Scientific Cyberinfrastructure Program

Jennifer Fostel, PhD
Division of the NTP
National Institute of Environmental Health Sciences

NTP Board of Scientific Counselors Meeting
August 4, 2021

Membership

Liaison

| Charles Schmitt ODS | Scott Auerbach PTB | Jeremy Erickson ODS | Jennifer Fostel ODS/CEBS | Nicole Kleinstreuer PTB/NICEATM | Kamel Mansouri PTB | Vickie Walker IHAB |

ODS: Office of Data Science
PTB: Predictive Toxicology Branch
CEBS: Chemical Effects in Biological Systems (data management system)
NICEATM: NTP Interagency Center for the Evaluation of Alternative Toxicological Methods
IHAB: Integrative Health Assessments Branch

The Scientific Cyberinfrastructure (SCI) Program

What is the SCI Program?

The SCI Program is distinguished from the other programs that have been introduced previously

- o SCI is 1 of 2 programs charged with providing cross-cutting capabilities for DNTP

- o Focus is on Scientific Cyberinfrastructure

What is Scientific Cyberinfrastructure?

Computing environments that

- support scientific data acquisition
- support data storage
  and management

- provide data integration, annotation
  and analysis

- provide other computing &
  information processing services

Staff supporting scientific discovery*.

*Derived from Wikipedia and NSF Cyberinfrastructure
Vision (https://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf)

What do we mean by Scientific Cyberinfrastructure?

Computing environments that

- support scientific data acquisition

- support data storage and management

- provide data integration, annotation and analysis

- provide other computing & information processing services

Staff supporting scientific discovery*.

Enterprise IT

*Derived from Wikipedia and NSF Cyberinfrastructure
Vision (https://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf)

What do we mean by Scientific Cyberinfrastructure?

Computing environments that

- support scientific data acquisition
- support data storage and management
- provide data integration, annotation and analysis
- provide other computing & information processing services

Staff supporting scientific discovery*.

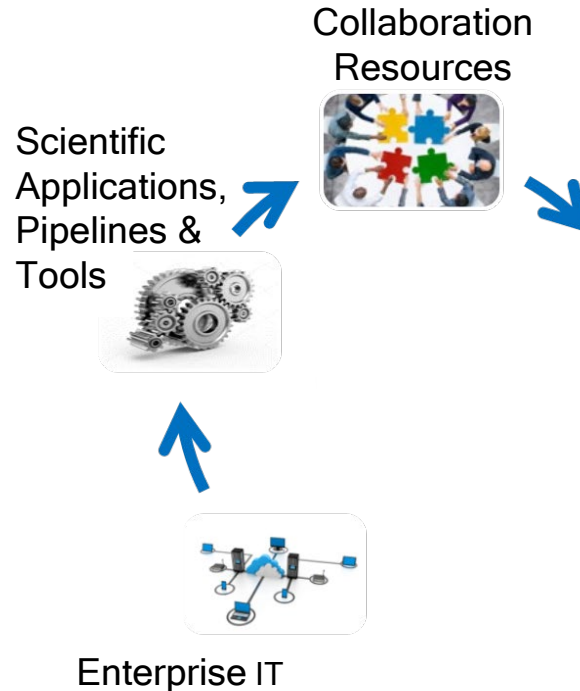Scientific Applications, Pipelines & Tools

Enterprise IT

What do we mean by Scientific Cyberinfrastructure?

Computing environments that

- support scientific data acquisition
- support data storage and management
- provide data integration, annotation and analysis
- provide other computing & information processing services
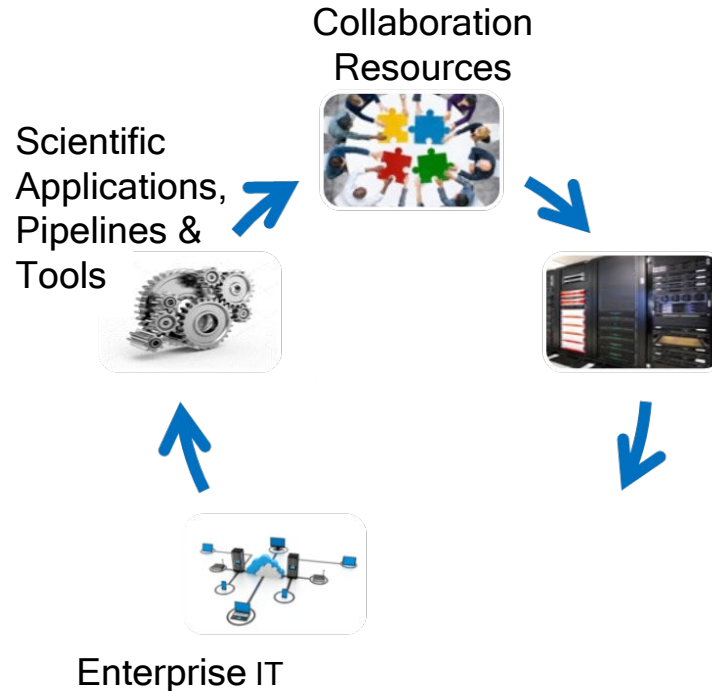
Staff supporting scientific discovery*.

Collaboration Resources

Scientific Applications, Pipelines & Tools

Enterprise IT

What do we mean by Scientific Cyberinfrastructure?

Computing environments that

- support scientific data acquisition
- support data storage and management
- provide data integration, annotation and analysis
- provide other computing & information processing services

Staff supporting scientific discovery*.

