Published in final edited form as:

Wiley Interdiscip Rev Syst Biol Med. 2013 March ; 5(2): 135–151. doi:10.1002/wsbm.1200.

Accelerating Cancer Systems Biology Research through Semantic Web Technology

Zhihui Wang¹, Jonathan Sagotsky², Thomas Taylor³, Patrick Shironoshita³, and Thomas S. Deisboeck^{2,§}

¹Department of Pathology, University of New Mexico, Albuquerque, NM 87131, USA

²Harvard-MIT (HST) Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA 02129, USA

³INFOTECH Soft Inc., Miami, FL 33131, USA

Abstract

Cancer systems biology is an interdisciplinary, rapidly expanding research field in which collaborations are a critical means to advance the field. Yet the prevalent database technologies often isolate data rather than making it easily accessible. The Semantic Web has the potential to help facilitate web-based collaborative cancer research by presenting data in a manner that is self-descriptive, human and machine readable, and easily sharable. We have created a semantically linked online Digital Model Repository (DMR) for storing, managing, executing, annotating, and sharing computational cancer models. Within the DMR, distributed, multidisciplinary, and inter-organizational teams can collaborate on projects, without forfeiting intellectual property. This is achieved by the introduction of a new stakeholder to the collaboration workflow, the institutional licensing officer, part of the Technology Transfer Office. Furthermore, the DMR has achieved silver level compatibility with the National Cancer Institute's caBIG®, so users can not only interact with the DMR through a web browser but also through a semantically annotated and secure web service. We also discuss the technology behind the DMR leveraging the Semantic Web, ontologies, and grid computing to provide secure inter-institutional collaboration on cancer modeling projects, online grid-based execution of shared models, and the collaboration workflow protecting researchers' intellectual property.

Keywords

Cancer modeling; model repository; oncology; Semantic Web; systems biology

SEMANTIC WEB AND COLLABORATIVE SCIENCE

In the post-genomic era, the life sciences have turned into a very data-intensive domain. Many web sites and web-based platforms with tools like wikis, blogs, and forums have been developed to aid in the collaboration of scientific research (e.g., WikiGenes ¹ and other research community developments ²). However, life sciences researchers are facing a serious challenge: *how to effectively utilize highly distributed, heterogeneous, and voluminous data sources?*

[§]Corresponding Author: Thomas S. Deisboeck, M.D., M.B.A., Complex Biosystems Modeling Laboratory, Harvard-MIT (HST) Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital-East, 2301, Bldg. 149, 13th Street, Charlestown, MA 02129, Tel: 617-724-1845, Fax: 617-726-7422, deisboec@helix.mgh.harvard.edu.

Conflict of Interest:

The authors declare that they have no competing interests.

While many groups intend to share data with their collaborators, commonly used database technologies such as relational database management systems (RDBMS) are not always conducive to the task. Relational databases are designed for static structures generally consisting of tabular data. But the table model fails in a number of cases. It does not easily represent sets of data. For example, an array of values could be stored as comma delimited text entry to be parsed once retrieved, but such approaches defeat the purpose of using a database. The popular convention for this problem is to create a separate table of values, which must be cross referenced³ (see Table 1A,B). No matter how well planned a database's schema is, it will become less manageable as layers and layers of tables and relations accumulate³. Accessing that data will require custom code and advanced knowledge of the database's peculiar structure⁴. Common problems with reverse engineering relational databases include the use of *implicit structures*, where constructs have not been explicitly declared either intentionally or due to the expressive weakness of the database management system (DBMS); *optimized structures*, where redundant and de-normalized constructs are added for time or space optimization; poor design, where novice developers can create incorrect structures or omit relevant relations; and *obsolete structures*, in which parts of a database are deprecated and are no longer in use⁵. Outside researchers looking in on the database may find it difficult to discern its unique structure to identify and extract useful data.

The **Semantic Web** offers an alternative to this. Its purpose is to provide semantic meaning across the web such that computers can make logical assertions based on coded semantics, rather than speculation based on heuristics applied to text intended for human readers⁶. This is achieved through the use of RDF (Resource Description Framework)⁷, which describes the relationship between two entities identified by URIs (Uniform Resource Identifier). Data is stored in triples – subject, verb, and object – which carry meaning with them. The relationships between entities are defined by ontologies, which can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms⁸.

Ontologies greatly aid biomedical research by providing a structured approach to capturing knowledge in a computer understandable way^{9, 10}. In a single document it is possible to apply descriptions from several ontologies to a single entity. This allows relationships and descriptions to be used in contexts beyond their original intention. As an example, an ontology describing people does not need to define an email address relationship because that relationship already exists in the Friend of a Friend (FOAF) ontology¹¹. No single ontology is ever responsible for defining everything; instead, documents can pick and choose elements from all available ontologies. Adding more data to an established database can be done without the need for building ad-hoc structures to cross-reference (see Table 1C). Aside from technical features like scalability, unordered sets, and multilingualism, the Semantic Web provides **self-descriptive** data, an obvious benefit to anyone who has ever tried to work with somebody else's data¹². The effect of this is that any outsider coming in to work with a preexisting database will find the Semantic Web database easier to navigate and understand.

Thus far, a number of Semantic Web-based open, integrated resources have been made available to the public. Some of the most successful ones include Open Biomedical Ontologies¹³, a collaborative effort to provide principles for ontology development in the biomedical domain; PathwayCommons, a single point of access to public biological systems represented in RDF; UniProt¹⁴, a comprehensive catalog of information on proteins; the Cell Cycle Ontology¹⁵, an integrated information resource for detailed analysis of cell cycle-related molecular networks; the Biological Pathway Exchange (BioPAX)¹⁶, a standard that enables integration, exchange, visualization, and analysis of biological pathway data; BioPortal¹⁷, a uniform access point to most of the biomedical ontologies

through a single user-interface and advanced tools to query over multiple, heterogeneous data resources; and workflow management systems like BioMOBY¹⁸ that take advantage of ontologies to support tasks such as sequence analysis and genome annotation; and lastly, the CardioVascular Research Grid (CVRG), a grid service that makes a wide range of open-source informatics and data analysis tools available to support the cardiovascular research community in collaborative research projects¹⁹. Furthermore, increasing efforts are being made to provide analytic tools to explore dispersed Semantic Web information resources. For example, RDFScope provides an interactive application to browse and query Semantic Web resources in an intuitive way and visualize and analyze biological networks²⁰.

