Overall: CTSI Resource Sharing Plan

RESOURCE SHARING PLAN

The University of Florida Clinical and Translational Science Institute (CTSI) and its CTSA partner, Florida State University, intend to broadly share research resources according to FAIR (findable, accessible, interoperable, and reusable) principles, meeting and even exceeding NIH requirements for sharing of research resources. Besides data and software, resources to be shared include model organisms, computable phenotypes, ontologies, policies, practices, education and training materials, model consent forms, and templates for memoranda of understanding and legal agreements and contracts. We are furthermore committed to incorporating in our translational workforce development programs competencies that enable workforce members to contribute to resource sharing and reuse, including the development of research consent language that permits broad data reuse.

We have a strong track record of resource sharing, with over 120, open, and free shared-resource repositories available at https://github.com/ctsit and https://github.com/ufbmi. Biomedical Informatics (BMI) Program faculty and staff also collaborate with other CTSA sites who are sharing resources via a free and open source model (e.g., see https://github.com/midas-isg, where the BMI Program maintains an ontology-based, searchable, FAIR repository of resources for the Modeling Infectious Disease Agents Study). The CTSI and OneFlorida have also shared resources such as computable phenotypes, data use agreement and other regulatory template documents, and software within the National Patient Centered Clinical Research Network, NCATS ACT, Implementing Genomics in Practice (IGNITE) Research Network, and the NIH All of Us Research Program. As part of the NIH Common Fund Metabolomics Program, the CTSI's Southeast Center for Integrated Metabolomics has submitted 60 datasets to the national Metabolomics Workbench data repository.

Model organisms sharing plan. Nothing in the plans of the linked U, K, and T proposals that constitute our renewal submission for a Clinical and Translational Science Award involves development of unique model organisms directly. However, we have implemented a policy whereby all pilot projects we fund that involve development of unique model organisms, regardless of funding source, must comply with NIH policies on sharing model organisms as a condition of making the pilot awards. Furthermore, we assist investigators to whom we make pilot awards with compliance with these policies. We will continue these efforts in the renewal award.

Data sharing plan. We plan to share research, clinical (including EHR), genomic, investigator, and program evaluation data as broadly as possible within the limits prescribed by law, regulation, and the principles of ethical human subjects research. Illustrative of this commitment, the CTSI and UF Health are already making de-identified clinical data available to the CTSA network for participant recruitment in the NCATS Accrual to Clinical Trials project. Through participation in the OneFlorida Clinical Research Consortium, CTSI has created multi-directional clinical data sharing among its members for the purposes of research. OneFlorida also requires investigators who use Consortium infrastructure for a study to commit to making HIPAA-limited study data sets available back to the Consortium at the end of the study. This broad data sharing in OneFlorida is accomplished through memoranda of understanding and executed data use agreements. Research using UF Health clinical data other than fully HIPAA de-identified data requires IRB approvals (or at least a non-human-subjects determination) and executed data use agreements, but under these conditions is open to investigators at UF, in OneFlorida, and the CTSI network. The CTSI and UF Health are committed to sharing clinical data needed for multi-site clinical trials including pragmatic trials and implementation science studies through the Research Electronic Data Importer (RED-I) software, as they have done already with HCV TARGET. OneFlorida also plans to make de-identified, aggregate site-level recruitment and research participation data available to all sites participating in studies using the OneFlorida Consortium infrastructure. All research data will be shared according to the NIH Genomic Data Sharing Policy. All clinical trials will comply with the NIH Policy on the Dissemination of NIH-Funded Clinical Trial Information and the International Committee of Medical Journal Editors (ICMJE) requirements and will be compliant with ClinicalTrials.gov requirements, including a data sharing plan approved by the IRB, with facilitative support available to investigators through the CTSI Service Center and Office of Clinical Research. Where needed, information regarding data sharing will be included in informed consent forms and HIPAA authorizations for research, as directed by the IRB. We are also broadly sharing UF investigator profiles in VIVO format as recommended by the CTSA network in

2011, and will continue to make available data from its Network Science program in this manner. Finally, we plan to share evaluation and tracking data to inform the science of translation.

We are also committed to stimulating research data sharing by helping investigators receive academic credit for generating and sharing data sets. To that end, the CTSI Biomedical Informatics Program will offer as a service to investigators the procurement of digital object identifiers or DOIs for their data sets. DOIs enable investigators and those who use their data sets to cite the data sets in publications and other academic writing, thereby generating citation statistics by which to measure the data's influence on subsequent work. To enable discoverability and promote the visibility of DOI-tagged data sets, the CTSI BMI Program will also work with Biomedical and healthCAre Data Discovery and Indexing Ecosystem (BioCADDIE) led by the University of California San Diego to incorporate data sets into the relevant data catalogs in an expeditious manner. As these developments of citable datasets and research data catalogs are emerging, we will also track the evolution of best practices in this regard in the overall NIH Big Data to Knowledge (BD2K) program and the NCATS CTSA Data to Health (CD2H) program.

