

# AiravataProposal

## Airavata Proposal for Apache Incubator

### Status

The Airavata Proposal has been [accepted](#)

### Abstract

Airavata is a software toolkit currently used to build science gateways but that has a much wider potential use. It provides features to compose, manage, execute, and monitor large scale applications and workflows on computational resources ranging from local clusters to national grids and computing clouds. Users can use Airavata back end services and build gadgets to deploy in open social containers such as Apache Rave and modify them to suit their needs. Airavata builds on general concepts of service oriented computing, distributed messaging, and workflow composition and orchestration.

### Proposal

Airavata will provide web interfaces and scalable Service Oriented Architecture based backend services to build or enhance Science Gateway (see <https://www.teragrid.org/web/science-gateways/>) and similar environments. Airavata will specifically focus on:

1. sophisticated server-side tools for registering and managing large scale applications on computational resources.
2. graphical user interfaces to construct, execute, control, manage and reuse of scientific workflows.
3. interfacing and interoperability with various external (third party) data and provenance management tools.

### Background

Working in close quarters with Apache Axis2 committers and inspired by the true open source community driven software development of ASF, Suresh Maru and Marlon Pierce have been pioneering the idea of a Science Gateways software-based Apache project since late 2008. Many Apache members have fostered these ideas and guided them to arrive at this proposal.

Currently the software is actively used in various science gateways. But the tools are general purpose and build upon widely used Apache tools like Axis2, ODE engine. The core team is motivated to expand the community and build a community welcoming both synergistic software components and also new usage scenarios.

It is perhaps worth noting that one of the three seed projects that make up the Apache Rave (Incubating) project is also the product of this same team and is derived from the same Science Gateways community.

### Rationale

The nature of computational problems has evolved from simple desktop calculations to complex, multidisciplinary activities that require the monitoring and analysis of remote data streams, database and web search and large ensembles of simulations. In the academic domain Science Gateways have emerged to address these needs and have built software platforms that provide a community of users with the ability to easily solve computational problems within a specific domain. The tools developed to support these gateways are potentially of value to any organisation needing to perform complex computations. Gateways provide a convenient interface to the underlying infrastructure without the need for a deep understanding of the intricacies that infrastructure.

We summarize the rationale for choosing The Apache Software Foundation (ASF) below. This is what we hope to gain from participating in the ASF.

1. **Broader impact:** our science gateway tool set is based on Service Oriented Architecture principles, and it has always been our goal to align our software with broader trends in the development of software for distributed systems. Participating in the ASF provides a concrete way to implement this idea. In particular, we have done extensive work on the workflow systems, messaging, and application management as Web services from the perspective of computational science use cases (i.e., high failure rates, very long running jobs, dynamic service creation, workflows not expressible as directed acyclic graphs, etc). These requirements and our work to implement them have already had direct impact on the Apache Axis 2 and Apache ODE projects. As an Apache project, it is hoped that our community will have an enhanced opportunity for collaboration and complementary development with Apache Hadoop (for scientific application management), Apache QPID (for messaging), Apache Rave (incubator - Open Social Container) and others. It is our goal to expand our software's usage beyond just science gateways to the broader enterprise community.
2. **Sustainability:** Science gateway software development (and cyberinfrastructure software generally) is primarily funded in the US by the National Science Foundation (NSF), so the long term sustainability of software across funding cycles is a longstanding problem. The NSF is attempting to solve this problem, and its vision for sustainable software is described here: <http://www.nsf.gov/pubs/2010/nsf10015/nsf10015.jsp>. Participating in the ASF is our project's vision for reaching software sustainability that underpins the NSF CF21 vision. As a successful ASF project (after incubation), we will have created a community led, rather than funding led, environment for the development of our software. This community, through our community engagement work and adoption of meritocratic principles, will expand beyond our current core team and existing project collaborations. This will greatly increase the chances that our software will continue to grow and improve beyond the participation of any individuals.
3. **Maturity:** much of the software included in this proposal was developed initially by graduate students as part of their Ph. D. work. The Open Grid Computing Environment has devoted significant effort (through salaried staff and volunteers from collaborating institutions) to convert these research projects into mature, reliable, well-written, packaged components. The code is currently hosted at [SourceForge](#), but we recognize the need to go beyond just the [SourceForge](#) support tools to participate in a real community of software engineering experts. It is our desire, through the Apache Incubator, to take our software engineering efforts to a higher level by learning from the substantial experience of appropriate Apache Committers. Apache mentors will provide initial guidance, as will the attraction of additional committers from the relevant Apache projects.

