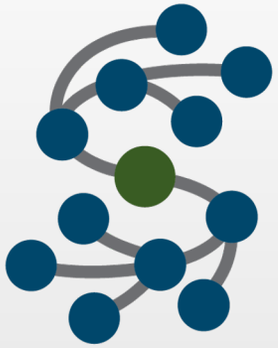


SEARCCH

Sharing Expertise and Artifacts for Reuse through Cybersecurity Community Hub

Toward Findable, Accessible, Interoperable, and Reusable Cybersecurity Artifacts

Presentation for the
15th Cyber Security Evaluation and Test Workshop (CSET 2022)

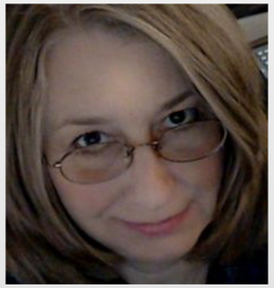
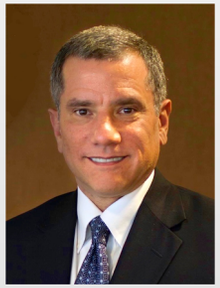
David Balenson, SRI International
Monday, August 8, 2022

This material is based upon work supported by the National Science Foundation under Grant Numbers 1925773, 1925616, 1925588, 1925564. 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.

Collaborators and co-authors


Terry Benzel, Jelena Mirkovic
USC-ISI
Marina Del Rey, CA, USA
benzel@isi.edu,
mirkovic@isi.edu

USC Viterbi Information Sciences Institute

Laura Tinnel, David Balenson
SRI International
Arlington, VA, USA
laura.tinnel@sri.com,
david.balenson@sri.com

SRI International®


Eric Eide, David Johnson
University of Utah
Salt Lake City, UT, USA
eeide@cs.utah.edu
johnsond@cs.utah.edu

THE UNIVERSITY OF UTAH

David Emmerich
University of Illinois
Urbana, IL, USA
davidpe@illinois.edu

ILLINOIS

The cybersecurity community is still far from an ecosystem in which artifacts are FAIR: Findable, Accessible, Interoperable, and Reusable

Researchers in experimental cybersecurity are increasingly sharing the code, data, and other artifacts associated with their studies

Encouraged and rewarded by conferences and journals through practices such as artifact evaluation and badging

Lack of established standards and best practices for sharing and reuse results in artifacts that are often difficult to find and reuse

Lack of community standards results in artifacts that may be incomplete and low-quality

FAIR principles

Findability - Relates to metadata, both for the artifact itself (e.g., a DOI) and for describing the content of the artifact

Accessibility - Means that the artifact is openly available to interested parties

Interoperability - Concerns the representation of the artifact, e.g., the use of standard languages and vocabularies for datasets, and the use of standard tools, libraries, and techniques for software

Reusability - Relates to accurate provenance, clear licensing terms, and adherence to other community standards

See https://www.go-fair.org/fair-principles/ and Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Challenges to achieving FAIR principles for experimental cybersecurity research

Findability challenge - Many artifacts, but little metadata, many channels, and hard to determine suitability

Accessibility challenge - Identifying and documenting relevant details about infrastructure and experimental workflow

Interoperability and Reusability challenge - Dependence on infrastructure such as programming languages and environment, hardware and software, private datasets, etc.

