Common Core Ontologies
Path to Standardization

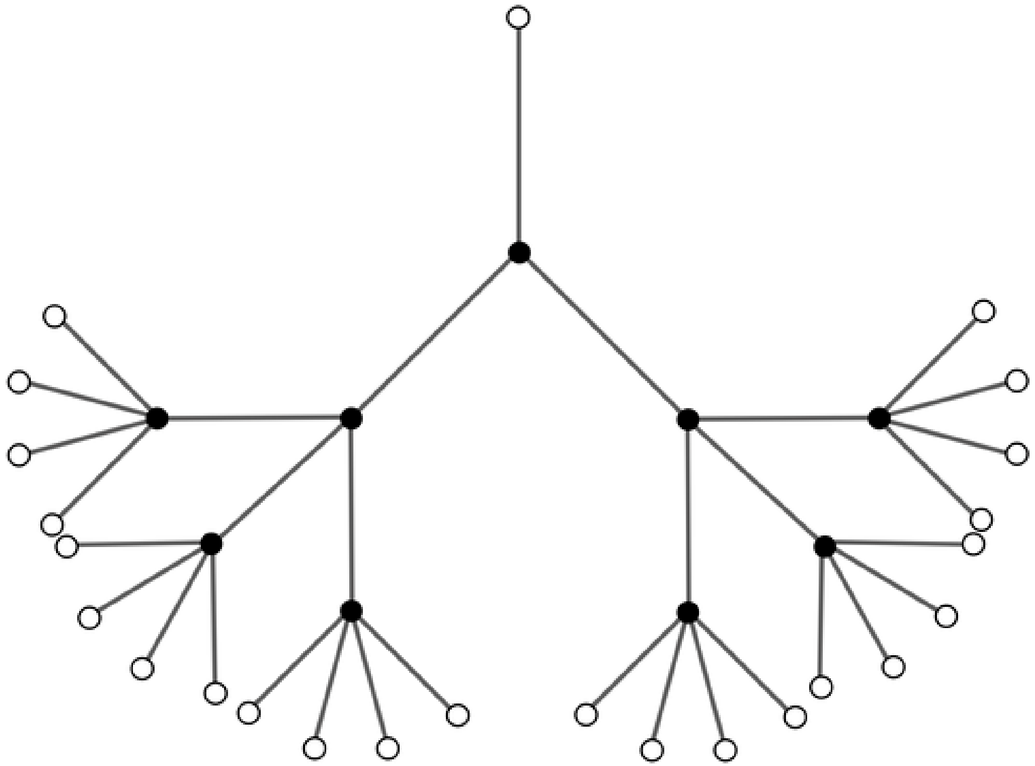
Mark Jensen
The DoD/IC Ontology Working Group
Ground Systems Architecture Workshop
Feb 28, 2024

Origin of the Common Core Ontologies (CCO)

