

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

## Journal of Biomedical Informatics

journal homepage: [www.elsevier.com/locate/yjbin](http://www.elsevier.com/locate/yjbin)

## The Biomedical Resource Ontology (BRO) to enable resource discovery in clinical and translational research

Jessica D. Tenenbaum<sup>a,\*</sup>, Patricia L. Whetzel<sup>e</sup>, Kent Anderson<sup>b</sup>, Charles D. Borromeo<sup>c</sup>, Ivo D. Dinov<sup>k</sup>, Davera Gabriel<sup>b</sup>, Beth Kirschner<sup>h</sup>, Barbara Mirel<sup>h,l</sup>, Tim Morris<sup>d</sup>, Natasha Noy<sup>e</sup>, Csongor Nyulas<sup>e</sup>, David Rubenson<sup>o</sup>, Paul R. Saxman<sup>n</sup>, Harpreet Singh<sup>c</sup>, Nancy Whelan<sup>c</sup>, Zach Wright<sup>f</sup>, Brian D. Athey<sup>n,h</sup>, Michael J. Becich<sup>c</sup>, Geoffrey S. Ginsburg<sup>i</sup>, Mark A. Musen<sup>e</sup>, Kevin A. Smith<sup>n</sup>, Alice F. Tarantal<sup>g</sup>, Daniel L. Rubin<sup>j</sup>, Peter Lyster<sup>m</sup>

<sup>a</sup> Duke Translational Medicine Institute, Duke University School of Medicine, Durham, NC, USA

<sup>b</sup> Clinical and Translational Science Center, University of California at Davis Health System, Sacramento, CA, USA

<sup>c</sup> Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA

<sup>d</sup> Research and Health Sciences IT Division, Emory University, Atlanta, GA, USA

<sup>e</sup> Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, USA

<sup>f</sup> Center for Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA

<sup>g</sup> Departments of Pediatrics and Cell Biology and Human Anatomy, School of Medicine, UC Davis Clinical and Translational Science Center, University of California, Davis, CA, USA

<sup>h</sup> National Center for Integrative Biomedical Informatics, University of Michigan, Ann Arbor, MI, USA

<sup>i</sup> Institute for Genome Sciences & Policy, Duke University, Durham, NC, USA

<sup>j</sup> Department of Radiology, Stanford University School of Medicine, Stanford, CA, USA

<sup>k</sup> Center for Computational Biology, Laboratory of Neuro Imaging, University of California, Los Angeles, Los Angeles, California, USA

<sup>l</sup> School of Education, University of Michigan, Ann Arbor, MI, USA

<sup>m</sup> Center for Bioinformatics and Computational Biology, National Institute of General Medical Sciences, Bethesda, MD, USA

<sup>n</sup> Michigan Institute for Clinical & Health Research, University of Michigan, Ann Arbor, MI, USA

<sup>o</sup> Stanford University Cancer Center, Stanford University, Stanford, CA, USA

### ARTICLE INFO

#### Article history:

Available online 16 October 2010

#### Keywords:

Ontology

Biositemaps

Resources

Biomedical research

Resource annotation

Resource discovery

Search

Semantic web

Web 2.0

Clinical and Translational Science Awards

### ABSTRACT

The biomedical research community relies on a diverse set of resources, both within their own institutions and at other research centers. In addition, an increasing number of shared electronic resources have been developed. Without effective means to locate and query these resources, it is challenging, if not impossible, for investigators to be aware of the myriad resources available, or to effectively perform resource discovery when the need arises. In this paper, we describe the development and use of the Biomedical Resource Ontology (BRO) to enable semantic annotation and discovery of biomedical resources. We also describe the Resource Discovery System (RDS) which is a federated, inter-institutional pilot project that uses the BRO to facilitate resource discovery on the Internet. Through the RDS framework and its associated Biositemaps infrastructure, the BRO facilitates semantic search and discovery of biomedical resources, breaking down barriers and streamlining scientific research that will improve human health.

© 2010 Elsevier Inc. All rights reserved.

### 1. Statement of the problem to be addressed

The biomedical research community uses a diverse set of resources to conduct research. These resources include computer software, animal models, regulatory expertise, facilities and cores, and training programs, to name just a few. Investigators who are

able to leverage these resources to facilitate their research can be more efficient and avoid duplication of effort. The inventory of resources within the research community is continually growing, changing, and evolving. Software packages are upgraded to include new functionality, technology cores are established, and new instruments are introduced. Information about these many valuable resources is scattered across institutional and laboratory websites, and may be highlighted only in publications or conference proceedings, if at all. Without a readily accessible inventory it is challenging, if not impossible, for investigators to be aware of the

\* Corresponding author. Address: PO Box 17969, Durham, NC 27715, United States. Fax: +1 919 668 7868.

E-mail address: [jessie.tenenbaum@duke.edu](mailto:jessie.tenenbaum@duke.edu) (J.D. Tenenbaum).

myriad resources available to facilitate their research, or to effectively perform resource discovery when the need arises. Many valuable federal or state-funded resources may be underutilized without information sharing, advertisement, and active promotion. There is also risk for unnecessary duplication of resources not only within institutions but among regional collaborating groups.

General-purpose Web search engines are useful and ubiquitous, but with millions of pages indexed, they lack specificity for searching complex and technologically advanced research resources. For instance, Google would return information, publications, and image results in response to a search for “animal models,” although the user may have been attempting to obtain information related to facilities with specialized expertise and histopathology-related resources. More important, a Google search would not distinguish between the thousands of textual Web pages that simply contain the words “animal models” and those that provide information relevant for a biomedical investigator in need of a particular facility.