Collaboration Resources

Scientific Applications, Pipelines & Tools

Enterprise IT

*Derived from Wikipedia and NSF Cyberinfrastructure Vision (https://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf)
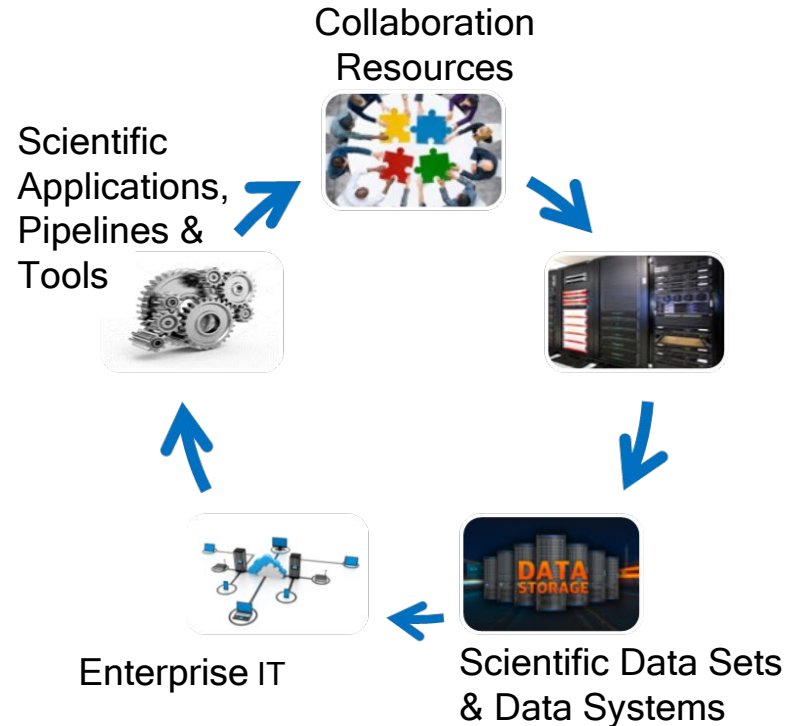
8

What do we mean by Scientific Cyberinfrastructure?

Computing environments that

- support scientific data acquisition
- support data storage and management
- provide data integration, annotation and analysis
- provide other computing & information processing services

Staff supporting scientific discovery*.

Collaboration Resources

Scientific Applications, Pipelines & Tools
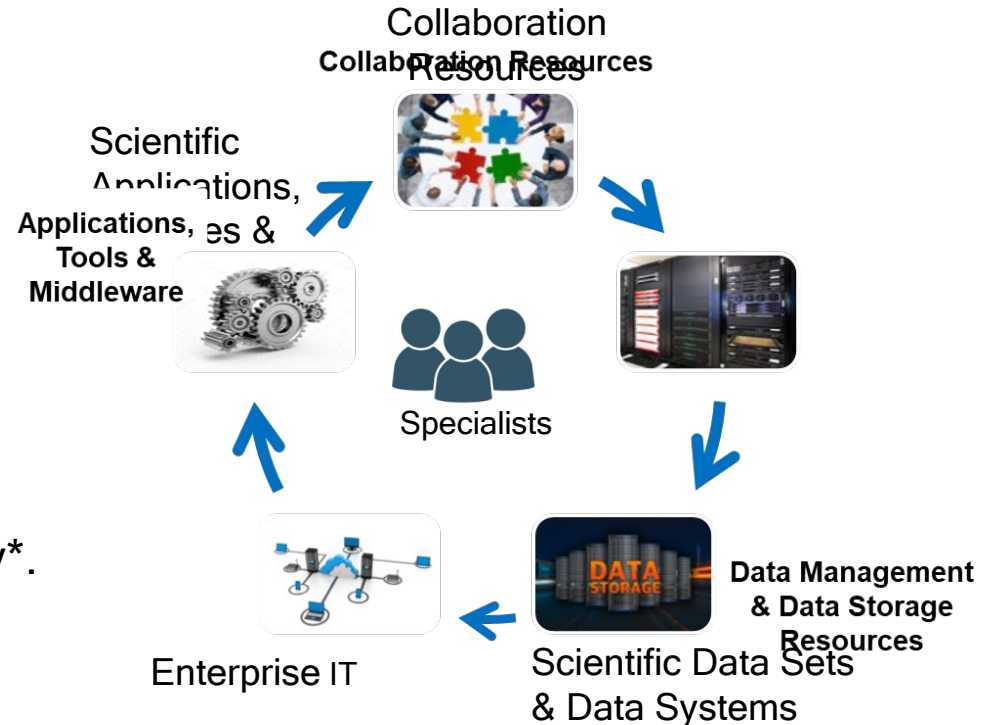
Enterprise IT

Scientific Data Sets & Data Systems

What do we mean by Scientific Cyberinfrastructure?

Computing environments that

- support scientific data acquisition
- support data storage and management
- provide data integration, annotation and analysis
- provide other computing & information processing services

Staff supporting scientific discovery*.

Scientific Applications,

Applications, Tools & Middleware

Collaboration
Collaboration Resources
Resources

Specialists

Enterprise IT

Data Management & Data Storage Resources

Scientific Data Sets & Data Systems

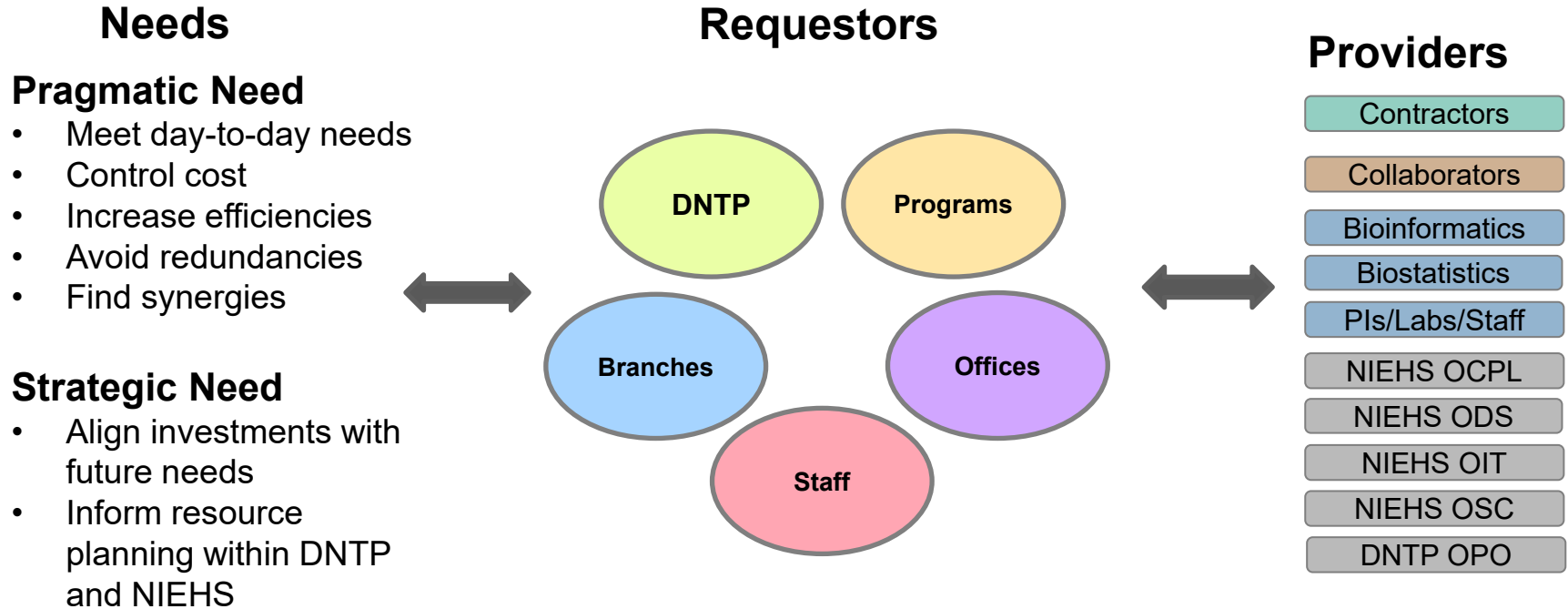# Definition of Scientific Cyberinfrastructure

