



# SciencePAD

## Open Software for Open Science

Alberto Di Meglio – CERN

- **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

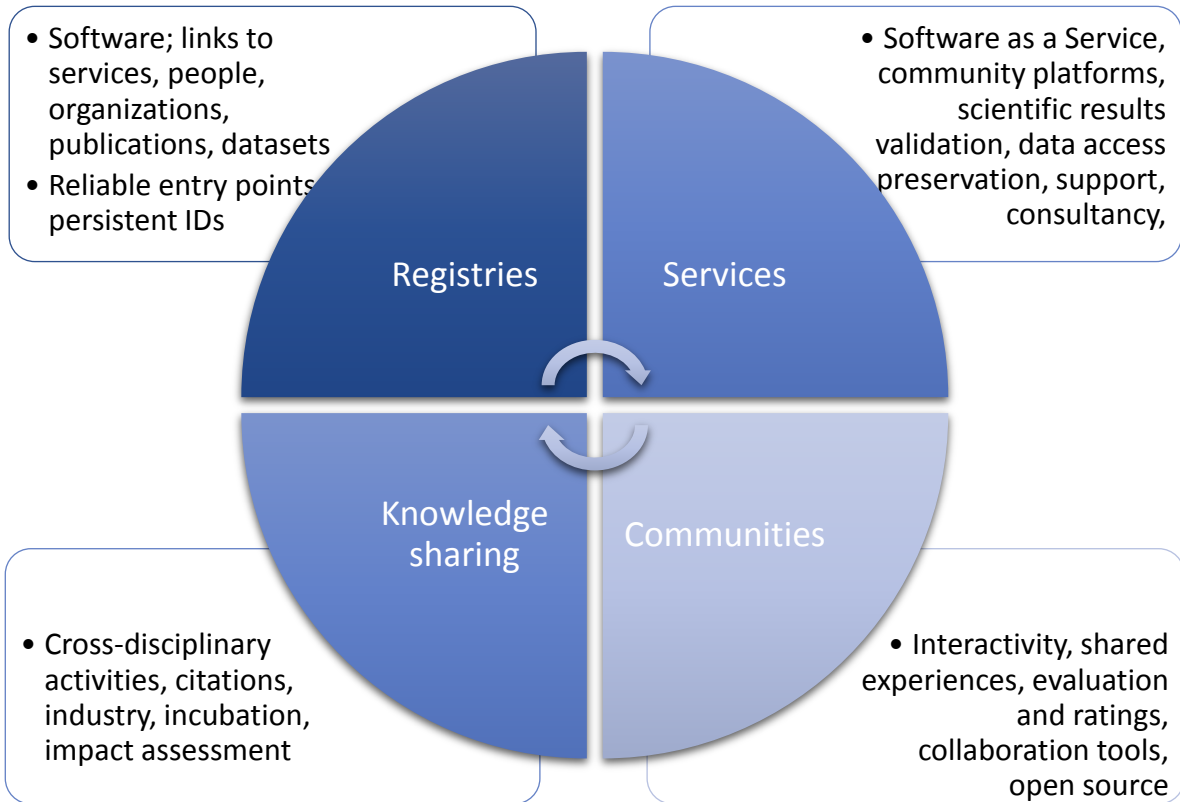
# Background

- SciencePAD (formerly know 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

# Critical issues

Problem area	Challenge description	Affected stakeholder groups
<b>Difficulty in identifying software and related activities</b>	Limited or complex ways of finding what exists already	Researchers, software developers
	Lack of visibility and recognition of development activities	Software engineers, developers
<b>Difficulty in evaluating software</b>	Lack of consistent real usage information and impact assessment	Development projects, infrastructure managers, funding bodies
	Limited access to other users' experience	Researchers, infrastructure managers and operators
<b>Difficulty in leveraging existing software through re-use</b>	Lack of continuity in development, coordination of software	Software engineers, developers, software development projects
	Lack of continuity in support of software	Researchers, research projects
	Non-optimal communication between users and developers	Researchers and software developers, R&D projects, infrastructure managers and operators
<b>Difficulty in justifying or proposing business case for development of new software</b>	No way of assessing the user "market" and potential revenues	SMEs
	Limited possibilities of influencing the production of software	Researchers, infrastructure managers and operators
	Limited commercial exploitation and support for technology transfer	Funding bodies, software engineers, SMEs

