Managing Scientific Literature with Software from the PORTAL-DOORS Project

1st Shiladitya Dutta
Brain Health Alliance
Ladera Ranch, CA, USA
sdutta@bhavi.us

2nd Pooja Kowshik
Brain Health Alliance
Ladera Ranch, CA, USA
pkowshik@bhavi.us

3rd Adarsh Ambati
Brain Health Alliance
Ladera Ranch, CA, USA
aambati@bhavi.us

4th Sathvik Nori
Brain Health Alliance
Ladera Ranch, CA, USA
snori@bhavi.us

5th S. Koby Taswell
Brain Health Alliance
Ladera Ranch, CA, USA
ktaswell@bhavi.us

6th Carl Taswell
Brain Health Alliance
Ladera Ranch, CA, USA
ctaswell@bhavi.us

***Abstract*—Scholarly research associated with finding and citing scientific literature in the 21st century requires new approaches to address the continuing problems that occur with the provenance of content in the literature as well as the peer and editorial review process for publishing this literature. The PORTAL-DOORS Project (PDP) has developed software for the Nexus-PORTAL-DOORS-Scribe (NPDS) cyberinfrastructure in support of identifying, describing, locating and linking things on the internet, web and grid with both lexical and semantic tools and applications. This presentation of our PDP software will highlight *Discoverable Data with Reproducible Results for Equivalent Entities with Accessible Attributes and Manageable Metadata* with the DREAM principles, and the *Fair Acknowledgment of Information Records* also called the *Fair Attribution to Indexed Reports* with the FAIR metrics. This software demonstration will explain use of the network of metadata repositories for scientific literature accessible from www.portaldoors.org, and use of the open source software that powers the NPDS cyberinfrastructure, PDP web sites and PDP web services. Our PDP software for the NPDS cyberinfrastructure will be released publicly at this presentation of the software where we will also discuss challenges in the peer review process that include plagiarism detection.**

***Index Terms*—DREAM principles, NPDS cyberinfrastructure, FAIR metrics, bibliometrics, peer review, semantic web.**

I. History of NPDS

Berners-Lee *et al.* defined the semantic web as “An extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation” [1]. However, this proposed semantic web remains underdeveloped as a result of insufficient content that has been published in the necessary semantic formats. To facilitate a transition from the original lexical web to the planned semantic web, the PORTAL-DOORS Project (PDP) began in 2006 focusing on the design and development of the Nexus-PORTAL-DOORS-Scribe cyberinfrastructure [2]–[4]. The lexical web uses the Internet Registry Information System (IRIS) to register domain names along with the Domain Name System (DNS) to translate human-readable domain names to computer-readable IP addresses. The NPDS cyberinfrastructure has been designed in a manner analogous to IRIS-DNS with a separation of concerns for the Problem Oriented Registries of Tags And Labels (PORTAL) as lexical registries acting as a registering system for identifying, tagging and labeling entities and the Domain Ontology Oriented Resource System (DOORS) as semantic directories acting as a publishing system for locating and describing entities [2].

Subsequent developments since the original PORTAL-DOORS design [2] have included the creation of Nexus diristries [3] that act as a combination of the operations of both PORTAL registries and DOORS directories, and the creation of Scribe registrars that separate read-write services at the registrars from the read-only services at the registries, directories and diristries [4]. Some examples of Nexus diristries include the BrainWatch diristry focused on the field of brain imaging, Avicenna on the field of clinical trial methods, and the SOLOMON diristry on dementia and neurodegenerative disorders [5]. Supported by a not-for-profit organization, PDP maintains the goal of building the NPDS cyberinfrastructure as a distributed network of open access repositories of data and metadata with sufficient flexibility, extensibility and decentralization to avoid and prevent monopolistic control by any single for-profit organization, thus to assure availability of information accessible to all.