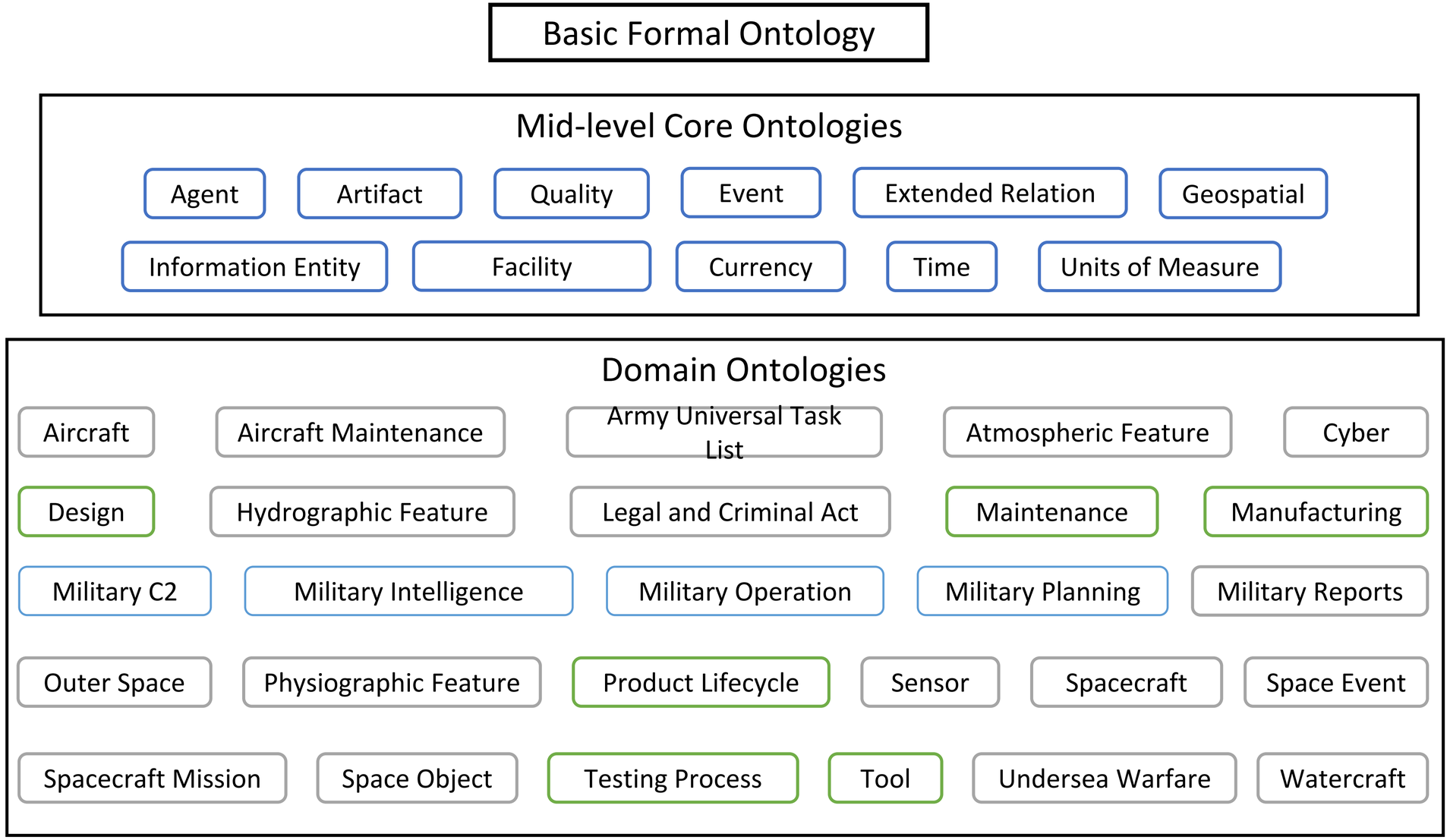
- Created as part of the IARPA Knowledge, Discovery and Dissemination Program (2010-2015).
- Requirement was to rapidly produce actionable intelligence from unfamiliar information sources having unanticipated domains.
- Created ontologies for describing types of things that are part of many, if not all, domains of interest.
  - Descriptions of new domains require only adding content specific to that domain.
- Ontologies are designed to function as a single point of integration, resulting in ***semantic interoperability*** across data sources.

Top-, Mid-, and Domain-Level Ontologies

- An ecosystem of highly interoperable ontologies
- Distinguished by layers of generality and have clearly defined scopes
- Top-levels provides a wire-frame that is repeatable and applicable to many use-cases
- Top- and mid-level are stable, highly reusable, carefully managed, do not overcommit
- Domain ontologies are require input from subject-matter experts, are perspectival, support specific use-cases, linked to databases.



The CCO Ecosystem



Expansion and Development

- CUBRC has supported the development, curation, and dissemination of CCO since its inception.
- Students and post-docs from University at Buffalo ontology programs have gained experience and training at CUBRC using and developing CCO.
- CCO was open-sourced in 2017.
- Refined and vetted via numerous domain and application ontologies across hundreds of projects in last 13 years.
- Recently endorsed by DoD-IC as a baseline ontology standard.
- Undergoing standardization as a IEEE mid-level ontology.
- Independent governance by leading experts in the field.

Common Core Governance Board

- Ensuring that CCO Mid is openly available, well-maintained, responsive to user needs, maintain pace with technological and theoretical developments, and remains independent of any undue influence imposed by a single project or organization.
- The board will have an established charter with transparent and documented procedures and will be composed of representatives from stakeholder organizations.
- Founding Members:
    - *CUBRC Inc*: Mark Jensen (Chair), Alexander Cox
    - *University at Buffalo*: John Beverley, Barry Smith
    - *JHU Applied Physics Laboratory*: J. Neil Otte
- A government liaison has been invited, pending approval.

