2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks

The Hierarchically Distributed Mobile Metadata (HDMM) Style of Architecture for Pervasive Metadata Networks

Carl Taswell
Global TeleGenetics, Inc.
Ladera Ranch, CA, USA
ctaswell@computer.org

***Abstract*—The Internet Registry Information Service (IRIS) registers domain names while the Domain Name System (DNS) publishes domain addresses with mapping of names to addresses for the original web. Analogously, the Problem Oriented Registry of Tags And Labels (PORTAL) registers resource labels and tags while the Domain Ontology Oriented Resource System (DOORS) publishes resource locations and descriptions with mapping of labels to locations for the semantic web. Both the IRIS-DNS System and the PORTAL-DOORS System share a common architectural style for pervasive metadata networks that operate as distributed metadata management systems with hierarchical authorities for entity registering and attribute publishing. Hierarchical control of metadata redistribution throughout the registry-directory networks constitutes an essential characteristic of this architectural style called Hierarchically Distributed Mobile Metadata (HDMM) with its focus on moving the metadata for *who what where* as fast as possible from servers in response to requests from clients.**

***Keywords*-architectural style; mobile metadata; hierarchical authority; distributed registry-directory system; PORTAL-DOORS System; IRIS-DNS System; HDMM.**

I. Introduction

According to Taylor *et al.* [1], "an architectural style is a named collection of architectural design decisions that (1) are applicable in a ... context, (2) constrain architectural design decisions [for] a system within that context, and (3) elicit beneficial qualities in each resulting system." The REpresentational State Transfer (REST) architectural style [2] serves as an important example of an architectural style for network-based applications on the web. Other styles, such as peer-to-peer, have been named and described for distributed and networked architectures [1].

However, not all styles of distributed and networked architectures have been appropriately characterized with an identifying name and detailed description of the principles that constitute the essential distinguishing aspects of the style. In particular, the architectural style that characterizes both IRIS-DNS and PORTAL-DOORS has not yet been elaborated with an explicit name and description even though the PORTAL-DOORS System for the semantic web was purposefully architected by the author [3] by analyzing and emulating the design principles and paradigm of the IRIS-DNS System for the original web. Therefore, this report formally names and describes the architectural style shared by both IRIS-DNS and PORTAL-DOORS as pervasive registry-directory networks with Hierarchically Distributed Mobile Metadata (HDMM), and then further discusses the architecture and usage of PORTAL-DOORS within the context of this HDMM architectural style.

II. Hierarchically Distributed Mobile Metadata as an Architectural Style

IRIS registries [4] and DNS directories [5] provide the model for the architectural style that inspired the design of PORTAL registries and DOORS directories [3]. The most essential characteristics of this HDMM architectural style can be summarized by the following principles:

1) Pervasively distributed and shared infrastructure, content, and control of content including distributed and shared control over both the contribution and distribution of the content.
2) A hierarchy of both authoritative and non-authoritative servers (root, primary, secondary, forwarding and caching) enabling global interoperable communication while permitting local control of policies.
3) A separation of concerns with registries for identification and directories for location.
4) A freedom of choice in the selection of identifiers with purposeful absence of any requirement to use the *same* top-level root name or label for *all* identifiers, thus enabling essentially unrestricted choice of naming or labeling schemes for identification and avoiding monopolistic control by any single organization.
5) A focus on moving the metadata for 'who what where' as fast as possible from servers in response to requests from clients that access non-authoritative local forwarding and caching servers updated regularly by the authoritative servers.

Users of today's web browsers may not be familiar with the engineering of the hidden infrastructure system that enables them to navigate to any web site around the world. But it is the IRIS-DNS infrastructure system, which is responsible for registering domain names and mapping them to numerical IP addresses, that makes it possible for the user to browse the web in such an effortless manner almost

978-0-7695-3908-9/09 $26.00 © 2009 IEEE
DOI 10.1109/I-SPAN.2009.73

always without ever typing, seeing, or even being aware of the existence of the numerical IP addresses.

