HUMAN COLOSSUS
FOUNDATION

# Overlays Capture Architecture (OCA)

## Global semantic harmonization

**Paul Knowles**
November 13th, 2020

# What is a Schema?

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

# What is a Schema Base?

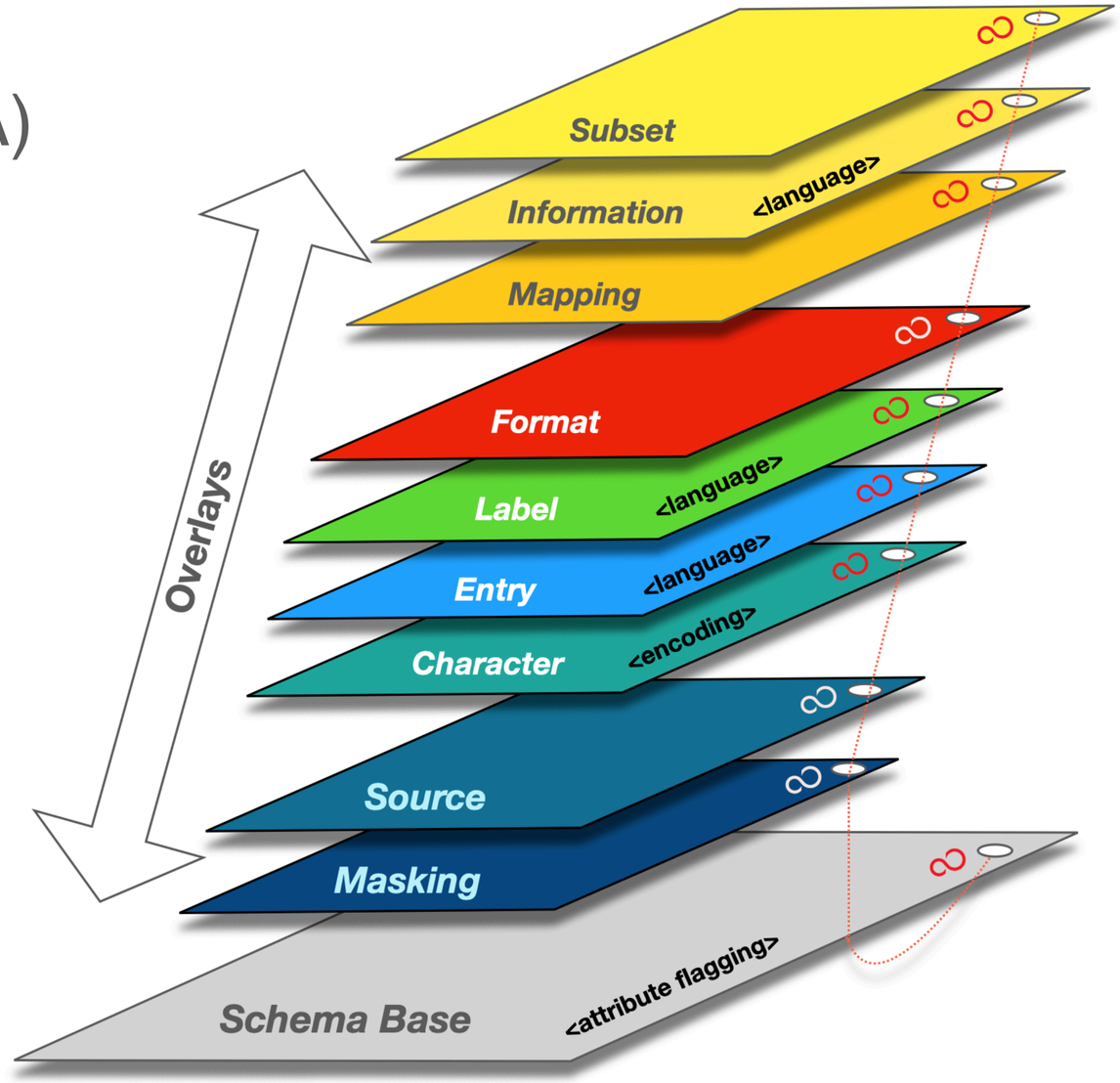
A stable base object that defines a single set of data in its purest form thus providing a standard base from which to decentralize data. A *schema base* contains a blinding block which allows the issuer to flag any attributes that could potentially unblind the identity of a governing entity.

# What is an Overlay?

A linked object that provides an extra layer of contextual and/or conditional information to a schema base. *Overlays* 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)

OCA is an architecture that presents a schema as a multi-dimensional object consisting of a stable *schema base* and interoperable *overlays*. Overlays are task-oriented linked data objects that provide additional extensions, coloration, and functionality to the schema base.



# 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 OCA, 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 governing 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.

# How OCA Differs from JSON-LD and Why You Need Both …

### *What is Linked Data?*

Linked Data is data expressed on a website that can traverse (via links) to other websites. Common formats of data include HTML, JSON, XML, CSV and RDFa. These common data formats can be linked together using **JSON-LD**.

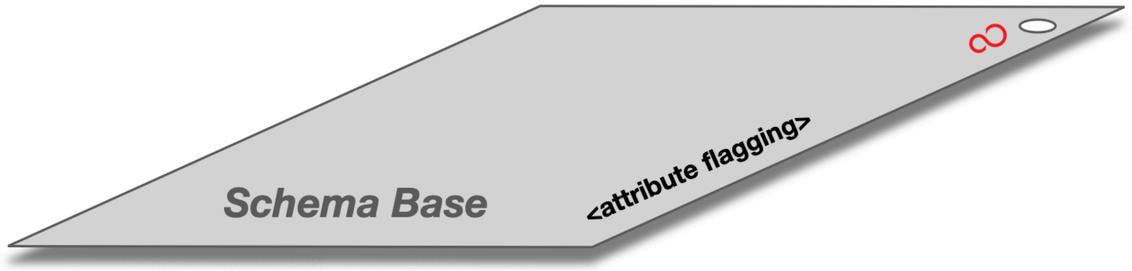
### *What is a Database Model?*

A database model shows the logical structure of a database, including the relationships and constraints that determine how data can be stored and accessed. Common kinds of data models include Hierarchical database model, Relational model, Network model, Object-oriented database model, Entity-relationship model, Document object model, Entity-attribute-value model, Star schema and Object-relational database model. The data semantics across these common database models can be harmonized using **OCA**.

# Overlays Capture Architecture (OCA)

- Issuer: **Schema Base**

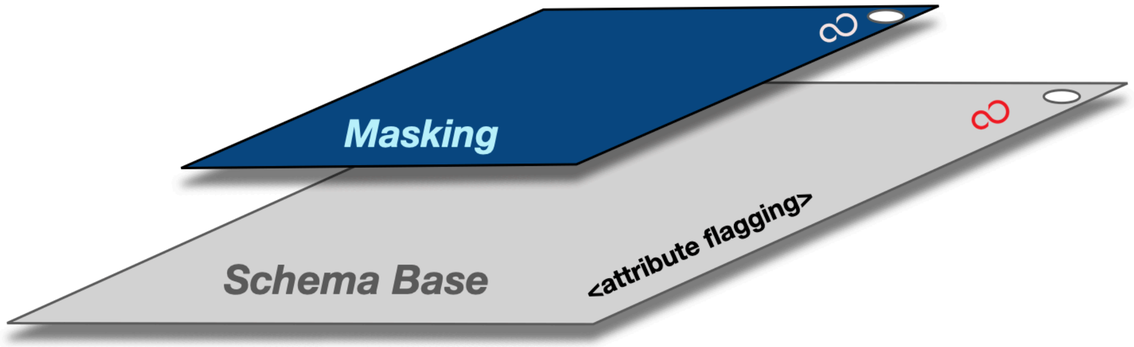
*A stable base object that defines a single set of data in its purest form thus providing a standard base from which to decentralise data. A **Schema Base** contains a blinding block which enables the issuer to flag attributes that could potentially unblind the identity of a governing entity.*



# Overlays Capture Architecture (OCA)

- Issuer: **Masking Overlay**

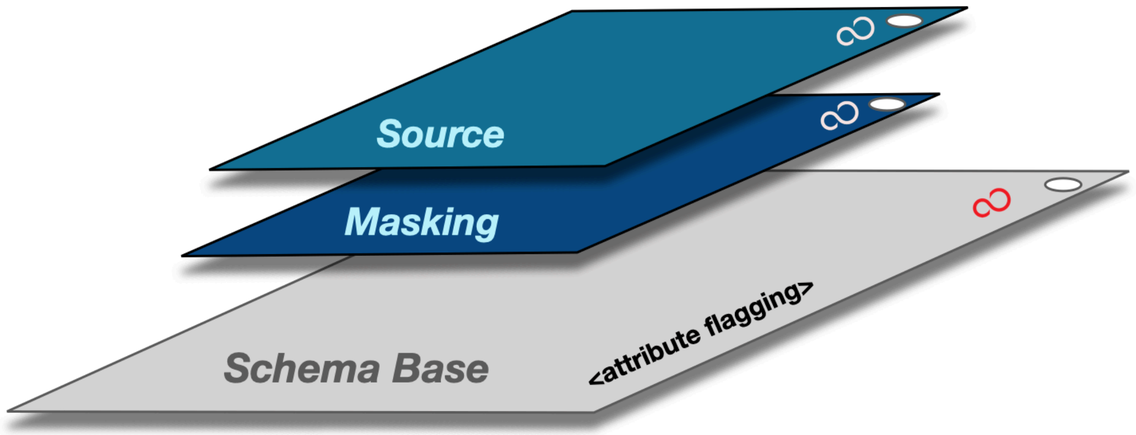
*A **Masking Overlay** is used to add data masks to schema attributes that have been flagged in the Schema Base.*



# Overlays Capture Architecture (OCA)

- Issuer: **Source Overlay**

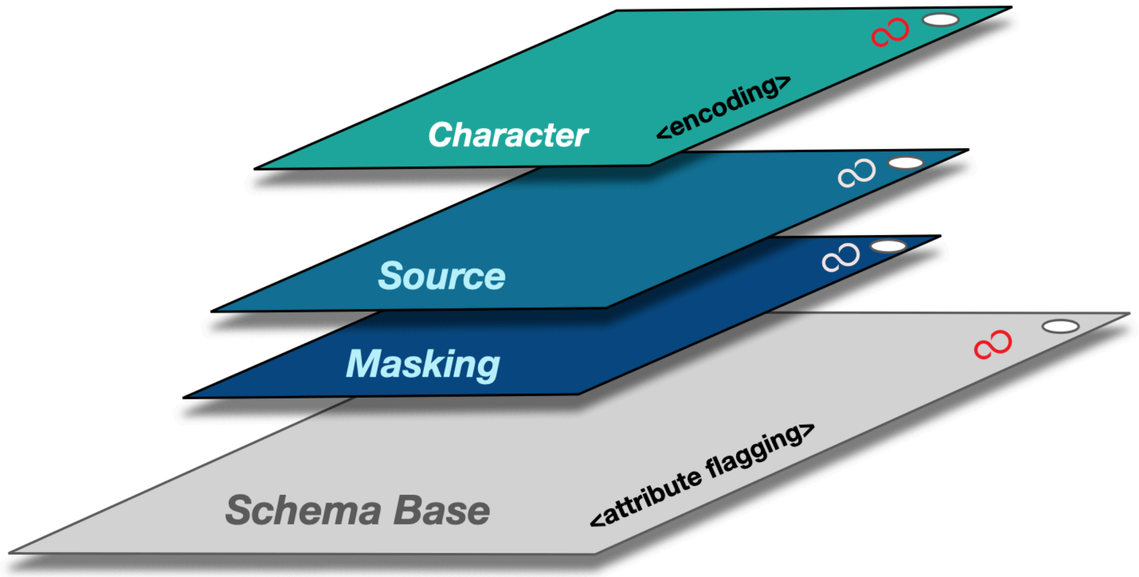
*A **Source Overlay** is used to specify endpoints where dynamic data or linked attachments can be located.*



