6.1 Recommendation #4: Standard Data Models and Elements

6.1.1 Introduction to this Interoperability Challenge

COVID-19 health data comes from multiple sources and in different formats, based on who, where, and how that data is collected in a given jurisdiction. But as with passports and payment cards, fully interoperable digital health pass systems need to support a common data model for data exchange, including specifying standard attributes and forms that can be supported by specific use case implementations.

The Good Health Pass Data Models & Elements recommendations are grounded in the work put forward by several cross-industry efforts, including:

- COVID-19 Credentials Initiative (CCI) [Linux Foundation Public Health (LFPH)], which has established a task force to develop form and schema specifications, Codes of Practice, technical guidelines, JavaScript Object Notation (JSON) code, and related components. This CCI work includes full reference to the eHealth Network's guidelines for COVID-19 certificates.
- W3C Credentials Community Group (CCG)'s Vaccination Certificate Vocabulary
- Smart Vaccine Certificate Working Group at World Health Organization (WHO)
- Canadian COVID Credentials Consortium (C4)
- International Air Transport Association (IATA): IATA Travel Pass Initiative
- International Civil Aviation Organization (ICAO)
- FHIR Focus Group at Trust over IP Foundation (ToIP)
- Vaccination Credential Initiative (VCI): SMART Health Cards Framework
- Centers for Disease Control and Prevention (CDC): COVID-19 Vaccination Reporting Specification (CVRS)
- UK's National Health Service (NHS): Testing for coronavirus (COVID-19) / Coronavirus COVID-19 vaccine
- ISO/IEC 18013-5 mdoc for eHealth

The Good Health Pass Collaborative (GHPC) would like to applaud the global efforts of all of these initiatives – each of which reflects vast participation and collaboration from public health authorities, private health vendors and consortiums, electronic health record (EHR) providers, pharma companies, and major pharmacy chains (US).

Through an extensive evaluation of the above cross-industry initiatives, this group worked to develop a set of data element recommendations that can accommodate requirements of different countries and jurisdictions around the world.

6.1.2 Objective of this Drafting Group

Given the global nature of health pass credentials and passes, the Good Health Pass ecosystem **MUST** be technologically agnostic to all data ingestions. As such, this document focuses recommendations predominantly on data elements, semantic harmonization, and common models for data exchange that can help us work toward interoperability without putting an undue burden on existing health systems and workflows.

6.1.3 Problem #1: Data Elements – Certificates, Credentials, and Passes

6.1.3.1 Problem Description

In healthcare contexts, data is held in a number of places, such as laboratories, health information systems, and clinical trial registries. In order for an issuer to create a credential or pass, based on the requirements of the activity they're undertaking, these institutions must make a version of that data available to that person in the form of a certificate or credential.

When creating standards for a globally-useable health pass, agreement on a minimum-necessary set of data, including how the data is named, organized, etc. is critical to ensuring interoperability. This is of particular importance when considering that the sources of the data (e.g., the aforementioned institutions) may collect or code data differently.

In general, a Good Health Pass (GHP) should only include the minimal set of information necessary to verify and confirm the holder's vaccination, testing, or recovery status as required for a verifier to make a particular trust decision. This is reinforced by Figure 7 from the Good Health Pass Guide to Key Concepts and Terminology:

