Pacio SSIM Thoughts

16 November 2018

Contents

[1 SSIM Objectives 2](#_Toc530456108)

[2 SSIM Components 3](#_Toc530456109)

[2.1 Data Element 3](#_Toc530456110)

[2.2 SSIM Id or Sid Service 3](#_Toc530456111)

[2.3 Directories Service 3](#_Toc530456112)

[2.4 Digital Id Service 4](#_Toc530456113)

[2.5 Ontologies 4](#_Toc530456114)

[2.6 Interfaces 5](#_Toc530456115)

[3 Ontologies 6](#_Toc530456116)

[4 Data Storage 8](#_Toc530456117)

[5 Sid Service 9](#_Toc530456118)

# SSIM Objectives

Provide the method of storing and accessing Pacio data in a standardised, semantic way which is efficient for scaling to large scale use.

SSIM must be able to interact with current systems such as XBRL, ODI etc.

To Do points for this doc

Deprecated concepts

Deleted concepts

Abstract and concrete

Context

Formatting in and out

Sid wo ontology

Transactions!!!!

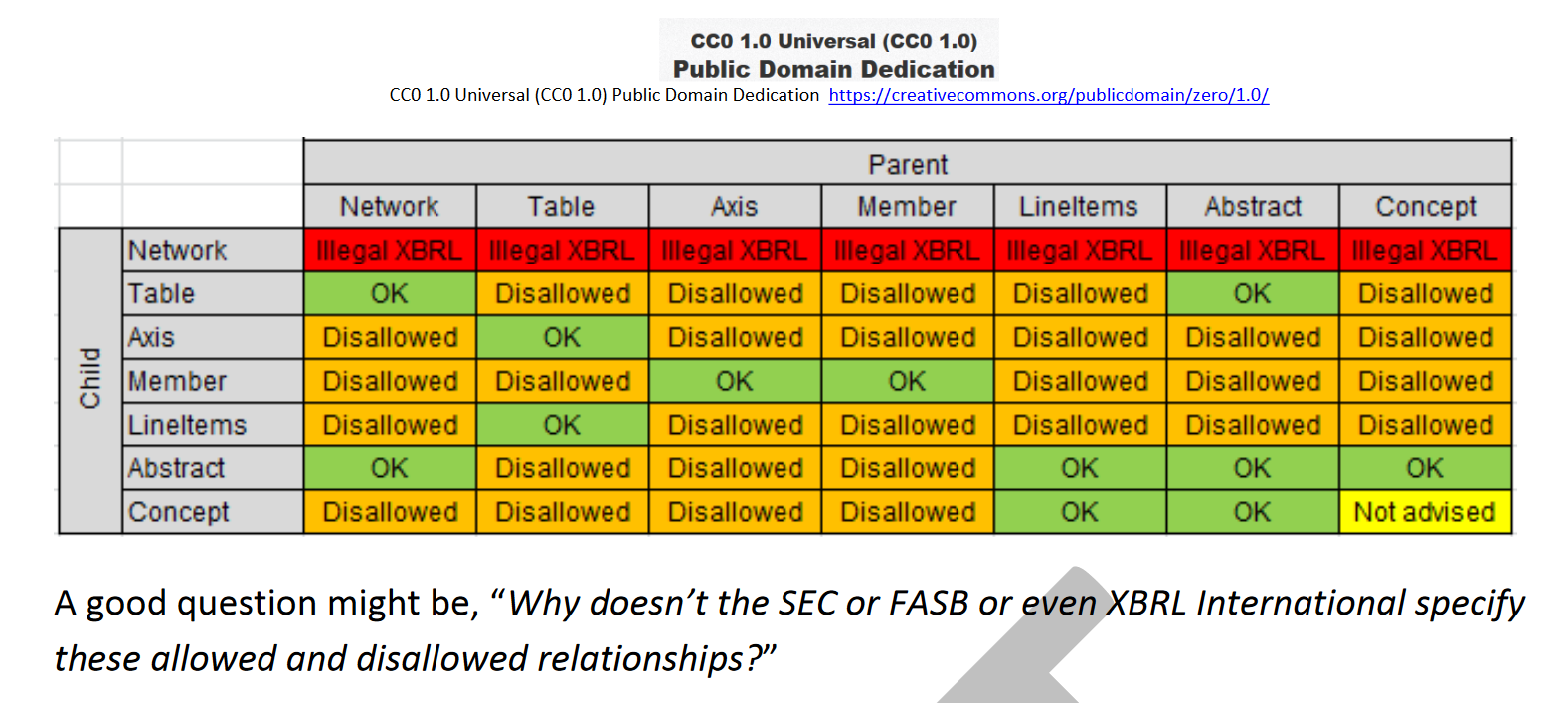
Periods

Dynamic – different tax treatments etc

The Data Commons Knowledge Graph (DCKG)  
<https://browser.datacommons.org/>

‘Has a’ etc as well as ‘is a’

Cover



and the previous “Pathological” XBRL in Chain of Capabilities Necessary to Automate Accounting Processes <http://xbrlsite.azurewebsites.net/2018/Library/ChainOfCapabilities.pdf>

# SSIM Components

SSIM (Standardised Semantic Information Model) will use the following components:

## Data Element

A single item or piece of data to be stored and referenced via SSIM is called an element. An element can be whatever the ontology (below) being used by an app supports, which can be any digital thing that a user or app wishes to store, retrieve, transfer, and be able to report on or use for analysis purposes.

