

DOORS to the Semantic Web and Grid With a PORTAL for Biomedical Computing

Carl Taswell, *Member, IEEE*

Abstract—The semantic web remains in the early stages of development. It has not yet achieved the goals envisioned by its founders as a pervasive web of distributed knowledge and intelligence. Success will be attained when a dynamic synergism can be created between people and a sufficient number of infrastructure systems and tools for the semantic web in analogy with those for the original web. The domain name system (DNS), web browsers, and the benefits of publishing web pages motivated many people to register domain names and publish web sites on the original web. An analogous resource label system, semantic search applications, and the benefits of collaborative semantic networks will motivate people to register resource labels and publish resource descriptions on the semantic web. The Domain Ontology Oriented Resource System (DOORS) and Problem Oriented Registry of Tags and Labels (PORTAL) are proposed as infrastructure systems for resource metadata within a paradigm that can serve as a bridge between the original web and the semantic web. Registers domain names while DNS publishes domain addresses with mapping of names to addresses for the original web. Analogously, PORTAL registers resource labels and tags while DOORS publishes resource locations and descriptions with mapping of labels to locations for the semantic web. BioPORT is proposed as a prototype PORTAL registry specific for the problem domain of biomedical computing.

Index Terms—Biomedical computing, BioPORT, cross-directory search, DNS, DOORS, IRIS, OWL, PORTAL, RDF, semantic web, XML.

I. INTRODUCTION

DEVISING more effective technologies and productive systems to accelerate the growth of the semantic web and grid remains a fundamental challenge for Internet engineers. In response to this challenge, this paper reports novel technologies called the Domain Ontology Oriented Resource System (DOORS) and the Problem Oriented Registry of Tags and Labels (PORTAL) intended for use with resource metadata. DOORS and PORTAL have been designed within a novel paradigm focused on *labeled resources* in analogy with existing Internet systems focused on *named domains*. This report further elaborates a prototype registry called BioPORT that is specific for the problem domain of biomedical computing. For cross-registry compatibility, problem-domain-specific registries such as BioPORT are designed to comply with the requirements of the generic root registry within the PORTAL system. These

Manuscript received October 31, 2006; revised June 11, 2007.

The author is with the University of California, San Francisco, CA 94143 USA. He is also with the Global TeleGenetics, Inc., Palo Alto, CA 94303 USA (e-mail: ctaswell@telegenetics.com).

Color version of one or more of the figures in the paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TITB.2007.905861

registries are proposed with scientific problem-oriented designs that avoid the engineering-technology-oriented restrictions of existing registries.

Sections II–IV review the background and motivation for DOORS, PORTAL, and BioPORT. Section II explains key concepts of the current semantic web and grid, and summarizes how they are driving the transformation of software architecture from designs based on closed-world computing to those based on open-world computing. Section III reviews the literature and current state-of-the-art in the life sciences web and grid, and summarizes the opinions of leading commentators in the bioinformatics community on existing barriers that impede development. Section IV defines the meaning and scope of biomedical computing as interpreted in this paper for BioPORT, and provides further motivation justifying the need for a new kind of metadata registry in biomedical computing. Sections V and VI review existing technologies, respectively, for domain naming and registering systems [including the domain name system (DNS), Internet Registry Information Service (IRIS), etc.] and for resource identifying and linking systems [including Uniform Resource Identifier (URI), Persistent Uniform Resource Locator (PURL), etc.] that serve as inspirations and/or foundations for DOORS and PORTAL.

Sections VII and VIII present the central novel contribution of this paper. Section VII provides a detailed exposition of the design principles and requirements necessary for both DOORS and PORTAL server functions and data records to operate as an effective infrastructure for registering resource labels and tags and publishing resource locations and descriptions intended for use by other semantic systems and applications. Similarly, Section VIII provides a description of the design principles and requirements for BioPORT as a registry for biomedical computing within the PORTAL-DOORS framework.

Sections IX–XIII provide further analysis and discussion of issues essential to DOORS and PORTAL. Section IX clarifies distinctions between the resource labels used in DOORS and PORTAL and the domain names used in DNS and IRIS. Section X discusses the importance of synergistic systems comprising synergies created not only among technology components but also between technologies and people necessary for the growth of the semantic web. Section XI discusses the importance of semantic search applications including their expected use within translational medicine. Section XII summarizes DOORS and PORTAL describing it as a hybrid with which to bootstrap and bridge from the original web to the semantic web. Section XIII summarizes some of the key advantages of DOORS and PORTAL in comparison with other systems and concludes with some remarks on future work.

II. SEMANTIC AND OPEN-WORLD COMPUTING

Recognized as the inventor of the World Wide web, and now the director of the World Wide web Consortium (W3C) [1], Berners-Lee has refocused his attention on development of the semantic web [2] and creation of a science of the web [3]. The semantic web extends the original web with technologies that provide syntactic structure [the extensible markup language (XML) [4]] and semantic meaning [the resource description framework (RDF) [5]] permitting the development of taxonomies and inference rules. When combined together as description logics languages [the DL variant and recent E and Eu extensions [6] of the web ontology language (OWL) [7]], they enable the compilation of knowledge representations or information collections known as ontologies [8], [9]. Several recent books [10]–[12] provide a comprehensive introduction to this rapidly changing field of semantic computing.

Regarding information, Berners-Lee *et al.* [3] observe that most data remain inaccessible (either hidden or locked in closed storage systems without communicating interfaces) rather than distributed via an open network of inter-referring resources. Regarding people, they note that scientists depend increasingly on the web but do not interact sufficiently with web technologists in a manner that would enable the engineers to build systems more suitable for use by the scientists. As a consequence, Berners-Lee *et al.* [3] conclude that accelerating the growth of the semantic web requires the development and support of a new interdisciplinary field called *web science*. They emphasize that this new field involves engineering novel infrastructure protocols and systems, developing more productive applications and user interfaces, and understanding the communities that use them.

A fundamental tenet underlying the web remains the *open-world* assumption that the computing environment is intrinsically open and continuously changing. However, traditional software development was based on a *closed-world* assumption. Baresi *et al.* [13] discuss the evolution of software architectures from being “static, monolithic, and centralized” in the closed-world setting to “dynamic, modular, and distributed” in the open-world setting. They provide an excellent summary of open-world computing with a review of existing solutions (including web services, publish/subscribe middleware, grid computing, and autonomic computing) and an outline for a research agenda (addressing specification, verification, monitoring, trust, implementation, and self-management). Zhuge [14], [15] provides another view of open-world computing with attention to its future interconnection environment, semantic grid, and e-science knowledge grid.

III. LIFE SCIENCES WEB AND GRID

Biomedical ontologies have benefited from significant development in the bioinformatics and clinical informatics communities [16]–[18]. In bioinformatics, the journal *Nucleic Acids Research* features an annual web Server Issue and an annual Database Issue. Recent articles include those on Euro-

pean Bioinformatics Institute’s (EBI’s) resources [19], the National Center for Biotechnology Information’s (NCBI’s) resources [20], the molecular biology database collection [21], the bioinformatics links directory [22], and the Online Bioinformatics Resources Collection (OBRC) [23]. Philippi and Köhler [24], [25] discuss the many problems impeding the semantic integration of these life science databases and ontologies. Asking the question, “A life science semantic web: Are we there yet?,” Neumann [26] and Neumann and Quan [27] provide another perspective featuring the Life Science Identifier (LSID) proposed standard [28], the Haystack semantic browser [29], and other initiatives such as the W3C Semantic web Health Care and Life Sciences Interest Group [30]. These influences have shaped the development of his prototype semantic web application BioDash [27] for drug discovery in pharmacogenomics and personalized medicine.

Cannata *et al.* [31] call for the organization of the “bioinformatics resourceome” arguing that investigators should have a comprehensive directory of algorithms, databases, and literature with sufficient annotation to facilitate appropriate use of the listed resources. They recommend the development of a distributed system for describing the availability and reliability of these resources. They envision a resource metadata system that would answer questions regarding the current location and availability of a resource and its quality as measured by objective benchmarks or subjective ratings.

Extending beyond bioinformatics to the wider expanse of all biomedical research, Buetow [32] reviews examples from the developing biogrid including myGrid [33], Biomedical Informatics Research Network (BIRN) [34], and the Cancer Biomedical Informatics Grid (caBIG) with caCORE [35]. He observes that biomedical informatics remains heterogeneous and serves disconnected medical, scientific, and engineering communities. He further explains that these communities speak different languages resulting in communication barriers that slow the cross-disciplinary transfer of knowledge. Considering existing technology alternatives including peer-to-peer systems, web services, and grid computing, he concludes that current efforts “have not yet crossed the threshold of demonstrated value.”

Buetow [32] recommends that a cyberinfrastructure of the future should: 1) transition smoothly from the current to future infrastructures; 2) adhere to open standards that promote platform agnosticism (i.e., neutrality and independence); 3) manage identity and control access; and 4) track data provenance, intellectual property, and academic credit. Most importantly, he admonishes against building a new infrastructure that simply replaces current silos with future cybersilos.

With another view in the larger context of e-science and e-business, DeRoure and Hendler [36] and DeRoure *et al.* [37] discuss their vision of the future infrastructure for the semantic web and grid. They provide a detailed review of requirements from “resource description, discovery, and use” to “integration with legacy IT systems” applicable in general. However, they also discuss several important case studies relevant to health care and the life sciences such as combinatorial chemistry, medical imaging, and medical devices [37].