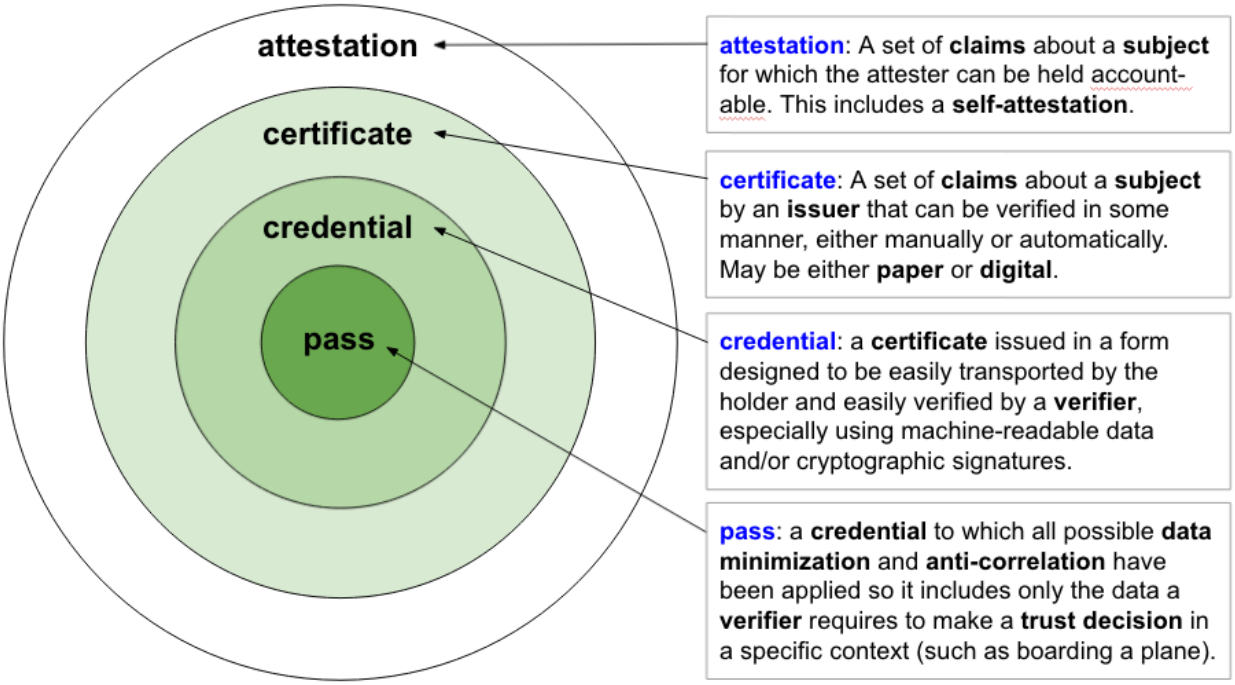

Figure 7: The four core terms for describing data containers for health data used for travel

6.1.3.2 Good Health Pass Design Requirements & Considerations

Within the Good Health Pass ecosystem, a health credential attests that a COVID-19 test event or vaccination event occurred and a health pass attests to the minimal set of data from one or more health records required for a specific verifier or class of verifiers to make a particular trust decision.

In February 2021, the COVID-19 Credentials Initiative (CCI) Schema Task Force began comparing data set recommendations from the cross-industry efforts listed above (see cross-reference links in Appendix

B for more details). This landscape began with the CDC's Technical Standards and Reporting Data – comprehensive documentation that contained recommendations beyond the minimum data requirements for credentials and passes.

When the World Health Organization published its core data set for the Smart Vaccine Certificate in March 2021, and when the EU's eHealth Network published its Value Sets for Digital COVID Certificates in April 2021, these were cross-referenced with the CDC standards to determine common attributes across the three entities, creating the foundation for the recommendations below.

6.1.3.3 Recommendations

6.1.3.3.1 Overall Recommendations

Based on this cross-review process, the GHPC RECOMMENDS that solution providers include the following minimum viable value sets for COVID-19 credentials for global interoperability. Because the GHPC focuses on credentials and passes for both vaccination and testing, these recommendations most closely resemble the EU specification. (The EU specification was also informed by WHO's core dataset, which focuses exclusively on vaccination).

Note that GHPC also recommends that CVX code and state/province of vaccination **SHOULD** be included as additional data fields for North America only (these two fields are not currently in EU recommendations).

6.1.3.3.2 Health Certificates

Verifiable Vaccination Certificates

These guidelines aim at preparing for interoperability between attestations of vaccination for medical purposes (also known as vaccination certificates). Other purposes of non-discriminatory use (e.g., in particular for travel purposes) **MAY** be decided upon by jurisdictional authorities, with attestations of vaccination reserved for ongoing global legal, ethical, scientific and societal discussions.