# Overlays Capture Architecture (OCA)

- Issuer: **Character Encoding Overlay**

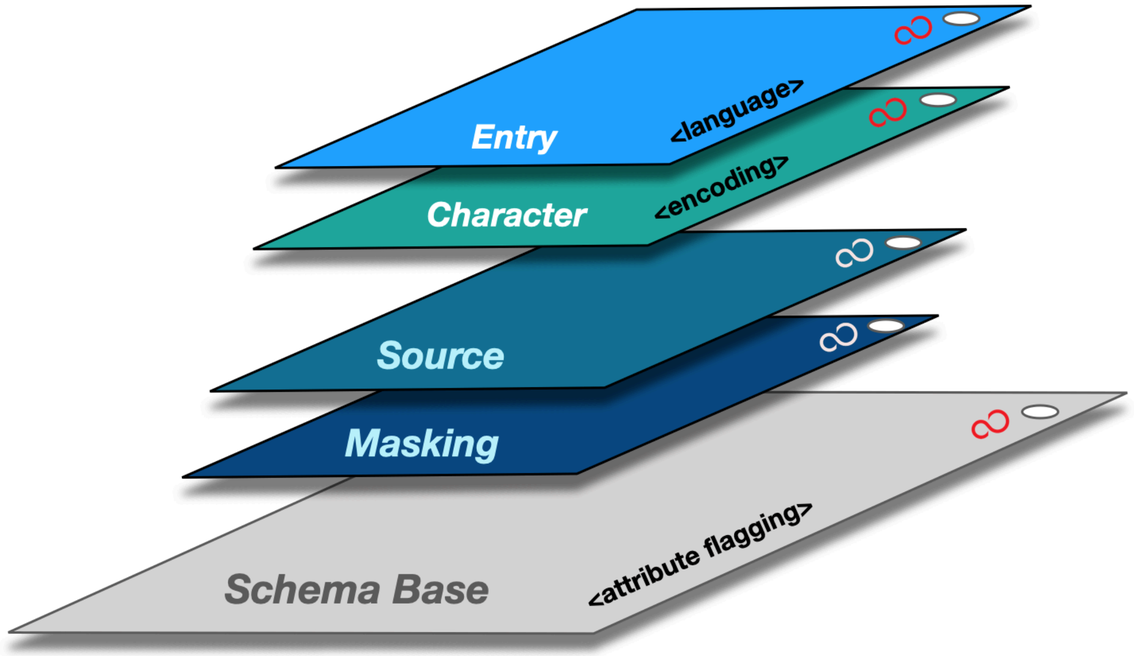
*A **Character Overlay** is used to define character set encoding (e.g. UTF-8, ISO-8859-1, Windows-1251, Base58Check, etc.).*



# Overlays Capture Architecture (OCA)

- Issuer: **Entry Overlay**

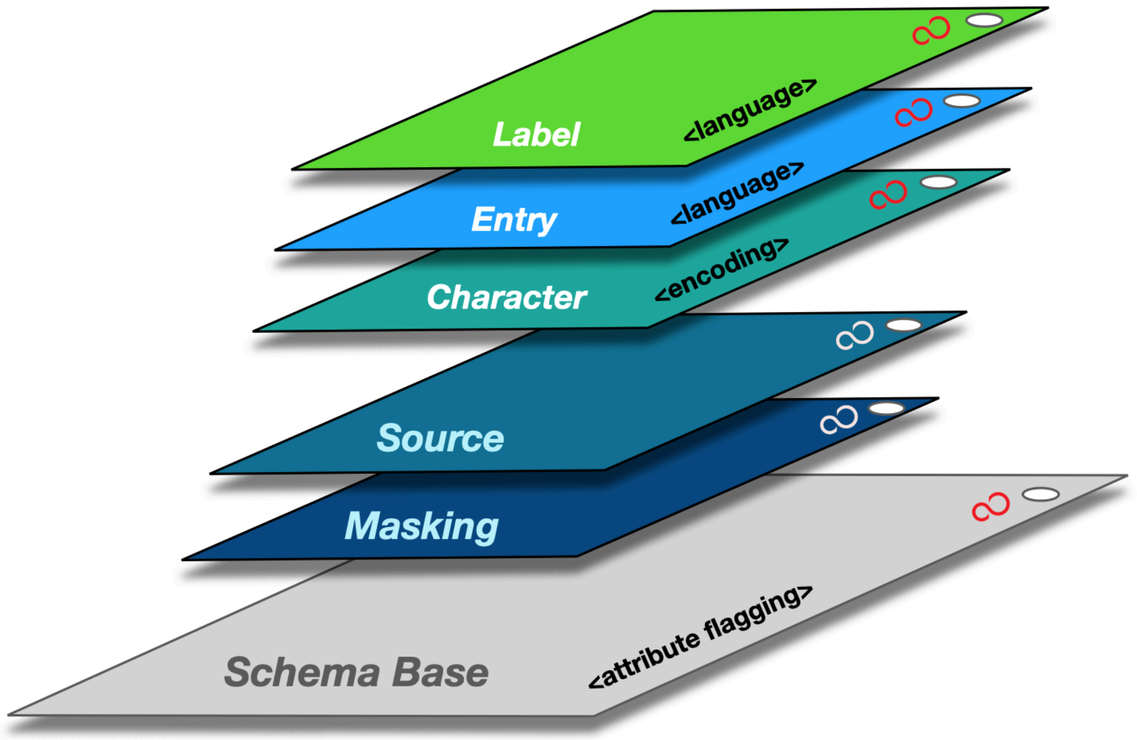
*An **Entry Overlay** is used to add predefined field values to schema attributes.*



# Overlays Capture Architecture (OCA)

- Issuer: **Label Overlay**

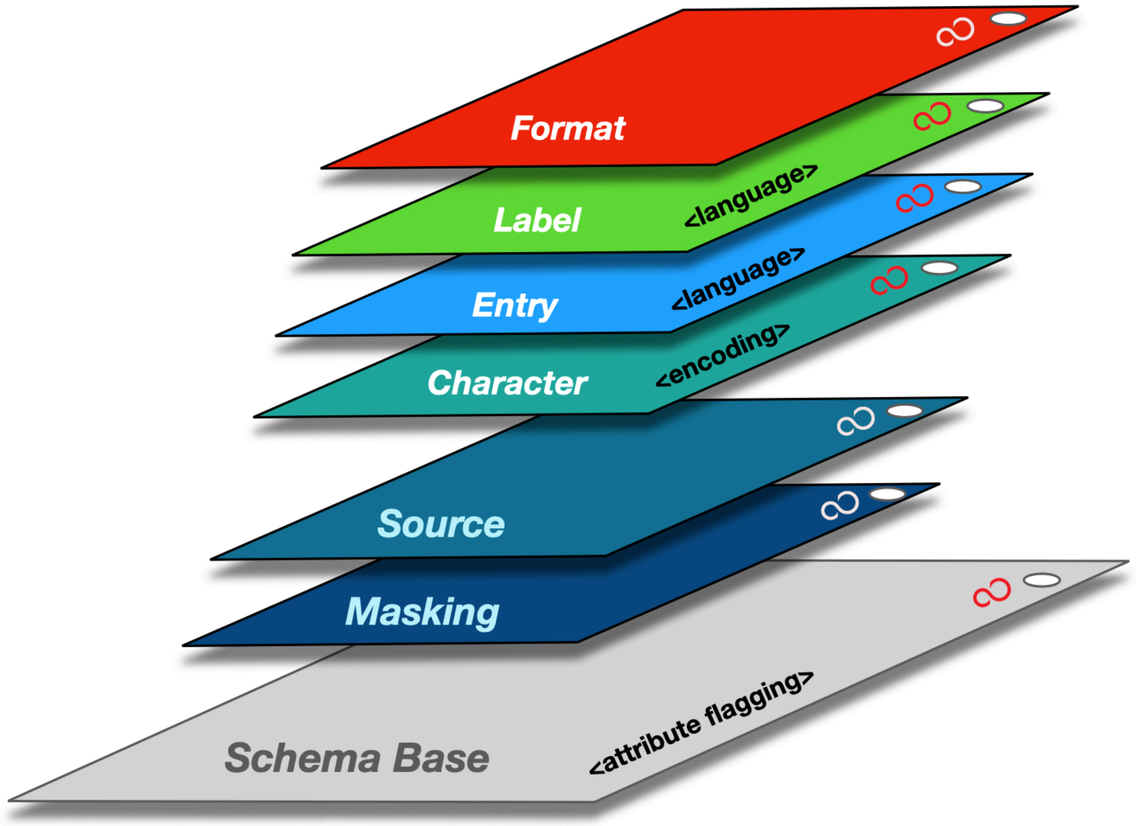
*A **Label Overlay** is used to add labels to schema attributes (incl. category labels).*



