Scientific Cyberinfrastructure Program

Presenter: Dr. Jennifer Fostel, Office of Data Science, NIEHS/DNTP

Program Management Team: Scott Auerbach, Jeremy Erickson, Jennifer Fostel, Nicole Kleinstreuer, Kamel Mansouri*, Charles Schmitt, Vickie Walker

**Temporary program management team member*

Problem Statement

The Division of the National Toxicology Program (DNTP) programs, branches, and offices rely heavily on the use of cyberinfrastructure resources (see Appendix A for definition) in support of work and project activities. This usage covers traditional enterprise information technology (IT) capabilities such as desktop computing and telecommunications, research computing capabilities that include high-performance computing and storage, and scientific computing capabilities for predictive modeling. This usage also includes informatics and data science capabilities for extracting, organizing, annotating, analyzing, and visualizing scientific data. These capabilities are both tactical in meeting day-to-day DNTP needs as well as strategic in enabling new ways to conduct toxicology research. These capabilities are also critical in DNTP's mission to provide access to the scientific data, knowledge, and interpretations that DNTP generates as part of its work.

As DNTP executes its mission to transform toxicology to be more predictive, human-relevant, and timely through innovative tools and strategies, we have opportunities to advance predictive toxicology, data-driven science, and computational science as well as to innovate in how we deliver data and knowledge products through investments enabling cyberinfrastructure tools, expertise, and new methods. Likewise, we have opportunities to improve laboratory and data information management processes by adopting improved data and metadata capture and processing tools and operating procedures.

These opportunities come at a cost as new capabilities are expensive to acquire, operate effectively, and sustain over time. A significant challenge is that capabilities are spread across a multitude of resource providers, including traditional IT offices, scientific computing and data science offices, offices and branches providing informatics support, and scientific contracts that have informatics, data science, and scientific computing capabilities. The Scientific Cyberinfrastructure (SCI) program will work to strategically plan for cyberinfrastructure capabilities in anticipation of evolving DNTP needs and coordinate efforts across providers, recognizing the interdependency of IT, scientific computing, data science, and informatics.

Objectives

The objectives of the SCI program are to:

1. Provide oversight, governance, and prioritization for DNTP SCI investments, ensuring that investments meet DNTP needs and that resource providers can plan for and meet requests.
2. Engage and partner with other SCI providers to increase capability and lower costs, including with the National Institute of Environmental Health Sciences (NIEHS) and National Institutes of Health (NIH) SCI providers, as well as external providers.

3. Provide strategic capabilities for the DNTP toxicology pipeline, including ensuring DNTP data are managed according to FAIR and TRUST principles[1], providing SCI resources to support predictive toxicology, and advancing automation of evidence-based informatics.

The SCI program is divided into key areas based on project type, the nature of resources involved, related staff expertise and skill sets, and the NIEHS and DNTP organizational structure and operational model. Each key area in turn has a team and associated coordination and governance processes that reside within the overall SCI program.

The current key areas include:

1. Core Key Area (Core): focuses on basic cyber infrastructure needed to support DNTP, including servers, storage, networking, research software, and middleware systems. The Core works closely with NIEHS and NIH IT providers to ensure DNTP needs are met.
2. Evidence Informatics Key Area (EI): focuses on developing and supporting tools and methods for the retrieval, extraction, and interpretation of evidence from literature and external knowledgebases.
3. ToxChem Informatics Key Area (TCI): focuses on developing and maintaining tools, databases, and methods supporting the application of chem-informatics and tox-informatics.
4. Data Management Key Area (DM): focuses on the collection, management, and publishing of National Toxicology Program (NTP) data.
5. Knowledge Management Key Area (KM): focuses on the creation and application of metadata to describe DNTP data and to facilitate linkage of NTP data to broader biomedical structured knowledge (e.g., ontologies, pathways, annotations).

Rationale

The rationale for maintaining a coordinated SCI effort and program management team is twofold.