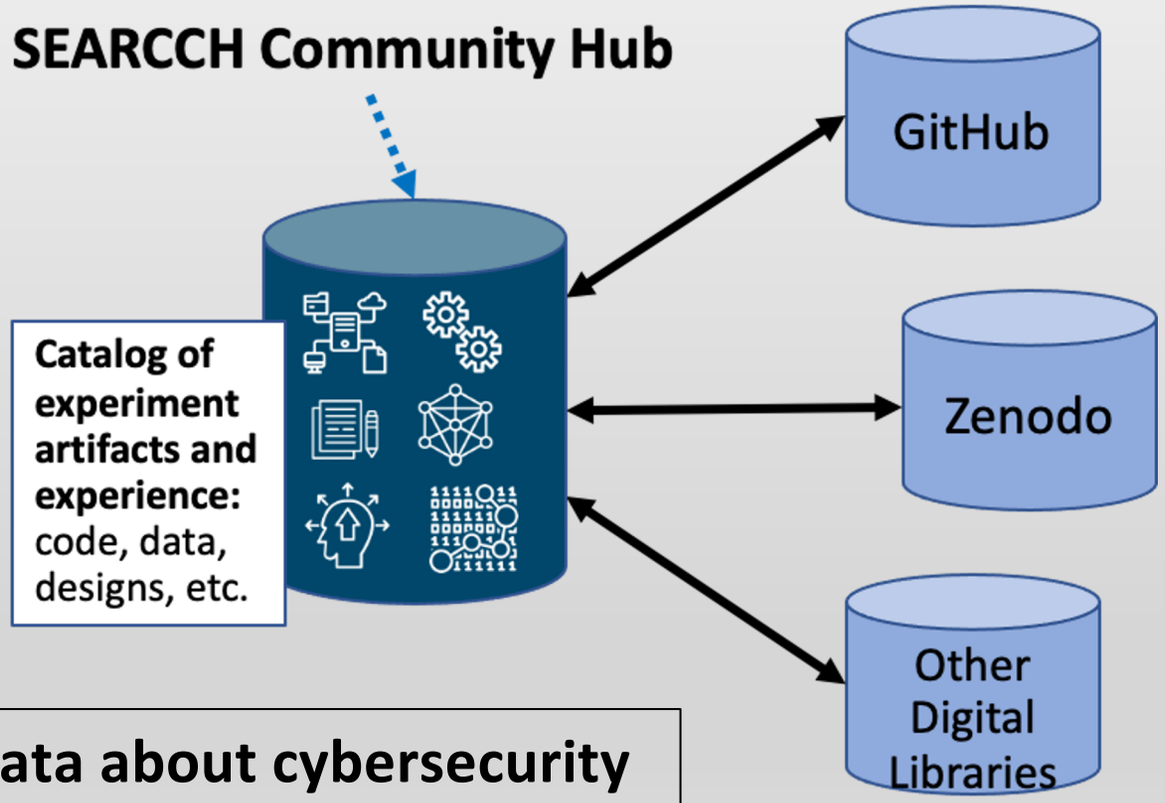
Motivational challenge - High effort for sharing and reuse, limited reward, often easier to create own environment

The community needs a common metadata format for capturing the knowledge that is necessary to adopt artifacts

SEARCCH is a web-based community portal that aims to improve the findability and reusability of cybersecurity artifacts

