

# ScienceSoft Workshop

CERN, 8 February 2012

## Participants (28):

AIFTIMIEI, Doina Cristina	Istituto Nazionale Fisica Nucleare (IT)	Padova ITALY
AIMAR, Alberto	CERN	Geneva SWITZERLAND
BEGIN, Marc-Elian	SixSq	Geneva SWITZERLAND
Dr. BRESSAN, Beatrice	Case Western Reserve University (US)	Geneva SWITZERLAND
BRYCE, Ciaran	Shared Object	Geneva SWITZERLAND
CECCANTI, Andrea	Istituto Nazionale Fisica Nucleare (IT)	Bologna ITALY
Dr. DI MEGLIO, Alberto	CERN	Geneva SWITZERLAND
ESTRELLA CAINGLET, Florida	CERN	Geneva SWITZERLAND
Dr. FOSTER, David	CERN	Geneva SWITZERLAND
Dr. FUHRMANN, Patrick	Deutsches Elektronen-Synchrotron (DE)	Hamburg GERMANY
GIORGIO, Emidlo	Istituto Nazionale Fisica Nucleare (IT)	Catania ITALY
Dr. HAKALY, Robert	D-Core Systems	Nyon SWITZERLAND
Mr. HOFMANN, Matthias	TU Dortmund University	Dortmund GERMANY
JONES, Bob	CERN	Geneva SWITZERLAND
KATZ, Daniel S.	University of Chicago	Chicago, IL USA
KONYA, Balazs	Lund University (SE)	Lund SWEDEN
Prof. LAGANA, Antonio	Università di Perugia	Perugia ITALY
Dr. LITMAATH, Maarten	CERN	Geneva SWITZERLAND
Dr. LÖFFLER, Frank	Louisiana State University	Baton Rouge USA
Mr. MANSET, David	MAAT France	Archamps FRANCE
Mr. MANZI, Andrea	CERN	Geneva SWITZERLAND
Prof. MARCO, Jesus	Universidad de Cantabria (ES) / CSIC	Santander SPAIN
Dr. MAZZUCATO, Mirco	INFN Sez. di Padova	Padova ITALY
Ms. NARDI, Nadia	Engineering Ingegneria Informatica spa	Roma ITALY
NEWHOUSE, Steven	EGL.eu	Amsterdam NETHERLANDS

NILSEN, Jon Kerr	University of Oslo (NO)	Oslo NORWAY
Dr. PAGANO, Pasquale	CNR – ISTI	Pisa ITALY
Mr. RIEDEL, Morris	JUELICH SUPERCOMPUTING CENTRE	JUELICH GERMANY

## Agenda:

The agenda is available at:

<https://indico.cern.ch/conferenceDisplay.py?ovw=True&confId=160503>

## Minutes

Alberto opens the workshop with a brief description of the agenda and the logistics. The format of the workshop is based on an open discussion of topics with as little presentation time as possible.

The proposed topics are:

- Setting the community scene: who is who
- ScienceSoft: vision, scope, features and services
- Organization and mandate
- Operation and funding mechanisms
- Timeline and next steps
- Summary

### *Setting the community scene: who is who*

Alberto shows some introductory slides to describe the background of the ScienceSoft initiative and what has happened since the initial discussions in September 2011 (the slides are available on the agenda page). A discussion follows on how the initiative positions itself with respect to existing communities. It is necessary to define what this community is going to contribute on top of what the larger open source and scientific communities are already doing.

The scope of the initiative should be better clarified. In addition also the scope of the initial participation in the initiative must be defined. What is this initial group is going to do? From the existing ScienceSoft document it's difficult to understand what this is. First define what this is or what this could be, and then we can see whether the community at large may be interested. What is the unique problem you are trying to solve? Alberto describes various problems that have been reported by people interviewed about their current experience with developing and using software within EMI and other DCI and research projects. ScienceSoft is not limited to a particular type of software, not limited to middleware. Also this is not just about grid, but also about cloud and high-performance computing. Many people involved in software production and usage in the scientific research projects at all levels share similar problems and can benefit from common solutions.

What is the motivation for creating this initiative? If the biggest community of EMI (WLCG) is not willing to continue with the current software we are producing, what is the motivation for making this permanent? Looking at this just from the EMI point of view is very restrictive, ScienceSoft is not maintaining the current EMI middleware, but rather to find what problems hinder the establishment of more active open source communities for scientific research. If WLCG doesn't want to continue using the current EMI middleware, they will use something else, but many of the problems presented in the slides still apply. In addition other communities (e.g. Computational Chemistry) are willing to pay in kind for continued access to the infrastructure and its services and are already setting up solutions to problems like giving credits for the contributions received.

### *ScienceSoft: vision, scope, features and services*

ScienceSoft can start from simple solutions. It is for example simple and easy to put together a step 0 for what we want to build on. Initially we could think of this as a kind of software forge. The question is what do we need in addition to that?

The creation a software marketplace is an interesting possibility to match offer and demand for software and services.

ScienceSoft could be like an "OpenAIRE for software" where software can be treated as a type of publication. The possibility of assigning identifier to software so it can be cited is a need felt by many developers. It could help referencing software in published papers, giving full academic recognition to software products.

Another benefit would be the increase in the openness of the software production and the possibility for users to switch technologies if they can compare and evaluate different options also based on other users' evaluations.

However, what is the market for the services we are offering? Is there really a demand? We need to clearly answers questions like, if I am a software provider, what do I gain? If I am a user, what do I gain?