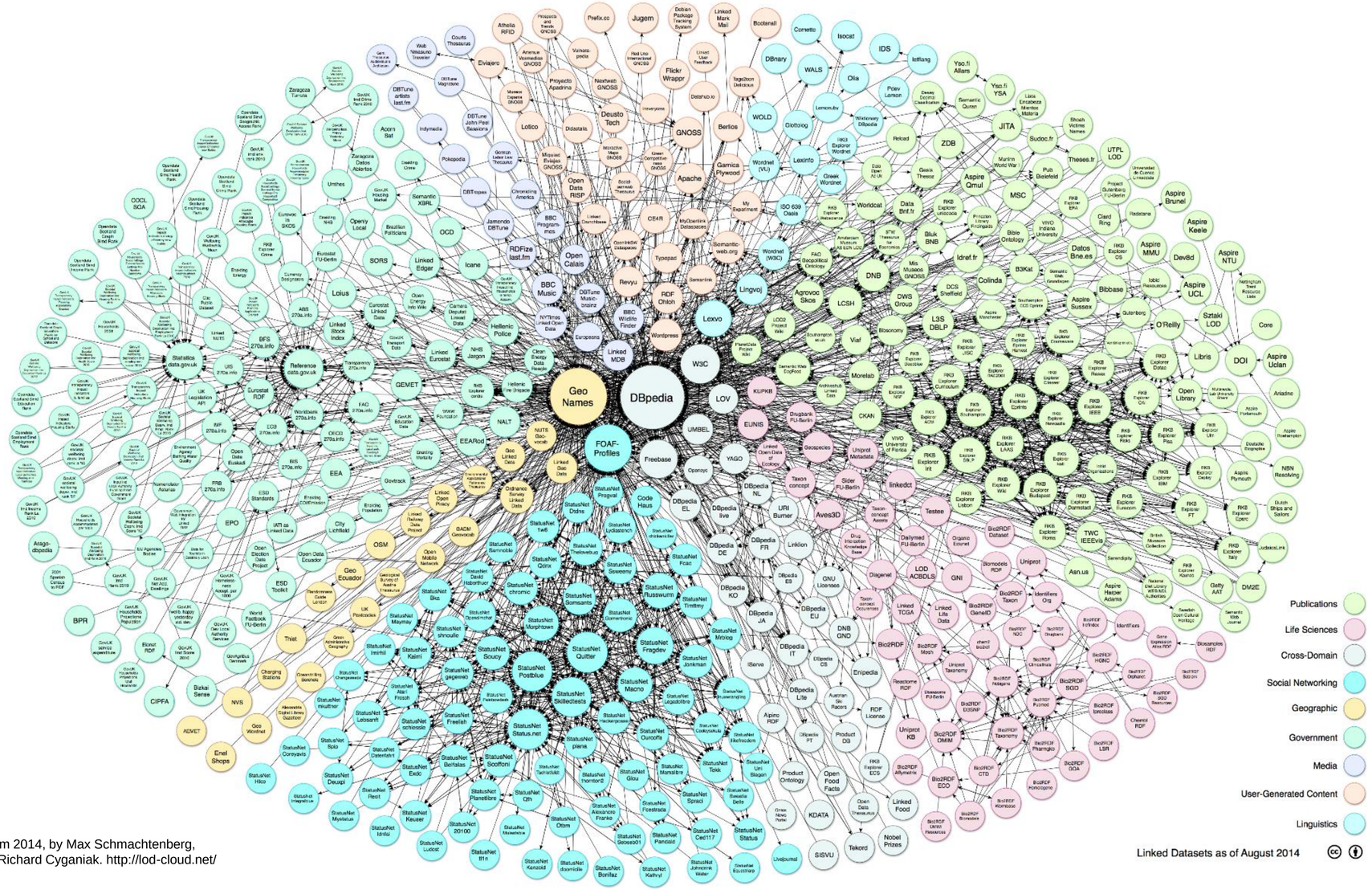
Some Users of CCO

- Aerospace Corp.
- AFRL/RIGB
- BAE
- Boeing
- Defense Intelligence Agency
- Dow Jones & Company
- I2WD
- Institute for Defense Analyses
- Integrated Solutions for Situation Awareness
- Johns Hopkins University APL
- Lincoln Labs
- Lockheed Martin
- MITRE
- MTConnect
- NGA
- OUSD Personnel & Readiness
- Parsons
- RTX
- SAIC
- Securboration
- SMC/SPGA
- Stevens Institute of Technology
- Texas State University
- United Technologies Corporation
- University at Buffalo
- University of Toulouse

