



Opening democratised portals and doors to the free flow of findable facts

- Communication barriers across different scientific communities have increased the need for interoperable information systems.
- Available tools and technologies for finding and accessing data across fields of study continue to face obstacles and have not yet risen to the challenge.
- Established in 2006, the PORTAL-DOORS Project (PDP) continues to work on developing and maintaining the Nexus-PORTAL-DOORS-Scribe (NPDS) cyberinfrastructure as an open information management system founded on the principle of supporting the free flow of findable facts for democracies around the world.

The year was 2006. The newest concept for the internet, termed 'Web 2.0' was about to be highlighted by *Time* magazine as a part of their person of the year, due to the popularity of the internet. Community and collaboration was the name of the game – at least on the surface – for the common user accessing Myspace or YouTube, both brand new and unexplored territory. Massive data stores of information crucial for advancing scientific research were locked up and hidden away in so-called 'cyber silos'. Cross-disciplinary research became challenging as outsiders might not be aware of the existence of such a repository, let alone have access to or know how to work with that data. To address this issue, the **PORTAL-DOORS Project** (PDP) was created by Dr Carl Taswell as an information management system to serve as a bridge between these many different fields.

Navigating the sea of information

Since then, the internet has gone from a bucket to a vast sea of information, making

it hard to find a specific resource on the web. A simple search using a common search engine can return hundreds of thousands of records of links, images, articles, videos, and more. In most cases, these search engines make use of both human-readable text and computer-readable supporting information. However, this only works efficiently when that supporting information is actually available.

Much of the web lacks robust and clear supporting information, such as author, creation/edit date, size, subtitle, and so on. Without this contextual information, data can become untrustworthy, but additionally, it makes it more difficult for search engines to efficiently find the most pertinent information for the user's search.

Moreover, specific to science and other technical fields, the reliance on similarity of text instead of similarity of underlying meaning can lead to cyber silos where, instead of data being hidden by security, information is hidden in plain sight. A user needs to know the jargon specific to the field they are interested in to find the correct data, and without that prior knowledge it can be difficult to find. As a result, many looking for information outside of their area of expertise rely on widely used monolithic search engines with context-free search algorithms that corporations guard as trade secrets.

Finding data with data

This supporting information is often called 'metadata' – referring to 'data about data'. Metadata characterises all attributes about a particular resource, but does not contain

Sharing data amongst interoperable systems ensures that it is not locked up in a specific cyber silo, enabling the free flow of information.

the stored data itself. For example, an image file on a computer would have metadata informing the user about the file format, size, user denoted tags and more, prior to showing the image itself. By having clean and neat metadata, information can become easier to find by enabling a search to look at metadata before delving deeper.

Improving data findability

The PORTAL-DOORS Project and NPDS Cyberinfrastructure utilise a Hierarchically Distributed Mobile Metadata (HDMM)

architectural design, the goal of which is to create a shared format for messages exchanged between data repositories supporting both the lexical web (based on strings of characters) and the semantic web (based on meanings of words). Individual computers can work in a distributed network of nodes throughout the internet and web, sharing metadata about resources with dynamically updated content, much like how libraries can obtain books that are not in their collections via interlibrary loans.

NPDS software can still search records for keywords, but the flexible, extensible, and open-source nature of the project means that anyone can create their own repository devoted to a given topic. For example, one repository could hold the recipes for cakes, descriptions of the ingredients, cooking implements, and the stores where you can find the ingredients and implements, all cross-referenced to each other.

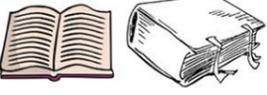
This approach allows anyone to publish their own online collections that can communicate with others, democratising the ability to publish facts, and make them findable. Sharing data amongst interoperable systems ensures that it is not locked up in a specific cyber silo, enabling the free flow of information.

Democratising access to findable facts with software from the PORTAL-DOORS Project

Several dozen repositories distributed across various websites with registrars maintained at portaldoors.net, brainhealthalliance.net, telegenetics.net, and other sites, contain a diverse set of records that demonstrate the diversity and extensibility of the system. The individual records are organised in topic-specific repositories, including key features such as physical and online location information, cross-references to records

using other record management systems, computer-readable semantic descriptions, and record provenance.

