

Provenance and the Question of “Equal or Equivalent Entities?”

Anousha Athreya, S. Koby Taswell, Adam Craig, Carl Taswell

Brain Health Alliance Virtual Institute, Ladera Ranch, CA, USA

2022 ASIS&T Workshop: October 9 @ 13:00 EDT
Who Are the Guardians of Truth and Integrity?



Comparisons and Contrasts

- Comparisons and Contrasts between two specific entities rely on the set of attributes of each entity
- If two entities with similar attributes are compared, then they have a degree or ranking of similarity
- Semantic and lexical attributes, as well as information affiliated with concepts (e.g. date, location, author) aid in verifying its origin
- Considering a claim or opinion within a manuscript, identifying its origin can help in distinguishing one has simply been voiced by a person, or one that has been validated by reproducible scientific research

Lexical vs. Semantic

- Lexical comparison processes two entities as character strings without embodying its semantic characteristics
- Semantic comparison processes two entities with reference to word meaning in the context of defined vocabularies, thesauri, or ontologies

Example 1: Fair versus Fare

- Lexical Comparison: comparing difference in characters
 - 'Fair' versus 'Fare'
- Semantic Comparison: sentence context and meaning
 - 'Fair' - noun, "an exhibition with accompanying entertainment and amusement"
 - 'Fair' - adjective, "in accordance with rules and standards, legitimate"
 - 'Fare' - noun, "a paying passenger on a public conveyance"
 - 'Fare' - verb, "to do, succeed"

Example 2: Dessert versus Desert

- Lexical Comparison: comparing difference in characters
 - 'Dessert' versus 'Desert'
- Semantic Comparison: sentence context and meaning
 - 'Dessert' - noun, "the sweet course eaten at the end of a meal"
 - 'Desert' - noun, "arid land with sparse vegetation"
 - 'Desert' - verb, "to withdraw from, usually without intent to return"

Equal or Equivalent?

| | 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> | 1 |
| 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> | 2 |
| 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> | 3 |
| 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> | 4 |
| | A | B | C | D | |

Tracking Versions Over Time

- Differentiating between two objects remains essential in many situations, whether the human instinct to contrast friend from foe, or the societal challenge of reproducibility in experimental science with the question ‘Same, similar, related, or different?’
- Related versions of an object must be compared in context of the domain-specific field [2]
- These related versions are key to identifying provenance, which can be used to track how a concept was changed, influenced, or adapted between its related versions

Information Archives for Cultural Heritage

- Large information archives are developed in museums and libraries of cultural objects for 'information literacy' [3]
- Artifacts found in certain regions can create a timeline for events that occurred to a specific population
- The field is heavily interrelated with historical events, using a collection of primary, secondary, and tertiary sources to establish an event has occurred [4]

Equivalent Entities in the DREAM Principles

- The PORTAL-DOORS Project (PDP) articulated a collection of design principles in 2006 for the continuing development of the Nexus-PORTAL-DOORS-Scribe (NPDS) cyberinfrastructure
- The PDP design principles were renamed in 2019 as the DREAM principles with the phrase *Discoverable Data with Reproducible Results for Equivalent Entities with Accessible Attributes and Manageable Metadata* [5]
- The Equivalent Entities principle remains at the core of the DREAM principles in support of scientific reproducibility and integrity in scholarly research [6]

Reproducibility in Scientific Research

Within the realm of experimental science, a foundational concept involves the reproducibility of scientific research:

“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.” [7]

NPDS Cyberinfrastructure

- The Nexus-PORTAL-DOORS-Scribe (NPDS) cyberinfrastructure provides a 'who what where' diristry-registry-directory system for identifying, describing, locating and linking things on the internet, web and grid [8]
- NPDS contains hybrid Nexus diristries which bridge lexical PORTAL registries and semantic DOORS directories
- In this hybrid system, we are able to store data as well as metadata about artifacts that relate to each domain-specific diristry
- Our main goal, when using provenance with NPDS, is to link two entities together by comparing their attributes and then tag them as same, similar, related, and different