BFO/CCO Methodology – Realism

- Ontologies shall represent our current best understanding of reality, grounded in evidence.
  - Mitigates conflicting perspectives.
  - Objective benchmark for settling modeling disputes.
  - Enables consistent semantics at varying levels of granularity.
  - An important step towards automated and reliable interoperability.

BFO/CCO Methodology – Avoid this



Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/

Motivating Standards for Semantic Artifacts

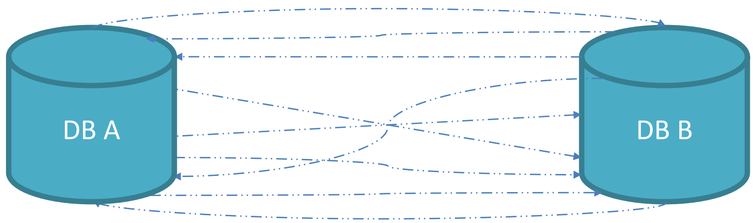
- Historically data has been modeled, structured and stored in a manner that best suits the information and performance needs of applications.
- Led to large amounts of siloed data.
- Achieving interoperability is a resource intensive task fraught with errors.
- Many standard data exchange formats exist now.
- Our big data problem is a horizontal one now.

From Syntactic to Semantic Standards

- Transform one way of referring to a thing so that it is understandable to an agent that speaks about the same thing differently.
- We do this by creating mappings that extract and manipulate source content into a form that's usable, ie., *understandable*, by a new user.
- Analogous to translating between spoken languages.
  - Not just vocabulary and grammar,
  - but context and the intended *meaning* of source data.
- Enterprise scale interoperability requires establishing meaningful links between data elements both within and between domains.
- A standardized vocabulary must break through the perspectives of data sources in order to link them semantically.

Map as You Go

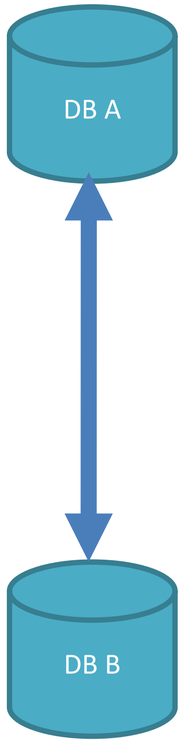
- Mappings are typically created manually when need arises, often application specific, incomplete, and labor intensive.



- Partial one-way mappings between databases allow one database to retrieve specific information from another, but the integration is fixed and rigid.

