

SDA and DAWN

Scientific Data Analysis

Mark Basham
Scientific Computing
Diamond Light Source



DAWN – General Project

- Combination of several elements in the eclipse framework
 - DAWB/Passerelle workflow tools
 - Fable/SDA Plotting/File loading
 - SDA Python/Jython
- This can be used as the underlying framework for many projects, with the ambition of reducing reimplementations of codes.

SDA-Core

- Core functionality for dealing with datasets in Java
 - NDAarray functionality which maps to a numpy ndarray
 - File Loading and saving functionality.
 - Maths, signal processing, fitting, Plotting
- Python/Jython modules
 - scisoftpy module, python/jython agnostic interface to all functionality, Based on numpy/scipy
 - XML-RPC and RMI serialization for all main dataset types allow communication between Java and Python ndarrays.
- Several generic GUI components :

Eclipse/RCP (Rich Client Platform)

The screenshot displays the Eclipse/RCP interface with several key components:

- File Navigator:** Shows a project structure with folders like 'i05', 'i11', 'i22-NCDRreduction', 'i22-nxs', 'MXData', 'MX-SDA-Training', 'BeamPosition', 'Collection', and 'DC'. The 'DC' folder contains files like 'DC_0001.cbf' through 'DC_0008.cbf'.
- Image Explorer View:** Displays a grid of diffraction images. A blue square highlights a specific image in the top-left corner.
- Dataset Plot:** Shows a grid of diffraction images. A green square highlights a specific image in the middle-left position.
- Header Dataset Plot:** Displays a table of metadata for the selected dataset.
- Image Playback:** Includes navigation buttons for image playback.
- Dataset Inspect:** Shows a 3D visualization of the dataset, including a resolution ring and a peak profile.
- Spot V:** Displays a table of spot data.

Arrows indicate the flow of data and interaction between these components:

- A blue arrow points from the 'DC' folder in the File Navigator to the Image Explorer View.
- A blue arrow points from the Image Explorer View to the Dataset Plot.
- A blue arrow points from the Dataset Plot to the Spot V table.
- A blue arrow points from the Header Dataset Plot to the Dataset Inspect 3D visualization.

Key	Value
# Count_cutoff	1220583
# Angle_increment	0.3000 deg
# Filter_transmission	1.0000
# Silicon sensor, thickness	0.00032
numPixels_y	1679
numPixels_x	1475
# Flat_field:	(nil)
# N_excluded_pixels =	255
# Detector_distance	0.22270 m
Unknown 1	# 2011-04-19T04:51:31.047
Unknown 0	# Detector: PILATUS 2M, S/N 24-01...
# Polarization	0.990
Unknown 5	# Image_path: /dls/i04-1/data/2011...
Unknown 4	# Trim_file: p2m0107_E14000_T700...
Unknown 3	# Gain_setting: low gain (vrf = -0.300)
Unknown 2	# Threshold_setting: 7000 eV
# Flux	0.0000
# Tau =	124.0e-09 s
# Start_angle	90.0000 deg.
# Wavelength	0.91730 Å
# Excluded_pixels:	badpix_mask.tif
# Pixel_size	172e-6 m x 172e-6 m
# Exposure_time	0.9964000 s

Y position	Data value	q X (1/Å)	q Y (1/Å)	q Z (1/Å)	2θ (°)	Re
680.5729	738.2269	45.0000	0.2802	0.5349	-0.0267	5.058

Generic File Exploring

The screenshot displays the DExplore software interface, which is used for comparing XAFS data. The main window title is "DExplore - Comparing /EXAFS data/Ptfoil3_1_566.nxs, /EXAFS data/Ptfoil3_2_567.nxs, /EXAFS data/Ptfoil3_3_568.nxs, /EXAFS data/Ptfoil3_4_569.nxs, /EXAFS data/Ptfoil3_5_570.nxs".

