TRUST Over IP FOUNDATION

Decentralized Semantics WG Weekly Meeting

14 July 2020

THE LINUX FOUNDATION

Antitrust Policy Notice

- Linux Foundation meetings involve participation by industry competitors, and it is the intention of the Linux Foundation to conduct all of its activities in accordance with applicable antitrust and competition laws. It is therefore extremely important that attendees adhere to meeting agendas, and be aware of, and not participate in, any activities that are prohibited under applicable US state, federal or foreign antitrust and competition laws.

- Examples of types of actions that are prohibited at Linux Foundation meetings and in connection with Linux Foundation activities are described in the Linux Foundation Antitrust Policy available at http://www.linuxfoundation.org/antitrust-policy. If you have questions about these matters, please contact your company counsel, or if you are a member of the Linux Foundation, feel free to contact Andrew Updegrove of the firm of Gesmer Updegrove LLP, which provides legal counsel to the Linux Foundation.

Agenda

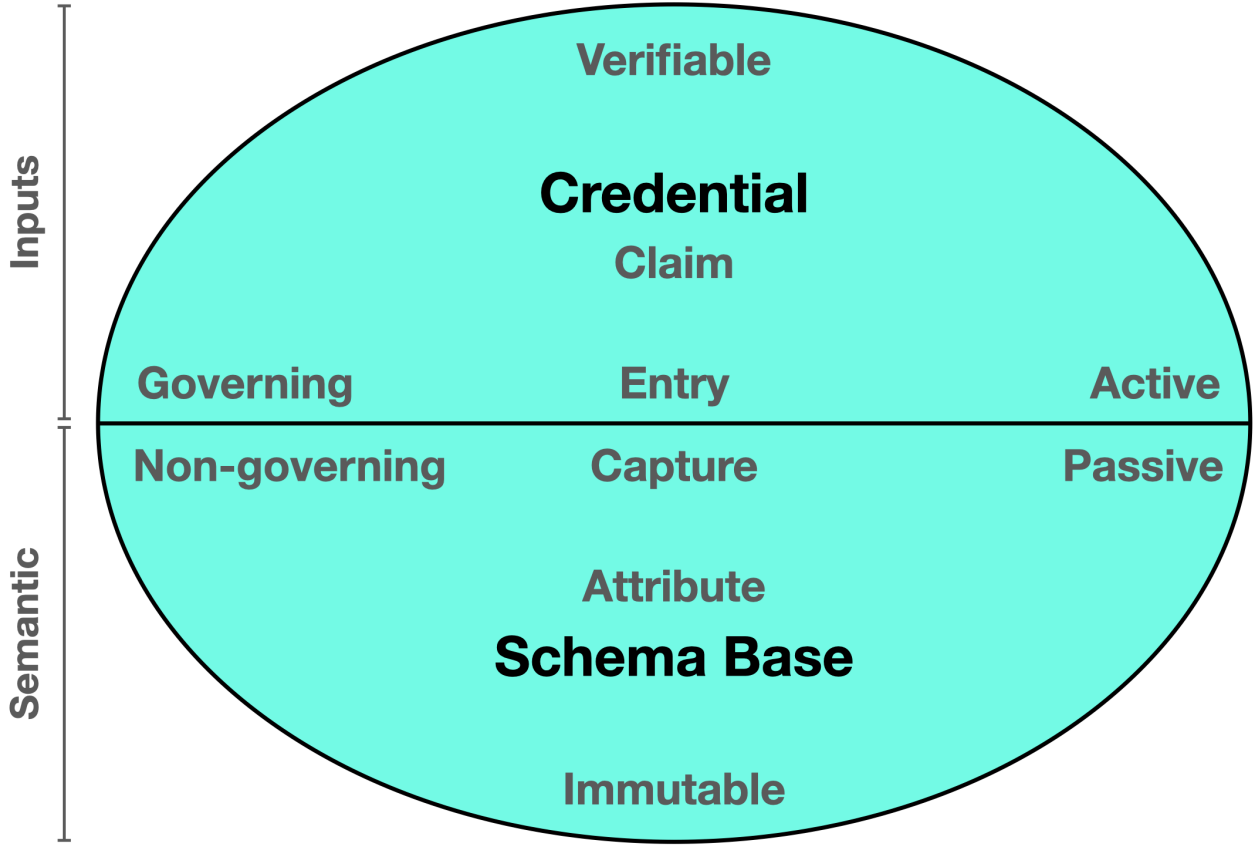
1. Welcome (Paul—5 min)
2. Mission and scope of this WG (Paul—5 min)
3. Newcomer Introductions (5 min)
4. Presentation: Overlays Capture Architecture (Paul—20 min)
5. Live demo: Overlays Capture Architecture (Robert—20 min)
6. Logistics (Paul—5 min)
    a. Chairs
    b. Meeting schedule

Mission and Scope of the Decentralized Semantics WG

Paul Knowles, Human Colossus Foundation

(5 min)

Characteristics of a decentralized network



Ref.: https://humancolossus.foundation/blog/active-and-passive-identifiers

Mission and scope

The mission of the WG is to define a data capture architecture consisting of immutable schema bases and interoperable overlays for Internet-scale deployment. The scope of the WG is to define specifications and best practices that bring cohesion to data capture processes and other Semantic Standards throughout the ToIP stack, whether these standards are hosted at the Linux Foundation or external to it. Other WG activities will include creating template Requests for Proposal (RFPs) and additional guidance to utility and service providers regarding implementations in this domain. This WG may also organise Task Forces to escalate the development of certain components if deemed appropriate by the majority of the WG members and in line with the overall mission of the ToIP Foundation.

Newcomer Introductions (30 seconds!)