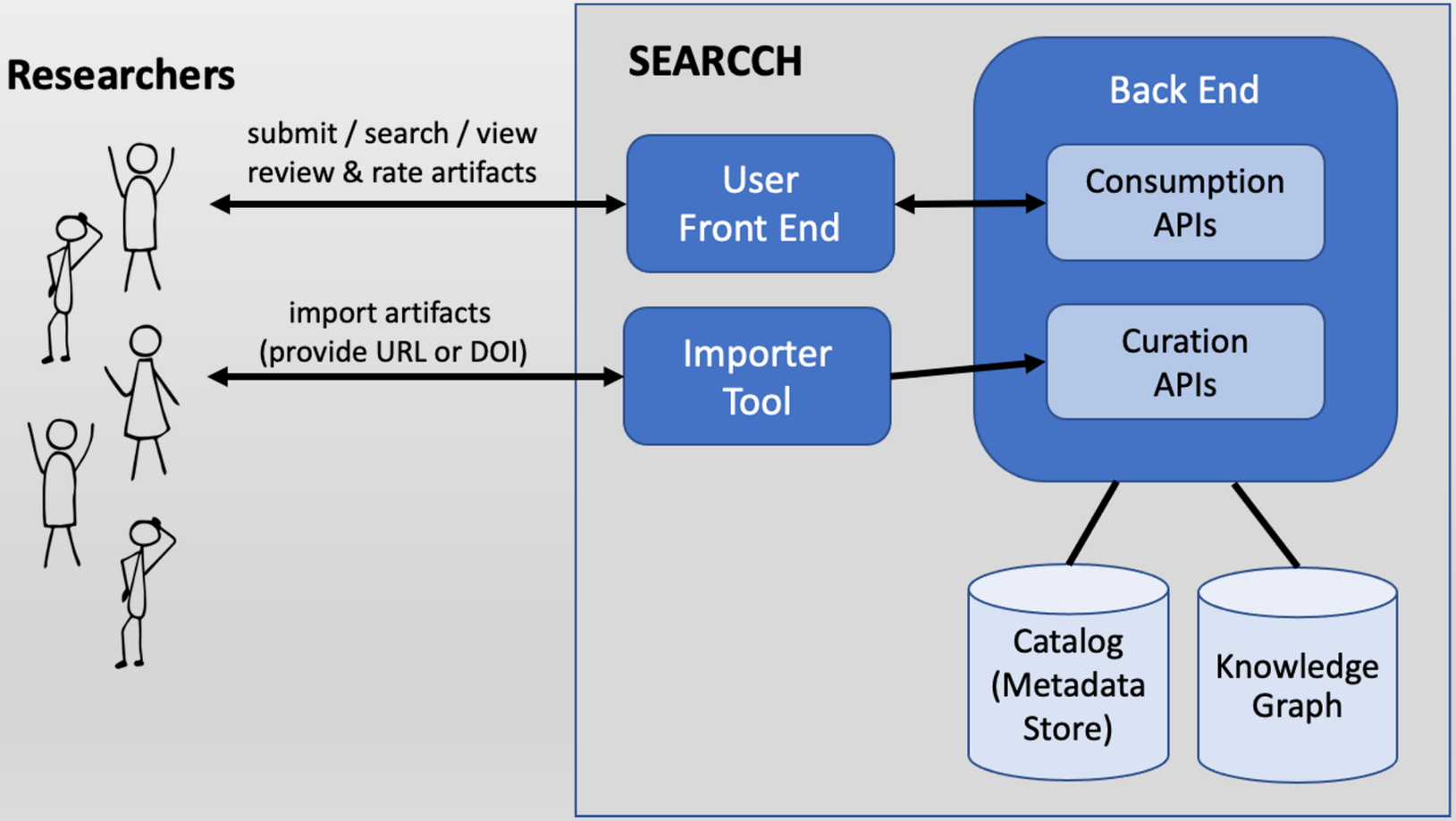
Catalog - Database of information about research artifacts located in different places on the Internet

Community - Allows researchers to extend the hub's content with new artifacts and discussion

