ITCR Training and Outreach Working Group

Thursday, July 13, 2017 Meeting Notes

4pm Eastern

Toll-free: 855-259-6342; \*\*Conference Code: 40152#; \*\* Security PIN: 256871#

|  |
| --- |
| **ITCR TOW WEBEX** |
| Thursday, July 13, 2017 |
| 4:00 am  | Eastern Daylight Time (New York, GMT-04:00)  | 1 hr |

<https://cbiit.webex.com/cbiit/j.php?MTID=m6f4ee5c5f35990ae6ec9e64600fe18f9>

|  |
| --- |
|  |

**Google Doc for the Meeting:**

<http://bit.ly/itcr-tow-july2017>

**Attendees: Please sign in!!**

Juli Klemm, NCI

Mervi Heiskanen, NCI

Leah Mechanic, NCI

Michael Reich, UCSD

Andrey Fedorov, BWH/HMS

Mary Goldman, UC Santa Cruz

Martin Morgan, RPCI

Mike Ryan, JHU / MD Anderson

Vince Carey, HMS

Bradley Broom, MD Anderson

Isaiah Norton, BWH

Rudi Pillich, UCSD

Nathalie Pochet, BWH/HMS/Broad

Daniel S. Katz, University of Illinois

Hiro Yoshida, MGH/HMS

**Agenda/Minutes**

1. Updates on training activities from the group
   * [25th NA-MIC Project week](https://na-mic.org/wiki/Project_Week_25): participation from 3D Slicer, QIICR, and OHIF ITCR teams
   * Bradley Broom gave seminar on NG-CHMs to Kyoto Medical University
   * Bioconductor [CSAMA](http://bioconductor.org/help/course-materials/2017/CSAMA/) training. Forthcoming: [Annual Conference](http://bioconductor.org/help/course-materials/2017/BioC2017/), July 26-8, Boston, MA.
   * NDEx: Participation as exhibitor at ISMB 2017 conference (Prague, July 21-25)
2. Presentation from [Daniel S. Katz](http://danielskatz.org/) (blog: <https://danielskatzblog.wordpress.com/>), Assistant Director for Scientific Software and Applications at the National Center for Supercomputing Applications (NCSA), Research Associate Professor in Computer Science (CS), Research Associate Professor in Electrical and Computer Engineering (ECE), Research Associate Professor in the School of Information Sciences (iSchool), and Faculty Affiliate in Computational Science and Engineering (CSE) at the University of Illinois Urbana-Champaign.

Title: Measuring software technology impact

Abstract: I will discuss the impact of software in research, and different efforts to measure it.  Software is becoming more important in research, both in terms of software that is part of a single research project, and in terms of software that enables many projects. The role of software gives rise to a large set of interrelated challenges, and some of these, such as incentives, citation, metrics, publication, communities, career paths, and catalogs are closely tied to the impact of specific software.  While there are many items that can be measured to understand the impact of specific software packages, those that are easiest to measure are not those that provide the most insight. New methods for measuring software's impact can be created, but it is very hard to bring them into widespread use. Alternatively, we can use measures that we already have in place for papers. This talk will discuss these issues, where we are today, and where we might go next to better measure the impact of research software.

Slides: <https://doi.org/10.6084/m9.figshare.5207824>

Codemeta: <https://github.com/codemeta/codemeta>

Software Ontology: [http://theswo.sourceforge.net](http://theswo.sourceforge.net/), <https://bioportal.bioontology.org/ontologies/SWO>

A couple of points.  First, DOI as a tool for identifying software artifacts seems quite natural.  It does not seem anywhere near widely adopted, and the skills required are not completely trivial.  Each version of the software gets a different DOI. Is it straightforward to determine that two different DOI refer to the “same” software?  Second, citation of software. It is common in the medical literature to find an analysis defined as “performed with” R version x.y, or SAS version w.z.  Referees and editors will have to have higher standards for clarity of citation if sufficient attribution is to occur through the literature. Potentially noteworthy is the level at which use should be identified.  After running “example(DESeq)” in R (to illustrate use of a procedure for differential expression analysis in RNA-seq), a sessionInfo() call reveals dependence on over 60 other packages, many of them contributed as research of one kind or another.  The general topic is of great interest in information science. But the quantity of effort required to establish citability/measurability of utility has to be low because the amount of time available to create and maintain software is itself in short supply.  -- Vince Carey