Throughout development since design, PDP has prototyped, built, and tested servers for the NPDS cyberinfrastructure which includes PORTAL registries, DOORS directories, Nexus diristries, and Scribe registrars with front-end web applications and back-end databases managed by both SQL and NoSQL database systems. PDP articulated a collection of foundational principles for the design of the NPDS cyberinfrastructure [2], [3]. The PDP principles were recently renamed the DREAM principles [6] as an acronym for the phrase *Discoverable Data with Reproducible Results for Equivalent Entities with Accessible Attributes and Manageable Metadata* (see Table I). Craig *et al.* [6] analyzed the similarity of the 2016 Wilkinson *et al.* FAIR principles [7] to the 2007 Taswell PDP principles [2], found a complete overlap with an item-by-item match between all of the FAIR principles with the previously published

Table I
SUMMARIZING PHRASES FOR THE DREAM PRINCIPLES AS THE FOUNDATION FOR THE PORTAL-DOORS PROJECT

Acronym	**Main Phrase**	**Alternative Phrases**			
D	Discoverable Data	Distributed	Data	Discoverable	Diristries
R	(with) Reproducible Results	Repositories	Records	Resources	(of) Reusable
E	(for) Equivalent Entities	(for) Enhanced	(for) Extensible	(for) Equivalent	Entity
A	(with) Accessible Attributes	Accessible	Accurate	Available	Attribute
M	(and) Manageable Metadata	Metadata	Metadata	Metadata	Metadata

PDP principles, and concluded that the FAIR principles were paraphrased from the PDP principles without citing any of the original sources [2], [3], [8]. Boeckhout *et al.* [9] posed the question whether "the FAIR guiding principles of data stewardship [were] fair enough?" Craig *et al.* [6] asked the simpler question whether the co-authors of the FAIR principles were fair?

Taswell [2] published the PDP principles almost a decade before Wilkinson *et al.* [7] paraphrased them as the FAIR principles. Key co-authors of [7] were aware of and knew about [2]. At least six of the fifty-three co-authors of [7] attended a scientific conference [10] in 2009 where direct face-to-face discussion occurred with conversations between the paraphrasing co-authors of [7] and the paraphrased author of [2]. Considering the sequence of events with these face-to-face conversations about PDP documented by the W3C 2009 F2F meeting attendance [10] and PDP presentation slides [11], and then the subsequent failure by Wilkinson *et al.* [7] to cite Taswell [2], we emphasize that *science will be neither reproducible nor fair without recognition, acknowledgment, attribution and citation of equivalent entities* regardless of whether those equivalent entities are considered to be scientific hypotheses, scientific experiments, scientific data, scientific results or published articles in the scientific literature.

Therefore, the most important of the DREAM principles should be those related to the reproducibility of equivalent entities and those embodied in the phrases *Fair Attribution to Indexed Reports* and *Fair Acknowledgment of Information Records*. We have implemented this principle of citation of equivalent entities with our use of the FAIR family of numerical quantitative metrics [12] for the promotion of truly fair citation and the prevention of plagiarism including the plagiarism of ideas. Moreover, the DREAM principles and FAIR metrics from PDP have been implemented in a variety of different software platforms and frameworks.