Moreover, from the user's perspective, what is most important now is that the speed of this conversion from domain name to IP address occurs so rapidly that the user does not experience it as a hindrance or delay in browsing. Even if the particular web page itself downloads and displays slowly, usually at least the web site address is found quickly. And that happens because the small amount of metadata (domain name and IP address) moves so quickly across the internet even if the larger amount of data (web page text and media) does not. Because of this important point, the phrase *Hierarchically Distributed Mobile Metadata* and acronym *HDMM* was introduced (9 May 2009 at www.portaldoors.org) as a name for this architectural style that characterizes both IRIS-DNS and PORTAL-DOORS.

Whereas IRIS-DNS implements the HDMM architectural style for the original web, PORTAL-DOORS extends and implements this style for the semantic web and grid. Further, PORTAL-DOORS extends the separation of concerns principle (see Item 3 above) to include the additional notion of separately optimising directories for semantic services (with use of logical reasoning, ontologies and the RDF/OWL/SPARQL stack of technologies) and the registries for lexical services (with use of character string processing, terminologies and only those XML technologies that do not require use of RDF triples). This separation of concerns enables the back-end use of traditional relational database stores for PORTAL registries and RDF-triple database stores for DOORS directories. Of course, XML stores and/or hybrid stores (such as OpenLink Virtuoso [6] which is an open source cross platform universal server) can also be used for both PORTAL and DOORS servers and services.

III. Architectural Design of the PORTAL-DOORS Infrastructure System

In accordance with the HDMM architectural style, PORTAL-DOORS has been designed to serve the semantic web and grid in a manner analogous to the way that IRIS-DNS has served the original web. The table and figures below have been adapted from the original 'blueprint' paper [3] and updated with revisions [7], [8]. Note that the original *separate* design of PORTAL registries and DOORS directories has been supplemented with a new bootstrapping *combined* design with integrated NEXUS registrars [8]. Both can coexist together.

Table I summarizes some of the similarities and differences between these paradigms from the perspective of considering both as distributed database systems with entity registering and attribute publishing implemented with the HDMM architectural style (see Section II). Figure 1 displays a diagram depicting the structure of data records at PORTAL registries and DOORS directories. Figure 2 displays a server network diagram for root, primary, and secondary DOORS directories interacting with root, primary, and secondary PORTAL registries.

Technical details of the PORTAL-DOORS paradigm are further elaborated in the publications [3], [7], [8] and at portaldoors.org. Some important characteristics include:

- A distributed network of registries and directories for resource metadata oriented by problem domain or specialist community rather than by technology format of the resource.
- A hierarchical system enabling local independence of communities while simultaneously maintaining global compatibility for communication between and search amongst different communities.
- A hybridized architecture with both XML Schemas and terminologies serving the original web and also RDF triples and OWL ontologies serving the semantic web to bridge and transition from the original web to the semantic web.
- Decentralization, distribution, and democratization to promote evolutionary adoption of componentized terminologies and ontologies (ie, survival of the fittest, not necessarily the first).
- Hierarchical authorities and globally unique identifiers to prevent namespace conflicts when identifying resources while maintaining autonomy of local communities with control over local policies.
- Designed to accomodate any resource — whether abstract or concrete, offline or online, semantic or non-semantic — with either non-semantic descriptions using tags referencing terminologies or semantic descriptions using RDF triples referencing ontologies.
- Supported with cross-references to other systems whether legacy or contemporaneous.

The PORTAL-DOORS System is *not* another attempt once again to create a so-called 'one stop shop' that claims to be the 'one and only' destination for 'all shopping needs'. In fact, the general philosophy of HDMM systems turns that notion upside down and argues that centralized 'one stop shops' cannot and will not solve the problems. Instead, there should be a multiplicity and diversity of registries and directories exchanging mobile metadata that becomes highly distributed, redistributed, and cached everywhere for the speed and efficiency of search and location which can be achieved effectively only by maintaining the interoperability of all registries and directories to communicate with each other transparently within the same infrastructure system.