The **Biositemaps** technology was developed as a collaboration between the National Centers for Biomedical Computing (NCBC) and Clinical and Translational Science Awards (CTSA) consortia [1]. In contrast to text-based search engines, it allows Web site authors to store structured information that enables special-purpose search engines to identify precisely those research-related resources that are of interest to investigators, and to provide specific information for accessing those resources. This paper describes the development and use of Version 3.0 of the **Biomedical Resource Ontology** (BRO) [2] to enable semantic resource annotation in the context of the **Resource Discovery System** (RDS) project [3–4] a federated, inter-institutional pilot initiative to facilitate resource discovery on the Internet. RDS (formerly the CTSA Informatics Inventory of Resources Web Presence, or CIRWP) uses the Biositemaps infrastructure [5] and was developed as a collaboration among six members of the Clinical and Translational Science Awards (CTSA) Consortium.

## 2. User requirements

Early in the course of this project, we conducted a series of six interviews with both translational researchers ( $N = 3$ ) and directors of translational technology resources ( $N = 3$ ). Interviews were done by phone and lasted approximately 1 h each. Findings from these interviews suggest that there are two general needs that motivate investigators to search for resources. The first is to gain access to resources the investigator requires in order to conduct his or her research. Many resources, for example complex scientific instruments, are expensive and available in only a few institutions; it would simply not be practical or feasible for each investigator in need of such a resource to consider purchase and maintenance. Where such instruments *have* been purchased, either by an individual investigator or as a shared resource, the more the resource can be leveraged by other users, the greater the return on investment to the scientific community. In addition, an increasing number of shared electronic resources are being made available in the public domain. Such resources may include software tools, computational algorithms, datasets, or high performance computing environments. These computational resources are generally exempt from the geographical considerations that might apply to, for example, a biobanking facility. It is therefore important to make their availability known to researchers throughout the country.

The second reason to search for external resources is to exchange information regarding use or management of a specific resource. Frequently the person responsible for a highly specialized technology faces domain-specific obstacles for which it would be helpful to connect with others who manage similar facilities in order to exchange best practices and lessons learned. To this end,

resource owners might seek contact information for personnel associated with a similar technology or offering.

Table 1 presents the top three use cases identified through both empirical experience and the formal investigator interviews. Use cases helped to drive system development and were used to evaluate the tool in design walk-throughs, usability inspections and expert reviews.

We note that Use Case 3 comprises two types of queries; the first is to identify candidate technologies for the application, and the second is to locate an instrument appropriate to the selected technology. A key strength of BRO and our RDS is that the user need not be aware of these queries; rather, the user can use our system to directly fulfill both needs.

## 3. Approach

To address the need for a federated, readily accessible inventory of research resources on the Internet, a consortium of investigators from six institutions within the Clinical and Translational Science Awards (CTSA) Consortium developed the Resource Discovery System (RDS) [3]. As part of the RDS project, the NCBC and CTSA teams collaboratively extended the development of the BRO and Biositemaps infrastructure as will be described in this paper. Following the first year of development, the RDS serves as an invaluable project implementing a federated resource annotation and semantic searching system, using information accumulated through other pilot projects on resources at multiple sites. It provides a number of lessons learned for moving this important area of investigation forward and highlights a number of challenges that remain to be addressed, both social and technological. Although still a pilot effort, the RDS [3], the Biositemaps infrastructure [5], and the BRO [2] are openly accessible for use, and we have already seen growth in content from a number of institutions and research groups.

In this manuscript, we describe the BRO and RDS projects and how they leverage the Biositemaps infrastructure. We address the issues and challenges that the project has uncovered, which remain active areas of ongoing investigation for the project team. Finally, we describe ongoing efforts toward harmonization with other related efforts which have different data models, including the Neuroscience Information Framework (NIF) [6], the Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC) [7], and the eagle-i Consortium [8].

The RDS project leveraged and built upon several existing NIH initiatives including CTSA Working Groups and Administrative Supplement Grants, and the NIH Roadmap National Centers for Biomedical Computing (NCBC). An overview of the relationships between the various components is shown in Fig. 1.

## 4. Design principles

We adopted four core principles as the basis for our design approach with the RDS project and associated extension of the BRO and Biositemaps infrastructure:

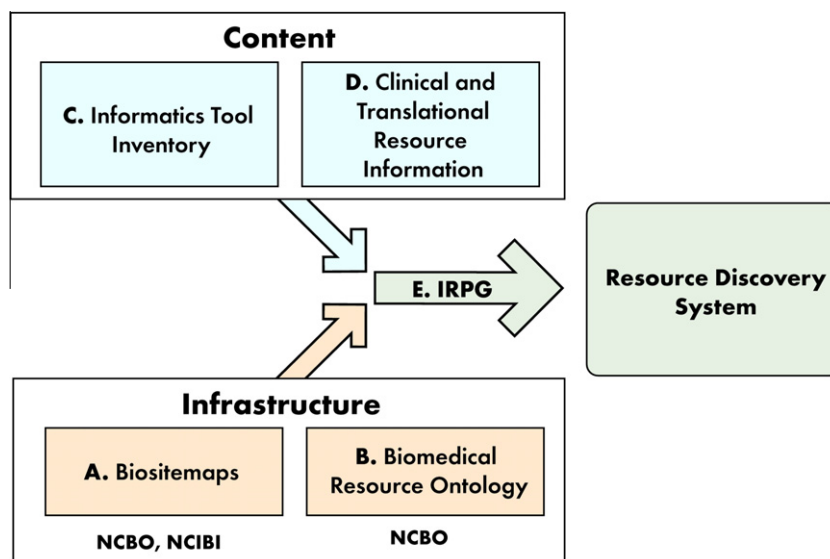
- (1) addressing real-world challenges faced by biomedical researchers;
- (2) leveraging existing technology;
- (3) design simplicity; and
- (4) employing iterative development to enable continuous refinement.