# Overlays Capture Architecture (OCA)

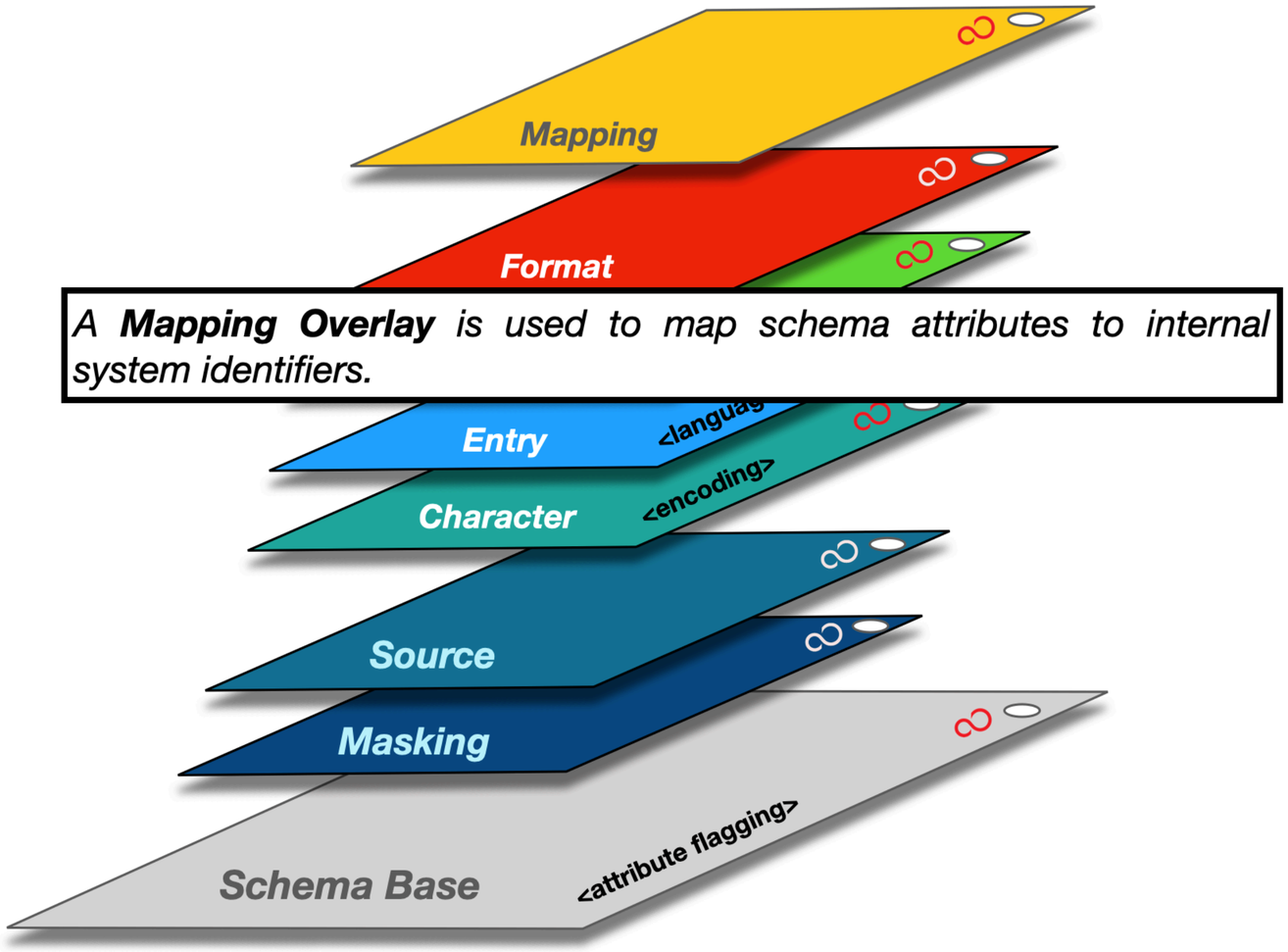
- Issuer: **Format Overlay**

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



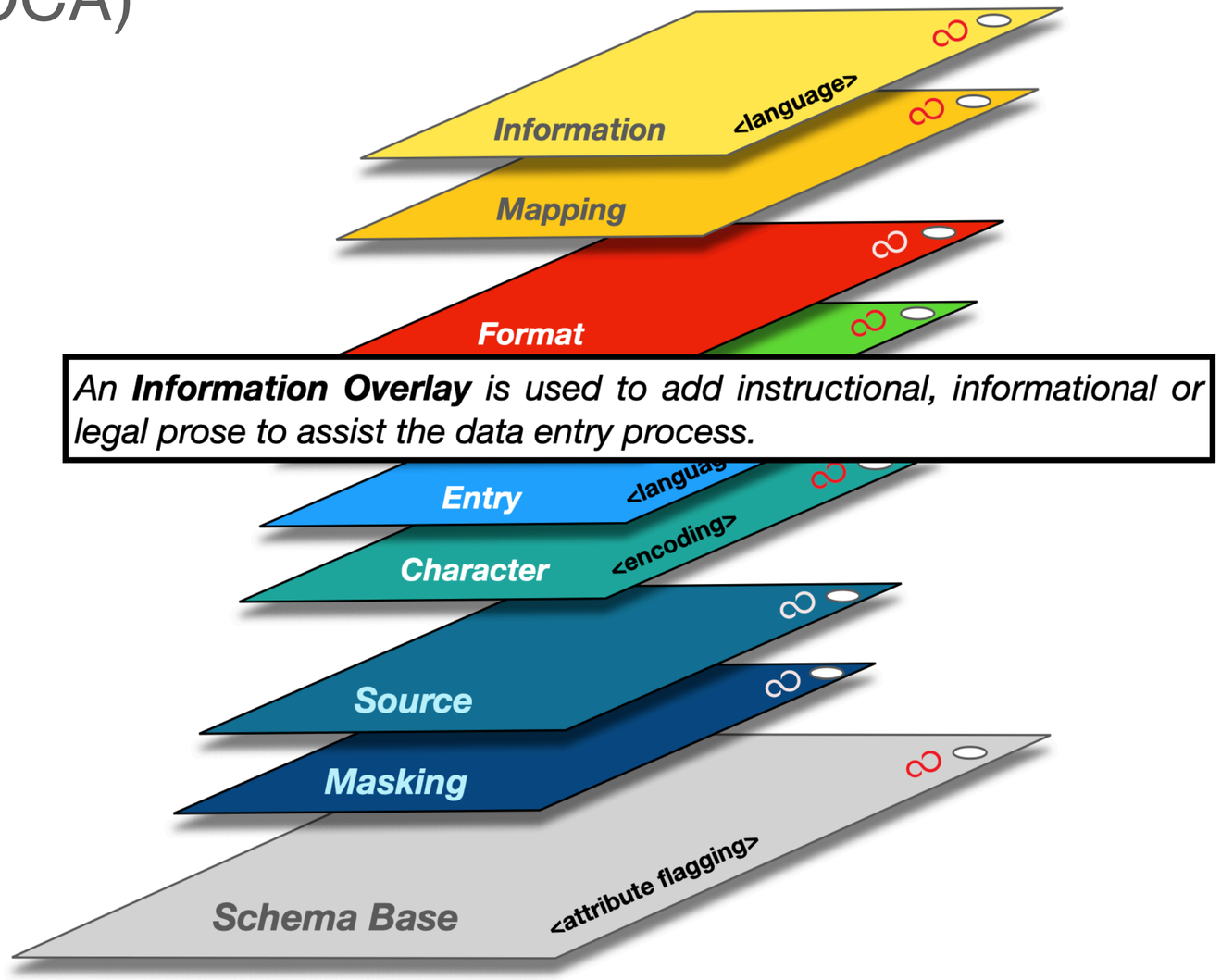
# Overlays Capture Architecture (OCA)

- Issuer: **Mapping Overlay**



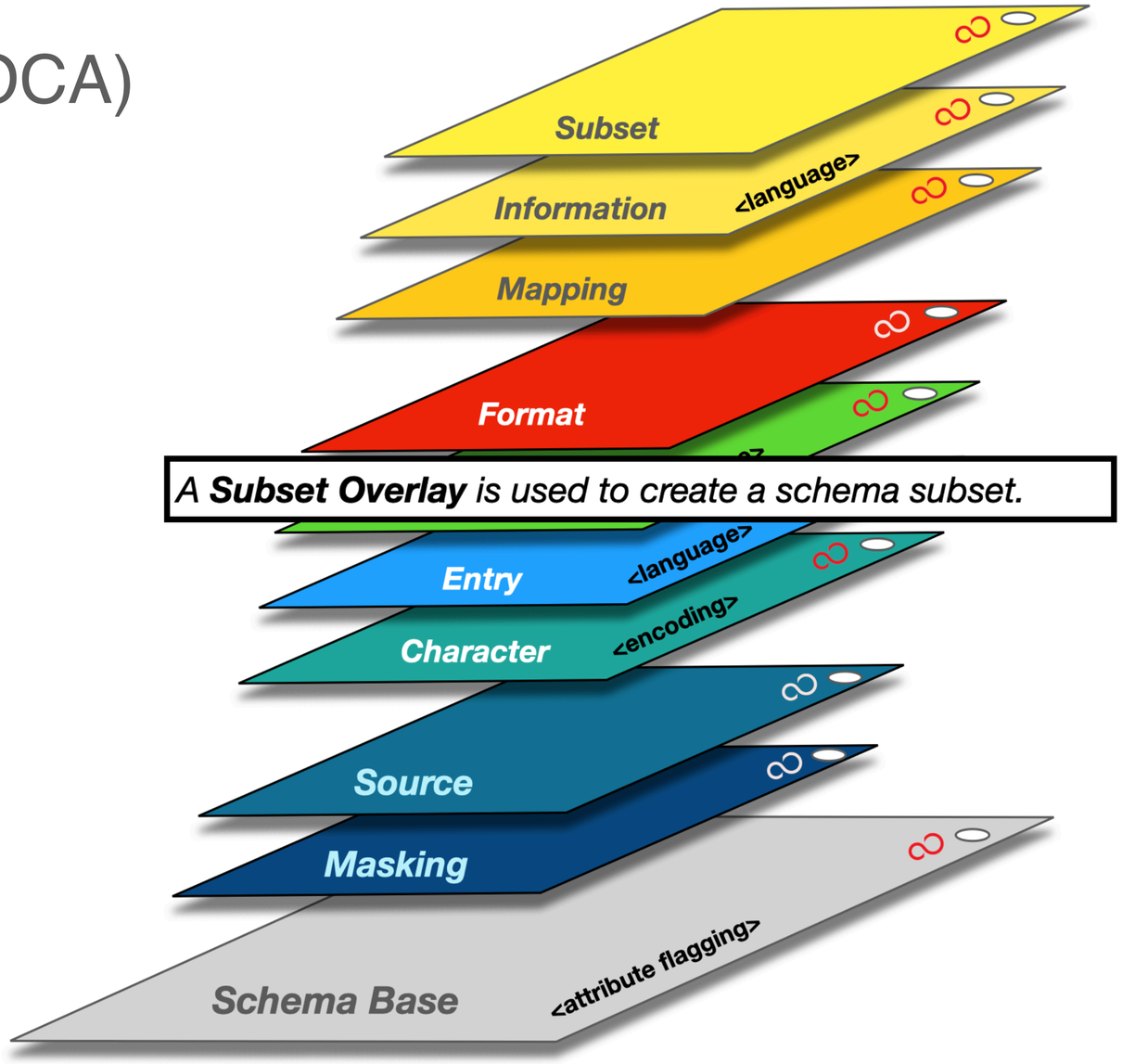
# Overlays Capture Architecture (OCA)

- Issuer: **Information Overlay**



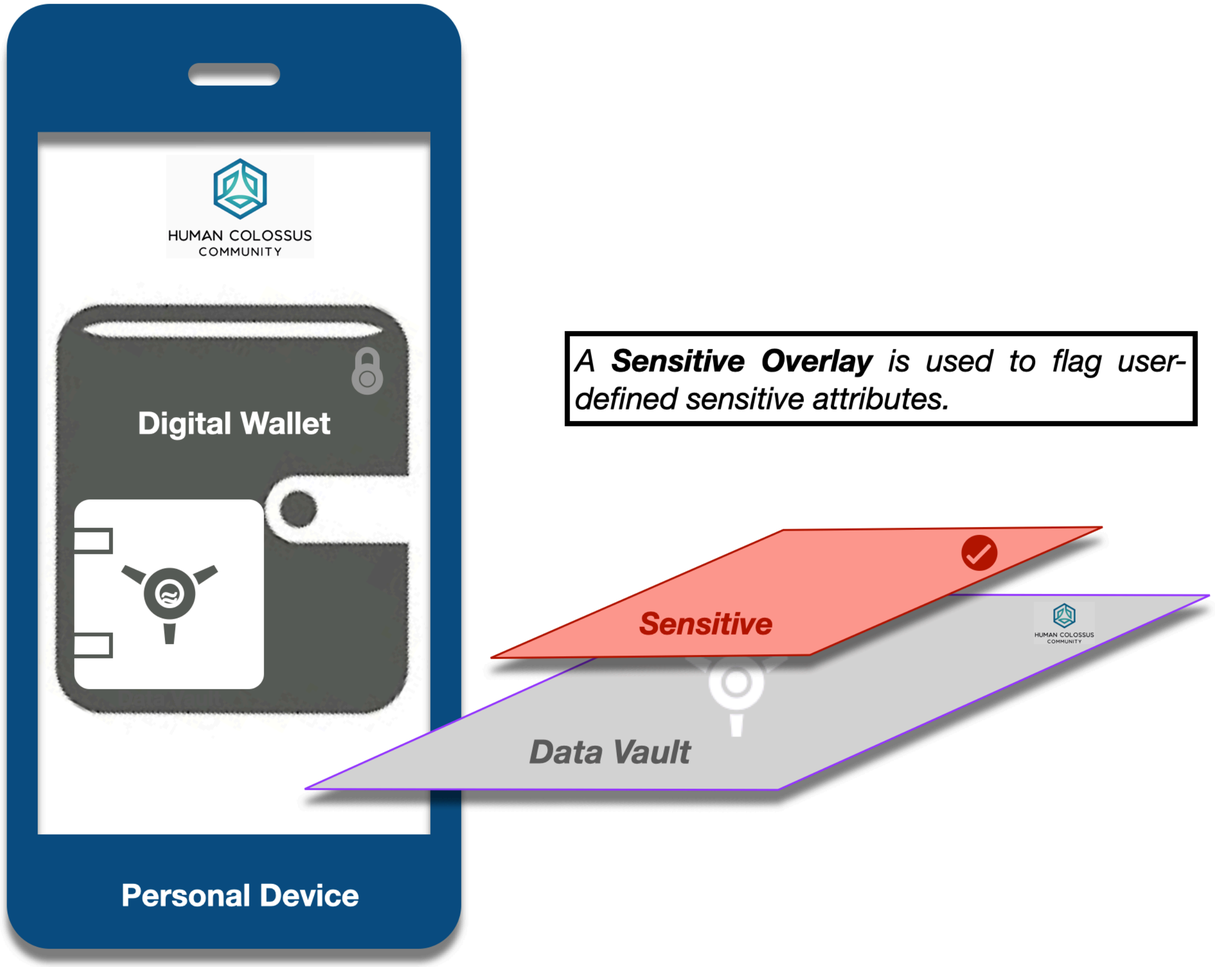
# Overlays Capture Architecture (OCA)

- Issuer: **Subset Overlay**



# Overlays Capture Architecture (OCA)

- Holder: **Sensitive Overlay**



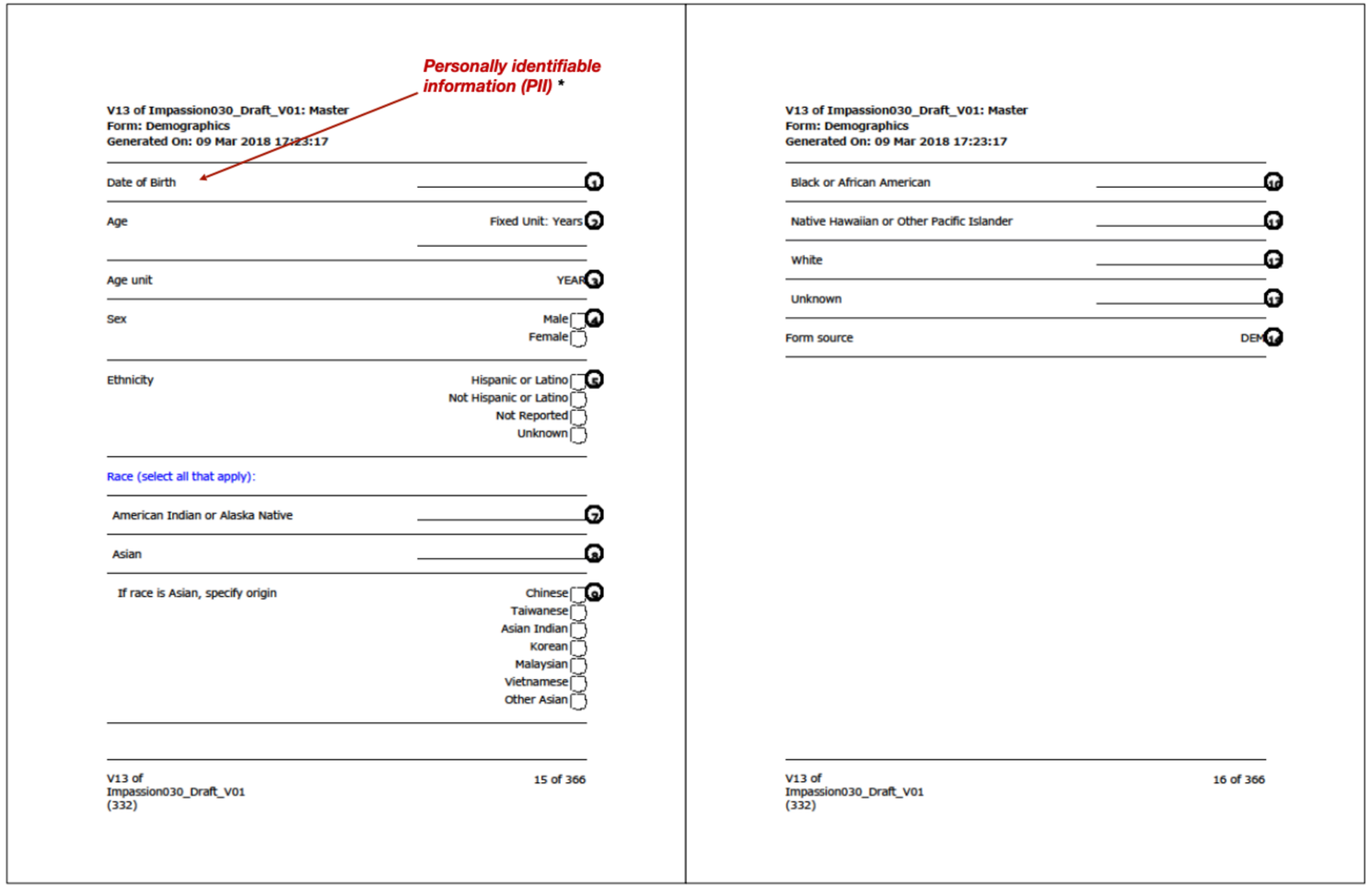
# Blinding Identity Taxonomy (BIT)

- 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.