Among these, provenance bears special importance, as it represents the journey of the object, whether physical or digital, from one owner to the next and the details of its original development, distinguishing an original creation from a derived work. This multifaceted approach to metadata provides more ways to keep track of relationships among different resources and spot discrepancies in records.

	Semantically Same	Semantically Similar	Semantically Related	Semantically Different
Lexically Same	 <p>Phone Phone</p> <p>Equal and Equivalent</p>	 <p>Drink Drink</p> <p>Equal, not Equivalent</p>	 <p>Letter Letter</p> <p>Not Equal or Equivalent</p>	 <p>Compact Compact</p> <p>Not Equal or Equivalent</p>
Lexically Similar	 <p>Color Colour</p> <p>Equivalent, not Equal</p>	 <p>Emigrate Immigrate</p> <p>Equivalent, not Equal</p>	 <p>Fair Fare</p> <p>Not Equal or Equivalent</p>	 <p>Click Clique</p> <p>Not Equal or Equivalent</p>
Lexically Related	 <p>Oral Verbal</p> <p>Equivalent, not Equal</p>	 <p>Affect Effect</p> <p>Equivalent, not Equal</p>	 <p>Light Bright</p> <p>Not Equal or Equivalent</p>	 <p>Dessert Desert</p> <p>Not Equal or Equivalent</p>
Lexically Different	 <p>Twelve Dozen</p> <p>Equivalent, not Equal</p>	 <p>Book Tome</p> <p>Equivalent, not Equal</p>	 <p>Device Gadget</p> <p>Not Equal or Equivalent</p>	 <p>Internet Desktop</p> <p>Not Equal or Equivalent</p>

Modified with permission from Athreya, et al. (2020) Brainiacs Journal.

The image illustrates some of the notions about 'equal or equivalent entities' with regard to lexical versus semantic similarities and differences. It demonstrates some of the difficulties encountered when the same word can be used in different contexts with completely different meanings.

How can we distinguish fake from real?

Managing provenance metadata has gained awareness with establishing a legal basis for intellectual property with technology methods such as the use of generative adversarial networks (GANs). Programmes such as GPT-4 and DALL-E have emerged with autonomous capabilities, while remaining open for public use. While our technology has adapted, current legislation remains strapped with archaic intellectual property laws and limited regulation for generated work. Without an author defined for AI-generated work, it falls into the public domain.

In the art world, provenance decides the value of a piece. Lensa AI art software was exposed for its generated artwork containing signatures that may have come

from the input artwork. However, since the artist did not create the generated artwork, it alters the painting's provenance. Beyond art, deep fakes can generate images or videos of people, such as public figures, saying or doing whatever the deep fake creator desires, making it even harder for the public to distinguish real news from fake. If generative art and videos are placed in the public domain or mistakenly attributed, how can creators protect their intellectual property?

Future steps

In support of our goal to democratise open access to the free flow of information,

we began the annual series of Guardians Workshops to encourage discussion and collaboration regarding topics in support of truth, integrity, and democracy. Brain Health Alliance will continue to support the larger community of scholars, policymakers, entrepreneurs, students, and concerned citizens across the globe to help build a more democratic internet and web for a more fair and truthful world where everyone who has the facts can not only share them but make them findable to those who need them.

For more information about the Guardians 2023 Workshop including our current call for papers, please visit BHA.VI.us/Guardians.

The annual Guardians Workshop encourages discussion and collaboration regarding topics in support of truth, integrity, and democracy.

Personal response

Was there any particular subject or field of research that brought this need to your attention initially?

The original example that motivated the development of the PORTAL-DOORS project for interoperable data management came from brain imaging informatics as described in a 2009 paper entitled 'Knowledge engineering for pharmacogenomic molecular imaging of the brain' and several SNMMI meeting posters from 2006 and 2010 on the ManRay Project in nuclear medicine and molecular imaging. These are available at www.portaldoors.org/pub/docs/SNM2006Taswell0605P1431.pdf and www.portaldoors.org/pub/docs/WRSNM2010Taswell1019.pdf.