IV. BIOMEDICAL COMPUTING

Bioinformatics and computational biology, biomathematics and mathematical physiology, or biostatistics and epidemiology are examples of pairs of related fields that have distinguishing definitions carefully crafted by the specialists in each of these related fields. In contrast, biomedical computing is defined here for the purposes of declaring the scope of BioPORT in Section VIII as the most general and comprehensive term referring to any multidisciplinary field that combines aspects of both the computational and life sciences.

Biomedical computing applies tools and methods from the computational world to answer questions in the biomedical world whether to discover and understand the nature of life or to promote health and prevent disease. In the sense of biomimicry [38], biomedical computing builds models of the animate world as a means of engineering systems in the inanimate world intended to emulate the efficiencies of nature created by evolution. Encompassing many alternative perspectives, a generalized definition of biomedical computing must incorporate all theoretical, computational, and experimental scientific and engineering approaches to the fusion of computers and computing with biology and medicine.

This generalized view extends analogously to computing itself defined here as execution by a machine of a program comprising algorithms operating on data *without* regard to type of data (numeric, symbolic, multimedia, etc.), class of algorithm (numerical simulation, database query, logical reasoning, computational complexity, etc.), machine (calculator, workstation, grid, etc.), platform (processor, operating system, programming language), implementation (hardware, firmware, software), director (human, software agent, other machine), or underlying theory (whether from mathematics, statistics, informatics, etc.). According to this multiperspectived view of computing, a resource can be anything from a simple utility that runs on a calculator isolated from the Internet to a sophisticated application that only runs on a distributed grid of supercomputers or massively parallel processing nodes.

Yet, as noted by Cannata *et al.* [31], scientists cannot necessarily find appropriate available resources even in their own fields of specialization. Moreover, in the life sciences as reviewed in Section III, most resource directories remain technocentric in the sense that each tends to collect information about resources of only one kind such as *database* or *web server* rather than all kinds of computing resources (including those not dependent on the Internet for operation) that might be relevant to the scientist's field of inquiry.

In order to prevent the replacement of current silos with future cybersilos forewarned by Buetow [32], a system of registries and directories for resources should be built in a manner analogous to that for the DNS [39]–[41] constructed for domain names. Thus, it would be unrestricted by either computing resource or application field just as DNS was unrestricted (see Section V). If successful, then a neuropharmacologist should be able to search a biochemistry or bioinformatics directory just as readily as a cardiologist might search an electrophysiology or cardiovascular drug trials directory. Each specialist should

be able to conduct cross-directory searches in related fields and find any relevant resource of interest whether a simple spreadsheet macro or an ontology-based expert system, regardless of location of the directory or registry governing the data record found for the resource metadata.

V. DOMAIN NAMING AND REGISTERING SYSTEMS

Purposeful avoidance of any requirement for the client or user to possess prior knowledge of a domain name's governing registry or authoritative directory (with the latter better known in DNS as a primary name server) has significantly contributed to the overwhelming success of DNS. With antecedents appearing as early as 1983 and the Internet Engineering Task Force (IETF) Request for Comments (RFC) 1035 approved as IETF Standard 13 in 1987 [39], DNS remains one of the most important pillars supporting the infrastructure of the Internet and the growth of many protocols (*telnet*, *gopher*, *ftp*, etc.) of which the most user-friendly and influential has been *http* spurring the growth of the web [40].

In simplest terms, DNS maps domain names (registered separately at a governing registry managed by a registrar) to numeric addresses identifying Internet locations. DNS operates with a system of root servers, authoritative primary servers, and nonauthoritative secondary servers known as name servers that, when accessed by clients known as resolvers, interact with recursive forwarding, caching, and “time-to-live” expiring, respectively, for querying, storing, and expunging record data. DNS has been further enhanced with support for security with DNS extensions implemented as the Domain Name System Security Extensions (DNSSEC) [42], [43] and for multilingualism with internationalized domain names implemented as the Internationalizing Domain Names in Applications (IDNA) and the Internationalizing Domain Names (IDN) standards [44], [45]. Despite the recent popularity of peer-to-peer technologies, it is difficult to imagine how a peer-to-peer-based alternative could reproduce the success of DNS and its enhancements DNSSEC and IDNA *without* its associated hierarchy of clients, caching servers, authoritative servers, and governing registries.

Originally motivated by the desire to build a replacement for the aging *whois* protocol [46], the IETF Cross Registry Information Service Protocol (CRISP) Working Group has been chartered [47] to

define a standard mechanism that can be used for . . . finding authoritative information associated with a label [and] a protocol to transport queries and responses for accessing that information . . . [which] provides uniform access to and view of data that may be held in disparate backend servers . . .

for registries [48]. The CRISP Working Group has already completed the initial draft *iris1* [49] of the IRIS Core Protocol, and drafts *dreg1* [50] and *areg1* [51] of several IRIS-dependent protocols for different types of registries. If approved, a pending update to IRIS called IRIS-XPC [52] will replace IRIS blocks extensible exchange protocol (BEEP) [53] by specifying XML pipelining with chunks (XPC) as the new default transport for IRIS and by providing full support for security and international languages. IETF's CRISP [48] should not be confused with

TABLE I
DISTRIBUTED HIERARCHICAL DATABASE SYSTEMS WITH ENTITY-ATTRIBUTE REGISTERING AND PUBLISHING

	IRIS-DNS	PORTAL-DOORS
Registering system	IRIS	PORTAL
— Entity registered	domain	resource
— Identified by	unique name	unique label with optional tags
Publishing system	DNS	DOORS
— Attributes published	address and aliases	location and description
— Specified by	IP number	URIs, URLs, ontologies, and semantic statements
Forwards requests	Yes	Yes
Caches responses	Yes	Yes
Serves original web	Yes via mapping of character name to numeric address	Yes via mapping of character label to URLs for IRIS-DNS
Serves semantic web	No	Yes via mapping of character label to semantic description
Crosslinks entities	No	Yes via mapping within description to other resources

the National Institute of Health's (NIH's) Computer Retrieval of Information on Scientific Projects [54]. Similarly, IETF's IRIS [49] should not be confused with the Interoperability and Reusability of Internet Services [55] or the International Rice Information System [56].

VI. RESOURCE IDENTIFYING AND LINKING SYSTEMS

As the core protocol for CRISP, IRIS has been designed to associate authoritative information with any arbitrary kind of label as declared and defined by the particular registry type [50], [51]. Theoretically, a label may be anything from a simple tokenized name to a more complex URI [57] or Internationalized Resource Identifier (IRI) [58]. These identifiers may specify either abstract or physical resources, neither of which are required to be accessible via the Internet. A URI that is resolvable to an Internet location is commonly known as a Uniform Resource Locator (URL) [59]. However, even a URI serving as an XML namespace identifier with the form `http://www.domain.org/namespace/` that appears as if it might also be a URL, and thus might resolve to a web site, is not required to do so.

Such namespace URIs are often associated with Internet-accessible web site directories that contain a collection of related resources supporting the namespace. The Resource Directory Description Language (RDDL) [60], built as an extension of extensible hypertext markup language (XHTML) [61] and XLink [62] with an added element `resource`, has been developed to provide both human- and machine-readable information describing the nature, purpose, and location of each resource in the directory with links to the resources targeted by the namespace URI.

RDDL with XLink may be related to the semantic web but does not constitute one of its inherent components; see [63] for a discussion of conversion from XLink-based resources to RDF-based resources more appropriate for the semantic web. The problem of interlinking and cross-linking resources has also been addressed by an independent consortium on XML Topic Maps (XTM) with its XML XTM specification [64].

A solution to another problem, that of persistent versus transient links, has been provided by the Online Computer Library Center (OCLC) with its PURL System (www.purl.org) for Persistent URLs [65]. The PURL system remains nonproprietary and available for use without fees. In contrast, the Handle System (www.handle.net) has been patented by the Corporation

for National Research Initiatives (CNRI), and does require registration and annual service fees. However, the Handle System [66], [67] provides a higher level of security than does the PURL System.

None of these linking systems (whether RDDL, XTM, PURL, or Handle) have been built with RDF and OWL enabling machine-understandable semantic relationships between linked resources. However, continuing refinements of RDFS and the mapping between RDF and OWL [68] strengthen RDF and OWL as the *de facto* languages of the semantic web. Therefore, none of the linking systems reviewed in this section can currently serve as infrastructure components immediately and directly suitable for the semantic web without first being revised and rebuilt with RDF/OWL and then appropriately embedded in semantic systems.

VII. PORTAL AND DOORS

As a protocol to facilitate interoperability of registries and registrars, CRISP with its core IRIS (in its current draft form with core protocol `iris1` [49], [52] and main registry type `dreg2` [69]) has been built primarily for the original web with a focus on the domain names of DNS. Extensions of IRIS and analogs of DNS can also be developed for the semantic web and grid with a focus on *labeled resources* instead of *named domains*. Thus, basic principles and requirements for data records and server functions are proposed here for a new infrastructure technology as an extension and analog of the existing IRIS-DNS framework. In this novel paradigm, the PORTAL operates as a resource label and tag registering system (i.e., IRIS extension) and the DOORS operates as a resource location and description publishing system (i.e., DNS analog).

Both the IRIS-DNS and PORTAL-DOORS frameworks can be viewed as analogous paradigms serving, respectively, the original web and the semantic web. Table I compares some of the similarities and differences of these paradigms from the perspective of considering both as distributed hierarchical database systems with entity-attribute registering and publishing. Detailed requirements of the PORTAL-DOORS paradigm are elaborated further in Sections VII-A to VII-E.

