

# Managing Scientific Literature with Software from the PORTAL-DOORS Project

Shiladitya Dutta, Pooja Kowshik, Adarsh Ambati,  
Sathvik Nori, S. Koby Taswell, and Carl Taswell

Brain Health Alliance, Ladera Ranch, California, USA

IEEE eScience 15th BC2DC'19 Workshop, San Diego, California  
24 September 2019



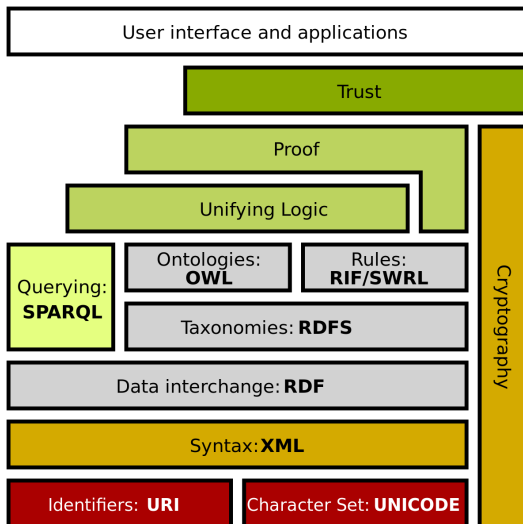
# Introduction to the PORTAL-DOORS Project

- PORTAL-DOORS operates as a distributed diristry, registry, directory network system of metadata and data repositories [1]
- PORTAL-DOORS for the semantic web modeled on the success of IRIS-DNS for the original lexical web [2]
- PORTAL-DOORS addresses diverse problems: information tsunami (needle in a haystack), informatics tower of babel (harmonization for interoperability), and cybersilos in scientific discourse
- As a cyberinfrastructure system for the semantic web, it has a universal design with applicability to any and all fields in a manner analogous to IRIS-DNS for the lexical web
- The Nexus-PORTAL-DOORS-Scribe (NPDS) cyberinfrastructure continues to be developed as a software implementation of the original PORTAL-DOORS Project principles

# Semantic versus Lexical Information Systems

- A lexical (“dumb”) system is an information system in which words are processed as character strings that have no meaning to the processing agent
- A semantic (“smart”) system is one in which words have defined meaning to the agent processing them with logic-based reasoners [3]
- Semantic search may be efficient, while lexical search inefficient, for the given search task:
  - How many hits returned in response to the search query?
  - Are there too many hits for a person to review?
  - Or if just a few hits, are they relevant?
  - Do the returned hits answer the search question directly or indirectly?
- Semantic information systems can be built with the XML, RDF, OWL, SPARQL stack of technologies for describing and querying resources

# W3C Semantic Web Stack



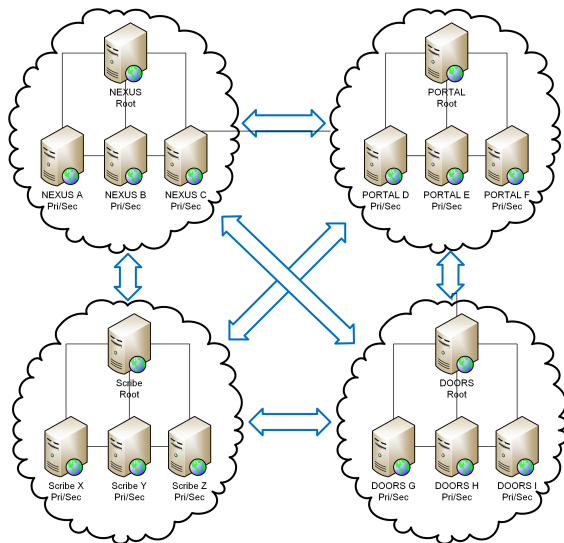
## PORTAL-DOORS compared to IRIS-DNS

- PORTAL-DOORS for the semantic web has been designed in a manner analogous to IRIS-DNS for the lexical web
- PORTAL (Problem-Oriented Registry of Tags And Labels) is an analogue of IRIS for naming and registering domains
- DOORS (Domain-Ontology Oriented Resource System) is an analogue of DNS for addressing and locating domains
- Using an analogous paradigm with labeled resources instead of named domains, PORTAL-DOORS designed to do for the semantic web what IRIS-DNS does for the lexical web
- PORTAL-DOORS built as a who-what-where diristry, registry, directory network system for identifying, describing, locating and linking things on the internet, web and grid

