# **MANTID**

# The Mantid Project

# The challenges of delivering flexible HPC for novice end users

Nicholas Draper SOS18









#### What Is Mantid

- · A framework that supports high-performance computing and visualisation of scientific data.
- · Manipulate and analyse Neutron and Muon data.
  - Neutron Scattering
    - · Diffraction, spectroscopy, small angle, reflectometry
  - Muon Spectroscopy
  - Could be applied to other techniques
- · Open Source
- Multiple Platforms
  - Windows, Linux, Mac







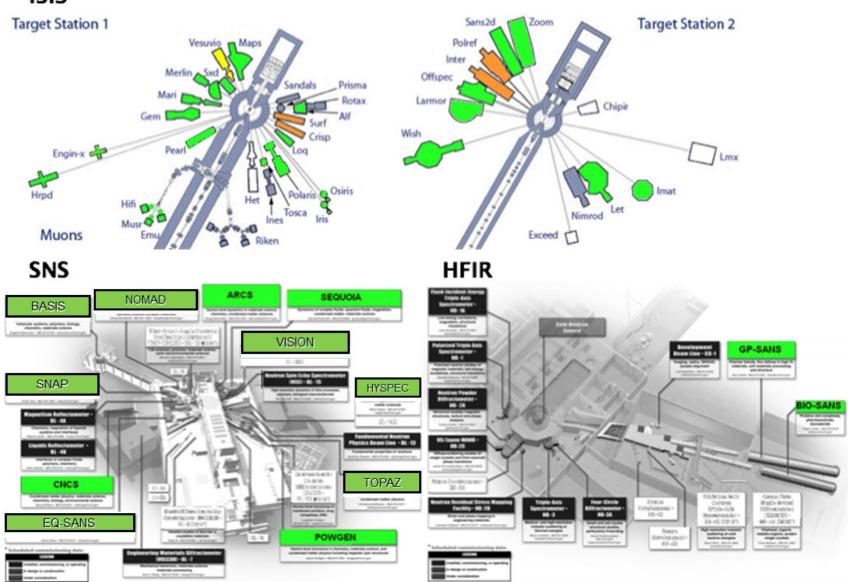
# **Project History**







#### **Members**



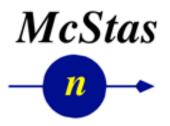




#### Contributors





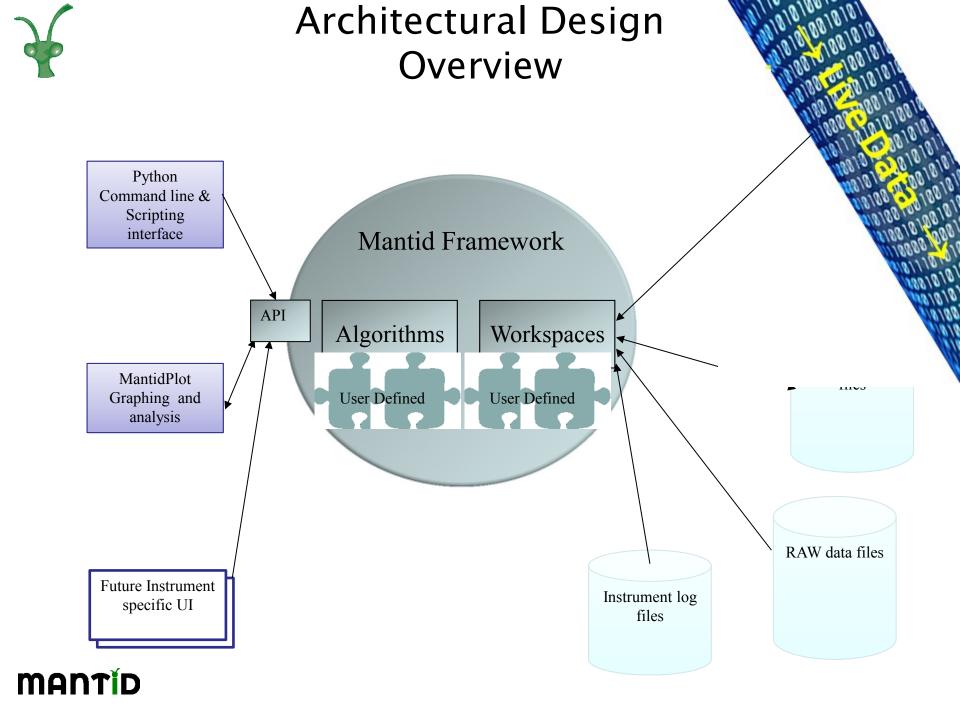














#### Plug in extensions

GUI

Algorithm Dialogs Custom Interfaces Custom Menus

Framework

Python scripts & libraries
Workflow Algorithms
Algorithms

Workspaces

Utility

**Unit Conversions** 

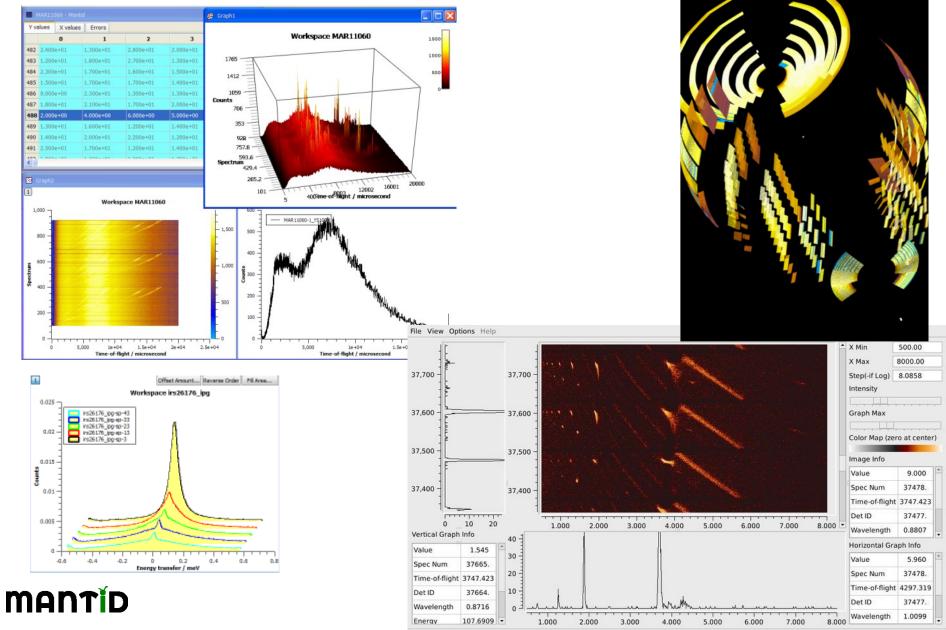
Fit Functions Cost Models Constraints Minimizer