A. DOORS Data Records

Fig. 1 displays a diagram summarizing the basic structure of a DOORS data record with both required and permitted

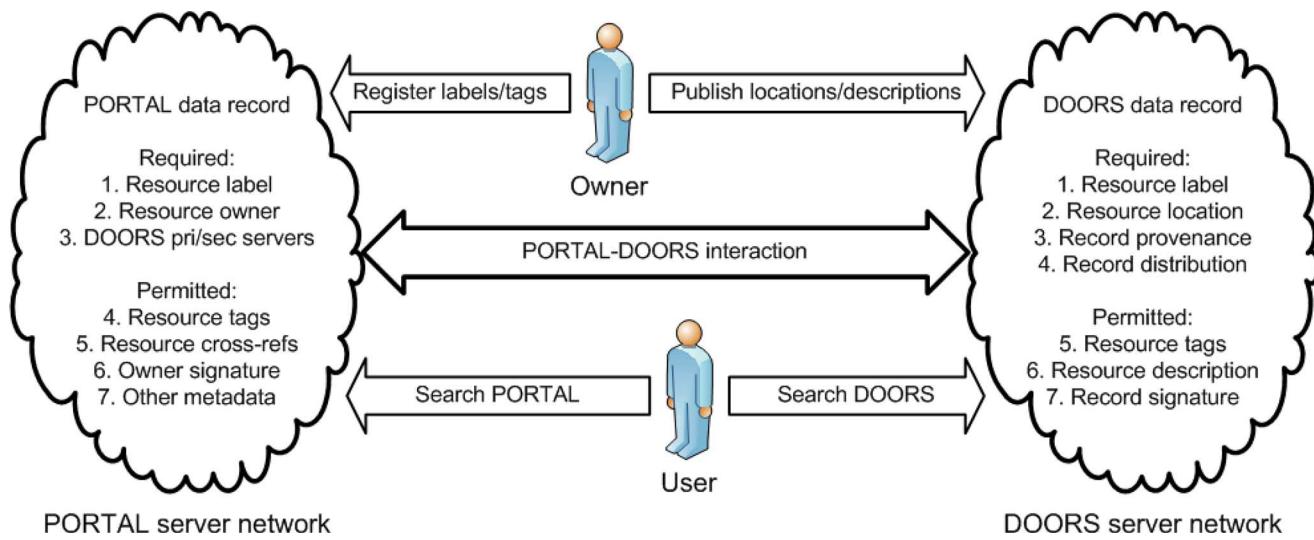


Fig. 1. Resource metadata registered and published by owners for search by users in the PORTAL-DOORS server networks. Fields within data records are considered *required* or *permitted* with respect to the schemas maintained by the root servers (see Fig. 2).

fields. Minimizing requirements remains imperative during the transition from original web to semantic web. Thus, resource label servers (as the analogs in DOORS of the domain name servers in DNS) should maintain database records with the following *required* metadata for each resource:

- 1) the *resource label* with a globally unique URI (or IRI) enabling nonsemantic string queries of labels;
- 2) the *resource location* with a URL (or IDN), possibly the same as the URI (or IRI) if resolvable, and any associated URLs (or IDNs) as explained in Section VII-C, enabling query responses;
- 3) the *record provenance* with identification of the: a) resource owner; b) authoritative master PORTAL registry; and c) authoritative primary DOORS server;
- 4) the *record distribution* with parameters for “time-to-live” caching and expiring as well as extent of redistribution for nonauthoritative secondary DOORS servers.

Given the operational features of both DOORS and PORTAL described, respectively, in Sections VII-C and VII-D, resource label servers should also maintain records with the following *permitted* metadata for each resource:

- 5) the *resource tags*, if registered at the governing registry, including a tokenized name and/or phrases enabling nonsemantic string queries of tags;
- 6) the *resource description* with an RDF mini-document, a collection of RDF triples that reference OWL ontologies, enabling semantic reasoning queries of descriptions;
- 7) the *record signature* with XML-Signatures [70] for the: a) resource owner; b) authoritative master PORTAL registry; and c) authoritative primary DOORS server.

As an informal demonstrative example, consider the following DOORS pseudorecord for a software application:

- 1) *resource label*: “<http://biomedicalcomputing.org/elida>”;
- 2) *resource location*: “<http://www.ellitron.com>”;

- 3) *record provenance*: a) resource owner: “Carl Taswell”; b) PORTAL master: “portal.biomedicalcomputing.org”; c) DOORS primary: “doors.biomedicalcomputing.org”;
- 4) *record distribution*: a) expiration time-to-live: “7 days”; b) redistribution extent: “all servers”;
- 5) *resource tags* (nonsemantic strings): “ELIDA”; “limiting dilution assays”; “biologically active particles”;
- 6) *resource description* (semantic statements): “ELIDA is downloadable freeware”; “ELIDA runs on workstations”; “ELIDA implements algorithms published in [71]”; “ELIDA analyzes limiting dilution assay data”; “ELIDA quantitates biologically active particles.”

This informal pseudorecord example contains the required unique label, three optional tags (of which the first is a tokenized name), and five semantic statements in the description. The label and tags can be searched with a nonsemantic string query while the description can be searched with a semantic reasoning query. A formal version of this record would be found at a DOORS server by a semantic search for “free software that analyzes limiting dilution assay data” initiated by a biologist at a DOORS client. For the implementation of the formal DOORS record as a valid XML document containing within itself a valid RDF mini-document for the semantic description, the five statements in this example should be expressed as RDF triples referencing OWL ontologies.

By requiring a DOORS record to reference its governing PORTAL registry, the DOORS server can access the schemas enforced for the record’s XML document and its RDF mini-document. Whenever resource metadata are stored or updated by the owner in records at the DOORS server, the metadata should always be validated for compliance with any schema imposed by the registry type of the governing registry. This design enables any DOORS server to maintain resource records governed by different PORTAL registries of varying specific PORTAL registry types all of which must comply with the generic PORTAL registry type (see Section VII-D). Usage patterns will

determine which servers accumulate records governed by which specific registry types.

B. PORTAL Data Records

Fig. 1 also displays the basic structure of a PORTAL data record with required and permitted fields. This structure is designed with the same principle of minimizing requirements as used in Section VII-A for DOORS data records. Thus, resource label registries (as the analogs in PORTAL of the domain name registries in IRIS) should maintain database records with the following *required* metadata for each resource:

- 1) the *resource label* with a globally unique URI (or IRI) required by the generic PORTAL registry type for identification of the resource in PORTAL-DOORS;
- 2) the *resource owner* with contact information for the personnel who own and manage the resource;
- 3) the *DOORS servers* with URLs (or IDNs) for the primary and secondary DOORS servers that publish the metadata not maintained at the PORTAL registry.

Given the operational features of both DOORS and PORTAL described, respectively, in Sections VII-C and VII-D, resource label registries should maintain records with the following *permitted* metadata for each resource:

- 4) the *resource tags* with character strings permitted by the policies of the specific PORTAL registry type;
- 5) the *resource cross-references* with any globally unique identifiers permitted by the policies of the specific PORTAL registry type for identification of the resource in other systems unrelated to PORTAL-DOORS;
- 6) the *owner signature* with the XML-Signature of the owner permitted by the generic PORTAL registry type;
- 7) any *other metadata* permitted by the policies of the specific PORTAL registry type.

Metadata items listed in Sections VII-A and VII-B are considered *required* or *permitted* with respect to the generic PORTAL registry type, not with respect to a semantic domain-specific PORTAL registry type (see Fig. 2). Thus, the schema imposed by the PORTAL root server (for the generic type) is least restrictive while a schema imposed by a PORTAL master server (for a specific type) may be more restrictive. An item considered *permitted* with respect to the generic PORTAL registry type may be considered *required* with respect to a specific PORTAL registry type if declared by its policies. Distinct registry types serving different semantic domains of inquiry may have very different policies regarding the manner in which unique labels and optional tags are created for each resource when registered.

For example, a specific registry type could allow each resource to be registered with a number of optional tags consisting of a single *principal tag* and multiple *supporting tags*. Registrants could then select a number of available tokenized names and phrases for the assignment to the resource being registered. In such a scenario (see Section III), one of the tokenized names should be noncolonized and designated as the principal tag for concatenation with a URI or an IRI namespace controlled by either the resource owner or by the registry type. This approach

would facilitate a policy in which both a locally unique resource tag (the principal tag) and a globally unique resource label (URI or IRI concatenated from namespace and tag) are guaranteed for each registered resource. Thus, even if the unique label for use by machines is long or complicated, multiple synonymous and simple tags are made available for use by humans who might not wish to remember or type the complex labels.

C. DOORS Server Functions

Just as DNS permits domain name owners to create and update records at name servers with the addresses for their domains, DOORS should permit resource label owners to maintain records with the locations of their resources. Just as DNS operates with a hierarchical system of forwarding and caching servers (see Section V) to map domain names to numeric addresses, DOORS should map resource labels to Internet locations with the following additional features.

- 1) *Map label to location*: Perform a lookup for a resource labeled uniquely by URI (or IRI) and return the associated URLs (or IDNs) required to be resolvable Internet locations for: a) the primary site and any mirror sites for the resource itself (a mapping via the associated URLs from the URI label to the resource itself); b) the URI (or IRI) namespace directory containing associated metadata maintained by the resource owner with descriptions in RDDL (a more indirect mapping via the associated URLs from the URI label to the metadata at the namespace directory linking to the resource); or c) the contact information maintained by the governing PORTAL registry if neither the resource itself nor its URI namespace is maintained online by the resource owner (the most indirect mapping via the associated URL from the URI label to the metadata at the registry enabling contact with the owner of the offline resource).
- 2) *Map tag to location*: Perform a lookup for a resource labeled uniquely by tag and return the associated URI and URLs subject to the constraint restricting the lookup to those resources governed by PORTAL registries of the same specific registry type with a policy that imposes uniqueness of a principal tag (see Section VII-B).
- 3) *Search nonsemantic strings in labels or tags*: Find resources by string query of character substrings in labels or tags and return the associated URIs and URLs recognizing that the search may yield nonunique results when performed across resources governed by registries of different registry types or of the same registry type without a policy imposing at least one unique tag.¹
- 4) *Search semantic statements in descriptions*: Find resources by semantic query with SPARQL [72] of semantic statements in descriptions and return the associated URIs and URLs recognizing that the search may yield unranked nonunique results.