SCI Program

- Core Key Area: Coordinate basic solutions common across DNTP SCI needs

- Data Management (DM) Key Area:
    - Collect, manage and publish all DNTP data
- Knowledge Management (KM) Key Area:
    - Adopt standardized language, metadata and data models within DNTP

- ToxChem Informatics (TCI) Key Area:
    - Organize, integrate, analyze and present diverse toxicology-related data

- Evidence Informatics (EI) Key Area:
    - Evaluate and support tools and methods for retrieval, extraction and interpretation of evidence from scientific literature and external knowledgebases

Challenge: Fulfilling needs across requestors and providers

Needs

Pragmatic Need
- Meet day-to-day needs
- Control cost
- Increase efficiencies
- Avoid redundancies
- Find synergies

Strategic Need
- Align investments with future needs
- Inform resource planning within DNTP and NIEHS

Requestors

- DNTP
- Programs
- Branches
- Offices
- Staff

Providers

- Contractors
- Collaborators
- Bioinformatics
- Biostatistics
- PIs/Labs/Staff
- NIEHS OCPL
- NIEHS ODS
- NIEHS OIT
- NIEHS OSC
- DNTP OPO

Objectives

- Support day to day and strategic needs

- Engage and coordinate stakeholders and partners

- Provide innovative tools and capabilities for DNTP mission

Oversight

Engagement & Partnerships

Strategic Capabilities

Governance teams
Oversight of generators
Prioritization advice
Informed decision-making

Oversight

Engagement & Partnerships

Strategic Capabilities
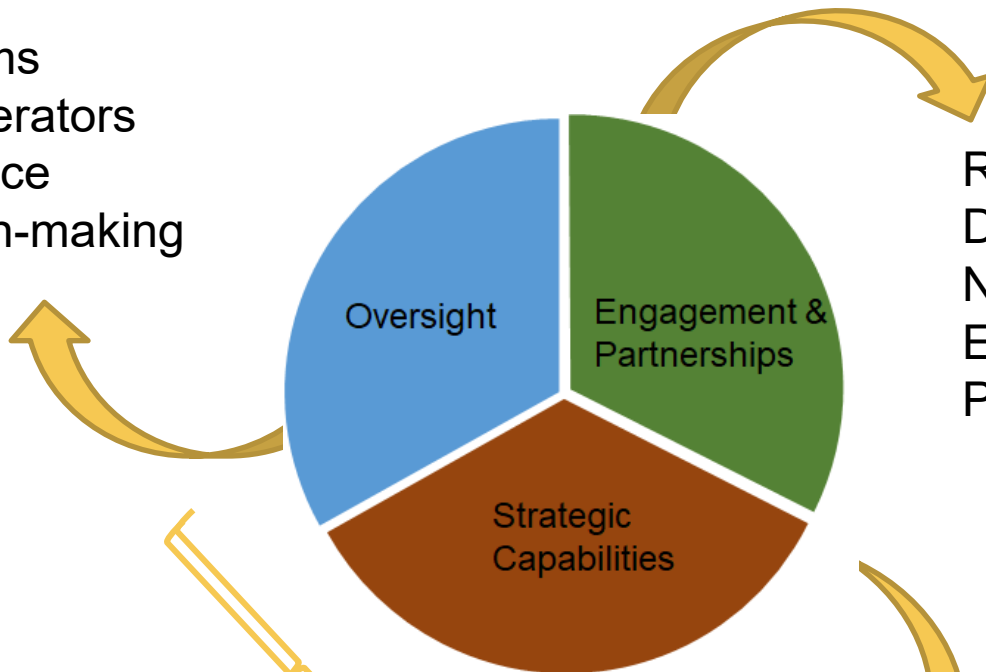
Resource providers
DNTP stakeholders
NIEHS / NIH
External stakeholders
Public users

Prioritized and coordinated capabilities across the DNTP pipeline

Data management pipelines
Knowledge management curation, annotation
ToxChem tools and analyses

Information for decision-making, prioritization, oversight of generators

- Lead Key Area Teams to provide advice, approval

- Information for decision-making by DNTP management

- Coordinate development of tools, analyses and pipelines

- Establish best practices for monitoring, leveraging and sustaining investment in scientific cyberinfrastructure

