# SAMBUCA Meeting Summary Tuesday 26 August 2014

Daniel’s first attempt at a list of activities and tasks (priority is loosely indicated by list ordering: higher priorities are ahead of lower priorities). Highlighted tasks are those I plan to start first.

1. Logistics / groundwork
   1. Gain access to PowerApp 5 and dedicated SAMBUCA VM on WRON infrastructure
   2. Test if existing SAMBUCA runs on Bragg using the SC IDL licences.
      1. Luke may be able to help with this
   3. Get data set for testing
      1. Small, quickly evaluated
      2. Well understood behaviour
      3. Known outputs (so I can set up some validation tests of Python SAMBUCA)
   4. Break down scope into two consecutive eResearch projects
   5. Provide information on upcoming Python training courses
   6. Do we need (or already have):
      1. JIRA for task/issue tracking?
      2. Confluence wiki?
   7. How many non-CSIRO collaborators are there? Can they all be provided with CSIRO partner accounts through [PUMA](https://wiki.csiro.au/display/CS/PUMA)? This will provide access to [Confluence](https://wiki.csiro.au/display/CS/Confluence), [JIRA](https://wiki.csiro.au/display/CS/JIRA), Stash etc.
   8. Get most recent design documents (workflow diagram etc)
   9. Double check proposed direction with key SAMBUCA stakeholders
2. Source control
   1. Definitely required for Python port. Daniel will be setting up a personal Git repo if a central one is not adopted.
   2. May be advantages to hosting the current IDL code, but this is less important
   3. Must be accessible to non-CSIRO collaborators
   4. Options:
      1. [Git on CSIRO Stash](https://wiki.csiro.au/display/CS/Git+Repository+Hosting+With+Stash)
         1. Accessible via CSIRO Partner accounts
         2. Should SAMBUCA be released as open source software in the future, a Git repo can be easily cloned to GitHub or similar service.
      2. CSIRO hosted Subversion
3. Python Port
   1. Feasibility study
      1. Primary blocks could be external dependencies of the IDL system. EG:
         1. ESRI and ENVI specific file formats
            1. Most can be handled by the GDAL Python library (ENVI .hdr rasters OK)
      2. Code inspection
      3. Gain understanding of current SAMBUCA system and workflows
   2. Goal: Loosely coupled Python modules for:
      1. SAMBUCA optimisation routines
      2. Data ingestion
         1. AGDC
         2. SIOP database
         3. Substrate database
         4. Bathymetry data
         5. Other sources
      3. Optional parallel execution
         1. Task decomposition
         2. Result aggregation
         3. Ideally on a cluster (NCI, CSIRO)
         4. If Workspace is adopted for workflow and GUI support, should it be relied on for parallel execution, or should we keep this in pure Python?
            1. Various tools exist for parallel execution of tasks via Python (eg: [IPython](http://ipython.org/)). Is this required, or should the first version exploit the total independence of each sub-task and simply launch separate jobs (perhaps in a PBS array job).
   3. Initial workflows driven by Python scripts, GUI added later
4. GUI
   1. May exceed effort available in current eResearch Collaboration Project (ERCP). A second ERCP for the first half of 2015 is recommended.
   2. [Workspace](http://www.csiro.au/Outcomes/ICT-and-Services/Workspace.aspx) is worth considering as a GUI solution
      1. Can be used to build custom GUI solutions, as well as the standard visual workflow interface
      2. Supports remote and parallel execution of tasks from within a workflow
      3. Blue-sky idea: Generic AGDC component may be able to attract development support from Workspace team.
      4. Daniel to clarify
         1. Workspace licencing
         2. Availability on NCI
         3. Restrictions on distribution to non-CSIRO colleagues
      5. Info:
         1. [CSIRO internal Workspace wiki](https://wiki.csiro.au/display/Workspace/Main)
   3. If Workspace is not adopted, what other GUI options are there?
      1. Custom GUI in Python, using an open source GUI library. EG:
         1. [TkInter](https://wiki.python.org/moin/TkInter): de-facto Python GUI standard
         2. [PySide](http://qt-project.org/wiki/PySide): LGPL Python bindings to the cross-platform Qt library.
5. Australian Geospatial DataCube (AGDC)
   1. SAMBUCA should be able to ingest data from the AGDC
   2. AGDC is in early development, and SAMBUCA may become one of the pilot projects
   3. Current access to AGDC data is through Python classes, so having some SAMBUCA code in Python will facilitate development of the AGDC ingestion modules
   4. Development effort from GA may be available to assist
   5. Activities
      1. AGDC metadata for use in GUI
      2. Ingestion of actual data
         1. Queries for data selection
         2. Possible format translation