¹All resources are required to be labeled with a globally unique URI or IRI. This label may be a URI containing *number* characters to identify distinct resources. Resources are not required to be tagged. Any such optional tags containing *letter* characters are not required to be unique.

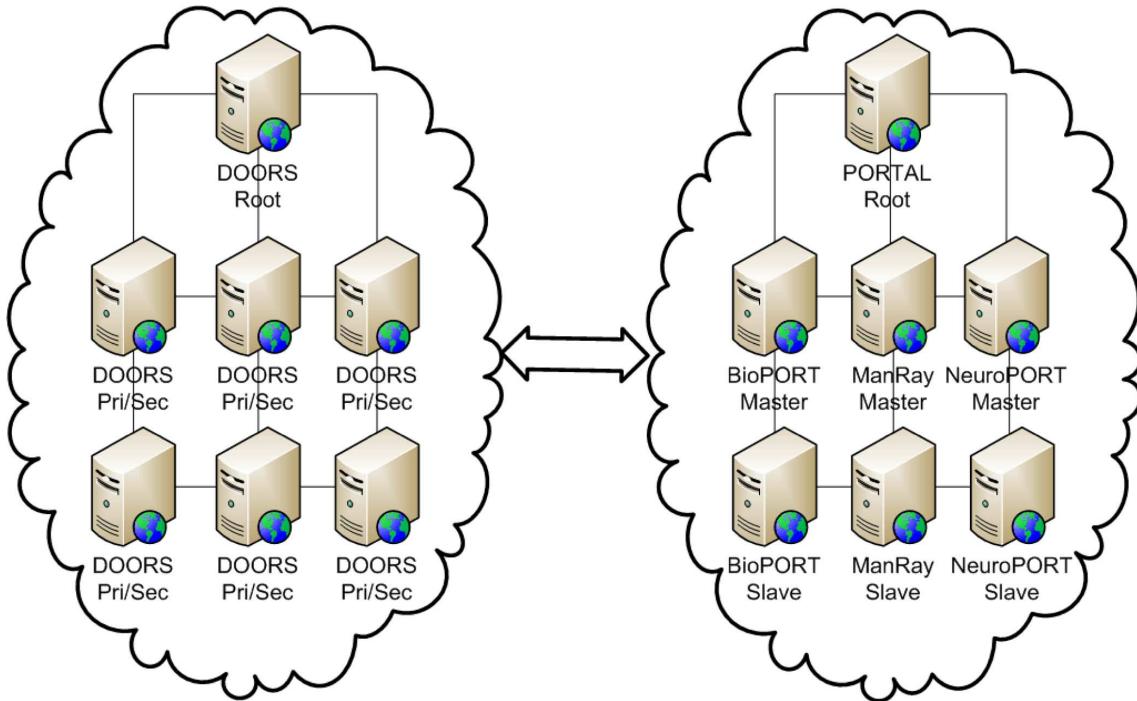


Fig. 2. Resource metadata server networks for PORTAL registering of labels and tags and DOORS publishing of locations and descriptions (see Fig. 1); analogous to domain metadata server networks for IRIS registering of names and DNS publishing of addresses. Both the PORTAL and DOORS server networks contain root, authoritative, and nonauthoritative servers all of which interact with each other. Authoritative and nonauthoritative servers for DOORS are called, respectively, *primary* and *secondary*, whereas those for PORTAL are called *master* and *slave*. The same DOORS server may operate as primary for some records while simultaneously as secondary for other records. Any problem-domain-specific PORTAL registry type [enforced by each of the BioPORT (see Section VIII), ManRay [73], and NeuroPORT master servers for the example registry types in the figure] must also adhere to the requirements imposed by the generic PORTAL registry type (enforced by the PORTAL root server) to maintain compatibility with cross-registry searches.

- 5) *Provide identification and authentication:* Include the provenance and signature of each resource record returned in the response to the lookup or query request.

Just as the network of DNS directories depends on a separate but related system of IRIS registries, DOORS depends on PORTAL. Both DOORS and DNS are directories, not registries. A DOORS search serves a fundamentally different purpose than a PORTAL search (see Section VII-D). Fig. 2 displays a diagram representing the PORTAL-DOORS distributed hierarchical database system with the PORTAL and DOORS networks of root, authoritative, and nonauthoritative servers all interacting with each other.

D. PORTAL Server Functions

Just as IRIS registries² publish the primary and secondary DNS servers for each registered domain name, PORTAL registries should publish the primary and secondary DOORS servers for each registered resource label with the following additional features.

- 1) *Comply with generic root schema:* Adhere to the schema required by root servers of the generic PORTAL registry type governing the interaction between servers of different specific PORTAL registry types.

- 2) *Comply with specific master schema:* Adhere to the schema required by master servers of the same specific PORTAL registry type governing the interaction between PORTAL and DOORS for the semantic domain of inquiry (i.e., the problem domain or specialty area) determined by declarations of the: a) ontologies controlling semantic statements in and queries of the resource description; b) policies establishing any additional requirements or options for the resource label, tags, and locations; c) policies establishing any additional requirements or options for the record provenance, distribution, and signature for the metadata maintained collectively at PORTAL registries and DOORS servers.

- 3) *Recommend related master PORTAL servers:* Provide a list of recommended PORTAL master servers of different specific registry types to facilitate cross-registry searches in related specialty areas.
- 4) *Recommend related primary DOORS servers:* Provide a list of recommended DOORS primary servers to facilitate recursive forwarding between DOORS servers for the set of recommended PORTAL master servers.
- 5) *Publish resource DOORS servers:* Perform a lookup of a registered resource by label or tag and return the assigned primary and secondary DOORS servers for the associated metadata record.
- 6) *Publish resource cross-references:* Perform a lookup of a registered resource by label or tag and return any

²See [48] for a discussion of the differences between “thick” and “thin” models for registrar/registry systems in which the registrar accepts registrations from registrants on behalf of the registries.

TABLE II
PORTAL-DOORS SERVER FUNCTIONS FOR RESOURCE OWNERS AND USERS

	PORTAL	DOORS
System	Enforce generic root and specific master schemas Recommend list of related master servers	Access PORTAL generic root and specific master schemas Redistribute list of related master PORTAL servers
Owner	Recommend list of related primary DOORS servers Create and maintain owner account at master server Create and maintain resource records at master server Register label and tags for each resource record Publish cross-references for each resource record	Redistribute list of related primary servers Create and maintain owner account at primary server Create and maintain resource records at primary server Publish location and description for each resource record Publish provenance and distribution for each resource record
User	Map label to primary and secondary DOORS servers Map label or tags to cross-references Map cross-reference to label and tags Search non-semantic strings in labels or tags Obtain owner information and other metadata	Map label to location Map tag to location Search semantic statements in descriptions Search non-semantic strings in labels or tags Obtain record identification and authentication

cross-references identifying the resource in other systems unrelated to PORTAL-DOORS.

- 7) *Publish resource owner and other metadata:* Perform other standard requests of registrar/registry systems [69] such as a lookup that returns all resources registered by an owner and a lookup that returns the contact information for the resource owner, managing personnel, or any other associated metadata.

Table II summarizes server functions for both PORTAL and DOORS servers from the perspective of each server network system as well as both the resource owner and the user.

E. Implementation of PORTAL and DOORS

PORTAL and DOORS could each be implemented as web services over `http`. However, doing so requires faith in a “one-size-fits-all” mantra currently promoted by some advocates of web services. Moreover, it precludes the possibility of using and optimizing a network infrastructure communications protocol for intended and related purposes rather than all possible purposes. Consequently, it would be better to consider the existing DNS and CRISP protocols for a primary implementation of DOORS and PORTAL after which an additional web service interface could be implemented.

Thus, DOORS could be implemented as an extension of either DNS or CRISP protocols since both have mechanisms enabling extensions. However, PORTAL should be implemented as an extension of the CRISP protocol because it lies so naturally within the scope of the stated goals for CRISP. Further, CRISP as an implementation framework for both PORTAL and DOORS functionalities would enable better interoperability of each with the other. Using the same framework for both functionalities would also more readily facilitate the development of a server suite that could be configured for deployment on a machine as both PORTAL and DOORS together or as either PORTAL or DOORS alone. Analogously, a client suite or an integrated client could also be developed capable of querying either DOORS or PORTAL servers.

If specifications for the DOORS and PORTAL systems are implemented as extensions of the CRISP framework, then they should be derived as XML schemas that depend upon the CRISP protocols. The schema DOORS1 for DOORS should inherit from the IRIS core protocol `iris1` with extensions to maintain

compliance with the requirements for DOORS data records (Section VII-A) and server functions (Section VII-C) while making it more suitable for use with semantic web applications. For example, the generic `bagType` and `bagsType` used in IRIS request and response transactions [49] must be modified to define an additional `rdfBagType` for bags with RDF content. The schema PORTAL1 for the generic PORTAL registry type should be implemented as an “RDF-ized” analog of the schema `dreg2` (for domain registry [69]) with modifications to maintain compliance with the requirements for PORTAL data records (Section VII-B) and server functions (Section VII-D). Individual schemas for specific registry types must inherit from the schema PORTAL1. Each could be named arbitrarily (e.g., `ManRay`) or in a manner reflecting its specialty area (e.g., BioPORT, GenePORT, NeuroPORT, CardioPORT, GeoPORT, AstroPORT).