Defining Provenance

According to Merriam Webster Dictionary, provenance is defined as "the history of ownership of a valued object or work of art or literature" or the "origin, source" [9]

These definitions of provenance often refer to a record of ownership of a work of art or a cultural artifact, however, we would like to define a broader scope for provenance

Within the Nexus-PORTAL-DOORS-Cyberinfrastructure (NPDS), provenance refers to the origin or source of an entity as well as the processes and methodology of its production and current state

This definition allows us to identify provenance in the context of art, history, and cultural artifacts, as well as data and metadata related to entities stored in NPDS, such as a brain scan, a piece of scholarly literature, or a physical object

Tools and Uses of NPDS Diristries

- The term "Garbage In, Garbage Out" (GIGO) was introduced by Charles Babbage in the early 19th century and later codified by William Mellin's remarks in the Hammond Times a century later
- GIGO refers to the concept of flawed data inputs producing flawed results in artificial intelligence applications (AIA)
- Brought to light the importance of developing and maintaining standards for reviewing the quality of curated data before applying the data within research [10]
- Within NPDS, there contains data repositories for a specific domain-specific field, in which transdisciplinary bridges can be built to compare the similarities and differences between and across fields

NPDS Software Release

- The [PDP-DREAM Software](#) for NPDS has been available as an open-source software since its release in 2021 (Taswell 2021)
- Three branches within the PDP-DREAM Software have been released, with dated releases in the future: Aoraki, Cervin, and Gangkhar that can be found in the PDP-DREAM GitHub repository
- The launch of open-source and open access software is in an effort to counter manipulation, censorship, and the falsification of data that limit the integrity of reproducible and validated scientific research

Ashurbanipal Diristry for Documenting Cultural Artifacts

- The Nexus-Portal-Doors-Scribe Cyberinfrastructure (NPDS) is an *archiving* system that, as part of the implementation measures to track related versions, can store metadata relating to the provenance of scholarly research and cultural artifacts
- The Ashurbanipal Diristry, as one of the hybrid Nexus diristries, contains the first use of tracking provenance
- The Ashurbanipal Diristry contains metadata records, documenting scholarly research about the preservation, restoration, and curation of cultural artifacts and cataloguing techniques

Ashurbanipal Diristry



www.BHAVI.us

Research for brain imaging and computing sciences

BHA ▾ STEMM ▾ BHAVI ▾ Files ▾ CsseViab ▾

NPDS ▾ Records ▾ aaathreya2 ▾ Agent ▾ Author ▾ Editor ▾

/NPDS/ScribeServer/AgentReseps from Ashurbanipal Diristry

« 1 2 3 ... 5 » Items per page 1 - 5 of 85 Items

+ Add new record

| Handle | Authorship | Type | Name | Nature | Private? | Shared? | Limited? | Released? | UpdatedOn |
|--------|------------|------|------|--------|----------|---------|----------|-----------|-----------|
| | | | | | false | false | false | false | |

EntityLabels SupportingTags SupportingLabels CrossReferences OtherTexts Locations Descriptions Provenances Distributions FairMetrics Services Snapshots Status

+ Add new record

| Index | Priority | Marked | Principal | EntityLabel | Resolvable | Private | TagToken | UpdatedOn | UpdatedBy |
|-------|----------|--------|-----------|-------------|------------|---------|----------|-----------|-----------|
| | | | | | | | | | |

« 0 » No items to display

| | | | | | | | | | |
|-------------|---------|----------------|-------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|-------|-------|-------|-------|------------------|
| ▶ ZBC83D613 | Release | PhysicalObject | Stradivarius Violin, The 'Kreutzer' | Antonio Stradivari was a craftsman of string instruments , including the Violin, The 'Kreutzer'. Violin is renowned for having a --> | false | false | false | false | 2021-12-22 02:37 |
| ▶ TB02CA7E | Release | PhysicalObject | Porcelain Bowl | | false | false | false | false | 2021-12-18 20:29 |