- Support DNTP compliance with NIH data sharing policy
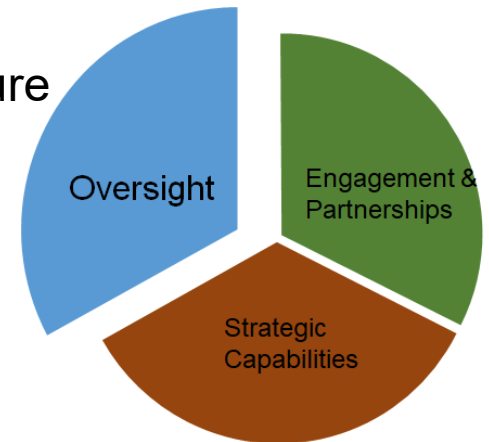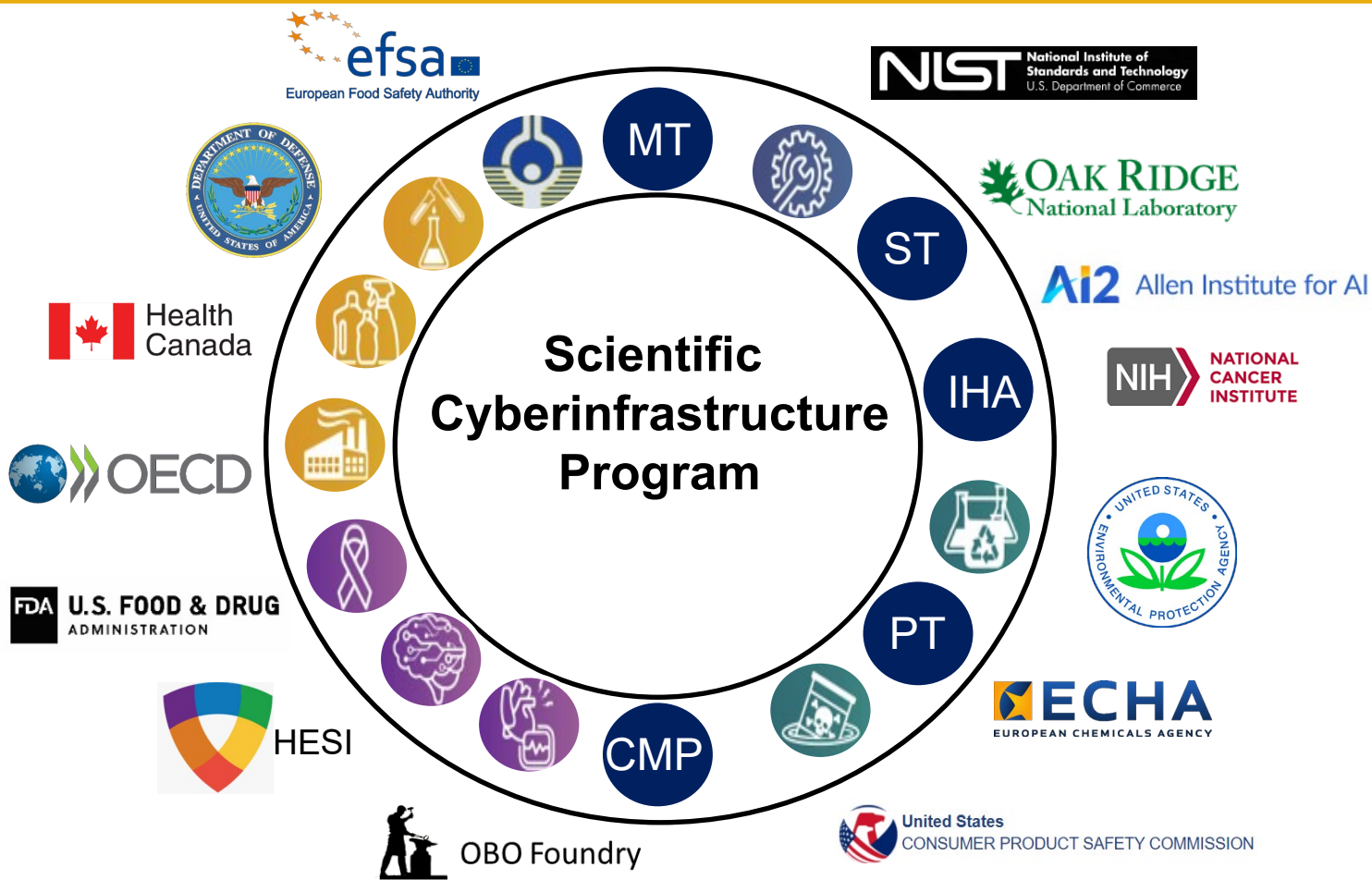
Engage and partner with SCI providers

- Data generators

- Tool builders

- Data managers

- Data and tool maintainers

- DNTP Branches

  - MTB: Mechanistic Toxicology Branch
  - STB: Systems Toxicology Branch
  - IHAB: Integrative Health Assessments Branch
  - PTB: Predictive Toxicology Branch
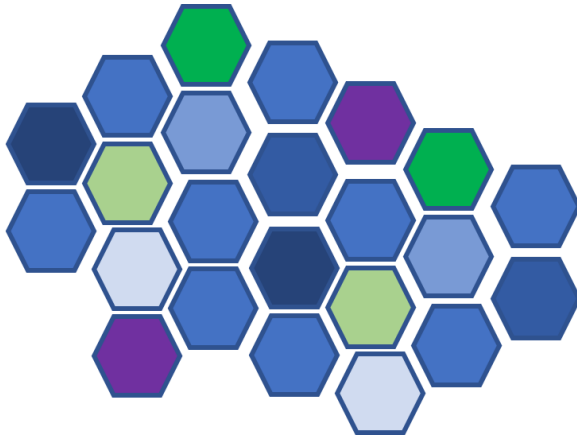 - CMPB: Cellular and Molecular Pathogenesis Branch