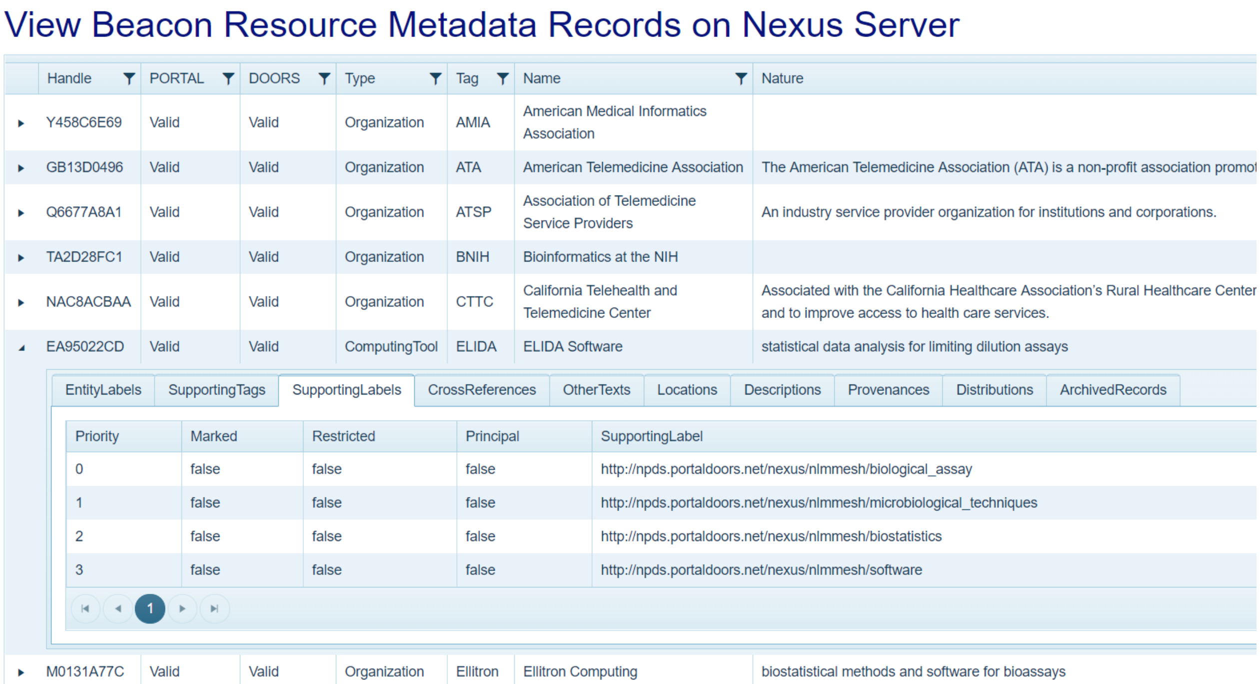
PDP maintains both an implementation for the Microsoft stack built with IIS Server, SQL Server, C# and .NET Core as well as a MEAN stack implementation that is JavaScript centric built with MongoDB, Express, Angular and Node. Another software stack, one that is Python centric, is now also under development. These three different implementations of NPDS with different software stacks oriented respectively to the programming languages of C#, JavaScript and Python have been named PDP-Aoraki, PDP-Meru, and PDP-Zunil.

For software agents and human users to manage metadata records, the Scribe registrars have been built which enable both agents and users to register, document and organize metadata records for entities in the PORTAL registries, DOORS directories, and Nexus diristries [13]. We have created implementations of these Scribe registrars as both web services that expose a RESTful read-write API intended for software agents, and web applications that provide a browsable user interface intended for human users (see Figure 1).

II. DESIGN OF NPDS

The NPDS schema for PDP has been designed from inception to be flexible, extensible, and customizable [2] with required, permitted, and optional fields supported in the metadata records. There have been some name changes for the fields between the original specification [2] and the subsequent specification [3], in particular, the change to clarify use of the words *resource* and *entity* within the framework of metadata levels and the concept of *metadata about metadata*. Some of the required and permitted fields include the following: entity label, entity tags, entity locations, entity description, record distribution, record provenance, record signature (see Figure 2). Whereas required and permitted fields are those named and defined by PDP for NPDS, optional fields are those named and defined by customized extensions for registries, directories, and diristries developed and maintained by independent scientific communities for their particular problem-oriented domains. When combined together as a valid infoset, all required, permitted, and optional fields in an NPDS metadata record provide essential information about the identification, ownership, location and description of an entity and the provenance, location, distribution and re-distribution of the entity's metadata and/or data both on the original authoritative primary server and on any non-authoritative secondary servers including caching servers.

A foundational tenet for PDP and NPDS has been independence from specific technology platforms for both the design principles and the software implementations. By building the NPDS cyberinfrastructure in different software stacks with a variety of programming languages and with both SQL and NoSQL database management systems, PDP will make the NPDS cyberinfrastructure available to a diversity of users, communities and organizations. Moreover, this variety of software stacks through which NPDS has been and will be implemented should demonstrate the important foundational tenet


View Beacon Resource Metadata Records on Nexus Server

Handle	PORTAL	DOORS	Type	Tag	Name	Nature
Y458C6E69	Valid	Valid	Organization	AMIA	American Medical Informatics Association	
GB13D0496	Valid	Valid	Organization	ATA	American Telemedicine Association	The American Telemedicine Association (ATA) is a non-profit association promot
Q6677A8A1	Valid	Valid	Organization	ATSP	Association of Telemedicine Service Providers	An industry service provider organization for institutions and corporations.
TA2D28FC1	Valid	Valid	Organization	BNIH	Bioinformatics at the NIH	
NAC8ACBAA	Valid	Valid	Organization	CTTC	California Telehealth and Telemedicine Center	Associated with the California Healthcare Association's Rural Healthcare Center and to improve access to health care services.
EA95022CD	Valid	Valid	ComputingTool	ELIDA	ELIDA Software	statistical data analysis for limiting dilution assays
M0131A77C	Valid	Valid	Organization	Ellitron	Ellitron Computing	biostatistical methods and software for bioassays

