CTSA Program Common Metric for Informatics Solutions

KRISTI HOLMES, PHD
DIRECTOR OF EVALUATION, NUCATS
DIRECTOR, GALTER HEALTH SCIENCES LIBRARY & LEARNING CENTER
NORTHWESTERN UNIVERSITY

CTSA PROGRAM STEERING COMMITTEE MEETING 6.12.2017

NCATS

NIH National Center for Advancing Translational Sciences

Metric Development Team

Informatics Metric Development Team NCATS Leads: Erica Rosemond and Ken Gersing C4 Coordinator: Stephanie Mayers		
Member	Name	Affiliation
Administrator	Thomas Fogg	Rochester
Administrator / Associate Director of Bioinformatics	Elizabeth Wood	Weill Cornell
Evaluator	Kristi Holmes	Northwestern
Evaluator	Patrick Barlow	University of Iowa
PI	Bob Clark	UTHSC at San Antonio
PI	Jiajie Zhang	UTHSC at Houston
Subject Matter Expert	Justin Starren	Northwestern

Informatics for the CTSA Program

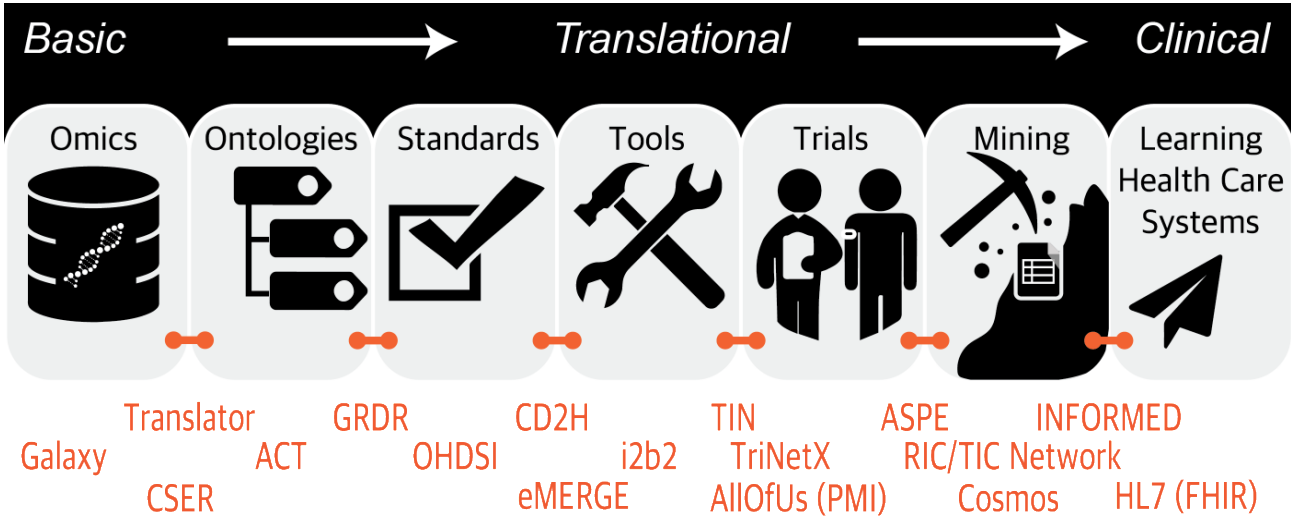
What is Informatics?

- The study and practice of creating, storing, finding, manipulating and sharing information

How does the CTSA Program Support Informatics Solutions? Through support and innovation in:

- **Data Standards:** *compatible research systems and use of standard terminologies to enable data harmonization*
- **Data Integration:** *integrating different types of data from different sources for discovery and improved health*
- **Data Access & Data Sharing:** *ability to query across sources and organizations and respond to diverse queries; enable data access, integration, and processing*
- **Data Quality:** *ensure data are fit for purpose, provide benchmarking for new tools and algorithms*
- **Data Security:** *user friendly infrastructure to assist investigators in ensuring the security of their data*

Informatics for the CTSA Program: Vision



Adapted from Ken Gersing, 2017 AMIA Summits
-figure by J. McMurry

Vision:

The CTSA Program is a collaborative and interoperable national research network that will leverage resources across multiple systems and unique expertise within our institutions to connect research to health care that results in better health through research

Informatics for the CTSA Program: Goal

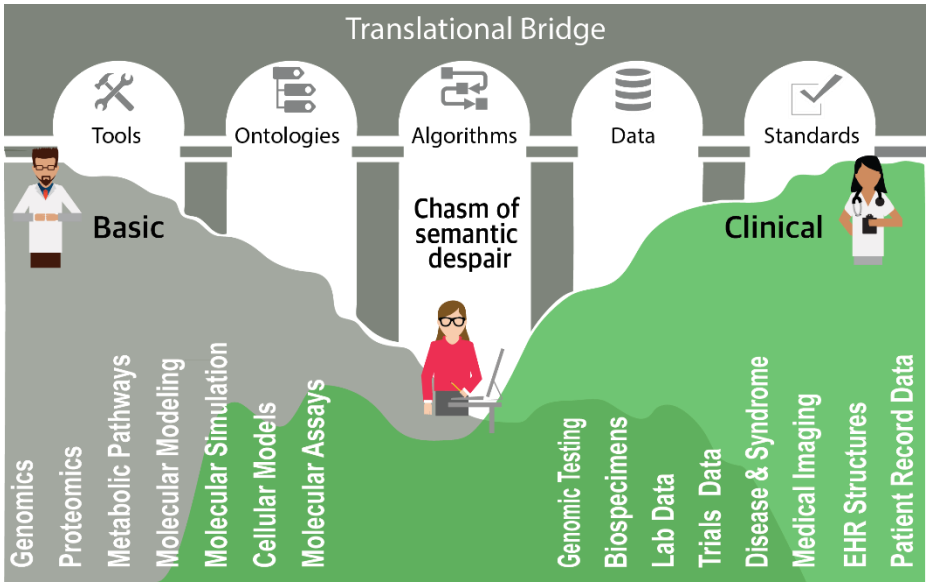
Goal:

Improve the interoperability of data within multiple systems by making the data adhere to the FAIR data principles to ultimately enable rich machine readable data:

- **Findable:** data are assigned a globally unique and eternally persistent identifier
- **Accessible:** data are retrievable by their identifier using a standardized communications protocol
- **Interoperable:** data use vocabularies that follow FAIR principles
- **Re-usable:** data have a plurality of accurate and relevant attributes

Specific Goals of the Informatics Common Metric

- Facilitate the **interoperability of research data models** through standards-based clinical data repositories.
- Facilitate **sharing** of the repository's assets for discovery.
- Encourage use of **standards-based data models**, not bound to a specific technology, but **harmonized** with other organizations, and initiatives (i.e., PCORI, ONC, FDA, and the Trial Innovation Network).
- The long-term goal is **machine-readable, interoperable data that adhere to the FAIR data principles.**



The common metric will help us understand where we can enable bridging of the chasm of semantic despair

-figure by J. McMurry

Metric: Data Repository Characterization

Operationalized Metric Title:

- Improving data access within and between CTSA Program hubs by improving clinical data repository completeness and standardization across the CTSA Program

Data Scope:

- Hubs will be asked to provide data about their local repository including the total number of unique patients and prevalence of standardized domain-specific data to describe the quantity and comparability of data in their local repository (optional to use different data models: OMOP, PCORnet, TriNetX, i2b2)

Metric: Clinical Data Repository Characterization

Metric Definition: Reporting on 8 common data domains within a clinical data repository

Repository Characterization = % of patients with a standard value in each domain

Each common data domain will include the following:

- Numerator: Count of unique patients with the standard value
- Denominator: Count of unique patients within the clinical data repository
- Metric: % of unique patients with that standard value

Examples:

- How many patients have records with laboratory tests?
  - Count of patients with a LOINC identifier/ count of unique patients within the data repository
- How many patients have records with medications?
  - Count of patients with a RxNorm identifier / count of unique patients within the data repository