# NPDS Server Network

- NPDS consists of 4 primary server types: Nexus, PORTAL, DOORS, and Scribe:
  - ① *Nexus diristries* (a diristry = directory + registry) operate as integrated combinations of virtual PORTAL and DOORS servers available on the same physical server
  - ② *Lexical Problem Oriented Registry of Tags And Labels (PORTAL) registries* register resources and publish labels and tags for the resources
  - ③ *Semantic Domain Ontology Oriented Resource System (DOORS) directories* publish locations and descriptions of the resources with the mapping of labels to locations for the resources
  - ④ *Scribe registrars* provide read-write access to data records for servers that support the read-only Nexus-PORTAL-DOORS services
- Communicate with each other using RESTful API interfaces [4]

# NDPS Server Network Overview



# DREAM Principles

- In 2006, PDP specified a comprehensive collection of foundational principles including identification, location, description, accessibility, interoperability, flexibility, extensibility, democratizability, provenance, distribution with a set of required, permitted and optional elements
- In 2019, PDP renamed these principles the DREAM principles with the phrase *Discoverable Data with Reproducible Results for Equivalent Entities with Accessible Attributes and Manageable Metadata*
- The *Equivalent Entities* principle remains essential for promoting reproducibility and preventing plagiarism: “science will be neither reproducible nor fair without recognition, acknowledgment, attribution and citation of equivalent entities”
- We must be able to ask: “What is same and what is different?”



# HDMM Cyberinfrastructure Principles

- The distributed network of servers in the NPDS cyberinfrastructure has been designed in accordance with the Hierarchically Distributed Mobile Metadata (HDMM) style of architecture [1], [5]
- Acts as a hybrid of hierarchical and peer-to-peer servers to assure that *who what where* information can be transmitted and exchanged both efficiently and effectively. HDMM has 5 main concepts:
  - 1 *Distributed infrastructure*: Pervasively distributed and shared infrastructure, content, and control
  - 2 *Hierarchical authorities*: A hierarchy of both authoritative and non-authoritative servers
  - 3 *Mobile metadata*: A focus on moving the mobile metadata for who what where as fast as possible
  - 4 *Separated concerns*: A separation of concerns with registries for identifying resources and directories for locating resources
  - 5 *Unrestricted identification*: Freedom of choice to select identifiers

# Implementation of NPDS

- PDP has prototyped, developed, and tested servers for the NPDS cyberinfrastructure which includes PORTAL registries, DOORS directories, Nexus directories, and Scribe registrars with front-end web applications and back-end databases managed by both SQL and NoSQL database systems.
- DREAM principles from PDP are implementation independent and adaptable to whichever software technology stack an organization chooses. Three implementations are currently being developed:
  - *PDP-Aoraki*: Microsoft stack that is C# centric, and built with IIS Server, SQL Server, and .NET Core
  - *PDP-Meru*: MEAN stack that is JavaScript centric, and built with MongoDB, Express, Angular and Node
  - *PDP-Zunil*: Python centric stack with the Django web framework
- PDP-Aoraki is the software stack currently operational at [www.portaldoors.org](http://www.portaldoors.org), and demonstrated in the screenshot images and video available with this presentation

# Registrars and Diristries

## BHA Nexus Diristries

Nexus diristries managed by BHA Scribe Registrar include:

- [BrainWatch Diristry](#) for Brain Health
- [Eywa Diristry](#) for Biodiversity and Wildlife Conservation
- [Gaia Diristry](#) for Biosurveillance and Environmental Protection
- [HELPME Diristry](#) for Health Education Law Public Policy and Medical Ethics
- [SOLOMON Diristry](#) for Sensory-Onset, Language-Onset, Motor-Onset Dementias

## GTG Nexus Diristries

Nexus diristries managed by GTG Scribe Registrar include:

- [Beacon Diristry](#) for Biomedical Computing
- [CTGaming Diristry](#) for Clinical TeleGaming
- [GeneScene Diristry](#) for Genetic Medicine
- [ManRay Diristry](#) for Nuclear Medicine
- [Osler Diristry](#) for Personalized Medicine

## PDP Nexus Diristries

Nexus diristries managed by PDP Scribe Registrar include:

- [Avicenna Diristry](#) for clinical trial analysis
- [DaVinci Diristry](#) for biomedical data integration

PDP web sites © 2007 - 2019 [PORTAL-DOORS Project \(PDP\)](#)  
8 Gilly Flower Street, Ladera Ranch CA 92694 USA. Tel. +1(949)481-3121.