EntityLabels SupportingTags SupportingLabels CrossReferences OtherTexts Locations Descriptions Provenances Distributions ArchivedRecords

Priority	Marked	Restricted	Principal	SupportingLabel
0	false	false	false	http://npds.portaldoors.net/nexus/nlmmesh/biological_assay
1	false	false	false	http://npds.portaldoors.net/nexus/nlmmesh/microbiological_techniques
2	false	false	false	http://npds.portaldoors.net/nexus/nlmmesh/biostatistics
3	false	false	false	http://npds.portaldoors.net/nexus/nlmmesh/software

Figure 1. Screenshot view of metadata records from Beacon Nexus diristry; metadata for each record includes EntityLabels, SupportingTags, SupportingLabels, CrossReferences, OtherTexts, Locations, Descriptions, Provenances and Distributions.

that the NPDS cyberinfrastructure will remain a platform-independent system. Thus, the design principles for PDP will continue to be articulated in UML models, XML schemas and OWL ontologies for the DREAM principles, FAIR metrics, and NPDS cyberinfrastructure.

Another foundational tenet of PDP and NPDS has been the principles of flexibility, extensibility and customizability [2] with support for the use of other vocabularies, ontologies, formats and/or standards such as PROV-O [14] or DCAT [15] that may be incorporated and used in the relevant fields of the NPDS metadata records exploiting the principle of metadata about metadata [3]. This flexibility and extensibility of NPDS has enabled it to remain both adaptable to evolving standards and customizable by different scientific communities [2].

Different registries, directories, and/or diristries for each scientific community's problem-oriented domain of specific focus area with ontologies developed for the defining concepts and relationships relevant to the domain constitute an essential and important aspect of the NPDS cyberinfrastructure. These ontologies are used to determine the relevance and appropriateness of inclusion of metadata records for resources within diristries, and to determine the scope of other types of semantic searches, such as automated searches for online resources and meta-analyses of published literature [16], [17]. This approach enables problem-oriented domains with repositories such as ManRay, BrainWatch or SOLOMON to be established in which independent and localized governance of content can be maintained. However, these independent localized problem-organized domain repositories can still access the NPDS distributed network of interoperable repositories in other problem-oriented domains.

In addition to the DREAM principles which guide record structure and server function in the NPDS cyberinfrastructure, the network of interoperable servers has been designed in accordance with the Hierarchically Distributed Mobile Metadata (HDMM) style of architecture for pervasive metadata networks in a manner analogous to IRIS-DNS [8]. NPDS features a distributed metadata management system operated via a hybrid of hierarchical and peer-to-peer servers to assure that *who what where* information can be transmitted efficiently. This architectural style was adopted due to concern that neither a strictly hierarchical network nor a strictly peer-to-peer network would be suitable for the requirements of NPDS in scientific use with continuously changing information, data and metadata repositories. Pure hierarchical networks suffer risks associated with single central points of control including monopolization and the absence of independent local autonomy. Pure peer-to-peer networks cannot efficiently search, find and retrieve data in a large and distributed network.

Thus, PDP and NPDS have adopted the HDMM architectural style principles as quoted from [3, pp. 163-164]:

"1. *Distributed infrastructure:* Pervasively distributed and shared infrastructure, content, and control of content including distributed and shared control over both the contribution and distribution of the content defined as the mobile metadata records.
2. *Hierarchical authorities:* A hierarchy of both authoritative and non-authoritative servers (root, primary, secondary, forwarding and caching) enabling global interoperable communication and exchange of the mobile meta-

