Software sustainability and The ScienceSoft Initiative

Alberto Di Meglio, CERN

1st ScienceSoft Workshop
08/02/2012

Welcome and Logistics

- Welcome

- Agenda

Format of the workshop

- Not a series of presentations, but an open discussion on a set of related topics

- At the end of the day, we will try to draw some conclusion

- Then iterate on the following weeks to produce a document describing an initiative (timelines to be discussed later)

EMI INFSO-RI-261611

Organization of Topics

Community → Vision, scope, services → Organization and mandate → Funding models → Timeline

How do we go about setting this up? What are the immediate next steps and the important milestones for the next months or years? Who can commit to the initiative and in what terms? What are the success criteria to be met and by when?

EMI INFSO-RI-261611

Organization of Topics

Community

Definition of an open source community
Does it apply to us?
Is there one or many?
Do they exist already or should be organized?
What problems or needs are they experiencing?

EMI INFSO-RI-261611

Sustainability

1. The capacity to endure

2. Long-term maintenance of responsibility, which has environmental, economic, and social dimensions

3. Good business sense

Be more efficient and competitive
Increase financial return and reduce risk
Attract and retain customers and employees
Strengthen community relations

EMI INFSO-RI-261611

Software Sustainability

- Three main sustainability models:
  - Option 1: Publicly funded projects
    - *Our main model so far, funds from EC or other public agencies*
  - Option 2: Community funded open source projects
    - *Funded (usually in effort) by the members for their own needs or through the provision of paid services*
  - Option 3: Sponsored open source projects
    - *Funded (in money and effort) by commercial companies as a business driver*

Software Sustainability

- Option 1 does not require a large active community

- Option 2 requires an active, motivated community

- Option 3 exploits an active community, which in turn receives services and benefits

Where are we?
Can we do something different? Do we need to?

EUROPEAN MIDDLEWARE INITIATIVE

Open source

- Most of the software is licensed under a valid OSI-compliant license

- OS is not just about the license or the code
  - Adoption of well-established release and distribution procedures and packaging formats
  - Distribution through mainstream operating systems or repositories
  - Bottom-up contribution process with high-level strategic coordination
  - Events and coordination activities and processes

Strong, active communities

- Communities do exist, but are not always interacting with each other
  - *Interaction between users, developers, service providers*
  - *Interaction between different scientific domains*
- Interaction often at project level, but at personal level is limited
- Very difficult to know who's using/doing what, how to contribute and how to get and give credit for it or rate the products and services (**honor system**)

Strong, active communities

- This prevents from establishing active communities and an efficient exchange of information

- In turn, the lack of visibly active, engaged communities makes sustainability difficult
  - How to justify the need for funding?
  - How to bring in commercial companies?
  - How to preserve useful software at the end of a project?

What brings us together?

- Can the common attribute of "doing scientific research" justify and require the establishment of an open source community around this attribute?

- What other common attributes or goals define our community?

Organization of Topics

Community → Vision, scope, services

Definition of an open source community

Given a definition of community, its composition, needs and goals, what services do their members need? What is the vision of an initiative that proposes to provide those services? Is it for everybody or for a subset or some other combination?

EMI INFSO-RI-261611

EUROPEAN MIDDLEWARE INITIATIVE

More interactions, exchange of information and knowledge sharing, higher efficiency and RoI

Developers

Users

EMI INFSO-RI-261611

Bottom-up Approach
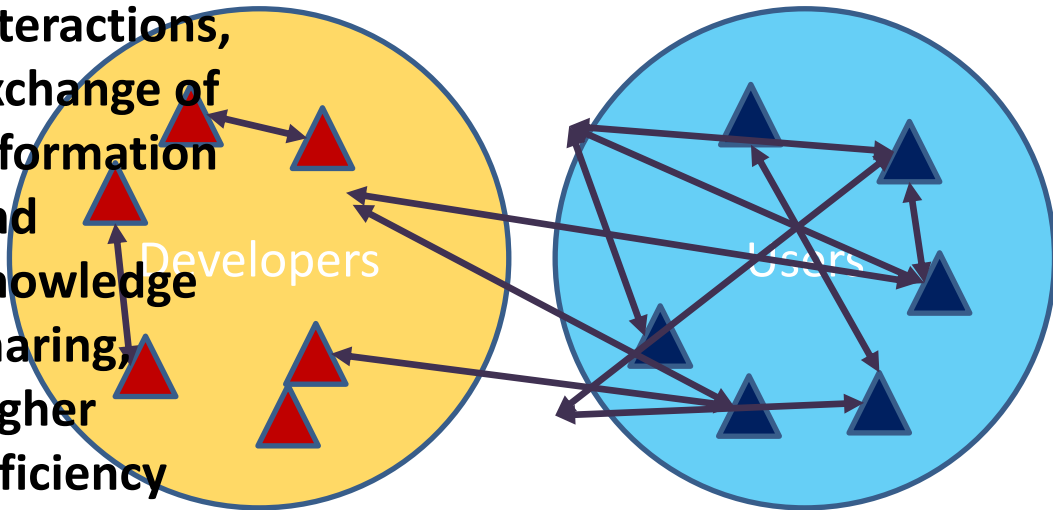
- Community-oriented not project-oriented
  - Committed contributors
  - Engaged users