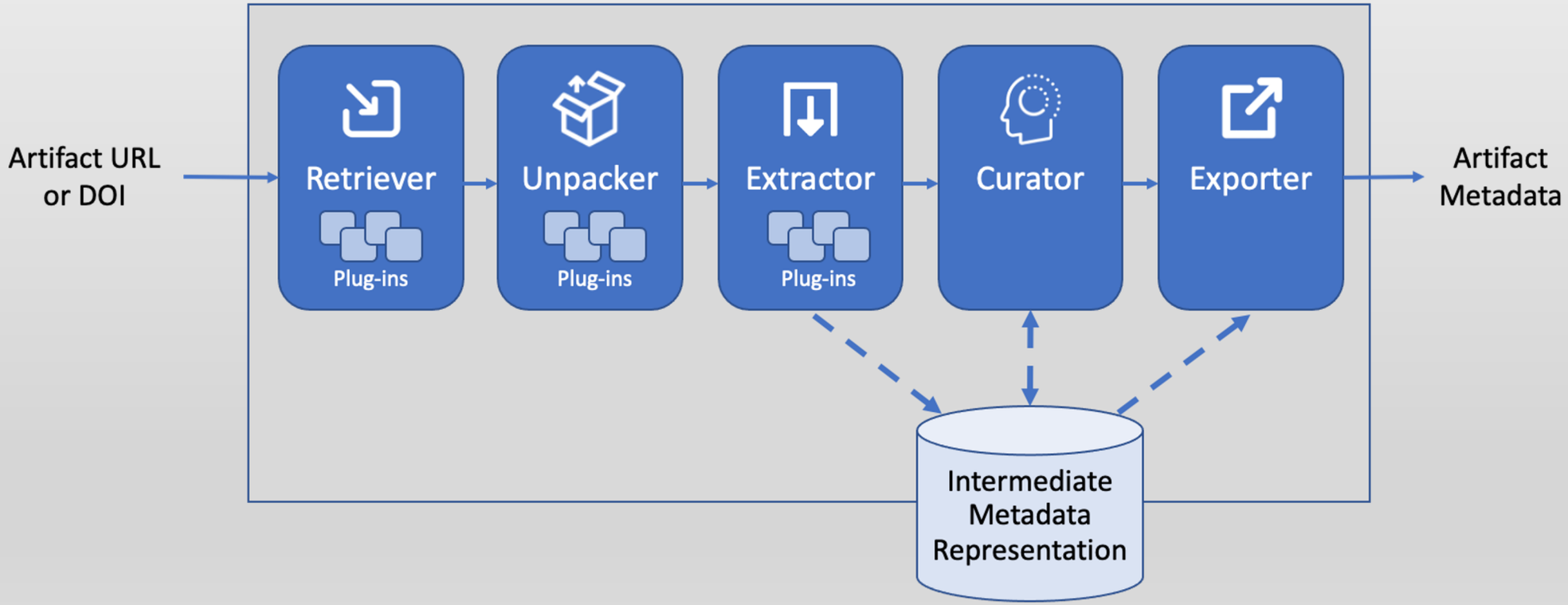


SEARCCH catalogs metadata about cybersecurity artifacts that are stored in separate repositories