INTEGRATIVE CANCER SYSTEMS BIOLOGY

The life sciences are now being supplemented by Semantic Web technologies for a variety of scientific purposes^{21–23}. As a newly emerging field in the life sciences, *cancer systems biology* uses an interdisciplinary approach to provide insight into the systemic understanding of cancer initiation, progression, and metastasis by investigating how individual components interact to give rise to the emergent, dynamic behavior of the cancerous system as a whole²⁴. This field needs to bring together researchers of diverse expertise to identify, articulate, and structure problems of interest, and involves intense interdisciplinary collaboration and resource (data and model) sharing²⁵. In this regard, the Semantic Web can be considered as an ideal platform for representing and linking the data and models produced in this diverse and interdisciplinary field.

A systems approach often involves mathematical and computational modeling, in addition to conventional laboratory-based experiments, to generate testable predictions²⁶. Such *in silico* methods are especially useful when it is not possible or practical to conduct biological experiments²⁷. In the past decade, many cancer models focusing on either a specific biological scale or spanning multiple scales in time and space have been developed (see^{28–34} for reviews on current cancer models and specific modeling methods). Particularly, modeling has emerged as a solution to shed light on the understanding of complex cancer diseases at a systems level^{35–39}, the development of novel diagnostic or therapeutic applications^{40–43}, and the identification of drug targets in cancer^{44–47}.

While undoubtedly progress has been made in cancer modeling, this sprawling field also led to some redundant work while some areas have drawn insufficient attention. Research activities sometimes overlap between different groups, and researchers do not want to build their models from scratch⁴⁸. Thus, the cancer modeling community, or even the cancer systems biology community at large, will benefit from an open, publicly accessible online system to efficiently share data, reuse modeling tools, and exchange information and knowledge within the context of distributed, multidisciplinary, and inter-organizational collaborative teams, while at the same time sufficiently protecting data ownership. In the following, we present a dynamic online interface for cancer modeling, the Digital Model Repository (DMR), based on Semantic Web technology. The DMR specifically seeks to offer a semantically linked platform for facilitating multidisciplinary and inter-organizational collaboration on cancer systems biology projects.

DEVELOPMENT OF DIGITAL MODEL REPOSITORY (DMR)

DMR Overall Architecture

The DMR is a Semantic Web platform for the exchange and collaboration of cancer research models (see Figure 1 for a schematic of the overall architecture). The DMR has been developed and is currently administrated by the Center for the Development of a Virtual Tumor (CViT)⁴⁹ at Massachusetts General Hospital; CViT provides an online community

space for discussions between cancer researchers at a global scale (a total of 335 registered investigators from 177 research institutions across 37 countries). The DMR's architecture enables researchers to share data and collaborate on projects. Users can upload their computational cancer models, associated data files, and corresponding results and share them with other DMR users. Each user is bound to an academic or non-academic institution, which retains legal ownership of all uploaded files associated with his/her account. As a special feature of the DMR, institutions' technology transfer offices and, as their executors, the 'licensing officers' (LOs) are responsible for approving user access to the repository as well as approving all publishing (i.e., sharing) requests within a given DMR account.

DMR Technology

The DMR is built using a number of Semantic Web technologies. In brief, it uses the FOAF ontology¹¹ for annotating data about its users; Dublin Core (DC)⁵⁰ for describing its users, organizations, and documents; and Web Ontology Language (OWL)³⁰ for entity classification. The CViT ontology interlinks terms from FOAF, DC, and OWL, supplemented by CViT terms to define the high-level entities managed by the DMR. The CViT ontology also contains a hierarchical classification of computational cancer model features to both help users annotate their models and search for relevant models. FOAF, DC, OWL, and CViT ontologies together provide the language that defines the DMR. In addition, as part of caBIG™ Silver-Level Compliance, DMR concepts are linked to terms from the NCI Thesaurus. Table 2 highlights how these ontologies are combined and linked with the NCI Thesaurus to construct the data structures of the DMR, while Table 3 describes the cancer model classification terms in the CViT ontology. Figure 2 shows the complete UML model of the DMR. The DMR uses OpenAnzo⁵² as its Semantic Web data store backed by the DB2 Express relational database. SPARQL (an RDF Query Language)⁵³ is used to query data in the DMR. The reference CViT ontology and data structures are stored in CVS version control system, and as concepts are added to the ontology and the DMR platform evolves, both the reference ontology in CVS and working ontology in OpenAnzo are simultaneously updated. What follows is a description of the **workflow** used by DMR members to collaborate.

Upon logging in to the DMR, a principal investigator (PI) will have the option to create a new entry. Entries are the basic container unit for data in the repository and are designed to house a single project. Each entry must be given a title and classification: *in vivo*, *in vitro*, *in silico*, or clinical (for clinical data, see Part A, paragraph 1b of the CViT license (ver. 2.1), a BSD-style open source license⁵⁴). Entries also have additional metadata sections for description, abstract, concept, hypothesis, conclusion, classification, and notes. These fields are optional and can be used or omitted at the discretion of the user, depending on their relevance to the project.

Data files can be uploaded to the entry (see Figure 3A for an example), which must be classified as an algorithm, source code, computational model, parameters, image, movie, or experimental data. References can also be attached, classified as papers, reviews, and books. Data and references can either be uploaded directly from a user's hard drive or linked by URL.

To utilize the capabilities of RDF to interlink content, files and references can also be selected from neighboring entries in the DMR or even other web-accessible sources across the internet rather than being uploaded from a user's computer. For example, a PI who is creating an entry can select files from another entry to which he has read access. Because those files are owned by the creator of the original entry, they are not necessarily readable by anyone browsing the new entry. A user would have to have read privileges for both entries in order to see the files that are linked to another entry. This feature is important

because it indicates to the system that two separate and distinct entries share a common background, thus defining a relation between the two. As the DMR matures these links will grow and a 'web of knowledge' will develop.

