

## STATEMENT OF WORK

### Knowledge-Based Resource for Linking Animal Models to Human Disease

#### A. Background Information

Pre-clinical research advances, driven by technology breakthroughs, are changing the approach to discover methods aimed at treating human disease. Because of these advances, there is greater urgency for effective integration of the growing number of disparate data sources. Access to and use of disease model information is currently an inefficient process owing, at least in part, to the large number and complex nature of existing disease models and associated data. Research data generated from use of animal models constitute part of the decisional data package used to determine when and if a basic science discovery should be considered as a potential therapy and accelerated through the developmental process—translational science. The absence of such decisional data or, just as importantly, the use of inaccurate data contributes to a costly “slowdown” in many therapeutic development efforts. Optimizing this process for translational research is expected to enhance the path to new and effective cures.

The proper modeling of human disease requires an understanding of how conditions in nonhuman species relate to human conditions. In very few cases are the conditions produced in animals equivalent to the human condition; it is more common for animal models to present one or more features that have relevance to a particular aspect of the human disease. In some cases, this relevance to a human condition is relatively straightforward; in others, it is quite complex. And it is not surprising—but is currently relatively uncommon—to find that models data generated during studies of one disease are also relevant to understanding another seemingly unrelated disease. A system that enables the efficient capture of these cross-over bits of information would be valuable to any therapeutic development process.

Therefore, an important motivation for developing better access to resources for animal models of human disease is the enhanced ability to search across multiple animal models as well as to capture specific model information that may be relevant to multiple diseases. More rigorous methods for describing animal models will allow researchers to identify and examine commonalities and differences across multiple animal models. As more animal models are generated, it is important to develop better mechanisms by which these models are mapped onto human conditions as well as better ways to capture and then access this newly generated information. Such methods must eventually use automated ways of linking various types of representations to identify equivalent, comparable, or related concepts.

Forward movement in developing a resource to address the issues discussed above will first require a means of identifying, locating and characterizing currently used resources and access methods for all disease models, including their respective data repositories, in a type of electronic directory of animal models of human disease. This directory should be built in an extensible fashion to accommodate future growth in content and functionality.

#### B. General Description of the Required Objectives and Desired Results

The objective of this contract is to create a resource for integration and sharing of data and information about animal models. The project will consist of two components:

The initial component of this project will address the development of a directory of available disease models. The envisioned directory would enhance information access and retrieval by assisting researchers to find information about animal models. This resource directory would provide reference materials, information about the resources, and information about the animal models themselves. Initially this component would focus on a few model animal species (e.g., mouse, zebrafish), and eventually expand to other species and along other dimensions (e.g., microbes, tissues). Resource database development would focus on the needs of both the animal model and human disease research communities and the characteristics of existing resources, and address issues such as data description standards, user interface, user services, etc.

The second component of this project of the planned initiative would take a knowledge-based approach to design a portal that provides a one-point-of-contact from which all animal models and associated data can be accessed and processed.

C. Detailed Description of the Technical Requirements

Independently, and not as an agent of the Government, the Contractor shall furnish all the necessary services, qualified personnel, material, equipment and facilities, not otherwise provided by the Government as needed to perform the statement of work below.

The project shall consist of two components. The deliverable of the first component shall be an electronic directory of animal models. In the second component, the Contractor shall design a portal for integration of data, models and knowledge about animal models.

The Contractor shall perform the following tasks:

**Component 1. Create an electronic directory of animal models. Tasks shall include, but are not limited to:**

1. Identify the sources where the data and information about animal models exist. These sources may include scientific publications, white papers, websites, and may represent a variety of organizations: academic, private sector and, government, are a few examples. The NCRR will help in identifying the main contributors in the arena of animal models and provide a list of initial contacts to the contractor that may lead to other specialists in the field;
2. Create a comprehensive meta-model for collecting data about the sources, the models and the data. The meta-data shall be sufficient to define the purpose for which the model was created, its usability, validity, reliability and channels of integration with other models. The meta-model shall enable efficient searches and integration at the second component of the project;
3. Implement the meta-model in a database;
4. Populate the database with the actual data about the identified models.

NOTE TO OFFERORS: Examples of the metadata are as follows:

1. Sources:
  - a. Website, scientific journal publication
  - b. Organization: academic, private sector
2. Models:
  - a. Peer reviewed
  - b. Model validation
  - c. Available or proprietary
  - d. The task(s) for which the model was created and used
  - e. Disease
  - f. Animal species
  - g. Taxonomy / ontology / controlled vocabulary utilized
  - h. Model origin / predecessor / history
3. Data:
  - a. Availability and accessibility, e.g. publicly available or proprietary
  - b. Data acquisition process
  - c. The purpose of data collection
5. Simultaneous to the identification of sources and metadata requirements, the Contractor shall survey the market of users to:
  - a. Identify motivations for contributing/sharing data and models;

