Panel: On the Curation of Artifacts in the Era of AI/ML for Cybersecurity

David Balenson
Senior Computer Scientist
Networking and Cybersecurity Division
USC Information Sciences Institute

ACSAC 2022
December 8, 2022

PIVOT Project

https://www.pivot-auto.org/

Coordinated effort to bring together a community around the development and sharing of robust automotive datasets to foster and support new, open cybersecurity research in smart city, transportation, and automotive applications

(1) Robust and reliable hardware/software platform upon which the system runs

(2) Curation and sharing of the data and contextual information

(3) Researcher centric services for sharing, securing, and evaluating datasets

(4) Common software-based tools to collect, transform, combine, filter, and visualize the data

(5) Extensive community outreach and engagement to improve the data utility using design feedback mechanisms

USC Viterbi School of Engineering Information Sciences Institute

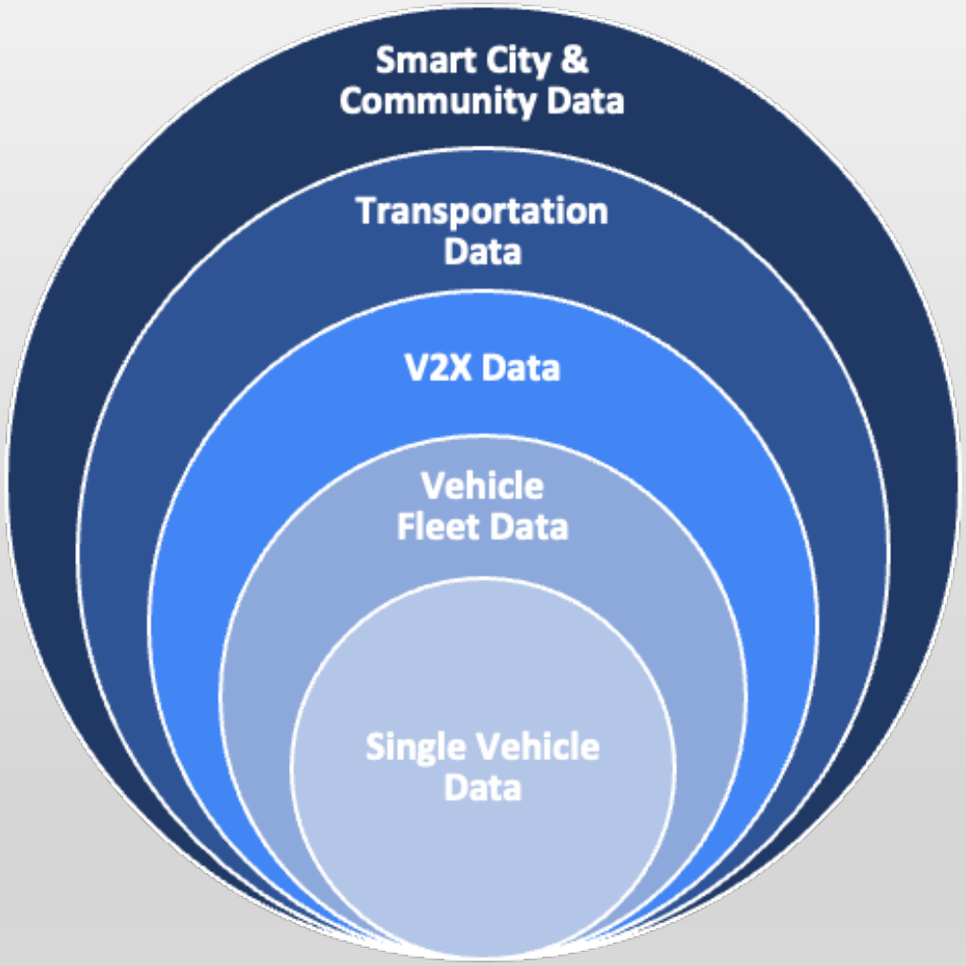
THE UNIVERSITY OF MEMPHIS

Colorado State University

GEOTAB

This material is based upon work supported by the National Science Foundation under Grant Numbers 2213733, 2213735, and 2245323. 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.

