Biomedical Data Translator Funding Opportunity Announcement

Table of Contents

Participating Organization(s)	**National Institutes of Health (NIH)**
Components of Participating Organizations	**National Center for Advancing Translational Sciences (NCATS)**
Funding Opportunity Title	**Biomedical Data Translator: Technical Feasibility Assessment of Reasoning Tool (OT2)**
Activity Code	**OT2: A single-component research award that is not a grant, cooperative agreement or contract using Other Transaction Authorities**
Announcement Type	**New**
Catalog of Federal Domestic Assistance (CFDA) Number	**93.350**
Number of Applications	**Only one proposal per team**
Funding Opportunity Purpose	**As part of the Biomedical Data Translator Program, NCATS is utilizing its Other Transactions authority to invite interested applicants to submit innovative proposals for research and development of reasoning software that autonomously integrates biomedical facts, models and inferences to answer important classes of translational research problems. Successful applicants will not only contribute their expertise and resources but also must be willing to collaborate to revolutionize translational science and propel new discoveries and best practices for practitioners across the translational spectrum, from biologists to chemists to computer scientists, from scientists doing target validation to clinicians seeing patients.**
Program Background	**www.ncats.nih.gov/translator**
Objective Review	**Objective review will be conducted in two phases. In the first phase, concept letters will be reviewed and successful candidates will receive written notification with instructions for submitting a full proposal and giving a virtual presentation to the review panel. Unsuccessful applicants will not receive written feedback from the reviews.**

Concept Letter Due Date	**September 22, 2017**
Earliest Start Date	**January 2018**
Funding Instrument	**Other: An assistance mechanism that is not a grant or cooperative agreement. Other Transactions awards are subject to the requirements of the NCATS Other Transaction Award Policy Guide for this initiative. Applicants may review this policy guide available at https://ncats.nih.gov/files/NCATS-Translator-OT-Policy-Guide.pdf**
Eligibility	**All U.S. and foreign organizations and U.S. citizens are eligible to apply. This funding opportunity is open to U.S. and foreign organizations, including academic institutions and commercial entities; subcontracts are allowed. U.S. citizens may also apply as individuals without an organizational affiliation and may be direct recipients of an award. Non-citizen individuals residing in the U.S. or foreign country who are not affiliated with either a U.S. or foreign organization are not eligible to be direct recipients of an award.** **See Eligibility section of this announcement.**
Funds Available and Anticipated Number of Awards	**NCATS intends to commit $3,000,000 in FY 2018 to fund up to 3 awards.** **Future year support is not applicable.**
Award Budget	**Awards will be approximately $1,000,000 total costs. Milestones should be proposed accordingly.** **Performance will be evaluated monthly and continued support will be based on the outcome of those evaluations.**
Award Project Period	**10 months**
Funding Opportunity Expiration Date	**September 23, 2017**
	Frequently Asked Questions

Glossary of terms

FOA: This document you are reading is a funding opportunity announcement; it provides readers with specific instructions on what funding NIH is offering and how to apply for the funding.

Challenge: In order to access the complete FOA before a required concept letter is due, prospective applicants must successfully complete a series of computational tasks. Successful completion of the tasks will demonstrate the appropriate skill set for developing a reasoning tool. The Challenge is Step I of the three-step application process.

Concept letter: Step II of the application process. The concept letter is a required letter that must be submitted by the deadline in the Key Events table. This FOA document contains instructions for the content of that letter. Concept letters are submitted to the email address specified in the instructions.

Full proposal: Step III of the three-step application process. This is the document submitted by the applicant in November that requests funding for the development of a reasoning tool prototype. Based on the objective review of the concept letters, select applicants are invited to submit a full proposal. Instructions for submitting a full proposal will be provided separately.

Proof-of-concept project: This is a project that applicants are expected to present as part of their full proposal in November and will serve as preliminary data for the reasoning tool prototype research, development and testing. This project will be presented for evaluation by reviewers in a virtual meeting. Financial support is not provided for proof-of-concept development.