The interface is divided into several panels:

- Project Explorer:** Shows a tree view of the project structure, including folders like "bashtest" and "EXAFS data", and sub-folders like "TS1". It lists various data files (e.g., Cofoil_1_841.nxs, Ptfoil3_1_566.nxs) with their sizes and timestamps.
- Comparing /EXAFS data/Ptfoil3_1_566.nxs, /EXAFS data/Ptfoil3_2_567.nxs, /EXAFS data/Ptfoil3_3_568.nxs, /EXAFS data/Ptfoil3_4_569.nxs, /EXAFS data/Ptfoil3_5_570.nxs:** A table showing the comparison results for each file. The table has columns for "Use", "File name", and "Value".
- Dataset Plot:** A plot showing the comparison of XAFS data for the selected files. The y-axis is labeled "I0[0=0], I0[1=1], I0[2=2], I0[3=4]" and ranges from 289371 to 390000. The x-axis ranges from 0 to 300. The plot shows four data series: I0[0=0] (black), I0[1=1] (blue), I0[2=2] (red), and I0[3=4] (purple).
- Data axes selection:** A panel showing the selected data axes. The "Name" is "I0; Rank: 2; Dims: [4, ...]". The "Dim" table shows the selected dimensions: Dim 1 is "index" (checked) and Dim 2 is "dim:2" (checked). The "Energy" dimension is not selected (marked with a red X).
- 1D scatter plot / 1D stack plot:** A panel showing the selected plot type. The "1D stack plot" option is selected.

Use	File name	Value
✓	/home/awa25/XAFS_Data/Ptfoil3_1_566.nxs	0
✓	/home/awa25/XAFS_Data/Ptfoil3_2_567.nxs	1
✓	/home/awa25/XAFS_Data/Ptfoil3_3_568.nxs	2
✓	/home/awa25/XAFS_Data/Ptfoil3_4_569.nxs	3
✓	/home/awa25/XAFS_Data/Ptfoil3_5_570.nxs	4

Name	Class	Dims
entry1	NXentry	
counterTimer01	NXdata	
Energy	SDS	846
I0	SDS	846
Iref	SDS	846
It	SDS	846

Dim	1	2	3
1	✓ index		
2	✓ dim:2	✗ Energy	✗ Iref

Legend for Dataset Plot:

- I0[0=0] (black line)
- I0[1=1] (blue line)
- I0[2=2] (red line)
- I0[3=4] (purple line)

Generic Tools: ICAT Explorer

The screenshot displays the ICAT Explorer software interface. The main window is titled "DExplore - /dis/l12/data/2011/cm2061-4/default/2279.nxs - sda". The interface is divided into several panels:

- Project Explorer:** Shows a tree view of data files, including folders for years (2007-2012) and specific datasets (CM2061-1 to CM2061-6, NT6397-31).
- Metadata Panel:** Displays key information for the selected dataset (2279.nxs), including creation time, description, unique ID, dataset ID, location, modification time, version, signature, ID, checksum, file size, and command.
- Dataset List:** A table listing various elements within the dataset, such as HDF5_Versio, NeXus_versio, file_name, file_time, and a series of EDXD_Element entries (01 to 14).
- Dataset Plot:** A plot showing the data distribution. The x-axis is labeled "edxd_energy_approx" and the y-axis is "data". The plot shows several sharp peaks.
- Side: Dataset Plot:** A panel for peak analysis, showing a table of peak parameters and a zoomed-in plot of a peak profile.

Four yellow callout boxes highlight specific features:

- Direct read from ICAT:** Points to the Project Explorer.
- NEXUS data elements:** Points to the Dataset List.
- DAWN/SDA peak plotting:** Points to the Dataset Plot.
- DAWN/SDA peak profile:** Points to the Side: Dataset Plot.

Visible	Peak Position	FWHM	Area	Name
<input checked="" type="checkbox"/>	49.83	0.45	42250.49	Gaussian
<input checked="" type="checkbox"/>	72.82	0.52	986.19	Gaussian
<input checked="" type="checkbox"/>	74.96	0.54	1838.64	Gaussian
<input checked="" type="checkbox"/>	81.39	0.54	15363.79	Gaussian

Generic Tools: DIVA

- ADXV capabilities.

The screenshot displays the DIVA software interface with several panels:

- File Navigator:** Shows a tree view of files, including a folder named 'DC' containing files like 'DC_0001.cbf' through 'DC_0008.cbf'.
- Image Explorer View:** Displays a grid of diffraction images. A blue square highlights a specific region in the top-left image.
- Dataset Plot:** Shows a grid of diffraction images. A green square highlights a specific region in the bottom-left image.
- Header Dataset Plot:** Displays a table of key-value pairs for the dataset.
- Dataset Inspect:** Shows a 3D visualization of the dataset, with a wireframe box and a grid of data points.
- Table:** A table with columns for X position, Y position, Data value, q X (1/Å), q Y (1/Å), q Z (1/Å), 2θ (°), and Re.