We also need to understand the role of funding agencies. If the national governments are willing to fund science, there is no need for this initiative. This is not entirely true. Funders can use the citations and software ratings to evaluate funding of software development proposals. In addition some funders are only supporting open source e.g. NSF (NSF is also working on software sustainability, worth creating links to address common problems). Also the EC has a policy of supporting open source development.

What would the structure be for this initiative? We had a common coordination group for the technical activities, but what is the commitment, who is interested to take part in this how? Who will maintain the software? In terms of software maintainability, in the Linux system maintainability is driven by the packages. The package maintainers are in charge of making sure the software can be properly released. What would be the equivalent role in ScienceSoft? Who looks after the quality of what gets released? ScienceSoft does not release software; it provides pointers, links to software. The contributors are still responsible for their software. However, one of the possible roles of ScienceSoft can be the drive

towards adopting proper open source release procedures from the leading operating system distributions.

What have we identified so far? We need to summarize:

0. To continue to provide what is available now beyond the duration of short-term projects
1. Move away/ not to be tied to EC funding
2. Keep alive common coordination groups for the technical activities
3. Dedicated to software production at large, not limited to middleware
4. Important to understand software usage. This should not be limited to download and sharing, but real measurements of how and where software is actually used
5. Weak connection between software developers and their users. Developers are not necessarily the ones interacting with users, need for more direct contacts and collaboration
6. Marketplace for software and service
7. Software rating/citation. This will help users switch to more mature software. Also include software performance metrics and other types of technical metrics
8. Need for peer-reviewed scientific software, similar to scientific publication. DOI-like system to allow software to be cited in scientific publications (see handle.net)
9. Limit fragmentation by finding what exists already and reusing as much as possible. Also limit fragmentation within research organisation by identifying commonalities across different proposals or projects
10. Reasonable level of coordination, control and regulation. People will not buy into something that will dictate what they should do. Regulation should put some level of quality control but should not discourage participation. Reducing fragmentation via coordination may reduce innovation. Balance between coordination and innovation is important.
11. It cannot be completely open. It should have reasonable criteria for participation e.g. using a valid OSI license should be a mandatory criterion. However, a lot of proprietary and commercial software is used in scientific research, how could this software be included? We could have different level of memberships for open source and non-open source members, with different "fees"
12. Create social "circles" around specific scientific topics and establish long-duration focus groups around themes of interest to a community or across several communities
13. Software catalogues are the base of most of the ScienceSoft features, it can also be extended to services and people

Using the concept of software catalogues, ScienceSoft could be the glue that binds the platform integrators with the software providers. Platform integrators could put together community specific software stacks using the information in the software catalogues. It can be envisaged to package the community stacks in pre-configured appliances to be shared within a community or across communities. There could be simple stacks to be used as building blocks for more complex or more specific appliances. Users could try directly a pre-configured, integrated stack using some kind of cloud service offered by community members through the ScienceSoft portal.

Other possible services to offer in addition to what has been discussed so far:

1. An incubation service is desirable for small projects or initial phases of larger projects. The issue is maintenance and support of the software and hardware. There could be integration with existing services like GitHub
2. Software integration and testing services could be offered through the portal, e.g. SixSq could offer access to their integration platform through this initiative with trial runs and a commercial service. The same model applies to any commercial provider of software related services.

### *Organization and mandate, Operation and funding mechanisms*

Funding is needed to take care of the infrastructure needs of the initiative. There are two phases, first to put the 'service' in place and then to keep it operational. We have to show credibility for funders to support this initiative. Can this be done only with voluntary contributions? It is proposed to configure this as a Foundation and use it to maintain the software to be provided. The two things are not the same. ScienceSoft is not about looking for funds to maintain the software, the ScienceSoft members have to have their own ways to fund themselves. ScienceSoft can help its members in finding evidence to be used when submitting proposals to the EC or other public funding agencies.

However, ScienceSoft itself can become a Foundation with a legal entity by itself. This would allow ScienceSoft to manage financial matters independently, for example it could participate in EC calls with proposals to develop parts of its services. It is agreed that establishing a Foundation now is too soon, first ScienceSoft has to prove it can provide useful services. However, the goal of it becoming a Foundation on the model of existing successful open source foundations is a valid one and the initial implementation should not prevent this from happening in the future.

How to ensure the commitment of participants to the ScienceSoft initiative? A discussion starts about establishing MoUs among participants, but the general opinion is that to build an open source initiative like ScienceSoft at the beginning there must be a very lightweight engagement from interested parties. Forcing MoUs at this stage when the ideas are not yet sufficiently defined would be too heavy.

### *Timeline and next steps, Summary*

The workshop is concluded with a discussion of the next steps:

1. Revise the existing ScienceSoft document to include points discussed at the workshop and circulate to participants for feedback (Alberto)
2. Produce a public version of the document that can be considered a sort of manifesto of the initiative (all)
3. Define a short-list of features and implementation steps and assign priorities (all)
4. Once priorities are assigned, issue calls for voluntary contribution in developing the features (all)
5. The general timeline is agreed to be as follows:
  - a. Between now and June 2012: define what ScienceSoft should be and design the most important features
  - b. Between July and December 2012 implement a prototype and invite test users
  - c. January 2013 onward run the ScienceSoft portal, iterate further design and implementation phases, look for funding sources

Next opportunities for meeting are OGF and EGI/EMI conference.