Prototype project: Reasoning tool prototypes will be developed & tested over a 10-month period with financial support from NCATS.

Other Transactions

The use of other transactions (which are not grants, contracts or cooperative agreements) enables NCATS to manage projects in which developments and integration of ideas and expertise from various disciplines are essential to achieve a programmatic goal. This means proposed projects and/or components of the projects submitted may be expanded, modified, partnered, not supported, or later discontinued based on program needs, emerging methods, technologies, or approaches, and availability of funds. All awardees will be expected to collaborate and cooperate with NCATS staff, one another and potentially other contributors to the overall program to maximize the exploration of the potential capabilities of Translator and to understand technical feasibility and challenges of having multiple groups build a single resource.

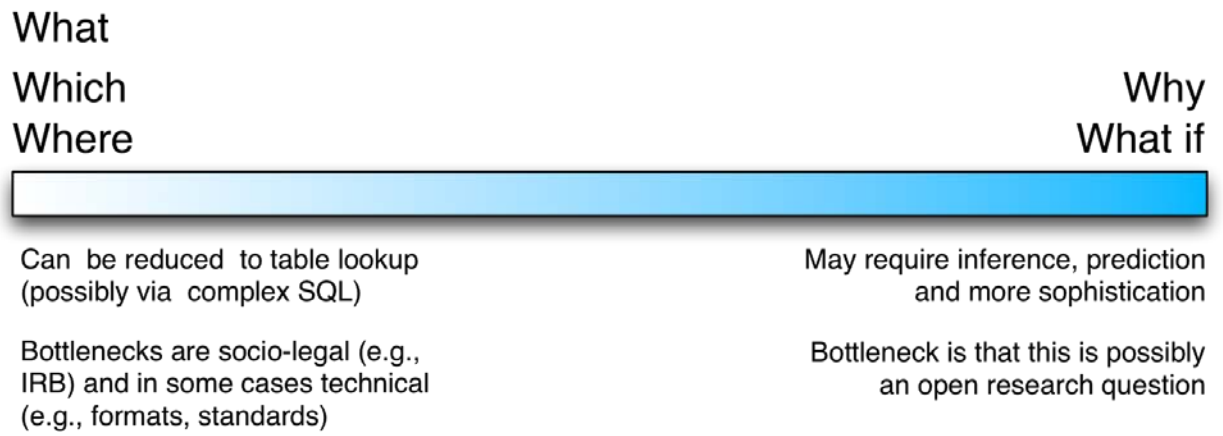
Eligibility:

This Other Transaction initiative requires successful completion of Step I, the challenge at: https://ncats.io/challenge/start. The challenge is designed to ensure that the applicants have the requisite skills to develop a reasoning tool prototype in a relatively short period of time. The

challenge is a multi-level computational exercise. Upon successful completion of each level, additional sections of the funding opportunity will be revealed. Prospective applicants must successfully pass through all levels of the challenge to gain access to the complete funding opportunity announcement. Inability to pass through all levels will result in an inability to access all of the instructions necessary to submit the requisite concept letter. All U.S. and foreign organizations and U.S. citizens are eligible to apply. This funding opportunity is open to U.S. and foreign organizations, including academic institutions and commercial entities; subcontracts are allowed. U.S. citizens may also apply as individuals without an organizational affiliation and may be direct recipients of an award. Non-citizen individuals residing in the U.S. or foreign country who are not affiliated with either a U.S. or foreign organization are not eligible to be direct recipients of an award.

Background: Data assessment and infrastructure design phase

In September 2016, NCATS issued the first five awards supporting investigators from 11 institutions for the Biomedical Data Translator (Translator). During the initial project period, Translator teams formulated a series of "translational" questions that should be answerable computationally, provided the right combination of data sources would be available. There are many types of questions possible, and we can place them on a spectrum of "engineering feasibility", schematically shown below.