Data Ownership

Protecting data ownership has not been paid enough attention in data sharing applications in life sciences where it is obviously as important for sustaining cross-institutional collaborations for any technical developments. The DMR introduces an innovative strategy to ensure that any institution that deposits data retains ownership of that data, by introducing a new stakeholder, i.e., the License Officer (LO, see above), to the data publishing and sharing architecture. Whenever a user interacts with the CViT license, the LO must approve that interaction. Currently this happens in the following two cases. *Case 1:* When a user first logs on to the DMR he is presented with a copy of the CViT open-source license to read. If he accepts, he is not immediately granted access; instead, the system sends a notification to the user's LO who must approve the user's acceptance of the license. This means that each institution registered in the DMR must have a LO in order for users to log in to the DMR. *Case 2:* Licensing officers are also part of the entry publishing process. When a PI wants to share his data entry with members of another institution, he can publish it to them to grant them read access. The PI will once again be presented with the CViT license. If accepted, the LO will be sent a message to approve the license. Once that approval is given, the entry is published to the CViT members the PI originally selected (Figure 3B). Approval records are visible in the Licensing History section of each published entry in the DMR, thereby employing RDF-conveyed provenance to ensure transparency of IP ownership throughout the sharing process, essential in an ever growing collaborative environment.

In addition to read access, a PI can give out write access to his entry by choosing contributors. Currently, a contributor must belong to the same institution as the PI. Selecting a contributor, such as post-doctoral fellows or graduate students that work in the PI's lab, does not invoke the LO workflow. Contributor privileges allow a user to write annotations and upload files to an existing entry (Figure 3C). The only privilege that is unique to an entry's owner is the ability to publish the entry and select additional contributors.

Online Computational Model Execution

A recent feature of the DMR is the Computational Model Execution Framework (CMEF). The CMEF has been developed as an extension to the DMR to enable the online, grid-based execution of the computational cancer models deposited within the DMR. Utilizing CMEF, members of the CViT community can select and configure models to be executed (Figure 4A), determine the data to be used to run these models (Figure 4B), and review simulation results (Figure 4C). The model shown in Fig. 4 is a two-dimensional multiscale mathematical model simulating lung cancer growth and invasion⁵⁵. This model spans molecular and multi-cellular scales, and can quantify the relationships between extracellular stimuli, intracellular signaling dynamics, and multicellular tumor growth and expansion. The reader should be able to get a brief idea of how a cancer model is launched through the CMEF. Note that the models stored in the DMR can be multiscale or scale-specific.

When a model has been added to the entry, anyone with appropriate permission can execute the model. Execution is scheduled on one of CViT's CMEF nodes, selected automatically by the CMEF based on runtime requirements (operating system, hardware architecture, and installed libraries) and node availability. When execution is complete, CMEF notifies the user who scheduled the run by email and the results are linked back to the original entry as downloadable files. The results of that run remain in the entry for future reference. Users can execute a model online on the CMEF grid, or can download the model and run it locally,

provided the user has the requisite computational infrastructure to execute the model. The CMEF server (see Figure 5 for its architecture and Figure 6 for its extensions to the DMR domain model) was specifically designed to support the computational cancer models in the DMR and integrates seamlessly into the CViT.org website providing grid-based model execution of Java, compiled executables, and R programs on 32- and 64-bit Windows and Linux nodes. Table 4 lists the major entities that have been added to the DMR domain to support the CMEF.

caBIG® Silver-level Compliant Data Service

The DMR infrastructure is supported beyond CViT as well. The cancer Biomedical Informatics Grid (caBIG®) is an NIH/NCI funded project that connects researchers, physicians, and patients in the cancer research community⁵⁶. caGrid provides the underlying infrastructure for accessing and integrating data and analytical tools deployed at different institutions within the caBIG® environment⁵⁷). In order to be enabled as a part of the grid, a service must comply with caGrid standards; this includes rigorous testing and mapping of API elements to the NCI Thesaurus. Over the course of several months the DMR team worked with caBIG® staff to curate the DMR domain model and with the Vocabularies and Common Data Elements (VCDE) workgroup to link DMR concepts with equivalent terms in the NCI Thesaurus; Table 2 includes the links between DMR entities and the NCI Thesaurus. The DMR has achieved silver level compatibility with caBIG®. This means that other services can be used to connect to the DMR to use its data. All DMR licensing restrictions are honored by the service, so users connecting through caGrid will be required to use their existing DMR account to access data. This data service allows CViT users to take any DMR models they can read and plug them directly into caGrid for processing, without ever logging in to the DMR website (see Table 5 for current DMR service functions).

DMR Usage

We have attempted to make data discovery in the DMR as easy as possible to locate, access, and review data that other scientists have published. Users are notified by email when they first receive read access to an entry. The DMR provides an online search feature for finding readable entries. Entries are also categorized into a topic tree, which appears in the navigation bar. There also exist options (e.g., the CViT mashup and DMR graph, described below) outside of the DMR itself that list all entries, not just currently readable ones.

Several tools have been developed to pull public data from the DMR and assemble it into additional resources for the previously mentioned CViT community. The CViT **mashup** is a display of the DMR across Google Maps (Figure 7A). It uses Javascript to fetch all institutions, their coordinates, and users from the DMR and places those on a geographic map. As each DMR user is bound to an institution, preparing the DMR for Google Maps involved finding latitude and longitude coordinates for each of those institutions. Each institution is given a map marker and placed accordingly on the map. When that marker is clicked, a message box appears listing all members of that institution. Markers are color coded according to the type of research their members take part in (experimental, computational, or both). The mashup also provides modes for displaying DMR projects or entries owned by an institution, instead of individual members. For example, the map can be changed to focus on a single entry, showing only those institutions and users that have access to that entry.

CViT's DMR **graph** (Figure 7B) graphically links users, entries, and institutions, dynamically illustrating the up-to-date Semantic Web of the repository. At the root of the graph is the DMR itself. Stemming off of the DMR are institutions. Surrounding the

institutions are its users and entries. Clicking once on any node brings up its info window, which displays its title and description as well as any relevant links. Clicking that node a second time refocuses the graph around the node. Depending on the context of the newly focused node, more information will be presented in the graph. Focusing on an entry shows color-coded links to users with access to the entry. Focusing on a user shows their access levels to other entries. These new nodes can of course be followed to traverse the entirety of the DMR. As the DMR grows, entries will begin to link to each other as references. As this happens, what will emerge is an illustration of the growth of knowledge within the DMR.

Another new feature for the DMR is **Analytics**, which can be used to measure user interest and entry presence. DMR Analytics also features CollaboRank (Figure 8), which is CViT's attempt to encourage data sharing by ranking how much sharing is being done by each institution. The sharing metric is simple: Analytics counts how many people have been given read access to each entry in an institution. Those values are summed to determine the institution's sharing score. Institutions are sorted and displayed according to this score, along with a breakdown of which entries they own and to whom those entries have been published. While already interesting, this metric is too simplistic at the moment; ultimately, quality needs to be factored in to CollaboRank and this can happen with user rated entries and/or with weighing impact factors from papers linked to entries.