IV. Infrastructure System versus Tools and Applications versus Content

PORTAL-DOORS as a lower-level infrastructure system must be distinguished from higher-level tools and applications built on the foundation of the infrastructure. PORTAL-DOORS as a mobile metadata management, communication, and distribution system must also be distinguished from

Table I
HIERARCHICALLY DISTRIBUTED MOBILE METADATA SYSTEMS WITH ENTITY REGISTERING AND ATTRIBUTE PUBLISHING

	IRIS-DNS System	PORTAL-DOORS System
Dynamic metaphor	A distributed communications network brain of nodal neurons continuously updating, exchanging, and integrating messages about 'who what where'	
Static metaphor	A simple phonebook	A sophisticated library card catalogue
Registering system	IRIS registries	PORTAL registries
— Entity registered	domain	resource
— Identified by	unique name	unique label (URI or IRI) with optional tags
Publishing system	DNS directories	DOORS directories
— Attributes published	address and aliases	location and descriptions
— Specified by	IP number	URIs, URLs, RDF triples referencing OWL ontologies
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 URL for IRIS-DNS
Serves semantic web	No (IRIS-DNS does not use RDF triples)	Yes via mapping of character label to semantic description
Crosslinks entities	No	Yes via mappings within DOORS descriptions to other resources
Crosslinks systems	No	Yes via mappings within PORTAL crossreferences to other systems

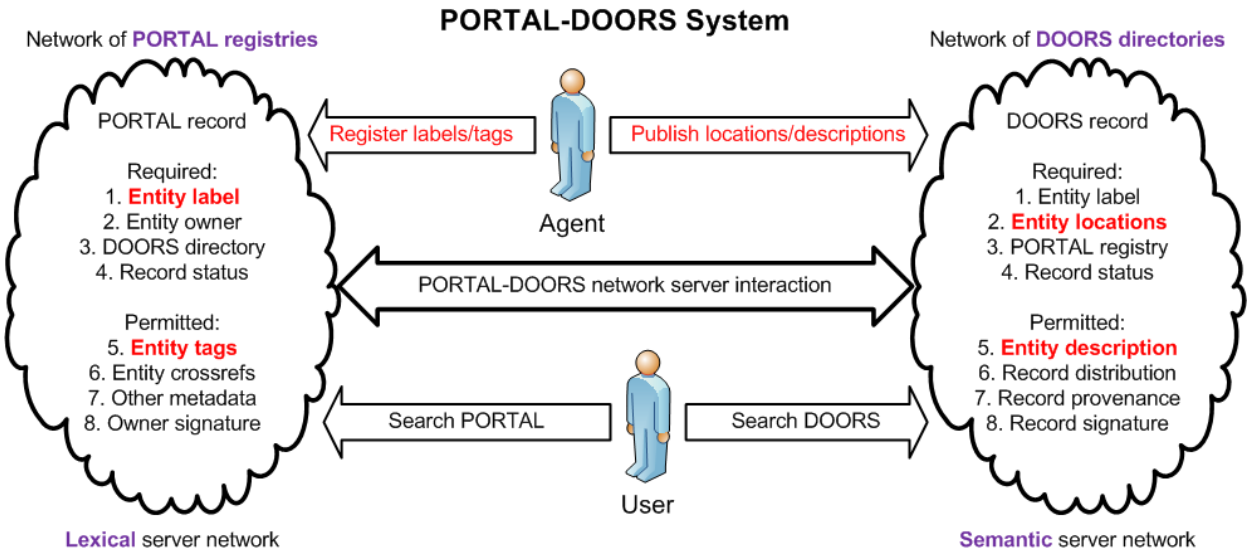

Figure 1. PORTAL-DOORS System Data Records: Resource metadata is registered and published by agents for search by users in the PORTAL-DOORS server networks. *Semantic* services here are defined as those using the RDF/OWL/SPARQL stack of technologies, whereas *lexical* services are defined as those using only character string processing, terminologies, or those XML technologies that do not require use of RDF triples. Fields within data records are considered *required* or *permitted* with respect to the schemas maintained by the root servers. The figure above displays only the most important fields; for all fields, see the reference model implemented with XML Schemas.