These principles are illustrated in the approach employed for system development:

**Table 1**

Top three use cases identified through both empirical experience and formal investigator interviews.

Use case	Description
1	A researcher wants to know where to send samples for full genome sequencing using a particular type of next generation sequencer.
2	A genomics core director wants to know what experiences others have had with two different data storage system vendors.
3	A researcher is studying physiology and metabolism. She already makes use of a calorimeter at her home institution, but is not aware of a double-labeled water technology to quantify oxidation, available at another institution – and useful for various applications within the study of metabolism and physiology.



**Fig. 1.** Relationship between the various contributing groups to the RDS initiative. The foundational technology infrastructure consists of: (A) The Biositemaps infrastructure with its associated Biositemaps Information Model [5] for resource meta-data broadcast and retrieval. (B) The Biomedical Resource Ontology (BRO), which provides a controlled terminology for annotation of resources. (C) Resource information and annotation provided by the Informatics Inventory Resource Working Group (IRWG), a voluntary effort through the Informatics Key Function Committee of the CTSA to provide local inventories of informatics tools from each of the 46 CTSA sites, and (D) Investigators at Duke University and the University of California Davis who worked to identify facilities, cores, and resources for translational research, and to develop and annotate a pilot inventory of such resources among seven CTSA sites. (E) The Informatics Inventory Resource Project Group (IRPG) enhanced and integrated these pre-existing efforts to implement the RDS, a Web-accessible inventory of biomedical research resources.

#### 4.1. Grounding in real-world challenges

Based on investigator interviews, we defined key use cases for resource discovery (Table 1). This ensured that the ultimate outcome would be informed by, and address, challenges faced by real-world users, and not simply by what is technically possible or easily achieved.

#### 4.2. Leveraging existing technologies

Rather than build a new system from the ground up, we chose to embrace and extend existing infrastructure – the Biositemaps technology, Biositemaps Information Model and Biomedical Resource Ontology for describing biomedical resources (see Section 5). By leveraging existing technologies, we were able to focus on functional requirements and save time required to build the initial infrastructure.

#### 4.3. Design simplicity

We aimed to keep the design as simple as possible to enable a decentralized approach in which resource owners or curators can easily describe their resources in a structured manner, and those searching for resources can successfully carry out the key use cases (Table 1). A more complex data model to describe resources would enable more complex queries and inferences regarding resources, however, it would also result in the need for added complexity

in resource annotation. While such complexity is less of an obstacle for a centralized solution in which resource annotation and curation is performed by dedicated personnel with training and support, RDS is intended for use by a broader, more heterogeneous community.

#### 4.4. Iterative development

We performed iterative development across all components of the RDS, BRO, and Biositemaps projects, harmonizing both within the project and with related initiatives. Iterative development, as opposed to creating and adhering to complete and final specifications developed up front, enables more rapid development as well as incorporation of lessons learned along the way.

### 5. System overview

Fig. 2 gives an overview of the end-to-end RDS. The various components are explained in more detail below.

#### 5.1. Biositemaps

The Biositemaps technology is a mechanism designed to enable basic scientists, bioinformaticians, clinicians, and translational scientists to broadcast, search, compare and retrieve meta-data about diverse computational biology resources [1,9]. Information describing biomedical resources is encoded in an RDF (Resource

Description Framework) file published on an institution's Web site. The approach is analogous to, and was inspired by, Sitemaps [10], which provide a means for webmasters to inform search engines about pages on their sites that are available for crawling. Biositemaps RDF files encode resource meta-data, which is made available to web-crawlers and Biositemaps query tools. Institutions, groups and individuals with biomedical research resources may publish Biositemaps files on their Web site and register the location of the file with the Biositemaps registry [11]. Each biositemap.rdf contains meta-data describing the institution's resources for biomedical research. Resources described in online-accessible Biositemaps files can be discovered, parsed, and catalogued by web crawlers and search agents such as the RDF query tool (Fig. 2). Originally, resources published through Biositemaps were limited to informatics-oriented datasets and tools, such as software, Web services, and algorithms. Our work has expanded the scope of the resources beyond computational tools to include basic, clinical and translational research more broadly.