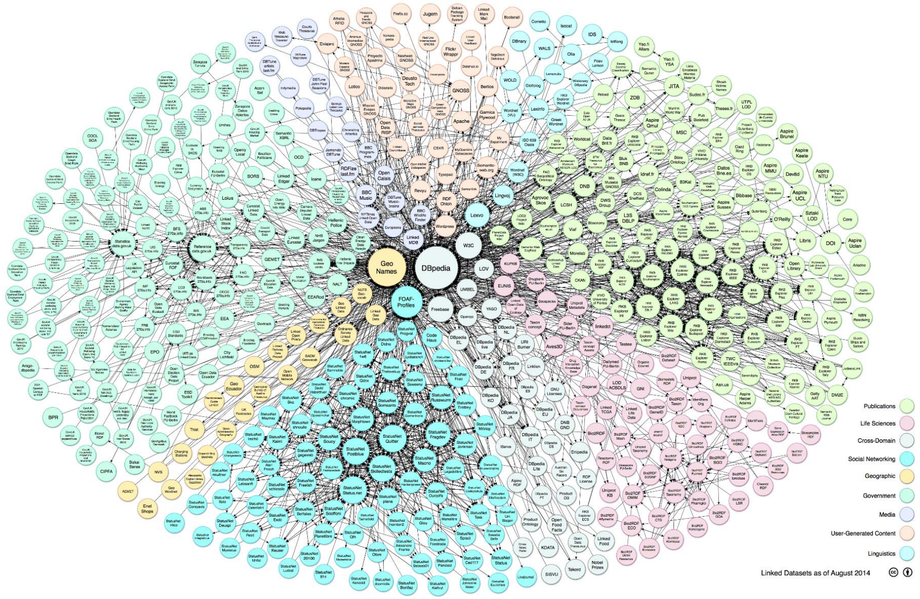
Transformational Mappings

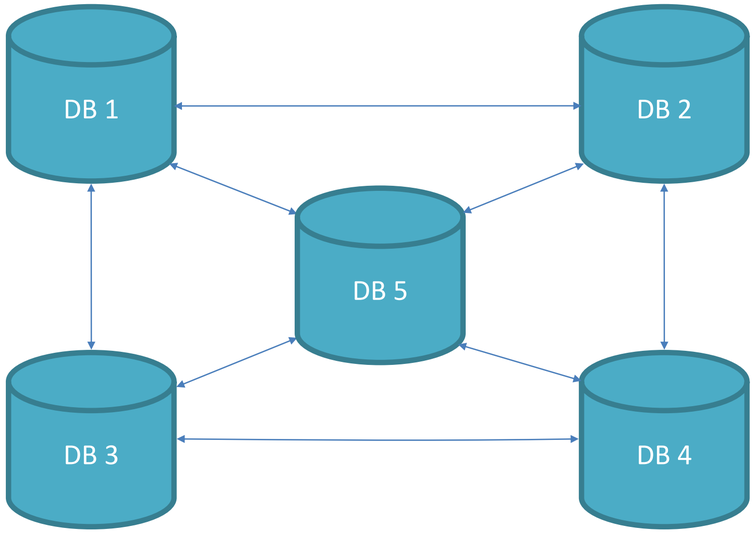
- A comprehensive bidirectional mapping between databases is what's needed to support interoperability across domains.
- Enables complex analytics, predicting courses of action, matching mission needs to asset capabilities, automated reasoning and learning, developing insight and understanding.
- Analogous to people who are completely fluent in each others' native language. Communication is seamless and intended meaning is preserved and carried through the transformation.



Mappings — Scalability Problems

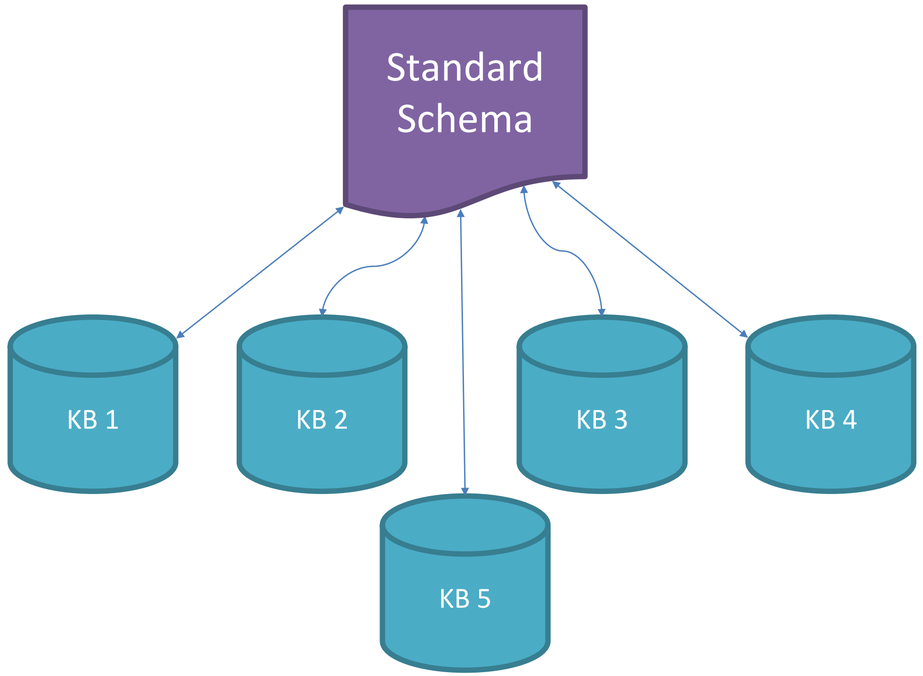
- Comprehensive bidirectional mappings are resource intensive
- A polynomial problem: $N^2 - N$
- 5 schemas = 20 complete and validated mappings
- 100 = 9900
- 10000 = 99,990,000
- …