## SSIM Id or Sid Service

An element is classified or using a single 128 bit (16 byte) number called an SSIM Id or Sid. Sids allow totally flexible data description in just 16 bytes, which will help make data storage and transaction transfers efficient. No long or variable length tags are involved.

A Sid is a key into a global database of sets of references or keys into the Directories, Digital Id Service, and Ontologies described in the next sections. Each set of references used by an app results in a single Sid. A Sid does not point to data directly. It is purely to provide reference or semantic content information.

A 128 bit sSid allows for 2128 – 1 or 3.4E+38 or 340,282,366,920,938,463,463,374,607,431,768,211,455 different sets of references which should cover the world’s needs for centuries. If ever that limit should be approached, then extending Sids to just 20 bytes (the length of Ethereum addresses) would add a few more centuries.

Sids once created will exist “forever”. They may become deprecated or no longer valid for new data, but will remain forever for historical analysis purposes.

A Sid by itself would not tell a human observer anything, but software will be easily able to show what its references, and to use it in searches or reports etc via the Pacio *Sid Service.*

Sids are not specific to Pacio i.e. they could also be used by non-Pacio systems.

## Directories or Facts Service

Pacio will develop and maintain directories of world facts or information types as below. These directories are for static information which does not change, or which changes infrequently. (An example of a ‘fact” which could change is a country grouping as for Brexit.) These directories will use existing e.g. ISO classification where possible, but extend them to be more encompassing, while allowing them to be integrated into a Sid.

Pacio will also provide dynamic data via its “Authenticated Accounting and Real Time Feed for Oracles or Apps” service. Such data will be categorised according to this directory service, but not itself form a “directory”.

Countries, states/provinces, jurisdictions, regions, cities/towns, and groupings of these

Languages

FIAT and crypto currencies

Exchanges (stock and currency/crypto)

Industry/business classifications

Units of weight and measure

Data types – number, string, date, document, image, video, file … any digital thing etc with optional restrictions and formatting information (in and out) where relevant, potentially varied according to jurisdiction and/or language

Activities – all business and human activities e.g. ‘retail sale’, ‘accounting’, programming, exercising etc, keyed to units where applicable e.g. truck driving and tonne kilometres

Entity types from sole proprietorship to public limited company with all the of many variations in use world wide

Functional roles – expense, sale, equity, fixed asset etc

People’s roles - director, partner, officer, remote worker etc

Report names or headings such as “Balance Sheet” allowing for language and jurisdictional variations

* Other groupings that be required or arise over time

The Pacio directories will be open source and accessible to all as a Directories Service.

A Sid could consist of just directory references.

## Digital Id Service

Pacio will provide a digital id service for people and entities which works in conjunction with other services:

Open source technology and standards as listed in [Decentralized Digital Identities and Blockchain](https://cloudblogs.microsoft.com/enterprisemobility/2018/02/12/decentralized-digital-identities-and-blockchain-the-future-as-we-see-it/):

* [Decentralized Identity Foundation (DIF)](http://identity.foundation/)
* [Decentralized Identifiers (DIDs)](https://w3c-ccg.github.io/did-spec/) – a W3C spec that defines a common document format for describing the state of a Decentralized Identifier
* [Identity Hubs](https://github.com/decentralized-identity/hubs/blob/master/explainer.md) – an encrypted identity datastore that features message/intent relay, attestation handling, and identity-specific compute endpoints.
* [Universal DID Resolver](https://medium.com/decentralized-identity/a-universal-resolver-for-self-sovereign-identifiers-48e6b4a5cc3c) – a server that resolves DIDs across blockchains
* [Verifiable Credentials](https://w3c.github.io/vc-data-model/) – a W3C spec that defines a document format for encoding DID-based attestations.

Existing digital identity participants:  
[Civic](https://www.civic.com/), [DID](https://decentralized.id/) (Decentralized ID), [Essentia](https://essentia.one/), Estonia, [Legal Entity Identifier (LEI)](https://www.globallei.com/), [OpenID](http://openid.net/connect/), [Persona](https://persona.im/), [uPort](https://www.uport.me/), the United Nations ID2020 programme, [VeriMe](https://www.verime.net/)

Other initiatives that gain traction

As with all aspects of Pacio, the Digital Identity service will be open, and designed to readily accommodate new standards or services as they become available.

A Pacio Digital Id can form part of a Sid.

## Ontologies

SSIM will use ontologies for the detailed description of data and the relationships between data.

Any element stored using SSIM will reference an Ontology via its Sid.

Any number of ontologies may be created.

An app will select or use the ontologies applicable to its needs.