# Identified areas of work



# Existing Activities



# Wide collaboration

ORCID

INVENIO

CRISP  
The Cluster of Research Infrastructures  
for Synergies in Physics

pandata

OpenAIRE  
Open Access Infrastructure for Research in Europe



ILL  
NEUTRONS  
FOR SCIENCE

EM  
EUROPEAN MICROELECTRONICS  
INITIATIVE



hloh

blackduck

EMBL-EBI

JISC

EXM

OLLIANCE  
Consulting  
A Division of Black Duck

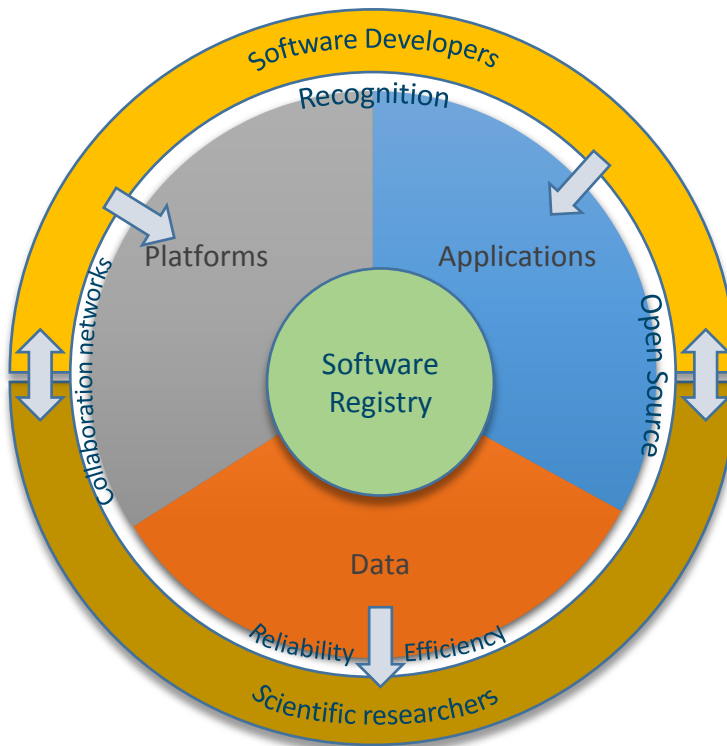
BioMedBridges

# Current status

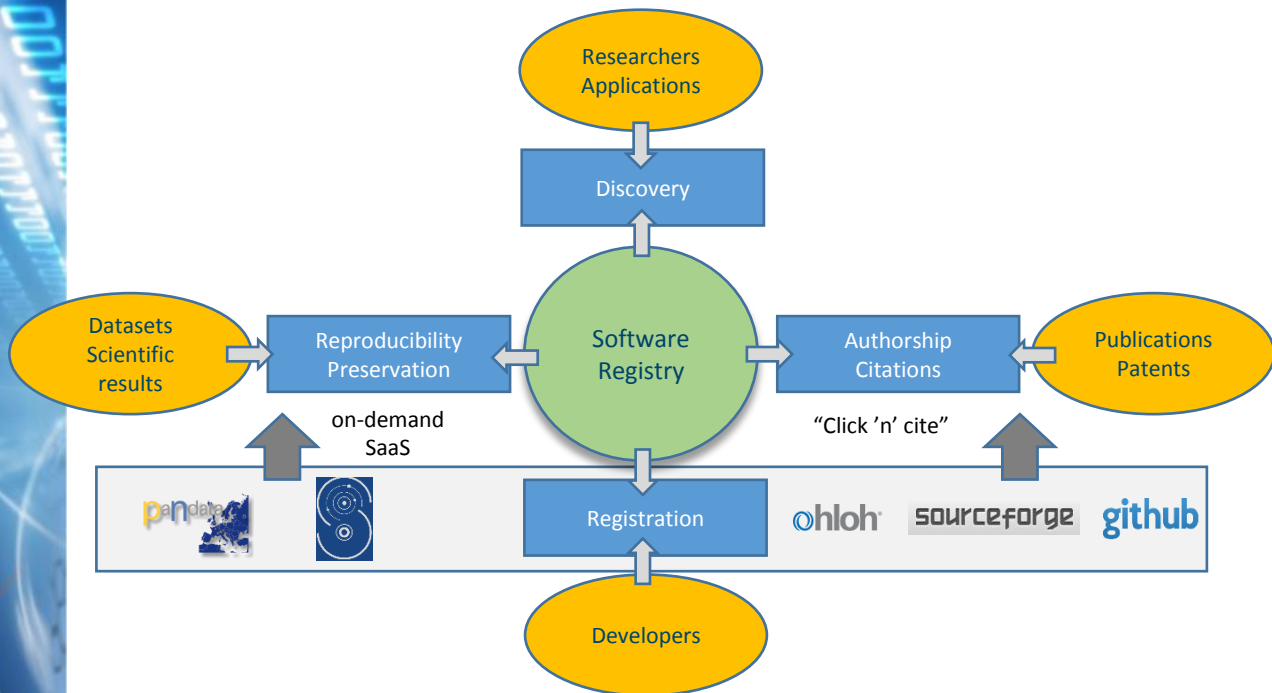
- Contacted most of the identified relevant projects and activities
- Contacted different scientific communities via some of their major Institutes and Research Centres
- Initial activities:
  - SciencePAD software registry (<http://sciencepad.org>)
  - Workshop on Persistent IDs for software (Jan 2013)
  - 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



# SciencePAD Activities



*“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.”*

**Jamie Shiers, Long-Term Data Preservation in High Energy Physics: A 2020 Vision, DPHEP, Feb 2013**

# 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



<http://sciencepad.org>



SciencePAD is an initiative by EMI, partially funded by the European Commission under Grant Agreement RI-261611