Examples of Automotive Datasets



- Oak Ridge ROAD and potential future datasets
- Korea University HCRL Datasets
- Bosch SynCAN (for CANet)
- TU Eindhoven Lab Automotive CAN Bus Intrusion
- CrySiS Lab CAN-Log Infector and Ambient CAN Traces
- Cephas Baretto Dataset
- Heavy Truck Datasets from Jeremy Daily @ CSU
- Autonomous driving datasets
- Geotab telematics data and Altitude analytics platform
- US Department of Transportation Public Data Portal
- SmartColumbus Datasets Curated for Visualization
- Wyoming DOT CV Pilot

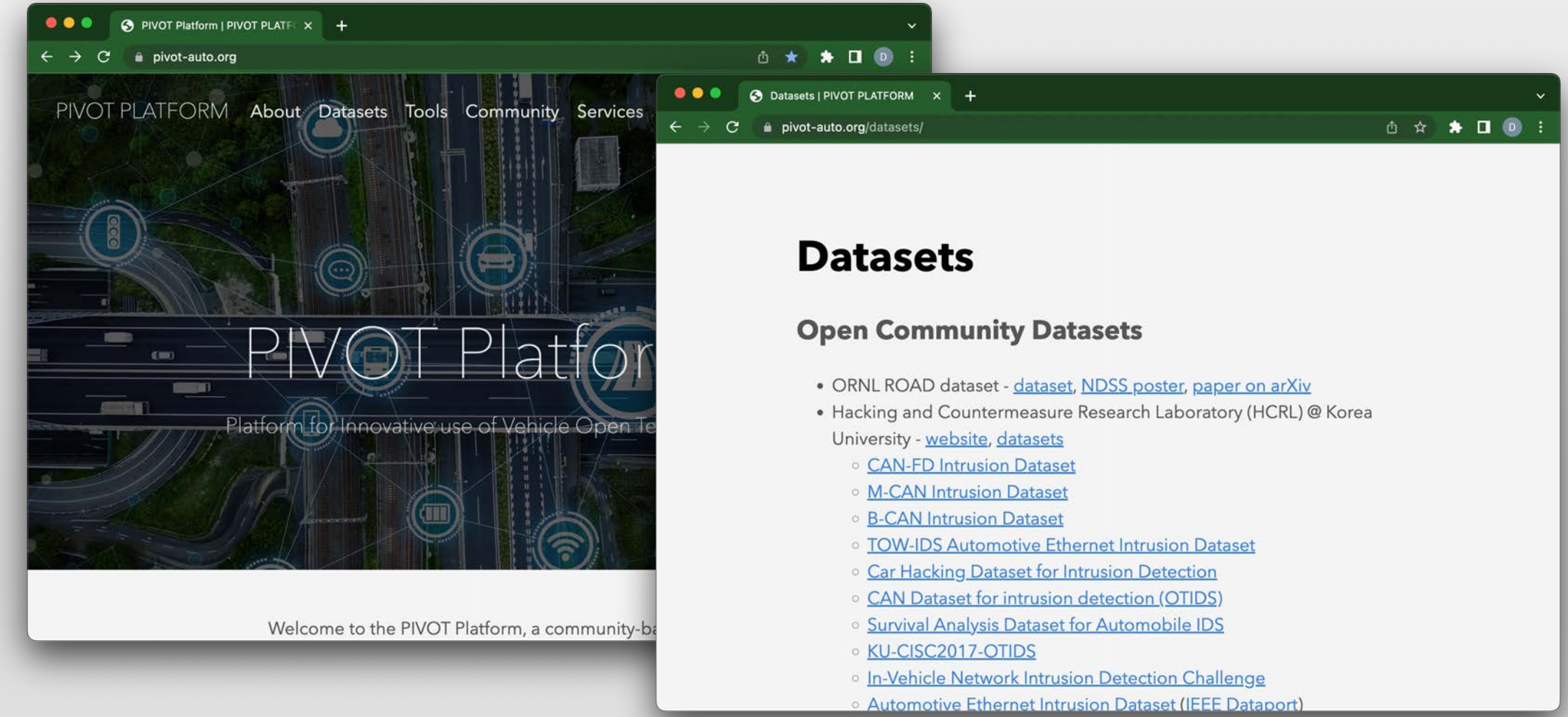
Potential Applications of Automotive Datasets