### 5.2. The Biositemaps Information Model (BIM)

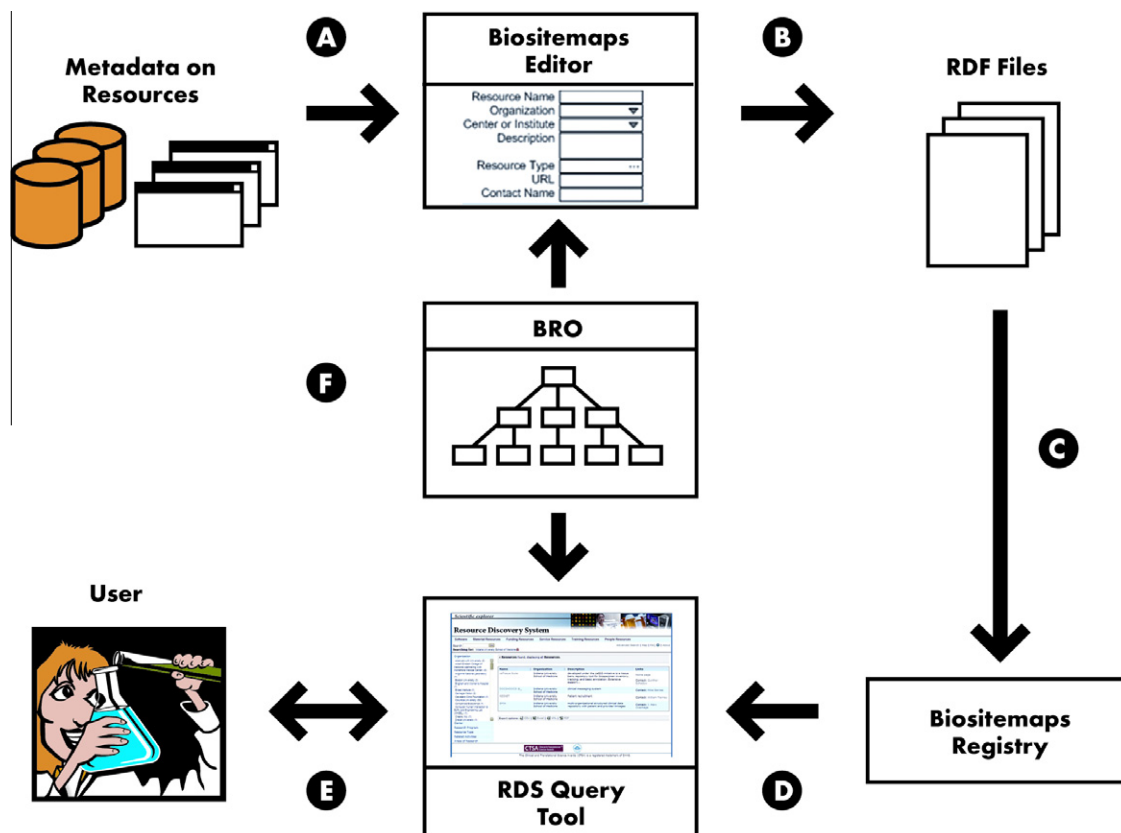
The BIM is a set of properties that are used to specify meta-data for a resource (*Resource name, Organization, etc.*). The formal specification is published on the Biositemaps Web site [5]. The initial version of the BIM was based on the requirements for the NCBC consortium. It was used to describe the informatics tools and resources offered by the seven funded NCBCs, or any other site that wished to contribute resources. In parallel with the development of the BIM (before the efforts were coordinated), the CTSA Informatics Inventory Resources Project Group (IRPG) developed a basic

information model comprising a list of attributes in an Excel spreadsheet used to collect information on informatics resources from all funded CTSA sites. The harmonized version of the BIM, consisting of coded values, free text, and ontological domains, can be found on the Biositemaps website [5]. Many elements are optional, applicable primarily to informatics resources. This is an area of active development; we have identified a core set of elements, and seek to define other "modules" of data elements, to be utilized for different resource types being described.

### 5.3. The Biomedical Resource Ontology (BRO)

The BRO [2] was developed as an ontology to classify types of biomedical resources. In the context of Biositemaps, the purpose of the BRO is to provide a controlled terminology to provide values for the BIM *Resource Type* attribute. In addition to a controlled list of names of resources, a taxonomy of the resources was needed to enable searches at varying levels of granularity and abstraction. For example, a researcher may wish to search for all imaging software, or for those packages that provide image segmentation. In addition to the taxonomic relationships, the ontologic structure of BRO allows for the possibility of adding other relations such as the manner in which individual resource types may be components of some composite resource types.

The first draft of BRO was built by conducting interviews with investigators in different disciplines and with different scientific backgrounds. The classes in BRO represented types of informatics resources. Subsequently, both the breadth and depth of the BRO have been expanded in an iterative process. Classes representing more specific types of resources have been added, and the range



**Fig. 2.** Overview of end-to-end RDS system. (A) A user (content generator) at an institution enters information about institutional resources using the Biositemaps Editor. (B) The Biositemaps Editor generates an RDF file, stored locally on a server at the institution. (C) The location of that RDF is registered through the Biositemaps registry. (D) The RDS query tool searches content from the registered RDF files. (E) A researcher uses the query tool to search for resources. (F) The BRO is used by the Biositemaps Editor to provide valid values for the Resource Type property in the editor, and by the query tool to provide a list of terms by which the user can easily search for resources.

of types of resources has been expanded in order to support the annotation of translational, non-informatics-oriented resources. A medical librarian augmented the content and reorganized the initial structure of BRO. The BRO was subsequently uploaded to the National Center for Biomedical Ontology (NCBO) BioPortal [2] where it is available for community review, comment, and use.

BRO development has been, and continues to be, an iterative process. While early versions of the BRO served as a useful controlled terminology for preliminary RDS testing, some aspects of the ontology lacked ontological rigor. For version 3.0 of the BRO, released in March 2010, a devoted task force was formed with two main goals: (1) formalize a set of principles to which the class hierarchy and definitions must adhere, and (2) BRO is *self consistent* in the sense that it follows an 'is\_a' hierarchy and the locations in the hierarchy are consistent with the class definitions. It is *complete* up to the following level of the resource hierarchy: Funding Resource, Information Resource, Material Resource, People Resource, Service Resource, Software and Training Resource (note that Software is synonymous with Software Resource as per agreement with NIF). Beyond that level it covers all the classes that are needed for current Biositemaps usage. However, it is not formally complete, e.g., under Software some branches of the hierarchy are deeper than others. BRO 3.0 was informed by real-world use cases and represents a significant improvement in ontological consistency from previous versions, however, it is a work in progress and we will continue to refine and make more complete as we receive community feedback and harmonize with other important technologies like eagle-i, NITRC, and NIF.