Researchers interact with SEARCCH through consumption and curation APIs



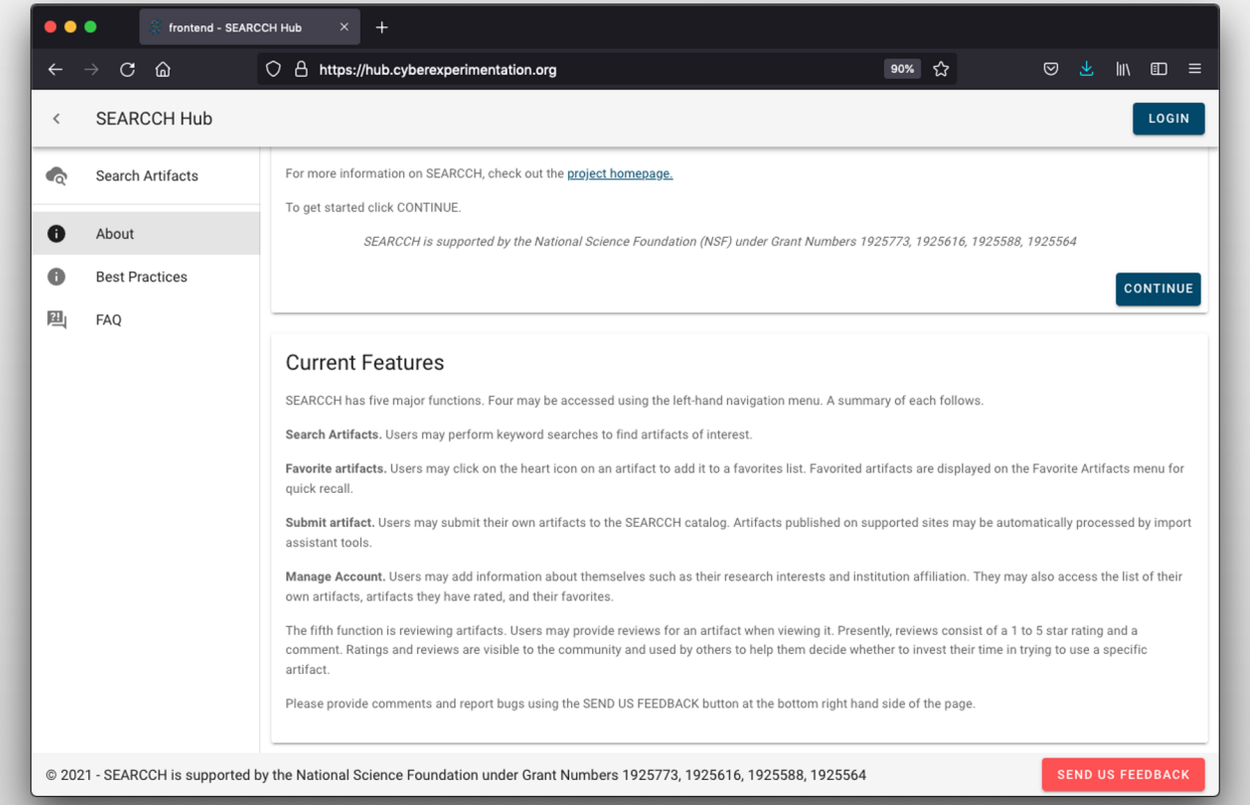
The SEARCCH importer tool uses “plug-in” modules to obtain metadata about artifacts stored in a variety of locations and formats



SEARCCH features and capabilities

- Submit artifact
- Search artifacts
- View artifacts
- Review and rate artifacts
- Favorite artifacts
- Manage account
- Best practices
- FAQ

https://hub.cyberexperimentation.org/



The SEARCCH catalog currently lists ~393 cybersecurity artifacts and publications: 206 software artifacts, 55 datasets, and 132 publications

Lessons learned

The importance of community outreach and engagement

- Develop a diverse and vibrant community of cybersecurity researchers using SEARCCH to instill a sense of ownership, raise awareness, and elicit design input and feedback
- Engage researchers early in the development process to identify and prioritize key features, support a diverse set of workflows, and motivate simpler user interfaces and easy-to-use tools

The need for improved community metadata standards

- Achieving a high degree of findability requires structured attributes that describe an artifact's domain of concern, contexts in which it is applicable, and relationships between artifacts
- Maximizing reusability depends on packaging and documenting the requirements of an artifact

Conclusions

SEARCCH

Sharing Expertise and Artifacts for Reuse through Cybersecurity Community Hub

The SEARCCH web-based community hub aims to improve the findability and reusability of cybersecurity artifacts

- Community input was essential for defining the features of the hub
- More community involvement will be needed to further advance FAIR principles through the development of new metadata standards for experiment artifacts

We encourage you to contribute to and make use of cybersecurity experiment artifacts in the SEARCCH hub

- Follow us on Twitter: @SEARCCH_Hub
- Visit us on the web: https://searcch.cyberexperimentation.org

Virtual Workshop: Encouraging the Production and (Re)use of Cybersecurity Datasets, Software, and Other Research Artifacts - September 15, 2022

Goal: Explore the issues and challenges researchers face in producing and reusing cybersecurity research artifacts, including datasets and software, as well as experiment designs, execution, and results. Participants will discuss and share experiences regarding current initiatives, infrastructure, and incentives for producing and reusing artifacts

Topics:

- FAIR principles and metadata standards
- SEARCCH and other repositories
- Artifact evaluation processes
- Beyond datasets and code
- Next steps - what can we do as a community

To participate: contact david.balenson@sri.com