- **Vehicles**: System monitoring and optimization, in-vehicle infotainment, predictive maintenance, route and trip planning, etc.; ADAS, CAVs, EVs, heavy trucks, etc.
- **Transportation and fleet management**: passenger safety, traffic management, ride sharing, multi-modal mobility, data-driven insurance
- **Smart cities**: infrastructure monitoring and mgmt, weather sensing and mapping, asset mgmt, etc.
- **Safety and cybersecurity**: CAN bus anomaly detection, sensor security, autonomous driving, etc.
- **NSF research communities**: SaTC, S&CC, CPS, CIVIC, etc.

PIVOT Datasets

https://pivot-auto.org/datasets/

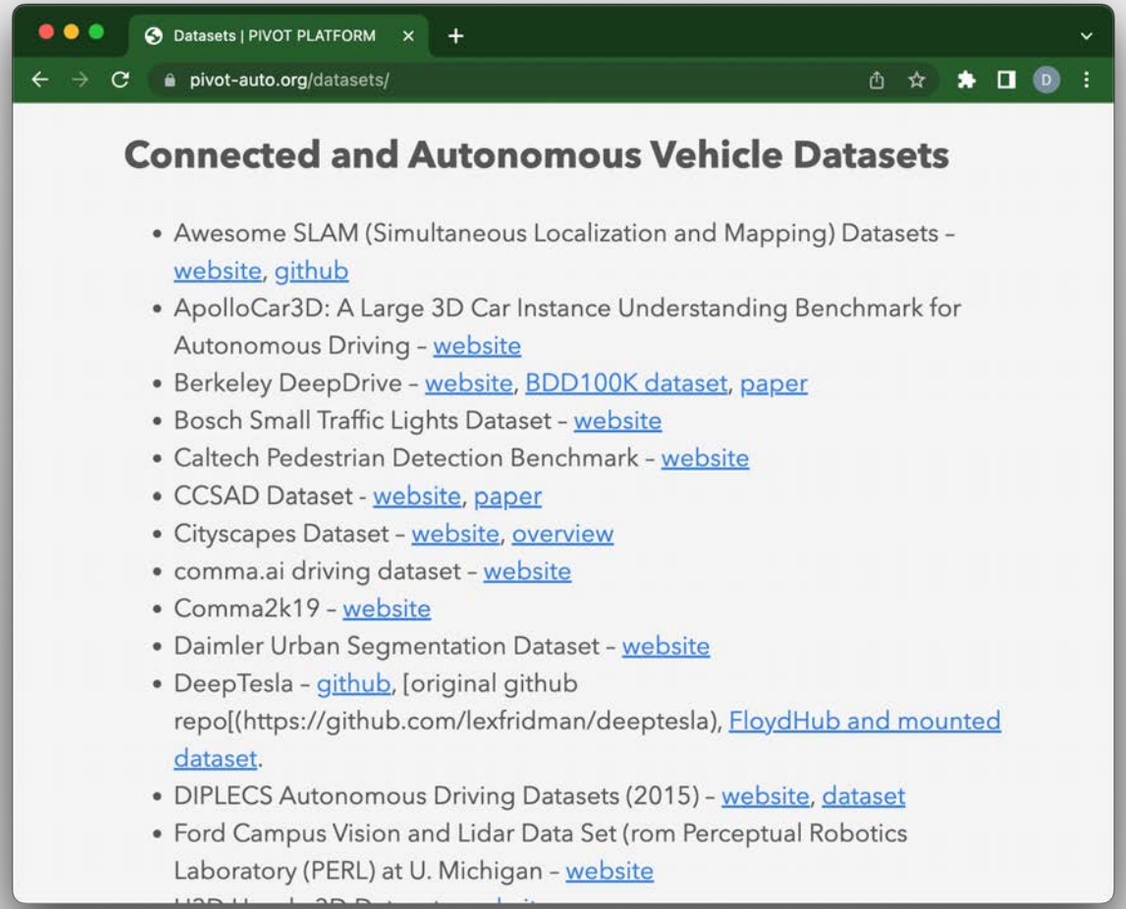
Curated list of automotive dataset on the PIVOT platform