As stated above, the BRO is an ontology of *resource types*. There are two key aspects of a resource type: first, there is the *means or method by which it provides access to something*, and second there is the *entity to which that access is being provided*. For example, data on a hard drive is not a resource unless there is some means of accessing that data. In that case, the resource is not the actual data itself, but rather the repository or Web service that provides access to that data. Such a repository or service would be classified as type *data resource* because it provides access to data. Analogously, a pet store would be considered an *animal resource* type, as opposed to the animals in the store, which are not in themselves resource types. Definitions of classes in BRO generally conform to the structure "A resource that provides..." or "A resource that provides access to..." In some cases, this structure is implied by referencing the parent class with further qualification. For example, a *facility core* is "A resource that provides instruments, technologies, facilities, and/or expert support for a specific area of research." Its child, *Fabrication Facility* is defined as "A facility core devoted to creating, manufacturing, building or assembling resources used in scientific research." Referencing the parent in the definition avoids excessive verbosity in definitions, and redundancy across sub-classes. The one other exception to general rule for definition structure is the top-level class *Resource*, defined as *mechanism that provides access (either in the open community or within an organization) to material, intellectual, financial, technological, or electronic means of carrying out research and development*. Second level (child) and subsequent levels are classes with an *is\_a* hierarchy (Fig. 2).

The foundational approach for the BRO may be described as Aristotelian in that the definition of a term conveys "what makes an entity of a given sort an entity of that sort" [12] and that it takes on the form "An A is a B which..." Like the Gene Ontology, however, and in contrast with the more rigorous Foundational Model of Anatomy, the BRO relies primarily on *is\_a* relationships [12–14]. A more complex model was consciously avoided in order to maximize ease of use by a distributed group of curators with varying levels of technical expertise, though in the future we will introduce a limited number of properties and additional relationship

types. While this design decision does not take full advantage of the richness and semantic complexity that an ontology affords over a list of terms and their definitions, it does simplify the terminology for ease of use by researchers. It also makes it difficult to avoid multiple inheritance. In this respect we have diverged from the Aristotelian ideal. While inferred multiple inheritance through the use of properties is generally considered preferable, and we plan to move in that direction in the future; there are currently certain cases where there is asserted multiple inheritance in order to improve usability of the BRO as a classification scheme. This trade-off was deemed worthwhile given that user navigation is the primary use case scenario for the BRO, as opposed to automated inferences over the ontological hierarchy.

To improve usability of the BRO, we have made use of SKOS (Simple Knowledge Organization System), a W3C standard [15]. In the BRO, SKOS provides lexical labels such as *preferred term* and *synonym*.

The Biositemaps Initiative reflects the growing interest among the biomedical research community in *understanding the scope and availability of biomedical resources and resource types*. The work of the Neuroscience Information Framework (NIF), an effort growing out of the 2004 NIH Blueprint for Neuroscience Research, is one such example. NIF is focused on a broad-based effort to bring neuroscience relevant information from across the research community to individual neuroscientists. This includes a component aimed at identifying resources relevant to neuroscience. Eagle-i is a new initiative funded by the American Recovery and Reinvestment Act (ARRA) that will focus on the identification of research resources within the nine institutions participating in the initiative. We are aware of some similar aims between NIF and Biositemaps and have made efforts to harmonize the ontologies when appropriate. In May 2009 the authors representing the NCBC Working Groups and CTSA/RDS efforts collaborated with the NIF team to harmonize second level class names (Table 2). Communication along these lines with the eagle-i team is ongoing. Ultimately it will be desirable to conduct a detailed systematic comparison of these and other related initiatives, to harmonize the approaches when there are common aims, and to highlight the distinctions to better support users with differing needs.

We evaluated the BRO in terms of its sufficiency for annotating biomedical resources that had been compiled by several research groups. The IRWG collected 370 informatics resources from 40 different CTSA sites. In addition, a group within the CTSA Translational Steering Committee collected a list of more than 450 translational resources by manually mining the websites of seven CTSA organizations. (The complete list can be found via the online query tool [3]). Despite the number and variety of the resources collected, almost all resources were able to be annotated with an existing BRO class. Where no suitable class existed, the BRO was extended. Extension of the BRO in this empirical manner is an ongoing activity as additional resource types are identified. Interested parties are encouraged to comment on BRO by posting a comment on the ontology or specific terms using the BioPortal Notes (Fig. 3) [2] and by joining the BRO discussion group [16].

#### 5.4. Biositemap RDF file generation and back-end data storage

The Biositemaps Editor [17] allows a user to generate a BIM-compliant RDF file describing a set of resources available within a given institution. The Editor provides a simple Web-based interface to collect descriptions of resources using text boxes and drop down menus as shown in Fig. 4, which displays the CTSA-specific editor. Other applications such as the iTools web-based navigator can also be used to generate Biositemaps from excel spreadsheet or by manually entering resource meta-data [9]. The BIM defines the set of Resource Properties shown within the Biositemaps Editor

**Table 2**

Second level BRO class names and definitions. Each of these classes is a child of the top level *Resource* class in an *is\_a* hierarchy.

Class name	Definition
<i>Funding resource</i>	A resource that provides monetary support in the form of grants, contracts, or gifts for research, training, or education.
<i>Information resource</i>	A resource that provides data, knowledge or narrative.
<i>Material resource</i>	A resource that provides items such as reagents, instruments, tissue samples or organisms.
<i>People resource</i>	A resource comprised of one or more individuals who have specific expertise and provide professional or expert advice.
<i>Service resource</i>	A resource that provides an act of helpful activity, consisting of an organized system of apparatus, appliances, technology, personnel, etc.
<i>Software</i>	Resource that provides access usually through repositories or the Internet to computer executables, libraries, plugins, or source code.
<i>Training resource</i>	A resource that provides access to educational materials or events, such as courses, workshops or graduate programs.

