"Preserving source code:

challenges and opportunities for the reproductibility of science"

Stefano Zacchiroli
(Università Paris - VII)

Giovedì 25 maggio 2017, ore 11:00 – 12:30
Sala riunioni n. 1 (3° piano) - DAUIN – Politecnico di Torino

Host: prof. Juan Carlos De Martin (DAUIN), demartin@polito.it

A vast amount of modern scientific and technological knowledge relies on the software that we have been collectively writing: deep knowledge from fields like mathematics, physics, chemistry, biology, medicine, finance, and social sciences is inextricably embodied into complex software systems, which model it at a level of detail that goes way beyond that of the usual scientific publications. Preserving this body of software is hence of paramount importance to preserve our knowledge. It is a necessary prerequisite to allow the replication of experiments, which is the foundation of the scientific method, as well as to ensure our ability to modify and correct the software components that are constantly being incorporated into critical systems that need to stay in production for decades.

In this talk we will review the challenges and opportunities we are facing, and discuss the role of long term archival of the entire body of Free/Open Source Software (FOSS) as its key enabler. We will present the Software Heritage project, launched in June 2017, which has the ambitious mission of collecting, preserving, and sharing all publicly available source code. Forever. Software Heritage has developed a uniform data model, based on a Merkle graph structure, capable of storing in a compact and fully deduplicated form all publicly available source code together with its development history, as captured by state-of-the art version control systems. Software Heritage is already the largest source code archive in existence, having collected more than 3 billion unique source files 700 million commits, coming from more than 50 million FOSS development projects.

Stefano Zacchiroli is Associate Professor of Computer Science at University Paris Diderot on leave at Inria. His research interests span formal methods, software preservation, and Free/Open Source Software engineering. He is co-founder and current CTO of the Software Heritage project. He is an official member of the Debian Project since 2001, where he was elected to serve as Debian Project Leader for 3 terms in a row over the period 2010-2013. He is a former director of the Open Source Initiative (OSI) and recipient of the 2015 O'Reilly Open Source Award.

https://www.softwareheritage.org/