



**Workshop EPFL-Inria
7 et 8 février 2017, Lausanne**

Speaker: Roberto Di Cosmo (Inria and University of Paris Diderot, France)

« What would you do with billions of source code files? Challenges and opportunities from software archival. »

Abstract:

From ten years of working on analysing the characteristics of large open source software repositories, we draw some lessons on the key properties we need for this kind of software engineering large scale studies. This led us to launching Software Heritage, the most ambitious project to date to build a universal source code software knowledge base. The size of this archive is daunting, with billions of unique source code files, coming from tens of millions of repositories. We will highlight some of the new challenges, both organisational and scientific, that Software Heritage brings up.

Roberto Di Cosmo holds a PhD in Computer Science and is currently Computer Science professor at University Paris Diderot, after teaching for almost a decade at Ecole Normale Supérieure in Paris, and spending a few years at INRIA.

He has been actively involved in research in theoretical computing, specifically in functional programming, parallel and distributed programming, the semantics of programming languages, type systems, rewriting and linear logic. He focus now on new scientific problems posed by the general adoption of Free Software, with a particular focus on static analysis of large software collections, that were at the core of the european reseach project Mancoosi.

He created the Free Software thematic group of Systematic in October 2007, and since 2010 he is director of IRILL, a research structure dedicated to Free and Open Source Software quality.

In 2016, he founded and directs Software Heritage, an initiative to build the universal archive of all the source code publicly available.

Learn more at <http://www.dicosmo.org>