ReproZip
Packing Experiments for Sharing and Publication

Fernando Chirigati, Juliana Freire | NYU-Poly
Dennis Shasha | NYU
Motivation

- Published articles are not made reproducible
- Computational reproducibility may be difficult to achieve

Some current solutions require the user to adopt a system

Other solutions rely on capturing information about the computational environment
- Virtual Machines
- CDE [4]

Author

How to encapsulate my experiment?
Too many dependencies…
Too many files to keep track…
Sigh.

How to compile this program?
How to execute it?
How to explore it?
Sigh.

Reviewers

Collaborators

Fernando Chirigati – NYU-Poly
ReproZip: Packing Experiments for Sharing and Publication

ReproZip

• ReproZip is a packaging solution
  o It makes it easier for authors to pack experiments and for reviewers to verify computational results

• It creates reproducible packages from existing experiments on computational environment $E$
  o No need to port experiments to other system
  o Leverages provenance of computational results

• It unpacks an experiment on computational environment $E'$

• It generates a workflow specification that encapsulates the execution of the experiment
  o Eases the verification process
  o Allows users to explore the experiment, while keeping track of provenance
ReproZip: Packing Experiments for Sharing and Publication

Overview

packing (on environment E)

Experiment → Provenance Tree → Workflow → Reproducible Package

files + binaries + workflow

unpacking (on environment E')

Reproducible Package → Experiment Extraction → files + binaries + workflow

verification and exploration

Fernando Chirigati – NYU-Poly
References

Thank You!

Fernando Chirigati
fchirigati@nyu.edu
http://vgc.poly.edu/~fchirigati