

# NPDSLINKS: The Nexus-PORTAL-DOORS-Scribe Learning Intelligence aNd Knowledge System

Shreya Choksi, Peter Hong, Sohyb Mashkooor, Carl Taswell

Brain Health Alliance Virtual Institute

IEEE TransAI Conference 23 September 2020  
ID: Paper IP2; Session 22: TransAI-8 @ 4:45 PM



# Outline

- Principle of GIGO
- NPDS Cyberinfrastructure with Data
- LINKS Applications with Algorithms
- NPDSLINKS & Anti-GIGO Quality Descriptors
- Live Demo of NPDSLINKS

# Introduction

- Charles Babbage once said “On two occasions I have been asked, ‘Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?’ ... I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.” (Babbage 1864)
- Army Specialist William Mellin described the inability of computers to think for themselves with the observation that “sloppily programmed” inputs inevitably lead to incorrect outputs (Mellin 1957)
- Garbage In, Garbage Out (GIGO) expresses the truism that absence of quality in input data yields absence of quality in output results despite presence of a perfect analytic algorithm
- For real world data sets in experimental science, validity and reliability of results depend on both quality and quantity of data

# Importance of GIGO in AIA

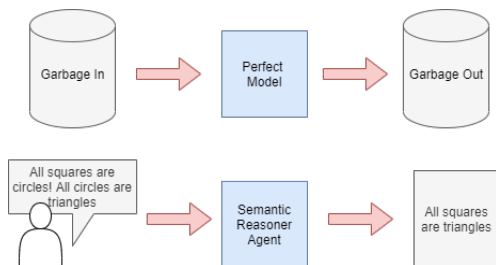


Figure: Garbage In, Garbage Out

- GIGO impacts all artificial intelligence applications (AIA)
- In big data, Gudivada *et al.* 2017 says "high-quality datasets are essential for developing machine learning models"

# Background of PDP

- AIA built with XML, RDF, OWL technology stack continue to confront challenges during ongoing transition from lexical to semantic web
- The PORTAL-DOORS project (PDP) specified a design for a distributed network of repositories for registering resource entities and publishing metadata about them in a manner analogous to IRIS-DNS for internet domains (Taswell 2007)
- PORTAL (Problem-Oriented Registry of Tags And Labels) registries identify resources with unique labels and lexical tags in a manner compatible with the lexical web
- DOORS (Domain-Ontology Oriented Resource System) directories specify locations and semantic descriptions for these identified resources in a manner compatible with the semantic web

# Original PORTAL-DOORS Design

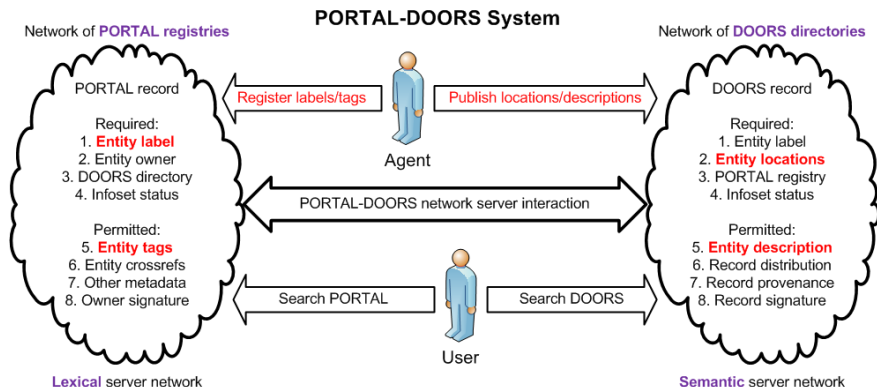


figure adapted from Taswell 2007 IEEE TITB

# Current Implementation of NPDS

- PDP developed the Nexus-PORTAL-DOORS-Scribe (NPDS) cyberinfrastructure as a "who what where" diristry-registry-directory system for identifying, describing, locating, and linking things on the internet, web, grid
- NPDS has been extended from the original PDP design with Nexus diristries (Taswell 2010) and Scribe registrars (Craig *et al.* 2016)
- NPDS provides a distributed and decentralized infrastructure where individuals and organizations can maintain independently managed problem domain repositories of semantic and lexical metadata and data for resource entities

# LINKS Applications with Algorithms

- Nexus-PORTAL-DOORS Scribe (NPDS) cyberinfrastructure of distributed network repositories for management of data referred to as “*NPDS cyberinfrastructure with data*”
- Learning Intelligence aNd Knowledge System (LINKS) applications and algorithms for analysis of the data referred to as “*LINKS applications with algorithms*”
- We use the acronym AIA for artificial intelligence applications and algorithms in general, and the acronym LINKS for the AIA developed by PDP and built on the foundation of the NPDS cyberinfrastructure
  - Automated search and meta-analysis of published scientific literature (Taswell *et al.* 2015)
  - Software agents with converters to generate metadata records (Taswell *et al* 2020)



# NPDSLINKS as a Knowledge Engineering System

- NPDS and LINKS designed to facilitate management with storage, retrieval, exchange and analysis of knowledge between and across different problem-oriented domains
- Different research communities can communicate and learn from each other, e.g., neuroscience and machine learning
- Via algorithms in LINKS applications and data in distributed NPDS repositories, with transdisciplinary bridges instead of silo walls, we envision building a knowledge engineering system capable of contrasting and comparing concepts across fields for better understanding of similarities and differences