## Initial Goals

- Implement a standalone version of the code base with a simple hello world service, workflow and gadget(s) to access the examples.
- Migration of documentation and design knowledge from existing SF project
- Re-architect Grid based security (GSI) dependencies and adopt more general purpose security implementations.
- Make sure Cloud (including hadoop) support is more first class.
- Aim to have the first Apache release within the first 6 months
- Verify with Apache Legal that some of the more esoteric licences in our dependencies are acceptable, or replace them as appropriate

## Current Status

The proposed tools are currently hosted on [SourceForge](http://sourceforge.net/projects/ogce/) at <http://sourceforge.net/projects/ogce/> (source at <https://ogce.svn.sourceforge.net/svnroot/ogce/ogce-xbaya-gui/>) and are described at <http://www.collab-ogce.org>.

## Meritocracy

A significant portion of initial committers are already ASF Committers/Members, and the entire team is well experienced with open source software development. The existing code base has resulted from multi-institutional collaborative projects. The developers are well aware of the Apache way and will honor the meritocracy policy of ASF foundation.

## Community

To date our focus has been serving our immediate partners needs rather than looking outwards in order to build a broader community with diverse needs. Whilst the core team area likely to remain focussed on the Science Gateways communities we are keen to welcome community members from other disciplines.

## Core Developers

Our core developers consist of participants from academic, not-for-profit and for-profit organisations. Many are already well versed in The Apache Way.

Amongst our initial team we have one or more committers on the following Apache top level projects; axis, geronimo, synapse, ws, ws-pmc, ws-woden as well as Apache Rave (Incubating).

## Alignment

Airavata software is built upon Apache Projects like Axis2, ODE, Rampart, Tomcat and Maven. We will try to closely align the project with ODE to ensure BPEL workflow compatibility. We will align with metadata management projects like Apache OODT. Web interfaces within the Airavata software will be synergistically developed with Apache Rave.

## Known Risks

### Orphaned products

We acknowledge the need to seek project contributions outside the current developers. The core team actively travels and conducts workshops and tutorials at relevant academic conferences like Supercomputing, [TeraGrid](#), Collaborative Technologies Systems and SciDAC. Previous experiences have showed that these tutorials and outreach efforts will bring in community participation. The general strategy will be to encourage users to be active in the community and develop patches and contribute. Also, the core developers use the Airavata software in multiple projects with a life span ranging from 2 to 10 years, so the risk of orphaned products is very minimal.

Furthermore, by opening our doors to non-academic organisations already adopting large scale computation related projects in the ASF we hope to be able to build community beyond the proposing teams Science Gateway interests.

### Inexperience with Open Source

The core team is very familiar with open source practices. The developers include existing Apache members who have long term experience with the Apache Way. The OGCE project has been an active open source project in [SourceForge](#) since November 2006. We welcome the new directions and are well prepared to follow the Apache way.

### Homogenous Developers

We have a semi-distributed development environment distributed among Indiana University and Lanka Software Foundation. We fully expect contributions from the partnering science gateways adding to the heterogeneous development.

### Reliance on Salaried Developers

The core developers are self motivated on the project and also are funded through various federal, state and endowment research grants. Participation in these research efforts based on Airavata software is mostly voluntary and above and beyond the requirements of the salaried jobs.

The Open Gateway Computing project, from which the initial code donation is sourced, is funded for the next 3 years and is mandated by the funding guidelines to open source software development - <http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=1032742>. We believe in the Airavata software capabilities and its vital role in providing sustainable middleware for Science Gateways. Nevertheless, the core team will actively build upon Airavata software and foster developer community outside the current core.

## Relationships with Other Apache Products

See “Alignment” above. Airavata is based on the concepts of Service Oriented Architecture and all services run within Tomcat container. The web services are based on Axis2. The orchestration of the scientific workflows uses Orchestration Director Engine. The software is built using Apache Maven.

## An Excessive Fascination with the Apache Brand

The Apache brand would certainly help promote the software suite, but gaining the brand is not the motivation for this project. Airavata is being proposed to Apache because of the belief in Apache’s meritocracy model for mentored, community-driven, open source software is the best way to develop sustainable software. See “Rational” above. Most importantly, The Apache Software Foundation will help us create an institution-neutral contribution venue and will help us build a long-standing community around Airavata to sustain and improve it beyond the span of specific, targeted research grants.

## Documentation

Existing documentation is available from the OGCE wiki, [http://www.collab-ogce.org/ogce/index.php/Main\\_Page](http://www.collab-ogce.org/ogce/index.php/Main_Page). In addition, there is abundance of presentation and self guided video tutorial material. Effort will be put in to collect all this information into meaningful documentation on the Apache websites.

