Software as Infrastructure

Daniel S. Katz
Assistant Director for Scientific Software & Applications, NCSA
Research Associate Professor, CS
Research Associate Professor, ECE
Research Associate Professor, iSchool
dskatz@illinois.edu, d.katz@ieee.org, @danielskatz

FORCE11 Scholarly Communications Institute
WT02: Software Citation: Principles, Usage, Benefits, and Challenges
2–3 August 2017
Infrastructure in research

• What is it?
  • Ideas
• Books and papers
• Instruments
• Objects (e.g. Samples)
• Reagents
• Tools
• Data
• Software
• Computers
Cyberinfrastructure

• Aka eScience infrastructure, eResearch infrastructure
• What is it?
  • Ideas?
• “Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks, to improve research productivity and enable breakthroughs not otherwise possible.”

Purposes of software in research

Research Infrastructure

- Create Hypothesis
- Acquire Resources (e.g., Funding, Software, Data)
- Perform Research (Build Software & Data)
- Publish Results (e.g., Paper, Book, Software, Data)
- Gain Recognition
- Acquire Resources (Data)
Research software vs. infrastructure software

• Some software is intended for research
  • Funded by many agencies, sometimes explicitly, often implicitly
  • Intended for immediate use by developer
  • Maybe archived for future use and reproducibility
• Other software is intended as infrastructure
  • Funded by some agencies, almost always explicitly
  • Intended for use by community
• We focus mostly on infrastructure software, but some issues cross between
  • Reproducibility causes the most overlap