From a pragmatic perspective, DNTP has ongoing requirements to maintain useful SCI resources as well as to develop new resources in response to evolving program, branch, and office needs. Often it is the case that needs can be met by more than one SCI resource provider, including those directly managed by DNTP as well as providers from within the NIH and federal government. It is important for DNTP to maintain a portfolio-wide perspective of existing and future investments in SCI resources as well as use of resource providers in order to prioritize investments, make good choices in selecting providers, make recommendations on implementation that fits with overall SCI investments, plan for providing future SCI capabilities, plan for sunsetting resources, and provide adequate training of staff.

DNTP programs and studies generate a wealth of toxicology data that include multiple measurement types (e.g., pathology, expression, behavioral, imaging) on multiple different in vivo and in vitro biological systems. From a data science perspective, ensuring DNTP maximizes the value of these data is important to ensure the investment in producing data is realized. There are multiple dimensions to this challenge. First, the data and metadata must be collected, described, and made available according to

[1] FAIR principles = findable, accessible, interoperable, and reusable; TRUST principles = transparency, responsibility, user focus, sustainability, and technology

FAIR principles[2], and DNTP data repositories must operate according to TRUST principles[3] to ensure DNTP meets community expectations for managing and delivering data and data-related products. Second, research data often have value beyond immediate study goals. For example, combining data across DNTP studies can allow data exploration tools to support researchers in recognizing commonalities across chemical and biological mechanisms and can support tools used to guide chemical and assay selection decisions to reduce redundancy in testing and fill important gaps in developing predictive models. Finally, DNTP consumes biomedical/toxicological data and knowledge from external sources (e.g., Comparative Toxicology Database, the U.S. Environmental Protection Agency's (EPA) Chemical Dashboard) and integration of these data with DNTP data is important to support the increasing use of informatics and predictive approaches with the DNTP toxicology pipeline. Ensuring investments in generating and consuming toxicological data—and ensuring that these investments are aligned with external efforts—requires DNTP-wide coordination and strategic planning.

Public Health Context

The SCI program directly supports DNTP programs, branches, and offices in meeting their public health goals and ensuring that DNTP-generated toxicology data, analysis results, and findings are made accessible and useable to advance public health.

Alignment with Mission, Goals, Strategic Pipeline

The SCI program provides strategic planning, coordination, and oversight of the tools and data generated and consumed by the toxicology pipeline. In addition, the program supports DNTP programs, branches, and offices in their alignment with DNTP's vision and mission and supports several portions of the DNTP toxicology program, including tools and methods for literature mining, data mining, predictive modeling, and data analytics, as well as data acquisition and processing pipelines and data repositories for handling data collected by laboratories.

Stakeholder Interest and Engagement

Steps Taken to Engage Stakeholders

The SCI program primarily serves DNTP programs and branches, DNTP staff, and select external stakeholder groups. These stakeholders are routinely engaged by SCI program members through a combination of regular project and work activities that involves stakeholders, requests made to the SCI program and to resource providers represented by the SCI program, and periodic interviews with stakeholders.

---

[2] Wilkinson, M. *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

[3] Lin, D. *et al.* (2020) The TRUST Principles for digital repositories. Sci Data 7, 144. https://doi.org/10.1038/s41597-020-0486-7

Ongoing and Continuing Interactions

Stakeholder*	Issue	Role of Stakeholder
EPA	EPA has similar challenges and goals in supporting toxicology practice; EPA staff collaborate with DNTP in the areas of evidence informatics, tox-informatics, data management, and knowledge management	Collaborators
International Regulatory Organizations (e.g., Health Canada, ECHA, EFSA, other OECD member countries agencies)	User of DNTP data, advisor to DNTP database and tool development (Chemical Effects in Biological Systems, Integrated Chemical Environment (ICE))	Partner; user
ICCVAM Federal Agency Partners (e.g., FDA, CPSC, DoD)	User of DNTP data and advisors on development of ICE-specific tools, data sets, and database	Users; advisors
Comparative Toxicology Database	User of DNTP data	Users
Ontology for Biomedical and Biological Investigations (OBI)	DNTP makes use of OBI ontology and standards and occasionally requires updates to the ontology	Collaborators
Developers of Evidence Informatics (e.g., Allen Institute for AI, Evidence Prime, Oak Ridge National Labs, NIST)	Developers of EI tools and methods	Collaborators; technical advisors
Developers of General Informatics (e.g., SCIOME, ICF)	Developers of general informatics tools and methods	Technical advisors