COVID-19 Antigen Test Certificate

A common list of COVID-19 rapid antigen tests, including those of which their test results are mutually recognised, and a common standardized set of data to be included in COVID-19 test result certificates.

6.1.3.3.3 Health Credentials

Vaccination Credential

Vaccination credentials are to be used primarily as a standardized and interoperable form of proof of vaccination for medical purposes. For other use cases, such as for the purpose of travel, one could consider situations where a person arrives in a country and a verifying authority confirms whether the person has been vaccinated against an infectious disease.

The following data fields **SHOULD** be included in the vaccination credential:

(a) name: surname(s) and forename(s), in that order;
(b) date of birth;
(c) disease or agent targeted;
(d) vaccine/prophylaxis;
(e) vaccine medicinal product;

(f) CVX code* *(North America only; not currently in EU recommendations)*;
(g) vaccine marketing authorization holder or manufacturer;
(h) number in a series of vaccinations/doses;
(i) date of vaccination, indicating the date of the latest dose received;
(j) state/province of vaccination* *(North America only; not currently in EU recommendations)*;
(k) country of vaccination;
(l) certificate issuer;
(m) a unique certificate identifier.

Test Credential

Robust testing strategies have been – and will continue to be – an essential aspect of preparedness and response to the COVID-19 pandemic, allowing for early detection of potentially infectious individuals and insight on infection rates and transmission within communities. They are also a prerequisite to adequate contact tracing that can help limit the spread through prompt isolation.

The unequal rollout of COVID-19 vaccines underscores the importance of continued, widespread COVID-19 testing as an essential public health tool – one that must continue alongside vaccination to ensure an equitable return to public life. Additionally, with increasing concerns over the circulation of SARS-CoV-2 variants, testing – and variant-specific testing – will continue to be vital for controlling and suppressing further spread of the virus.

The following data fields **SHOULD** be included in the COVID-19 antigen test credential:

(a) name: surname(s) and forename(s), in that order;
(b) date of birth;
(c) disease or agent targeted;
(d) the type of test;
(e) test name *(optional for NAAT test)*;
(f) test manufacturer *(optional for NAAT test)*;
(g) date and time of the test sample collection;
(h) date and time of the test result production *(optional for rapid antigen test)*;
(i) result of the test;
(j) testing centre or facility;
(n) state/province of test* *(North America only; not currently in EU recommendations)*;
(k) country of test;
(l) certificate issuer;
(m) a unique certificate identifier.

Recovery Credential

According to current evidence, although still testing positive for SARS-CoV-2, infected individuals who have recovered from COVID-19 may still be infectious. In those particular cases, negating the virus may not be viable, with a limited risk of transmission to others.

However, by not presenting a negative test result for unrestricted movement, jurisdictional authorities would prevent those individuals from crossing borders. On balance, the evidence suggests that those who have recovered from COVID-19 have a reduced risk of infection.

The following data fields **SHOULD** be included in the COVID-19 citizen recovery credential:

(a) name: surname(s) and forename(s), in that order;
(b) date of birth;
(c) disease or agent the citizen has recovered from;
(d) date of first positive test result;
(e) state/province of test* *(North America only; not currently in EU recommendations)*;
(f) country of test;
(g) certificate issuer;
(h) certificate valid from;
(i) certificate valid until *(not more than 180 days after the date of first positive test result)*;
(j) a unique certificate identifier.

6.1.3.3.4 Good Health Passes

While the above health credentials attest that a COVID-19 test event or vaccination event occurred, one could consider situations where a person may be required to provide proof of COVID-19 status and may not want – or should not be required to – share all of the information that is captured in a credential.

In these scenarios, such as for international travel, a health pass **MAY** be used to attest to the minimal set of data from one or more health records required for a specific verifier or class of verifiers to make a particular trust decision.

