

Subject: [11th_chairs_mtg] Updates on the activities related to software (SSC IG and SCID WG)
Date: Wednesday, June 26, 2019 at 11:55:49 AM Mountain Daylight Time
From: roberto@dicosmo.org@rda-groups.org on behalf of rdicosmo
To: 11th_chairs_mtg@rda-groups.org
CC: roberto@dicosmo.org, Daniel S. Katz, Martin Fenner
Attachments: ATT00001.htm

Dear fellow IG/WG chairs,

I'm sharing below a summary of updates on the current activities around software that I would have loved to discuss with you: I hope this will help in your meeting.

All the best

--

Roberto
High level summary

- + there is growing interest worldwide in software as a first class research output, and the number of subscribers in the IG/WG groups show that this subject is attracting interest in RDA too (~100 subscribers to the IG, 65 for the WG);
- + there are a variety of communities sharing interest in research software, so we should strive to join forces as much as possible, to avoid dispersion of efforts, and improve the chances of seeing our outputs adopted
- + as a concrete step in this direction, the Source Code Identification WG has been created as a joint effort with Force 11, and kicked off at the P10 Philadelphia plenary and we have decided on the following plan:
 - kickoff at RDA 13 (Philadelphia, April 2019) (done)
 - second meeting at Force 11 (Edinburgh, 16-17 October 2019)
 - third meeting at RDA 15 (April 2020)
 - 4th and final meeting + outcomes: October 2020 both at Force 11 and RDA
- + research software has attracted a broader attention over the past years, but there are places where it has been at the centerstage for decades: we need to learn from their experience and best practice

Notes from the kickoff of the Source Code Identification Working Group

The Kickoff of the WG was quite intense, and here are a few takeouts:

- identifying source code is a complex issue
- there are at least two distinct main objectives for source code identification: reproducibility and credit; each has clearly different timescales and requirements

- we reviewed the general framework for identifiers systems that distinguishes DIOs and IDOs (see <https://hal.archives-ouvertes.fr/hal-01865790/>); there was general agreements on the following:
 - + IDOs are necessary for reproducibility,
 - + DIOs are needed for credit and citation
 - + both are needed in a scholarly environment

- we reviewed four different existing approaches that provided ground for thought
 - 1) Software Heritage and its IDOs (the SWH-IDs), for reproducibility/traceability
 - 2) the ASCL registry, with its DIOs and the detailed curation process
 - 3) the swmath.org approach to software identification (DIOs with curated metadata extracted from publications)
 - 4) the moderated scientific software deposit in Software Heritage via the HAL open access portal (curated metadata via moderation, with DIOs for the metadata and IDOs for the traceability of software, that can evolve asynchronously)

- we drew some important lessons from this work :
 - IDOs allow to identify a precise version of a software project, and require no source of authority; for reproducibility, that's the way to go

 - for attribution, credit, and in general the metadata associated to a software project, we need a registry, and DIOs to identify the proper entries

 - moderation of metadata is essential to get the quality results needed for proper attribution and credit in the scholarly world

 - there are many decisions that go into creating proper metadata, and these require a source of authority, which is not necessarily the owner of the software project

- we concluded that there is a need to learn more about the source of authority, and the methods used to identify software that exist already in the various institutions.

--

Roberto Di Cosmo

Computer Science Professor
(on leave at INRIA from IRIF/University Paris Diderot)

Director
Software Heritage <https://www.softwareheritage.org>
INRIA
Bureau C328 E-mail : roberto@dicosmo.org
2, Rue Simone Iff Web page : <http://www.dicosmo.org>
CS 42112 Twitter : <http://twitter.com/rdicosmo>
75589 Paris Cedex 12 Tel : +33 1 80 49 44 42

GPG fingerprint 2931 20CE 3A5A 5390 98EC 8BFC FCCA C3BE 39CB 12D3