## Initial Source

The initial source of the project is in [SourceForge](#). The source is available for anonymous check out from svn at <https://ogce.svn.sourceforge.net/svnroot/ogce/ogce-xbaya-gui/>

## Source and Intellectual Property Submission Plan

Indiana University is the current holder of Intellectual Property rights for the software. The university has approved the code donation and signed trustees approval, Corporate Contributor Licence Agreement and Software Grant Agreement have been emailed to ASF secretary and received acknowledgement.

Specifically Indiana University will donate 4 components into Airavata project.

1. XBaya Scientific Workflow Suite - includes a GUI for workflow composition and monitoring. The composed workflow can be exported to various workflow languages like BPEL, SCUFL, Condor DAG, Jython and Java. The defacto workflow enacting engine used is Apache ODE. 2. GFac - an application wrapper service that can be used to wrap command line-driven science applications and make them into robust, network- accessible services. This component is build on Axis2 web service stack. 3. XRegistry - a registry service for storing deployment information about wrapped application services and constructed workflows. 4. WS-Messenger - a “publish-subscribe” based message broker implemented on top of Apache Axis2 web services stack. It implements the WS-Eventing and WS-Notifications specifications and incorporates a message box component that facilities communications with clients behind firewalls and overcomes network glitches.

## External Dependencies

Following the guideline -<http://www.apache.org/legal/resolved.html>, the following are the dependent software and all of them are in binary format in java archive (jar files).

- CDDL license - Javax activation, JSR311, Portlet-API, Servlet-API
- Apache V2: cog-jglobus, globus, caster, gridsphere, Woodstox, xmpp, xsul, sigiri, atomixmiser, weps-beans.
- BSD: puretts,
- MIT: bcprov, hsqldb, dom4j, slf4j
- PSFL: Jython
- GPL 2.0: mysql-connector-java
- Other:
  - cryptix32, cryptix-asn1 (<http://www.cryptix.org/LICENSE.TXT>)
  - backport (public domain)
  - jaxen (<http://jaxen.codehaus.org/license.html>)

Licence incompatibilities (GPL) will be resolved during incubation.

## Cryptography

The software does not implement any cryptographic algorithms. However, to perform secured messaging and data movement and SSL communications, the software depends upon third party security libraries. These external libraries depend in turn on Java Security, Puretts, Cryptix and Bounce Castle libraries. Apache Cryptographic steps will be followed to register the use of these libraries.

## Required Resources

### Mailing lists

1. airavata-dev
2. airavata-commits
3. airavata-private

## Subversion Directory

<https://svn.apache.org/repos/asf/incubator/airavata>

## Issue Tracking

We intend to make use of Jira for issue tracking. Proposed key: AIRAVATA

## Other Resources

We intend to manage our website using the Apache CMS.

## Initial Committers

Names of initial committers with affiliation and current ASF status:

| Name                 | Email                  | Affiliation               | ICLA    | ASF Status       | Apache Id |
|----------------------|------------------------|---------------------------|---------|------------------|-----------|
| Suresh Marru         | smarru@cs.indiana.edu  | Indiana University        | On File | Apache Committer | smarru    |
| Marlon Pierce        | mpierce@cs.indiana.edu | Indiana University        | On File | Apache Committer | mpierce   |
| Srinath Perera       | hemapani@apache.org    | Lanka Software Foundation | On File | Apache Member    | hemapani  |
| Aleksander Slominski | aslom at us.ibm.com    | IBM                       | On File | Apache Member    | aslom     |
| Raminderjeet Singh   | ramifnu@indiana.edu    | Indiana University        | On File | Apache Committer | raminder  |
| Archit Kulshrestha   | akulshre@indiana.edu   | Indiana University        | On File | N/A              | N/A       |
| Chathura Herath      | chathura@apache.org    | Indiana University        | On File | Apache Committer | chathura  |
| Eran Chinthaka       | chinthaka@apache.org   | Indiana University        | On File | Apache Member    | chinthaka |
| Thilina Gunaratne    | thilina@apache.org     | Indiana University        | On File | Apache Committer | thilina   |
| Wathsala Vithanage   | wathsala@opensource.lk | Lanka Software Foundation | On File | N/A              | N/A       |

All the parties are affiliated with companies and organizations that are familiar with the development of open source. We expect that the amount of volunteer work will increase, and more developers will come on board.

## Champion

Ross Gardler, Apache Software Foundation

## Nominated Mentors

- Ross Gardler, Member, Apache Software Foundation
- Alek Slominski, Member, Apache Software Foundation
- Ate Douma, Member, Apache Software Foundation
- Sanjiva Weerawarna, Member, Apache Software Foundation
- Paul Fremantle, Member, Apache Software Foundation
- Chris A. Mattmann, Member, Apache Software Foundation

## Sponsoring Entity

Apache Incubator Project.