A Good Health Pass **SHOULD** only include the minimal set of information necessary to verify and confirm the holder's vaccination, testing, or recovery status as required for a verifier to make a particular trust decision. For travel, this **SHOULD** include the following basic elements:

(a) name: surname(s) and forename(s), in that order;
(b) date of birth;
(c) pass status

All defined semantic content **MUST** be uniquely identified by cryptographic hash functions to ensure that the same message always results in the same hash.

6.1.4 Problem #2: Data Exchange – Common Data Model

6.1.4.1 Problem Description

Common data models are used to standardize and facilitate the exchange, sharing, or storing of data from multiple sources. They are frequently used in healthcare, where there is a need to share data from disparate sources for a particular use, such as clinical research. Common data models can also help avoid the need to share patient-level data – a privacy preserving feature with great relevance for the Good Health Pass ecosystem.

Common data models are designed to promote interoperability between systems that encode healthcare data in different ways. At the same time, there are a multitude of common data models designed for specific contexts – necessitating thoughtful choices based on the proposed use case.

6.1.4.2 Recommendations

The EU, World Health Organization (WHO), and Vaccine Credentialing Initiative (VCI) have all recommended FHIR as the common data model of choice for COVID-19 credentialing initiatives. To further promote interoperability, the GHPC also RECOMMENDS that health credentials and passes

SHOULD use the **HL7 FHIR data model** as described below. In doing so, credential and pass developers can guarantee that data elements included will represent the necessary data requirements covering specific jurisdictions and usage types.

Figure 8: HL7 FHIR Logo

Fast Healthcare Interoperability Resources (FHIR) is a standard for health care data exchange that ultimately aims to get legacy EHRs (electronic **health records**) to a point where they can be interoperable. FHIR allows EHRs the versatility to be used in mobile devices, web-based applications, cloud communications, and EHR data-sharing using modular components. This makes it easier for third-party developers to integrate medical applications into existing systems.

While the FHIR data model aligns Good Health Pass implementers with the data format selected by the EU/WHO/VCI initiatives, it **SHOULD** be recognized that FHIR's adoption is still uneven and there are many healthcare systems that have not, or could not adopt this data model. To address this, the GHPC RECOMMENDS that solution providers **SHOULD** provide guidelines and/or tools to convert non-FHIR formatted EHR data into FHIR formatted records as a pre-processing step. Example guidance on importing csv data into FHIR can be found here.

This RECOMMENDATION assumes that the candidate source EHR data meets the minimum data field requirements called for in the Health Credentials: Vaccination Credential, Test Credential, Recovery Credential sections above.

6.1.5 Problem #3: Semantic Harmonization

6.1.5.1 Problem Description

In exploring the minimal datasets required to differentiate between certificates, credentials, and passes, it is clear that COVID-19 related data will often come from disparate, siloed sources.

When data for credentials and passes comes from multiple sources, it can be difficult to facilitate efficient processing during verification, use an automated rules engine and/or decision support system for pass issuance (based on jurisdiction requirements), and support multiple languages. All of these are key considerations for any proposed solution that would be used to support international travel. Further, dictionary coding differs across jurisdictions, underscoring the importance of having flexible overlays.

As such, data and semantic harmonization are critical to ensuring interoperability of health credentials and passes. Data harmonization involves transforming datasets to fit together in a common architecture. Semantic harmonization is the process of ensuring that – as part of data harmonization – the meaning and context of data remains uniformly understood by all interacting actors, regardless of how the data was collected originally. This also means that transient objects such as credentials and passes can be more easily resolved into multiple languages.

6.1.5.2 Good Health Pass Design Requirements & Considerations

Advances in decentralized data modeling can enable data harmonization across different models and representation formats. The introduction of decentralized technologies in identity, semantics, and governance has the added benefit of providing a safeguard against online exploitation, surveillance, or potential abuse by redistributing digital control away from centralized platforms and placing control back in the hands of a person – a key principle of the GHPC.