* CPSC = U.S. Consumer Product Safety Commission; DoD = U.S. Department of Defense; ECHA = European Chemicals Agency; EFSA = European Food Safety Authority; EPA = U.S. Environmental Protection Agency; FDA = U.S. Food and Drug Administration; ICCVAM = Interagency Coordinating Committee on the Validation of Alternative Methods; NIST = National Institute of Standards and Technology; OECD = Organisation for Economic Co-operation and Development

Input Received

Generally, feedback to different program areas by stakeholders and collaborators has been very favorable toward plans. In many cases, DNTP staff in the SCI program areas have ongoing collaborations through which feedback continues to influence program directions. For example, DNTP works closely with EPA on systematic review approaches and data sharing, and some degree of co-evolution of products/methods is informed by these collaborations. In other cases, stakeholders have provided specific feedback that is now adopted into program plans, such as providing access to individual animal data through the CEBS [Chemical Effects in Biological Systems] website.

Milestones and Metrics

Milestones and metrics are specific to key areas and projects, and each key area has its own set of objectives and processes that drive the choice of projects and related activities. The SCI program reviews objectives across key areas to make recommendations both to the key area teams and to DNTP management.

Core Key Area

The Core area seeks to provide basic solutions that are common across cyberinfrastructure needs within DNTP. Many of the resources for DNTP core needs come from existing core providers, in particular the NIEHS Office of Information Technology, the NIEHS Office of Scientific Computing, the NIEHS/NTP Office of Data Science, the NIH Center for Information Technology, and the NIH Biowulf. The SCI program serves to coordinate communications with these providers.

Milestones and measures of progress for achieving the Core objective is outlined below. Milestones are organized into short-term (1 year), medium-term (2–3 years), and long-term (4–5 years) targets.

Short-term:

- Provide monitoring of scientific hardware and middleware (e.g., web servers, databases) usage by DNTP to inform lifecycle planning for scientific servers and storage.
- Finish the transition to a common suite of tools and platforms for development and deployment of scientific tools and workflows, including RStudio Connect, Knime, Postgresql, nginx, Python Django/Flask, Oracle, and GitHub/GitLab, as well as common chem-informatics and tox-informatics tools (e.g., rdkit).

Medium-term:

- Increase efficiency in delivering scientific tools by developing security, privacy, and compliance standard operating procedures and reducing effort to address security issues.
- Provide targeted training for using common tools and platforms.
- Increase integration of high-performance computing and storage into tools and workflows.
- Plan and promote greater usage of cloud computing.
- Establish and enact best practices for sustaining investments in scientific tools and databases.

Long-term:

- Provide a data science training curriculum that is customized for toxicology and environmental health data sets, including those generated by DNTP.
- Transition to a hybrid infrastructure solution that includes on-premise, cloud, and contractor hosted solutions.

Evidence Informatics Key Area

The EI area focuses on the extraction and synthesis of existing scientific knowledge that is encoded in published scientific literature (e.g., PubMed, gray literature). The domain of EI often requires intensive human effort to search for, parse through, capture, and analyze literature to enable synthesis of existing knowledge to support decision-making, hypothesis generation, and communication of findings. For instance, conducting a systematic review generally requires human experts to read through thousands of documents to identify and extract research findings that are relevant to the review being conducted. Human experts ensure high-quality results but also pose limits on how frequently EI approaches can be applied and on the scope of application. The volume of scientific knowledge is growing rapidly, compounding the overall challenge. The EI area seeks to identify and advance the use of automated and semi-automated approaches that can assist human experts in their workflows and offers the possibility of expanding the use of EI approaches as part of the translational toxicology pipeline to improve efficiency while maintaining superior quality.

