Since it started in 2006, the Workshop on Public Data about Software Development (WoPDaSD) has grown as a venue where researchers can foster the analysis of publicly available data sources and the exchange of data between different research groups, specifically in the FLOSS (free, libre, open source software) area.

Those data sources are enabling researchers in many ways: reproduction of results is easier, projects based on larger analyses are now possible, results are now traceable and explainable with a common vocabulary, it is possible to export to semantic web browsers, it is quicker to obtain results, and the retrieval process is simplified for non-experts.

This workshop is aimed specifically at three different possible types of research:

- Analysis of some data collections about software development.
- Retrieval process and exchange formats of public available data collections about software development.
- Data mining activities and new retrieval tools.

With respect to the first of these, projects such as FLOSSmole and FLOSSMetrics are compiling huge quantities of data about FLOSS development. The availability of these data in formats suitable for analysis by third parties are enabling researchers to focus on the study of the data, and not on data retrieval activities. This is fortunate, since data retrieval from software development repositories is becoming more and more complex, especially when reliable and detailed information from many projects is needed.

Studies and research results based on these kinds of datasets have already been presented in workshops, conferences and journals, but rarely the focus is on how to benefit from the datasets, or on the problems derived from their use. In addition, the details of how to use the datasets for different purposes, or specific results from their analysis, are not published elsewhere.

This workshop has been for five years a place to discuss all these topics, and to present research results developed with these ideas in mind: how these large datasets about FLOSS software development are retrieved, how can they be analyzed and mined, how they can be published, exchanged and extended, which lessons are we learning from their use, and which results are being obtained from their analysis.

With this background, it seemed natural
to join efforts with some leading publication in the field, to publish extended, updated and enhanced versions of the best papers presented in the workshop. The result is this special issue, which includes the following papers:

- “Repositories with Public Data about Software Development” by Jesús González Barahona, Megan Squire and Daniel Izquierdo-Cortazar aims to explain, from an introductory point of view, what a repository of repositories means and how this is increasing its importance thanks to fields of study such as software evolution. Indeed, most of the studies in that context have to deal with the data retrieval process. The authors explain the importance of those meta-repositories and focus on two of them, FLOSSMole and FLOSSMetrics, providing examples.

- “Flat for the few, steep for the many: Structural cohesion and rich-club effect as measures of hierarchy and control in FLOSS communities” by Guido Conaldi, addresses an empirical approach to the social network analysis using directly data from the FLOSSMetrics dataset. He aims to explain the paradox between the decentralized communities, working around the world and the real hierarchical structure that they show after being studied. Conaldi proposes a measure of structural cohesion based on network node connectivity to check if sub-groups of developers are equally centralized or are not.

- “Weaving a Semantic Web across OSS repositories: unleashing a new potential for academia and practice” by Olivier Berger, Valentin Vlasceanu, Christian Bac, Quang Vu Dang and Stéphane Lauriere investigates the potential of using semantic web technologies in navigating among different bug tracking systems scattered in the FLOSS ecosystem. This idea could help unify the way the bug trackers work; by having a centralized way to help, for instance, researchers can avoid a fragmented or unrealistic view of the bugs across projects.

- “Impact of programming language fragmentation on developer productivity: A SourceForge empirical study” by Jonathan L. Krein, Alexander C. MacLean, Charles D. Knutson, Daniel P. Delorey and Dennis L. Eggett, presents a metric, language entropy, as a way to characterize the effort of developers at the level of different programming languages. They found a correlation between the language entropy and the size of the contributions based on a population of 500 projects taken from SourceForge.

In summary, this special issue is the result of many years of work and the invaluable help of the WoPDaSD Program Committee through their effort in the reviewing papers and selecting the most adequate for each of the edition of the workshop. Many thanks to all the authors and the reviewers for their hard work. We sincerely appreciate the excellent support given by the OSS organizers and the support and cooperation of Stefan Koch, Editor-in-Chief of the International Journal of Open Source Software and Processes.

Jesús M. González-Barahona
Daniel Izquierdo Cortázar
Megan Squire
Guest Editors
JOSSP
Jesus M. Gonzalez-Barahona teaches and researches in Universidad Rey Juan Carlos, Móstoles (Spain). He started to be involved in the promotion of libre software in 1991. Since then, he has carried on several activities in this area, including the organization of seminars and courses, and the participation in working groups on libre software, both at the Spanish and European levels. Currently he collaborates with several libre software projects (including Debian) and associations, writes in several media about topics related to libre software, and consults for companies and public administrations on issues related to their strategy on these topics. His research interests include understanding libre software development, where he has published several papers, and is participating in some international research projects. He is also one of the promoters of the idea of an European master program on libre software, and has specific interest in the education in that area.

Daniel Izquierdo-Cortazar is a PhD student at the Universidad Rey Juan Carlos in Móstoles, Spain. He earned a degree in computer science from the same U. and obtained his master degree in computer networks and computer science systems in 2006. His research work is centered in the assesment of libre software communities from an engineering point of view and especially with regard to quantitative and empirical issues. Right now he holds a grant from the Universidad Rey Juan Carlos to dedicate part of his time to his PhD's thesis. He is also involved in European-funded projects such as QualOSS or FLOSSWorld. He also teaches in Universidad Rey Juan Carlos, Móstoles (Spain) in the Master on Free Software.

Megan Squire is an associate professor in the Department of Computing Sciences at Elon U. (North Carolina, USA). Her primary research focus is on data mining and large database systems, particularly for open source and software engineering data. She is one of the leaders of the FLOSSmole project, which is a collection of tools designed to gather and analyze data about open source software development.