Questions on the left side correspond to (possibly complex) database queries which can be achieved if the appropriate data sources are merged, or connected. Fundamentally, these queries seek to pull out appropriate subsets of data from a set of data sources.

Questions on the right hand side are somewhat open ended and seek to identify an answer that may not be explicitly listed in any data source. These types of questions may involve inference or prediction. The reasoning tool is envisioned to ultimately be able to address questions on the right hand side of this spectrum.

This exercise is helping to not only identify high value data sources that would be needed for a comprehensive Translator, but also the infrastructure needed to support the interoperability of these data sources and methods for deriving knowledge from the integrated data.

It is envisioned that the output of the Translator would not simply be a statement addressing the question posed ("What is the answer to life, the universe and everything? Answer: 42"), but rather it would deliver a dossier of information that would allow the user to easily identify sources of information/knowledge and how the sources relate to one another. Thus, it enables the user to examine the information more deeply and draw their own conclusion or concur with the Translator's conclusion when an analysis is provided.

This FOA is soliciting delivery of software, a reasoning tool, which can build dossiers by integrating biomedical facts, models and inferences that the existing Translator teams have assembled. Working with these existing components, the reasoning tool will optimize the application of the facts, models and inferences to a variety of important classes of translational research problems.

Applicants are encouraged to participate in a public meeting of the Biomedical Data Translator investigators in North Carolina, October 25, 2017 to learn about the current status of research efforts and infrastructure development.

Multi-step application process

There are three steps to the application process:

1. Successful completion of all **the challenge** tasks
2. Submission of a **concept letter** describing a short **proof-of-concept** software project that if selected, will be presented for evaluation in November
3. Successful applicants will receive written notification with instructions for submitting a full **proposal to develop a reasoning tool** and giving a virtual presentation and to **demonstrate the operation of their initial proof-of-concept** software to an objective review panel

Key Events	**Dates**	**Action needed by applicants**
Challenge opens (Step I)	September 7, 2017	Solve challenge puzzle to complete Step I of the process
Concept letters due (Step II)	September 22, 2017	Submit concept letter following instructions provided through the challenge before 11pm EDT* to complete Step II of the process
Objective review of concept letters completed	October 2, 2017	Receive written notification with instructions for submitting a full proposal and giving a virtual presentation to the review panel if the outcome of objective review is favorable
Biomedical Data Translator public meeting	October 25, 2017	Attend in person (Optional) Webex will be available
Written, full proposal for reasoning tool development that includes milestones due (Step III)	November 20, 2017	Submit by email the 10-month plan and milestones by 5pm local time* [action taken by AOR for organizations or signing representative for individuals] to initiate Step III of the process.
SAM and DUNS number submitted	November 27, 2017	**Candidates e-mail their **DUNS** number and **SAM** account information
Virtual presentations of proof-of-concept software and 10-month project plan	November 28-29, 2017	Participate in the virtual meeting with the review panel to complete Step III of the process. ***

Milestone negotiations begin	December 4, 2017	Candidates and AOR (if applicable)
Awards announced	January 2018	

*Letters and proposals received after these times will not be accepted.

**DUNS and SAM number registration can take 6 weeks or more. Candidates should begin the registration process at least 6 weeks prior to this deadline to ensure completion in time to provide these to NCATS.

***Applicants should save-the-date to ensure availability for the virtual interview.

Reasoning tool proof-of-concept

In November 2017, applicants invited for a virtual interview will present:

1. a 10-month plan to build a reasoning tool and
2. their proof-of-concept software

The proof-of-concept software will serve as preliminary data to give the review panel a sense of the approach that will be taken, the quality of the software and speed with which development is able to progress.