VIII. BioPORT

Within the PORTAL system, the schema BioPORT1 for the specific registry type BioPORT is derived from the schema PORTAL1 for the generic registry type PORTAL. BioPORT focuses its semantic lens on biomedical computing as the problem domain of inquiry. The policies imposed by the BioPORT registry type are intended to be as flexible as possible to allow graceful evolution with respect to changing biomedical and computing ontologies. These flexible policies should facilitate the development of applications built upon the infrastructure services exposed by DOORS servers publishing locations and descriptions of resources with labels and tags registered at PORTAL registries of the BioPORT type.

This flexibility entails allowing the RDF triples of the resource description to reference any version of any biomedical or computing ontology when making a semantic statement about the resource. To limit the “payload” size of resource records redistributed throughout the DOORS server network, and to limit the search space for DOORS semantic queries of resource descriptions, the number of RDF triples allowed per resource description must be constrained. It is arbitrarily set at a maximum of nine in BioPORT with a minimum of two, of which one must be a biomedical statement and the other a computing statement, that reference a simple ontology on biomedical computing integrated within BioPort. However, each of the seven other

semantic statements per record can reference different external ontologies without restriction. Moreover, a resource owner can modify the set of semantic statements in the resource description at any time. Similarly, the owner can modify the resource location at any time in a manner analogous to changing the IP address numbers for a domain name record at an authoritative DNS server.

BioPORT's flexibility also entails permitting optional PORTAL cross-references and tags to be associated with the unique required PORTAL label as defined in Section VII-B. Although PORTAL requires neither cross-references nor tags, BioPORT permits cross-references, permits *supporting tags*, and requires a *principal tag*. The latter must be a noncolonized and tokenized name unique within BioPORT. Imposing this policy enables BioPORT registration of resources with a required PORTAL label that defaults to a concatenation of BioPORT's namespace with the principal tag for the resource. This default labeling scheme benefits those owners with offline resources and/or without their own supported URI or IRI namespaces. *Supporting tags* may be any word or phrase strings registered optionally in BioPORT as additional tags. *Cross-references* may be any URIs or IRIs stored optionally in BioPORT for identifying the resource in other systems. This flexibility should encourage the development of applications that exploit DOORS string searches on resource labels and tags in addition to DOORS semantic searches on resource descriptions while maintaining cross-links between resources in PORTAL-DOORS and cross-references to other systems.

As a metadata registry intended for biomedical computing resources, BioPORT will not only be limited to resources implemented primarily as web services or as grid services, but will also be available for the registration of resources that do not require the Internet for operation, are not now implemented as web or grid services, or perhaps may never be at any time in the future. Some representative examples of such resources include Dalal's NUTMEG [74] in neuroimaging and Taswell's ELIDA [71] in biostatistics (see pseudorecord example in Section VII-A). Other examples abound for offline resources that would contribute to more productive research if registered with appropriate web-enabled semantic interlinks to the scientific literature and other resources. Finally, the registration of a label, principal tag, and/or supporting tags for a resource at a BioPORT registry does not preclude registration at other registries (see Fig. 2 with BioPORT, ManRay [73], and NeuroPort registries) in a manner analogous to the registration of the same name in different *.com, *.net, and *.org generic top-level domain name registries.

IX. RESOURCE LABELS VERSUS DOMAIN NAMES

The PORTAL-DOORS framework proposed here has been designed as part of the infrastructure for the semantic web. As an infrastructure system based on resource metadata with labels, tags, locations, and descriptions, it facilitates most readily the development of semantic search applications. However, it can also serve as part of the foundation for developing other kinds of semantic web applications. Just as distributed cached copies of DNS records with domain names and addresses helped to spur

the growth of the original web, so too will distributed cached copies of DOORS records with resource labels and locations help further the growth of the semantic web. Just as people were motivated to register and assume the responsibility and ownership of domain names, so too will people be motivated to register and assume the responsibility for resource labels, especially if appropriate resource-label-driven browsers for the semantic web are developed analogous to the domain-name-driven browsers for the original web.

Resource labels are different from domain names in many ways not least of which remains the greater universality of resource labels with their associated tags and descriptions. As defined here, resource labels are much more general and flexible than domain names. A resource may be any entity whether abstract or concrete, whether offline or online. Its label may be any URI or IRI. Its nonsemantic tag may be any tokenized name or phrase including anything from multiword phrases to restricted noncolonized names capable of serving as an unqualified XML name for an XML tag. Its semantic description may be any set of RDF triples referencing ontologies. As a compelling example of a registry type fully exploiting the capabilities of the resource label system proposed here, a patent and trademark office could develop a registry type with policies that accommodate the registration of resources that may be products, services, and patented devices or methods. These resources may be assigned unique labels with associated tags consisting of one or more trade or service marks, and with associated descriptions referencing ontologies for patent and trademark classes and the semantic definitions for entities within those classes.

Moreover, just as domain names have served many different Internet communication protocols from telnet to http, so too can resource labels serve many different currently evolving communication systems whether web services or distributed grid. If implemented, supported, and maintained as a separate independent infrastructure, the PORTAL-DOORS framework can be tuned and optimized for metadata while the distributed grid and web services will perhaps continue to be optimized, respectively, for scientific and business purposes, each with different kinds of messaging requirements for the different kinds of data (i.e., not just metadata) exchanged.

An infrastructure optimized for resource metadata and semantic searches with messages of small size limited by design should not necessarily be the same as one optimized for messages of unlimited and potentially large size whether tuned for grid computing with binary data or for secure commerce with text data. By integrating into a common infrastructure framework, the capability for both string search on the resource labels and semantic search on the resource descriptions, the DOORS and PORTAL systems enable a graceful transition from the original web to the semantic web as the ontologies for the semantic web continue to evolve.

X. SYNERGISTIC SYSTEMS

As discussed in Sections II and III, the semantic web has not yet achieved the goals set by its visionaries. Good and Wilkinson [75] assert that "most, if not all, of the standards

and technologies" have been established, and suggest that the barriers to progress remain "social rather than technological." The opinion advanced here in this paper remains contrary, i.e., *not enough of the necessary infrastructure has yet been designed and built.* As reviewed by many authors (including [25], [26], [31], and [32]), semantic web systems currently in place do not suffice. They remain far too complicated to motivate most users, and even many developers, to become involved and participate in building the semantic web.

Other authors have also called for the development of additional necessary technologies, systems, and applications. Quan and Karger [29], Quan *et al.* [76], Dzbor *et al.* [77], and Alani *et al.* [78] have all argued for a semantic web browser or a "killer app" as the necessary key to unlock the doors to the semantic web. Despite choosing the suggestive acronyms DOORS and PORTAL for the infrastructure systems proposed here, they do not suffice alone any more than would the best conceived "killer app" or the most zealous social will.

The original web succeeded as a consequence of the amazing synergism between DNS as a domain name system with registries and name servers, http as a communications protocol, and web browsers to view web pages published by people motivated to register domain names for their web sites. The semantic web will succeed analogously when a similar dynamic synergism can be created between a resource label system with registries and label servers, all of the appropriately optimized communications protocols, and the necessary semantic web browsers and label search clients to access resources published by people motivated to register labels and maintain descriptions for their resources.

Until then, searches on the web (whether at google.com, yahoo.com, or even a specialty search engine when a particular database record locator is not input) will continue to yield irrelevant or innumerable results too often. These results then lose practical usefulness because they consume too much time for the user who attempts to peruse them. New versions of search engines such as Swoogle [79] and OntoLook [80] will hopefully enable useful search of the semantic web in the future. But the current semantic web itself requires sufficient growth and development with enough metadata annotation of enough resources and documents before a threshold of practical use can be attained.

XI. SEMANTIC SEARCH AND ANALYSIS APPLICATIONS

Kazic [81] agrees with the "genuine need for fast, accurate delivery of relevant information in ways that do not overwhelm humans" in her insightful and enlightening analysis on factors influencing the adoption of the semantic web. Regarding this point, she emphasizes "accuracy, relevance, and comprehensibility." But in discussing technology adoption, Kazic omits the mention of the original authoritative work of Rogers [82] and other key investigators including Fichman [83], who have contributed to the empirical field that studies technology innovation, diffusion, and assimilation.

Thus, Kazic's own search of the literature performed as part of her analysis [81] does not meet the declared criterion of

"comprehensibility." A more comprehensive search would have benefited from the semantic association networks (SANs) discussed by Börner [84] for improving scholarly knowledge and expertise management. SANs integrated with digitalic libraries would enable investigators to cross disciplines and search fields outside of their main area of expertise without being required to know in advance key words such as the phrase "diffusion of innovations" that would have been relevant to Kazic's search of the literature.

From the perspective of biomedical computing in health care and life sciences, the topic of technology innovation, diffusion, and assimilation provides an interesting test case for continuing development of semantic search and SANs. Can theoretical and computational models with results from experimental research for technology and knowledge diffusion [82], [83], [85], [86] be related to work on predator-prey interactions [87], the spread of epidemics [88], and speciation in phylogenetics [89] such that these scholarly fields are interlinked within SANs? Will the creation of SANs in this manner contribute to cross-fertilization that yields more productive research in each of the fields participating?

Productive cross-fertilization with improved communication between basic and clinical science remains the primary goal of translational research and drug discovery [90]. As an important use case for semantic web technologies (with infrastructure components and services for machines) and semantic search, decision support, and knowledge management applications (with user interface tools for humans), informatics for translational medicine on the semantic web and grid will be driven by the compelling needs and powerful finances of the health care and pharmaceutical industries. Continuing development of the infrastructure, tools, and applications will be guided by benchmarks and measures for ontology evaluation [91]–[94], knowledge ranking [95], system performance [96], and surely other new metrics yet to be invented, as well as by the legal and social issues pertaining to semantic web standards [97].