# “Demographics” Form



*Personally identifiable information (PII)* *

V13 of Impassion030_Draft_V01: Master
Form: Demographics
Generated On: 09 Mar 2018 17:23:17

Date of Birth 1

Age Fixed Unit: Years 2

Age unit YEAR 3

Sex Male 4
Female

Ethnicity Hispanic or Latino 5
Not Hispanic or Latino
Not Reported
Unknown

Race (select all that apply):

American Indian or Alaska Native 7

Asian 8

If race is Asian, specify origin Chinese 9
Taiwanese
Asian Indian
Korean
Malaysian
Vietnamese
Other Asian

V13 of Impassion030_Draft_V01 (332) 15 of 366

V13 of Impassion030_Draft_V01: Master
Form: Demographics
Generated On: 09 Mar 2018 17:23:17

Black or African American 10

Native Hawaiian or Other Pacific Islander 11

White 12

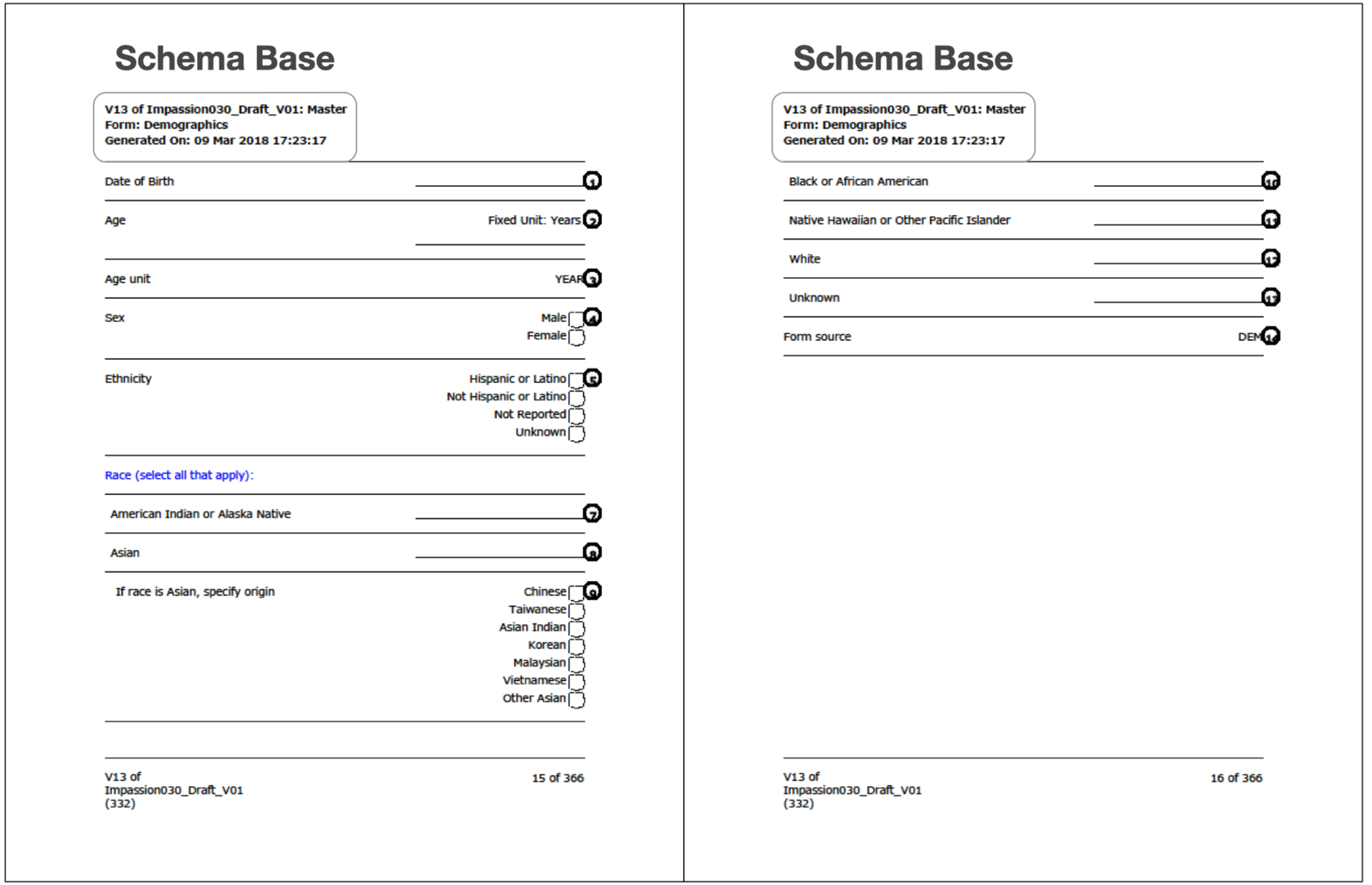
Unknown 13

Form source DEM 14

V13 of Impassion030_Draft_V01 (332) 16 of 366

* BIT element: *Dates and timestamps (e.g. Date of Birth, transaction dates, etc.)*

# “Demographics” Form



## Schema Base

**V13 of Impassion030_Draft_V01: Master**
**Form: Demographics**
**Generated On: 09 Mar 2018 17:23:17**

Date of Birth

Age — Fixed Unit: Years

Age unit — YEAR

Sex — Male / Female

Ethnicity — Hispanic or Latino / Not Hispanic or Latino / Not Reported / Unknown

Race (select all that apply):

American Indian or Alaska Native

Asian

If race is Asian, specify origin — Chinese / Taiwanese / Asian Indian / Korean / Malaysian / Vietnamese / Other Asian

V13 of Impassion030_Draft_V01 (332) — 15 of 366

## Schema Base

**V13 of Impassion030_Draft_V01: Master**
**Form: Demographics**
**Generated On: 09 Mar 2018 17:23:17**

Black or African American

Native Hawaiian or Other Pacific Islander

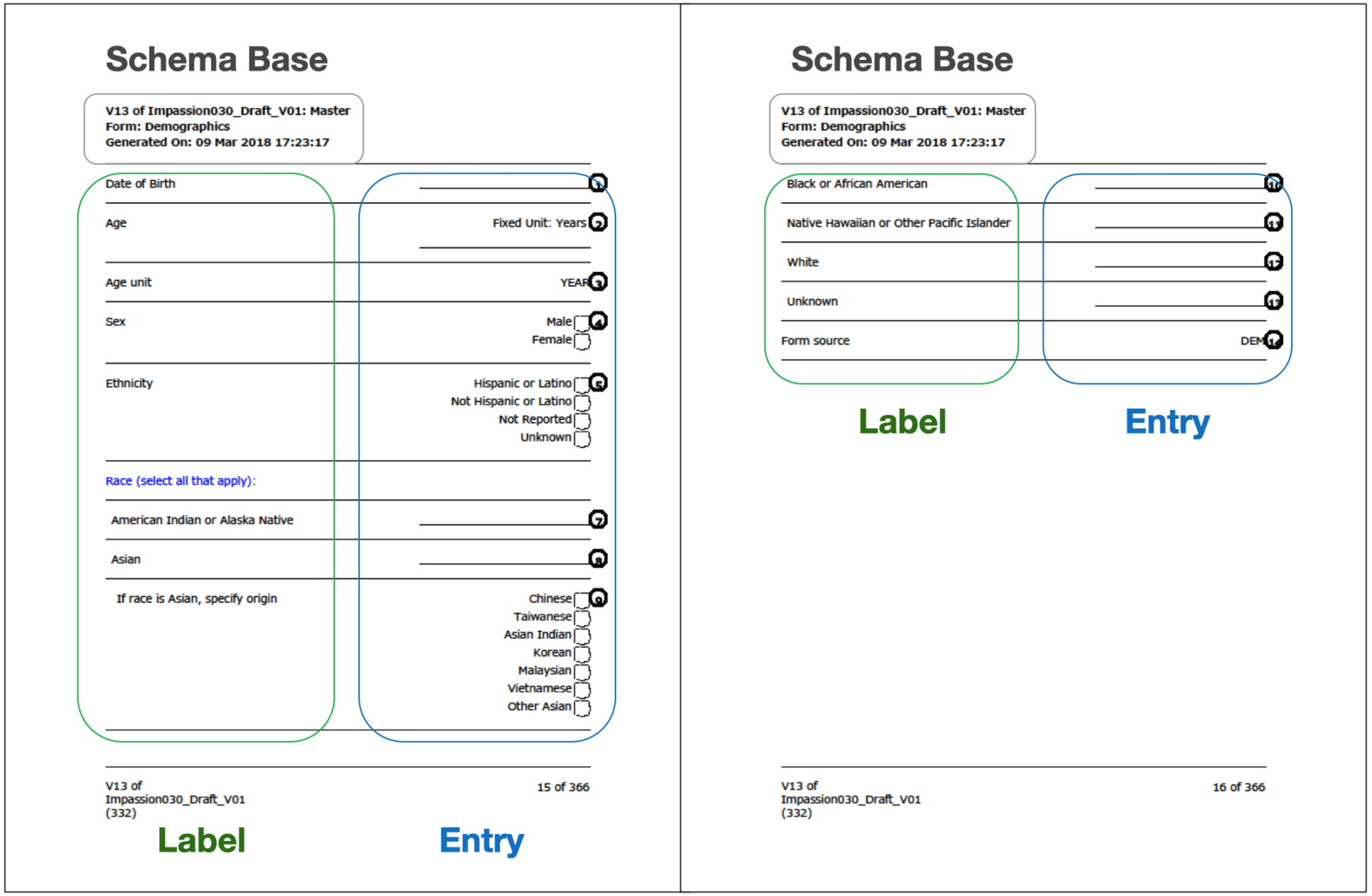
White

Unknown

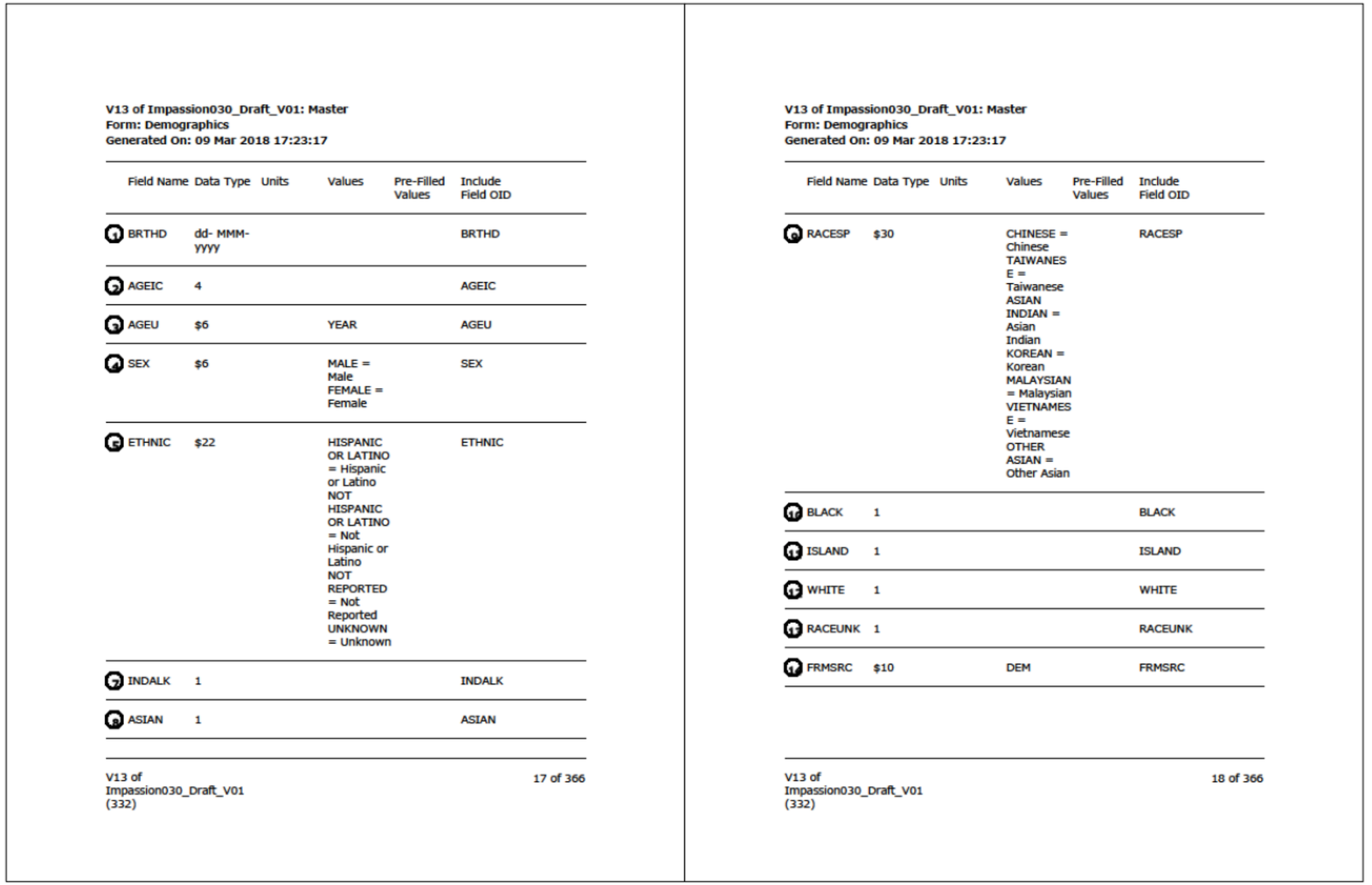
Form source — DEM

V13 of Impassion030_Draft_V01 (332) — 16 of 366

# “Demographics” Form



# “Demographics” Form - Field Definitions



V13 of Impassion030_Draft_V01: Master
Form: Demographics
Generated On: 09 Mar 2018 17:23:17