(Fig. 4). The Biositemaps Editor accesses the BRO from BioPortal [2] in order to populate the list of *Resource Type* choices in the editor. The editor also provides drop down lists of values for properties such as *Organization* and *Center or Institute* and free text fields where appropriate. The RDF file generated by the editor is saved locally and posted to a publicly available folder or directory on the author's web site. The RDF author then publishes the location of this RDF file through the Biositemaps registry [11].

The publicly-available corpus of Biositemaps files can be queried in a number of ways. For the RDS implementation, the list of URLs in the registry is used by Mulgara technology [18] to build a data store of resource meta-data using the data from the RDF file found at each published URL. The Mulgara application stores the RDF data in a graph-based data structure, which can then be queried by agents such as the Web-based RDS query tool. Mulgara can also perform inferencing on the RDF data, which allows the RDS query tool to return not only resources directly associated with a given BRO term, but also those terms that are children, grandchildren, etc. of a term in the hierarchy. Thus a user can search for resources using the BRO at both broad and specific levels of granularity.

### 5.5. The resource discovery system (RDS) query tool

The RDS query tool was designed to search Biositemaps resource descriptions [3] as shown in Fig. 5. The query tool was designed to be user-friendly for non-technical users. The interface enables the user to perform a basic search involving both free text and faceted search, or advanced search in which free text may be combined with multiple criteria based on properties in the BIM. Included in the free text search is all free text fields designated for a given resource. In the next version of the tool, search by synonyms will be added to improve recall. Results are displayed in a column-sortable list view, with links to a detail page, the home page for the resource, and a mailto: link for the contact person. In addition, results may be exported to a variety of formats including Excel and CSV.

We iteratively designed and implemented the RDS query tool through heuristic evaluations and cognitive walkthroughs, i.e., demonstrated support of the documented use cases (see Table 1). Heuristic evaluations (HE) rate interfaces based on their accord with established usability standards. After early HE of some RDS proof-of-concept tools and comparable systems, a visual

The screenshot displays the BioPortal interface for the Biomedical Resource Ontology (BRO) - Resource View, Version 3.0. The top navigation bar includes links for Browse, Search, Projects, Annotate, All Mappings, All Resources Alpha, and Sign In. The main content area shows the ontology structure on the left and details for the selected 'Material Resource' class on the right.

**Left-hand panel (View Ontology Summary):**

- Resource
  - Funding Resource
    - Federal Funding Resource
    - Private Funding Resource
    - State Funding Resource
  - Information Resource
    - Data Resource
    - Narrative Resource
    - Portal
    - Semi-Structured Knowledge Resource
    - Structured Knowledge Resource
    - Unstructured Knowledge Resource
  - Material Resource (selected)
  - People Resource
  - Service Resource
  - Software
  - Training Resource

**Right-hand panel (Details):**

- ID: BRO:Material\_Resource
- Full Id: [http://bioontology.org/ontologies/BiomedicalResourceOntology.owl#Material\\_Resource](http://bioontology.org/ontologies/BiomedicalResourceOntology.owl#Material_Resource)
- Comment: As per alignment with NIF resource type hierarchy.
- Pref Label: Material Resource
- Definition: A resource that provides items such as reagents, instruments, tissue samples or organisms.

**Fig. 3.** BRO Hierarchy in NCBO BioPortal. The figure shows the tree view of the *is\_a* hierarchy of the BRO in BioPortal (left-hand panel) and the details for each class selected (right-hand panel).

**Fig. 4.** CTSA Biositemaps RDF Editor. Resource entries may be created, copied, or deleted using the buttons in the bottom left of the editor. Once created, each resource entry is listed in the left panel by name. Clicking on a given resource name enables the user to modify the various attributes in the main editor pane through text fields, drop-down controls, or ontology term selection controls. Finally, the user can save an RDF file containing the description of the resources the user created, to be posted on a Web site.

mock-up was created to serve as a functioning prototype. Informal user observation sessions were conducted using this prototype. Collaborators in the project – none of whom had been involved in query tool design – participated in these sessions, interacting with the RDS interface for actual search and retrieval tasks. The users thought aloud in this one hour work session as they searched and observers took notes. Users also participated in post hoc interviews in which – from their expertise in using resource search systems – they provided critical comments on the prototype. User feedback has guided further design as well as requirements for the end-to-end system and many resulting feature requests are in queue for future versions. Formal usability studies are planned for the future with translational researchers not already engaged in the project; i.e., the target audience for the RDS initiative.

## 6. Future directions

Our work to date comprises a pilot effort, and more work remains to produce a final database of searchable resources. Expansion and improvement of the BRO and BIM are ongoing activities, being performed in parallel with an additional “deep dive” of resource inventoring taking place at the University of Pittsburgh, and in collaboration with NIF and the eagle-i Consortium for resource discovery. Other active areas of development include improving performance and usability of the query tool, enhancement of the BIM with specific modules for specific

high-level resource types, and a new version of the Biositemaps Editor that will enable batch editing and import.