Key	Value
# Count_cutoff	1220583 counts
# Angle_increment	0.3000 deg.
# Filter_transmission	1.0000
# Silicon sensor, thickness	0.000320 m
numPixels_y	1679
numPixels_x	1475
# Flat_field:	(nil)
# N_excluded_pixels =	255
# Detector_distance	0.22270 m
Unknown 1	# 2011-04-19T04:51:31.047
Unknown 0	# Detector: PILATUS 2M, S/N 24-01...
# Polarization	0.990
Unknown 5	# Image_path: /dls/i04-1/data/2011...
Unknown 4	# Trim_file: p2m0107_E14000_T700...
Unknown 3	# Gain_setting: low gain (vrf = -0.300)
Unknown 2	# Threshold_setting: 7000 eV
# Flux	0.0000
# Tau =	124.0e-09 s
# Start_angle	90.0000 deg.
# Wavelength	0.91730 Å
# Excluded_pixels:	badpix_mask.tif
# Pixel_size	172e-6 m x 172e-6 m
# Exposure_time	0.9964000 s

X position	Y position	Data value	q X (1/Å)	q Y (1/Å)	q Z (1/Å)	2θ (°)	Re
680.5729	738.2269	45.0000	0.2802	0.5349	-0.0267	5.058	10

Python and PyDev

The screenshot displays the PyDev IDE interface with the following components:

- Editor:** A Python script named `multiplot.py` is open, showing code for loading a dataset, defining regions of interest (ROIs), and plotting data. The code includes comments and function calls like `dnspio.load`, `dnspio.surface`, and `roi` operations.
- Dataset Plot:** A 2D plot showing a grayscale image of a dataset with a green rectangular ROI and a smaller green grid ROI. Below the plot is a table of data values:

X position	Y position	Data value	q X (1/Å)	q Y (1/Å)	q Z (1/Å)	2θ (°)	Resolution	Dataset name
1146.0000	2083.0000	21.0000	2.4249	-3.5536	-0.1475	3.926	1.4596	pixi_000000_0000.tif

- maxVals:** A 2D line plot showing two data series (blue and black) over an X-axis. The Y-axis ranges from 1 to 3.
- Plot 1:** A 3D surface plot showing a complex, multi-peaked surface structure. The Z-axis ranges from 0 to 1.2.
- Plot 2:** A 3D surface plot showing a similar surface structure to Plot 1, but with a different color scheme and Z-axis range.

Generic vs. Specific

- General Eclipse interface moves towards the generic
 - Users can get confused with the amount of functionality, very steep learning curve.
 - Everything is available, make the tools work as you want
- These issues can be addressed by providing bespoke views and tools which provide the users with a single interface
 - More intuitive for users
 - Not as flexible

PEEMA: PEEM image analysis

The screenshot displays the PEEMA software interface, which is used for PEEM image analysis. The interface is divided into several panels:

- Live Plot:** Shows a large circular PEEM image with a color map. A green square highlights a region of interest. Below the image is a table with columns for X position, Y position, Data value, and Dataset name. The dataset name is "image-00".
- Peem Analysis View:** Contains settings for image processing. It includes fields for "Images location" (C:\Data\peema), a "Filter" (*), and "Mode" (2 or 4). There are also options for "Align and average", "CD threshold" (0.00), and "Output location".
- Plot 1:** Shows a zoomed-in view of the region of interest from the main image. Below it is a table with columns for X position, Y position, Data value, and Dataset name. The dataset name is "diff".
- ImageExplorer View:** Provides "Image Playback" controls, including play, stop, and zoom buttons.
- Plot 1 (Right):** Shows a histogram of the intensity values. The x-axis is "Intensity Value" (ranging from -0.6 to 0.2) and the y-axis is "Count" (ranging from 0 to 6000). The histogram is overlaid with a color scheme and a histogram range (Minimum: 0.58, Maximum: 0.10).