- Interactions are stronger if supported by motivated individuals within the more general interests of Institutes or Companies

- Institutes have of course to have policies in place to encourage and reward this behaviour

Main Services

- Collect and publish information about software
  - who develops it, who uses it, what licences are used, who likes or dislikes it, etc.
  - stats by developer, by user, by Institute, by geographical region, by scientific community/domain, etc.

Vision, scope, services

EMI INFSO-RI-261611

Main Services

- Provide info and services to Institutes and companies to

  – Assess their software production (how much does it cost, what legal requirements do I have, is any optimization possible?)

  – Promote their worth (how much is my software used, by whom? Can I better support my funding requests? Increase users?)

  – Access funding resources

Main Services

- Provide communication channels
  - among developers, among users, between users and developers, special-interest groups, etc.

- Organize and promote events
  - Conferences, workshops, contributors gatherings

- Provide an open marketplace of products, services and people in the context of scientific research
  - Act as a broker between demand and supply

EMI INFSO-RI-261611

Benefits

- Enhance the motivational aspects of software development and usage
  - Software is well advertised, its use is known and acknowledged
  - Users can share experiences, get help and rate the software
  - Developers can be more responsive to user needs
- Increase cross-domain interaction
  - A solution/problem found in one domain or community doesn't need be found again and again

Benefits

- Increase visibility and opportunities:
  - Being able to show that software is used and is well-rated helps with sustainability (getting funds)
  - Promoting one's own skills and competencies, improves job opportunities

Benefits

- Improve opportunities for commercial usage
  - Companies get knowledge of potential markets
  - Can propose additional services to interested users
  - Suitably equipped Research Institutes could also provide additional value-added services

Organization of Topics

Community → Vision, scope, services → Organization and mandate

Definition of an open source community

Given a definition of community, its composition

Given a definition of the community and its needs, a vision and a set of services and function, what are the most suitable organization, mandate, structure, business model for an entity to coordinate such a community and provide the desired services?

Successful examples

- Open source foundations or communities already exist

- They fall into one of four "scope" categories
  - Technological: Apache, Drupal, Eclipse, OpenStack, etc.
  - Operating System: Fedora, Debian, etc.
  - General purpose: SourceForge, GitHub, etc.
  - Domain-specific: e.g. Nanohub

EMI INFSO-RI-261611

Successful examples

- Most open source initiatives are independent not-for-profit legal entities
  - Required to manage funds directly
- Do we need that?
- What other models are possible?
  - But consider the follow-up discussion on funding models and the possible constraints or requirements

EMI INFSO-RI-261611

Organization of Topics

Community → Vision, scope, services → Organization and mandate → Funding models

Definition of an open source community
Given a definition of community, its composition
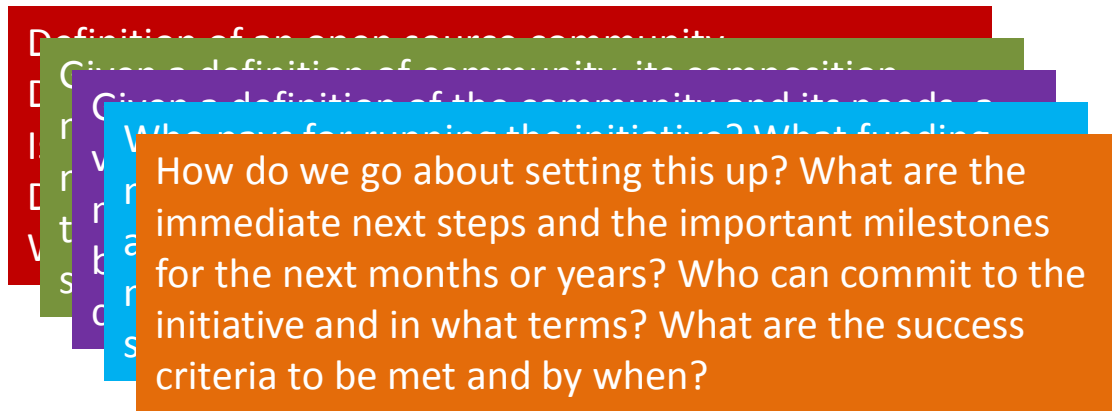Given a definition of the community and its needs, a