Archive Searching LiveData Listeners Data Catalogs



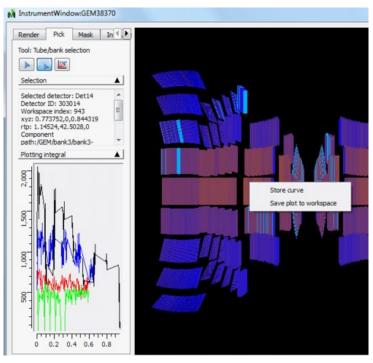


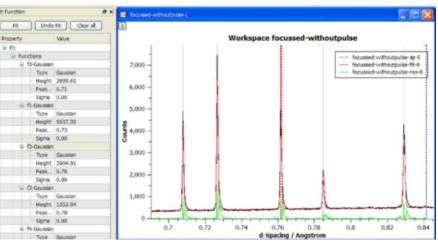
#### What can we do - Visualization

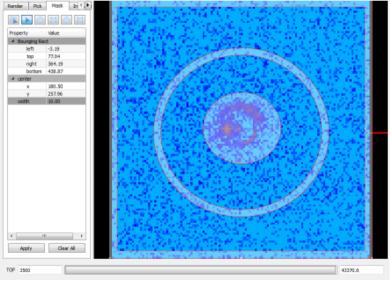


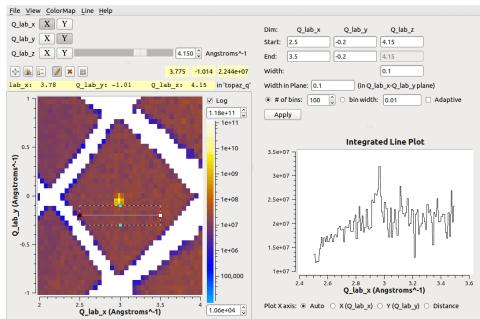
# Y

#### What we can do - Tools





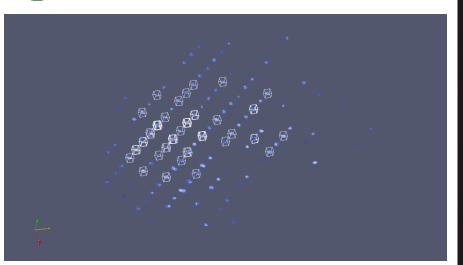


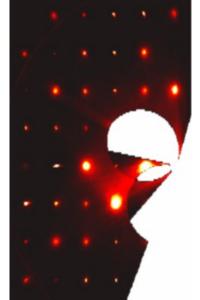


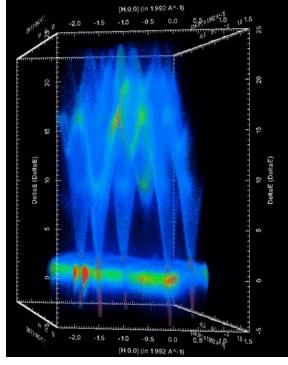


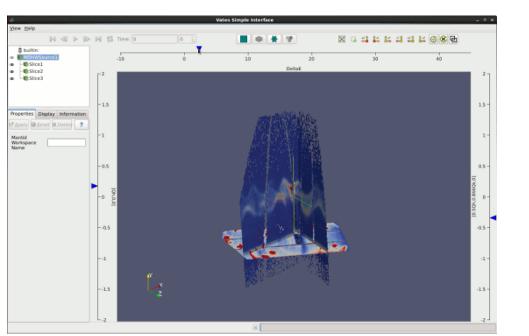


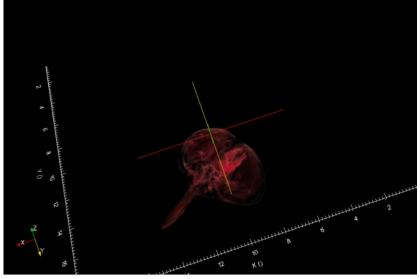
# What can we do - Visualization







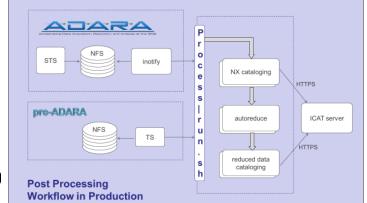






### What we can do - Scripting

- Python Control
  - Within MantidPlot or command line
  - iPython Shell
  - Integrated numPy
  - Python Plugins
    - · Algorithms
    - Fitting functions
    - User interfaces
  - Automated reduction



```
MantidPlot: Python Window
                                                                                                                                  _ D X
File Edit Execute Window
Squares.py RunPythonScript.py calpdf.py
    3 # Mantid-PDF Calculator for POWGEN and NOMAD
    6 # (1) Hard coded for POWGEN NI
    8 # Last Edit Location
   10 # * 2011.06.16 MBF
   11#
   12 #
   16 DBPATH = "/home/wzz/Projects/Mantid-Project/Tests/PDF-01/"
   18 binningQ = (0.02,0.02,40)
   20 HUGE = 1.0E18
   22 def prototypemain():
       *** Main method
       # vanruns = ["PG3_2548_event.nxs", "PG3_2577_event.nxs"]
# canruns = ["PG3_2583_event.nxs", "PG3_2584_event.nxs"]
        # bkqdruns = ["PG3_2585_event.nxs", "PG3_2586_event.nxs"
        sampleruns = ["PG3_2581_event.nxs"]
        vanruns = ["PG3_2548_event.nxs"]
        canruns = ["PG3 2583 event.nxs"]
        bkgdruns = ["PG3 2585 event.nxs"]
Wed 13. Jun 10:40:58 2012: Script execution started
Status: Executing.
```

```
Rebin-[Notice] Rebin started

>>>
>>> print "Rebinned workspace has " + str(rebinned.getNumberHistograms()) + " hi
stograms"
Rebinned workspace has 2584 histograms
>>> print "Spectrum 450's X data size = " + str(len(rebinned.readX(450))) + " bi
n boundaries"
Spectrum 450's X data size = 4901 bin boundaries
>>> # Perform some algorithms
... testWs = Load("HEIT5869.raw")
Load-[Notice] Load started
>>> converted = ConvertUnits(testWs, "dSpacing")
ConvertUnits-[Notice] ConvertUnits started
>>> rebinned = Rebin(converted, "0.1,0.001,5")
Rebin-[Notice] Rebin started
>>> print "Rebinned workspace has " + str(rebinned.getNumberHistograms()) + " hi
stograms"
Rebinned workspace has 2584 histograms
>>> print "Spectrum 450's X data size = " + str(len(rebinned.readX(450))) + " hi
n boundaries"
Spectrum 450's X data size = 4901 bin boundaries
>>> print "Spectrum 450's X data size = " + str(len(rebinned.readX(450))) + " bi
```