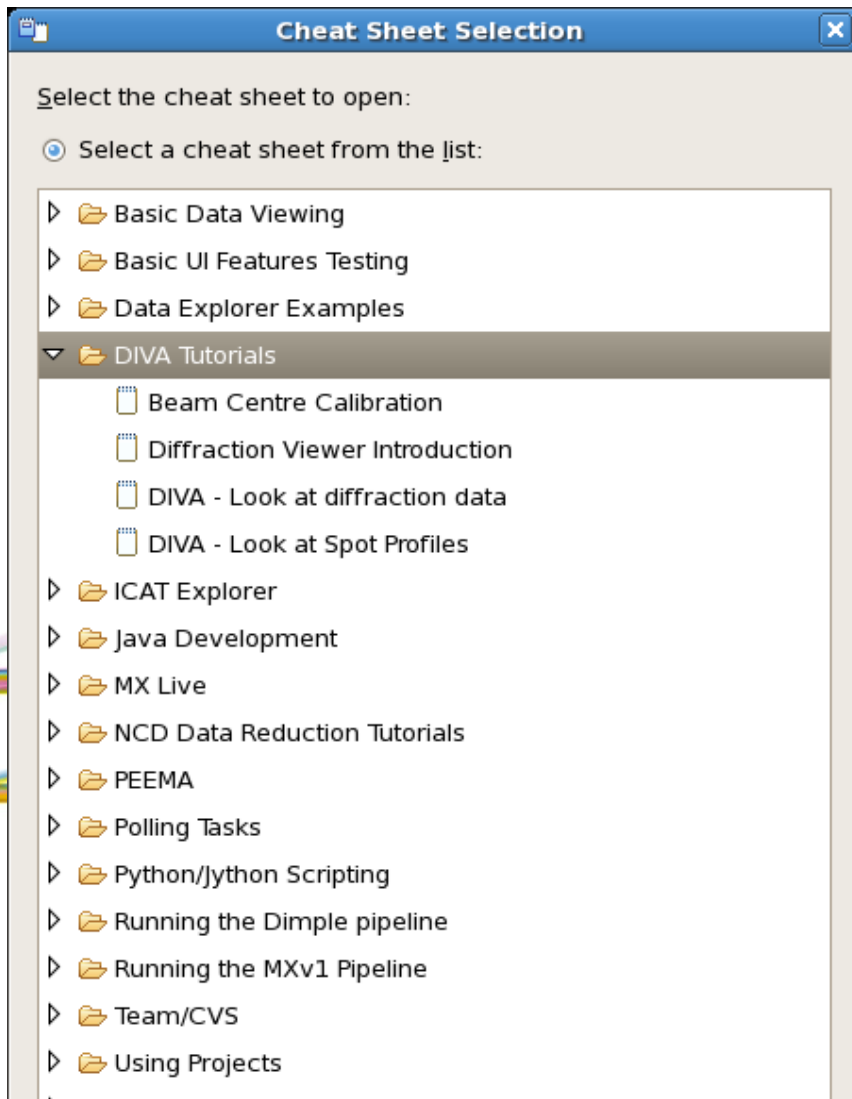
NCD Calibration and Reduction

The screenshot displays the NCD software interface with several key components:

- Project Explorer:** Lists project files (i22-4075.nxs to i22-24132.nxs) and a tree view under 'NCD' containing 'data' and 'entry1'.
- Dataset Plot:** Shows a 2D detector image with a red shaded region and axes labeled 'dim:3' and 'dim:1'.
- Side: Dataset Plot:** Displays a histogram of 'Distance (mm)' and 'Azimuthal angle (degrees)'.
- NCD Data Reduction Parameters:**
 - Data reduction pipeline:**
 - 1. Normalisation
 - 2. Background subtraction
 - 3. Detector response
 - 4. Sector integration
 - 5. Invariant
 - 6. Average
 - Reference data:** Normalisation Data: Scalers, Channel 1, Abs. Scale.
 - Background Subtraction File:** /home/SAS_Data/i22-24131.nxs, Bg. Scale 0.1
 - Detector Response File:** /home/SAS_Data/i22-24125.nxs
 - Background frame selection:** First 1, Last 2
 - Data frame selection:** First, Last, :1;3-5
 - Grid data averaging:** Average dimensions (2-3)
 - Experiment Parameters:**
 - Detectors:** WAXS (Hotwax) 1D, SAXS (Rapid2D) 2D, pixel (mm) 0.383
 - Energy (keV):** 12.4
 - Results directory:** /tmp
- Saxs Q-axis Calibration:**