Within DNTP, EI can be applied across three broad areas:

- Fit-for-purpose literature-based assessments that employ the rigor and transparency of systematic review methodologies. Scoping review and systematic evidence maps summarize,

categorize, and display the extent and types of evidence to inform decision-making. Systematic reviews use the multistep process to synthesize, integrate, and develop hazard conclusions.
- Ensuring NTP researchers and staff have access to best-of-breed tools for evidence informatics, including ongoing monitoring and evaluation of new commercial and open-source solutions.
- Impact assessments that look to quantify the impact of DNTP products (e.g., technical reports, monographs) on policy makers, science, and the public.

Milestones and measures of progress for achieving the EI objective are outlined below. Milestones are organized into short-term (1 year), medium-term (2–3 years), and long-term (4–5 years) targets.

Short-term:
- Finalize Dexter, a custom tool to support environmental health data extraction projects that combines human-based and machine-based capabilities for annotating text, extracting concepts, grouping concepts into higher order constructs, and mapping text to standard terms.
- Identify, investigate, and procure commercial knowledge tools (e.g., Pharmapendium, Causaly, Qinsight) to support and improve existing evidence informatic pipelines.

Medium-term:
- Extend the capabilities of Dexter to include other evidence streams (e.g., epidemiological studies and in vitro or mechanistic data) and expand ontologies and standard vocabularies.
- Extend capabilities to extract information from tables and figures of published literature and automate the process of identifying high-quality scientific studies in the literature.
- Release an initial version of Citation Finder, a custom tool to aid in identifying references to DNTP publications within published and gray literature to support impact assessment.
    - Evaluate use of Citation Finder in practice and develop plans for any subsequent modifications.
- Develop training data sets to construct, train, and test new machine learning models.

Long-term:
- Building on previous work, leverage DNTP capabilities to automate/streamline literature-based approaches to provide living systematic evidence maps to monitor trends and inform decision-making.
- Explore the potential to partner and advance publishing practices of scientific literature to improve the availability of machine-readable data formats.

ToxChem Informatics Key Area

The TCI area focuses on organizing, integrating, analyzing, and presenting a diverse collection of toxicology-related data. This scope entails the creation, updating and maintenance of databases, analysis pipelines and software tools for access, visualization, analysis, and modeling of data. These tools can be accessed through expert channels, and web-based tools have been created that democratize access for all DNTP/NIEHS scientists. Overall, the TCI efforts are intended to facilitate access to a complex collection of information in easy to-use formats to allow for data exploration, mining, and knowledge development. The specific products that emerge from these efforts are diverse in nature and continue to evolve with technological changes; however, at a fundamental level, the efforts are focused on finding linkages between different types of information—from chemistry through molecular targets/pathways, phenotypic changes, functional changes, and disease. These efforts support a variety of needs, including:

- Chemical information gathering.
- Screening level characterization of chemicals; chemical testing prioritization.
- Structure-activity relationship modeling.
- Dose-response modeling.
- In vitro and in vivo high dimensional models predictive of toxicity and disease.
- In vitro to in vivo extrapolation (IVIVE) and forward toxicokinetic modeling.

Major components of the TCI portfolio include:

- The Integrated Chemical Environment (ICE), which includes a public-facing database of model-ready data sets, as well as tools and computational workflows (e.g., IVIVE, PBPK [physiologically based pharmacokinetic modeling], Chemical Characterization) that support alternative testing methods.
- BMDExpress, which includes capabilities for conducting benchmark dose analysis on toxicogenomic data, including integration with pathway analysis.
- A Tox21 toolbox that includes databases of Tox21 data and tools for analyzing and visualizing Tox21 dose-response curves.
- A chem-informatics toolbox that includes tools for processing chemical structures and mapping chemical space.
- A tox-informatics toolbox that includes the DrugMatrix and TG-Gates databases, tools for exploring relationships of biological measures (expression, pathology, clinical chemistry, phenotype), and tools for structural and biological read-across and association mapping across different levels of biological/toxicological complexity.
- BioChemDB, a database that integrates external data sources on chemical and biological annotations that are frequently used in analysis.
- DNT DIVER [Developmental NeuroToxicity Data Integration and Visualization Enabling Resource] tool set for reporting and exploring in vitro and zebrafish-based neurotoxicology data sets.
- OPERA [OPEn structure-activity/property Relationship App], a suite of QSAR/QSPR [quantitative structure-activity relationship/quantitative structure–property relationship] models predicting physicochemical properties, environmental fate, ADME [absorption, distribution, metabolism, and excretion], and toxicity endpoints of organic chemicals.

