SciencePAD

• Platforms, Applications, Data for Science

• Collaboration among a number of research centres, research projects and companies

• Its goals are to investigate:
  • The requirements of scientific communities in terms of software information management
  • The formalization of such information and the integration with other digital objects (publications, people, datasets, etc.)
  • The requirements for long-term preservation and re-use especially related to data
  • The prototype a data-driven “Software as a Service” platform for scientific research
Background

• SciencePAD (formerly known as ScienceSoft) started as a lightweight investigation in Sep 2011

• Involved a number of research and computing projects (EMI, iMarine, StratusLab, OpenAIRE, IGE), infrastructures (EGI, WLCG) and companies (Maat, Sixsquare, SharedObjects, DCore)

• Asked questions about the issues and challenges in preserving, discovering, accessing software in relation to scientific research
<table>
<thead>
<tr>
<th>Problem area</th>
<th>Challenge description</th>
<th>Affected stakeholder groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty in identifying software and related activities</td>
<td>Limited or complex ways of finding what exists already</td>
<td>Researchers, software developers</td>
</tr>
<tr>
<td></td>
<td>Lack of visibility and recognition of development activities</td>
<td>Software engineers, developers</td>
</tr>
<tr>
<td>Difficulty in evaluating software</td>
<td>Lack of consistent real usage information and impact assessment</td>
<td>Development projects, infrastructure managers, funding bodies</td>
</tr>
<tr>
<td></td>
<td>Limited access to other users’ experience</td>
<td>Researchers, infrastructure managers and operators</td>
</tr>
<tr>
<td>Difficulty in leveraging existing software through re-use</td>
<td>Lack of continuity in development, coordination of software</td>
<td>Software engineers, developers, software development projects</td>
</tr>
<tr>
<td></td>
<td>Lack of continuity in support of software</td>
<td>Researchers, research projects</td>
</tr>
<tr>
<td></td>
<td>Non-optimal communication between users and developers</td>
<td>Researchers and software developers, R&amp;D projects, infrastructure managers and operators</td>
</tr>
<tr>
<td>Difficulty in justifying or proposing business case for development of new software</td>
<td>No way of assessing the user “market” and potential revenues</td>
<td>SMEs</td>
</tr>
<tr>
<td></td>
<td>Limited possibilities of influencing the production of software</td>
<td>Researchers, infrastructure managers and operators</td>
</tr>
<tr>
<td></td>
<td>Limited commercial exploitation and support for technology transfer</td>
<td>Funding bodies, software engineers, SMEs</td>
</tr>
</tbody>
</table>
Identified areas of work

- Software; links to services, people, organizations, publications, datasets
- Reliable entry points, persistent IDs

- Cross-disciplinary activities, citations, industry, incubation, impact assessment

- Interactivity, shared experiences, evaluation and ratings, collaboration tools, open source

- Software as a Service, community platforms, scientific results validation, data access preservation, support, consultancy,

Registries

Services

Knowledge sharing

Communities
Wide collaboration
Current status

• Contacted most of the identified relevant projects and activities
• Contacted different scientific communities via some of their major Institutes and Research Centres

• Initial activities:
  • SciencePAD software registry ([http://sciencepad.org](http://sciencepad.org))
  • Workshop on Persistent IDs for software (Jan 2013)
  • Workshop on Software Registries and Metadata (Apr 2013)
  • Submitted funding proposal for a support action to coordinate the collaboration activities (outcome expected in May 2013)
SciencePAD Goals

Software Developers

Recognition

Platforms

Applications

Open Source

Software Registry

Data

Collaboration networks

Reliability

Efficiency

Scientific researchers
SciencePAD Activities

- Researchers Applications
- Discovery
- Software Registry
- Authorship Citations
- Publications Patents
- Datasets Scientific results
- Reproducibility Preservation
- Registration
- Developers
- on-demand SaaS
- "Click ‘n’ cite"
- "Click ‘n’ cite"
- Publications Patents
- "Click ‘n’ cite"
- "Click ‘n’ cite"
- "Click ‘n’ cite"
“Of all the different elements of a successful long-term data preservation strategy, by far the most complex – and least studied to date – is that of maintaining the software and associated environment usable for long periods and in adapting it to changes, particularly in the period when the original authors and experts are no longer available.”

SciencePAD and DPHEP

- Collect and analyse requirements from HEP community
- Work on a strategy to formalize and preserve software information and related configuration and environment properties
- Create and manage links between software, data and authors
- Ease use, porting and adaptation by making the information clearly available, discoverable and accessible
- Prototype an “on-demand” SaaS platform to perform data validation, re-use, benchmarking.
Next steps

• Keep collecting more information and feedback from developers, research communities, and other interested parties

• Discuss with other projects about common activities and extending/reusing/integrating methodologies and functionality

• Prepare for the more formal operational phase
  • Design of generic and community-specific software registries
  • Software metadata, formats, ontologies (generic and community-specific)
  • Prototypes of software services with interested communities
SciencePAD
Platforms · Applications · Data

http://sciencepad.org