6.1.5.3 Recommendations

6.1.5.3.1 Overall Recommendations

The GHPC RECOMMENDS using Overlays Capture Architecture (OCA) as a solution for semantic harmonization between data models and data representation formats that has been specifically devised for data object interoperability, acting as a catalyst for standardized credential issuance. An OCA schema stack consists of a stable capture base and interchangeable task-oriented overlays. This level of object separation within the schema introduces a dynamic solution to presenting credentials in different languages, based on a user's locale and other language preferences (See Appendix C for more information).

OCA offers an optimal level of both efficiency and interoperability in alignment with FAIR principles (Findability, Accessibility, Interoperability, and Reusability). The architecture provides a stable infrastructure to facilitate seamless semantic harmonization and interoperability processes, not only between internal departments and functions but also between external organizations working under an **ecosystem governance framework (EGF)** (e.g., **Good Health Pass Ecosystem Governance Framework (GHP EGF)**) as defined by an industry collaborative or consortium.

The GHPC is also aware of other semantic architectures being developed in open source communities, which will be continuously evaluated and may feature in future GHPC publications.

6.1.5.3.2 Processing Recommendations

The GHPC RECOMMENDS using the FHIR-OCA data processing pipeline (see Appendix D). The FHIR-OCA tool provides for the conversion of COVID-19 vaccination, testing, or recovery JSON formatted resource bundles via a FHIR JSON-LD context that enables FHIR resource elements to be represented in JSON-LD. This approach can also include the pre-processing conversion of non-FHIR formatted data into a processable FHIR format for the FHIR-OCA tool.

This conversion has several advantages, including allowing the FHIR element definitions to be referenced in a linked data vocabulary, enabling purpose-of-use data framing, and allowing the application of Zero Knowledge Proof and linked data signature encryption techniques to the health certificate and related verifiable credentials. Details of this pipeline can be found in Appendix C.

6.1.6 Phased Approach

With the publication of this document, implementers would take a phased approach toward adoption of recommended standards. These phases would be defined along timelines measured in 30-, 90-, and 180-day segments and further accounted for across industries with different levels of agility based on established development cycles.

The standards and technology patterns being developed in this specification, and broadly as part of the GHPC activities, are designed to be used not just for the immediate needs of this phase of the COVID pandemic but to be a potential framework to handle a variety of verifiable credential use cases that may arise in the future. To that end, for implementers that require a longer lead time, (e.g., larger companies), announcements of participation via product road maps, with regular updates until official launch, would demonstrate their commitment to GHP principles to the community as a whole.

6.1.6.1 Phase One (30 Day Horizon)

1. The GHPC **RECOMMENDS** that solution providers, issuers and attesters start setting the minimum viable value sets for COVID-19 credentials required for global interoperability.
2. The GHPC also **RECOMMENDS** that CVX code and state of vaccination be included as additional data fields for North America only.
3. To further promote interoperability, the GHPC **RECOMMENDS** that health credential and pass solution providers/issuers/developers **SHOULD** start using the HL7 FHIR data model. For healthcare systems that have not, or could not adopt the FHIR data model, the GHPC **RECOMMENDS** that solution providers **SHOULD** provide guidelines and/or tools to convert non-FHIR formatted EHR data into FHIR formatted records as a pre-processing step.

6.1.6.2 Phase Two (90 Day Horizon)

1. The GHPC **RECOMMENDS** that solution providers enable semantic harmonization between data models and data representation formats to create health certificates, health credentials, and GHPs for jurisdictions they are affiliated with.
2. The GHPC **RECOMMENDS** using the FHIR-OCA data processing pipeline for the conversion of COVID-19 vaccination, testing, or recovery JSON formatted resource bundles via a FHIR JSON-LD context that enables FHIR resource elements to be represented in JSON-LD.
3. The GHPC **RECOMMENDS** that solution providers start testing their health certificates, health credentials and GHPs against rules engines for interoperability with other jurisdictions.

6.1.6.3 Phase Three (180 Day Horizon)

The GHPC **RECOMMENDS** that solution providers work with jurisdictions rolling out health certificates, health credentials and GHPs.