Engage with stakeholders

- Common terms

- Coordinate data models, knowledge organization systems across partners
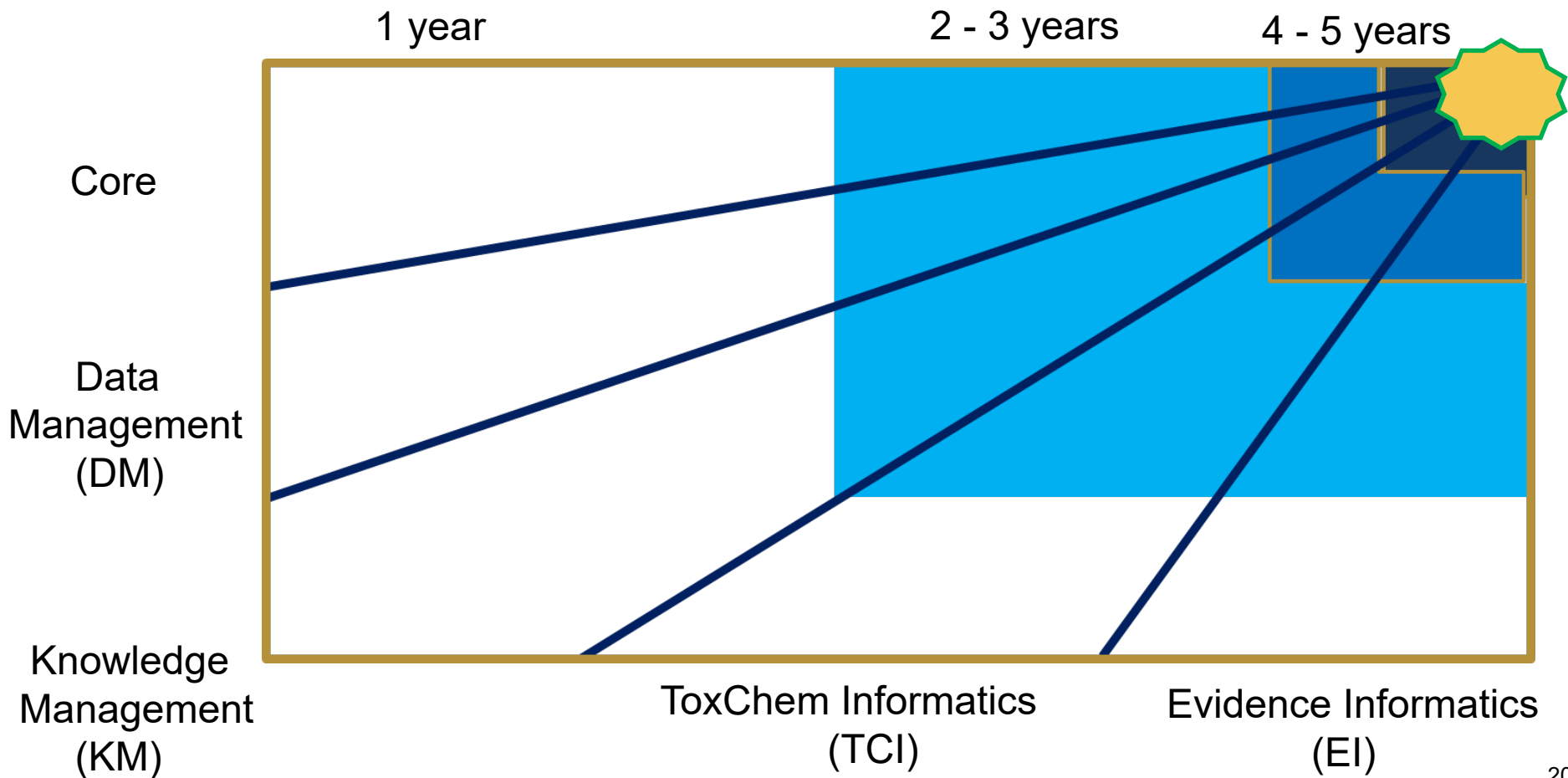
- Enhance tools for stakeholders

Faster, better, smarter

- SCI resources to support predictive toxicology

- Advance automation of evidence-based informatics

- Champion automation for data management

- Advance automation of curation and annotation

- Support knowledge-discovery from integrated DNTP data

ToxChem Informatics
(TCI)

Evidence Informatics
(EI)

Common solutions across SCI needs

4-5 years

2-3 years

Data Science Curriculum
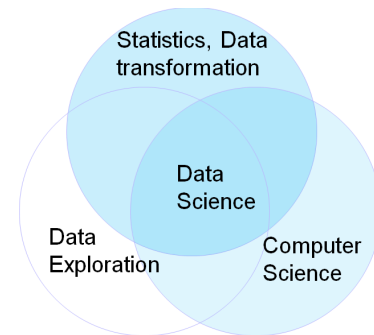Hybrid infrastructure

Establish
best practices to sustain investments
Targeted training in DNTP tools, data and platforms

1 year

Transition to a common tool set

Provide monitoring to inform lifecycle planning

Statistics, Data transformation

Data Science

Data Exploration

Computer Science

Leveraging data science

BMD Express 2.2
NTP · EPA · Health Canada · Sciome

Training on DNTP tools

Developing the DNTP IT Tool Kit

Access to best-of-breed tools for evidence informatics

4-5 years

Extend DEXTR to other areas of science; and formats such as graphs

2-3 years

Release Citation Finder; & evaluate performance
Partner for machine-readable literature

1 year

Provide evidence-based review standards
Finalize DEXTR phase 1
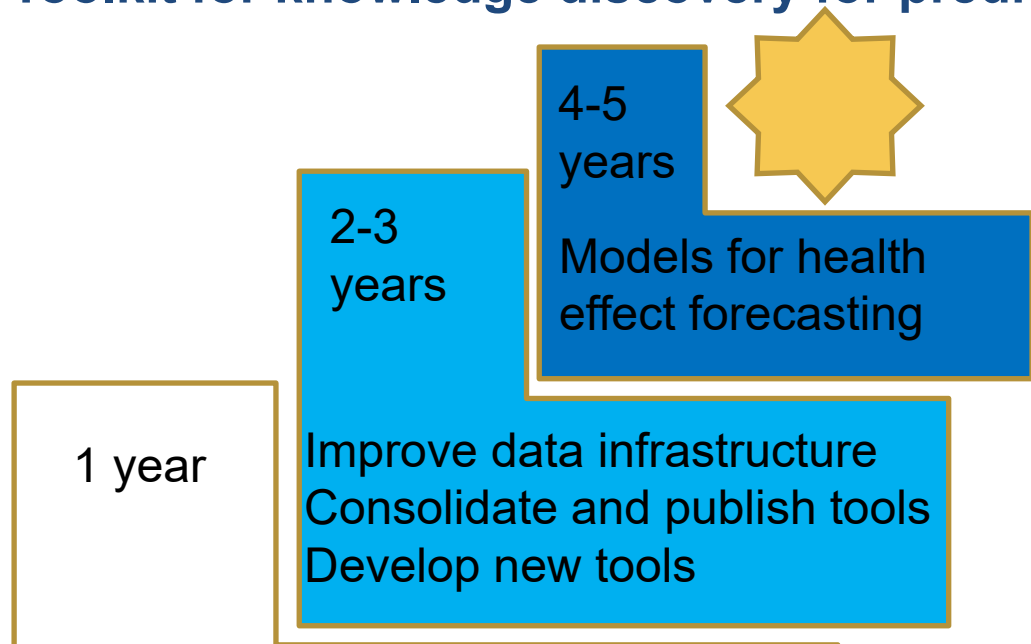Evaluate commercial knowledge tools
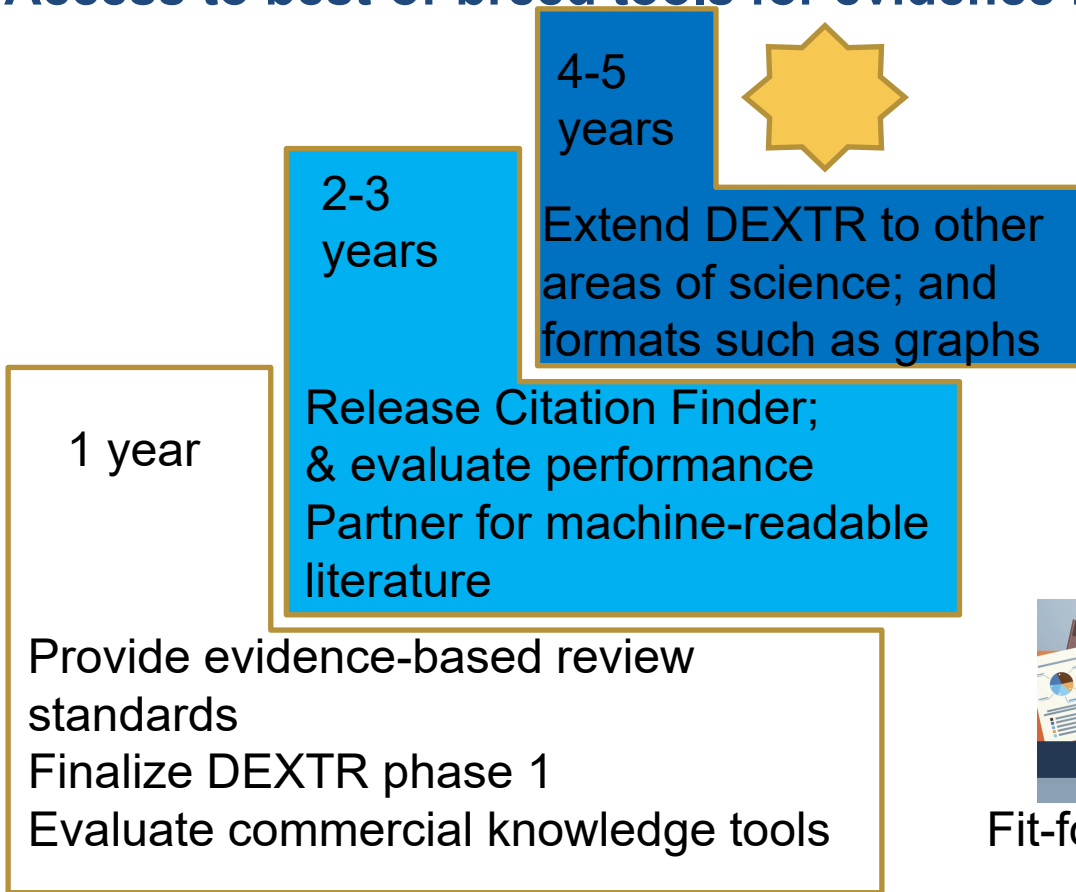
Access to essential tools