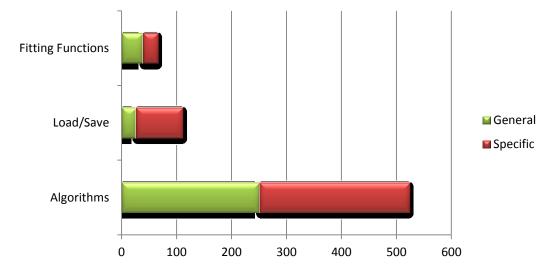




### What we can do - Algorithms

# 703

- Arithmetic
- Correction Functions
  - Absorption
  - Background
  - Efficiency
- Technique
  - Single Crystal
  - Inelastic
  - SANS
  - Powder diffraction
- Data handling



- · Diagnostics
- Event Filtering
- Optimization
- Transforms
  - Masking
  - Grouping
  - Smoothing
  - Unit conversions





#### The Mantid Environment

#### Users

- From scientific experts, who will understand HPC to some extent, but have limited time.
- To visiting scientists, who just want results, and have little time to understand systems or learn new processes.

#### Compute environments

- Local computer only
- Powerful workstations
- Facility HPC facilities
- University facilities
- Commercial Cloud resources
- National HPC facility





### Mantid and Distributed Computing

- Not all operations in Mantid would benefit from distributed or HPC computing
  - Small data volume
  - Large data volume, local to client, simple operations
- For some operations the need is clear
  - Large Data Volume, fast access by cluster
  - Complex scalable operations
    - Monte Carlo instrument simulations
    - Absorption corrections
  - Use of third party codes
    - Molecular dynamics simulations
  - Computationally expensive optimisations of models to experimental data





### What is a Mantid job?

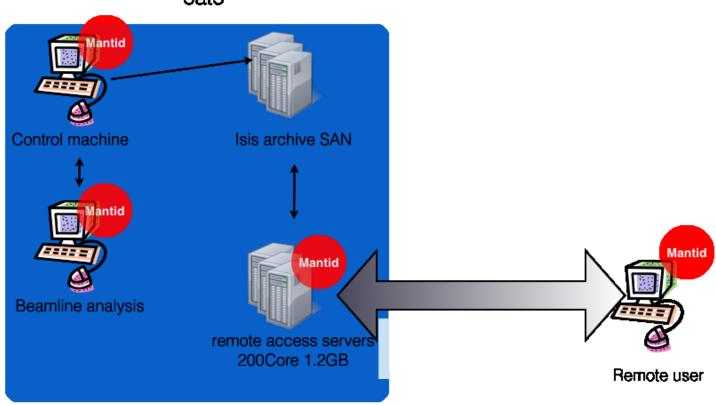
- · Input data
  - Shared File Location
  - Uploaded File
- Python Script
- · Output data
  - Shared File Location
  - Downloaded File