DISCUSSION AND FUTURE DIRECTIONS

Systems biology typically involves the integration of data and models across disciplinary boundaries in order to solve complex behavior of biological systems^{58, 59}. The Semantic Web approach is precisely aiming at enhancing information exchange and integration by providing standardized formats, such as RDF, RDFS and OWL, to achieve a formalized computational environment²¹. In fact, the importance of properly semantically annotated data and services indeed has been recognized by the systems biology community⁶⁰. However, regardless of how the Semantic Web has revolutionized the way the new data and the corresponding models can be stored, vast amounts of existing scientific data are still present in an unstructured form; some of them are even provided as plain text, which can only be easily understood and interpreted by humans. Therefore, the process of manual curation of literature by human experts is indispensable, at least at the current stage. An ever growing number of applications aiding large-scale manual curation in the biomedical research have been developed. Representative examples are UniProt¹⁴, Reactome⁶¹, and Gene Ontology Annotation⁶². Especially, BioModels.net collection, built on the well-known BioModels Database⁶³ (a free platform for storing, searching, browsing and retrieving published and curated quantitative kinetic models of biochemical and cellular systems), is also working on providing Semantic Web Services to ease the integration and composition of web services that it provides⁶⁴. However, one has to keep in mind that, while the knowledge extracted through curation is typically of high quality, these efforts do not scale up to the current needs of life sciences due to the exponential growth in the amount of biological and clinical data⁶⁵. DMR's own team does not have the ability to curate all the entries they receive, so the DMR instead is self-curated by its users and the amount of usage of these data will be an expression of the data quality at entry.

We envision that the existence of the DMR will contribute to an improvement in the quality of the computational cancer modeling by establishing a model evaluation process. In other words, quality control is seen as important as curation for the DMR. As mentioned in the Analytics/CollaboRank discussion, in future releases users will need to have the ability to grade or rate any entry they read, not unlike Amazon or eBay. This rating will be factored into CollaboRank's scoring system. It will also be useful for other users trying to find interesting or worthwhile data. Furthermore, having the models and data available in

standardized formats with clearly stated dependencies will improve the utility of models and facilitate the creation of workflows that can generate model results to compare model predictions with experimental data, all in an automated fashion. This task can be facilitated by collaborating with other groups developing standard formats for model exchange at different biological scales, such as the Systems Biology Markup Language (SBML⁶⁶), CellML⁶⁷, BioPAX¹⁶, and FieldML⁶⁸.

Most important of all is the establishment of a large user base. The more users that publish data, the more attractive the DMR becomes, thus encouraging more users to join – a classic network externalities phenomenon. The problem of growth in a web community, especially in an expert community (e.g., the cancer modeling community) where experts are a limited resource, is probably sociological rather than technical⁶⁹. From a technical perspective, we have taken every measure to ensure that DMR is open and accessible to new users. It is based on Semantic Web technology, which benefits collaborators and researchers trying to work with it. On the other hand, computational biologists and cancer modelers need to understand the importance of IT infrastructure systems like the DMR that allow them to create semantic data now, so, in comparison with traditional (or even open access) publishing methods, their data and models can be published and then shared with others more efficiently and securely, according to their intentions and within their controls. For example, through the DMR, a model can be accessed and directly executed online or may be validated against other types of experimental data published by other groups.

What does the future hold for the DMR? One of the next steps is the logical combination of the caGrid data service and CMEF. The data service allows repository data to be plugged into analytical services on caGrid. CMEF allows users to plug data into models stored on the DMR. Combining these two features would allow caGrid data to be interfaced with DMR models. We note here that, although the NCI has recently decided to replace the caBIG with a new program, National Cancer Informatics Program (NCIP), the founding principles of the caBIG will remain. Specifically, the NCIP will continue to support (1) the development of community-driven standards for data exchange and application interoperability and (2) semantic infrastructures that allow data to be integrated across cancer centers. The development of the DMR is along the lines of those efforts. Our point here is that, if the DMR is built upon and inter-linked with a bigger online community (whether it is caGrid, NCIP, or the NSF Science Gateways⁷⁰), the DMR can function as an analytical web service for the community instead of a source for data only.

The experience with CViT and the DMR also is being applied to a European Commission-supported, EU-US collaborative project, Transatlantic Tumour Model Repositories (TUMOR⁷¹), aimed at developing a European clinically oriented semantic-layered cancer digital model repository from existing EU projects that will be interoperable with the US grid-enabled semantic-layered digital model repository platform at CViT.org. This interoperable, DMR interfaced environment will offer a range of services to international cancer modelers, biomedical researchers, and eventually clinicians, aimed at supporting both basic cancer quantitative research and individualized optimization of cancer treatment. This ‘transatlantic’ project will therefore be the starting point for an international validation environment, which will support joint applications, verification and validation of the clinical relevance of cancer models.

In summary, the Semantic Web is reaching new areas in life sciences, such as cancer systems biology. It provides a set of standards and technologies to semantically characterize and link heterogeneous and distributed data and models. As a pioneering development in the cancer modeling field, the DMR offers a semantically linked online data store for cancer systems biology researchers to share their models. Researchers not only own the models that

they deposited on the DMR, but are able to grant a variety of access rights to their collaborators upon gaining their institutions' LO's permission. Furthermore, the CMEF allows researchers to execute cancer models online without having to prepare custom computing resources.

In its current version, the DMR only allows a model to be executed as an independent computer application, but in the future, it is planned that the DMR will enable seamless integration of different computational modules based on the use of various accepted ontologies and semantically annotated objects/parameters exchanged between applications. As a result, different research groups may independently develop a part of a big project as a plug-in, and the final computer program is dynamically 'created' online at run-time by integrating these distinct modules. While this feature requires an additional middle-layer application at the DMR to process the ontologies and to reveal the possible interactions, it will facilitate the process of cross-model integration thus reduce redundancy and harness collaborative expertise, both critical prerequisites to accelerate computational cancer system biology.

Acknowledgments

This work has been supported in part by NIH grant CA 113004 and by the Harvard-MIT (HST) Athinoula A. Martinos Center for Biomedical Imaging and the Department of Radiology at Massachusetts General Hospital. Both CViT.org and DMR/CMEF are currently hosted at INFOTECH SOFT Inc.