the actual metadata that the infrastructure is designed to send, receive, and exchange throughout the system. Fundamentally, the PORTAL-DOORS System establishes an interoperable, platform-independent, application-independent, interface standard for information exchange over the internet with a design that is guided by the HDMM architectural style, further specified to fulfill additional requirements in order to serve both the original web and semantic web as described in the 'blueprint' paper [3], and currently partially detailed in a draft reference implementation written in XML Schema *.xsd files.

Work to complete a reference implementation must clarify not only the structural data model for metadata records, but also the functional behavioral model for the PORTAL and DOORS services in response to requests from clients. Servers and clients must also communicate over transport protocols. The PORTAL-DOORS Project maintains a vision of serving more than one transport protocol as discussed in

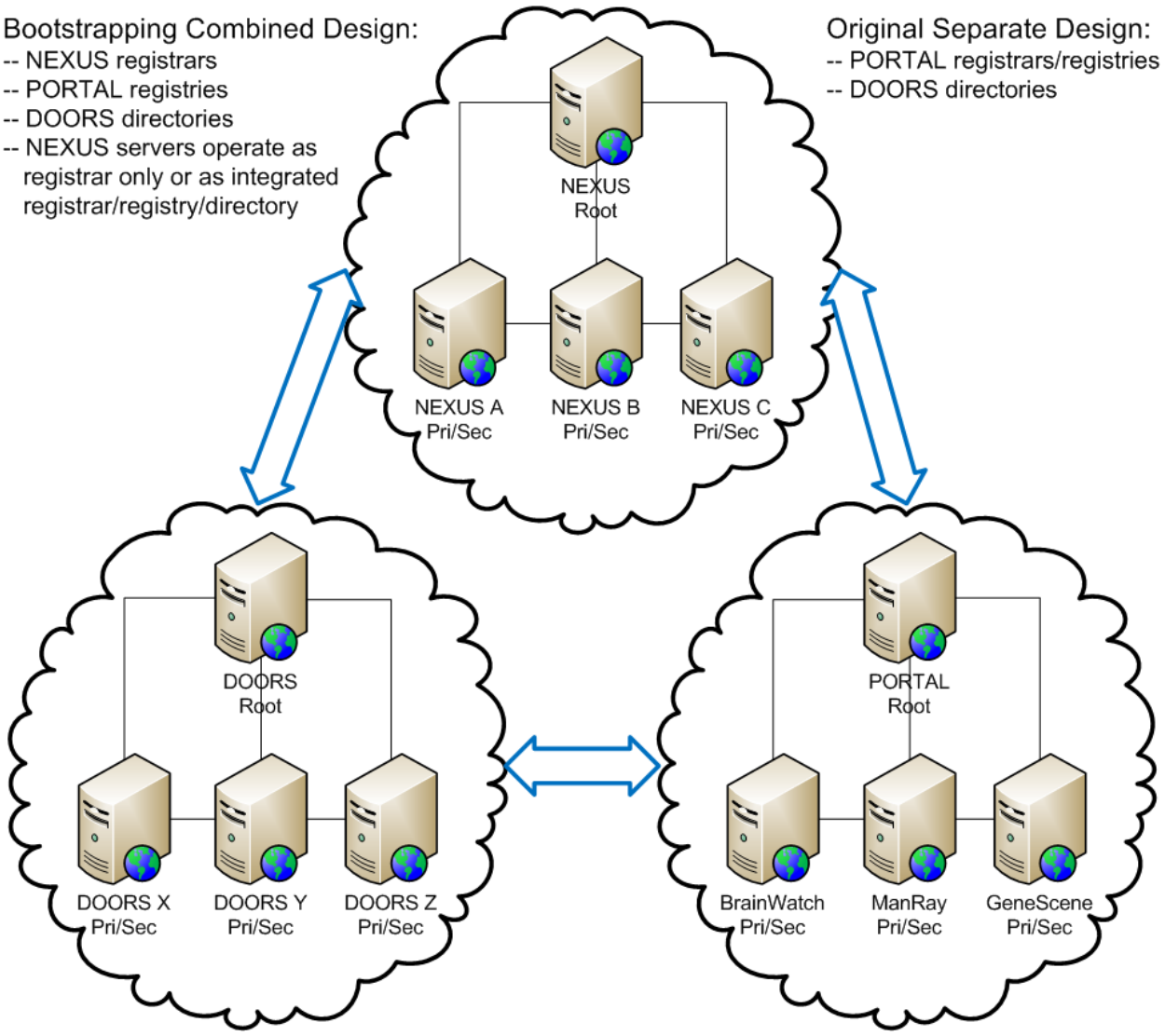

Figure 2. PORTAL-DOORS System Server Network: Resource metadata server networks for PORTAL registering of labels and tags and DOORS publishing of locations and descriptions are analogous to domain metadata server networks for IRIS registering of names and DNS publishing of addresses. Primary PORTAL registries may be established by any individual or organization which maintains any local policies governing registration of resources at that particular primary PORTAL registry. Examples shown here (GeneScene, BrainWatch, ManRay) implement policies with a problem-oriented focus on their respective specialty domains. Specific criteria for registration are determined by the local schema of the PORTAL primary which must nevertheless comply with the global requirements of the PORTAL root in order to assure interoperability between different PORTAL primaries.