| Field Name | Data Type | Units | Values | Pre-Filled Values | Include Field OID |
|---|---|---|---|---|---|
| BRTHD | dd- MMM-yyyy | | | | BRTHD |
| AGEIC | 4 | | | | AGEIC |
| AGEU | $6 | | YEAR | | AGEU |
| SEX | $6 | | MALE = Male<br>FEMALE = Female | | SEX |
| ETHNIC | $22 | | HISPANIC OR LATINO = Hispanic or Latino<br>NOT HISPANIC OR LATINO = Not Hispanic or Latino<br>NOT REPORTED = Not Reported<br>UNKNOWN = Unknown | | ETHNIC |
| INDALK | 1 | | | | INDALK |
| ASIAN | 1 | | | | ASIAN |

V13 of Impassion030_Draft_V01 (332) 17 of 366

V13 of Impassion030_Draft_V01: Master
Form: Demographics
Generated On: 09 Mar 2018 17:23:17

| Field Name | Data Type | Units | Values | Pre-Filled Values | Include Field OID |
|---|---|---|---|---|---|
| RACESP | $30 | | CHINESE = Chinese<br>TAIWANESE = Taiwanese<br>ASIAN INDIAN = Asian Indian<br>KOREAN = Korean<br>MALAYSIAN = Malaysian<br>VIETNAMESE = Vietnamese<br>OTHER ASIAN = Other Asian | | RACESP |
| BLACK | 1 | | | | BLACK |
| ISLAND | 1 | | | | ISLAND |
| WHITE | 1 | | | | WHITE |
| RACEUNK | 1 | | | | RACEUNK |
| FRMSRC | $10 | | DEM | | FRMSRC |

V13 of Impassion030_Draft_V01 (332) 18 of 366

# “Demographics” Form - Field Definitions

## Schema Base

**V13 of Impassion030_Draft_V01: Master**
**Form: Demographics**
**Generated On: 09 Mar 2018 17:23:17**

| Field Name | Data Type | Units | Values | Pre-Filled Values | Include Field OID |
|---|---|---|---|---|---|
| BRTHD | dd- MMM-yyyy | | | | BRTHD |
| AGEIC | 4 | | | | AGEIC |
| AGEU | $6 | | YEAR | | AGEU |
| SEX | $6 | | MALE = Male<br>FEMALE = Female | | SEX |
| ETHNIC | $22 | | HISPANIC OR LATINO = Hispanic or Latino<br>NOT HISPANIC OR LATINO = Not Hispanic or Latino<br>NOT REPORTED = Not Reported<br>UNKNOWN = Unknown | | ETHNIC |
| INDALK | 1 | | | | INDALK |
| ASIAN | 1 | | | | ASIAN |

17 of 366

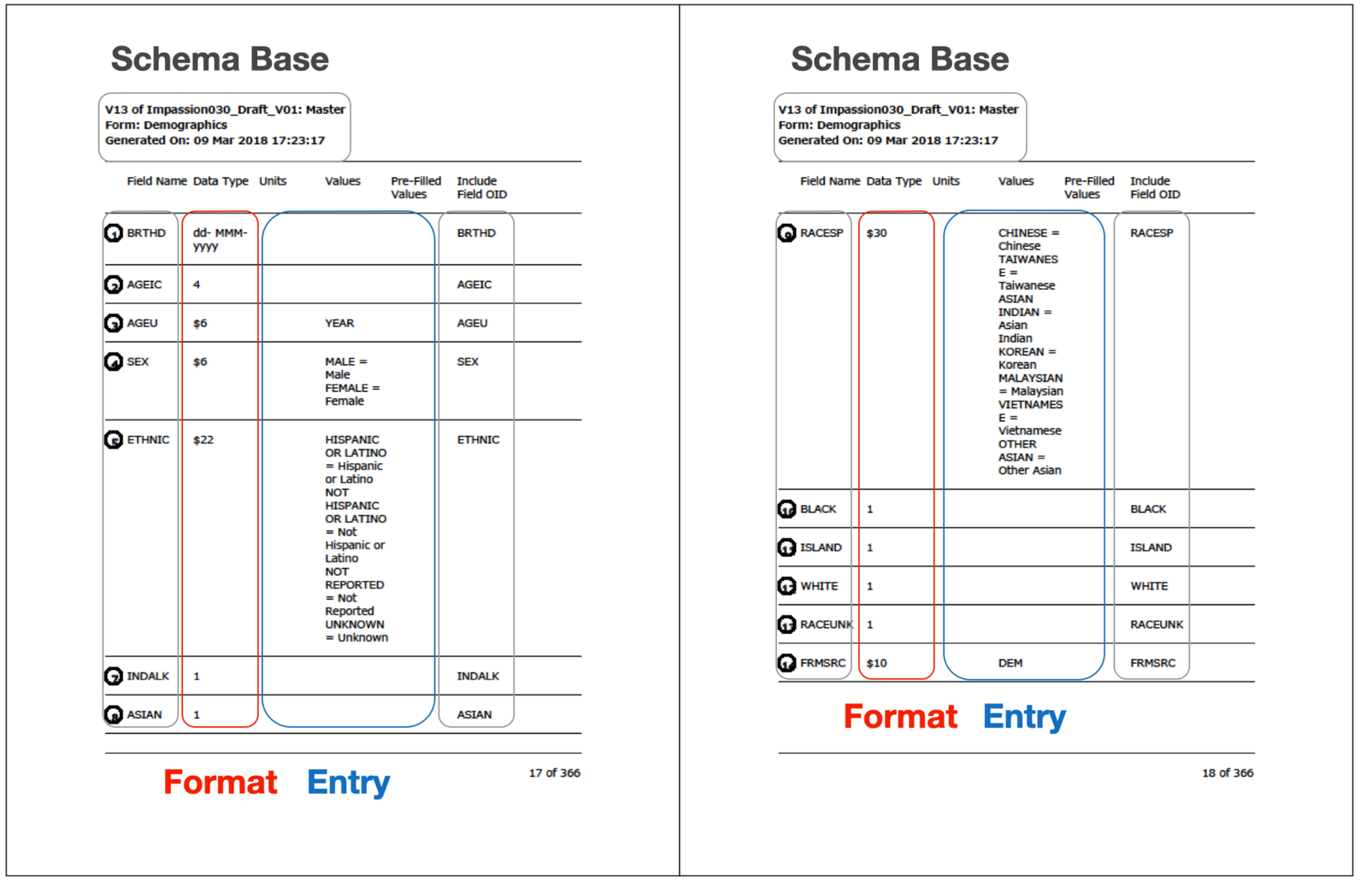
## Schema Base

**V13 of Impassion030_Draft_V01: Master**
**Form: Demographics**
**Generated On: 09 Mar 2018 17:23:17**

| Field Name | Data Type | Units | Values | Pre-Filled Values | Include Field OID |
|---|---|---|---|---|---|
| RACESP | $30 | | CHINESE = Chinese<br>TAIWANESE = Taiwanese<br>ASIAN INDIAN = Asian Indian<br>KOREAN = Korean<br>MALAYSIAN = Malaysian<br>VIETNAMESE = Vietnamese<br>OTHER ASIAN = Other Asian | | RACESP |
| BLACK | 1 | | | | BLACK |
| ISLAND | 1 | | | | ISLAND |
| WHITE | 1 | | | | WHITE |
| RACEUNK | 1 | | | | RACEUNK |
| FRMSRC | $10 | | DEM | | FRMSRC |

18 of 366

# “Demographics” Form - Field Definitions



## Schema Base

**V13 of Impassion030_Draft_V01: Master**
**Form: Demographics**
**Generated On: 09 Mar 2018 17:23:17**

| Field Name | Data Type | Units | Values | Pre-Filled Values | Include Field OID |
|---|---|---|---|---|---|
| BRTHD | dd- MMM-yyyy | | | | BRTHD |
| AGEIC | 4 | | | | AGEIC |
| AGEU | $6 | | YEAR | | AGEU |
| SEX | $6 | | MALE = Male<br>FEMALE = Female | | SEX |
| ETHNIC | $22 | | HISPANIC OR LATINO = Hispanic or Latino<br>NOT HISPANIC OR LATINO = Not Hispanic or Latino<br>NOT REPORTED = Not Reported<br>UNKNOWN = Unknown | | ETHNIC |
| INDALK | 1 | | | | INDALK |
| ASIAN | 1 | | | | ASIAN |

**Format** **Entry**

17 of 366

## Schema Base

**V13 of Impassion030_Draft_V01: Master**
**Form: Demographics**
**Generated On: 09 Mar 2018 17:23:17**

| Field Name | Data Type | Units | Values | Pre-Filled Values | Include Field OID |
|---|---|---|---|---|---|
| RACESP | $30 | | CHINESE = Chinese<br>TAIWANESE = Taiwanese<br>ASIAN INDIAN = Asian Indian<br>KOREAN = Korean<br>MALAYSIAN = Malaysian<br>VIETNAMESE = Vietnamese<br>OTHER ASIAN = Other Asian | | RACESP |
| BLACK | 1 | | | | BLACK |
| ISLAND | 1 | | | | ISLAND |
| WHITE | 1 | | | | WHITE |
| RACEUNK | 1 | | | | RACEUNK |
| FRMSRC | $10 | | DEM | | FRMSRC |

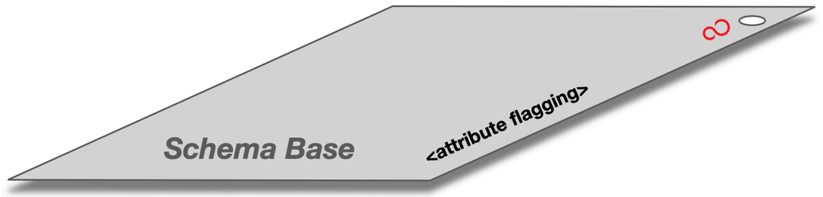
**Format** **Entry**

18 of 366

# Overlays Capture Architecture (OCA)

## - Issuer: **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 decentralise data. A **Schema Base** contains a blinding block which enables the issuer to flag attributes that could potentially unblind the identity of a governing entity.*





```
{
   "@context":"https://oca.tech/v1",
   "name":"Demographics",
   "type":"spec/schema_base/1.0",
   "description":"The Demographics domain includes a
   "classification":"GICS:35202010",
   "issued_by":"",
   "attributes":{
      "BRTHDTC":"Text",
      "AGE":"Number",
      "AGEU":"Text",
      "SEX":"Text",
      "RACE":"Text",
      "ETHNIC":"Text"
   },
   "blinding_attr":[
      "BRTHDTC"
   ]
}
```


Date of Birth

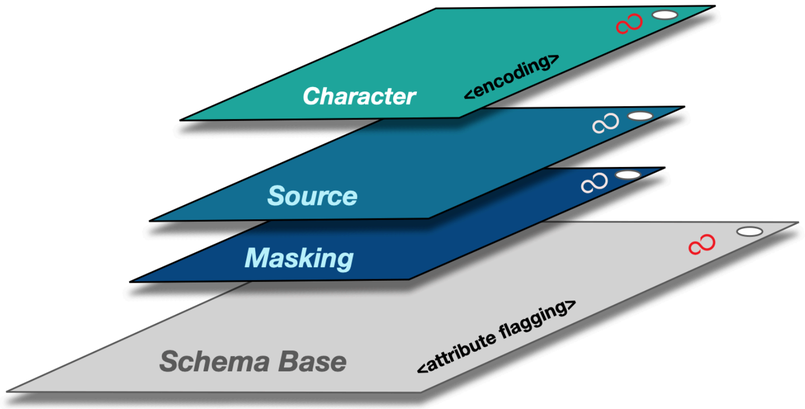
Date of Birth *

* BIT element: *Dates and timestamps (e.g. Date of Birth, transaction dates, etc.)*