References

- Hoffmann R. A wiki for the life sciences where authorship matters. *Nat Genet.* 2008; 40:1047–1051. [PubMed: 18728691]
- Sagotsky JA, Zhang L, Wang Z, Martin S, Deisboeck TS. Life Sciences and the web: a new era for collaboration. *Mol Syst Biol.* 2008; 4:201. [PubMed: 18594519]
- James, J. It's Time to Consider Alternatives to RDBMS. Available at: <http://blogs.techrepublic.com.com/programming-and-development/?p=1563>
- Borysowich, C. Some Pros & Cons of Relational Databases. Available at: <http://it.toolbox.com/blogs/enterprise-solutions/some-pros-cons-of-relational-databases-24144>
- Chiang R, Barron T, Storey V. A framework for the design and evaluation of reverse engineering methods for relational databases. *Data & Knowledge Engineering.* 1997; 21:57–77.
- Berners-Lee T, Hall W, Hendler J, Shadbolt N, Weitzner DJ. Computer science. Creating a science of the Web. *Science.* 2006; 313:769–771. [PubMed: 16902115]
- Manola, F.; Miller, E. RDF Primer. Available at: <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>
- Shawver LK, Slamon D, Ullrich A. Smart drugs: tyrosine kinase inhibitors in cancer therapy. *Cancer Cell.* 2002; 1:117–123. [PubMed: 12086869]
- Bodenreider O, Stevens R. Bio-ontologies: current trends and future directions. *Brief Bioinform.* 2006; 7:256–274. [PubMed: 16899495]
- Ruttenberg A, Clark T, Bug W, Samwald M, Bodenreider O, Chen H, Doherty D, Forsberg K, Gao Y, Kashyap V, et al. Advancing translational research with the Semantic Web. *BMC Bioinformatics.* 2007; 8 (Suppl 3):S2. [PubMed: 17493285]
- FOAF-project. Introducing FOAF. Available at: <http://www.foaf-project.org/original-intro>
- Bergen, Pv. Semantic web marvels in a relational database. Available at: <http://techblog.procurios.nl/k/news/view/34300/14863/Semantic-web-marvels-in-a-relational-database---part-I-Case-Study.html>
- Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol.* 2007; 25:1251–1255. [PubMed: 17989687]

14. Patient S, Wieser D, Kleen M, Kretschmann E, Jesus Martin M, Apweiler R. UniProtJAPI: a remote API for accessing UniProt data. *Bioinformatics*. 2008; 24:1321–1322. [PubMed: 18390879]
15. Antezana E, Egana M, Blonde W, Illarramendi A, Bilbao I, De Baets B, Stevens R, Mironov V, Kuiper M. The Cell Cycle Ontology: an application ontology for the representation and integrated analysis of the cell cycle process. *Genome Biol*. 2009; 10:R58. [PubMed: 19480664]
16. Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, Wu G, D'Eustachio P, Schaefer C, Luciano J, et al. The BioPAX community standard for pathway data sharing. *Nat Biotechnol*. 2010; 28:935–942. [PubMed: 20829833]
17. Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T, Musen MA. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res*. 2011; 39:W541–545. [PubMed: 21672956]
18. Kerhornou A, Guigo R. BioMoby web services to support clustering of co-regulated genes based on similarity of promoter configurations. *Bioinformatics*. 2007; 23:1831–1833. [PubMed: 17496321]
19. The CardioVascular Research Grid. Available at: <http://cvrgrid.org/>
20. Splendiani A. RDFScape: Semantic Web meets systems biology. *BMC Bioinformatics*. 2008; 9 (Suppl 4):S6. [PubMed: 18460179]
21. Post LJ, Roos M, Marshall MS, van Driel R, Breit TM. A semantic web approach applied to integrative bioinformatics experimentation: a biological use case with genomics data. *Bioinformatics*. 2007; 23:3080–3087. [PubMed: 17881406]
22. Sioutos N, de Coronado S, Haber MW, Hartel FW, Shaiu WL, Wright LW. NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J Biomed Inform*. 2007; 40:30–43. [PubMed: 16697710]
23. Luciano JS, Andersson B, Batchelor C, Bodenreider O, Clark T, Denney CK, Domarew C, Gambet T, Harland L, Jentzsch A, et al. The Translational Medicine Ontology and Knowledge Base: driving personalized medicine by bridging the gap between bench and bedside. *J Biomed Semantics*. 2011; 2 (Suppl 2):S1. [PubMed: 21624155]
24. Deisboeck TS, Berens ME, Kansal AR, Torquato S, Stemmer-Rachamimov AO, Chiocia EA. Pattern of self-organization in tumour systems: complex growth dynamics in a novel brain tumour spheroid model. *Cell Prolif*. 2001; 34:115–134. [PubMed: 11348426]
25. Kreeger PK, Lauffenburger DA. Cancer systems biology: a network modeling perspective. *Carcinogenesis*. 2010; 31:2–8. [PubMed: 19861649]
26. Popel AS, Hunter PJ. Systems Biology and Physiome Projects. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*. 2009; 1:153–158. [PubMed: 20835988]
27. van Riel NA. Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Brief Bioinform*. 2006; 7:364–374. [PubMed: 17107967]
28. Deisboeck TS, Wang Z, Macklin P, Cristini V. Multiscale cancer modeling. *Annu Rev Biomed Eng*. 2011; 13:127–155. [PubMed: 21529163]
29. Lowengrub JS, Frieboes HB, Jin F, Chuang YL, Li X, Macklin P, Wise SM, Cristini V. Nonlinear modelling of cancer: bridging the gap between cells and tumours. *Nonlinearity*. 2010; 23:R1–R9. [PubMed: 20808719]
30. Wang Z, Deisboeck TS. Computational modeling of brain tumors: Discrete, continuum or hybrid? *Scientific Modeling and Simulation*. 2008; 15:381–393.
31. Frieboes HB, Chaplain MA, Thompson AM, Bearer EL, Lowengrub JS, Cristini V. Physical oncology: a bench-to-bedside quantitative and predictive approach. *Cancer Res*. 2011; 71:298–302. [PubMed: 21224346]
32. Gatenby RA, Gillies RJ. A microenvironmental model of carcinogenesis. *Nat Rev Cancer*. 2008; 8:56–61. [PubMed: 18059462]
33. Tracqui P. Biophysical models of tumour growth. *Reports on Progress in Physics*. 2009; 72.
34. Edelman LB, Eddy JA, Price ND. In silico models of cancer. *Wiley Interdiscip Rev Syst Biol Med*. 2010; 2:438–459. [PubMed: 20836040]