Section VII.E. of [3]. Initial drafts of the PORTAL-DOORS System schema files (prior to version 0.5) assumed use of the IRIS core protocol. The current draft (version 0.5) addresses only the structural data model. The next draft (version 0.6) will re-introduce use of a specific transport protocol but replace the IRIS core protocol with an http protocol using RESTful web services. At present, in a bootstrapping stage of development for PORTAL-DOORS, RESTful web services provide a more favorable environment for promoting adoption of the system. However, a fully dedicated and optimized protocol specifically for PORTAL-DOORS may ultimately prove necessary to achieve the speed and efficiency comparable to that which exists now for IRIS-DNS.

As PORTAL-DOORS continues to be developed and implemented, any client tool, application, or web site that accesses PORTAL-DOORS must be distinguished from the system itself. The PORTAL-DOORS System should not be considered either a single site or repository any more than the IRIS-DNS System of domain name registries and directories could be construed to be a single site or repository. For both IRIS-DNS and PORTAL-DOORS infrastructure

systems, server-side data stores and services and client-side tools and applications can be written in any language on any platform. Client tools are necessary for agents to edit the information maintained at an individual server data store. Client tools are also necessary for agents and users to navigate, search and query the information stored not only at a particular server but also throughout the entire network of servers. These tools include faceted browsers, keyword search utilities, and SPARQL query interfaces.

Even more complex applications can be built in which the navigation, search, and query tools may be embedded within more sophisticated applications that hide these tools from the user interface. An important example is an application component that would provide natural language answers to natural language questions in the context of the overall function of the software application. In this example, the component converts the user's natural language question to a SPARQL query submitted to PORTAL-DOORS, and then converts the query response from PORTAL-DOORS back to a natural language answer for presentation to the user.