Peak Position (mm)	Two Theta (deg)	d Spacing (nm)	Index (hkl)
56.77	0.98	5.84	(0 0 1)
115.42	1.96	2.92	(0 0 2)
174.26	2.94	1.95	(0 0 3)
- Calibration Function:** Gradient 0.018320, Intercept 0.03684
- Camera Length:** Camera Distance 3.43 m
- Calibration Controls:** Standard Ag Behena, Maximum reflection index 3
- Progress:** Radial Intergration, 1D Fitting, Calibrate
- Data axes selection:** Name: data, Rank: 4, Dims: 1, 1, 511. Dim 1 checked, Dim 2 checked, Dim 3 checked, Dim 4 checked.
- 2D scatter plot:** x-axis dim:4, y-axis dim:3
- Dataset slicing:** Table with columns Dim, Start position, Start value, Items, Step size.

More User Help – Cheat Sheets/Documentation.

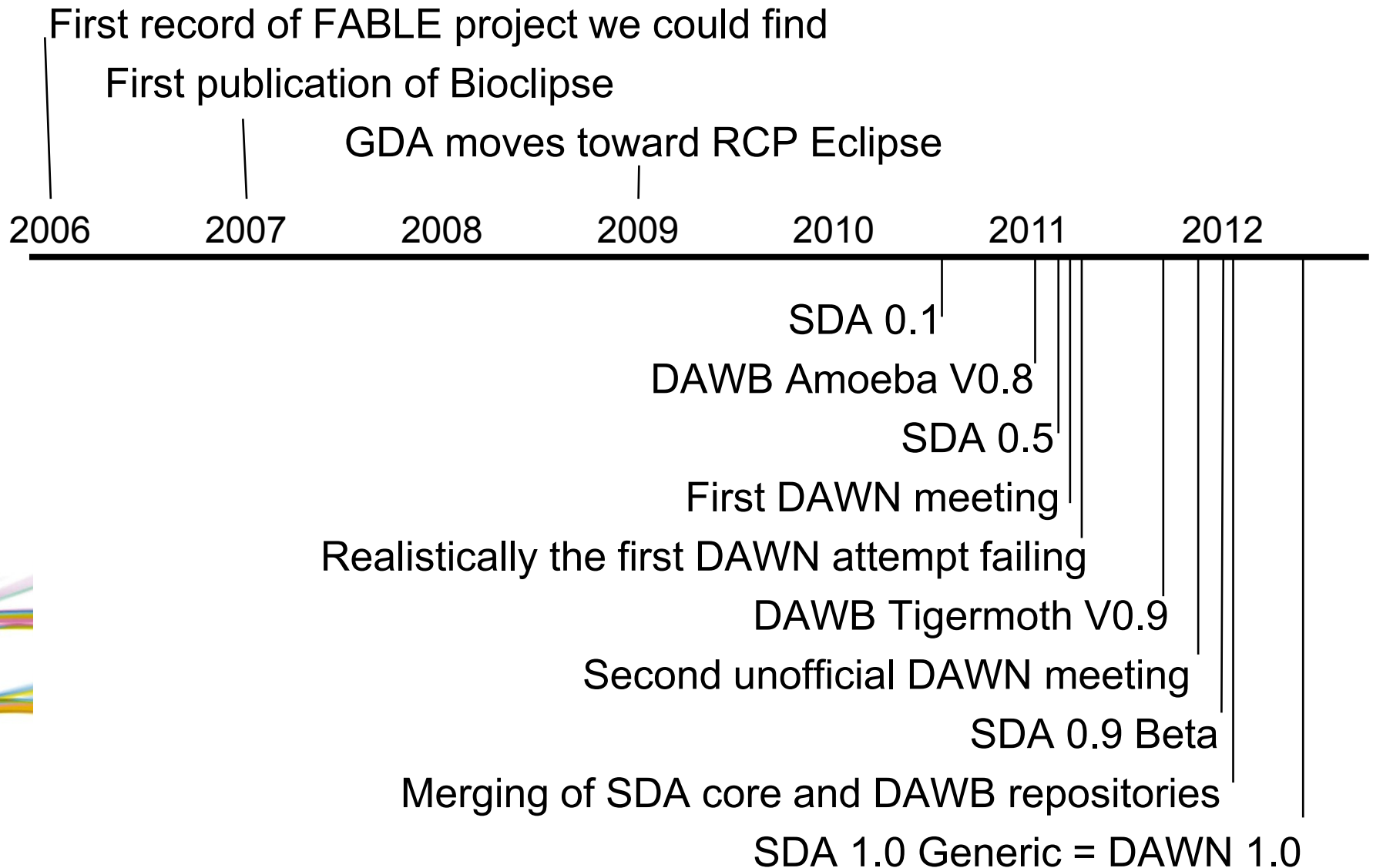


- Gives a workflow which the user can follow to understand how to achieve a goal
- Cheat sheets form major part of release testing.