35. Bauer-Mehren A, Furlong LI, Sanz F. Pathway databases and tools for their exploitation: benefits, current limitations and challenges. *Mol Syst Biol.* 2009; 5:290. [PubMed: 19638971]
36. Rejniak KA, Anderson AR. Hybrid models of tumor growth. *Wiley Interdiscip Rev Syst Biol Med.* 2011; 3:115–125. [PubMed: 21064037]
37. Schnell S, Grima R, Maini PK. Multiscale modeling in biology - New insights into cancer illustrate how mathematical tools are enhancing the understanding of life from the smallest scale to the grandest. *American Scientist.* 2007; 95:134–142.
38. Wang Z, Birch CM, Sagotsky J, Deisboeck TS. Cross-scale, cross-pathway evaluation using an agent-based non-small cell lung cancer model. *Bioinformatics.* 2009; 25:2389–2396. [PubMed: 19578172]
39. Martin NK, Gaffney EA, Gatenby RA, Maini PK. Tumour-stromal interactions in acid-mediated invasion: a mathematical model. *J Theor Biol.* 2010; 267:461–470. [PubMed: 20816684]
40. Araujo RP, Liotta LA, Petricoin EF. Proteins, drug targets and the mechanisms they control: the simple truth about complex networks. *Nat Rev Drug Discov.* 2007; 6:871–880. [PubMed: 17932492]
41. de Pillis LG, Radunskaya AE, Wiseman CL. A validated mathematical model of cell-mediated immune response to tumor growth. *Cancer Res.* 2005; 65:7950–7958. [PubMed: 16140967]
42. Haeno H, Gonen M, Davis MB, Herman JM, Iacobuzio-Donahue CA, Michor F. Computational modeling of pancreatic cancer reveals kinetics of metastasis suggesting optimum treatment strategies. *Cell.* 2012; 148:362–375. [PubMed: 22265421]
43. Basanta D, Gatenby RA, Anderson AR. Exploiting evolution to treat drug resistance: combination therapy and the double bind. *Mol Pharm.* 2012; 9:914–921. [PubMed: 22369188]
44. Feala JD, Cortes J, Duxbury PM, Piermarocchi C, McCulloch AD, Paternostro G. Systems approaches and algorithms for discovery of combinatorial therapies. *Wiley Interdiscip Rev Syst Biol Med.* 2010; 2:181–193. [PubMed: 20836021]
45. Wang Z, Bordas V, Deisboeck TS. Discovering Molecular Targets in Cancer with Multiscale Modeling. *Drug Dev Res.* 2011; 72:45–52. [PubMed: 21572568]
46. Wang Z, Bordas V, Sagotsky J, Deisboeck TS. Identifying therapeutic targets in a combined EGFR-TGFbetaR signalling cascade using a multiscale agent-based cancer model. *Math Med Biol.* 2012; 29:95–108. [PubMed: 21147846]
47. Kumar N, Hendriks BS, Janes KA, de Graaf D, Lauffenburger DA. Applying computational modeling to drug discovery and development. *Drug Discov Today.* 2006; 11:806–811. [PubMed: 16935748]
48. Henkel R, Endler L, Peters A, Le Novere N, Waltemath D. Ranked retrieval of Computational Biology models. *BMC Bioinformatics.* 2010; 11:423. [PubMed: 20701772]
49. Deisboeck TS, Zhang L, Martin S. Advancing cancer systems biology: introducing the Center for the Development of a Virtual Tumor, CViT. *Cancer Inform.* 2007; 5:1–8. [PubMed: 19390664]
50. DCMI. Metadata Basics. Available at: <http://dublincore.org/metadata-basics/>
51. NCI Thesaurus. Available at: <http://ncit.nci.nih.gov/ncitbrowser/>
52. DuCharme, R. Getting Started with Open Anzo. Available at : <http://www.snee.com/bobdc.blog/2009/03/getting-started-with-open-anzo.html>
53. Prud'hommeaux, E.; Seaborne, A. SPARQL Query Language for RDF. Available at: <http://www.w3.org/TR/rdf-sparql-query/>
54. CViT. CViT License. Available at: <https://www.cvit.org/license.pdf>
55. Wang Z, Zhang L, Sagotsky J, Deisboeck TS. Simulating non-small cell lung cancer with a multiscale agent-based model. *Theor Biol Med Model.* 2007; 4:50. [PubMed: 18154660]
56. Saltz J, Oster S, Hastings S, Langella S, Kurc T, Sanchez W, Kher M, Manisundaram A, Shanbhag K, Covitz P. caGrid: design and implementation of the core architecture of the cancer biomedical informatics grid. *Bioinformatics.* 2006; 22:1910–1916. [PubMed: 16766552]
57. Saltz J, Kurc T, Hastings S, Langella S, Oster S, Ervin D, Sharma A, Pan T, Gurcan M, Permar J, et al. e-Science, caGrid, and Translational Biomedical Research. *Computer (Long Beach Calif).* 2008; 41:58–66. [PubMed: 21311723]