The TCI area needs to balance the development of highly custom, fit-for-purpose tools, software libraries, and workflows with the development of common capabilities to gain efficiencies. To support this balance, the involved teams are building a common infrastructure that includes: 1) custom public-facing tools using common libraries/languages, 2) common reusable workflows for data processing and executing predictive models, 3) core set of databases that aggregate frequently used chem-bio-tox data and model outputs, and 4) common development/operations environment (e.g., containers, packaging, testing, deployment).

TCI activities and products are targeted at short-term needs, so milestones are organized into either short-term (1 year) and medium-term (2–3 years) targets.

Short-term:

- Provide ongoing updates to ICE based on the Interagency Coordinating Committee on the Validation of Alternative Methods/NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (ICCVAM/NICEATM) roadmap.
- Provide ongoing updates to BMDExpress based on user requests.
- Consolidate Tox21 tools and data sets in the Tox21 toolbox and make publicly available.
- Release in-development tools from the chem-informatics and tox-informatics toolboxes.

- Finish phase 1 of BioChemDB.
- Extend DNT DIVER for additional data and endpoints.

Medium-term:

- Improve underlying infrastructure, including databases and tools, through increased use of APIs [application programming interfaces], service mechanisms, packaged libraries, and packaged workflows.
- Consolidate and document chem-informatics tools and data sets in the chem-informatics toolbox and make portions publicly available.
- Consolidate and document tox-informatics tools and data sets in the tox-informatics toolbox and make portions publicly available.
- Continue the development and release of tools as needed to support the DNTP toxicology pipeline and branch, program, and office needs.
- Explore the use of cloud-based and on-premise hybrid cloud solutions for tool development and deployment.

Data Management Key Area

The DM area focuses on the collection, management, and publication of all DNTP data. At present, this process has automated components that streamline data management and reporting, particularly for the DNTP testing program. However, in other areas this process is disjointed and ad hoc, leading to inconsistencies and gaps; this is a situation we are striving to improve. The goal is to provide a two-tiered solution for data. First, all DNTP-generated data will be accessible through the CEBS repository. CEBS provides a TRUST-worthy repository and permits the raw data to be tracked to the published results, and the published results to be available for new knowledge generation in the context of other DNTP data. Second, select data are available through custom data systems, such as ICE, which aggregates DNTP data and findings with external toxicology data to provide custom data targeted for alternatives-based toxicological assessments and development of predictive modeling, and the DNT DIVER application, which supports exploration and visualization of developmental neurotoxicology data.

Milestones and measures of progress for achieving the DM objective are outlined below. Milestones are organized into short-term (1 year), medium-term (2–3 years), and long-term (4–5 years) targets.

Short-term:

- Complete DNTP DM plan template and define process for integration of the DM plan with program proposals and the science to support them.
- Complete the design and delivery of the Histopathology Slide Review Module, an in-house application to streamline the pathology review process, capture DNTP decisions in an auditable database, and integrate DNTP tables with contract lab reports.
- Complete the automated data management pipeline for in vitro DNTP data, including pilots for DNTP laboratories and immunotoxicology studies.
- Design a proposal to extend the current repository for DNTP data files, and add metadata, to streamline deposition, retrieval, and reuse of unstructured data.

Medium-term:

- Support the creation and adoption of DM plans early in the project design process.
- Support program compliance with NIH Data Sharing Policy as it goes into effect in 2023.
- Complete the system to permit access to unstructured DNTP data and add tools for users to find, share, and reuse data.