1. Name
2. Location / time zone
3. Affiliation(s)
4. One-sentence summary of your interest in Decentralized Semantics (or one particular semantics-related issue you personally want to see solved)

Presentation: Overlays Capture Architecture (OCA)

A standardized global solution for data capture

July 14th, 2020

Paul Knowles
Stem Cell, The Human Colossus Foundation
Co-convener of the Decentralized Semantics WG

Email: paul.knowles@humancolossus.org

What is a schema?

A machine-readable definition of the semantics of a data structure.

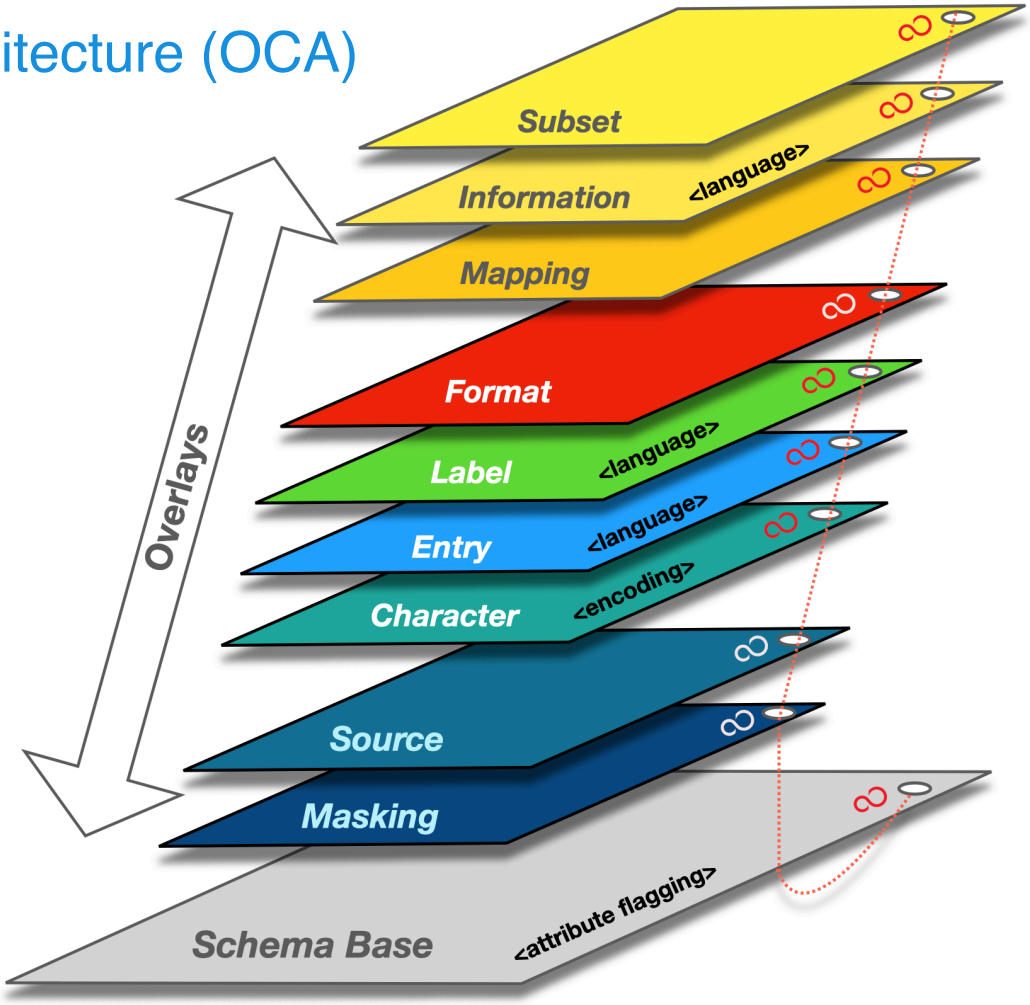
What is a schema base?

A stable base object that defines a single set of data in its purest form thus providing a standard base from which to decentralize data. A Schema Base facilitates a Blinding Identity schema object which allows the issuer to flag attributes that could potentially unblind the identity of a governed entity.

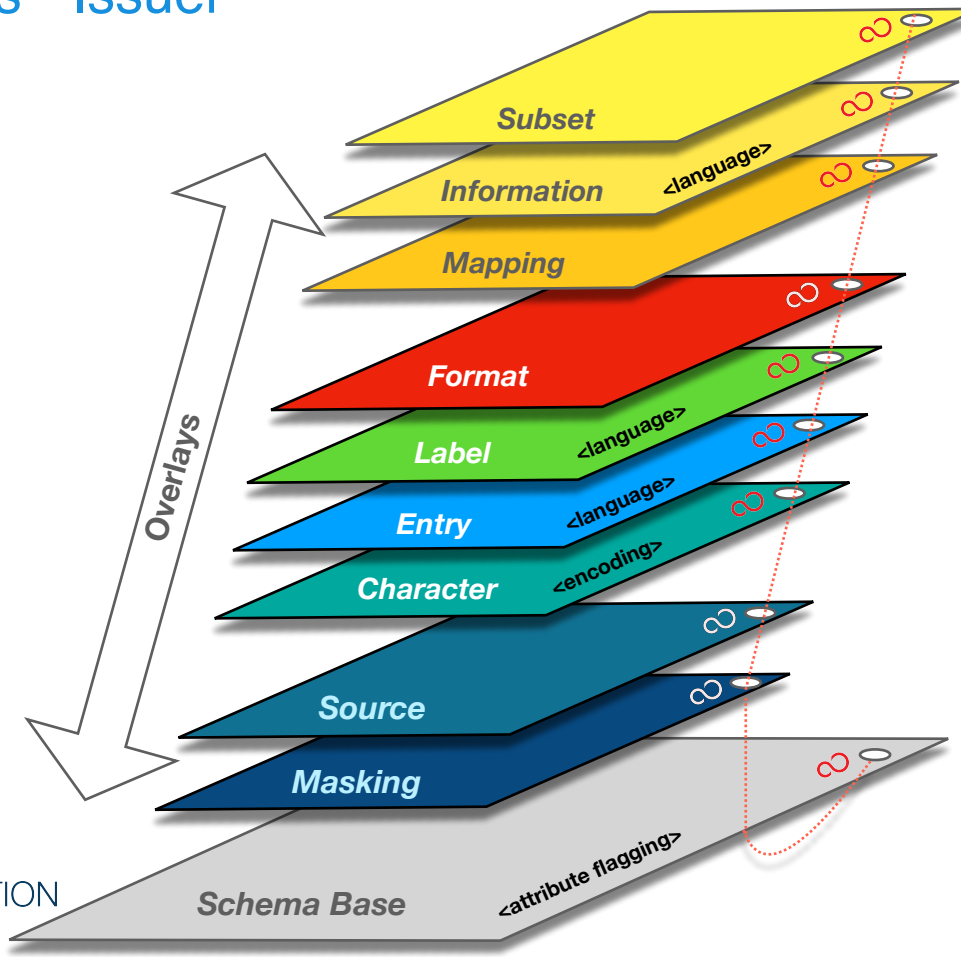
What is an overlay?