# Overlays Capture Architecture (OCA)

## - Issuer: **Character Encoding Overlay**

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



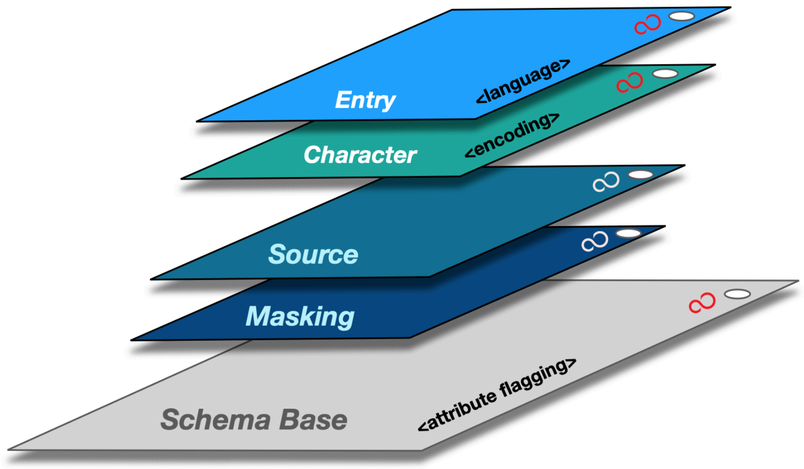


```
{
  "@context":"https://oca.tech/overlays/v1",
  "type":"spec/overlay/character_encoding/1.0",
  "issued_by":"",
  "role":"",
  "purpose":"",
  "schema_base":"hl:fqSVkf4H6JKuEgu6AUUkD3iYzcBofhKPiuNBTxoL9Dw1",
  "default_character_encoding":"utf-8",
  "attr_character_encoding":{
    "BRTHDTC":"utf-8",
    "AGE":"utf-8",
    "AGEU":"utf-8",
    "SEX":"utf-8",
    "RACE":"utf-8",
    "ETHNIC":"utf-8"
  }
}
```

# Overlays Capture Architecture (OCA)

## - Issuer: **Entry Overlay**

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



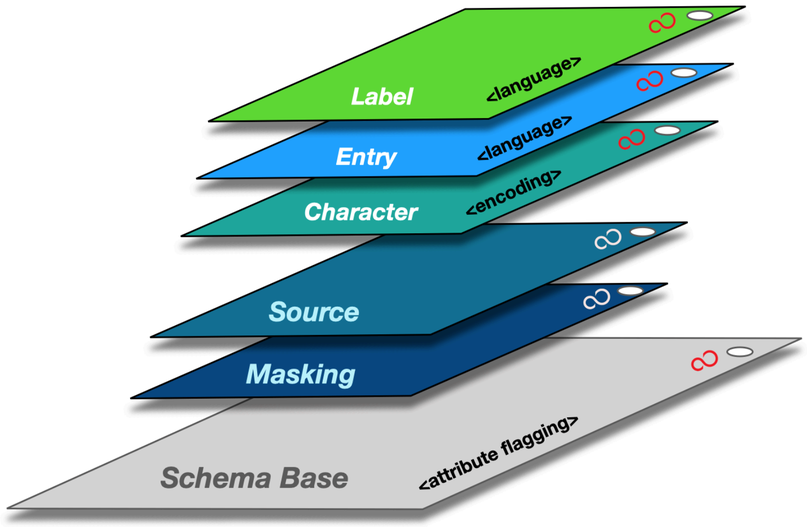


```
{
  "@context":"https://oca.tech/overlays/v1",
  "type":"spec/overlay/entry/1.0",
  "issued_by":"",
  "role":"",
  "purpose":"",
  "schema_base":"hl:fqSVkf4H6JKuEgu6AUUkD3iYzcBofhKPiuNBTxoL9Dw1",
  "language":"en_US",
  "attr_entries":{
    "AGEU":[
      "YEARS"
    ],
    "RACE":[
      "American Indian or Alaska Native",
      "Asian",
      "Black or African American",
      "Native Hawaiian or Other Pacific Islander",
      "White"
    ],
    "ETHNIC":[
      "Hispanic or Latino",
      "Not Hispanic or Latino"
    ]
  }
}
```

# Overlays Capture Architecture (OCA)

- Issuer: **Label Overlay**

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



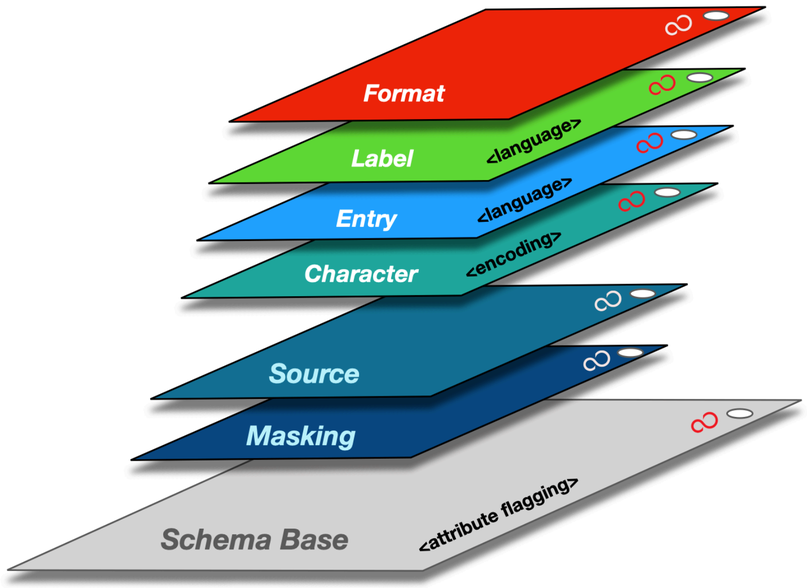


```
{
   "@context":"https://oca.tech/overlays/v1",
   "type":"spec/overlay/label/1.0",
   "issued_by":"",
   "role":"",
   "purpose":"",
   "schema_base":"hl:fqSVkf4H6JKuEgu6AUUkD3iYzcBofhKPiuNBTxoL9Dw1",
   "language":"en_US",
   "attr_labels":{
      "BRTHDTC":"Date/Time of Birth",
      "AGE":"Age",
      "AGEU":"Age Units",
      "SEX":"Sex",
      "RACE":"Race",
      "ETHNIC":"Ethnicity"
   },
   "attr_categories":[
      "_cat-1_"
   ],
   "cat_labels":{
      "_cat-1_":""
   },
   "cat_attributes":{
      "_cat-1_":[
         "BRTHDTC",
         "AGE",
         "AGEU",
         "SEX",
         "RACE",
         "ETHNIC"
      ]
   }
}
```

# Overlays Capture Architecture (OCA)

## - Issuer: **Format Overlay**

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

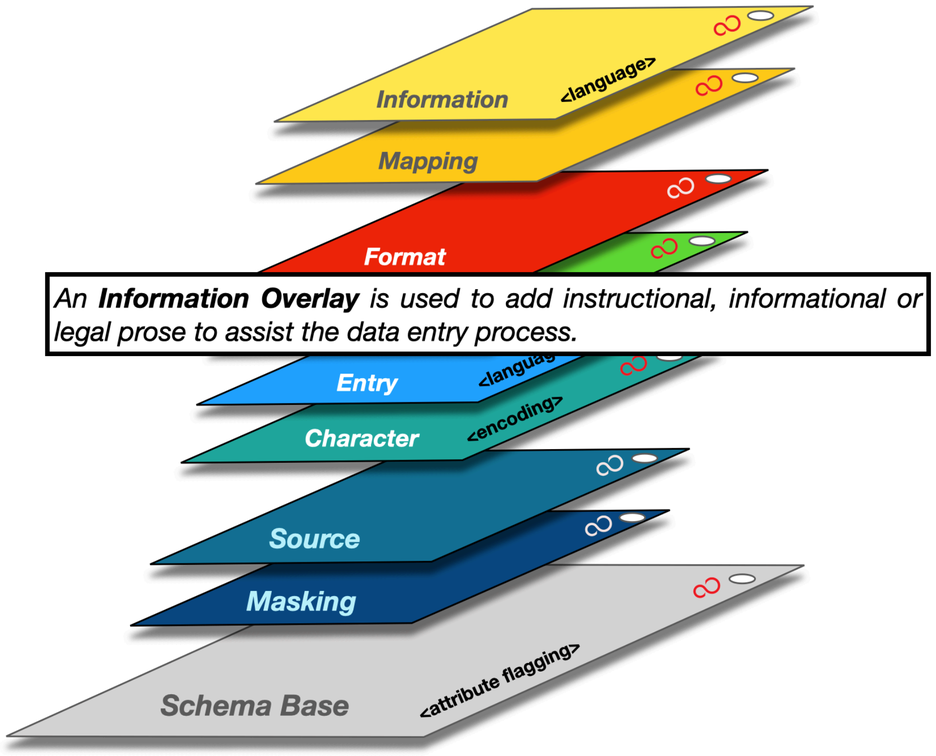




```
{
   "@context":"https://oca.tech/overlays/v1",
   "type":"spec/overlay/format/1.0",
   "issued_by":"",
   "role":"",
   "purpose":"",
   "schema_base":"hl:fqSVkf4H6JKuEgu6AUUkD3iYzcBofhKPiuNBTxoL9Dw1",
   "attr_formats":{
      "BRTHDTC":"ISO 8601",
      "AGEU":"CDISC/NCI: AGEU",
      "SEX":"CDISC/NCI: SEX",
      "RACE":"CDISC/NCI: RACE",
      "ETHNIC":"CDISC/NCI: ETHNIC"
   }
}
```

# Overlays Capture Architecture (OCA)

- Issuer: **Information Overlay**

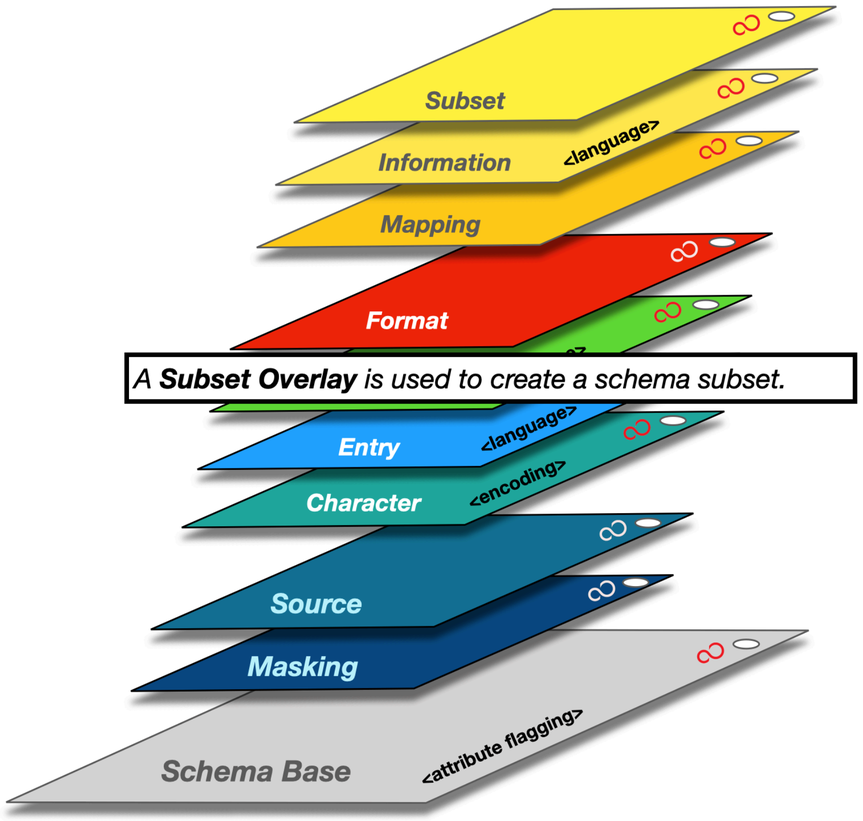




```
{
  "@context":"https://oca.tech/overlays/v1",
  "type":"spec/overlay/information/1.0",
  "issued_by":"",
  "role":"",
  "purpose":"",
  "schema_base":"hl:fqSVkf4H6JKuEgu6AUUkD3iYzcBofhKPiuNBTxoL9Dw1",
  "language":"en_US",
  "attr_information":{
    "BRTHDTC":"Date/time of birth of the subject.",
    "AGE":"Age expressed in AGEU. May be derived from RFSTDTC and BRTHDTC
    "AGEU":"Units associated with AGE.",
    "SEX":"Sex of the subject.",
    "RACE":"Race of the subject. Sponsors should refer to “Collection of
    "ETHNIC":"The ethnicity of the subject. Sponsors should refer to “Col
  }
}
```

# Overlays Capture Architecture (OCA)

- Issuer: **Subset Overlay**





```
{
   "@context":"https://oca.tech/overlays/v1",
   "type":"spec/overlay/subset/1.0",
   "issued_by":"",
   "schema_base":"hl:fqSVkf4H6JKuEgu6AUUkD3iYzcBofhKPiuNBTxoL9Dw1",
   "attributes":[
      "BRTHDTC",
      "AGE",
      "AGEU",
      "SEX"
   ]
}
```

