

2022-12-13 DMRWG Meeting Notes Template

Meeting Date

13 Dec 2022 The DMRWG meets bi-weekly on Tuesdays at 12:00-13:00 PT / 16:00-17:00 UTC. Check the [ToIP Calendar](#) for meeting dates.

Zoom Meeting Link / Recording

- [Recording of Meeting](#)

Attendees

- [Neil Thomson](#)
- [Carly Huitema](#)
- [Zaida Rivai](#)

Main Goal of this Meeting

TBD

Agenda Items and Notes (including all relevant links)

Time	Agenda Item	Lead	Notes
5 min	<ul style="list-style-type: none">• Start recording• Welcome & antitrust notice• Introduction of new members• Agenda review	Chair	<ul style="list-style-type: none">• Antitrust Policy Notice: <i>Attendees are reminded to adhere to the meeting agenda and not participate in activities prohibited under antitrust and competition laws. Only members of ToIP who have signed the necessary agreements are permitted to participate in this activity beyond an observer role.</i>• New Members:
5 mins	Announcements	TF Leads	<p>News or events of interest to WG members:</p> <ul style="list-style-type: none">• A presentation is planned for mid-Jan 2023 by James Schoening (IEEE member and chair of the Ontology WG) on the IEEE MyOntology project, which is working with the IAM project - in an integrated personal data system (iPDS) in a joint project.<ul style="list-style-type: none">◦ The MyOntology project is an open-source project open (free) to anyone who joins the IEEE and creates an account.◦ From the GitLab project description: "The My Ontology (MO) is a standards-based data format for the individual, which allows uniquely formed data from vendors to be transformed into this common format and better utilized by the individual."• We are looking for Data related topics and presentations for 2023.<ul style="list-style-type: none">◦ The target is to look for a series of projects to hold working meetings on delivering results and presentations of any kind on Data in an SSI world, particularly for data other than what is stored in Verifiable Credentials, particularly how to integrate SSI Trust, Data Exchange (joint work with the DIF Data Agreements group) and Authentic Data principles into existing data exchange, publication and pipeline models in current and future practice

20 mins	Possible general solutions to modeling data across multiple schemas (including multiple standards)	Al	<p>Nov 29, 2022, meeting featured a presentation by Burak Serdar included a medical data transformation system which is based on work by a number of US Medical Data groups and projects which had created a model (OMOP) which allows data extraction and transformation from data sets using multiple "concrete" schema standards (e.g., standard relational DB schemas) without having to resort to "YAS" yet another standard.</p> <p>As a base to understand today's discussion, it is suggested to review the recording and the materials listed (repeated here):</p> <ul style="list-style-type: none"> • Presentation • Related Paper • Meeting Recording (skip to 6:10 - 6 min, 10 seconds) • Also, see Data Transformation and Exchange Concepts. <p>The approach used was to build a common model which abstracts data across standard schemas into a few high-level "concepts", which allows the grouping of related data into a row-based vs. column-based approach. A column-based (typical relational DB) approach is that each type and variation of personal data, medical measurements, conditions, and observations is a separate column in one or more record types. A row-based approach has a simplified structure that allows differences in measurement types and categories to become a small number of "self-describing" metadata columns that provide type and classification information for the actual data value.</p> <p>Another approach to aid in different language locales was to take any of the data values and, where they were not a numerical value to assign them a discreet index number (integer) such that translation to other languages is a lookup of the index number for that particular language.</p> <p>This approach resulted in over 6,000,000 (yes, 6 million) separate discreet index numbers, categories, and data types.</p> <p>Its advantage is that over time, it has proven that supporting additional medical data standards only adds to the index number values, which doesn't change the abstract/meta-model, showing that in this medical data "ecosystem," the model is stable.</p> <p>This model is more complex to extract data for a particular study (or to transform to a different concrete relational or graph data schema, however, creating multiple pass queries is a well-known approach that can be optimized. The stability of the abstract/meta-model also brings stability to queries built for extraction and transformation.</p> <p>By a multi-pass query, querying an abstract/meta-model requires the first pass to identify the metadata and a second to extra the data specified by that metadata.</p> <p>The promise is to explore this approach as a way of simplifying data exchange in general - in a world dominated by multiple international and local "standards" - for next generation SSI based applications but also to potentially assist in resolving interoperability for Verifiable Credentials within a given Credential Ecosystem (e.g., academic, medical).</p> <p>Good Health Pass as a POC - As a project to explore this technique, it was suggested to revisit the Good Health Pass project, which was developed in 2020 as the COVID crisis was exploding and impeding travel across international borders. The project ultimately did not achieve its goal of producing a "health pass" related to COVID, designed to be universally acceptable by each country.</p> <p>This project did not succeed in being accepted, in part due to the inability of countries around the world to cooperate in defining a single acceptable set of claims and credentials for COVID-related travel and a set of common evaluation rules (and common rules evaluation engine) to determine if a traveler met a given national health standard for entry into that country. This was understandable as a travel solution was required almost immediately, so an attempt at an international agreement was quickly abandoned.</p> <p>The goal of the project (possibly as a Task Force) is</p> <ul style="list-style-type: none"> • To get a full understanding of the GHP, including the proposed solutions and understanding the technical and human reasons that it was not accepted. This includes understanding all the various individual national requirements for what health data was being asked for and evaluation rules of that data to allow a traveler to enter the country. • Understand the possible concrete schemas for each country/jurisdictions data • Attempt to discover a common abstract/metamodel that can support the requested health data across all nations, plus be able to support arbitrary rules (or at least samples of them across the different national requirements. • Discover/develop requirements for the tools required to implement a VC evaluation system for the Good Health Pass goals. • Understand the limits of technology to solve (e.g., common presentation format across different VC implementation technologies) vs. ones that humans need to solve • Define what tools and techniques can be used to assist in resolving the "human"(political) interoperability problems for health VCs
5 mins	Trust Registry issues in Academia	Al	<p>Carly provided some insight into the academic domain on potential issues for "authentic data", particularly for provenance to all the human individuals involved in collecting, processing, governing, and publishing the data, as well as academic papers based on the data.</p> <p>Data collected by (grad (Masters, PhD) students is assigned to a student by the academic institution identity and possibly their own personal identity (but possibly only indirectly). The same is also true of the range of positions within post-graduate studies, including various levels of researchers (from "teaching assistants" to full-tenured professors).</p> <p>Post-grad students (like employees in many organizations) are only there for a short time, yet they may be involved with data, where they are cryptographically signing the data in one of several roles. That data may be used for many years, where the student/employee has moved on from their role at the university/employer, yet a user of the data may be looking to establish the quality of the data based on the reputation of the people involved in the all the steps in processing the data.</p> <p>Over time the problem (in a distributed network) becomes where does that (now inactive) information on signatories to the authentic data change is stored and how can it be found - both of which are currently understood to be responsibilities of trust registries and their relationship with "authentic" objects (identities, VCs, and general data, including medical records), which have yet to be defined.</p>
5 mins	<ul style="list-style-type: none"> • Review decisions /action items • Planning for the next meeting 	Chairs	Line up topics and presentations for meetings in 2023