General purpose ontologies will be developed and maintained by Pacio, but other entities may also create specific purpose ontologies if they so wish.

Ontologies may be created for any purpose, but the initial ones to be built by Pacio will be intended for business and financial data. These ontologies may be structured in accordance with accounting standards e.g. a US GAAP focused one, or an IFRS focussed one, but this is not necessary given the interfaces of the following section. It will, in fact, be a goal of Pacio to produce more general or fundamental ontologies, which are universal without the anglo-american biases of the current IFRS and US GAAP views of world, and which can be used to generate reports according to any desired accounting standard.

SSIM ontologies are described further in section *3 Ontologies.*

## Interfaces

Interfaces will allow conversion of data organised via one ontology to other systems, at least to the extent that other systems can replicate the information depth of the SSIM ontology. In cases where the target system lacks equivalents for some SSIM ontology features, information will necessarily be lost on conversion.

Interfaces will be built for:

One SSIM ontology to another e.g. a “US GAAP” ontology to an IFRS one.

Ontology upgrades i.e. from the 2018 ontology to the 2019 ontology

SSIM ontology to XBRL taxonomy and thence to XBRL reports

SSIM ontology to W3C type ontologies

SSIM ontology to schemas as in [Schema.org](https://schema.org/)

SSIM ontology to any other required "ontology" as per [Ontologies Ontologies Everywhere – but Who Knows What to Think?](https://protege.stanford.edu/conference/2006/submissions/slides/1.2_Uschold.pdf) for which an interface can be written

In some cases, according to need, the interfaces could be both ways i.e. export and import.

# Ontologies

SSIM ontologies will define how data is structured, organised, and related. Apps will use ontologies to categorise the data they are creating, and, if alternatives should exist for a particular item, to prompt a user for a decision, all without users (other than ontology creators/maintainers) needing to know anything about the ontology being used.

Ontologies will use:

Directory references from the directories service. For example, functional roles can be attached to any ontology component, within the limitations that the component might apply. Such functional roles are similar to XBRL roles but are universal, not particular to any one ontology.

Realms, which define groupings of domains, next item. Realms are somewhat like the XBRL concept of hypercubes, but are more powerful and mandatory for all data, unlike for XBRL hypercubes re XBRL categorised data.

Domains provide the next level down of categorisation, and a ‘home’ for elements, the next item. A domain can be a member of multiple realms and must be a member of at least one.

Elements or concepts are the lowest level of categorisation. Elements are members of domains. An element can be a member of only one domain. Elements have attributes for:

* Data type (number, string etc) from the Data Type Directory, which could also provide formatting information
* Concrete or not, where concrete elements can hold data. Non-concrete elements can provide structure or headings etc but not hold data.
* Level from 0 upwards for presentation and summing purposes. For presentation purposes a higher level element would typically be indented vs a lower level element preceding it. For summing purposes a set of numeric elements at a given level would typically sum to the lower level element preceding the set, though other attributes below may be used to override or vary this.
* Read only (or Report only) or not. Read only elements cannot be used for a data being added to the database, but they could, if of an appropriate data type, be the target of a summing operation.
* Optional summing rules for numeric elements
* Optional exclusivity or mutual exclusivity rule vs other elements
* Optional additional to rule – such elements cannot be used by themselves

SSIM Data Objects or Sidos pronounced “sid oh s”)

which provide properties with order, hierarchy, and relationships as per structured (linked data) ontologies

while also allowing for properties without order or hierarchy as per tags in the article Goodbye, Object Oriented Programming

and its conclusion for the way to overcome problems with inheritance in OO languages of using tags:

Tags have no order or hierarchy. (This solves the Diamond Problem too.)

Tags are analogous to interfaces since you can have multiple types associated with the document.

The author's tags are also analogous to our properties, properties being a tag plus attributes and members.

SOSO properties will be dynamic unlike XBRL ones e.g. Director.MaryLou, and support semantic triple type relationships as per WC3's RDF but done using test based tags and relationship operators without complex url/XML schemes.

Ontology "truth" re structure and relationships will be achieved by a consensus methodology to be developed by Pacio - a wikipedia like open source world wide group of interested people but with better consensus courtesy of blockchain to avoid trolling and barrow pushing. Inspired in part by the FOAM Project and the Noam Chomsky Knowledge Graph project. Must allow for updating though. Cannot be immutable.

Ontologies will allow importing of existing ontologies, relevant knowledge graphs, and taxonomies especially XBRL ones.

Ontologies will be stored in a "Global DB". DB to be decided. BigchainDB or TiDB/TiKV as per ont.io

There will be many ontologies, ultimately thousands of them for all types financial reporting requirements.

These smart object/DB based ontologies will be much more powerful than other methods of expressing semantic data relationships such as W3C's OWL ontologies or XBRL Taxonomies yet will be easier to work with and understand.

Developing and maintaining the ontologies will be one of the things that PCL will do, under a nominally different name akin to the original idea of having a separate Pacio Foundation for this sort of stuff. Pacio.org is not available though it is not in use. Pacio.app is available though.

# Data Storage

# Sid Service