Real-time Software Citations – #OpenSciFi Horizons

Real-time software citations that can load software into a VM and render a cited research data set in the browser.

The question is then what kind of future science scenarios open up? Where does this take us?

Future scenario:

(in a not to distant future — as this model has been in use for a few decades, for example by packet management systems like Debian's Dpkg, etc.)

A researcher finds a reference to a 'research data set' with an accompanying 'software citation'.

Systems are advanced to a point where the researcher can load the research data from a 'persistent identifier' (PID) link in the browser, and the browser will locate and load the exact version of the cited software into a 'virtual machine' (VM) of some sorts, allowing full use of the 'research data set' in the browser.

The hypothetical workflow:

1. Software is stored in a Git repository with 'Continuous Validation' (CI) running.
2. A 'software citation' file is in the repository.
3. 'Research data' is publishing, having been validated using a 'content version' of a CI system, and including a 'software citation' file for running the software.
4. A researcher finds the published 'research data' with the related 'software citation' for the software needed to run the data.
5. The researcher wants to check the 'research data' that they have found. In real-time and using a 'persistent identifiers' (PIDs) for the 'research data' and for the 'software citation' the researcher is able to load the data in their browser and have the software run via a 'virtual machine' and used in the browser.

Notes:

The tool chain – It can be imagined that something like @ProjectJupyter is used where data and code can be edited and run, with #binder https://mybinder.org/ or @CodeOceanHQ or @SWHeritage being extended to offer VM services to the software makers. The glue in the citations for the 'research data' and 'software citations', can be DOIs from @ZENODO_ORG that have been wrapped in a future version of the #CFF Citation File Format @stdruskat, or #codemeta https://github.com/codemeta @cboettig

Scenarios

2014 Ebola outbreak it was pointed out that after the epidemic subsided data was scattered, not available and you can assume the knowledge about software systems used also lost. See: *Dr Simon Hodson, Executive Director CODATA, France & Chair of the EC Expert Group on FAIR Data.* **Slide 17**, https://www.open-science-conference.eu/wp-content/uploads/2018/03/OSC2018_Hodson.pdf from Open Science Conf https://www.open-science-conference.eu/programme/ March 2018 LG S2.0.