- b. Gather requirements from the potential users, e.g., what information is important to the user about animal models; what would the ideal resource look like and how would user use it; what functionality would be helpful;
  - c. Collect data about the surveyed users, i.e., whether they represent academia or the private sector, their scientific domain, computer skills, current research; and
  - d. Identify what resources users currently use
6. Develop and maintain a website with the description of the project. Main requirements include, but are not limited to:
    - a. The Website shall be designed to attract attention and be user friendly;
    - b. The Website shall address main challenges formulated by the community of users;
    - c. The Website shall concisely explain the conceptual model of the project; and
    - d. The Website's content shall remain current and compliant with Section 508 of the Rehabilitation Act (29 U.S.C. 794d)
  7. Identify relevant taxonomies, ontologies, minimal information and other standards required to include a particular model resource in the electronic directory.
  8. Identify reference sources and knowledge bases for future integration.
  9. Popularize the resource and develop tutorials and online help for bringing new users up to speed quickly, therefore lowering the technological barrier and increasing community involvement. The NCRR will collaborate closely and guide the process of popularization. Objectives and activities of the resource shall be announced in scientific journals such as Science and Nature. The initial announcement shall be approved by the NCRR Project Officer prior to placement. Any substantive changes to the initial announcement or to acquisition or distribution policies shall be approved by the Project Officer.
  10. In conjunction with NCRR, develop the policies for using the resource.
  11. Develop a model of community involvement which may include, but not be limited to:
    - a. Model publishing whereby contributors can check-in their models themselves via a user-friendly interface and uploading of models and data sets can be done via a curator;
    - b. Peer review and discussion of models and datasets that can be performed in a wiki style; and
    - c. Mechanism for continuous listening and adjusting to the new community developments and requirements.

**Component 2. Develop mechanisms for data, models and knowledge integration. Tasks shall include, but are not limited to:**

1. Design a portal for integration of data, models and knowledge about animal models. The portal shall represent one point of contact from which all models and data can be accessed and processed together. The portal shall allow for inclusion of new informatics tools, data, models, reference resources, and knowledge bases.
2. Develop a blueprint for integration of the electronic directory of models with existing knowledge structures, i.e. ontologies, taxonomies, controlled vocabularies, knowledge bases and libraries.
3. Identify potentially advantageous new methods for data and knowledge integration that shall include, but not be limited to, model merging, data and text mining, and various approaches to knowledge discovery. An inventory of these methods shall be developed.
4. Identify a 'tracking' system that shall allow recording, managing and comparing various pathways of data and knowledge processing and integration.

5. The target, first-level functionality for this web-enabled database should explain how the deliverables of the first Component facilitate movement to the full functionality implementation.
  - a. The query path shall be recorded to enable a user to go back or explore different paths.
  - b. A user shall always be able to see available dimensions relevant to the current search path and information granularity. This concept is known as zoomable multidimensional maps and it employs multiple information scales and metagraphs (e.g., see Z. Hu et al, Nature Biotechnology, v 25, n 5, 2007).
  - c. A visualization tool shall be employed to present the search paths. There are many available from which to choose. Examples of visualization tools are ([www.touchgraph.com](http://www.touchgraph.com)), THINKMAP ([www.thinkmap.com](http://www.thinkmap.com)), and Visual Thesaurus ([www.visualthesaurus.com](http://www.visualthesaurus.com)).
  - d. A user shall always be presented with a list of available keywords/variables for every chosen dimension to eliminate/reduce guesswork.
  - e. Relevant standard vocabularies, ontologies and thesauri shall be utilized to make searches efficient.

NOTE TO OFFERORS: The Contractor may choose to develop a plan for a focused but extensible prototype—using a single animal model—to demonstrate the above requirements.

**Relevant to both components:**

1. Participate as a member of the Steering Committee (SC). The SC shall be comprised of the Principal Investigator, the NCRR Project Officer, and any other personnel deemed necessary by NCRR. **For solicitation purposes, assume two (2) Contractor staff will attend four (4) one-day meetings (2 per year) to be held on the NIH campus over the course of the project. Assume costs for 10 annual conference calls of the SC.**
2. Participate in site visits to Contractor facilities by NCRR staff and SC members. Produce all source documents and be prepared to discuss them. **For solicitation purposes, assume two sites visits at contractor facilities for each of the two years of the period of performance.**
3. The Contractor shall develop a plan to assure an orderly transition at the time of contract expiration, in the event there is a follow-on contract and the Contractor is replaced by a successor, either the Government or another contractor. To affect an orderly and timely transfer of all models-related systems developed to any such successor, the Contractor shall perform the following activities during the last two months of the contract:
  - a. Cooperate fully with the successor;
  - b. Participate in meetings and/or telephone conversations with the NCRR Project Officer and the successor in which current procedures and activities are discussed in detail; and
  - c. Deliver to the successor, in accordance with the delivery schedule and shipping procedures specified by the NIH Contracting Officer, the electronic directory of animal models, any plans or blueprints developed in Components 1 and 2, and any Government furnished property. All delivery costs shall be borne by the Contractor.