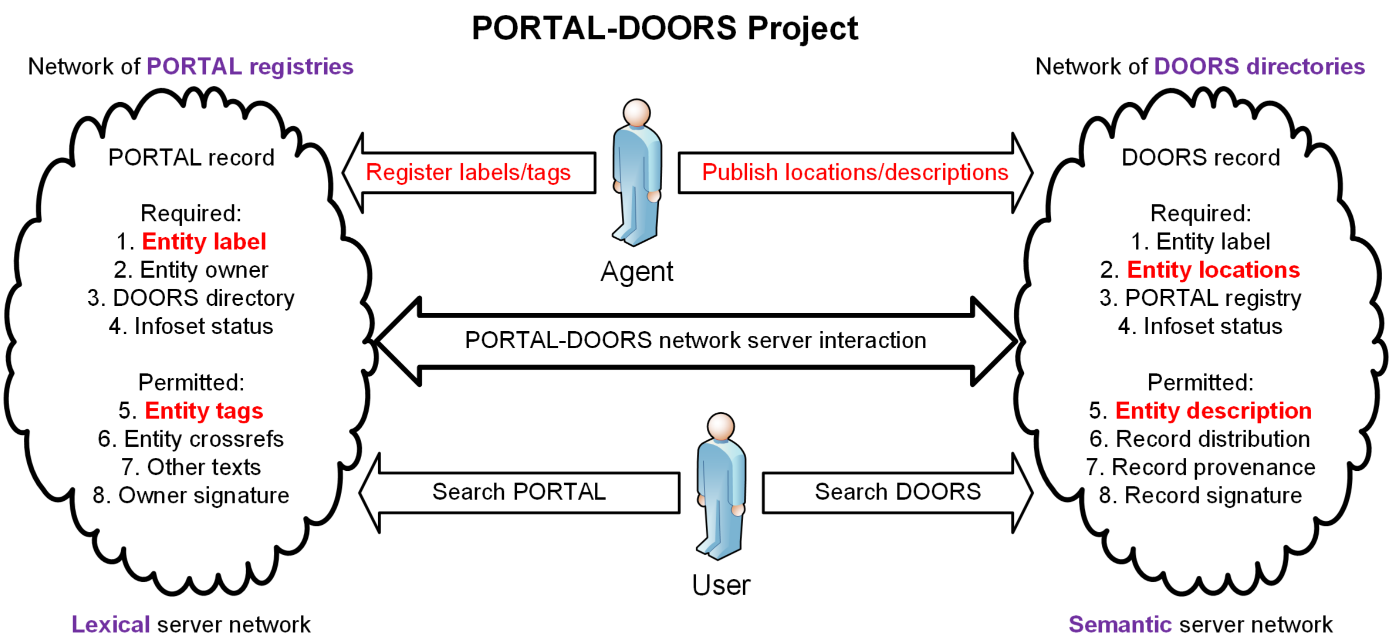


Figure 2. Diagram of relationships between lexical PORTAL registries, semantic DOORS directories, agents and users. A Nexus diristry operates as the integrated combination of a DOORS directory with a PORTAL registry.

data records while permitting independent administrative control of local policies governing the publication and distribution of the metadata records.

3. *Mobile metadata:* A focus on moving the mobile metadata for who what where as fast as possible with pervasive distribution and redistribution from servers in response to requests from clients that access non-authoritative local forwarding and caching servers updated regularly by the authoritative servers.
4. *Separated concerns:* A separation of concerns with registries for identifying resources and directories for locating resources that have been globally uniquely identified in the registries.
5. *Unrestricted identification:* A relative freedom of choice in the selection of identifiers with purposeful absence of any requirement to use the same root name or label for all identifiers, thus enabling essentially unrestricted choice of naming or labeling schemes for identification and thereby avoiding monopolistic control by any single organization."

III. Future of NPDS

NPDS has been designed and built as a hybrid, bridge and link between the lexical and semantic webs, thereby creating a cyberinfrastructure platform for a diverse variety of different computerized approaches and applications with artificial intelligence on the semantic web. These range from organized, accessible databases of complex, rapidly changing fields of biomedical sciences such as research on neurodegenerative diseases to an automated search engine query of other kinds of literature. Currently, the PDP development team continues to build the following new components 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 customized 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 as widely used in current library and online repositories so that citations of references can be easily imported to and exported from NPDS servers.
- *Concept-Validating Search Engine Agent (CoVaSEA):* An automated query engine and search agent that interoperates with NPDS and populates NPDS diristries with metadata records about resources.