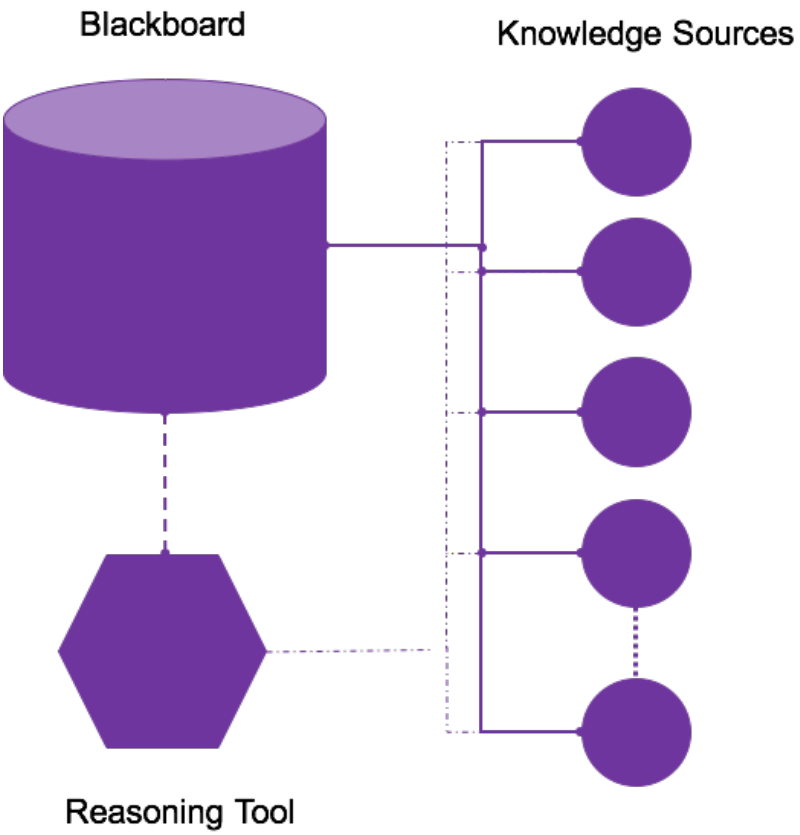
Proof-of-concept expectations

A key feature of any proposed reasoning tool is the ability to automate the process of answering a question. Currently, a question such as "*how does this drug induce a clinical outcome?"* could be answered by manually identifying the relevant data sources and data types followed by manually constructing API calls or SQL queries to retrieve the appropriate subsets of data. This would be followed by joining one or more datasets and then finally presenting the data or some computed output as the solution. Examples of such manually-constructed workflows can be found in the form of Python notebooks developed as part of the current project [link and link].

Proof-of-concept software is expected to provide reasonable and relevant answers for specific classes of research questions (see proof-of-concept questions below). Note that a specific input format for the reasoning tool is not being prescribed. While the questions, such as above, will be specified in plain English, the reasoning tool may parse it in any way that is appropriate (including manually), though automated and semi-automated approaches will be preferred. Operation of proof-of-concept software may require the assistance of the applicant to initiate its operation, but must be transparent to NIH staff. Once initiated, software is expected to complete its tasks without further operator input within a working period of twelve hours on cloud computing instances that will be provided by NCATS. The initial proof-of-concept may implement access to novel resources and algorithms, but development after November must be interoperable with existing components of the Translator (see *Further Information* below).

Prior to its demonstration in the virtual presentation, applicants are strongly encouraged to provide the objective review panel access to documentation or source code for those parts of the software responsible for substantive aspects of its operation and document sources of knowledge used. Accessibility of code will be a consideration of the evaluation (see *Collaboration, Sharing, and Intellectual Property Expectations* below).

Further information on Biomedical Data Translator



The general architecture currently employed by the Biomedical Data Translator is based on the Blackboard architecture. This architecture is comprised of three components

- Knowledge Sources (KS) - this component corresponds to specific data or information sources. This can correspond to traditional databases such as Uniprot, ChEMBL or Entrez Gene. But a KS can also be an algorithm that has well defined inputs and outputs. See https://github.com/NCATS-Tangerine/translator-knowledge-beacon/blob/develop/api/knowledge-beacon-list.yaml for a description of some KS's currently used by the Biomedical Data Translator.
- Blackboard (BB) - this component represents a space within which data queried from KS's is collated. The Biomedical Data Translator project has two BB implementations https://tkbio.ncats.io/ and https://translator.ncats.io/blackboard
- Reasoning tool - this component accepts a query, prioritizes KS's based on relevance to the query, coordinates requests to the KS's and computes on the retrieved data to produce an answer or a dataset that can be used by the user to formulate the answer. Currently, the reasoning tool function of the Translator is addressed by humans, and the actions of the reasoning tool are recorded in a Python Notebook. See here for a series of examples.

