Living Artifacts
Enabling Vertical Progress in Cybersecurity Research

Jelena Mirkovic and David Balenson (mirkovic, balenson@isi.edu)
University of Southern California Information Sciences Institute

Presented to: MetaCRISP Workshop, May 21, 2026

SPHERE is based upon work supported by the National Science Foundation under Grant #2330066. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Motivation and Need

- Like many scientific fields, cybersecurity struggles with reproducibility and reuse
- Artifact evaluation efforts at top conferences since at least 2017
  - Successfully evaluated artifacts earn badges, displayed on papers
  - Helps motivate artifact sharing: Olszewski et al. found that the percentage of papers with artifacts has increased by 36% from 2013 to 2023
- Sharing is necessary, but not sufficient for successful reuse
  - Code artifacts may not easily compile and run
  - Data preprocessing, output validation, hyperparameter tuning may be missing
  - Datasets may be hard to interpret
- Successful reuse is important
  - Allows future independent researchers to build on prior published work
  - Over time, leads to more sophisticated and impactful solutions ⇒ **vertical progress!**

25-40%
released code artifacts

Olszewski et al. 2023, 2025

40-44%
ran w/o modifications

Current Artifact Evaluation

- Human-centered practice
  - Poorly standardized metadata - sometimes in PDF
  - Loose packaging instructions
  - README files instead of installation and run scripts
- Focus on author claim validation
  - Similar to paper review - double-check that the authors' work is valid
  - **A few other researchers** were able to validate claims **on their own hw** or on authors' hw
  - Possibly with authors' help, but fixes may not propagate or apply to other platform
- May lead to incomplete artifacts
  - Private data, scaled-down models, portion of results
- Resulting artifacts are
  - Hard to find - appendix and metadata are separate from the artifact and paper
  - Hard to reuse - hw, sw, paid APIs, packaging challenges
  - May become less reusable over time

Interoperable ❌

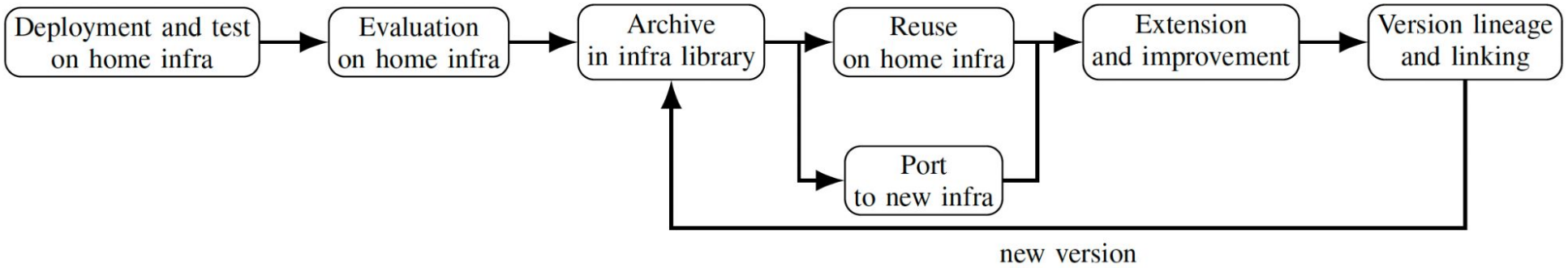
standard metadata and install/running scripts

Findable ❌
Accessible ❌
Reusable ❌

code installs and runs

Living Artifacts

- Standardized metadata
  - Including links to paper, hw/sw dependencies, install scripts, claim-verify scripts
- Packaged and evaluated on a public research infrastructure (home)
  - Author interactions feed into the final artifact
  - Easy to reuse on the home infrastructure, can be archived there
- Others can build on and version the artifact
  - Add new features, port to new infrastructure, modernize over time
  - Facilitates vertical science
- Some artifacts will die out
  - Remain in permanent repository, can be revived by future researchers