In addition to expanded breadth and depth of BRO classes, we will be pursuing three other efforts, each intended to better facilitate the guiding use cases for the project. Recall Use Case 3 from Table 1 above: A researcher is studying physiology and metabolism. She already makes use of a calorimeter at her home institution, but is not aware of a double-labeled water technology to quantify oxidation, available at another institution – and useful for various applications within the study of metabolism and physiology. In the current version, this researcher is able to search for a physiology core facility, and might learn from contacting that core about the double-labeled water technology they use. A number of enhancements to the BRO could help better support a researcher in this situation by incorporating knowledge about the pertinent technologies with the resources.

The first potential enhancement, as mentioned above, is to include additional properties within the BRO itself. While still adhering to our design principle of simplicity, augmenting BRO beyond the current *is\_a* relationships would enable richer querying capabilities. A relation such as “*used\_for*” could be added, enabling resources that are used for similar purposes to be inferred.

Second, we plan to link to ontologies that include richer relationships than the BRO *is\_a* classification. For example, we plan an extension of the BRO to include instrument terms from OBI [19]. As addressed above, the domain of the BRO is limited to class names and a hierarchy that satisfy “Resources that provides...” or

**Scientific explorer**

# Resource Discovery System

Software   Material Resources   Funding Resources   Service Resources   Training Resources   People Resources

Search   Advanced Search | Map | FAQ | About

Searching for: Indiana University School of Medicine

Organization

- Aberystwyth University (2)
- Albert Einstein College of Medicine (partnering with Montefiore Medical Center) (6)
- Argonne National Laboratory (1)
- Boston University (3)
- Brigham and Women's Hospital (1)
- Broad Institute (1)
- Carnegie Mellon (2)
- Cleveland Clinic Foundation (1)
- Columbia University (58)
- Compendia Bioscience (1)
- Computer Human Interaction & Software Engineering Lab (CHISEL) (1)
- Creare, Inc. (1)
- Drexel University (1)
- Center
- Research Program
- Resource Type
- Related Activities
- Areas of Research

4 Resources found, displaying all Resources.

Name	Organization	Description	Links
caTissue Suite	Indiana University School of Medicine	developed under the caBIG initiative is a tissue bank repository tool for biospecimen inventory, tracking, and basic annotation. Extensive support...	Home page Contact: Gunther Schadow
DOCS4DOCS @_	Indiana University School of Medicine	clinical messaging system	Contact: Mike Barnes
RESNET	Indiana University School of Medicine	Patient recruitment	Contact: William Tierney
SPIN	Indiana University School of Medicine	multi-organizational structured clinical data repository with patient and provider linkages	Contact: J. Marc Overhage

Export options:

**CTSA** Clinical & Translational Science Awards

The Clinical and Translational Science Awards (CTSA) is a registered trademark of DHHS.

**Fig. 5.** Web-based Resource Discovery System query tool. In this search screen, the user is able to filter resources by free text search using the text box at the top left of the interface and/or by clicking on a “faceted” value at the left, e.g., a specific institution or specific term from the BRO. Results are displayed under sortable column headings. Applied search criteria are listed as “breadcrumbs” above the results pane. A user may click on the resource name, Home page link, or contact name to navigate to a resource detail page, resource website, or new email message respectively. Buttons at the bottom of the page enable the user to export the results to a number of different file formats.

“Resources that provide access to...” Ontologies such as instrument types or classifications do not fall into these definitions, but have richer relationships such as *has\_component*. Incorporating these richer relationships into the BRO and the BIM will also help to build connections with other research inventory initiatives such as NIF, NITRC, and the eagle-i Consortium.

The final potential enhancement involves extension of the BIM. Through our formal interviews, we found that in addition to searching explicitly by resource type, researchers were interested in the possibility of searching by their area of research or by the type of activity in which they were involved. This researcher would not have known to search for *double-labeled water* technology by name, but might have found that and other useful resources had she been able to search by her areas of research, physiology and metabolism. Our initial idea for how to enable this functionality was to create two additional top level classes in the BRO as siblings of *Resource: Related Area of “Research”*, and “*Related Activities*”. In fact we did create these additional classes, and these are implemented as properties in the current specification of the BIM. However, it soon became apparent that many of the terms that would belong in these branches of the hierarchy already exist in other existing, more mature and previously developed terminologies. Instead of continuing to develop these branches, we plan to investigate some likely candidates for existing terminologies that would serve this purpose, for example the Ontology of Biomedical Investigations (OBI), MeSH or the NCI Thesaurus.

## 7. Discussion and conclusions

With the proliferation of research resources and online biomedical analysis tools, there is a pressing need to catalog

available resources to enable investigators to find the resources they need to carry out their research. In our work, we built upon the Biositemaps and BRO infrastructure to develop RDS, a system that enables institutions to describe and publish structured resource descriptions as well as enabling semantic search and discovery of key tools in biomedical research. This framework can help deliver biomedical resources to the broader research community and reduce redundancy in resource development.

A number of important lessons have been observed empirically through real-world use. For example, real-world deployment has highlighted the importance of training, support, and motivation for participation. Institutions and individual researchers must see direct benefit to them before they are likely to put in the time and effort required to collect and record resource meta-data. While it is easy to see how one benefits from *others* publishing their information, it is harder to see how the home institution can benefit. Investigators do tend to be enthusiastic about the ability to locate resources within their own institution, as these are often not methodically captured, nor documented in any organized way. For this reason, RDS was developed as an open system that can be deployed not only at the national level, but at local institutions as well. Information collected for a local deployment of the system can then be easily repurposed to share more broadly. Future planned functionality is the ability to designate resources as “public” or “private” at an institutional level. This type of continuing on-the-ground observation will continue to inform our system design moving forward.