58. Hoehndorf R, Dumontier M, Gennari JH, Wimalaratne S, de Bono B, Cook DL, Gkoutos GV. Integrating systems biology models and biomedical ontologies. *BMC Syst Biol.* 2011; 5:124. [PubMed: 21835028]
59. Chen H, Yu T, Chen JY. Semantic Web meets Integrative Biology: a survey. *Brief Bioinform.* 2012
60. Lamprecht AL, Naujokat S, Margaria T, Steffen B. Semantics-based composition of EMBOSS services. *J Biomed Semantics.* 2011; 2 (Suppl 1):S5. [PubMed: 21388574]
61. Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, et al. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* 2009; 37:D619–622. [PubMed: 18981052]
62. Barrell D, Dimmer E, Huntley RP, Binns D, O'Donovan C, Apweiler R. The GOA database in 2009—an integrated Gene Ontology Annotation resource. *Nucleic Acids Res.* 2009; 37:D396–403. [PubMed: 18957448]
63. Le Novère N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, Schilstra M, Shapiro B, et al. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.* 2006; 34:D689–691. [PubMed: 16381960]
64. Li C, Courtot M, Le Novère N, Laibe C. BioModels.net Web Services, a free and integrated toolkit for computational modelling software. *Brief Bioinform.* 11:270–277. [PubMed: 19939940]
65. Howe D, Costanzo M, Fey P, Gojobori T, Hannick L, Hide W, Hill DP, Kania R, Schaeffer M, St Pierre S, et al. Big data: The future of biocuration. *Nature.* 2008; 455:47–50. [PubMed: 18769432]
66. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics.* 2003; 19:524–531. [PubMed: 12611808]
67. Lloyd CM, Halstead MD, Nielsen PF. CellML: its future, present and past. *Prog Biophys Mol Biol.* 2004; 85:433–450. [PubMed: 15142756]
68. Christie GR, Nielsen PM, Blackett SA, Bradley CP, Hunter PJ. FieldML: concepts and implementation. *Philos Transact A Math Phys Eng Sci.* 2009; 367:1869–1884. [PubMed: 19380316]
69. Deisboeck TS, Sagotsky J. Professional networks in the life sciences: linking the linked. *Cancer Inform.* 9:189–195. [PubMed: 20838607]
70. Wilkins-Diehr N, Gannon D, Klimeck G, Oster S, Pamidighantam S. TeraGrid Science Gateways and Their Impact on Science. *Computer.* 2008; 41:32.
71. TUMOR: Transatlantic Tumour Model Repositories. Available at: <http://tumor-project.eu/>

Figure 1.

A high-level architecture of the DMR. CViT's DMR is built on top of an RDF data store that stores data in an IBM DB2 database. RDF allows CViT to semantically annotate the content of the repository and store hyperlinks to other resources. The CViT.org website provides the graphical user interface to the repository. Through CViT.org, scientists can add new models, share models with other researchers, and discuss model simulations. The DMR Grid Service provides a caBIG Silver-level compliant interface to the DMR, which allows caBIG client applications to securely upload and access models and model metadata within the repository. The CMEF Server (bottom left) provides capabilities to execute the models stored in the repository on a computational grid.

Figure 2.

The initial Digital Model Repository domain model was constructed to expose the content of the CViT repository through a caBIG silver-level compliant data service. It took several months of interaction with caBIG's work groups to refine the model and annotate it with NCI Thesaurus concepts.

Figure 3. Snapshots of key DMR workflow features. (A) An example of a complete Entry in the DMR with title, description, references, etc. (C) User publishes his entry to other users in the DMR. (D) New contributors are added to a DMR entry.

Figure 4. Snapshots of key CMEF workflow features. **(A)** New model wizard. **(B)** User submits a job – simulation task. **(C)** Job results page. The model added as an example is a two-dimensional simulation model for investigating lung cancer growth across molecular and multicellular scales⁵⁵.

Figure 5.

CMEF components of the DMR architecture. The CMEF extends the functionality of the DMR by enabling any model stored within the repository to be executed within a grid-based execution environment. It consists of the following services. (1) The Model Administration Service lets the PI add semantic metadata necessary to describe the parameters for model execution. (2) The Model Execution Service handles the requests from users to execute models and is in charge of linking the input data specified with the model, and of submitting the execution to the Job Scheduler. (3) The Job Scheduler interacts with external Grid Execution Engines in order to send a model for actual execution. (4) The Execution Monitoring Service receives events from the models under execution and from the Grid Execution Engine, and provides information to the user through the CViT website. Upon completion of execution of a model, it stores the simulation results within the DMR.

Figure 6. Extended Elements of the DMR Domain Model to support the CMEF. The Computational Model Execution Framework (CMEF) domain model extends from the CViT Digital Model Repository (DMR) domain model to introduce additional classes to support the requirements of the CMEF.

Figure 7.
Snapshots of the CViT mashup (A) and DMR graph (B).

Figure 8. Snapshot of CollaboRank displaying registered institutions' sharing scores.

TABLE 1

Different layouts of databases. **(A)** and **(B)** represent a pair of RDBMS tables which must be cross referenced to retrieve user email addresses. **(C)** Depicts how a Semantic Web can show all the same information without creating a separate layer.

<i>A: SQL Names</i>		
id	fname	Lname
usr1	Abby	Aronson
usr2	Bert	Buford
usr3	Carry	Crawford

<i>B: SQL Email Addresses</i>	
id	Email
usr1	aa@domain.edu
usr1	abby@domain.com
usr1	a.aronson@domain.net
usr2	bert@domain.gov
usr2	bbuford@domain.org

<i>C: Semantic Web Names & Email Addresses</i>		
usr1	fname	Abby
usr1	lname	Aronson
usr1	email	aa@domain.edu
usr1	email	abby@domain.com
usr1	email	a.aronson@domain.net
usr2	email	bert@domain.gov
usr2	email	bbuford@domain.org

TABLE 2

Ontologies used in the DMR. An Entity organizes a set of Attributes defined in an Ontology in order to describe data objects in the DMR. Attribute terms are selected from FOAF, DC, or OWL or are added to the CViT ontology.

Entity	Ontology	Attribute	NCI Concept Code	Description
User	CViT		C25190	User profile
		Depiction	C54273	Profile image of the user
	FOAF	Firstname	C40974	User's first name
		Homepage	C19467	Link to user's website
		Mbox	C42775	User's email address
		Phone	C40978	User's phone number
		Surname	C40975	User's last name
		Title	C25354	User's degree, e.g., Ph.D., M.D.
	CViT	Address	C70946	User's street address
		City	-	User's city
		Fax	C42879	User's fax number
		Group	C41167	DMR grouping
		inRole	-	Special privileges such as institution membership, PI-status, LO-status
		Organization	C19711	The institution to which a user belongs
		Position	C19067, C25193	Position within an institution
		Research	C25284, C15429	Paragraph describing user's research
		researchInterests	C48910, C15429	An enumeration of common user interests
		Seniority	C25554, C25193	Expertise level, e.g., grad student, post doc, faculty
		State	-	User's home state
		Zipcode	-	User's zip code
	OWL	Type	C25190	Asserts that user is a user and not an entry
Organization			C19711	An institution. Most of these will be colleges, universities, and research institutes.
	DC	Source	C19467	Link to institution's web page
		Title	C42614	Institution's name
	OWL	Type	C25284, C15429	Asserts that this organization is of the organization type
	CViT	Description	C25365	Optional description for institution
		Geocode	C25341, C68643, C68642	Latitude and longitude of organization's campus
Entry		researchType	-	Description of research being done at this organization
			C47885	An entry in the repository contains relevant, uploaded information regarding a project.
	OWL	Type	C25284, C25474, C47885	Asserts that entry is of the entry type
	DC	Contributor	C25190	ID of users who have write access to this entry
	CViT	Description	C25365	Description of entry
		Title	C42774	Title of entry
		principalInvestigator	C25190, C25190	Owner of entry

Entity	Ontology	Attribute	NCI Concept Code	Description
		Abstract	C60765	A brief summary of the project's description
		Concept	C48910	Background and basic idea of this project
		Hypothesis	C28362	What assumption(s) will be proved by this experiment
		Conclusion	C54033	The outcome and significance of the project
		Data	C25474	Data files and computational models associated with this project
		Categories	C25372, C47885	Collection of Entry Categories from the CViT ontology classifying the project
		References	C25641	Publications or other Web resources related to this experiment

TABLE 3

CViT computational cancer model classification ontology.