A linked object that provides an extra layer of contextual and/or conditional information to a Schema Base. This extra context can be used by an issuer to transform how information is displayed to a viewer or to guide a verifier or holder in how to apply a custom process to schema data.

Overlays Capture Architecture (OCA)



Overlays Types - Issuer



*A **Subset Overlay** is used to create a schema subset.*

*An **Information Overlay** is used to add instructional, informational or legal prose to assist the data entry process.*

*A **Mapping Overlay** is used to map schema attributes to internal system identifiers.*

*A **Format Overlay** is used to add formats to schema attributes (incl. field lengths).*

*A **Label Overlay** is used to add labels to schema attributes (incl. category labels).*

*An **Entry Overlay** is used to add predefined field values to schema attributes.*

*A **Character Overlay** is used to define character set encoding (e.g. UTF-8, ISO-8859-1, Windows-1251, Base58Check, etc.).*

*A **Source Overlay** is used to specify endpoints where dynamic data or linked attachments can be located.*

*A **Masking Overlay** is used to add data masks to schema attributes that have been flagged in the Schema Base.*

*A stable base object that defines a single set of data in its purest form thus providing a standard base from which to decentralize data. A **Schema Base** facilitates a Blinding Identity schema object which allows the issuer to flag attributes that could potentially unblind the identity of a governing entity.*

Why is OCA useful?

Data pooling. Decoupling can occur at any time as overlays are linked objects. With all coloration stored in the overlays, combining data from related sources becomes much easier. Overlays can be removed from the base objects before the data merging process begins and reapplied to ensure consistent coloration post data pooling.

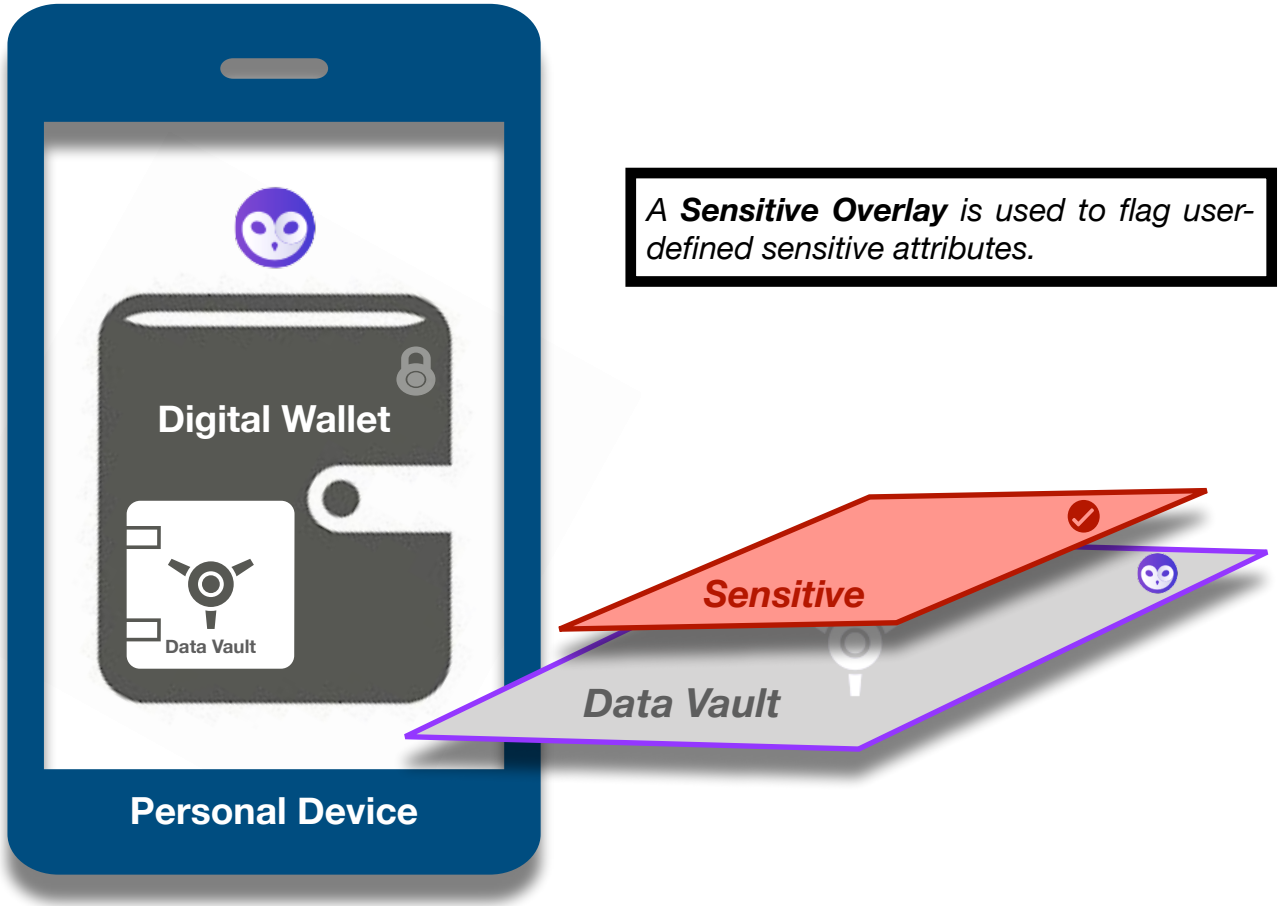
Stable schema bases. Most schema updates tend to be done at the application stage. In the case of ODCA, all extension, coloration, and functionality definitions are applied in the overlays. This enables issuers to edit one or more of the linked objects to create simple updates rather than having to reissue schema bases on an ongoing basis.