# One possible Infrastructure

ISIS user compute infrastructure 1Gbit cat5







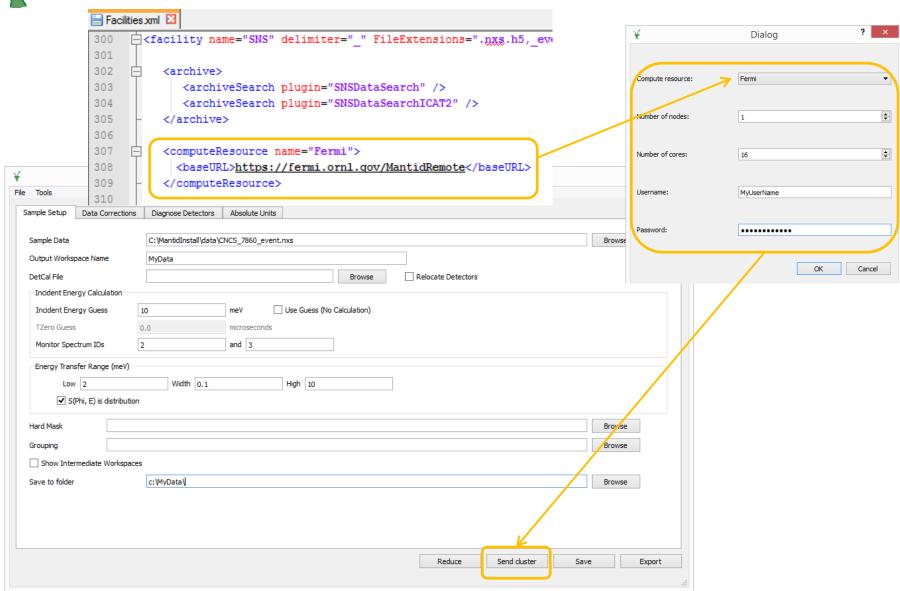
### Challenges

- Different facility infrastructures
  - Authentication
    - Username/password, certificates, need to physically turn up with ID.
  - Job Schedulers
    - Few accessible via web services
    - Of those that do not few work well with C++ cross platform
  - Locating resources





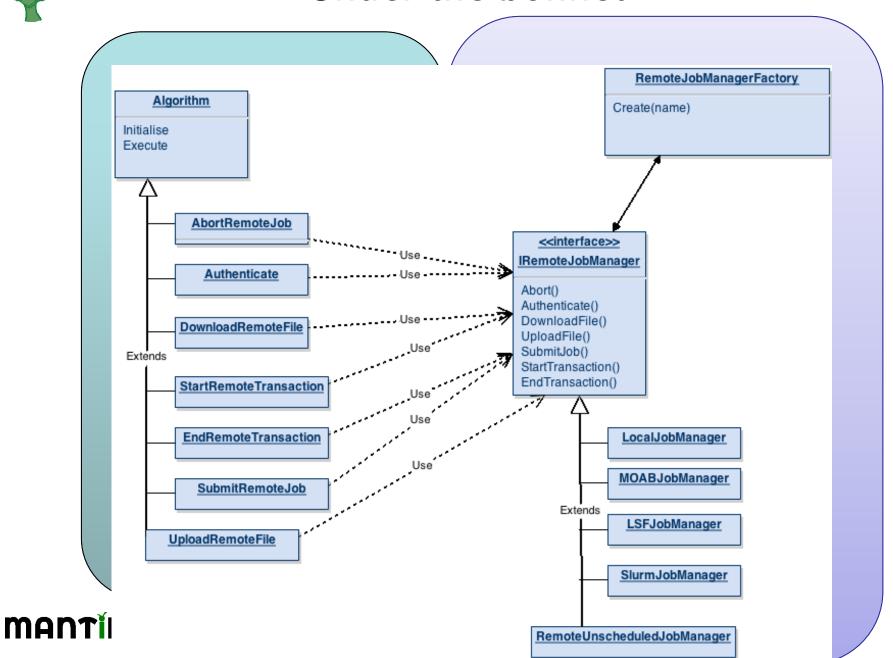
# Interfaces - Keep it simple







#### Under the bonnet





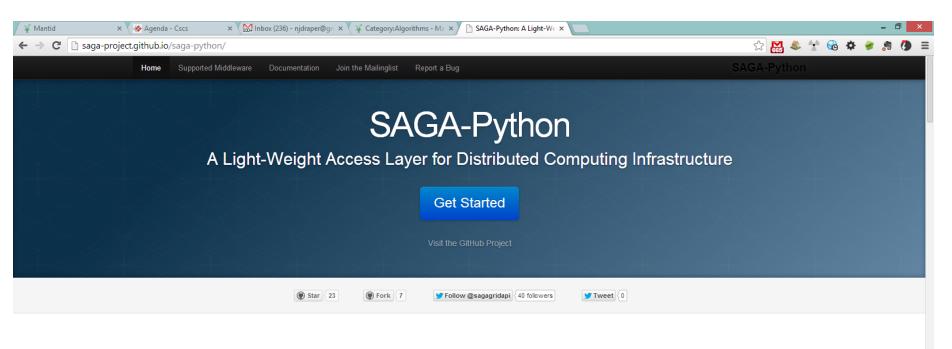
#### The Ideal Middleware

- Removes the need to care what job scheduler is in use.
- · Compiles easily on Windows, Mac, Linux
- Packages easily for deployment
- APIs
  - C++, Python
  - others
- · Needs to handle network proxies, firewalls etc
- Does not need Admin or special permissions
- Easy to add additional Job Managers
- · Just works ...