V. General Usage Scenarios for the PORTAL-DOORS System

PORTAL-DOORS has been designed to be as flexible as possible with both backward and forward compatibility from Web 1.0 to Web 3.0. Given the partition with lexical non-semantic services on the PORTAL side and semantic services (with use of the RDF/OWL/SPARQL stack) on the DOORS side, and also the partition with both *required* and *permitted* elements for each of PORTAL and DOORS, there are many possible scenarios for usage of the entire PORTAL-DOORS System. Some examples include:

- Minimal use of required elements for both PORTAL registries and DOORS directories: This scenario essentially reduces use of the system to an alternative equivalent to PURLs [9] (and other similar services). However, it does so *without* requiring use of a predetermined URL identifier root like purl.oclc.org and instead allowing use of any identification scheme as long as it is a URI or IRI.
- Maximal use of permitted elements for PORTAL registries but minimal use of required elements for DOORS directories: This scenario enables exploiting the full metadata management facilities of the PORTAL non-semantic services (which include provisions for tags, micro-formats, cross-references, etc) without any obligation to use the DOORS semantic services (that necessitate use of the RDF/OWL/SPARQL stack of technologies and tools). This scenario enables resource agents to publish metadata now in non-semantic formats and defer until later any possible transition to semantic formats which would then be facilitated by the prior staging in the non-semantic formats.
- Minimal use of required elements for PORTAL registries but maximal use of permitted elements for DOORS directories: This scenario serves those situations where there is no barrier to transition the metadata from original web formats to semantic web formats, and the resource owner and agent do not wish to maintain the metadata in both semantic and non-semantic formats. This scenario requires that the resource agent registering and publishing the metadata already has access to established ontologies that can be referenced by semantic tools for describing the resource.
- Maximal use of permitted elements for both PORTAL registries and DOORS directories: This usage scenario provides the significant benefit of exposing as much metadata as possible to as many clients as possible including both older non-semantic as well as newer semantic tools and applications.

Enabling these usage scenarios constitutes an important goal for the PORTAL-DOORS Project which also includes the following tasks:

- Complete development of a specification model for the PORTAL-DOORS System as the interoperable informatics infrastructure using the Hierarchically Distributed Mobile Metadata (HDMM) architectural style for a distributed network of registries and directories.
- Complete implementation of a reference model with XML Schemas for the interoperable communication interface standards and with RESTful web services for the transport protocol.
- Build open source software clients and servers for multiple platforms, operating systems and programming languages according to the detailed roadmap (see Sec. VII) for continuing development of the previously published designs and prototypes.

The PORTAL-DOORS Project for development of the PORTAL-DOORS System thus serves to build the necessary foundation, the core infrastructure for an information-seeking support system [10] upon which higher-level applications are constructed.

VI. Specific Use Cases for the PORTAL-DOORS System

The original PORTAL-DOORS 'blueprint' paper [3] discussed the following use cases:

- Assisting with organization of the "bioinformatics resourceome" and the description, discovery and use of resources for e-science and e-medicine in health care and life sciences (see [3] Sec. III).
- Cataloguing resources for biomedical computing (see [3] Sec. IV and VIII).
- Cataloguing patents and trademarks and relating them to products and services for e-business (see [3] Sec. IX).
- Assisting with semantic search, decision support and knowledge management applications in translational

research and drug discovery for personalized medicine (see [3] Sec. XI).

More detailed descriptions of examples in the context of biomedical translational research include the following use cases of PORTAL-DOORS as an information-seeking support system for:

- Pharmacogenomic molecular imaging [11].
- PET and SPECT brain imaging [12].

Although originally conceived and described in the context of health care and life sciences, the diversity of possible use cases for PORTAL-DOORS remains as universal as the diversity of possible use cases for IRIS-DNS.