Flagged attributes for encryption. Using the Blinding Identity Taxonomy (BIT) as a reference, issuers can flag attributes in the schema base that could potentially unblind the identity of a governed entity. With attributes flagged at the base object layer, all corresponding data can be treated as sensitive throughout the data lifecycle and encrypted or removed at any stage making associated entity identification impossible.

Data decentralisation. Schema base definitions can remain in their purest form thus providing a standard base from which to decentralise data. Once the data holder has given adequate consent, data controllers can contribute anonymous data to decentralised data lakes upon which 3rd parties can trigger accurate criteria searches for matched data. This eliminates the need for data silos and encourages consented data sharing. The data holder is empowered by self-determination regarding secondary use of their personal data.

Internationalisation. As character set encoding definitions are captured in a separate linked data object, a single report definition can contain different attribute forms for different languages available to users, based on a user's locale and other language preferences.

Overlays Types - Holder



*A **Sensitive Overlay** is used to flag user-defined sensitive attributes.*

Blinding Identity Taxonomy (BIT)

General Data Protection Regulation (GDPR) :

Deficiencies from a tech implementation perspective

- The need for a common standard to help protect the identity of a governed entity.
- Introducing the …

Blinding Identity Taxonomy (BIT)

Ref.: https://kantarainitiative.org/download/blinding-identity-taxonomy-pdf/

Blinding Identity Taxonomy (BIT)

- Names (incl. First Names, Last Names, Full Names, Entity Names)
- Physical Addresses
- E-mail Addresses
- Telephone Numbers
- Postal Codes
- Personal Software Application Handles (e.g. Skype, Slack, Hyperledger Chat, etc.)
- Profile Pages
- Passport Numbers
- Social Security Numbers
- National Insurance Numbers
- Driving License Numbers
- Vehicle Registration Numbers
- Bank Account Numbers
- Credit (or Debit) Card Numbers
- Personal Identification Numbers (PIN)
- Private Keys / Master Keys
- Symmetric Keys
- Public Keys
- Link Secrets
- Employee Identifiers
- Account Identifiers
- Governmental Identifiers
- Membership Identifiers (e.g. Trade Union Membership, etc.)
- Institutional Identifiers (e.g. Private Health Care Identifiers, etc.)
- Case Identifiers (e.g. Case ID Numbers, Benefit Plan Participation Identifiers, etc.)
- User Identifiers (e.g. User IDs, Logins, etc.)
- Passwords
- Signatures
- Digital Certificates
- Photos
- Videos
- Images
- Vocal Sound Bites
- Dates and timestamps (e.g. Date of Birth, transaction dates, etc.)*
- Genetic Identifiers (incl. chromosomal, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) data)
- Biometric Identifiers (incl. voiceprints, iris scans, facial imaging and dactyloscopic (fingerprint) data)
- Internet Protocol (IP) Addresses
- Media Access Control (MAC) Addresses
- Service Set Identifiers (SSID) (incl. local WiFi SSIDs)
- Bluetooth Device Addresses (BD_ADDR)
- Locational Information (incl. Global Positioning System (GPS), 3 word address, etc.)
- Cookie Browser Identifiers
- Radio Frequency Identifiers
- IoT Identifiers (incl. smart meter data)
- International Mobile Equipment Identity (IMEI)
- International Mobile Subscriber Identity (IMSI)
- Social media interactive elements, posts and comments (incl. likes, emojis and polling results)
- Free-Form Text Fields / Unstructured Data**

* Note: Not all captured dates will reveal identity but some will so, if in doubt, encrypt.

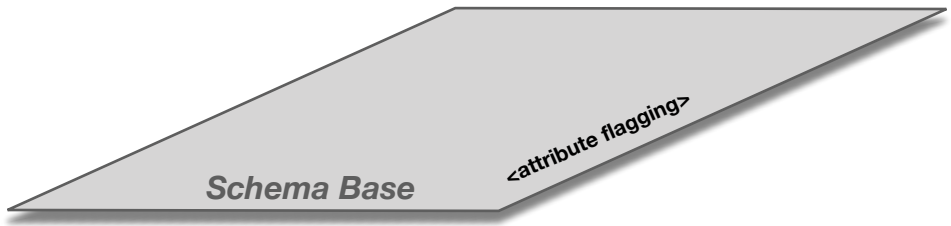
** Defn.: Text which does not have a given structure, nor which is entered in any specific format. Note: All free-form text fields should be encrypted.