# Entity Types

Organization	any named organization	true	true	true	true
Person	any named person	true	true	true	true
OfflineRealEntity	any offline or real resource	true	true	true	true
PhysicalObject	a real world thing	true	true	true	true
ChemicalSubstance	a real world substance	true	true	true	true
BiologicalBeing	a real world life form	true	true	true	true
GeographicLocation	a real world place	true	true	true	true
OnlineVirtualEntity	any online or virtual resource	true	true	true	true
TerminologyItem	an item in a controlled terminology	false	false	true	true
TaxonomyItem	an item in a controlled taxonomy	false	false	true	true
ThesaurusItem	an item in a controlled thesaurus	false	false	true	true
DataRecord	datum in a data repository (record in database)	true	true	true	true
ComputingTool	client-based software tool/application for processing data	true	true	true	true
ComputingService	server-based online service for processing data	true	true	true	true
Publication	any publication including books, articles, papers, etc	true	true	true	true
AudioItem	audible sound item	true	true	true	true
ImageItem	static visible image item	true	true	true	true
VideoItem	dynamic visible image sequence	true	true	true	true
MultiMediaItem	mixed (text/audio/image/video) multimedia item	true	true	true	true
OneDimSequence	any one dimensional sequence of symbols	true	true	true	true
MultiDimSignal	any multidimensional signal, dynamic volumetric imaging, etc	true	true	true	true
MetaResourceEntity	resource that describes another resource	true	true	true	true

# Editing Records as an Author

## Edit Author's Avicenna Resource Metadata Records on Nexus Server

The screenshot shows the Nexus Server interface for editing metadata records. The interface includes a header with navigation options and a main table of records. A red box highlights the 'Update' and 'Cancel' buttons for a record in the 'EntityLabels' tab.

**EntityLabels Tab:**

Priority	Marked	Restricted	Principal	Phone Number	Email Address	Web Address	Street Address	City / Locality	State / Region	Country	Postal Code	Update
0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>									<input type="button" value="Update"/> <input type="button" value="Cancel"/>

**Main Table:**

ID	Priority	Marked	Restricted	Principal	Phone Number	Email Address	Web Address	Street Address	City / Locality	State / Region	Country	Postal Code	Update
NB498CBB9	Pending	Pending	ComputingTool	M40451F09B3	DotNetRDF	DotNetRDF is a general purpose triple store developed in Java. However, it included numerous other add-ons such as an RDF editor tool and a SPARQL utility. These are accessible as independent applications.							<input type="button" value="Edit"/> <input type="button" value="Validate"/>
X9EB27E60	Pending	Pending	ComputingTool	Y9FBBE8BA81	FluentEditor	FluentEditor is an ontology builder developed by Cognitum. It is downloadable free and includes the ability to parse RDF/XML files.							<input type="button" value="Edit"/> <input type="button" value="Validate"/>

# Viewing and Requesting Authorship as an Agent

## View Agent's DaVinci Resource Metadata Records on Nexus Server

Handle	PORTAL	DOORS	Authorship	Type	Tag	Name	Nature
▶ MB259F7EB	Invalid	Valid	Request	Publication	Q39846B3F64	A Comparative Analysis and Evaluation of Bibliographic Ontologies	Analysis of bibliographic ontologies by Maria Teresa Biagetti
▶ V4B38DA73	Invalid	Valid	Request	Publication	ACUO	A Comparison of Upper Ontologies.	A summary of some existing comparisons drawn among subsets of the 7 Upper Ontologies.
▶ CA46A02DF	Valid	Valid	Request	Publication	C8CECB0E315	A Developer's Guide to the Semantic Web, 2nd Ed.	Book by Liyang Yu
▶ TEE52B39F	Invalid	Valid	Request	Publication	CEA1C86413A	A Distributed Infrastructure for Metadata about Metadata: The HDMM Architectural Style and PORTAL-DOORS System	Dr. Taswell's 2010 Future Internet Paper
▶ UC9BE3742	Valid	Valid	Request	Publication	MetaMetaNPDS	A Distributed Infrastructure for Metadata about Metadata: The HDMM Architectural Style and PORTAL-DOORS System.	2010 NPDS revision paper by Carl Taswell, outlining the updates made to PORTAL-DOORS System.
▶ VFD278E64	Pending	Pending	Request	Publication	WEA9685C737	A fast and accurate dependency parser using neural networks	Paper describing the dependency parsing method used by the StanfordCoreNLP engine. Dependency parsing is the process to create a tree-based representation of a sentence.
▶ G89543AD2	Valid	Valid	Request	Publication	W31953994EE	A framework for mining Life Sciences data on the Semantic	Paper that describes an implementation of how Semantic Web technologies offer both a means to define rich semantics

