

## **Workshop: From data curation to software curation: enhancing reproducibility and sustainability of data and software**

Date: Monday February 4, 1.30pm to 5.00pm. Lunch provided: 12:30-1:30

Location: The University of Melbourne, Arts West North Wing 453 - Collaborative eSeminar

This workshop will enable understanding of approaches to software curation, including its importance in enabling reproducible research. While a strong focus on data curation already exists, software curation is an emerging practice of equal importance. The workshop will present case studies to stimulate group discussion on how those engaged in facilitating reproducible research can support or actively engage with this.

The workshop will enable those concerned with reproducible research to understand issues in software curation. Software is fundamental to research, and plays a key role in creating and facilitating access to trusted, research outputs. "Software curation encompasses the active practices related to the creation, acquisition, appraisal and selection, description, transformation, preservation, storage, and dissemination/access/reuse of software over short- and long- periods of time." (Chassanoff, [Building a Model for Software Curation](#)).

This workshop seeks to assist participants to understand the challenges in software curation, the support and resources needed by researchers to facilitate software reproducibility and re-use, and to consider how those involved in digital curation can engage with researchers and research software engineers in support of software curation, including on aspects of software curation related to provenance.

Whilst discussions around reproducibility and open science have often focussed on research data, research software is critical to research. Nangia and Katz note that "a survey of academic faculty and staff at British universities found that 92% use research software, with 69% saying that their research would not be practical without it." Similarly, their Nature paper survey reveals that 32 of the 40 papers examined mention software, totalling 211 mentions of distinct pieces of software" ([Understanding Software in Research](#)). Other studies clearly show the need for increased understanding of how to curate software. In a 2017 PresQT research study, over 88% of respondents polled from among US National Science Foundation funded researchers and others likely to participate in data intensive research, reported that they create and/or use software, code, or scripts in their research. 76% of the 1700+ participants acknowledged that their software, code or scripts are needed for reproducing their results by third parties. Yet, less than 20% of respondents reported that they felt "more than moderately familiar with tools used to share, publish, cite and preserve data or software". (Gesing, Johnson, Meyers & Wang. [PresQT Needs Assessment](#)).

The workshop will use a range of speakers and interactive small and large group activities to enable participants to explore and understand:

- the importance of both software curation (alongside data curation) to achieving gains in reproducibility
- software curation best practice through case studies and policies
- software citation and aspects of software curation related to provenance
- challenges in software curation
- how those engaged with data curation can support or actively engage with these issues to cultivate and build capacity for collaborative efforts to collect, care for, and preserve software
- demos and discussions on how tools, platforms and projects like ReproZip, CodeOcean, Gigantum, and WholeTale can be used in the context of software curation and reproducible research.

## Draft workshop agenda

Time	Session
12:30-13:30	Lunch & networking
13:30-13:35	Welcome, introductions, housekeeping
13:35-13:50	Icebreaker activity: who's who in the room
13:50-14:10	From data curation to software curation: challenges and opportunities.
14:10-14:30	Software curation best practice: tools and platforms for source code <ul style="list-style-type: none"><li>• Case study 1: GitHub, Bit Bucket and more</li></ul>
14:30-15:00	Software curation best practice: tools and platforms for curating software in the context of publication and re-use. Possible case studies: <ul style="list-style-type: none"><li>• Case study 2: CodeOcean &amp; Jupyter</li><li>• Case study 3: WholeTale</li><li>• Case study 4: ReproZip</li></ul>
15:00-15:30	Afternoon tea break
15:30-16:00	Software curation best practice - case studies & policy <ul style="list-style-type: none"><li>• Case study 5: software provenance and software citation</li></ul> Policy initiatives to support software curation
16:00-16:30	Software curation best practice - support networks and research collaborations. Possible case studies: <ul style="list-style-type: none"><li>• Case study 6: URSSI</li><li>• Case study 7: Software Preservation Network</li><li>• Case study 8: EAASI</li><li>• Case study 9: PresQT</li></ul>
16:30-16:50	Table discussions <ul style="list-style-type: none"><li>• Reflections on case studies</li><li>• My action plan</li></ul>
16:50-17:00	Wrap up and evaluation
17:00	Thank you and close