XII. HYBRID, BOOTSTRAP, AND BRIDGE

To gain traction, PORTAL-DOORS should initially focus on development as an infrastructure for search applications with the tangible benefit of saving people time, and on application contexts where many people would be motivated to promote and publicize their resources, e.g., named entities including information databases and computing applications in scholarly research, or trademarks and named products and services in commercial business. If so, then the PORTAL-DOORS infrastructure proposed here could serve as a *bootstrap* to help further jumpstart the semantic web and the development of more sophisticated systems (e.g., agent-driven composition of services [98]) that require search as just one piece of the puzzle. As a bootstrap, the PORTAL-DOORS framework eschews debates about formal ontologies versus informal folksonomies and microformats [99]–[101]. Instead, it creates a *hybrid* with labels (URIs and IRIs) and tags (key word and phrase strings) for the original web, and with descriptions

(RDF triples) for the semantic web, that also serves as an effective *bridge* between the original web and the semantic web.

Concomitant development of resource label and tag editors for PORTAL registry records, resource location and description editors for DOORS server records, and semantic search clients or plugin modules for web browsers should be guided by designs intended to help motivate resource owners by simplifying for them the task of label registration and description maintenance for their resources. Resource descriptions will evolve over time as owners replace semantic statements referencing older ontologies with statements referencing newer ontologies. The collective wisdom of large number of people will thus determine the popularity of ontologies and the usage patterns of the RDF triples for the semantic statements contained within the resource descriptions. The analysis of these usage patterns and their reflection of human thought and behavior will enable the development of improved ontologies. This kind of investigation will constitute another manifestation of the new *science of the web* [3].

Web science itself must be pursued with full cognizance of *antedisciplinary versus interdisciplinary* science and multidisciplinary teams versus individuals [102], [103]. In particular, Eddy [102] emphasizes that, “Progress is driven by new scientific questions, which demand new ways of thinking. You want to go where a question takes you, not where your training left you.” As web science matures, and ontologies improve, the PORTAL-DOORS framework can be enhanced and refined. If it is fully implemented and adopted as part of the infrastructure foundation for the semantic web sufficiently popularized by both developers and users, then PORTAL and DOORS will contribute to building the pervasive web of knowledge envisioned by the founders of the semantic web.

XIII. CONCLUSION

PORTAL and DOORS are proposed as systems, respectively, for registering resource labels and tags and publishing resource locations and descriptions. They are thus analogous to IRIS and DNS, respectively, for registering domain names and publishing domain addresses. PORTAL and DOORS are designed to serve the semantic web just as IRIS and DNS are designed to serve the original web (see Table I). The PORTAL-DOORS paradigm favors a flexible and modular approach promoting collaborative networks of cross-linking resources and interreferencing ontologies capable of evolving dynamically with any changing standards for RDF and OWL that add future extensions for ordered relationships, probabilistic reasoning, or other refinements.

In contrast with existing directories such as OBRC [23] that endeavor to become a “one-stop gateway” to resources, the PORTAL-DOORS paradigm seeks to build a decentralized and distributed infrastructure that supports mass collaboration tapping the power of wikinomics [104] and enabling “all webizens to create, share, distribute, and enjoy ideas and information” [105]. In contrast with existing nonsemantic systems such

as PURL and Handle (see Section VI), the PORTAL-DOORS framework is built upon the XML/RDF/OWL foundations of the semantic web. In contrast with existing semantic systems such as SAN [84] or SemBOWSER [106], the PORTAL-DOORS framework does not limit the registration of resources to those of only a particular format or technology such as literature documents [84] or web services [106].

Resources for which unique labels with optional tags are registered in PORTAL are *not* required to be semantic resources in and of themselves. Rather, it is only their descriptions published in DOORS that contain semantic metadata referencing ontologies. The resources themselves may be anything whether abstract or concrete, offline or online. Pending development of user interfaces with label, tag, location, and description editors, resource owners should be able to maintain their own data records at PORTAL and DOORS servers *without* the intervention of expert curators or administrators.

In analogy with the IRIS-DNS framework and its multiplicity of registries for top-level domains such as *.com, *.net, and *.org, the PORTAL-DOORS framework enables a multiplicity of registries for different problem-oriented domains such as BioPORT (see Sections IV and VIII), ManRay [73], and NeuroPORT (see Fig. 2). Cross-registry searches will be facilitated by the common shared semantic foundation throughout the PORTAL-DOORS server networks.

Open source projects for BioPORT, PORTAL, and DOORS will be hosted at www.biomedicalcomputing.org and at www.portaldoors.org. Working drafts of BioPORT1, PORTAL1, and DOORS1 schemas will serve as the formal specifications for BioPORT as a prototype metadata registry within a system of PORTAL registries for resource labels and tags and DOORS servers for resource locations and descriptions. Initial root servers for PORTAL and DOORS will be maintained, respectively, at portal.portaldoors.org and at doors.portaldoors.org. The author invites and welcomes collaborators to contribute to these projects.

ACKNOWLEDGMENT

The author thanks S. R. Eddy, M. F. Huerta, and all who promote and support antedisciplinary and multidisciplinary approaches to science. In addition, he thanks the editors and anonymous reviewers of the IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE (TITB) for their helpful comments and assistance with the preparation of this manuscript.

REFERENCES

- [1] I. Jacobs, About the World Wide Web consortium (W3C) [Online]. Available: <http://www.w3.org/Consortium/about-w3c>
- [2] T. Berners-Lee, J. Hendler, and O. Lassila, “The semantic web,” *Sci. Amer.*, vol. 284, no. 5, pp. 34–43, May 2001.
- [3] T. Berners-Lee, W. Hall, J. Hendler, N. Shadbolt, and D. J. Weitzner, “Computer science. Creating a science of the web,” *Science*, vol. 313, no. 5788, pp. 769–771, Aug. 2006.
- [4] Extensible markup language (XML). (2006, Aug.). W3C Recommendation [Online]. Available: <http://www.w3.org/XML/>.
- [5] Resource description framework (RDF). (2004, Feb.). W3C Recommendation [Online]. Available: <http://www.w3.org/RDF/>