# Web Service API Example

```
<?xml version="1.0" encoding="UTF-8" xmlns="http://npds.portalddoors.org/nsv0/npdsystem#" Version="1.1.0 (2018-10-14)">
<!--
Nexus-PORTAL-DOORS System (c) 2006-2018 Carl Taswell and the PORTAL-DOORS Project Team
-->
<Message>
  <ClientRequest>
    <URL>
      http://npds.bhahealth.net/Nexus/Beacon?ef=1&vf=1&lc=11
    </URL>
    <Note/>
    <Question/>
  </ClientRequest>
  <ServerResponse>
    <Status/>
    <Note/>
  </ServerResponse>
  <Answer>
    <NexusService Count="11">
      <NexusResRep>
        <EntityMetadata>
          <Name>ABMS List of Member Boards</Name>
          <Nature/>
          <CanonicalLabel EntityType="OnlineVirtualEntity">http://npds.bhahealth.net/nexus/helpme/abmslmb</CanonicalLabel>
          <AliasLabels Count="11">
            <AliasLabel Priority="255">http://npds.bhahealth.net/nexus/helpme/dbaaale2c75</AliasLabel>
          </AliasLabels>
          <PrincipalTag>ABMSLMB</PrincipalTag>
          <SupportingTags Count="0"/>
          <SupportingLabels Count="1">
            <SupportingLabel Priority="0" IsRestricted="false">http://pds.bhahealth.net/helpme/abms</SupportingLabel>
          </SupportingLabels>
          <CrossReferences Count="0"/>
          <OtherEntity/>
          <OtherTexts Count="0"/>
          <Contact/>
          <Owner/>
          <Locations Count="1">
            <Location Priority="0">
              <anyXmlTag> WARNING: valid XML format required </anyXmlTag>
            </Location>
          </Locations>
          <Descriptions Count="0"/>
        </EntityMetadata>
        <RecordMetadata IsCachedCopy="false">
          <CreatedOn>2013-01-03T10:49:32Z</CreatedOn>
          <UpdatedOn>2013-01-03T15:00:05Z</UpdatedOn>
          <Signature/>
          <Distributions Count="0"/>
          <Provenances Count="0"/>
        </RecordMetadata>
      </NexusResRep>
    </NexusService>
  </Answer>
</Message>
</-->
</-->
```

## FAIR Metrics

- In support of the *Equivalent Entities* principle, PDP has developed the FAIR metrics with the acronym FAIR for the *Fair Attribution to Indexed Reports* and *Fair Acknowledgment of Information Records*
- Rewards good citation practices while highlighting instances where authors fail to cite properly the relevant literature [6]
- Categorical analysis with 4 types of research article claims derived from a 2x2 table with old versus new and valid versus invalid

Claim	Valid	Invalid
Old	Quoted	Misquoted
New	Novel	Plagiarized

- An *invalid new claim* may exist in the presence of paraphrasing without citing, plagiarism of ideas, or verbatim plagiarism
- Counts of the claim types are used to calculate the 4 ratio metrics ( $F_1$ ,  $F_2$ ,  $F_3$ , and  $F_4$ ) measuring different kinds of citation practices



# Future of NPDS

- NPDS has been designed as a hybrid bridge between the lexical and semantic webs, thereby creating a platform for a diverse variety of different computerized approaches and applications with artificial intelligence on the semantic web
- Currently, the PDP development team continues to build the following automated utilities for the NPDS cyberinfrastructure:
  - *Content Management Systems (CMS)*: A general purpose CMS with content formatted for semantic analysis built on the foundation of the NPDS cyberinfrastructure; and a customised CMS built on the NPDS foundation for use with peer review and publishing of scholarly research
  - *Citation Format Converter Utilities*: A convenient set of tools that convert citations of references in NPDS format back and forth to the prevailing standard formats for citing literature references
  - *Concept-Validating Search Engine Agent (CoVaSEA)*: An automated query engine and search agent that interoperates with NPDS and populates NPDS directories with metadata records about resources

# Brainiacs Journal

- These utilities along with the FAIR metrics will support development of our new journal called the Brainiacs Journal of Brain Imaging and Computing Sciences ([www.BrainiacsJournal.org](http://www.BrainiacsJournal.org)).
- The customized version of our CMS will provide the applications and tools required for the Brainiacs Journal
- We plan to integrate our FAIR metrics into the manuscript submission, peer and editorial review process in order to evaluate the citation practices of submitted manuscripts
- All references cited by submitted papers will be cataloged in Nexus directories of the relevant subject area for use with CoVaSEA and FAIR metrics in the peer review process for the Brainiacs Journal