A key feature of Biositemaps is that each resource owner publishes descriptions of the resources on the Web, annotated using terms from BRO. These resource descriptions are then discoverable by semantic web-based search engines. This decentralized



approach is scalable, and does not require curation of a central database. Our decentralized approach necessitated certain design decisions in order to ensure an intuitive interface and usability by a diverse set of users. Thus far, we have made tradeoffs between complexity and functionality on the one hand, and ease of comprehension and use on the other.

Over the past five months since we started tracking usage, without any formal marketing or publicity for the site, RDS has seen over 1000 visits and over 6000 page views from almost 500 individual users. Feedback to date has been very positive, both through formal usability evaluation and anecdotal evidence, however, the task is far from complete. With the infrastructure now solidly established, significant work items have been identified. Perhaps the most substantial item will be the work required to expand the existing inventory. While a portion of the work may be automated, through Web scraping agents and text mining algorithms, a considerable amount of manual curation will be necessary at each individual site. Even once existing resources have been described, ongoing curation of resources will be needed to ensure current, accurate information.

As the list of resources continues to expand, the BRO will also expand in parallel in order to describe the rapidly evolving landscape of resources for biomedical research. Our motivation for documenting the current state of RDS and the BRO is not to present this system as a finished product, but rather as a thriving and evolving project. We believe that the BRO will continue to evolve, and that there will ultimately be important applications beyond the RDS and Biositemaps efforts. The current work seeks to achieve forward compatibility with other efforts through ongoing discussions with NIF, the Neuroimaging Informatics Tools and Resources Clearinghouse, and the eagle-i Consortium, who are developing rich data models for centralized annotation. BRO development and extension has been, and will continue to be, a transparent and collaborative process. Further, the utility of the BRO depends on input from a broad range of stakeholders, from resource owners to semantic tool developers, to organizations with needs for semantic tools beyond what we can even anticipate at this time. We encourage interested parties to join the designated Google group *Biomedical Resource Ontology (BRO) Discussions* [19].

## Acknowledgments

Numerous BRO classes and definitions have been shared with and borrowed from the Neuroscience Information Framework

(NIFSTD) ontology in an ongoing partnership. Funding for this work was provided by grants 5UL1RR024128-03S1 and 1UL1RR024128-01 (JDT, GSG), UL1RR024146 (KA, DG, AFT), 1UL1RR025008-01 (TM), 1UL1RR024986-01 and 3U54DA021519-04S1 (BK, BA, KS), 1UL1RR024153-01 and 3UL1RR024153-03S1 (NW, MB, CB, HS), 3UL1 RR024153-03S1 (PW, CN, NN, DR, MM, NW, MB, CB, HS, BK, BA, KS), U54 RR021813 (IDD), U54DA021519 (BDA, BK), 3U54HG004028-04S1 and 5U54HG004028-05 (PW, CN, NN, DR, MM).

## References

- [1] Biositemaps White Paper. 2008; Available from: [http://biositemaps.ncbcs.org/Biositemaps\\_white\\_paper\\_v4.1.pdf](http://biositemaps.ncbcs.org/Biositemaps_white_paper_v4.1.pdf).
- [2] Biomedical Resource Ontology. Available from: <http://purl.bioontology.org/ontology/BRO-Core>.
- [3] Resource Discovery System (RDS) query tool. Available from: <http://biositemaps.org/rds>.
- [4] Morris TT, Borromeo CD, Kirschner B, Singh H, Tenenbaum JD, Whelan NB, et al. CTSA Inventory Resource Web Presence (CIRWP). 2010 AMIA Summit on Translational Bioinformatics; San Francisco, CA2010.
- [5] Biositemaps System. Available from: <http://www.biositemaps.org>.
- [6] Neuroscience Information Framework. Available from: <https://confluence.crbs.ucsd.edu/display/NIF/NIF+Ontologies+and+Terminologies>.
- [7] The Neuroimaging Informatics Tools and Resources Clearinghouse. Available from: <http://www.nitrc.org>.
- [8] Eagle-I Project. Available from: <http://www.ohsu.edu/xd/education/library/about/departments/ohsu-library-eagle-1.cfm>.
- [9] Dinov ID, Rubin D, Lorensen W, Dugan J, Ma J, Murphy S, et al. iTools: a framework for classification, categorization and integration of computational biology resources. *PLoS One* 2008;3(5):e2265.
- [10] Sitemaps.org: consortium of Microsoft, Google, and Yahoo to improve web searching Available from: <http://www.sitemaps.org/>.
- [11] Biositemaps Registry. Available from: <http://biositemaps.ncbcs.org/biositemap.registry>.
- [12] Smith B. The Logic of Biological Classification and the Foundations of Biomedical Ontology. 10th International Conference in Logic Methodology and Philosophy of Science. Oviedo, Spain: Elsevier; 2003.
- [13] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet.* 2000;25(1):25–9.
- [14] Rosse C, Mejino Jr JL. A reference ontology for biomedical informatics: the foundational model of anatomy. *J Biomed Inform* 2003 Dec;36(6):478–500.
- [15] SKOS Simple Knowledge Organization System Reference W3C Recommendation Available from: <http://www.w3.org/TR/2009/REC-skos-reference-20090818>.
- [16] Biomedical Resource Ontology (BRO) Discussions. Available from: <http://groups.google.com/group/bro-discuss>.
- [17] Biositemaps Editor. Available from: <http://biositemaps.bioontology.org/editor/?conf=ctsa>.
- [18] Mulgara: a scalable open source RDF database. [database on the Internet]. Available from: <http://www.mulgara.org/>.
- [19] The Ontology of Biomedical Investigations. Available from: [http://obi-ontology.org/page/Main\\_Page](http://obi-ontology.org/page/Main_Page).