Figure: Current Status of Ashurbanipal Diristry

Managing Provenance in NPDS

- For implementation of the provenance format, we have demonstrated explicit conceptual and templated examples in the fields of art history, music, performing arts, historical events, and cultural heritage, and plan to curate others relating to provenance in architecture and computer science principles
- This aligns with our goal of furthering diristries aligning with domain-specific fields to aid the comparison of
- Importance for NPDS diristries to manage provenance that is interoperable with existing bibliographic multimedia
- Records within NPDS can be redundant and overlapping in various diristries and with the tag of 'same, similar, related, different' between entities with similar or same attributes

Concluding Notes

- Asking the essential enquiry “Equal or Equivalent Entities?” about two things enables their identification and characterization as 'Same, Similar, Related, or Different'
- Attention to the Equivalent Entities principle in scholarly research supports the reproducibility of experimental science and the pursuit of truth for the common good
- The goal of launching PDP-DREAM as an open-source and open access software is to counter the falsification of data that minimizes reproducible and validated research
- We must pay attention to the GIGO principle to ensure the quality of curated data for AIA and flag false equivalences

References

- [1] A. Athreya, S. K. Taswell, S. Mashkoor, *et al.*, "The essential enquiry 'equal or equivalent entities?' about two things as same, similar, related, or different," *Brainiacs Journal of Brain Imaging And Computing Sciences*, vol. 1, PEDADC885, pp. 1–7, 1 Dec. 30, 2020.
- [2] S. Gañçarski and G. Jomier, "Managing entity versions within their contexts: A formal approach," in *International Conference on Database and Expert Systems Applications*, Springer, 1994, pp. 400–409.
- [3] K. Baker, *Information literacy and cultural heritage: developing a model for lifelong learning*. Elsevier, 2013.
- [4] K. V. Korostelina, "Understanding values of cultural heritage within the framework of social identity conflicts," *Values in Heritage Management: Emerging Approaches and Research Directions*, p. 83, 2019.
- [5] A. Craig, A. Ambati, S. Dutta, *et al.*, "DREAM principles and FAIR metrics from the PORTAL-DOORS Project for the semantic web," in *2019 IEEE 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, (Jun. 28, 2019), Pitesti, Romania: IEEE, Jun. 2019. DOI: [10.1109/ECAI46879.2019.9042003](https://doi.org/10.1109/ECAI46879.2019.9042003). [Online]. Available: www.portaldoors.org/pub/docs/ECAI2019DREAMFAIRO612.pdf.
- [6] S. K. Taswell, C. Triggles, J. Vayo, *et al.*, "The hitchhiker's guide to scholarly research integrity," *Proceedings of the Association for Information Science and Technology*, 2020. [Online]. Available: www.portaldoors.org/pub/docs/ASIST2020HHGGuide0610.pdf.
- [7] S. Dutta, P. Kowshik, A. Ambati, *et al.*, "Managing scientific literature with software from the PORTAL-DOORS Project," in *2019 IEEE 15th International Conference on eScience (eScience)*, (Sep. 24, 2019), San Diego, California: IEEE, Sep. 2019. DOI: [10.1109/eScience.2019.00081](https://doi.org/10.1109/eScience.2019.00081). [Online]. Available: www.portaldoors.org/pub/docs/BCDC2019PdpDemo0806.pdf.
- [8] 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/.
- [9] M.-W. Dictionary, "Merriam-webster," *On-line at www.merriam-webster.com/dictionary/*, 2021.
- [10] S. Choksi and C. Taswell, "The nexus-portal-doors-scribe (npds) learning intelligence and knowledge system (links)," *Brainiacs Journal of Brain Imaging And Computing Sciences*, vol. 1, B61CA3D89, pp. 1–9, 1 Dec. 30, 2020.