Assess impact of DNTP products
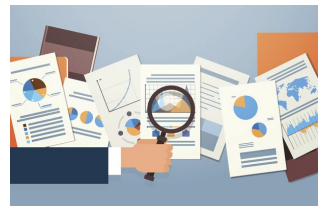
Fit-for-purpose literature assessments

Toolkit for knowledge discovery for predictive toxicology

4-5 years

Models for health effect forecasting

2-3 years

**Improve data infrastructure
Consolidate and publish tools
Develop new tools**

1 year

Update DNTP tools:
(ICE, BMDExpress, Tox21 toolbox,
ChemTox-informatics toolbox,
OPERA, BioChemDB, DNT DIVER)

Integrated Chemical Environment

BMD Express2.2

OPERA OPEn (q)saR App

NTP · EPA · Health Canada · ScitoVation

CEBS
Chemical Effects in Biological Systems

Tox21
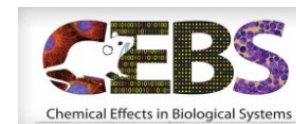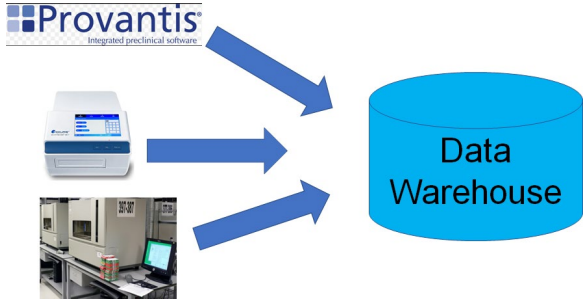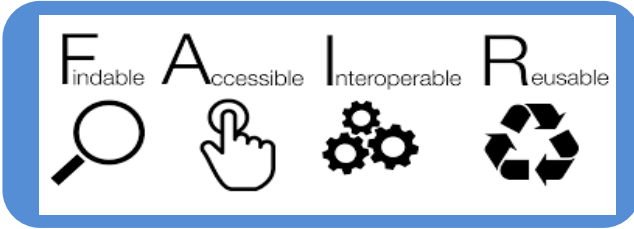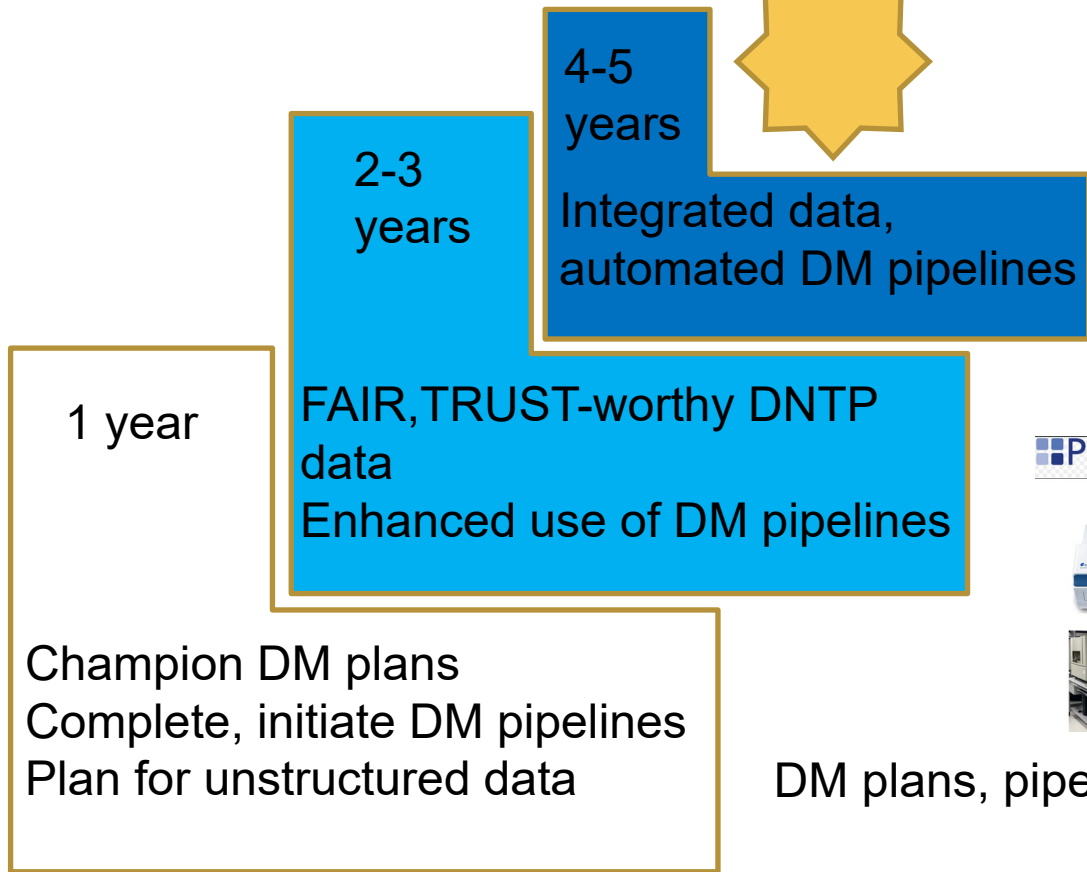
Developmental NeuroToxicity Data Integration and Visualization Enabling Resource (DNT-DIVER)
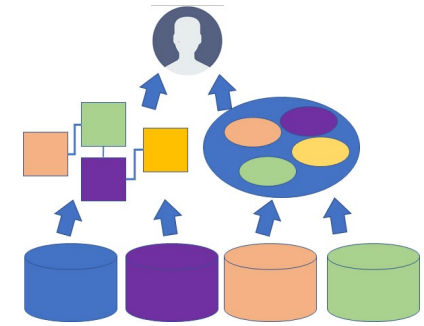
Automated data ecosystem