Classifier	Definition
Genetics	The branch of biology that deals with heredity, especially the mechanisms of hereditary transmission and the variation of inherited characteristics among similar or related organisms
Cell Cycle	The complex series of phenomena, occurring between the end of one cell division and the end of the next, by which cellular material is divided between daughter cells. The cell cycle is an ordered set of events, culminating in cell growth and division into two daughter cells. The stages of cell cycle includes G1-S-G2-M. The G1 stage stands for "GAP 1". The S stage stands for "Synthesis". This is the stage when DNA replication occurs. The G2 stage stands for "GAP 2". The M stage stands for "mitosis," and is when nuclear (chromosomes separate) and cytoplasmic (cytokinesis) division occur.
Signaling Pathway	An elaboration of the known or inferred interactions involved in a signal transduction pathway.
Proliferation	Growth and reproduction of new similar forms, e.g. cells, buds, or offspring.
Motility	Cell Motility consists of active translocation of a whole cell, or cell body, from one site to another; distinct from cell motion that involves movement of cell processes (e.g., axons, microvilli, etc.).
Angiogenesis	Development of new blood vessels.
Metastasis	Metastasis is the spread or migration of cancer cells from one part of the body (the organ in which it first appeared) to another. The secondary tumor contains cells that are like those in the original (primary) tumor. For example, breast cancer cells may spread (metastasize) to the lungs and cause the growth of a new tumor. When this happens, the disease is called metastatic breast cancer.
Immunology	The study of the immune system and its reaction to pathogens, as well as its malfunctions (autoimmune diseases, allergies, rejection of organ transplants).
Microenvironment	The complex relationships between tumor cells and the neighboring cells in the host environment. Cellular signaling within the microenvironment can promote the continuing survival and growth of tumor cells, or apoptosis (cell death). Research is currently underway to manipulate this relationship by altering the host environment in ways that silence or inhibit pro-survival signals concurrent with standard therapies.
Treatment	A type of study protocol designed to evaluate intervention(s) for disease treatment.
Statistics	A branch of applied mathematics concerned with the collection and interpretation of quantitative data and the use of probability theory to estimate population parameters.
Discrete	Representation of data as a chronological sequence of events. Each event occurs at specific instant and marks a change of state in a system.
Continuum	The mathematical practice of applying a model to continuous data which has a potentially infinite number, and divisibility, of attributes.
Hybrid	The integration of both discrete and continuum modeling techniques in representing distinct dynamics and topographic regions of a system.

TABLE 4

Major entities added to the DMR domain model to support the CMEF.

Entity	Attribute	Description
Computational Model		Extends from DataClassification to encapsulate the content and metadata of an executable computational model
	Name	The name of the model
	Description	A short description of how the computational model works
	commandLine	Command line used to execute the model. Values enclosed in angle brackets are replaced by corresponding ParameterValues before the model is executed
	Version	Software version of the computational model
	Files	Source files that constitute the computational model and requisite executable files.
	Documentation	User's Guide or documentation describing the use of the computational model
	Computer	Describes model program execution hardware constraints
	Program	Program execution language constraints (Java, Perl, R, C++, etc.) required by the model
	Parameters	Program parameters that can be set for the model
Parameter		Defines the metadata for describing an individual input value for a computational model. Parameters may be input files or values entered on the command line.
	Name	The name of the parameter – for a file, it should be the file's name, for a command line parameter, it should match the name specified in commandLine, for example "<name>"
	Description	Description of the parameter or value constraints
	dataType	Defines the required parameter value type (Text Integer Float File)
	Prefix	A command line prefix that will be added if the value is present. For example "-F" or "-o"
	Choices	Set of values that constrain the input parameter
	defaultValue	Default value used during model execution if no value is specified
	isOptional	Indicates that the parameter value can be omitted (if true)
	isFile	Indicates that the parameter value is a file
Computation Job		Extends from DataClassification to encapsulate the content and metadata of an executing (or executed) computational model
	dateSubmitted	The date/time that the computation job was submitted for execution
	dateCompleted	The date/time that the computation job execution completed
	jobNumber	System-assigned number identifying the model execution job
	jobStatus	Result of running the model (Success Failure)
	jobParameterValue	Values set for each Parameter of the model for the job
	jobFiles	Files produced by executing the model computation job, may include execution log, console output, and output files

TABLE 5

caBIG™ Silver-level compliant data service functions currently provided by the DMR.

Method	Description
Entry addEntry(Entry newEntry, Organization fundingOrganization)	Stores a new Entry in DMR. Returns the stored Entry with DMR-assigned Entry id.
DataClassification addDataToEntry(DataClassification data, Entry sourceEntry)	Adds the provided Data to the DMR and sets association to Entry identified by sourceEntry. If Data already exists in DMR, method will only set association. If Data does not exist in DMR, Data will be added. Returns the DataClassification with DMR-assigned DataClassification.id.
Reference addReferenceToEntry(Reference reference, Entry sourceEntry)	Adds the provided Reference to the DMR and sets association to Entry identified by sourceEntry. If Reference already exists in DMR, method will only set association. If Reference does not exist in DMR, Reference will be added.
void updateEntry(Entry entry)	Updates the annotation fields of the given Entry within the repository.
void updateData(DataClassification data)	Updates the annotation fields of the given Data within the repository.
void updateReference(Reference reference)	Updates the annotation fields of the given Reference within the repository.
CQLResult query(CQLQuery cqlQuery)	Executes a caBIG® Query Language (CQL) query against the information in the repository.