- [6] J. Z. Pan, "A flexible ontology reasoning architecture for the semantic web," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 2, pp. 246–260, Feb. 2007.
- [7] Web ontology language (OWL). (2004, Feb.). W3C Recommendation [Online]. Available: <http://www.w3.org/2004/OWL/>
- [8] T. Berners-Lee, D. Fensel, J. A. Hendler, and H. Lieberman, *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. New York: MIT Press, 2005.
- [9] L. W. Lacy, *OWL: Representing Information Using the Web Ontology Language*. Victoria, BC, Canada: Trafford, 2005.
- [10] V. Geroimenko and C. Chen, Eds., *Visualizing the Semantic Web: XML-Based Internet and Information Visualization*, 2nd ed. New York: Springer-Verlag, 2006.
- [11] J. Davies, R. Studer, and P. Warren, Eds., *Semantic Web Technologies: Trends and Research in Ontology-Based Systems*. New York: Wiley, 2006.
- [12] C. J. O. Baker and K.-H. Cheung, Eds., *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*. New York: Springer-Verlag, 2007.
- [13] L. Baresi, E. D. Nitto, and C. Ghezzi, "Toward open-world software: Issues and challenges," *IEEE Comput.*, vol. 39, no. 10, pp. 36–43, Oct. 2006.
- [14] H. Zhuge, "Future interconnection environment," *Computer*, vol. 38, no. 4, pp. 27–33, Apr. 2005.
- [15] H. Zhuge, "Semantic grid: Scientific issues, infrastructure, and methodology," *Commun. ACM*, vol. 48, no. 4, pp. 117–119, 2005.
- [16] M. A. Harris, *et al.*, "The Gene Ontology (GO) project in 2006," *Nucleic Acids Res.*, vol. 34, pp. D322–D326, 2006.
- [17] O. Bodenreider and R. Stevens, "Bio-ontologies: Current trends and future directions," *Brief Bioinf.*, vol. 7, no. 3, pp. 256–274, Sep. 2006.
- [18] F. Pinciroli and D. M. Pisanelli, "The unexpected high practical value of medical ontologies," *Comput. Biol. Med.*, vol. 36, no. 7/8, pp. 669–673, 2006.
- [19] C. Brooksbank, G. Cameron, and J. Thornton, "The European Bioinformatics Institute's data resources: Towards systems biology," *Nucleic Acids Res.*, vol. 33, pp. D46–D53, Jan. 2005 (Database issue).
- [20] D. L. Wheeler *et al.*, "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Res.*, vol. 33, pp. D39–D45, Jan. 2005.
- [21] M. Y. Galperin. (2007, Jan.). The molecular biology database collection: 2007 update. *Nucleic Acids Res.* [Online]. 35, D3–D4. Available: <http://dx.doi.org/10.1093/nar/gkl1008>
- [22] J. A. Fox, S. McMillan, and B. F. F. Ouellette, "A compilation of molecular biology web servers: 2006 update on the bioinformatics links directory," *Nucleic Acids Res.*, vol. 34, pp. W3–W5, Jul. 2006.
- [23] Y.-B. Chen, A. Chattopadhyay, P. Bergen, C. Gadd, and N. Tannery, "The online bioinformatics resources collection at the university of pittsburgh health sciences library system: A one-stop gateway to online bioinformatics databases and software tools," *Nucleic Acids Res.*, vol. 35, pp. D780–D785, Jan. 2007.
- [24] S. Philippi and J. Köhler, "Using XML technology for the ontology-based semantic integration of life science databases," *IEEE Trans. Inf. Technol. Biomed.*, vol. 8, no. 2, pp. 154–160, Jun. 2004.
- [25] S. Philippi and J. Köhler, "Addressing the problems with life-science databases for traditional uses and systems biology," *Nat. Rev. Genet.*, vol. 7, no. 6, pp. 482–488, Jun. 2006.
- [26] E. Neumann, "A life science semantic web: Are we there yet?," *Sci. STKE*, vol. 2005, no. 283, p. pe22, May 2005.
- [27] E. K. Neumann and D. Quan, "Biodash: A semantic web dashboard for drug development," in *Proc. Pac. Symp. Biocomput.*, 2006, pp. 176–187.
- [28] T. Clark, S. Martin, and T. Liefeld, "Globally distributed object identification for biological knowledgebases," *Brief Bioinf.*, vol. 5, no. 1, pp. 59–70, Mar. 2004.
- [29] D. Quan and D. R. Karger, "How to make a semantic web browser," in *Proc. 13th Int. Conf. World Wide Web*. New York: Association for Computing Machinery Press, 2004, pp. 255–265.
- [30] W3C semantic web health care and life sciences interest group [Online]. Available: <http://www.w3.org/2001/sw/hcls/>
- [31] N. Cannata, E. Merelli, and R. B. Altman, "Time to organize the bioinformatics resourceome," *PLoS Comput. Biol.*, vol. 1, no. 7, p. e76, Dec. 2005.
- [32] K. H. Buetow, "Cyberinfrastructure: Empowering a 'third way' in biomedical research," *Science*, vol. 308, no. 5723, pp. 821–824, May 2005.
- [33] C. Wroe, C. Goble, M. Greenwood, P. Lord, S. Miles, J. Papay, T. Payne, and L. Moreau, "Automating experiments using semantic data in a bioinformatics grid," *IEEE Intell. Syst.*, vol. 19, no. 1, pp. 48–55, Jan./Feb. 2004.
- [34] V. Astakhov, A. Gupta, S. Santini, and J. S. Grethe, "Data integration in the Biomedical Informatics Research Network (BIRN)," in *Proc. Data Integr. Life Sci. (DILS 2005)* (Lecture Notes in Computer Science), vol. 3615, pp. 317–320.
- [35] J. Phillips, R. Chilukuri, G. Fragozo, D. Warzel, and P. A. Covitz, The caCORE software development kit: streamlining construction of interoperable biomedical information services, *BMC Med. Inf. Decis. Making* [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16398930
- [36] D. De Roure and J. Hendler, "E-science: The grid and the semantic web," *IEEE Intell. Syst.*, vol. 19, no. 1, pp. 65–71, Jan./Feb. 2004.
- [37] D. De Roure, N. Jennings, and N. Shadbolt, "The semantic grid: Past, present, and future," *Proc. IEEE*, vol. 93, no. 3, pp. 669–681, Mar. 2005.
- [38] J. M. Benyus, *Biomimicry: Innovation Inspired by Nature*. New York: William Morrow, 1997.
- [39] P. V. Mockapetris. (1987, Nov.). STD13 RFC1035: Domain names—Implementation and specification. IETF Full Standard [Online]. Available: <http://www.ietf.org/rfc/rfc1035.txt>
- [40] C. Liu and P. Albitz, *DNS and BIND*, 5th ed. Sebastopol, CA: O'Reilly Media, 2006.
- [41] G. Lawton, "Stronger domain name system thwarts root-server attacks," *IEEE Comput.*, vol. 40, no. 5, pp. 14–17, May 2007.
- [42] D. Eastlake. (1999, Mar.). RFC2535: Domain name system security extensions. IETF RFC 2535 Working Draft [Online]. Available: <http://www.ietf.org/rfc/rfc2535.txt>
- [43] O. Kolkman and R. Gieben. (2006, Sep.). RFC4641: DNSSEC operational practices. IETF RFC 4641 Working Draft [Online]. Available: <http://www.ietf.org/rfc/rfc4641.txt>
- [44] P. Faltstrom, P. Hoffman, and A. Costello. (2003, Mar.). RFC3490: Internationalizing domain names in applications (IDNA). IETF RFC 3490 Working Draft [Online]. Available: <http://www.ietf.org/rfc/rfc3490.txt>
- [45] ICANN. (2003, Jun.). Guidelines for the implementation of internationalized domain names (version 1.0). Internet Corporation for Assigned Names and Numbers [Online]. Available: <http://www.icann.org/general/idn-guidelines-20jun03.htm>
- [46] A. Newton, "Replacing the whois protocol: IRIS and the IETF's CRISP working group," *IEEE Internet Comput.*, vol. 10, no. 4, pp. 79–84, Jul./Aug. 2006.
- [47] IETF Working Group Charter. (2006, Aug.). Cross registry information service protocol (CRISP) [Online]. Available: <http://www.ietf.org/html.charters/crisp-charter.html>
- [48] A. Newton. (2004, Feb.). RFC3707: Cross registry internet service protocol (CRISP) requirements. IETF RFC 3707 Working Draft [Online]. Available: <http://www.ietf.org/rfc/rfc3707.txt>
- [49] A. Newton, and M. Sanz. (2005, Jan.). RFC3981: IRIS: The internet registry information service (IRIS) core protocol. IETF RFC 3981 Working Draft [Online]. Available: <http://www.ietf.org/rfc/rfc3981.txt>
- [50] A. Newton, and M. Sanz. (2005, Jan.). RFC3982: IRIS: A domain registry (dreg) type for the internet registry information service. IETF RFC 3982 Working Draft [Online]. Available: <http://www.ietf.org/rfc/rfc3982.txt>
- [51] E. Gunduz, A. Newton, and S. Kerr. (2006, Oct.). RFC4698: IRIS: An address registry (areg) type for the internet registry information service. IETF RFC 4698 Working Draft [Online]. Available: <http://www.ietf.org/rfc/rfc4698.txt>
- [52] A. Newton. (2007, Jan.). XML pipelining with chunks for the information registry information service. IETF Working Draft [Online]. Available: <http://tools.ietf.org/html/draft-ietf-crisp-iris-xpc-05>
- [53] A. Newton, and M. Sanz. (2005, Jan.). RFC3983: Using the internet registry information service (IRIS) over the blocks extensible exchange protocol (BEEP). IETF RFC 3983 Working Draft [Online]. Available: <http://www.ietf.org/rfc/rfc3983.txt>
- [54] Computer retrieval of information on scientific projects (CRISP). NIH database [Online]. Available: <http://crisp.cit.nih.gov/>
- [55] U. Radetzki, U. Leser, S. C. Schulze-Rauschenbach, J. Zimmermann, J. Lussem, T. Bode, and A. B. Cremers, "Adapters, shims, and glue-service interoperability for in silico experiments," *Bioinformatics*, vol. 22, no. 9, pp. 1137–1143, 2006.
- [56] R. M. Bruskewich, A. B. Cosico, W. Eusebio, A. M. Portugal, L. M. Ramos, M. T. Reyes, M. A. B. Sallan, V. J. M. Ulat, X. Wang, K. L. McNally, R. S. Hamilton, and C. G. McLaren, "Linking genotype to phenotype: The international rice information system (IRIS)," *Bioinformatics*, vol. 19, no. 1, pp. i63–i65, 2003.