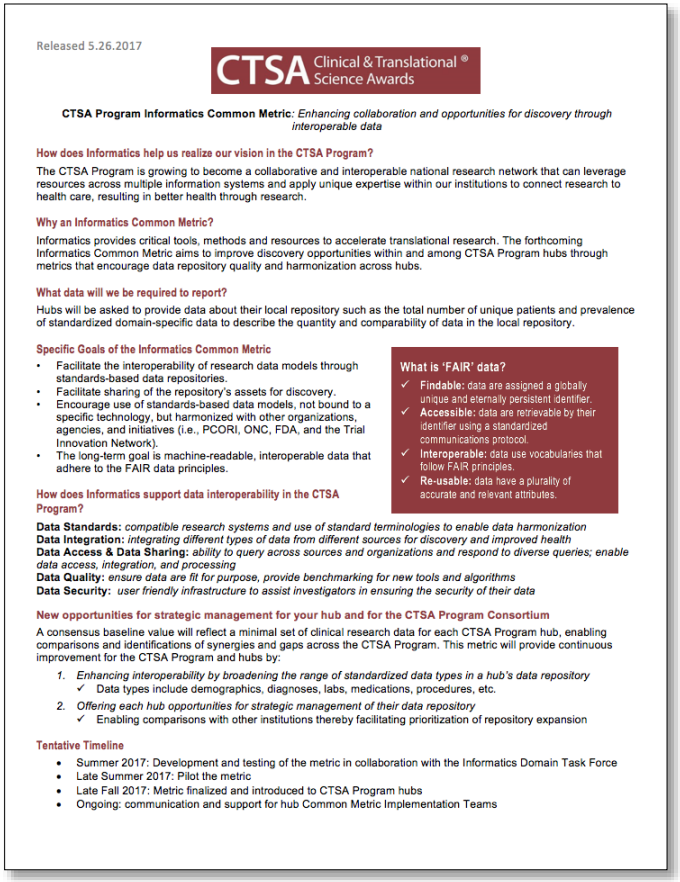
Metric: Clinical Data Repository Characterization

Data Domain	Standard Value	Numerator	Denominator	Metric
Patient	N/A	count of unique patients with a age/DOB value	Count of **all** patients in the Data Repository	% patients with age/DOB value
Patient	Administrative Gender	count of unique patients with gender value		% of total with gender value
Labs	LOINC ID	count of unique patients with a LOINC value		% of patients with LOINC value
Medications/ Drugs	RxNorm ID	count of unique patients with a RxNorm value		% of patients with RxNorm value
Conditions / Diagnosis	ICD 9/10 or SNOMED	count of unique patients with an ICD 9/10 value		% of patients with ICD 9/10 value
Procedures	ICD 9/10 CPT	count of unique patients with an ICD 9/10 or CPT procedure value		% of patients with ICD 9/10 or CPT procedure value
Notes / Narrative	N/A	count of unique patients with free text data		% of patients with free text data value
Notes / Narrative	N/A	count of unique patients with NLP		% of patients with NLP

Metric Development - Process

- ✓ Identify the baseline set of data domains with standard value set leveraging existing common data models (OMOP, PCORnet, TriNETX, i2b2) to be used to query a CTSA Program hub's data repository for completeness
- Develop or modify existing tools or scripts to support the characterization and quality assessment of a hub's data repository
  - Script/Tool is to be developed in collaboration with the iDTF
  - Data domains and standards/values will be agreed upon in collaboration with the iDTF
- Piloting of the metric will assist in testing the script/tool and the functionality
- Automation of the script/tool will lower burden for reporting of this metric at the CTSA Program hubs

A Collaborative Approach to Metric Development



Released 5.26.2017

CTSA Clinical & Translational Science Awards

CTSA Program Informatics Common Metric: *Enhancing collaboration and opportunities for discovery through interoperable data*

How does Informatics help us realize our vision in the CTSA Program?

The CTSA Program is growing to become a collaborative and interoperable national research network that can leverage resources across multiple information systems and apply unique expertise within our institutions to connect research to health care, resulting in better health through research.

Why an Informatics Common Metric?