The goal of this solicitation is to support the research and development of a reasoning tool prototype to automate Translator operation as far as possible, and if not achievable, identify critical bottlenecks that must be resolved before such a reasoning tool can be implemented.

Goals for 10-month translator reasoning tool prototypes

- Demonstrate the potential utility of integrating different types of knowledge sources to address translational research questions, including biomedical facts, results from models of biology and inferences either imputed or extrapolated from existing data
- Demonstrate progress on the development of algorithms that enable answering "how?", "what if?" and "why?" classes of translational research questions autonomously through:
    - the identification of relevant knowledge sources to answer a question
    - the construction of queries to those identified knowledge sources that retrieves relevant data
    - the analysis of the retrieved data to produce a final answer or dossier of information
- Demonstrate the potential to identify gaps in existing data sets and their associated metadata and strategies to address these deficits. It is especially important to identify such gaps that are critical bottlenecks for the proposed reasoning mechanism(s)

- Define the requirements for a comprehensive Translator -- its architecture and development path, that will catalyze getting more treatments to more patients more quickly

Collaboration, sharing, and intellectual property expectations

This project is intensely collaborative amongst research partners and NIH staff, including the unrestricted exchange of source code and software tools written as part of this program. Software is published into a program-directed source code repository, facilitating its reuse by others.

NIH believes that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health. The NIH expects and supports the timely release and sharing of final research data, software and tools that support the publication of these data from NIH-supported programs for use by other researchers. The goal of this programmatic effort is to produce data, software and tools that are open source and completely publicly available for any user. The use of proprietary resources or tools will be considered if no equivalent resources are available without use restrictions.

Concept letter content (Due September 22, 2017)

All pages should be Arial 11pt, single space with 1" margins.

Concept letters are required.

Cover Page

Identify the author of the concept letter and provide their email address and phone number. Also provide the email address used to register for the challenge and associated with the completed challenge. Indicate if a full proposal (if invited) would be submitted by an individual or an organization: if from an individual, identify country of citizenship of the individual; if from an organization, identify the organization.

Summary Vision Statement (1 page)

A single page describing the core methodology to be employed and the class of research questions that it is designed to answer. In particular, it should highlight how the proposed methodology addresses limitations and bottlenecks in the current state of the art. Highlight prior expertise in the proposed methodology and previous applications (if any) that have employed it. Finally, briefly discuss the extensibility of the proposed methodology to classes of research questions not considered in the original design of the methodology.

Project plan (3 page)

The project plan is separate from the vision statement and should clearly describe methodology that will be implemented in the proof-of concept tool for November 2017. The plan is not to exceed 3 pages. Any graphs, pictures, or data tables must be included within the 3-page limit. List(s) of cited references will not count against the summary vision statement and project plan page limits.

The project plan must address the following points:

1) A description of the methodology to be used to answer translational research questions in an autonomous manner. Consider framing this description in terms of the three steps outlined in the sub-bullets under the 10-month Goals described above.

2) Anticipated functionality for a reasoning tool proof-of-concept which implements "**Proof-of-Concept Project: Questions to be Addressed**" and is to be presented in November, 2017.

3) A description of the proposed software stack. Highlight software stack components that are available under Open Source licenses versus those components that are proprietary or otherwise not freely reusable or if license status is unclear.

4) A description of how the proposed software will interact with the Translator architecture components described in this FOA or how it will be enhanced to do so in the future. See "**Further Information on Biomedical Data Translator"** section above for a listing of the current Translator components.