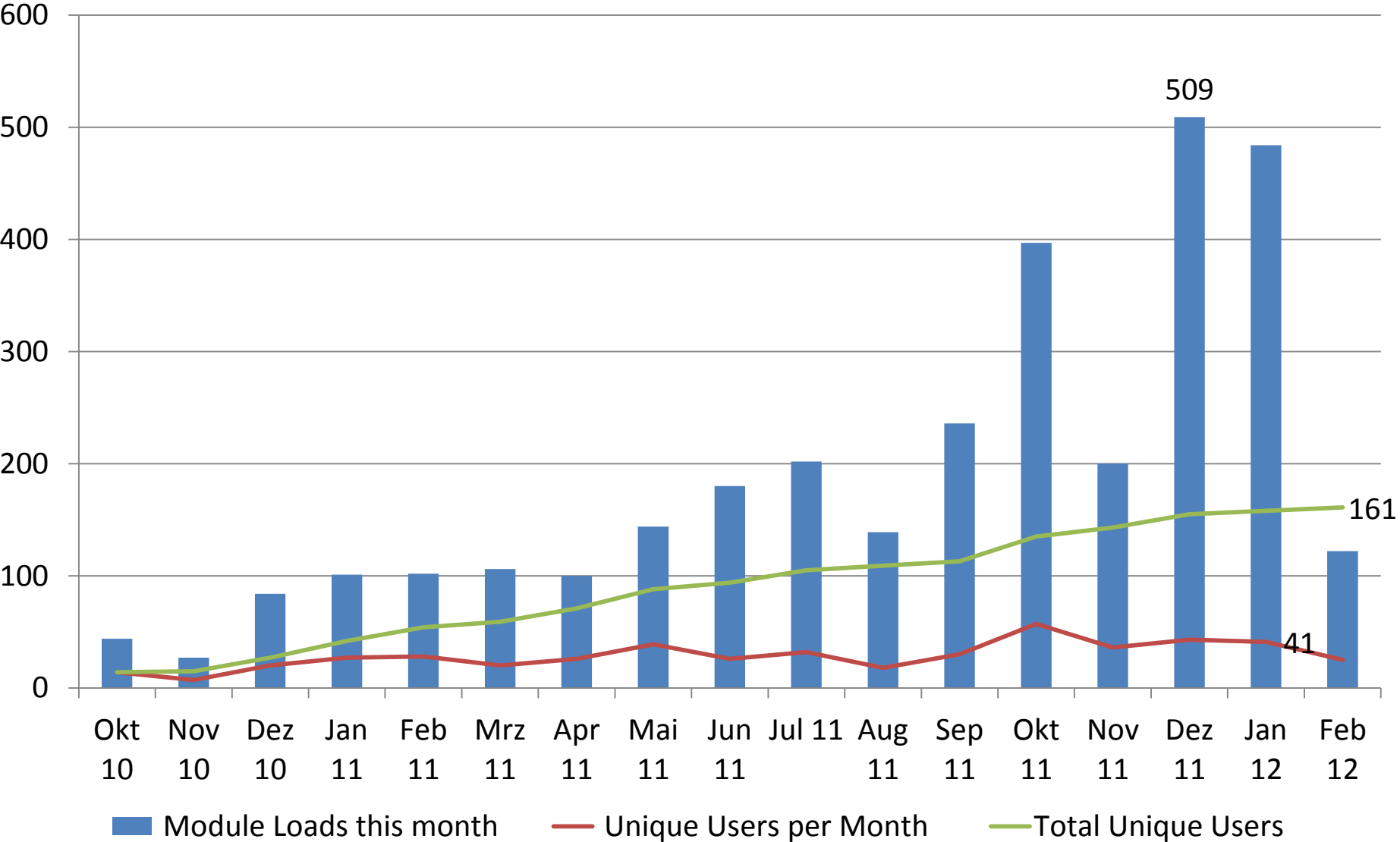
Future of SDA Generic - DAWNScience

- Eclipse project for:
 - Viewing and plotting scientific data.
 - Integrated python environment for scripting.
 - Workflow tools for pipelines.
- Open collaboration with :
 - Initial members Diamond, ESRF and EMBL-Grenoble
 - Expressions of interest from Soleil, GlobalPhasing, SLS and APS.

Future of SDA Generic - DAWNScience



Unique users and Loads over time for the Diamond SDA product.



Any Questions?



DAWN Project



DAWN – General Project

- Combination of several elements in the eclipse framework
 - DAWB/Passerelle workflow tools
 - Fable/SDA Plotting/File loading
 - SDA Python/Jython
- This can be used as the underlying framework for many projects, with the ambition of reducing reimplementation of codes.

DAWN – How will this work?

- Being based on the Eclipse Framework makes extension for other projects easy to achieve.
- Individuals or groups can work on extensions and other cutting edge tools based on the framework.
- Core functionalities are all worked on collaboratively by all DAWN contributors to keep the core tools up to date and bug free

DAWN – Code Management

- DAWNSCIENCE github repository, Main repository for the core project elements
- Local repositories for local plugins, however being in GIT would be useful for migration to the core if required at a later date.

DawnScience – Git Site

The screenshot shows the GitHub organization page for DawnScience. At the top, the GitHub logo and navigation links (Explore, Gist, Blog, Help) are visible. The user 'markbasham' is logged in. The organization name 'DawnScience' is prominently displayed, along with a notification that the user is an owner and a link to edit the profile. Below this, statistics show 16 public repositories, 0 private repositories, and 0 members. The 'Repositories (16)' section lists four repositories: 'scisoft-ui', 'scisoft-core', and 'dawn-ui', each with a commit history chart. The 'Organization Members (0)' section shows a yellow banner indicating no public members and a list of private members with 'Publicize membership' buttons.

github Search... Explore Gist Blog Help markbasham