# Doesn't this sound a bit like ... SAGA



#### Introduction

A Simple API for Grid and Distributed Applications.

SAGA (Simple API for Grid Applications) defines a high-level interface to the most commonly used distributed computing functionality. SAGA provides an access-layer and mechanisms for distributed infrastructure components like job schedulers, file transfer and resource provisioning services. Given the heterogeneity of distributed infrastructure, SAGA provides a much needed interoperability layer that lowers the complexity and improves the simplicity of using distributed infrastructure whilst enhancing the sustainability of distributed applications, services and tools.

SAGA-Python provides a Python module that is compliant with the OGF GFD.90 SAGA specification. Behind the API façade, SAGA-Python implements a flexible adaptor architecture. Adaptors are dynamically loadable modules that interface the API with different middleware systems and services. Most application developers use the adaptors that are already part of SAGA-Python, but you can easily implement your own in case your backend system is not supported yet.





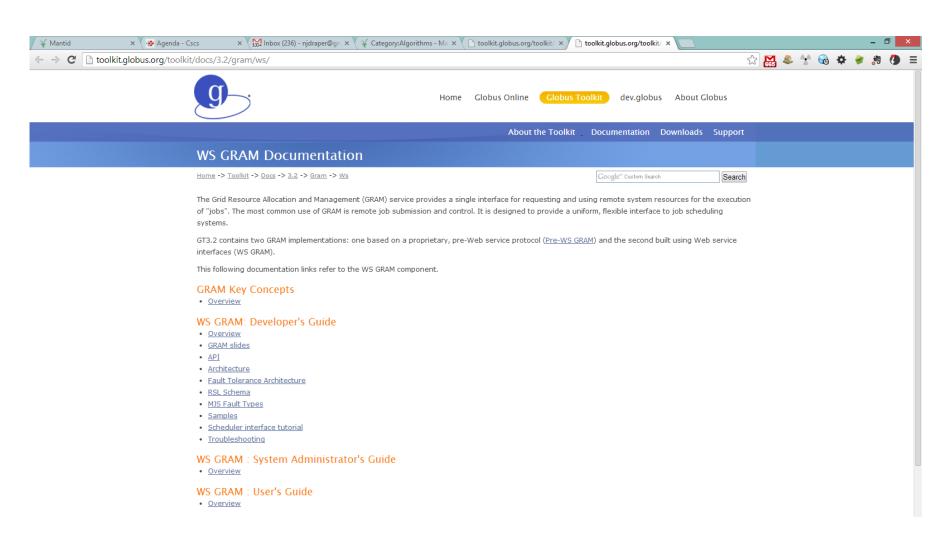
# Doesn't this sound a bit like ... SAGA

- Removes the need to care what job scheduler is in use.
- · Compiles easily on Wincxws, M?, Linx
- · Packages easily for deployment
- APIs Development stopped 2 yrs ago
  - CX., Python
  - others Java
- · Needs to handle network proxies, firewalls etc SSH only
- Does not need Admin or special permissions
- Easy to add additional Job Managers
- · Just works ..?





# Doesn't this sound a bit like ... WS\_GRAM







# Doesn't this sound a bit like ... WS\_GRAM

Removes the need to care what job scheduler is in use.
 Not local or remote workstation

**WS\_GRAM** 

Web services

- · Compiles easily on Windows, Mac, Linux
- Packages easily for deployment
- · APIs
  - C++, Python
  - others
- Needs to handle network proxies, fire etc
- Does not need Admin or special permissions
- Easy to add additional Job Managers
- · Just works ..?







#### Further information

- · Project web page www.mantidproject.org
- $\cdot$  Many Thanks to the Project Sponsors  $\, \operatorname{ISIS} \,$
- · And the development team