Building on the current web applications for NPDS servers, we plan to implement a general purpose CMS that includes a blog, forum, and gallery and make it available as an open source application that interfaces with the NPDS cyberinfrastructure. Beyond development of a generic CMS, we also plan to build customized versions of our generic CMS tailored to tracking and managing the data associated with specialized focus areas such as clinical telegaming and clinical trials. We have also begun work on a customized CMS with the application and tools necessary to support our new journal called the Brainiacs Journal of Brain Imaging and Computing Sciences (www.BrainiacsJournal.org). With the future launch of our Brainiacs Journal, we plan to integrate our FAIR metrics into the manuscript submission, peer and editorial review process. For this new journal, we will impose the use of the FAIR metrics [12] as a requirement in the peer review process to detect plagiarism prior to approval for publication of any submitted manuscripts. With FAIR as an acronym for the phrases *Fair Acknowledgment of Information Records* and *Fair Attribution to Indexed Reports*, social engineering with the FAIR metrics will be used to incentivize fair citation of the published literature and to prevent plagiarism.

The FAIR metrics have been introduced in response to the current deficiencies of evaluating scientific merit of a particular publication based mostly on the number of citations it receives [12], [18], [19]. This approach to tallying counts may create a perverse motivation for some authors not to cite work of potential competing rivals in a given research field. Thus, the FAIR metrics have been created to encourage those authors who properly cite previously published work while signaling alerts to peer reviewers for those authors who fail to do so. In pursuit of this goal, the FAIR metrics must be able to distinguish intentional plagiarism of content and ideas from unintentional omission of citation.

Various forms of intentional plagiarism (ie, with purposeful non-citation of the original source) persist in the scientific literature [20]–[22]. This intentional plagiarism should be prevented and stopped. Plagiarism includes reproduction of figures and diagrams and copying of text passages word for word from source material without citation. Most current plagiarism detection systems recognize only verbatim copying with this lexical plagiarism and not semantic plagiarism. The latter, commonly known as the *plagiarism of ideas*, can be analyzed with improving but not yet adequate methods for automated detection algorithms [23]–[26].

The FAIR metrics of Craig *et al.* [12] have been designed to solve this problem of semantic plagiarism detection with the use of semantic RDF triple comparison and concept similarity detection while simultaneously also positively incentivizing good citation practices. These analyses will be summarized and reported with the family of FAIR metrics: F_1, F_2, F_3, and F_4. Each of the metrics in the FAIR family measures a different aspect of the authors' and article's adherence to proper citation practices. With social engineering and the use of incentivization with the FAIR metrics, we hope that NPDS will cultivate a more fair, honorable and truthful environment for publishing and citing scientific literature.

A challenge associated with populating content in NDPS servers remains the labor intensive task for expert curators to review semantic descriptions in metadata records for published articles from the scientific literature. However, the use of CoVaSEA, a query engine and search agent that finds and converts references to NPDS metadata records, provides an approach to facilitating better efficiency for this task [27]. Currently, we are developing two versions of CoVaSEA. The first implementation takes a lexically focused approach intended to retrieve and register resources into their relevant Nexus diristries via the use of lexical keywords [28]. The second implementation takes a semantically focused approach with SPARQL queries designed to retrieve and register resources into their relevant Nexus diristries via the use of semantic descriptions with RDF triples extracted from the free-form text [27]. Agents and users can then search not only externally the literature found outside in non-NPDS data stores, but also internally the literature found inside in the NPDS lexical and semantic data stores.

IV. PDP Software for NPDS

Our presentation of the DREAM principles, FAIR metrics, NPDS cyberinfrastructure, and the PORTAL-DOORS project will provide a comprehensive demonstration of PDP software especially the C# centric implementation with PDP-Aoraki. We will also provide a brief update on the JavaScript centric implementation PDP-Meru and Python centric implementation PDP-Zunil. Then we will present a detailed demonstration of the FAIR metrics and their use for plagiarism detection with examples from the literature of published papers that are known to have been plagiarized and retracted. Traditional lexical comparison tools are effective in detecting the plagiarism that results from copy-and-paste word copying by the plagiarizer with character string matching by the detection tool. However, they fail to recognize and detect the plagiarism of ideas and concepts. Semantic comparison tools offer an alternative approach that can detect plagiarism based on similarity and matching of concepts even if the ideas plagiarized are not copied word for word. The approach with use of the FAIR metrics, which enable a semantic comparison tool, has been implemented with a family of metrics based on the possible relationships between the concepts [12].