Traditional Versus Living Artifacts

Property	Traditional	Living
Publicly available	✓	✓
Runs during evaluation	✓	✓
Standardized execution environment		✓
Easy reuse after publication		✓
Machine-readable metadata		✓
Claim-level execution scripts		✓
Supports versioned extension		✓
Traceable lineage		✓
Designed for long-term reuse		✓

Living Artifacts and Vertical Science

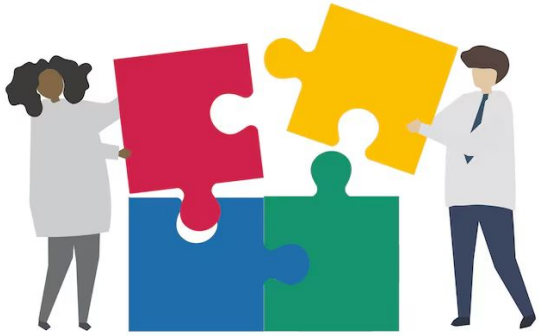
- Living artifacts should
  - Be easier to reuse on their home infrastructure
  - Easier to extend, port and incorporate into future security solutions
  - Facilitate vertical science where teams build on each other's work
  - Improve visibility of evaluators, authors and contributors
- Packaging and metadata standards are also needed
  - Enable automated search over metadata - better findability
  - Enable easy extension (build scripts instead of VM/Docker images)
- Necessary, but not sufficient for vertical science
  - Current paper reviews view work building on prior research as incremental
  - We could define specific criteria for evaluation of work that advances prior results, reusing and improving existing artifacts

Reusable ✅

Findable ✅
Accessible ✅
Interoperable ✅

Path Forward

- Artifact evaluation committees can:
  - Require standardized metadata and packaging
  - Require evaluation on public research infrastructure, wherever possible
- Authors can:
  - Package their artifacts on public research infrastructure, making it more reusable
- Research infrastructures can:
  - Provide specialized support to the AECs
  - Support artifact archiving, cloning, modification and version linkage
- Research community can:
  - Reward artifact reuse and building on prior work

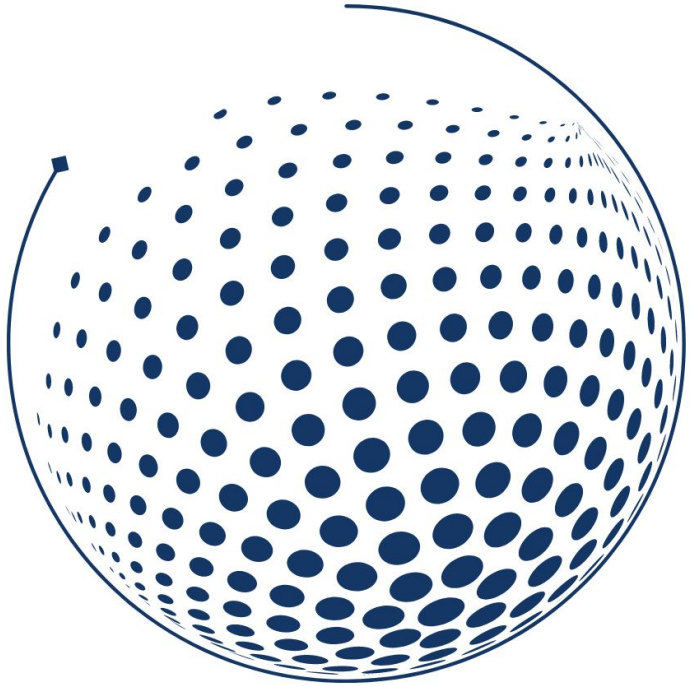


QUESTION: How should living artifacts handle work that depends on private infrastructure, private data, or specialized hardware?

Thank you!

https://sphere-project.net
https://sphere-testbed.net

contact@sphere-project.net



SPHERE

RESEARCH
INFRASTRUCTURE