CAV Datasets

https://pivot-auto.org/datasets/

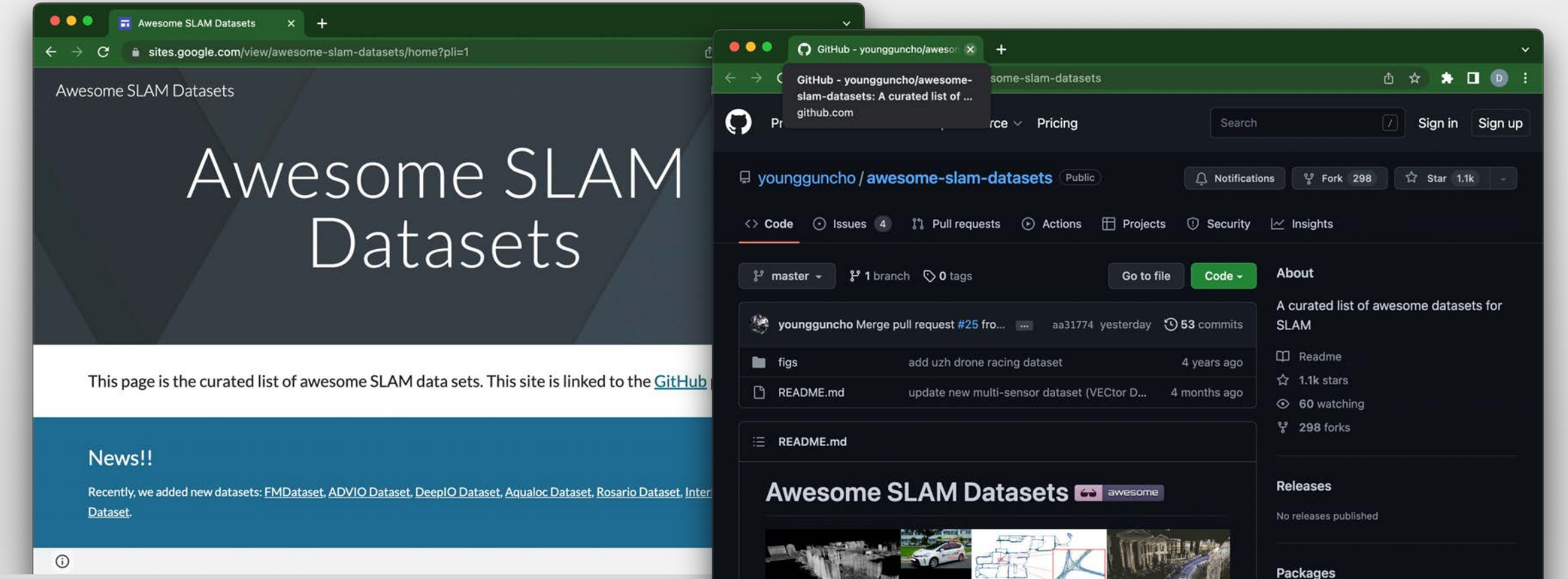
List of connected and autonomous vehicle datasets in the PIVOT platform



Awesome SLAM Datasets

https://sites.google.com/view/awesome-slam-datasets/home
https://github.com/younggguncho/awesome-slam-datasets