# Industry Sector Categorisation

Option

GICS: Global Industry Classification Standard

*“The Industry Standard”*

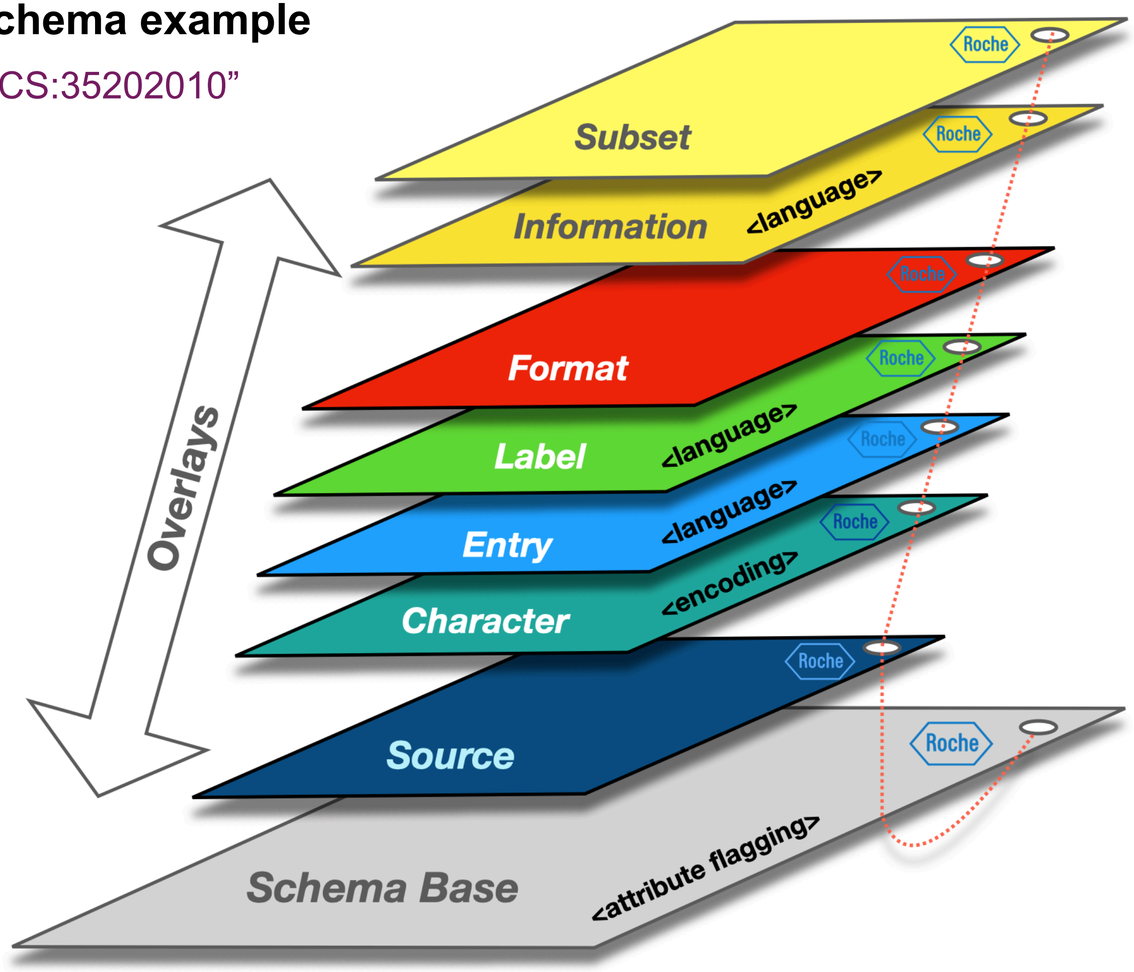

The **GICS** indices is an industry taxonomy for use by the global financial community as a basis to assign companies to a sub-industry, and to an industry, industry group, and sector, by its principal business activity.

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

# Object interoperability within an Industry Sector

**“Demographics” schema example**

“classification”: “GICS:35202010”




GICS:35202010

Sector code:
35 - Health Care

Industry group code:
3520 - Pharmaceuticals, Biotechnology & Life Sciences

Industry code:
352020 - Pharmaceuticals

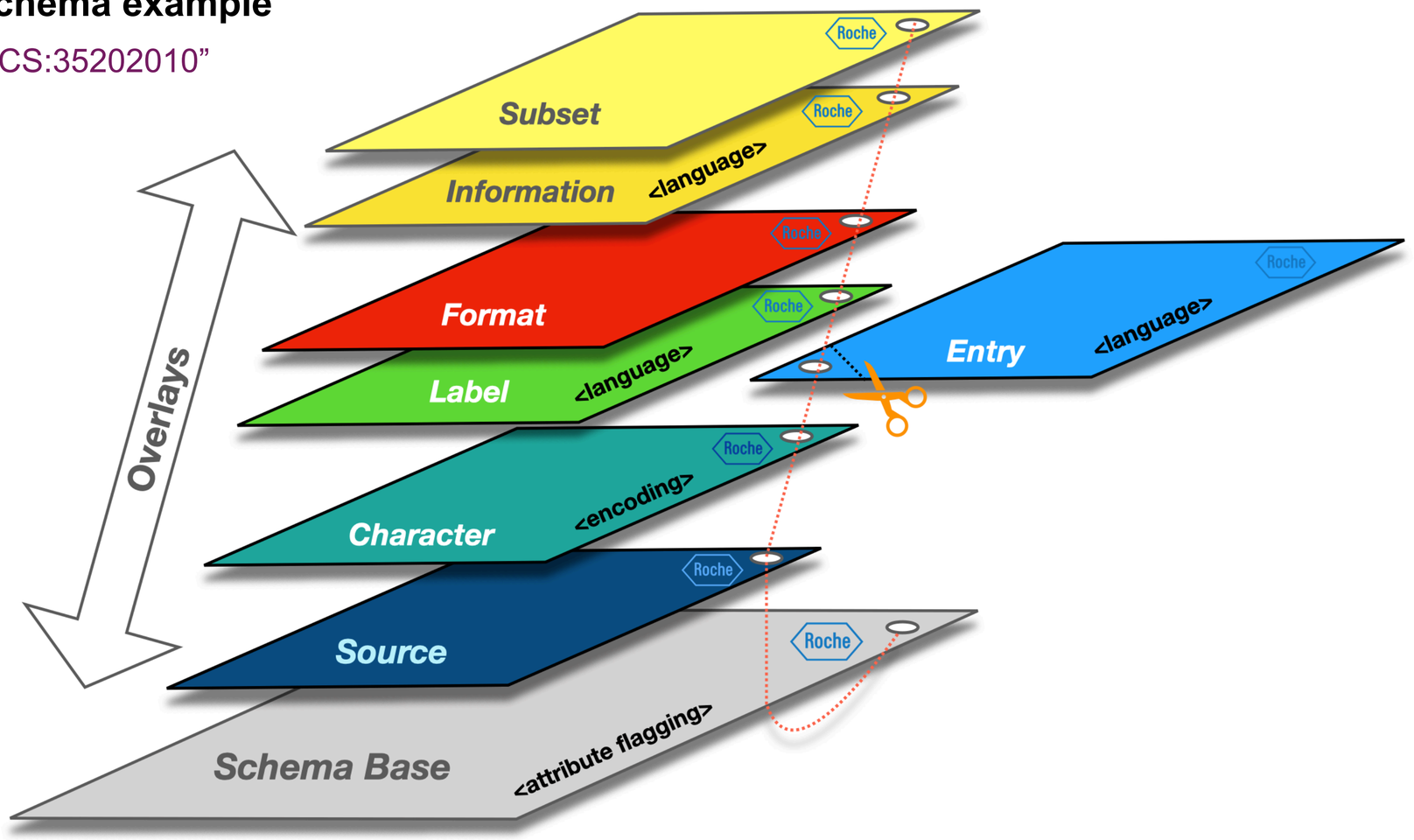
Sub-industry code:
35202010 - Pharmaceuticals

Description:
Companies engaged in the research, development or production of pharmaceuticals. Includes veterinary drugs.

# Object interoperability within an Industry Sector

**“Demographics” schema example**

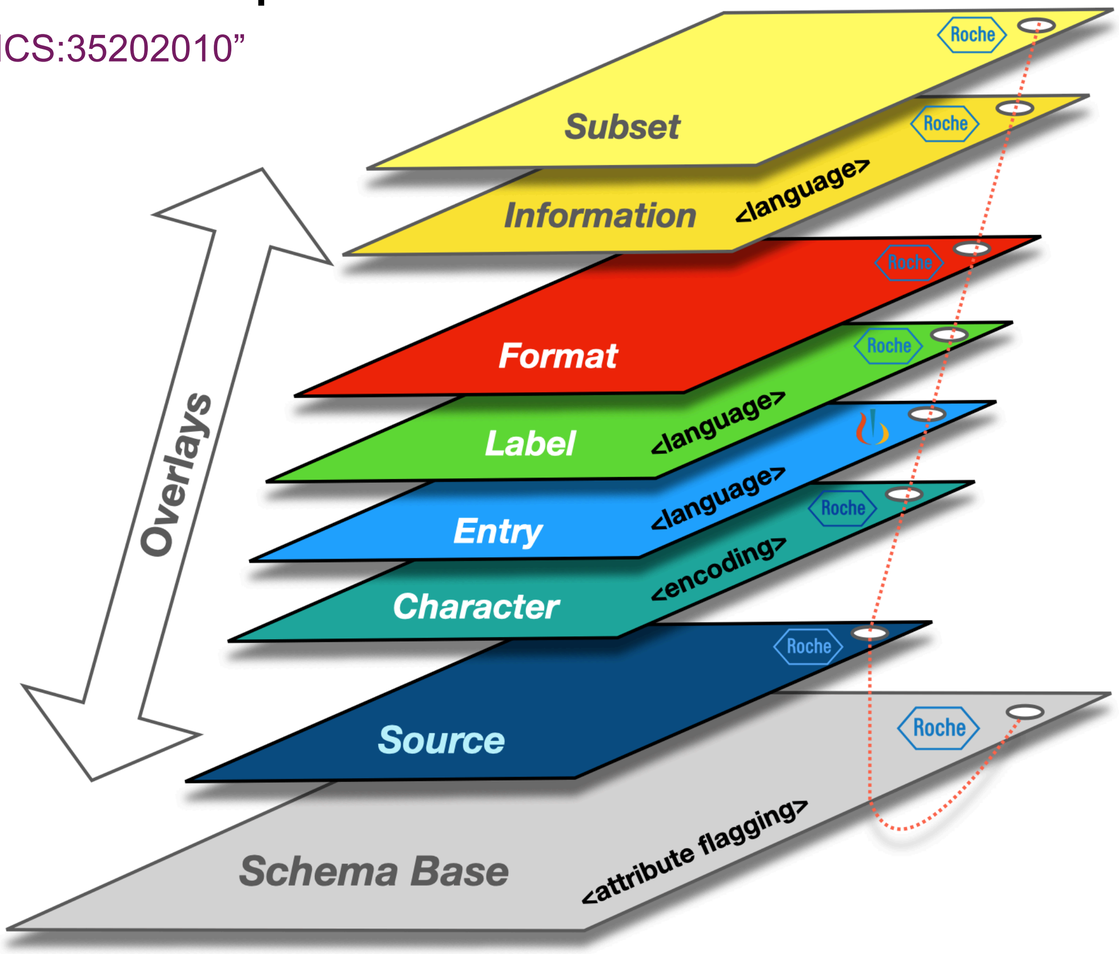
“classification”: “GICS:35202010”