4-5 years
Integrated data, automated DM pipelines

2-3 years
FAIR,TRUST-worthy DNTP data
Enhanced use of DM pipelines

1 year
Champion DM plans
Complete, initiate DM pipelines
Plan for unstructured data

F indable   A ccessible   I nteroperable   R eusable

TRUST
Transparency   Responsibility   User focus   Sustainability   Technology

Transparency
Responsibility
User focus
Sustainability
Technology

Provantis
Integrated preclinical software

Data Warehouse

DM plans, pipelines automate data management

Curated data in biological context

4-5 years

DNTP data are integrated
Add biological pathways

2-3 years

Data Models and Knowledge Organization Systems (KOS)
Standard DNTP metadata
Standard DNTP data format

1 year

Establish Community of Practice
Initiate work on DNTP Data Dictionary
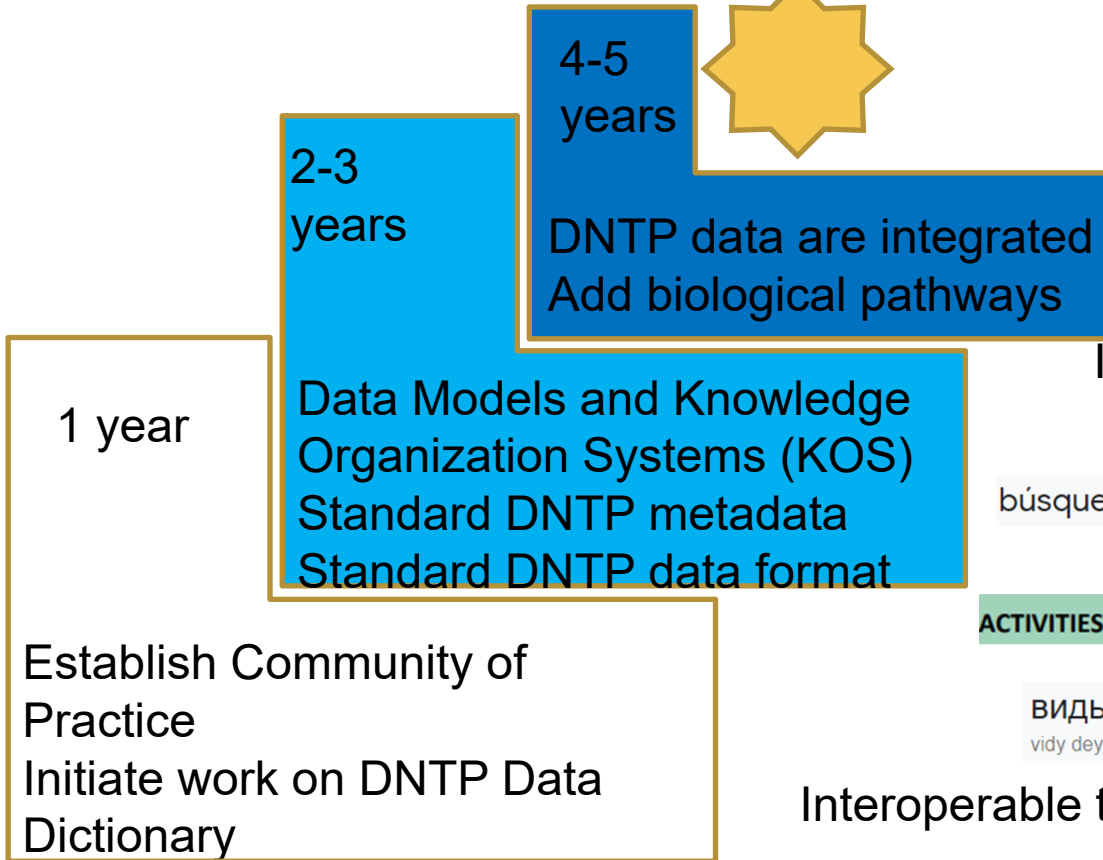
Interoperable data models, KOS

pursuits'