Mapping to a Standard Reference Model

- Applications and databases communicate via use of a shared reusable set of ontologies whose main feature is to support interoperability via design principles that enforce modularity, extensibility, transparency, and automation.
- 2N vs N^2–N mappings*
- 10000 = 20000 (vs ~100 million)



* Some pairwise mappings required, but nonetheless have a head-start and open to automation

Next Step: A Standard for Mid-level Ontologies

- Demand signal for interoperable ontologies is increasing, scalabilty across the enterprise.
- Knowledge graphs are increasingly relavant as our adoption of LLMs is exploding, training, validation, guardrails, prompts, domain knowledge, ….
- Data centricity requires reliability, automation, validation, transparency, scalability, neutrality.
- Syntactic standards are giving way to ones for semantics.

IEEE Standards Association, PAR Approved

- A working group for developing ontology standards.
- https://standards.ieee.org/ieee/3195/11025/
- P3195 will prescribe a standard for what criteria any mid-level ontology must have
- P3195.1 will describe how the CCO satisfies P3195

Ubiquitous Screenshot

P3195

Standard for Requirements for a Mid-Level Ontology and Extensions

Active PAR

This standard specifies the requirements for a mid-level ontology and for the creation of conforming extensions and modules (i.e., subsets) therefrom. A mid-level ontology is a set of well-defined terms and relations used across multiple domains, which enables conforming extensions

CDAO and IC CDO Directive

MEMORANDUM FOR CHIEF DIGITAL AND ARTIFICIAL INTELLIGENCE OFFICER COUNCIL MEMBERS
INTELLIGENCE COMMUNITY CHIEF DATA OFFICER COUNCIL MEMBERS

SUBJECT: Baseline Standards for Formal Ontology within the Department of Defense and the Intelligence Community

In April 2023, the Chief Digital and Artificial Intelligence Officer Council and the Intelligence Community Chief Data Officer Council chartered the joint Department of Defense (DoD) and Intelligence Community (IC) Ontology Working Group (DIOWG). It was tasked with developing coordinated ontologies to set the agreed definitions and standard necessary to make data machine understandable. Based on the DIOWG's recommendations, both Councils direct the use of three baselines: Top-Level Ontology, Basic Formal Ontology, and Common Core Ontology. These will set the baseline standards for formal DoD and IC ontology.

By aligning the DoD and IC ontologies to a common set of top and mid-level standards, the combined enterprise will realize significant gains in data interoperability, federated search

CDAO and IC CDO Directive

By aligning the DoD and IC ontologies to a common set of top and mid-level standards, the combined enterprise will realize significant gains in data interoperability, federated search and discovery, decreased analytic timelines, and better cost efficiency. This common approach to data ontology is key to deriving value from shared data assets at speed and scale. The DIOWG has provided additional background information on these international ontological standards in Attachment A.

The nation's warfighters and intelligence professionals will need to have a decisional advantage in the immediate future and that can only be unlocked through the sharing of interoperable data. The next steps for the DIOWG are to codify recommended principles and governance processes to manage the DoD-IC Ontology Foundry. The DIOWG collaboration site can be accessed by visiting https://www.trmc.osd.mil/wiki/display/DIOWG/.

WADE LORI C VYTRP0

Digitally signed by WADE LORI C VYTRP0
Date: 2024.01.25 14:33:16 -05'00'

Lori Wade
Intelligence Community Chief Data Officer
Office of the Director of National Intelligence

MARTELL.CRAIG.HARRY.1269768998

Digitally signed by MARTELL.CRAIG.HARRY.1269768998
Date: 2024.01.04 15:11:45 -08'00'

Dr. Craig H. Martell
Chief Digital and Artificial Intelligence Officer
Department of Defense

Thanks!

CCO = mark.jensen@cubrc.org

BFO = johnbeve@buffalo.edu

IEEE OSWG = james.schoening@ieee.org