Another focus of our presentation will be the planned use of the NPDS cyberinfrastructure and FAIR metrics for the Brainiacs Journal. All references cited by submitted papers will be cataloged in Nexus diristries corresponding to the relevant subject area for use with CoVaSEA and FAIR metrics in the peer review process for the Brainiacs Journal. Our demonstration will feature an explanation of the work flow process for submitting and publishing a scholarly research paper in the Brainiacs Journal to highlight the collaborative nature of the NPDS system. Additionally, our presentation will also include an explanation of the scientific literature accessible from the network of metadata repositories available at www.portaldoors.org. We will facilitate an interactive discussion with the audience regarding problematic issues and potential solutions for the peer review process. We believe this discussion will benefit the escience and information science community concerned about improving quality in the peer and editorial review process for scholarly research publishing.

References

[1] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American*, vol. 284, no. 5, pp. 34–43, May 2001. DOI: 10.1038/scientificamerican0501-34.

[2] C. Taswell, "DOORS to the semantic web and grid with a PORTAL for biomedical computing," *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.

[3] ——, “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. [Online]. Available: www.mdpi.com/1999-5903/2/2/156/.

[4] A. G. Craig, S.-H. Bae, and C. Taswell, “Bridging the semantic and lexical webs: Concept-validating and hypothesis-exploring ontologies for the Nexus-PORTAL-DOORS System,” *Journal of Systemics, Cybernetics and Informatics*, vol. 15, no. 5, pp. 8–13, Jul. 11, 2017. [Online]. Available: www.iiisci.org/journal/sci/FullText.asp?id=BA947YN17.

[5] M. Skarzynski, A. Craig, and C. Taswell, “SOLOMON: An ontology for Sensory-Onset, Language-Onset and Motor-Onset dementias,” in *Proceedings International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, Institute of Electrical and Electronics Engineers (IEEE), Nov. 2015, pp. 969–972. DOI: 10.1109/bibm.2015.7359814.

[6] A. Craig, A. Ambati, S. Dutta, *et al.*, “DREAM principles and FAIR metrics from the PORTAL-DOORS Project for the semantic web,” in *Proceedings 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, (Jun. 28, 2019), in press, Pitesti, Romania: IEEE, 2019. [Online]. Available: www.portaldoors.org/pub/docs/ECAI2019DREAMFAIR0612.pdf.

[7] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, *et al.*, “The FAIR guiding principles for scientific data management and stewardship,” *Scientific Data*, vol. 3, p. 160 018, Mar. 2016. DOI: 10.1038/sdata.2016.18.

[8] 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.

[9] M. Boeckhout, G. A. Zielhuis, and A. L. Bredenoord, “The FAIR guiding principles for data stewardship: Fair enough?” *European Journal of Human Genetics*, vol. 26, no. 7, pp. 931–936, May 2018. DOI: 10.1038/s41431-018-0160-0.

[10] W3C, “Semantic Web HCLSIG F2F meeting program and attendees,” (Apr. 30, 2009), World Wide Web Consortium, Cambridge, Massachusetts, 2009. [Online]. Available: https://www.w3.org/wiki/HCLSIG/Meetings/2009-04-30_F2F.

[11] C. Taswell, “Biomedical informatics for brain imaging and gene-brain-behavior relationships,” in *W3C Semantic Web HCLSIG F2F Meeting*, (Apr. 30, 2009), World Wide Web Consortium, Cambridge, Massachusetts, 2009. [Online]. Available: https://www.w3.org/wiki/images/9/94/HCLSIG$$Meetings$$2009-04-30_F2F$W3CHCLS20090430CTaswell.pdf.

[12] A. Craig, A. Ambati, S. Dutta, *et al.*, “Definitions, formulas, and simulated examples for plagiarism detection with FAIR metrics,” in *Proceedings ASIS&T 82nd Annual Meeting*, (Oct. 19, 2019), in press, Melbourne, Australia: ASIS&T, 2019. [Online]. Available: www.portaldoors.org/pub/docs/ASIST2019FairMetrics0611.pdf.

[13] A. Craig, S. H. Bae, T. Veeramacheneni, *et al.*, “Web service APIs for Scribe registrars, Nexus diristries, 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.