búsqueda

ocupaciones

ACTIVITY

ACTIVITIES

External reference: OBO Foundry

External reference: UMLS

виды деятельности
vidy deyatel'nosti
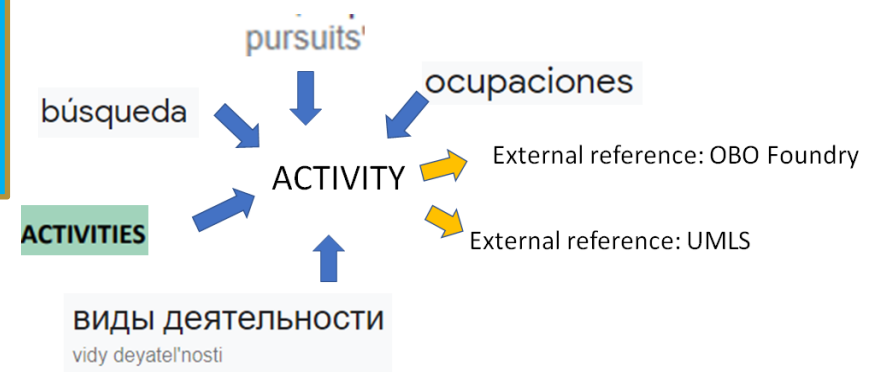
Interoperable terms linked to external reference
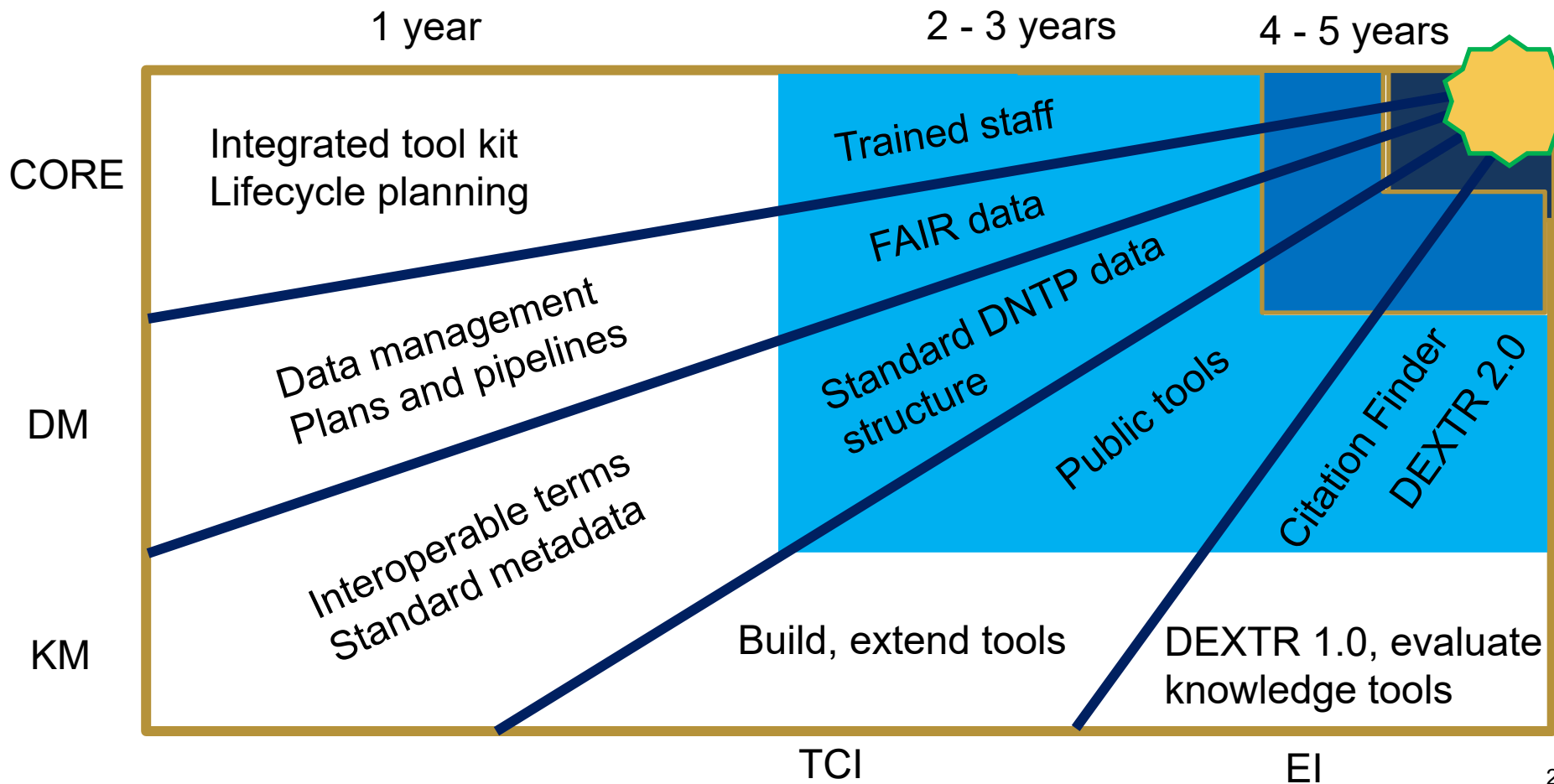
DNTP Staff

Mark Cesta
Frank Chao
Mike Conway
Jeremy Erickson
Steve Ferguson
Stephanie Holmgren
Jui-Hua Hsieh
Suril Mehta
Kelly Shipkowski
Vicky Sutherland
Amy Wang
Mary Wolfe

NIEHS Office of
Information Technology

NIEHS Office of
Scientific Computing

Management Team:
Warren Casey
Andy Rooney

Contractors

Arctic Slope Regional Corporation (ASRC)
Battelle
Burleson Research Technologies (BRT)
Charles River Laboratories
Experimental Pathology Laboratories (EPL)
Evidence Prime
ICF, International
Integrated Life Sciences (ILS)
Kelly Government Services
Maximus | Attain
Social and Scientific Systems (SSS)
Sciome
Southern Research

Thank You!

Membership

Liaison

| Charles Schmitt ODS | Scott Auerbach PTB | Jeremy Erickson ODS | Jennifer Fostel ODS/CEBS | Nicole Kleinstreuer PTB/NICEATM | Kamel Mansouri PTB | Vickie Walker IHAB |

ODS: Office of Data Science
PTB: Predictive Toxicology Branch
CEBS: Chemical Effects in Biological Systems (data management system)
NICEATM: NTP Interagency Center for the Evaluation of Alternative Toxicological Methods
IHAB: Integrative Health Assessments Branch

Abbreviations and Acronyms

IT	information technology
OCPL	Office of Communications and Public Liaison
OIT	Office of Information
OSC	Office of Scientific Computing
OPO	Office of Program Operations