# Object interoperability within an Industry Sector

**“Demographics” schema example**

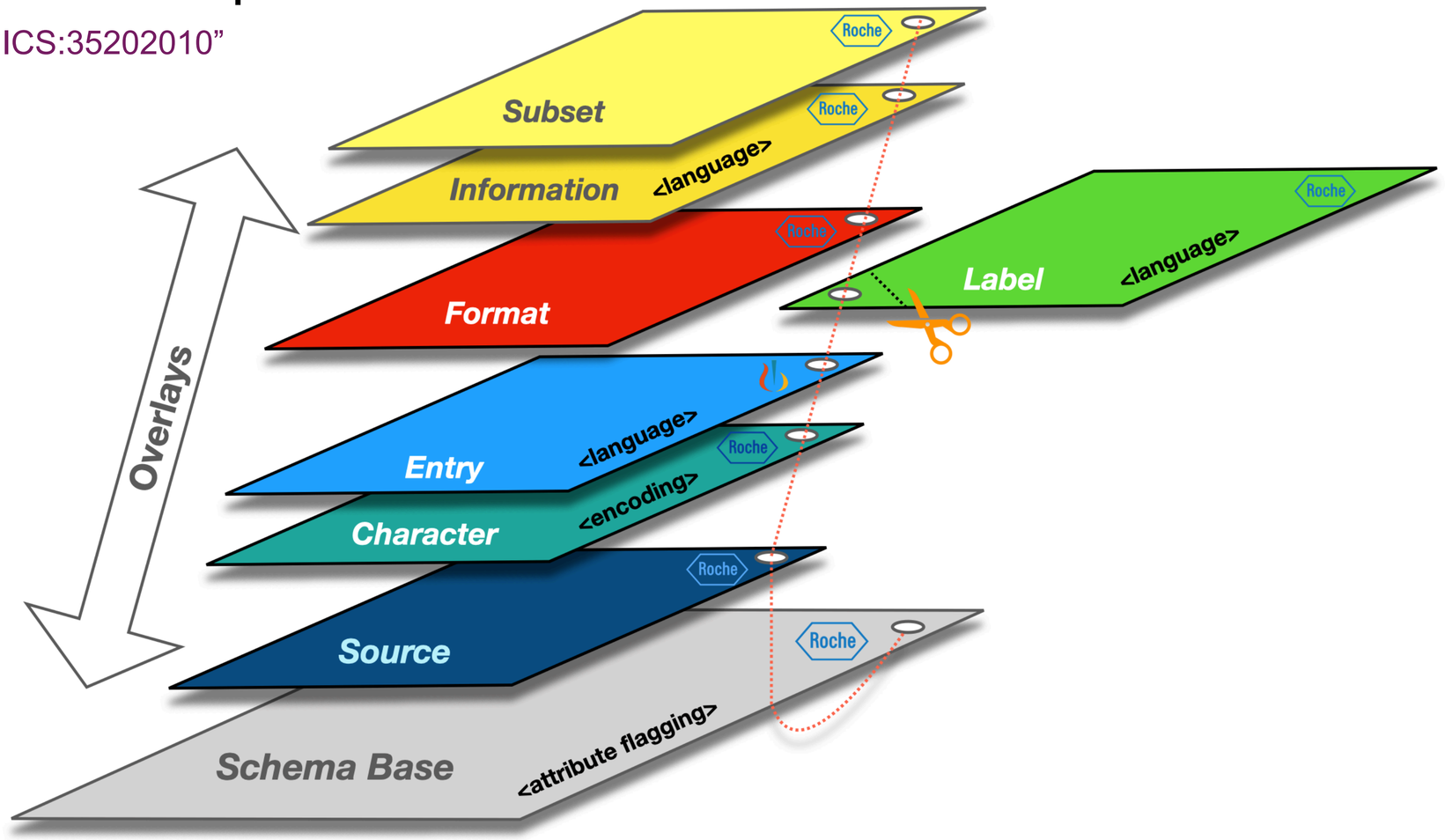
“classification”: “GICS:35202010”



# Object interoperability within an Industry Sector

**“Demographics” schema example**

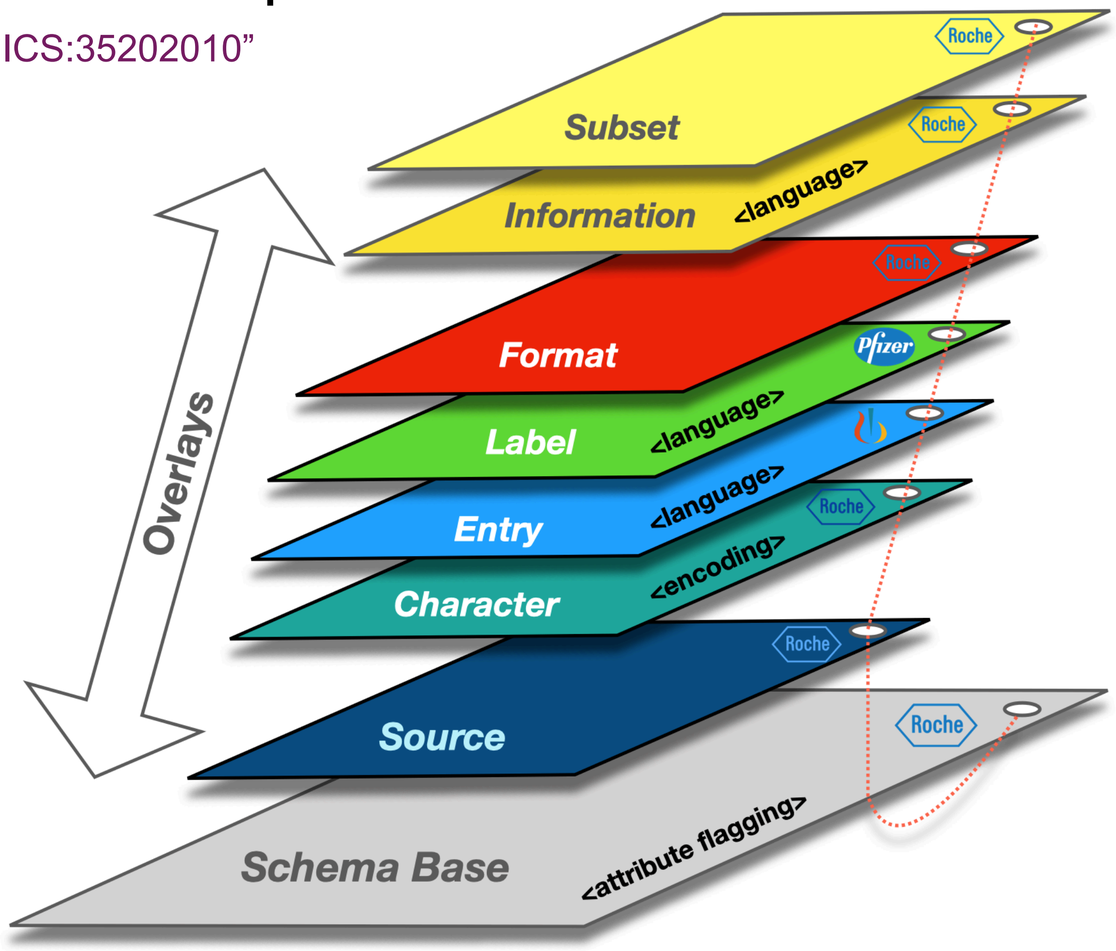
“classification”: “GICS:35202010”



# Object interoperability within an Industry Sector

**“Demographics” schema example**

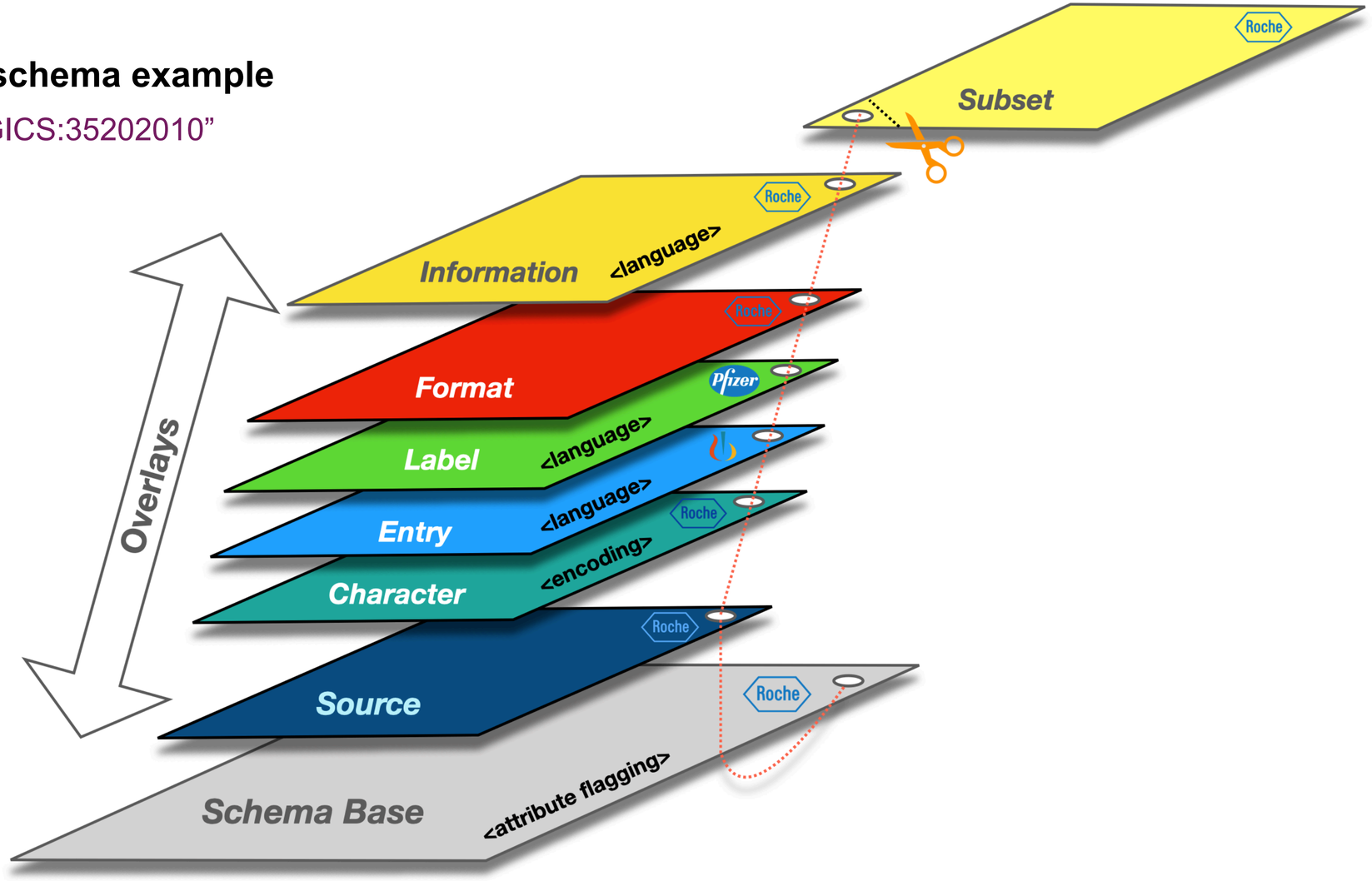
“classification”: “GICS:35202010”



# Object interoperability within an Industry Sector

**“Demographics” schema example**

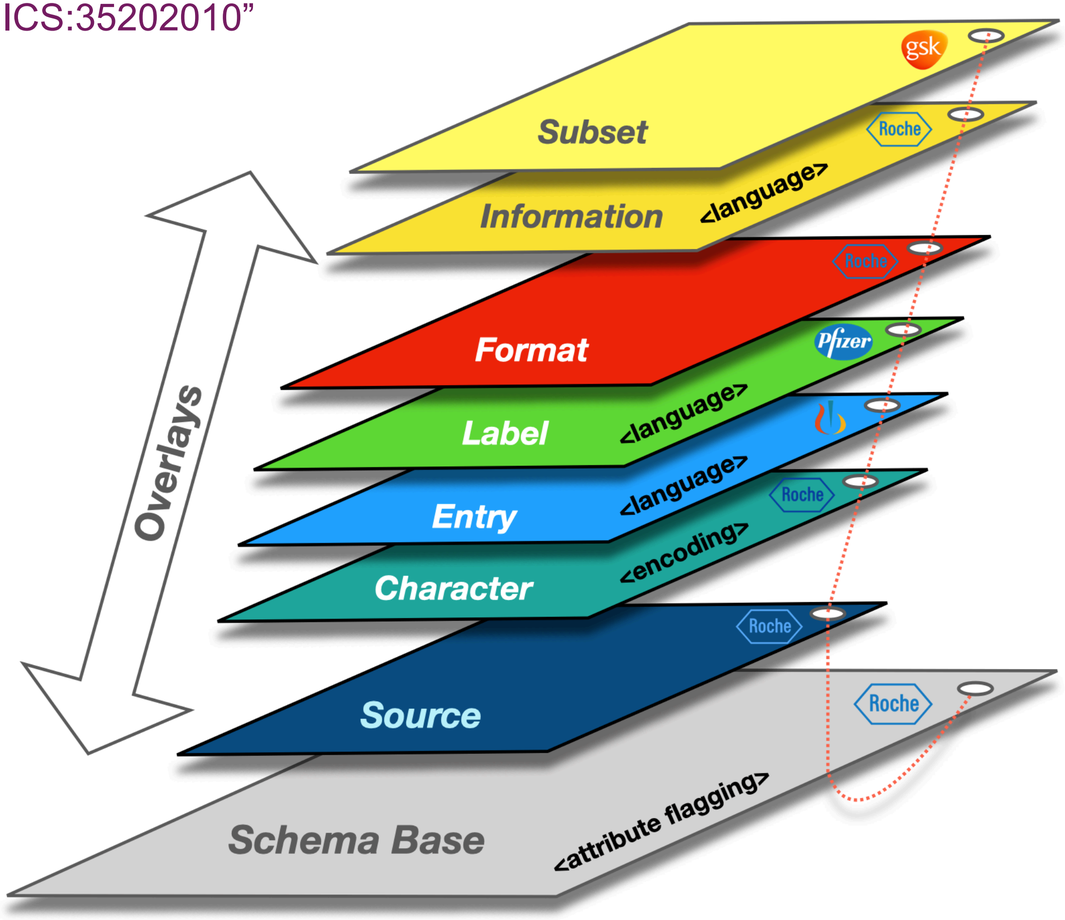
“classification”: “GICS:35202010”



# Object interoperability within an Industry Sector

**“Demographics” schema example**

“classification”: “GICS:35202010”



HUMAN COLOSSUS
FOUNDATION


- Paul Knowles paul.knowles@humancolossus.org