Ref.: https://kantarainitiative.org/download/blinding-identity-taxonomy-pdf/

Creating a Schema Base with linked Overlays

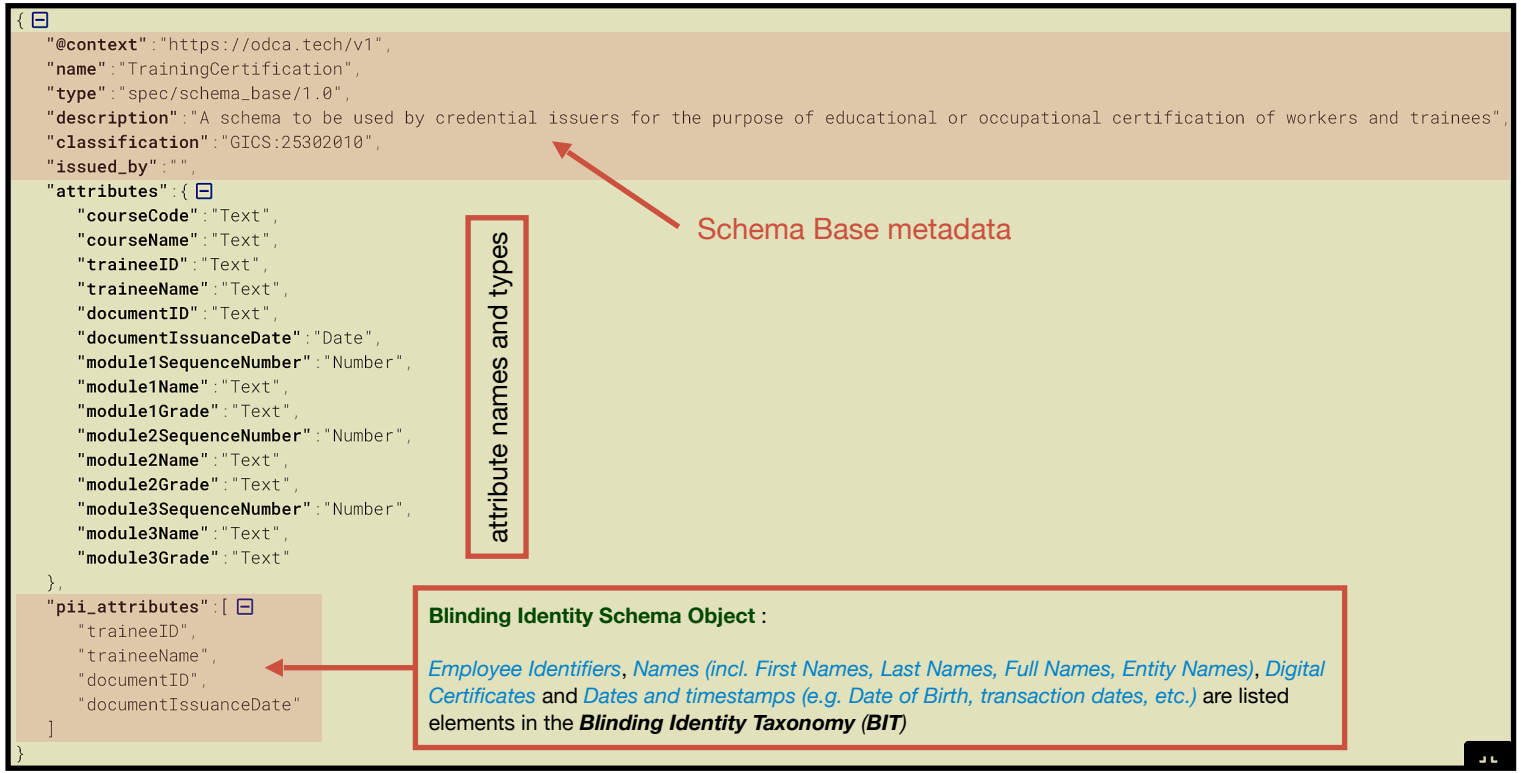
Schema Base

*A stable base object that defines a single set of data in its purest form thus providing a standard base from which to decentralize data. A **Schema Base** facilitates a Blinding Identity schema object which allows the issuer to flag attributes that could potentially unblind the identity of a governed entity.*

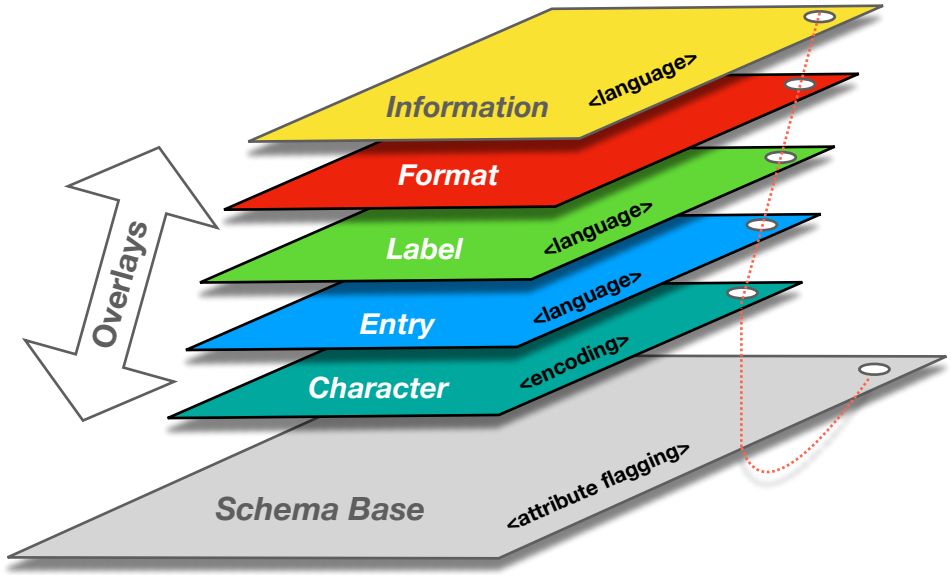


Schema Base

“Training Certification”