Software sharing plan. All software from this project will be made available freely to researchers and their institutions for educational, research, and non-profit purposes. Furthermore, DOIs are also available to software libraries and BD2K has sought from the research community its input into a software discovery index, through a request for information. Already, the BMI Program has obtained DOIs for nearly 20 software releases including RED-I, the *deduper* record-linkage/entity-resolution software created for OneFlorida, and *nacculator* software for converting REDCap-based datasets into the *National Alzheimer's Coordinating Center* data format. The CTSI plans to continue the procurement of DOIs for the software artifacts created in this project, and to incorporate them into the anticipated BD2K-sponsored software discovery index.

Licensing. All software developed under this project will be freely available for educational, research, and non-profit purposes under the terms of permissive software licenses, including MIT, Apache 2.0, and the Berkley Software Distribution version 3 (BSD-3) software license. These licenses meet all three requirements of the RFA: 1) they make the software freely available to researchers, healthcare systems, and research institutions and organizations; 2) they permit the dissemination and commercialization of enhanced and/or customized versions of the software and the incorporation of the software or components of it into other software packages; and 3) they include the ability of individuals outside UF and its collaborating institutions to modify the source code and to share modifications with other colleagues. Under BSD-3, no claims of suitability, and no warranty of any kind, are made. Under BSD-3, software can be modified, but under no conditions can anyone assert ownership over the code or its modifications, including commercial entities.

Availability. We commit to making the software developed in this project freely available under the terms of one of the aforementioned licenses to all biomedical researchers, educators and institutions in the non-profit sector such as education, research institutions and government laboratories. Software will be available for free, unrestricted download from UF and/or an open-source repository like SourceForge and GitHub. For example, the REDCap Electronic Data Integrator (RED-I) software discussed in the Research Plan is freely downloadable under BSD-3 from the GitHub site at: https://github.com/ctsit/redi/.

Open source community. In addition to making source code for the software publicly available, the CTSI Biomedical Informatics (BMI) Program will cultivate an active open-source development community by providing extensive developer documentation and plug-in architectures enabling others to contribute new functionality to the software. For example, the BMI Program collaborated with Stanford University to deliver their click-able image map software as a REDCap module that is delivered via REDCap's new External Module framework.

Enhancements. We anticipate that a community of practice will develop around each major software application released during this work, and that this community will support the software application after the proposal period. Community activity includes the submission of enhancements for inclusion in future releases. The R Project for Statistical Computing and VIVO project are examples of vibrant open source communities supporting complex software systems for research and translation. The communities fostered will operate in a similar fashion, establishing an archive and providing mirror sites for downloads, as well as on-line technical

support through a blog and wiki. The VIVO community, now hosted by the Duraspace organization, speaks to the CTSI's ability to create self-sustaining open-source communities.

Required software components. Software developed by the CTSI strives to reuse other open source components and not require any commercial software to run or host the software. Specifically, VIVO requires the use of the open-source Apache Tomcat, the Jena semantic web open-source library, and the open-source Virtuoso triple store. Shibboleth is required for support of federated identity use cases. VIVO, RED-I and their required components can be run on a wide variety of operating systems, both open source and commercial, and on a wide-range of commercially available hardware. It is strongly recommended that all CTSI-created software be deployed in accord with all institutional information security and privacy requirements.

Other research resources sharing plan. We also plan to share other research resources with the CTSA network and beyond, including policies, practices, education and training materials, model consent forms, and templates for memoranda of understanding and legal agreements and contracts. In addition to free and open-source software, we are also committed to free and open access to data standards including biomedical ontologies. Just as researchers are free to use SI units of measure and the chemical symbols of the elements without restriction or need to purchase a license, they ought to be free to standardize scientific data using ontologies in a free and open manner. Furthermore, the BD2K program has a strong interest in creating and fostering scientific communities that create and support such open standards. These communities enable the standards to change at the pace at which science changes, and to be consistent with the latest scientific knowledge. We therefore commit to releasing under the Creative Commons CC-BY license all ontological artifacts whose creation and ongoing development and maintenance is under its control, including the demographics ontology, the drug ontology, and the ontology of medically related social entities. The CC-BY license allows free use of the ontology without restriction, including commercial use, with proper attribution of the source of the ontology. Furthermore, personnel in the biomedical informatics program of the CTSI are active participants in the communities supporting other ontologies released as free and open resources, including the ontology for general medical science, the ontology of biomedical investigations, the biobanking ontology, and more.