- Complete the migration of all DNTP legacy data to integrated standardized resource for knowledge generation (pathology statistical results, genetic toxicology, reproductive and developmental toxicology, immunotoxicology, neurobehavioral responses, expression data), thereby enhancing compliance with FAIR data standards; provide support to DNTP, stakeholders, and the public in retrieving data from this resource.
- Complete automated data management pipelines (moving data and metadata to DNTP summary tables and integration of results with DNTP repository) for neurobehavioral data, qPCR data, and expression data.
- Initiate work on new automated data management pipelines identified by the program.
- In collaboration with other SCI program areas, work toward enhanced appreciation for data integration and data reuse at DNTP by adding tools to permit exploration and knowledge generation from DNTP integrated data.

Long-term:

- Complete the migration of all current DNTP data to an integrated standardized resource for knowledge generation, ensuring full compliance with FAIR standards
- Complete work to make CEBS a fully TRUST-worthy repository.
- Replace existing legacy code used to support DNTP testing program with streamlined and automated data management pipeline to reduce resources needed to manage these data.
- Complete interoperability of CEBS with other DNTP data resources for a seamless user experience.
- Working with DNTP colleagues to establish a culture to proactively respond to new data types so as to have data management plans and automated data management pipelines in place as data are gathered to avoid delays and data gaps.
- Migrate CEBS to a cloud-based platform that aligns with the goals of NIH in developing an overarching biomedical data ecosystem where data can be readily accessed, shared, and integrated.

Knowledge Management Key Area

The KM area is focused foremost on adopting standardized language, including controlled vocabularies and ontologies, in describing data sets as a necessary foundational step in fulfilling FAIR principles and facilitating integration of data from disparate sources. This goal is important both to facilitate integration and use of data within DNTP for analysis and decision-making, but also as a part of DNTP's responsibility to the broader community as a generator and provider of important toxicology data sets that are accessible and interpretable by both humans and machines, and that can be readily integrated with other data. The KM area also seeks to advance the adoption and application within DNTP of knowledge organization systems (KOS; e.g., knowledge bases, ontologies) that can support applications in predictive toxicology, computational modeling, and systems-levels tox-informatics. In addition, the KM area seeks to advance the translation of DNTP-generated knowledge into codified representation and systems (e.g., adverse outcome pathways (AOPs), Gene Ontology (GO)-causal activity models) that can be used broadly by the toxicology and biomedical community.

The KM portfolio includes:

- Defining common data elements and standardized vocabularies for describing DNTP data and metadata, in particular for: histopathology, clinical pathology, assays, and endpoints/treatment-related effects.

- Increasing adoption of common data elements and standardized vocabularies within DNTP databases, tools, and systematic review workflows, including development of semi-automated and automated process/pipelines.
- Providing tools to support collaborative development and sharing of vocabularies, within DNTP and with external collaborators.
- Aligning legacy terms used in information systems with modern terms.
- Incorporating external curated IDs into DNTP data systems to allow linkage to other knowledgebases and databases, especially for treatments (e.g., chemicals).
- Developing reference toxicological data sets using standardized terms to support broad community use (e.g., Tox21, ICE).
- Providing human- and computer-readable annotations for DNTP data sets (e.g., annotations for Tox21 assay endpoints).
- Constructing mappings between DNTP exposure measures and biomedical ontologies to facilitate knowledge-based analytical methods and tools.
- Extending biomedical ontologies to incorporate toxicology processes mapped out by DNTP programs.

Milestones and measures of progress for achieving the KM objective are outlined below. Milestones are organized into short-term (1 year), medium-term (2–3 years), and long-term (4–5 years) targets.

Short-term:

- Establish a community of practice to coordinate efforts with EPA on standards.
- Align existing terminologies across related efforts (DNTP, NICEATM, and collaborating EPA programs) into a DNTP data dictionary.
- Continue work to establish links between DNTP metadata and external IDs to support interoperability (ongoing as data increases and the data dictionary grows).
- Incorporate knowledge management component to data management plan (how to define new metadata and incorporate standard metadata, how data will add value to DNTP).