Who pays for running the initiative? What funding models are realistically possible in the short, medium and long term? What constraints do different funding models impose on the initiative organization and structure?

EMI INFSO-RI-261611

Funding models

- In terms of funding models, the successful open source initiatives usually are:
  - Self-funding through the sale of services (or different membership levels)
  - Sponsored by commercial companies as business drivers or incubators
- Do either apply to us?

EMI INFSO-RI-261611

Funding models

- Should the funding of the initiative be coupled or decoupled from the contributors funding?

- What about membership?
  - Should becoming a contributor be free (I'd say so or we may end up with an empty community)
  - Can we consider levels of membership for different types of members (e.g. academic vs. commercial)?

Funding models

- Should the contributors or members pay for the services to be provided? Or maybe just for special added-value services (Advertising? Custom reports or information packages? Organization of dedicated events?)

Organization of Topics

Community → Vision, scope, services → Organization and mandate → Funding models → Timeline

How do we go about setting this up? What are the immediate next steps and the important milestones for the next months or years? Who can commit to the initiative and in what terms? What are the success criteria to be met and by when?

EMI INFSO-RI-261611

Current Status

- First draft of a document describing the current ideas and discussion
  - Originated within EMI
  - Distributed to user and developers outside EMI
- Group of people being formed to improve the document
  - EMI, EGI, StratusLab, iMarine, OpenAIRE, Maat, SixSq, dCore Systems, SysFera, (PRACE), (HEP), …
  - work out more details of possble mandate, scope, functions, governance, funding model, etc.

Current Status

- Tentative name and motto:
  - ScienceSoft: Open Software for Open Science
- Domain name registered:
  - sciencesoft.org
- Place-holder web site:
  - http://sciencesoft.web.cern.ch

EMI INFSO-RI-261611

Timelines

- Survey being designed
  - to be sent to developers and users communities
  - collect ideas and feedback on the proposed initiative, whether there is a perceived need for it, what it should do

- Expected to be sent out at the beginning of February for 3 or 4 weeks

Timelines

- Presentation in Amsterdam on Jan 25th at the EGI workshop on sustainability

- A first workshop being organised at CERN on 8th February to discuss about the initiative and refine it

- Regular workshops during the year

- Official presentations/sessions at ISGC, EGI/EMI Conference, OGF, CHEP, …

Timelines

- Implementation of
  - Organization
  - Technical features

- Start in the second half if 2012, make it operational in 2013

- Incremental approach

Timelines

- How do we organize this initiative?
  - Working groups on specific topics to be better analysed?
  - Contribution of individual experts?
  - External consultants on open source?

- Who can commit to:
  - Take part in its setup
  - Use it as test users
  - Promote it

EMI INFSO-RI-261611

- Software, service, people catalogs
  - Tagging, taxonomies, technical metrics
- Marketplace for products, services, people
  - Match demand and offer
- Collection of statistics on:
  - Usage, deployment, geographical distribution
- Honour system
  - Who likes what and why, rating system
- Citation system, software referenced in papers
- Hooks into additional added-value services:
  - Support
  - Technical services (testing, deployment, etc)
- Platform integration support
  - Definition of community-specific profiles
  - Validation services (third-party)

- Organization of events
- Sub-community and interest groups (technical coordination at this level?)
- Admission/registration criteria:
  - By license?
  - By release/distribution method?
  - Peer-review of contributed projects?
  - Focus on "scientific research"
  - Required documentation or documented QA process
- Can this be used at a policy level to access fundings? (OpenAIRE experience)
- Start operations now, but work towards establishing a legal entity within a year timeframe
- Establish an MoU (to do what?)