Overlays


Information
<language>
Format
Label
<language>
Entry
<language>
Character
<encoding>
Overlays
Schema Base
<attribute flagging>

Character Encoding Overlay

*A **Character Overlay** is used to define character set encoding (e.g. UTF-8, ISO-8859-1, Windows-1251, Base58Check, etc.).*

```
{
   "@context":"https://odca.tech/overlays/v1",
   "type":"spec/overlay/character_encoding/1.0",
   "issued_by":"",
   "role":"",
   "purpose":"",
   "schema_base":"hl:38h8993Xy8gPx2pzpHfUbqde8B5pwuW8sZarWXUqxJHp",
   "default_character_encoding":"utf-8",
   "attr_character_encoding":{
      "courseCode":"utf-8",
      "courseName":"utf-8",
      "traineeID":"utf-8",
      "traineeName":"utf-8",
      "documentID":"utf-8",
      "documentIssuanceDate":"utf-8",
      "module1SequenceNumber":"utf-8",
      "module1Name":"utf-8",
      "module1Grade":"utf-8",
      "module2SequenceNumber":"utf-8",
      "module2Name":"utf-8",
      "module2Grade":"utf-8",
      "module3SequenceNumber":"utf-8",
      "module3Name":"utf-8",
      "module3Grade":"utf-8"
   }
}
```

Overlay metadata

Hashlink (cryptographic hyperlink) referencing the “Training Certification” schema base

Default character set encoding

Format Overlay

*A **Format Overlay** is used to add formats to schema attributes (incl. field lengths).*

```
{
   "@context":"https://odca.tech/overlays/v1",
   "type":"spec/overlay/format/1.0",
   "issued_by":"",
   "role":"",
   "purpose":"",
   "schema_base":"hl:38h8993Xy8gPx2pzpHfUbqde8B5pwuW8sZarWXUqxJHp",
   "attr_formats":{
      "documentIssuanceDate":"MMM YYYY"
   }
}
```

← Overlay metadata

← Hashlink (cryptographic hyperlink) referencing the “Training Certification” schema base

Label Overlay

*A **Label Overlay** is used to add labels to schema attributes (incl. category labels).*

Entry Overlay

Overlay metadata

Hashlink (cryptographic hyperlink) referencing the “Training Certification” schema base

Language definition

*An **Entry Overlay** is used to add predefined field values to schema attributes.*

Information Overlay

*An **Information Overlay** is used to add instructional, informational or legal prose to assist the data entry process.*


```
{
   "@context":"https://odca.tech/overlays/v1",
   "type":"spec/overlay/information/1.0",
   "issued_by":"",
   "role":"Issuer",
   "purpose":"Record",
   "schema_base":"hl:38h8993Xy8gPx2pzpHfUbqde8B5pwuW8sZarWXUqxJHp",
   "language":"en_US",
   "attr_information":{
      "module1SequenceNumber":"Sequence number in transcript log",
      "module1Grade":"Select from the dropdown list",
      "module2SequenceNumber":"Sequence number in transcript log",
      "module2Grade":"Select from the dropdown list",
      "module3SequenceNumber":"Sequence number in transcript log",
      "module3Grade":"Select from the dropdown list"
   }
}
```