What would you do differently if you were starting a new project in this same space today?

If anything, the growing interest in decentralised finance and decentralised social media apps shows an even greater desire among ordinary people to democratise aspects of the web where a few platforms that big corporations control currently serve as choke points.

If we were starting over, we would want to move faster, put more emphasis on ease of use, and promote our software not only to the scientific community, but also to anyone with an interest in making it easier to find truthful and useful information online.

If an average person were to ask, 'How is the PORTAL-DOORS project relevant to my life? Why should I take an interest in it? How can I use it for my business or personal life?' How would you respond?

While installing PDP software for an NPDS repository currently

requires technical computing skills, we are working to make it easier for other organisations to install and maintain their own repositories of data records. The NPDS cyberinfrastructure has been designed to be sufficiently flexible and extensible for keeping track of information relevant to any general business affairs, recreation activities, or any other topic of interest. We anticipate development of simpler easy-to-use plug-and-play installers of our free open source software in the near future.

How do you see the Brain Health Alliance progress over the next decade or so?

We currently have several major initiatives related to use of PDP software and NPDS cyberinfrastructure:

- Managing data for a clinical trial of entire-body PET imaging for monitoring multiple sclerosis. We hope to see the results of this clinical trial published in both conventional scholarly articles and as NPDS records over the next 3-5 years, and we expect this rich dataset to still be yielding new and valuable insights more than 10 years into the future.
- Organising and hosting the Guardians of Truth and Integrity workshops in collaboration with IEEE Computer Society. In 10 years, we hope that this annual event will grow in size with increasing numbers of attendees and manuscript submissions.
- Continuing to develop and enhance our PDP software while improving ease of installation and ease of use. We are actively advocating for deployment of the software at multiple independent research institutions. In 10 years, we hope to have numerous sites maintaining their own repositories freely exchanging data and information.

Details



Anousha Athreya



Adam Craig



S Koby Taswell



Carl Taswell

e: ctaswell@BrainHealthAlliance.org
 t: +1 949 481 3121
 w: www.BrainHealthAlliance.org

Funding

Brain Health Alliance

Bio

Founded in 2007, Brain Health Alliance is a not-for-profit collaboration of

individuals and organisations with the collective aim of building bridges for better brain health. The Alliance's mission is to advance theranostics (therapeutics and diagnostics) with precision medicine and molecular imaging for disorders of the brain and nerves.

Further reading

- Athreya, A, et al, (2020) The essential enquiry 'Equal or Equivalent Entities?' about two things as same, similar, related, or different. *Brainiacs Journal*, 1.1, Edoc:PEDADC885
- Athreya, A, et al, (2021) Management software for monitoring related versions of cultural heritage artifacts for libraries and museums. *Proc ASIST*, 58.1, 676-678
- Craig, A, and Taswell, C, (2021) PDP-DREAM software for integrating multimedia data with interoperable repositories. *Brainiacs Journal*, 2.1, Edoc:HA46280EF
- Craig, A, et al, (2022) Motivating and maintaining ethics, equity, effectiveness, efficiency, and expertise in peer review. *Brainiacs Journal*, 3.1, Edoc:I5B147D9D

- Custers, B, and Fosch-Villaronga, E, (2022) Law and artificial intelligence: Regulating AI and applying AI in legal practice. *ITLS*, 35. Springer Nature
- Taswell, C, (2008) DOORS to the semantic web and grid with a PORTAL for biomedical computing. *IEEE Trans Inform Technol Biomed*, 12.2, 191-204
- Taswell, C, (2010) A distributed infrastructure for metadata about metadata: The HDMM architectural style and PORTAL-DOORS system. *Future Internet*, 2.2, 156-189
- Taswell, SK, et al, (2021) Truth in science. *Brainiacs Journal*, 2.1, Edoc:M85EC99EE