## Conclusion

- The PORTAL-DOORS Project (PDP) has developed software for the NPDS cyberinfrastructure in support of identifying, describing, locating and linking things on the internet, web and grid
- The PDP foundational principles published in 2007 have been renamed the DREAM principles in 2019
- NPDS consists of Nexus directories, PORTAL registries, DOORS directories, and Scribe registrars
- The *Equivalent Entities* principle remains an essential DREAM principle for promoting reproducibility and preventing plagiarism
- FAIR metrics have been designed as a family of metrics to promote good citation practices
- Applications and utilities, such as a CMS and citation converter, are being built to support the launch of the new Brainiacs Journal

# Bibliography



C. Taswell, "A distributed infrastructure for metadata about metadata: The HDMM architectural style and PORTAL-DOORS system," *Future Internet*, vol. 2, no. 2, pp. 156–189, 2010, In Special Issue on Metadata and Markup., ISSN: 1999-5903. DOI: [10.3390/fi2020156](https://doi.org/10.3390/fi2020156). [Online]. Available: [www.mdpi.com/1999-5903/2/2/156/](http://www.mdpi.com/1999-5903/2/2/156/).



—, "DOORS to the semantic web and grid with a PORTAL for biomedical computing," eng, *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 2, pp. 191–204, 2 Mar. 2008, In the Special Section on Bio-Grid published online 3 Aug. 2007, ISSN: 1089-7771. DOI: [10.1109/TITB.2007.905861](https://doi.org/10.1109/TITB.2007.905861).



T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American*, vol. 284, no. 5, pp. 34–43, May 2001.



A. Craig, S. H. Bae, T. Veeramacheni, *et al.*, "Web service APIs for Scribe registrars, Nexus directories, PORTAL registries and DOORS directories in the NPD system," in *Proceedings 9th International SWAT4LS Conference*, Amsterdam, Netherlands, 2016. [Online]. Available: [ceur-ws.org/Vol-1795/paper4.pdf](http://ceur-ws.org/Vol-1795/paper4.pdf).



C. Taswell, "The hierarchically distributed mobile metadata (HDMM) style of architecture for pervasive metadata networks," in *Proceedings ISPAN 10th International Symposium on Pervasive Systems, Algorithms and Networks*, IEEE, Dec. 2009, pp. 315–320. DOI: [10.1109/i-span.2009.73](https://doi.org/10.1109/i-span.2009.73).



A. Craig and C. Taswell, "The FAIR metrics of adherence to citation best practices," in *Proceedings ASIS&T 81st Annual Meeting SIGMET Workshop*, Vancouver BC, Canada, Nov. 10, 2018. [Online]. Available: [www.portaldoors.org/pub/docs/SigMet2018ACCT1108poster.pdf](http://www.portaldoors.org/pub/docs/SigMet2018ACCT1108poster.pdf).

## For More Info

- [www.PORTALDOORS.org](http://www.PORTALDOORS.org)
- [www.BrainHealthAlliance.org](http://www.BrainHealthAlliance.org)
- [ctaswell@BrainHealthAlliance.org](mailto:ctaswell@BrainHealthAlliance.org)
- We welcome collaborators interested in contributing to the PORTAL-DOORS Project and NPDS cyberinfrastructure

## For More Information

- DREAM Principles and FAIR Metrics from the PORTAL-DOORS Project for the Semantic Web – *Presented June 2019, 11th IEEE ECAI*  
[portalddoors.org/pub/docs/ECAI2019DREAMFAIR0612.pdf](http://portalddoors.org/pub/docs/ECAI2019DREAMFAIR0612.pdf)
- Managing Scientific Literature with Software from the PORTAL-DOORS Project – *Presented September 2019, 15th IEEE eScience*  
[portalddoors.org/pub/docs/BCDC2019PdpDemo0817.pdf](http://portalddoors.org/pub/docs/BCDC2019PdpDemo0817.pdf)
- Definitions, Formulas, and Simulated Examples for Plagiarism Detection with FAIR Metrics – *To Be Presented October 2019, 82nd ASIS&T*  
[portalddoors.org/pub/docs/ASIST2019FairMetrics0611.pdf](http://portalddoors.org/pub/docs/ASIST2019FairMetrics0611.pdf)
- DREAM Principles from the PORTAL-DOORS Project and NPDS Cyberinfrastructure – *Submitted, under peer review*
- PDP software demo video available at  
[portalddoors.org/pub/mp4/PdpDemoVideo20190924.zip](http://portalddoors.org/pub/mp4/PdpDemoVideo20190924.zip)

## 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

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

Manuscript received 10/31/2006; online 8/3/2007; print 3/5/2008