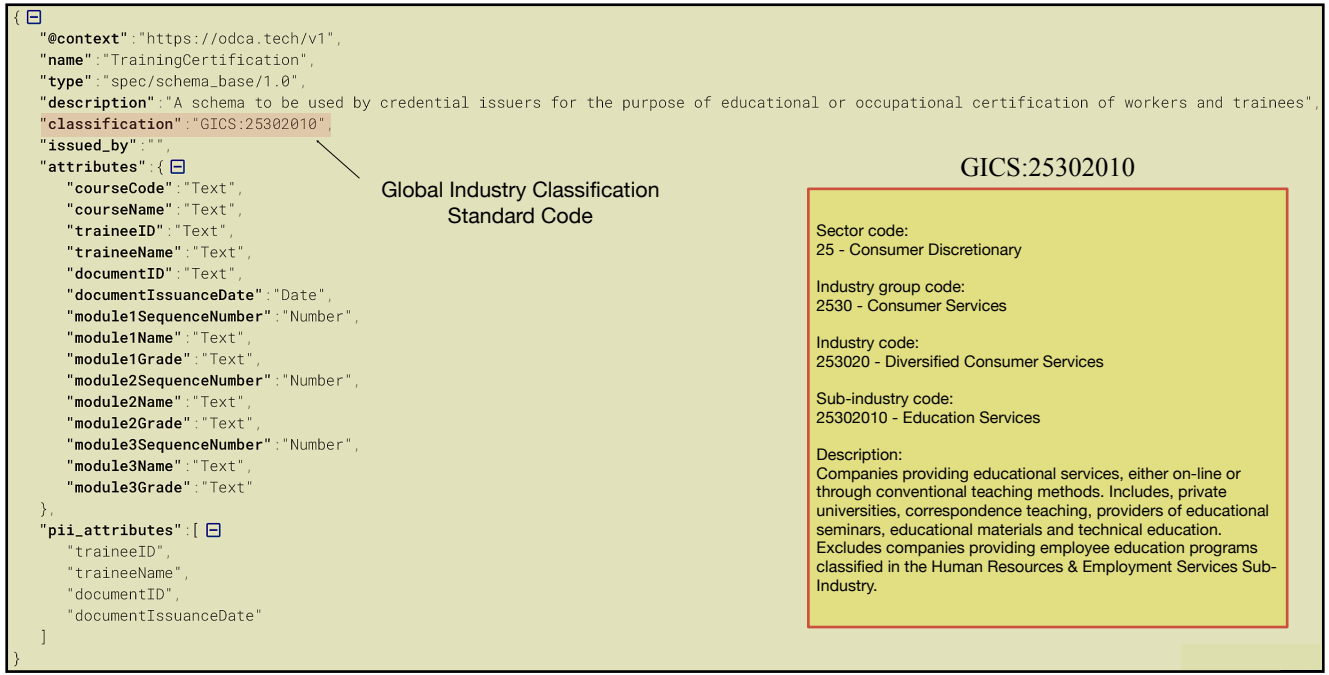

Overlay metadata

Hashlink (cryptographic hyperlink) referencing the “Training Certification” schema base

Language definition

Object interoperability within an Industry Sector

Schema Base: “Classification” Meta Tag



Industry Sector Classification

Option 1

GICS: Global Industry Classification Standard

"The Industry Standard"

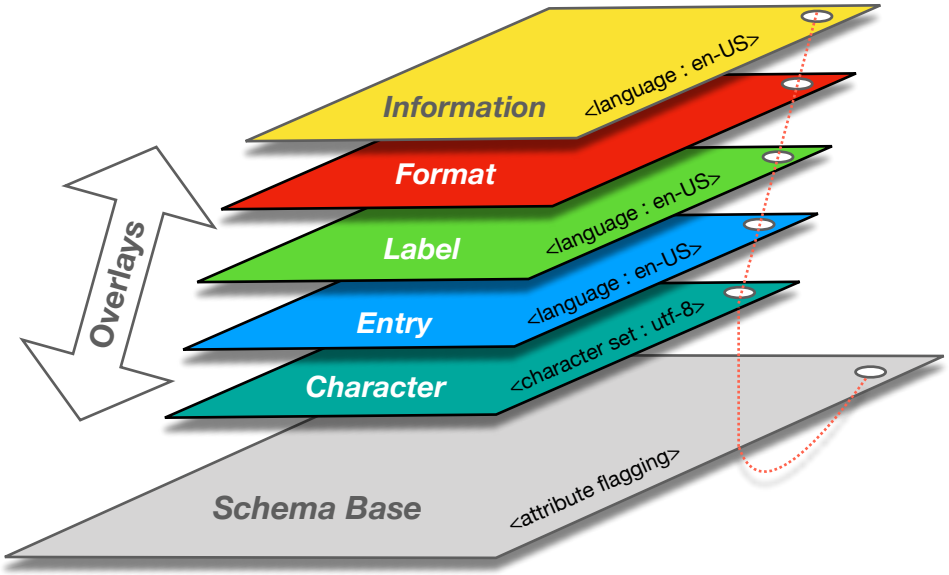
The **GICS** indices is an industry taxonomy for use by the global financial community as a basis to assign companies to a sub-industry, and to an industry, industry group, and sector, by its principal business activity.

- 11 Sectors
- 24 Industry Groups
- 69 Industries
- 158 Sub-Industries

Object interoperability within an Industry Sector

"Training Certification" schema example

"classification": "GICS:25302010"



GICS:25302010

Sector code:
25 - Consumer Discretionary

Industry group code:
2530 - Consumer Services

Industry code:
253020 - Diversified Consumer Services

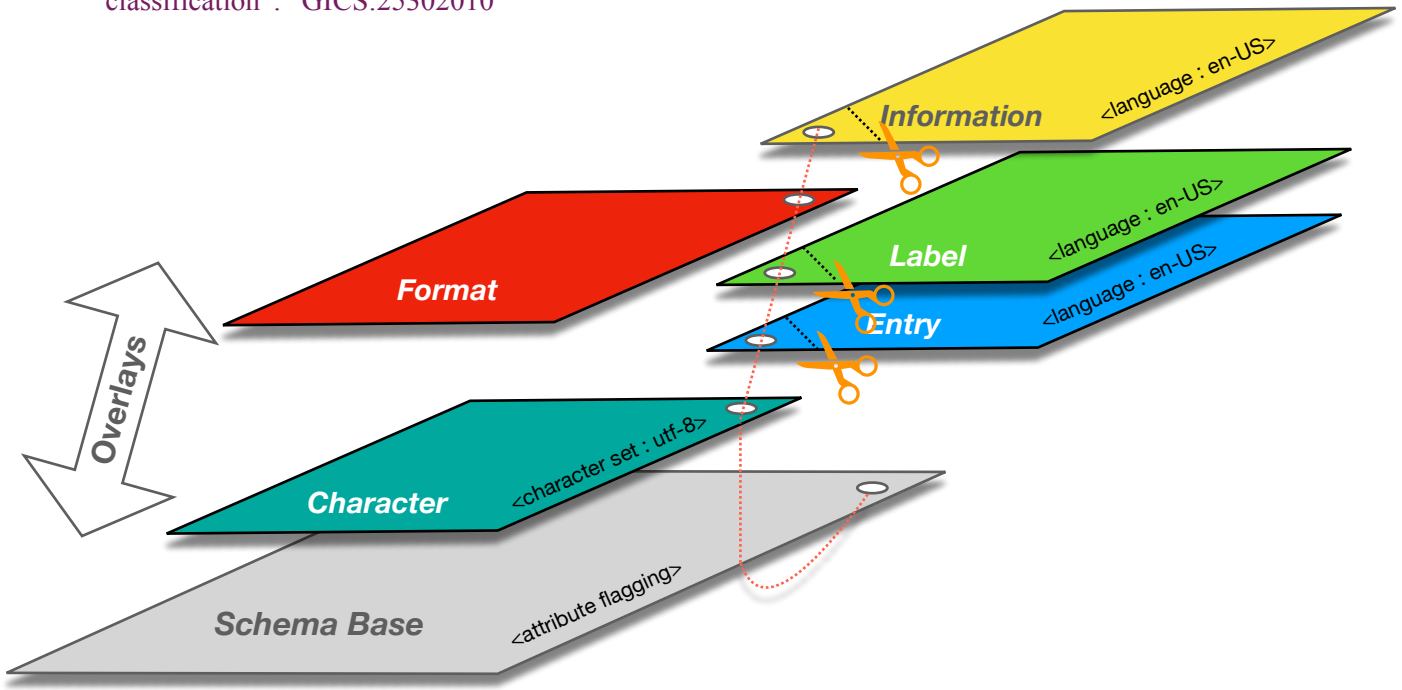
Sub-industry code:
25302010 - Education Services