Appendix B: COVID-19 Credentials Initiative (CCI) Schema Task Force Data Specification Repositories

CCI Schema Reference Documents:
https://drive.google.com/drive/u/1/folders/1h4vF79KzUY6KipBt3A6kMEboiqcSYtv7

Global Reference Documents:
https://drive.google.com/drive/u/1/folders/1y-nr-YlacLvv7P54N6EUYVrBNt7YA51T

Jurisdiction Specific Reference Documents:
https://drive.google.com/drive/u/1/folders/1Gmapb1ktvm-v3NDH65AZEUJ-__-ouvcg

Appendix C: OCA Background

OCA presents a schema as a multi-dimensional object consisting of a stable capture base and interoperable overlays. Overlays are task-oriented linked data objects that provide additional coloration to the capture base. This degree of object separation enables issuers to make custom edits to the overlays rather than to the capture base itself. In other words, multiple parties can interact with and contribute to the schema structure without having to change the capture base definition. With capture base definitions remaining stable and in their purest form, a common immutable base object is maintained throughout the capture process, which enables data standardization.

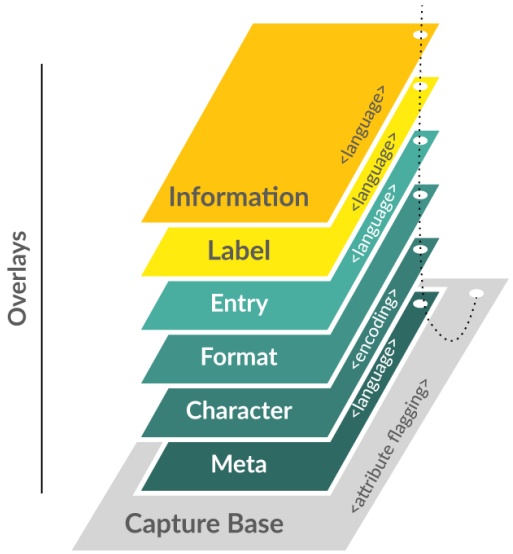

Figure 24: Overlay Stack

Explainer on OCA Objects:

Capture Base

A capture base is 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.

Attribute names and types are defined in the capture base. The construct also contains a blinding block which allows the issuer to flag any attributes that could potentially unblind the identity of a governing entity. With these attributes flagged at the base layer, all corresponding data can be treated as sensitive throughout the data lifecycle and encrypted or removed at any stage thus reducing the risk of identifying governing entities in blinded datasets.

Meta Overlay

A meta overlay is a core linked object that can be used to add contextual meta-information about the schema, including schema name, description and broad classification schemes.

Character Encoding Overlay

A character encoding overlay is a core linked object that can be used to define the character set encoding (e.g. UTF-8, ISO-8859-1, Windows-1251, Base58Check, etc.). This overlay type is useful when implementing solutions that facilitate data inputs across multiple languages.

Format Overlay

A format overlay is a core linked object that can be used to add formats, field lengths, or dictionary coding schemes to schema attributes.

Entry Overlay

An entry overlay is a core linked object that can be used to add predefined field values in a specified language to schema attributes. To minimize the risk of capturing unforeseen PII, the implementation of free form text fields is best avoided. This overlay type enables structured data to be entered thereby negating the risk of capturing and subsequently storing dangerous data.

Label Overlay

A label overlay is a core linked object that can be used to add labels in a specified language to schema attributes and categories. This overlay type enables labels to be displayed in a specific language at the presentation layer for better comprehensibility to the end user.