Medium-term:

- Ensure DNTP-generated research data sets are described using standardized metadata that renders them searchable and retrievable.
- Develop semi-automated processes for annotating DNTP-generated research data sets using DNTP Dictionary and other controlled vocabularies and ontologies.
- Enable the integration of DNTP data sets with one another for enhanced data value to DNTP and the public.
- Define standard format across DNTP scientific databases to house data in a way so as to facilitate modeling and knowledge generation.
- Work with stakeholders to coordinate data models and KOS to facilitate interoperability.
  - Support the use of KOS and natural language processing to support systematic review (in coordination with NTP's Integrative Health Assessments Branch).
  - Increase harmonization and use of KOS in ICE, CEBS, DNT DIVER, and SEAZIT projects; coordinate with EPA's Chemical Dashboard team to increase interoperability.

Long-term:

- Ensure all NTP-generated research data sets meet community-based FAIR principles.
- Enable the integration of well-annotated NTP-generated research data sets with external data resources.

- Leverage existing resources (e.g., OBI, NextBio, GO, KEGG, Ingenuity, WikiPathways, AOPs) and communities (OBO Foundry, Monarch, AOP) to map data set targets to biological pathways and facilitate modeling activities.

Value Proposition and Summary

The SCI program is fundamentally about coordinating and communicating internal efforts to build up an informatics and data infrastructure that is needed to support DNTP program and branch work and to ensure that DNTP meets its community responsibility in providing research data, knowledge, and tools to the public.

DNTP generates unique data in reporting the responses of biological systems—such as cells, organoids, and organisms—to controlled exposure of agents of unknown toxicity. Integrating these data with data from reference sources and then applying DNTP tools and analyses permits the discovery of underlying pathways and syndromes, which leads to increased understanding of toxic effects of classes of test articles and supports translation to human risk. Integration and annotation of DNTP data also make the data more available for knowledge discovery. Moreover, use of external standards allows the wider community to access the information using common terms and integrates DNTP data with other data and information sources.

When a new chemical exposure is detected, rapid response with solid and interpretable information about the effects of the chemical in model systems is essential to understanding potential human risk. By putting DM pipelines in place to manage and report data rapidly, and applying KM for curated annotation, DNTP can rapidly prepare scientific data for presentation and interpretation. With its spectrum of tools and other tox-informatics approaches, DNTP can rapidly translate experimental findings into the context of other toxicological data and thereby inform potential human health effects.

The coordination of infrastructure across DNTP groups is not without challenges or demands on project timelines and staff time commitments. However, without coordination of efforts and resources, the benefits of DNTP's investment in generation of data and application of its deep expertise to extracting knowledge and producing expert-based interpretations will not be fully realized.

Appendix A: Terminology

Scientific Cyberinfrastructure: computing environments that support scientific data acquisition[4], data storage[5], data management[6], data integration[7], data mining[8], data visualization[9], and other computing[10] and information processing services[11] and staff employed for supporting scientific discovery (derived from Wikipedia).

SCI resources: tangible components of scientific cyberinfrastructures, e.g., tools, applications, libraries, middleware, databases, data sets, hardware systems, ontologies and terminologies, and related training materials and courses.

SCI providers: providers of SCI components, e.g., commercial vendors, contract resources, government entities, open-source projects, academic labs, and independent contractors.

[4] https://en.wikipedia.org/wiki/Data_acquisition
[5] https://en.wikipedia.org/wiki/Data_storage_device
[6] https://en.wikipedia.org/wiki/Data_management
[7] https://en.wikipedia.org/wiki/Data_integration
[8] https://en.wikipedia.org/wiki/Data_mining
[9] https://en.wikipedia.org/wiki/Data_visualization
[10] https://en.wikipedia.org/wiki/Computing
[11] https://en.wikipedia.org/wiki/Service_(economics)