Description:
Companies providing educational services, either on-line or through conventional teaching methods. Includes, private universities, correspondence teaching, providers of educational seminars, educational materials and technical education. Excludes companies providing employee education programs classified in the Human Resources & Employment Services Sub-Industry.

Object interoperability within an Industry Sector

"Training Certification" schema example

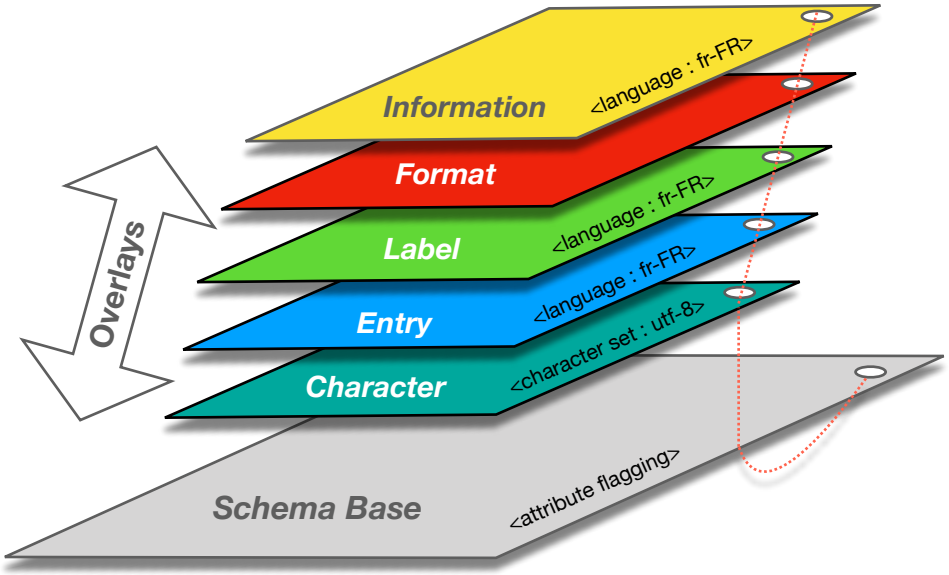
"classification": "GICS:25302010"



Object interoperability within an Industry Sector

"Training Certification" schema example

"classification": "GICS:25302010"



Live demo: Overlays Capture Architecture (OCA)

A standardized global solution for data capture

July 14th, 2020

Robert Mitwicki
Stem Cell, The Human Colossus Foundation
Founding sponsor of the Hyperledger Aries Project
Co-convener of ToIP Decentralized Semantics WG

Email: robert.mitwicki@humancolossus.org

Chairs

- As a Working Group, we elect our own chairs
  - At least one, and up to three
  - Two or three is recommended to spread out the load
- We can periodically rotate chairs as needed
- Volunteers now?

Meeting schedule

- Call timing
  - **ToIP Decentralized Semantics WG**

    Every Tuesday starting

    09:00 PT, 12:00 ET, 17:00 UK, 18:00 CET
- Next meeting
  - July 21st, 2020

Closing Q & A

THE LINUX FOUNDATION

Legal Notices

The Linux Foundation, The Linux Foundation logos, and other marks that may be used herein are owned by The Linux Foundation or its affiliated entities, and are subject to The Linux Foundation's Trademark Usage Policy at https://www.linuxfoundation.org/trademark-usage, as may be modified from time to time.

Linux is a registered trademark of Linus Torvalds. Please see the Linux Mark Institute's trademark usage page at https://lmi.linuxfoundation.org for details regarding use of this trademark.

Some marks that may be used herein are owned by projects operating as separately incorporated entities managed by The Linux Foundation, and have their own trademarks, policies and usage guidelines.

TWITTER, TWEET, RETWEET and the Twitter logo are trademarks of Twitter, Inc. or its affiliates.

Facebook and the "f" logo are trademarks of Facebook or its affiliates.

LinkedIn, the LinkedIn logo, the IN logo and InMail are registered trademarks or trademarks of LinkedIn Corporation and its affiliates in the United States and/or other countries.

YouTube and the YouTube icon are trademarks of YouTube or its affiliates.

All other trademarks are the property of their respective owners. Use of such marks herein does not represent affiliation with or authorization, sponsorship or approval by such owners unless otherwise expressly specified.

The Linux Foundation is subject to other policies, including without limitation its Privacy Policy at https://www.linuxfoundation.org/privacy and its Antitrust Policy at https://www.linuxfoundation.org/antitrust-policy. each as may be modified from time to time. More information about The Linux Foundation's policies is available at https://www.linuxfoundation.org.

Please email legal@linuxfoundation.org with any questions about The Linux Foundation's policies or the notices set forth on this slide.