Information Overlay

An information overlay is a core linked object that can be used to add instructional, informational or legal prose to assist the data entry process.

https://wiki.colossi.network/en/Technologies/OCA/Developer/Adopt

Appendix D: FHIR-OCA Data Pipeline

HL7 FHIR formatted data is being specified as a preferred exchange format for EHR data being processed into vaccination, test or recovery certificates for the EU DGC, WHO SVC, VCI and other initiatives. The respective FHIR communities are creating FHIR Implementation Guides that provide technical direction on how to create and structure the relevant FHIR resources to be processed. This support from the FHIR communities helps to ensure that FHIR formatted record sets will be widely available for processing. HL7 has several other formats for health data exchange; these formats 2.x, 3.x, and CDA have libraries and mappings to support the conversion of data formatted from their respective formats into a FHIR format. The conversion of data in pre-FHIR formats and the conversion of non-FHIR formatted data into a processable FHIR format for the FHIR-OCA tool, can be implemented as a set of pre-processing steps prior to pipeline processing.

Being that FHIR EHR data records represent clinical information from the clinical application source, there are compliance, security and audit features available via FHIR that facilitate the accurate capture of state for the health event being described. This in turn provides for a transparent usage of the data keeping it aligned with the original reason for its extraction and capture. The FHIR-OCA approach to processing assumes that these mechanisms have been applied by the clinical application.

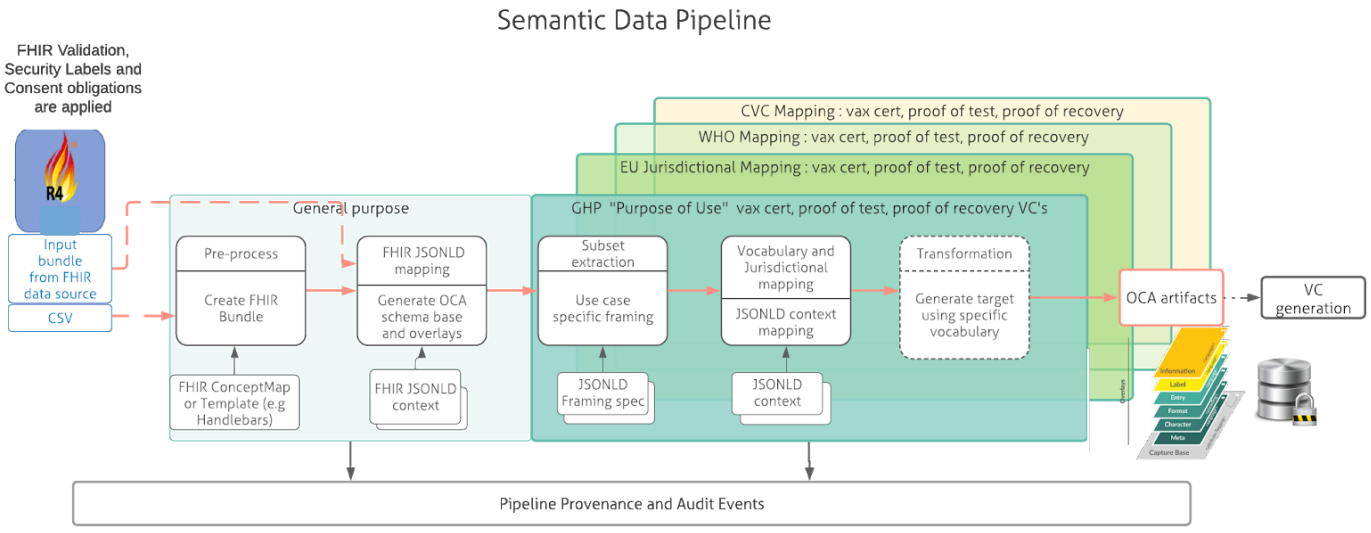

Figure 25: Semantic Data Pipeline

In the FHIR JSON-LD (JavaScript Object Notation for Linked Data) transform, the data elements are also structured into an OCA capture base. Next, the capture base elements have application context extracted and transformed into a set of overlay capture objects, the set of overlay objects are pre-defined to align with the purpose of use of the data set.

Finally, based on the purpose of use, jurisdictional and vocabulary mappings are applied and the target specific OCA artifacts are created. This set of OCA artifacts are then available for further encryption and persistence processing. https://github.com/SemanticClarity/oca-fhir-cli