Screenshots/Diagrams

Cloud Privacy Labs

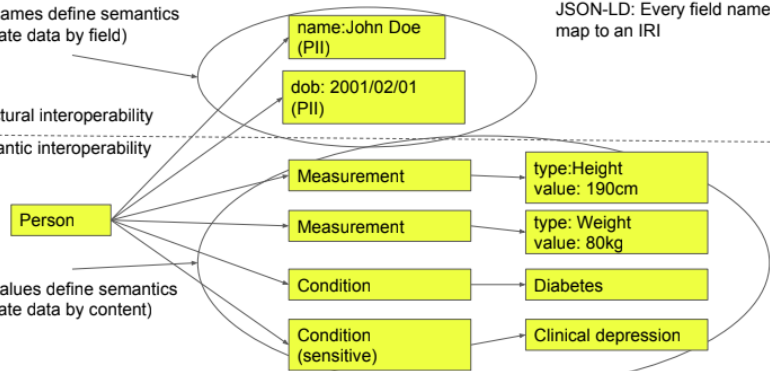
Semantics of Data?

Field names define semantics
(annotate data by field)

structural interoperability
semantic interoperability

Field values define semantics
(annotate data by content)

RDF: Each concept is an IRI
JSON-LD: Every field name
map to an IRI



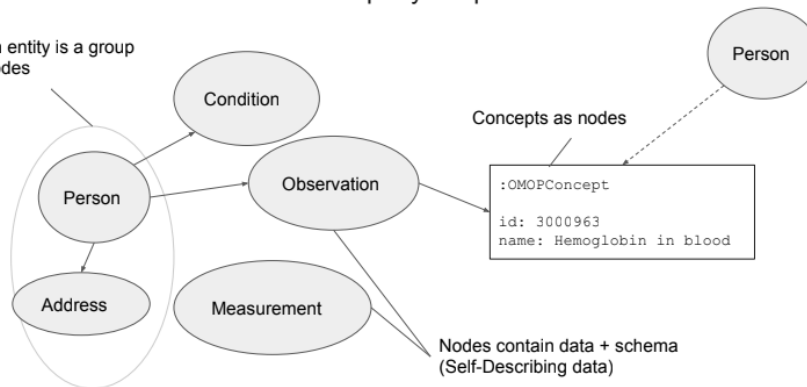
An illustration of taking concrete "column data" (e.g. each Measurement, Condition or Person data type is defined as a data column) which are categorized in a "row data" model where the category and type of data are "self-describing" metadata rows in columns in the meta/abstract tables using the model below.

Cloud Privacy Labs

Semantics of Data: LSA Approach

Labeled Property Graphs

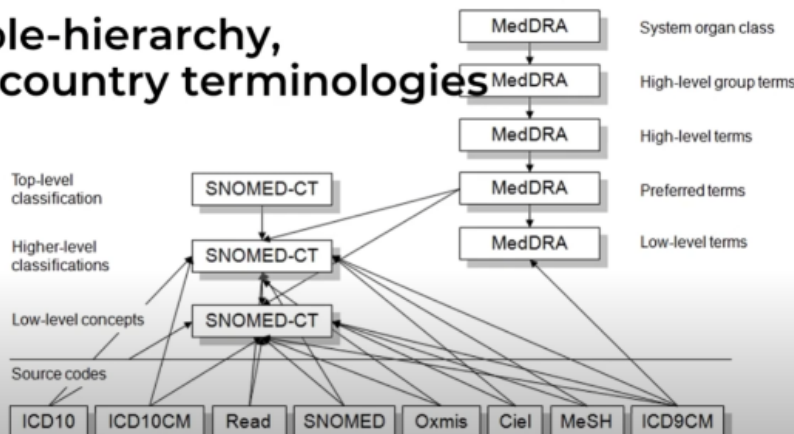
Each entity is a group
of nodes



The underlying "meta/abstract" model derived from multiple medical data standards for data exchange in the OMOP model



Multiple-hierarchy, multi-country terminologies



OMOP model - how multiple standards and terms (terminologies) can be combined in a single "meta/abstract" model that allows the extraction of data from multiple standards in a scalable way