CRC
UNIVERSITY of NOTRE DAME
CENTER for RESEARCH COMPUTING

DASPOS and some thoughts on software sustainability

Jarek Nabrzyski (on behalf of the DASPOS consortium)

naber@nd.edu

DASPOS

- Data And Software Preservation for Open Science (daspos.org)
  - $1.8M NSF-funded multi-disciplinary effort, funding started 9/12
  - Notre Dame, Chicago, UIUC, Washington, Nebraska, NYU
  - Links HEP effort (DPHEP+experiments) to Biology, Astrophysics, Digital Curation, and other disciplines
- Includes physicists, digital librarians, computer scientists
- Aims to achieve some commonality across disciplines in
  - meta-data descriptions of archived data
  - What's in the data, how can it be used?
  - Computational description (ontology/metadata development/computational experiment observation pattern)
  - How was the data processed?
  - Can computation replication be automated?
  - Impact of access policies on preservation infrastructure

We develop preservation software!

- Umbrella – a reproducible environment creator
  - Part of CCL tools (Doug Thain, http://ccl.cse.nd.edu/)
- Prune – Git for workflows, not just directories
  - Also CCL tools
- Smart container – Ontology and provenance for containers (@crc.nd.edu)

Is this software sustainable?

Some thoughts on software sustainability

- Need a user community that depends on your software.
- The software needs to facilitate, enhance, provide added value to researchers' research process.
- Plan in advance to work with your community of users. Make sure you plan your budget adequately.
- How about a preservation software? It can be sustainable only if it adds value to the community of users and is transparent to their workflow (adds no additional effort to preserve "research objects")