Personnel

Include a CV or NIH biosketch that is no more than one (1) page for each of the key personnel who have committed to participating in the award process. In the context of this program, it is

especially important to highlight contributions of personnel to existing open source projects, standards, and initiatives as well as evidence of ability to work collaboratively.

Letters of support

Letters of support should not be submitted.

Evaluation of concept letters

Letters will be evaluated by objective review of federal staff under a non-disclosure agreement. Evaluation will be based on the programmatic alignment of the vision statement with the Biomedical Data Translator, strength of the project plan, distinction of the approach from other competitive proposals and successful completion of the challenge.

The letters will be used to identify a subset of prospective applicants who will be invited to submit a full proposal in November and to present a proof-of-concept reasoning tool that addresses the following proof-of-concept questions.

Full proposal content (Due November 20, 2017)

Specific instructions for submitting the written portion of the full proposal and virtual presentation to the review panel will be provided in writing to successful candidates. Proposal components will include the completed proof-of-concept project, and a 10-month project plan for the development of a reasoning tool prototype with milestones to measure and evaluate progress. The plan and milestones will not exceed 7 pages.

Note that proof-of-concepts will be evaluated by an objective review panel based on the degree of autonomy of the software, avoiding the construction of manual workflows as much as possible to identify (1) relevant knowledge sources to answer a question, (2) retrieve relevant data from knowledge sources and (3) analyze data to produce a final answer or dossier of information to answer the questions provided below. The proof-of-concept project will be the most heavily weighted aspect of the proposal. If development for both questions cannot be completed, having partial solutions that address both classes of questions is more important than a comprehensive answer to only one class of question.

Proof-of-concept project: Questions to be addressed

Applicants will present proof-of-concept software as part of their full proposal in November. This project will be presented to reviewers in a virtual meeting and will serve as preliminary data for the reasoning tool prototype research, development, and testing. Financial support is not provided for proof-of-concept development. The proof-of-concept software should provide responses to two translational research questions. The first is listed below, and a second question is listed on the following page. The two questions will require different approaches. Therefore, if you are unable to provide complete answers to both questions, partial answers to each question will be more important than a complete answer to one.

1. Genetic condition protects from a disease

Sometimes a condition that causes adverse health effects also helps protect against other diseases. For example, the sickle cell trait offers some protection against malaria infection – heterozygous carriers of the sickle cell trait have a slight survival advantage in published studies.
https://www.cdc.gov/malaria/about/biology/sickle_cell.html
https://www.newscientist.com/article/dn20450-how-sickle-cell-carriers-fend-off-malaria/
For each of the diseases listed below, list which other genetic conditions observed in the human population might offer protection AND WHY. Please also provide a confidence for each assertion.

[List of diseases]

Proof-of-concept project: Questions to be addressed – continued

2. Clinical outcome pathways for drug-condition pairs

We do not understand how all approved drugs work, but for a majority of them we can construct a clinical outcome pathway that explains the molecular pathophysiology of their action. A clinical outcome pathway can be thought of as the efficacy equivalent of an adverse outcome pathway, it explains how a molecular initiating event precipitates a series of key events which manifest as a clinical outcome; it begins with 1) a molecular initiating event physically interacting with 2) a biological target, which affects 3) a biological pathway or series of pathways that are relevant to 4) a particular cell type or tissue that manifest as 5) a clinical phenotype or endpoint which together reflect 6) a disease or condition.

For each of the drug-condition pairs listed below, construct a clinical outcome pathway that best explains how the drug effects its action. Please also provide a confidence for each clinical outcome pathway assertion. Please report "clinical outcome pathway is not understood" in cases where the pathophysiology is not known in the literature and also provide a confidence for this assertion. For each element of a clinical outcome pathway, resolve the relevant entity to an existing dictionary ID/ontology term. Finally, can you predict which drugs used off-label might help to treat diseases without any marketed products?

[List of drug-condition pairs]