Informatics provides critical tools, methods and resources to accelerate translational research. The forthcoming Informatics Common Metric aims to improve discovery opportunities within and among CTSA Program hubs through metrics that encourage data repository quality and harmonization across hubs.

What data will we be required to report?

Hubs will be asked to provide data about their local repository such as the total number of unique patients and prevalence of standardized domain-specific data to describe the quantity and comparability of data in the local repository.

Specific Goals of the Informatics Common Metric

- Facilitate the interoperability of research data models through standards-based data repositories.
- Facilitate sharing of the repository's assets for discovery.
- Encourage use of standards-based data models, not bound to a specific technology, but harmonized with other organizations, agencies, and initiatives (i.e., PCORI, ONC, FDA, and the Trial Innovation Network).
- The long-term goal is machine-readable, interoperable data that adhere to the FAIR data principles.

What is 'FAIR' data?

- ✓ **Findable:** data are assigned a globally unique and eternally persistent identifier.
- ✓ **Accessible:** data are retrievable by their identifier using a standardized communications protocol.
- ✓ **Interoperable:** data use vocabularies that follow FAIR principles.
- ✓ **Re-usable:** data have a plurality of accurate and relevant attributes.

How does Informatics support data interoperability in the CTSA Program?

Data Standards: *compatible research systems and use of standard terminologies to enable data harmonization*
Data Integration: *integrating different types of data from different sources for discovery and improved health*
Data Access & Data Sharing: *ability to query across sources and organizations and respond to diverse queries; enable data access, integration, and processing*
Data Quality: *ensure data are fit for purpose, provide benchmarking for new tools and algorithms*
Data Security: *user friendly infrastructure to assist investigators in ensuring the security of their data*

New opportunities for strategic management for your hub and for the CTSA Program Consortium

A consensus baseline value will reflect a minimal set of clinical research data for each CTSA Program hub, enabling comparisons and identifications of synergies and gaps across the CTSA Program. This metric will provide continuous improvement for the CTSA Program and hubs by:

1. *Enhancing interoperability by broadening the range of standardized data types in a hub's data repository*
   - ✓ Data types include demographics, diagnoses, labs, medications, procedures, etc.
2. *Offering each hub opportunities for strategic management of their data repository*
   - ✓ Enabling comparisons with other institutions thereby facilitating prioritization of repository expansion

Tentative Timeline

- Summer 2017: Development and testing of the metric in collaboration with the Informatics Domain Task Force
- Late Summer 2017: Pilot the metric
- Late Fall 2017: Metric finalized and introduced to CTSA Program hubs
- Ongoing: communication and support for hub Common Metric Implementation Teams

Metric Informational Sheet

The development of this Informatics Solution Common Metric has been highly collaborative:

- ✓ Development team represents all key stakeholders involved with a successful implementation (evaluation, informatics, administration, hub leadership)
- ✓ Stepwise process with constant engagement and feedback from the iDTF
- ✓ Pre-pilot – 2 models (OMOP & PCORnet)
- ✓ Strong engagement effort: 1-page information document, FAQs, recorded webinar, additional engagement with evaluators, open ongoing engagement with evaluators, iDTF, CTSA Program consortium, NCATS, (future: "steps for success" support materials for hub implementation teams, virtual open office hours)
- ✓ Data-driven focus

Strategic Management

At the end of the implementation of the Informatics Solution metric:

- The CTSA Program will have established a path toward consensus / standard of excellence / baseline value that reflects a minimal standard for a searchable, centralized electronic data repository at hubs within the CTSA Program
- The CTSA Program will be able to track progress toward an **interoperable national research network** as it pertains to clinical data (e.g. TIN)

Strategic Management: Opportunities

This metric will provide **continuous improvement** for **both** the CTSA Program **and** individual hubs:

- This metric will provide continuous improvement for the CTSA Program by:
  - Enhancing interoperability by increasing different types of data in a hub's clinical data repository
  - Adding new data domains with standards/values
  - Adding different types of data: imaging, genetics, etc.
- Each hub will be able to strategically manage their data warehouse to:
  - **Enhance quality** of the data within the data repository
  - **Increase interoperability** within a hub and between hubs

QUESTIONS?

NCATS