[14] T. Lebo, S. Sahoo, and D. McGuinness, Eds. (Apr. 30, 2013). PROV-O: the PROV ontology, W3C Recommendation, The World Wide Web Consortium, [Online]. Available: www.w3.org/TR/prov-o/.

[15] F. Maali and J. Erickson, Eds. (Jan. 16, 2014). Data catalog vocabulary (DCAT), W3C Recommendation, The World Wide Web Consortium, [Online]. Available: www.w3.org/TR/vocab-dcat/.

[16] C. Taswell, “Concept validating methods for maintaining the integrity of problem oriented domains in the PORTAL-DOORS system,” in *IDAMAP 2010: Intelligent Data Analysis in Biomedicine and Pharmacology*, S. Swift and K. Phillips, Eds., 2010. [Online]. Available: www.portaldoors.org/pub/docs/IDAMAP2010p78-79Taswell0830.pdf.

[17] S. K. Taswell, A. Craig, D. Leung, *et al.*, “Hypothesis-exploring methods for automated meta-analyses of brain imaging literature,” in *Proceedings Annual Meeting of the Western Region Society of Nuclear Medicine*, Monterey CA, 2015. [Online]. Available: www.portaldoors.org/pub/docs/WRSNM2015TnT1p1020.pdf.

[18] 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.

[19] ——, “Formulation of FAIR metrics for primary research articles,” in *Proceedings International Conference on Bioinformatics and Biomedicine (BIBM)*, Madrid, Spain: IEEE, Dec. 3, 2018. DOI: 10.1109/bibm.2018.8621399.

[20] H. A. Maurer, F. Kappe, and B. Zaka, “Plagiarism - a survey.,” *Journal of Universal Computer Science*, vol. 12, no. 8, pp. 1050–1084, 2006. DOI: 10.3217/jucs-012-08-1050.

[21] I. Masic, “Plagiarism in scientific publishing,” *Acta Informatica Medica*, vol. 20, no. 4, p. 208, 2012. DOI: 10.5455/aim.2012.20.208-213.

[22] M. G. Tolsgaard, R. Ellaway, N. Woods, *et al.*, “Salami-slicing and plagiarism: How should we respond?” *Advances in Health Sciences Education*, vol. 24, no. 1, pp. 3–14, Feb. 2019. DOI: 10.1007/s10459-019-09876-7.

[23] N. Meuschke and B. Gipp, “State-of-the-art in detecting academic plagiarism,” *International Journal for Educational Integrity*, vol. 9, no. 1, 2013. DOI: 10.21913/IJEI.v9i1.847.

[24] S. A. Hiremath and M. S. Otari, “Plagiarism detection - different methods and their analysis: Review,” *International Journal of Innovative Research in Advanced Engineering*, vol. 1, pp. 41–47, 7 Aug. 2014, ISSN: 2349-2163. [Online]. Available: www.ijirae.com/volumes/vol1/issue7/AUCS10085.06.pdf.

[25] N. Meuschke, N. Siebeck, M. Schubotz, *et al.*, “Analyzing semantic concept patterns to detect academic plagiarism,” in *Proceedings of the 6th International Workshop on Mining Scientific Publications - WOSP 2017*, ACM, ACM Press, 2017, pp. 46–53. DOI: 10.1145/3127526.3127535.

[26] S. Vehviläinen, E. Löfström, and A. Nevgi, “Dealing with plagiarism in the academic community: Emotional engagement and moral distress,” *Higher Education*, vol. 75, no. 1, pp. 1–18, Feb. 2017. DOI: 10.1007/s10734-017-0112-6.

[27] S. Dutta and C. Taswell, “SPARQL-based search engine and agent for finding brain literature and converting references to NPDS metadata records,” in *Proceedings 11th International Conference on Brain Informatics*, Austin, Texas, Dec. 9, 2018. [Online]. Available: www.portaldoors.org/pub/docs/ICBI2018B277SDCT1209.pdf.

[28] S. H. Bae, A. G. Craig, and C. Taswell, “Expanding Nexus diristries of dementia literature with the NPDS concept-validating search engine agent,” in *Proceedings 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Jeju Island, Korea, Jul. 14, 2017, pp. 3704–3707. DOI: 10.1109/EMBC.2017.8037662.