DawnScience You are an owner of this Organization! Edit Dawn Science's Profile

Member Since Jan 10, 2012

16 Public Repos **0** Private Repos **0** Members

Repositories (16)

Find a repository...

All Repositories Public Private Sources Forks Mirrors

- scisoft-ui** Java 5 2
Scientific Software - user interface components
Last updated 2 days ago
52 week participation
- scisoft-core** Java 6 4
Scientific Software - core components
Last updated 2 days ago
52 week participation
- dawn-ui** Java 4 3
User interface plugins and features, including visualization feature
Last updated 4 days ago
52 week participation

Organization Members (0)

No public members

- andygotz (Andy Gotz) 1 public repos, 0 followers Publicize membership
- gerring (Matthew Gerring) 0 public repos, 0 followers Publicize membership
- mwebber (Matthew Webber) 3 public repos, 2 followers Publicize membership
- olofsvenson (Olof Svensson) 7 public repos, 0 followers Publicize membership
- PeterC-DLS 0 public repos, 0 followers Publicize membership
- markbasham 4 public repos, 2 followers Publicize membership
- DiamondLightSource-build-server (Source build server) 0 public repos, 0 followers Publicize membership

DAWN – Maintenance

- At DLS we run continuous Integration builds and tests for our product, currently the SDA, but soon will be DAWN + DLS
- This tests some but not all of the functionality, checking that other builds are also working is possibly difficult?
- This needs to be resolved somehow?

DAWN – Release Cycles

- Individual groups should be able to release whenever they need, due to beam-time or other operational requirements.
- The core should be released under a different schedule, but this needs to be decided by the DAWN collaboration.
- The proposed release for DAWN v1.0 will be June/July based on the majority of the work being done by DLS Developers.

DAWN – Product

- There has been much debate over the DAWN product, at the DAWN meeting in Feb 2011 the decision was made that there would be no standard release.
- This seems to have been reconsidered recently, and it is now proposed that there is a central DAWN product, which can be extended as required.