by parsing the SUMMARY file with the CCP4 loggraph parser - this would give all of the information which is displayed in loggraph, which could save some parsing of the standard output.