Curated list of Simultaneous Localization and Mapping (SLAM) datasets



(Some) Common CS Repositories

PAPERS

ACM DL DIGITAL LIBRARY

IEEE Xplore DIGITAL LIBRARY

Springer

ScienceDirect ELSEVIER

zenodo

usenix THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

arXiv.org

CODE & DATA

GitLab

aws Code Artifacts

Bitbucket

GitHub

codebase by KRYSTAL

MENDELEY DATA Put your research data online today

Microsoft Azure Artifacts

SOURCEFORGE

Google Code Archive

zenodo

(Some) Research Artifact Search Indices / Catalogues

Scientific research

Google Scholar ResearchGate & CONNECTED PAPERS ➢ papers

Google Dataset ➢ data

Mendeley ➢ papers, data (also repository)

OpenAIRE EXPLORE ➢ papers, data, and code

Computer science

FindResearch.org (no longer maintained) ➢ papers, data, and code

Machine learning research

Papers with Code ➢ papers, data, code, methods

UCI Machine Learning Repository ➢ data

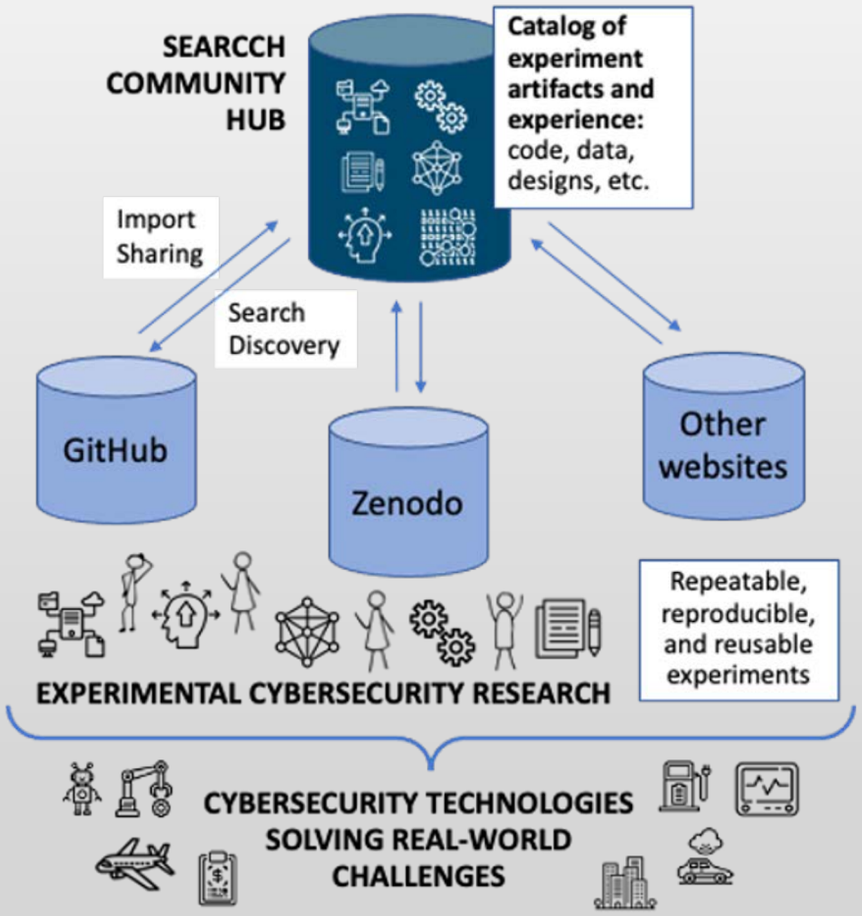
Cybersecurity research

SEARCCH ➢ data, code, papers (secondary)

SEARCCH Project

https://searcch.cyberexperimentation.org/

- Community-driven platform lowers barrier to sharing and reusing research artifacts
- Rich metadata representation enables researchers to better describe and find relevant artifacts
- Import, curation, and search functions enable greater scientific quality of cybersecurity research



USC Viterbi School of Engineering Information Sciences Institute · SRI International · The University of Utah · Illinois

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