- [57] T. Berners-Lee, R. Fielding, and L. Masinter. (2005, Jan.). STD66 RFC3986: Uniform resource identifier (URI) generic syntax. IETF Full Standard RFC [Online]. Available: <http://www.ietf.org/rfc/rfc3986.txt>.
- [58] M. Duerst, and M. Suignard. (2005, Jan.). RFC3987: Internationalized resource identifiers (IRIs). IETF RFC 3987 Working Draft [Online]. Available: <http://www.ietf.org/rfc/rfc3987.txt>.
- [59] URIs, URLs, and URNs: Clarifications and recommendations 1.0. (2001, Sep.). W3C Note, accessed version 20010921 [Online]. Available: <http://www.w3.org/TR/uri-clarification/>.
- [60] J. Borden, and T. Bray. (2002, Feb.). Resource directory description language (RDDL) [Online]. Available: <http://www.rddl.org/>.
- [61] XHTML 1.0 the extensible hypertext markup language (2nd ed.) (2002, Aug.). W3C Recommendation [Online]. Available: <http://www.w3.org/TR/xhtml1/>.
- [62] XML linking language (XLink) version 1.0. (2001, Jun.). W3C Recommendation [Online]. Available: <http://www.w3.org/TR/xlink/>.
- [63] Harvesting RDF statements from XLinks. (2000, Sep.). W3C Note, accessed version 20000929 [Online]. Available: <http://www.w3.org/TR/xlink2rdf/>.
- [64] XML topic maps (XTM) 1.0 (2001, Aug.). [Online]. Available: <http://www.topicmaps.org/xtm/1.0/>.
- [65] K. Shafer, S. Weibel, E. Jul, and J. Fausey. (1996, Mar.). Introduction to persistent uniform resource locators [Online]. Available: <http://purl.oclc.org/docs/inet96.html>.
- [66] R. Kahn and R. Wilensky, "A framework for distributed digitalic object services," *Int. J. Digitalic Libraries*, vol. 6, no. 2, pp. 115–123, 2006.
- [67] S. Sun, L. Lannom, and B. Boesch. (2003, Nov.). RFC3650: Handle system overview. IETF RFC 3650 Working Draft [Online]. Available: <http://www.ietf.org/rfc/rfc3650.txt>.
- [68] J. Z. Pan and I. Horrocks, "RDFS(FA): Connecting RDF(S) and OWL DL," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 2, pp. 192–206, Feb. 2007.
- [69] A. Newton, and F. Neves. (2006, May). Domain registry version 2 for the internet registry information service. IETF Working Draft [Online]. Available: <http://tools.ietf.org/wg/crisp/draft-ietf-crisp-iris-dreg2/>.
- [70] D. Eastlake, J. Reagle, and D. Solo. (2002, Mar.). RFC3275: Extensible markup language XML-Signature syntax and processing. IETF RFC 3275 Working Draft [Online]. Available: <http://www.ietf.org/rfc/rfc3275.txt>.
- [71] C. Taswell, "Limiting dilution assays for the determination of immunocompetent cell frequencies. I. Data analysis," *J. Immunol.*, vol. 126, no. 4, pp. 1614–1619, Apr. 1981.
- [72] SPARQL query language for RDF. (2006, Oct.). W3C Working Draft [Online]. Available: <http://www.w3.org/TR/rdf-sparql-query/>.
- [73] C. Taswell, B. Franc, and R. Hawkins, "The ManRay project: Initial development of a web-enabled ontology for nuclear medicine," presented at the 53rd Annu. Meeting Soc. Nuclear Med., San Diego, CA, Jun. 2006.
- [74] S. S. Dalal, J. M. Zumer, V. Agrawal, K. E. Hild, K. Sekihara, and S. S. Nagarajan, "Nutmeg: A neuromagnetic source reconstruction toolbox," *Neurol. Clin. Neurophysiol.*, vol. 2004, p. 52, 2004.
- [75] B. M. Good and M. D. Wilkinson, "The life sciences semantic web is full of creeps," *Brief Bioinf.*, vol. 7, no. 3, pp. 275–286, Sep. 2006.
- [76] D. Quan, D. Huynh, and D. R. Karger, "Haystack: A platform for authoring end user semantic web applications," in *Semantic web (ISWC 2003)*. Germany, Berlin: Springer-Verlag, vol. 2870, pp. 738–753.
- [77] M. Dzbor, J. Domingue, and E. Motta, "Magpie—Towards a semantic web browser," in *Proc. Int. Semantic Web Conf. (ISWC 2003)* (Lecture Notes in Computer Science), vol. 2870, pp. 690–705.
- [78] H. Alani, Y. Kalfoglou, K. O'Hara, and N. Shadbolt, "Towards a killer app for the semantic web," in *Proc. Int. Semantic Web Conf. (ISWC 2005)*, (Lecture Notes in Computer Science), vol. 3729, pp. 829–843.
- [79] L. Ding, T. Finin, A. Joshi, R. Pan, R. S. Cost, Y. Peng, P. Reddivari, V. Doshi, and J. Sachs, "Swoogle: A search and metadata engine for the semantic web," in *Proc. 13th ACM Int. Conf. Inf. Knowl. Manage. (CIKM 2004)*. New York: ACM Press, pp. 652–659.
- [80] Y. Li, Y. Wang, and X. Huang, "A relation-based search engine in semantic web," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 2, pp. 273–282, Feb. 2007.
- [81] T. Kazic, "Factors influencing the adoption of the semantic web in the life sciences," in *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, C. J. O. Baker and K.-H. Cheung, Eds. New York: Springer-Verlag, 2007, ch. 17, pp. 399–412.
- [82] E. M. Rogers, *Diffusion of Innovations*, 5th ed. Washington, DC: Free Press, 2003.
- [83] R. G. Fichman, "The diffusion and assimilation of information technology innovations," in *Framing the Domains of IT Management Projecting the Future . . . Through the Past*, R. W. Zmud, Ed. Cincinnati, OH: Pinnaflex Educ. Resour., 2000, pp. 105–127.
- [84] K. Börner, "Semantic association networks: Using semantic web technology to improve scholarly knowledge and expertise management," in *Visualizing the Semantic Web: XML-Based Internet and Information Visualization*, V. Geroimenko and C. Chen, Eds., 2nd ed. New York: Springer-Verlag, 2006, ch. 11, pp. 183–198.
- [85] H. Zhuge, "Discovery of knowledge flow in science," *Commun. ACM*, vol. 49, no. 5, pp. 101–107, 2006.
- [86] H. Zhuge, "Exploring an epidemic in an e-science environment," *Commun. ACM*, vol. 48, no. 9, pp. 109–114, 2005.
- [87] A. A. Berryman, "The origins and evolution of predator-prey theory," *Ecology*, vol. 73, no. 5, pp. 1530–1535, 1992.
- [88] D. J. Daley and J. Gani, *Epidemic Modelling: An Introduction*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
- [89] J. Sullivan and P. Joyce, "Model selection in phylogenetics," *Annu. Rev. Ecol. Evol. Syst.*, vol. 36, pp. 445–466, 2005.
- [90] V. Kashyap and T. Hongsermeier, "Can semantic web technologies enable translational medicine?," in *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, C. J. O. Baker and K.-H. Cheung, Eds. New York: Springer-Verlag, 2007, ch. 12, pp. 249–279.
- [91] C. Patel, K. Supekar, Y. Lee, and E. K. Park, "Ontokhoj: A semantic web portal for ontology searching, ranking and classification," in *Proc. 5th ACM Int. Workshop Web Inf. Data Manage. (WIDM 2003)*, ACM Press, New York, pp. 58–61.
- [92] J. Brank, M. Grobelnik, and D. Mladenic, "A survey of ontology evaluation techniques," presented at the SIKDD 2005, Ljubljana, Slovenia, Oct. 2005.
- [93] M. Sabou, V. Lopez, E. Motta, and V. Uren, "Ontology selection: Ontology evaluation on the real semantic web," presented at the WWW 2006, Edinburgh, U.K., May.
- [94] L. Obrst, W. Ceusters, I. Mani, S. Ray, and B. Smith, "The evaluation of ontologies," in *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, C. J. O. Baker and K.-H. Cheung, Eds. New York: Springer-Verlag, 2007, ch. 7, pp. 139–158.
- [95] L. Ding, R. Pan, T. Finin, A. Joshi, Y. Peng, and P. Kolari, "Finding and ranking knowledge on the semantic web," in *Proc. Int. Semantic Web Conf. (ISWC 2005)*, (Lecture Notes in Computer Science), vol. 3729, pp. 156–170.
- [96] Y. Guo, A. Qasem, Z. Pan, and J. Heflin, "A requirements driven framework for benchmarking semantic web knowledge base systems," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 2, pp. 297–309, Feb. 2007.
- [97] D. Greenbaum and M. Gerstein, "Semantic web standards: Legal and social issues and implications," in *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, C. J. O. Baker and K.-H. Cheung, Eds. New York: Springer-Verlag, 2007, ch. 18, pp. 413–433.
- [98] M. Burstein, C. Bussler, T. Finin, M. N. Huhns, M. Paolucci, A. P. Sheth, S. Williams, and M. Zaremba, "A semantic web services architecture," *IEEE Internet Comput.*, vol. 9, no. 5, pp. 72–81, Sep./Oct. 2005.
- [99] R. Khare, "Microformats: The next (small) thing on the semantic web?," *IEEE Internet Comput.*, vol. 10, no. 1, pp. 68–75, Jan./Feb. 2006.
- [100] R. McCool, "Rethinking the semantic web. Part 2," *IEEE Internet Comput.*, vol. 10, no. 1, pp. 93–96, Jan./Feb. 2006.
- [101] N. Shadbolt, W. Hall, and T. Berners-Lee, "The semantic web revisited," *IEEE Intell. Syst.*, vol. 21, no. 3, pp. 96–101, Jan./Feb. 2006.
- [102] S. R. Eddy, "'Antedisciplinary' science," *PLoS Comput. Biol.*, vol. 1, no. 1, p. e6, Jun. 2005.
- [103] M. F. Huerta, G. K. Farber, E. L. Wilder, D. V. Kleinman, P. A. Grady, D. A. Schwartz, and L. A. Tabak, "NIH roadmap interdisciplinary research initiatives," *PLoS Comput. Biol.*, vol. 1, no. 6, p. e59, Nov. 2005.
- [104] D. Tapscott and A. D. Williams, *Wikinomics: How Mass Collaboration Changes Everything*. New York: Portfolio, 2006.
- [105] K.-J. Lin, "Building web 2.0," *IEEE Comput.*, vol. 40, no. 5, pp. 101–102, May 2007.
- [106] S. S. Sahoo, A. Sheth, B. Hunter, and W. S. York, "SemBOWSER: Adding semantics to biological web services registry," in *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, C. J. O. Baker and K.-H. Cheung, Eds. New York: Springer-Verlag, 2007, ch. 14, pp. 318–340.



Carl Taswell (M'90) received a 1978 B.A. degree in biochemistry from Harvard University, a 1984 M.Sc. degree in mathematics and a 1985 M.D. degree in medicine both from New York University, and a 1991 M.Sc. degree and 1995 Ph.D. degree both in scientific computing and computational mathematics from Stanford University.

He has completed clinical training in psychiatry at University of California San Diego and in nuclear medicine at University of California San Francisco. Since receiving a top ten scholarship award in the 1974 Westinghouse Science Talent Search, he has published dozens of research papers spanning the fields of biochemistry, immunology, medicine, statistics, mathematics, and informatics. Ever passionate about biomedical computing, he has designed and built software including LDACPA, ELIDA, WavBox, and GeneScene.

Dr. Taswell is also a member of the Society for Industrial and Applied Mathematics, the American Medical Informatics Association, and the Society of Nuclear Medicine.