- Website at [www.NPDSLINKS.net](http://www.NPDSLINKS.net) launched to serve as the root of the NPDS cyberinfrastructure with a Scribe registrar intended for NPDS components including Nexus registries, PORTAL directories, DOORS directories, Scribe registrars
- Website at [www.NPDSLINKS.org](http://www.NPDSLINKS.org) will serve as the home for our PDP work on LINKS applications with algorithms with a focus on anti-GIGO quality descriptors

[View Client's NPDS-Root Resource Metadata Records on Scribe Registrar Service](#)

	Handle	PORTAL	DOORS	Type	Tag	Name	Nature
▶	H13A4D256	Invalid	Invalid	NpdsDiristry	Avicenna	Avicenna Nexus Diristry	NPDS diristry for clinical trial data analysis (including information, management, methodology)
▶	IC1E5F9B9	Invalid	Invalid	NpdsDiristry	Beacon	Beacon Nexus Diristry	NPDS diristry for biomedical artificial intelligence
▶	K584523B5	Valid	Valid	NpdsDirectory	BHA-DOORS	BHA DOORS Directory	unrestricted NPDS directory for resources at BHA Registrar
▶	W59EBB3B5	Invalid	Invalid	NpdsDiristry	BHA-Nexus	BHA Nexus Diristry	unrestricted NPDS diristry for resources at BHA Registrar
▶	U6F42772D	Valid	Unknown	NpdsRegistry	BHA-PORTAL	BHA PORTAL Registry	unrestricted NPDS registry for resources at BHA Registrar
▶	XEE74E51F	Valid	Valid	NpdsRegistrar	BHA-Scribe	BHA Scribe Registrar	NPDS registrar for health care and life sciences managed by Brain Health Alliance
▶	XF374D4EE	Valid	Valid	NpdsDiristry	BioPORT	BioPORT Nexus Diristry	NPDS diristry for biomedical computing (including biomedical informatics mathematics statistics)
▶	E531ACD41	Valid	Valid	NpdsDiristry	BrainACS	BrainACS	Journal of Brain Imaging And Computing Sciences
▶	Y32E19FE9	Valid	Valid	NpdsDiristry	BrainWatch	BrainWatch Nexus Diristry	NPDS diristry for brain imaging informatics science medicine and health
▶	XD5192355	Valid	Valid	NpdsDiristry	CTGaming	CTGaming Nexus Diristry	NPDS diristry for diagnostic and therapeutic clinical telegaming
▶	G24472710	Valid	Unknown	NpdsDiristry	DaVinci	DaVinci Nexus Diristry	NPDS diristry for biomedical data integration, semantic web, knowledge engineering
▶	J5D0FCDf8	Valid	Valid	NpdsDiristry	Eywa	Eywa Nexus Diristry	NPDS diristry for biodiversity ecology and conservation
▶	L160322F6	Invalid	Invalid	NpdsDiristry	Fidentinus	Fidentinus Nexus Diristry	NPDS diristry for cases of plagiarism and intellectual property theft
▶	U8BC51886	Valid	Valid	NpdsDiristry	Gaia	Gaia Nexus Diristry	NPDS diristry for biosurveillance, toxicovigilance, environmental health and environmental protection
▶	Q2A61603F	Valid	Valid	NpdsDiristry	GeneScene	GeneScene Nexus Diristry	NPDS diristry for genetic medicine and science

# Descriptors and Measures of Data

- NPDS records should be evaluated with anti-GIGO quantitative measures and qualitative descriptors for individual fields and respective groups of fields (Nexus, PORTAL, DOORS, Scribe)
- Logical indicators are true or false with boolean values and reflect whether the concept is present/absent
- Quantitative measures are simple counts with integer values of field-defined items or float values for ratios like the FAIR metrics
- Qualitative descriptors are categorical variables with enum values that can check content for level of compliance with ranks or scores, or as *none*, *lax*, or *strict*

# Evaluations for Specific NPDS Fields

- DOORS Locations field
  - URL addresses can be resolved, pinged, assessed for response media type (application, image, text) and as JSON, XML, HTML, etc.
  - Postal service mail addresses and geophysical addresses can be validated by a geolocation service
- DOORS Descriptions field
  - Evaluations with reasoning agent or engine that tests for logical consistency of claims in content and inferences from claims
- DOORS Provenances field
  - Validation checks can verify sources and origins for resource entity
  - Example of all cited references for a published research paper

# Conclusion

- Variety of qualitative descriptors and quantitative measures can be used for evaluating the content in NPDS metadata records for both the quality and quantity of the data
- Full implementation of anti-GIGO approaches and methods for the data in the NPDS repositories will lead to greater confidence in results obtained from LINKS applications with algorithms for transdisciplinary artificial intelligence

# Contact 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, the NPDS cyberinfrastructure, and LINKS applications with algorithms for TransAI

# References



C. Babbage, *Passages from the Life of a Philosopher*. Good Press, 2019.



W. Mellin, "Work with new electronic 'brains' opens field for army math experts," *The Hammond Times*, vol. 10, p. 66, 1957.



V. Gudivada, A. Apon, and J. Ding, "Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations," *International Journal on Advances in Software*, vol. 10, no. 1, pp. 1–20, 2017.



C. Taswell, "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).



——, "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/).



A. 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](http://www.iiisci.org/journal/sci/FullText.asp?id=BA947YN17).



S. 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.