VII. Development Roadmap for the PORTAL-DOORS System

Current plans envision following a PORTAL-DOORS Project roadmap with iterative software development for the PORTAL-DOORS System with these milestones:

- Version 0.5: Current live implementation with back-end database and front-end web browser client for partial PORTAL server functionality and partial DOORS server functionality.
- Version 0.6: Implementation as RESTful web services with both ASP.Net based clients enhanced with "user-friendly" graphical user interfaces and editors for managing (entering and updating) data records at PORTAL-DOORS servers on Microsoft Windows platforms.
- Version 0.7: Implementation as RESTful web services with JAVA based servers and clients for Linux and Mac OS X platforms.
- Version 0.8: Completion and revision of lexical PORTAL functionality (including terminology tools) for all platforms.
- Version 0.9: Completion and revision of semantic DOORS functionality (including ontology tools) for all platforms.
- Version 1.0: Official release of PORTAL-DOORS System models and schemas for an authoritative server at a single site.
- Version 2.0: Multi-site functionality (including security) for distributed interacting authoritative servers.
- Version 3.0: Multi-site functionality (including provenance) for distributed interacting non-authoritative servers operating with request forwarding and response caching amongst the distributed servers.

VIII. Conclusion

As part of an ongoing iterative re-assessment and revision of the architectural design for the PORTAL-DOORS System, the architectural style common to both PORTAL-DOORS and IRIS-DNS as pervasive registry-directory networks for 'who what where' metadata management, respectively for the semantic web and original web, has been named the Hierarchically Distributed Mobile Metadata (HDMM) style. This HDMM style has been characterized with a description of the design principles and constraints that define it. Moreover, the current status and future plans of the PORTAL-DOORS System and the PORTAL-DOORS Project have also been detailed in this report.

References

[1] R. N. Taylor, N. Medvidovic, and E. M. Dashofy, *Software Architecture: Foundations, Theory, and Practice*. Wiley, 2009.

[2] R. T. Fielding and R. N. Taylor, "Principled design of the modern web architecture," *ACM Trans. Internet Technol.*, vol. 2, no. 2, pp. 115–150, 2002.

[3] C. Taswell, "DOORS to the semantic web and grid with a PORTAL for biomedical computing," *IEEE Trans Inform Technol Biomed*, vol. 12, no. 2, pp. 191–204, Mar 2008, in the "Special Issue on Bio-Grid".

[4] A. Newton and M. Sanz, "RFC3981: IRIS: The internet registry information service (IRIS) core protocol," IETF RFC 3981 Working Draft, Jan 2005. [Online]. Available: http://www.ietf.org/rfc/rfc3981.txt

[5] P. V. Mockapetris, "STD13 RFC1035: Domain names - implementation and specification," IETF Full Standard, Nov 1987. [Online]. Available: http://www.ietf.org/rfc/rfc1035.txt

[6] Virtuoso universal server. OpenLink Software, Inc. Burlington, MA. [Online]. Available: http://www.openlinksw.com/

[7] C. Taswell, "Implementation of prototype biomedical registries for PORTAL-DOORS," in *Proceedings of the American Medical Informatics Association Summit on Translational Bioinformatics, San Francisco, CA*, Mar 2009, AMIA-0036-T2009.

[8] ——, "Alternative bootstrapping design for the portal-doors cyberinfrastructure with self-referencing and self-describing features," in *Semantic Web*. Vienna, Austria: IN-TECH Publishing, 2009, in press.

[9] K. Shafer, S. Weibel, E. Jul, and J. Fausey, "Introduction to persistent uniform resource locators," Mar 1996. [Online]. Available: http://purl.oclc.org/docs/inet96.html

[10] G. Marchionini and R. W. White, "Information-seeking support systems [guest editors' introduction]," *IEEE Computer*, vol. 42, no. 3, pp. 30–32, 2009.

[11] C. Taswell, "PORTAL-DOORS infrastructure system for translational biomedical informatics on the semantic web and grid," in *Proceedings of the American Medical Informatics Association Summit on Translational Bioinformatics, San Francisco, CA*, Mar 2008, p. 43.

[12] ——, "Biomedical informatics for brain imaging and gene-brain-behavior relationships," in *W3C Semantic Web HCLSIG F2F Meeting 4/30-5/1*, 2009. [Online]. Available: http://esw.w3.org/topic/HCLSIG/Meetings/2009-04-30_F2F