Data Management Plan

Description of data to be generated in this project:

During the course of the proposed project, textual data generated that should persist will be training materials and other public documentation (e.g., best practice guides, lessons learned educational curriculum, engagement reports). Trusted CI does not expect to generate or capture experimental or other data that would necessitate a relational database or specific data file formats for programmatic access from computer models. We expect that all of the textual data generated by this project can be projected into the Adobe Portable Document Format (PDF) and preserved as described below. PDF documents are commonly full text indexed by search engines and are available to text mining and natural language processing systems. The project team expects that the ability to consume and manage content in the PDF file format will outlive the meaningfulness of the data generated by Trusted CI.

The project developed and maintains the Trusted CI Framework, along with a suite of associated guidance, templates, and tools that are designed to be adopted and used by third-party organizations. These published materials will be stored in a version-controlled repository with a corresponding DOI, such that all previously published versions are accessible. (Please see a discussion of Zenodo, below.)

Videos for Trusted CI Webinars, Fellows sessions, Summit talks, and Trusted CI presentations at conferences are publicly available on YouTube (https://www.youtube.com/TrustedCI or www.youtube.com/@rrcop5071).

Videos for Trusted CI's software training materials (https://research.cs.wisc.edu/mist/SoftwareSecurityCourse/) are currently publicly available on Vimeo (https://vimeo.com/uwswsecurity). Source materials for these videos (videos, captioning, and scripts) and hands-on materials (disk images) are currently archived at University of Wisconsin-Madison. The video storage format is currently mp4 and includes closed-captioning for accessibility. We will periodically reevaluate the choice of a video platform that is used in order to ensure that the platform best meets the needs of the project.

The project will generate some data, related to its work in software assurance, regulated work, and community engagement activities, which will not be immediately public until we have had a chance to work with involved parties to perform responsible disclosure and satisfy agreed-upon redactions, after which time the data will become public. This process is well-established within Trusted CI and described in our "Trusted CI Collaborator Information Policy", which defines "Collaborator Information" and describes how collaborator information will be handled, along with listing exemptions to the policy.

Data created in support of Controlled Unclassified Information (CUI) can often contain more sensitive data, such as system diagrams, ports, VLAN or IPs. Trusted CI makes collaborators aware of data

sensitivity concerns or privacy matters, as the intent is to broadly share the results. Under no circumstances will Trusted CI store, transmit, or process CUI or Classified Data.

Responsibility for data management:

Ultimate responsibility for data management within Trusted CI will reside with PI Sean Peisert; however, Trusted CI team members will each be responsible for the management of data within activities they lead. This responsibility includes ensuring that the materials have appropriate distribution statements, search terms and metadata; have project, grant, and partner attribution; have the Trusted CI license declaration; have been cataloged as a project artifact; and have been preserved according to the policies within this data management plan.

License for data generated as a result of this project:

All materials *de novo* generated as part of this project that will be distributed will be distributed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0). The full terms of this license are available at https://creativecommons.org/licenses/by-nc/4.0/deed.en. This license includes the following terms: You are free to share – to copy, distribute and transmit the work and to remix – to adapt the work under the following conditions: attribution – you must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work). For any reuse or distribution, you must make clear to others the license terms of this work.

Data preservation, dissemination, and public use:

Google Drive, GitHub, and Slack are used to produce, share and collaborate on intermediate data products.

Trusted CI's aim is to make as much material that it generates as possible publicly available. For long-lived, public-facing documents, Trusted CI plans to leverage Zenodo (https://zenodo.org/) — a respected open access repository that is maintained by CERN — for data preservation and assigning of DOIs to public artifacts. Existing Trusted CI artifacts on Zenodo are collected in the Zenodo Trusted CI Community (https://zenodo.org/communities/trustedci/). There may be cases where Dryad (https://datadryad.org/) is preferable. The Dryad open data publishing platform is a general-subject repository for research data. Dryad is a 501(c)(3) originally funded by a grant from the U.S. National Science Foundation and with a long history and close partnership with the California Digital Library (CDL). Berkeley Lab, UIUC, and UW-Madison are all institutional members of Dryad, meaning all researchers at those institutions can submit research data (up to 300 GB per dataset) to Dryad at no cost. As with Zenodo, publishing data in Dryad ensures long-term preservation and availability.

Trusted CI will also make its products accessible via the center's public website and will take steps to ensure the NSF community and public are aware of these products.

There are some public-facing artifacts generated by Trusted CI for which DOIs may not be the preferred mechanism. Examples of such artifacts might include individual videos in Trusted CI's software assurance training materials, For data preservation of these materials, long-lived institution library

repositories are used. For example, UW-Madison Libraries support a service called MINDS@UW (https://www.library.wisc.edu/research-support/minds/) for archiving data online. It includes a long-term commitment of availability and permanent URLs. Similar repositories exist at each Trusted CI organization.