ISO Datatype Support

Scope Document

Version No: 1.1

Last Modified: 3/3/2010

Team : caCORE SDK, caGrid, caAdapter

Client : National Cancer Institute -   
 Center for Bioinformatics and Information Technology,

National Institutes of Health,

US Department of Health and Human Services

Document History

Revision History

| **Version Number** | **Revision Date** | **Author** | **Summary of Changes** |
| --- | --- | --- | --- |
| 0.1 | 2/19/2010 | Scott Oster | Initial Draft |
| 0.2 | 2/19/2010 | Satish Patel | Added content for SDK and caAdapter, applied template |
| 0.3 | 2/19/2010 | Shannon Hastings | Cleanup |
| 0.4 | 2/19/2010 | John Eisenschmidt | Comments, copy edits |
| 0.5 | 2/22/2010 | Satish Patel | Reviewed and incorporated comments |
| 0.6 | 2/22/2010 | Scott Oster | Detailed query plan; removed metadata tasks |
| 0.7 | 2/24/2010 | Ye Wu | Clarified caAdapter section |
| 0.8 | 3/1/2010 | Satish Patel | Reconciled comments and changes |
| 0.9 | 3/1/2010 | Satish Patel | Added SDK constraints in Appendix A |
| 1.0 | 3/3/2010 | John Eisenschmidt | Added Security Appendix |
| 1.1 | 3/3/2010 | Satish Patel | Added SDK security in appendix. Added RESTful assumption |

Review

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Team/Role** | **Version** | **Date Reviewed** | **Reviewer Comments** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Related Documents

More information can be found in the following related documents:

|  |
| --- |
| **Document Name** |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

Table of Contents

1. Introduction 4

2. Problem Statement 4

3. Stakeholder and User Descriptions 5

3.1 Stakeholder Summary 5

4. Timeline Requirement 6

5. Assumptions 6

6. Deliverables 6

6.1 ISO21090 Java Library 6

6.2 Introduce Support 7

6.3 caAdapter Support: 7

6.4 SDK Support: 8

6.5 Service Development Process Documentation: 8

7. Release Non-Functional Requirements 9

8. APPENDIX 10

8.1 Appendix A – SDK Constraints on Database Table Configuration Options 10

8.2 Appendix B - Security Support in this Solution 12

8.2.1 Application Security (caCORE SDK) 12

8.2.2 Grid Security 12

Scope for ISO Datatype Support

# Introduction

The purpose of this document is to identify

1. Artifacts needed in order to support the ISO datatype in the core infrastructure tooling
2. High level features of the artifacts
3. Responsible teams to complete the delivery of the artifacts.

# Problem Statement

The Patient Outcomes Data Service (PODS) currently utilizes a custom implementation of a localization of the ISO 21090 data types. This localization provides the capability to serialize to, and deserialize from, the canonical XML representation (as defined by the ISO XML Schema) for the purposes of transport on the grid. Use of this implementation in clients and other implementations is non-trivial, and needs to be supported by caBIG® infrastructure and tooling.

Longer term, more complete localizations may be required to facilitate other use cases or domain usage and a similar, but more comprehensive, solution will be required. Additionally, the full impact of their use on the semantic infrastructure needs to be evaluated, and may introduce additional requirements.

# Stakeholder and User Descriptions

## Stakeholder Summary

|  |  |  |
| --- | --- | --- |
| **Name** | **Description** | **Responsibilities** |
| George Komatsoulis | Deputy Director, NCI |  |
| Charlie Mead | Chief Technology Officer, NCI CBIIT |  |
| Avinash Shanbhag | Director, Core Infrastructure Engineering | Oversees CBIIT caCORE software engineering, and caGrid development |
| Sichen Liu | Associate Director, Core Infrastructure Engineering | Oversees caAdapter, and SDK product management and development |
| Libby Prince | ISO 20190 Project Manager |  |
| PODS Project Team | The PODS implementation is the reference use case for this effort |  |
| COPPA Project Team | Contributor of the ISO datatype’s localized library | Provide support to the development teams to adopt the localized library, provide guidance on the library |
| caGrid Project Team | Development team | Provide support for the ISO datatypes in their managed tools |
| caAdapter Project Team | Development team | Provide support for the ISO datatypes in their managed tools |
| caCORE SDK Project Team | Development team | Provide support for the ISO datatypes in their managed tools |

# Timeline Requirement

* Patient Outcomes Service: Release Late March/Early April
* Infrastructure Tooling: “shortly thereafter”

# Assumptions

* COPPA/PODS localization (30 ISO datatypes, datatype attributes, and inward/outward XML serialization mapping ) is sufficient for the initial iteration of support
* caDSR registration, and other semantic infrastructure representations, are out of scope for the initial iteration.
* OCL constraint enforcement is not required of the ISO datatype Java library
* Data Service (CQL) support is not required for the initial iteration, though it will be in the longer term solution and so should inform the technical direction in the short term.
* SDK will support one specific manner in which each datatype will be mapped to the database. In longer term, additional database structures for each datatype will be supported. (See Appendix A for more details)
* SDK will support only Java API interface to access the data. RESTful and web service interface will not be supported for this release.
* Security will be supported at the Grid service layer only. (See Appendix B for more details).

# Deliverables

## ISO21090 Java Library

An independent software project will be developed by a cross-project team, in support of the following:

1. Java Beans for Localization
2. Serialization support for to/from those Beans to standard ISO XSD
3. Hibernate support for ORM mapping those Beans to database

The existing COPPA implementation will be evaluated for fitness as a starting point, and enhanced as necessary. Minimally, the removal of the intermediary JAXB-generated beans (via more complex mapping descriptions) will be explored.

As mentioned in the assumption, the Hibernate mapping layer for the datatypes will allow one database structure for each datatype. Additional database structures for each datatype will be supported in incremental releases.

**Key Personnel:**

* Scott Oster
* Shannon Hastings
* Steve Langella
* Satish Patel
* Todd Parnell
* Steve Lustbader
* Abe Evans-el

**Dependencies:**

* Existing Patient Outcomes Service implementation

**Deployment Model:**

A software distribution consisting of documentation, jar files, xsd’s, and configuration files will be provided.

## Introduce Support

An Introduce datatype extension will be developed to provide *point and click* support for using this library, and ISO data types in service interface (and other schemas which reference them).

**Key Personnel:**

* Scott Oster
* Shannon Hastings
* Steve Langella

**Dependencies:**

* The ISO21090 Java Library developed above

**Deployment Model:**

The new extension will be provided for immediate consumption (caGrid 1.3.0.2) via the Introduce Software Update site, and later, shipped with the official caGrid 1.4.

## caAdapter Support:

caAdapter’s Model Mapping Service (MMS) module will be enhanced to load object models and data model with ISO 21090 datatyps. caAdapter will also perform adequate customization and validation. In addition, a drag and drop interface will be enhanced to perform mapping between object models and data models with ISO datatypes.

**Key Personnel:**

* Eugene Wang
* Ye wu

**Dependencies:**

* The ISO21090 Java Library developed above
* UML tag value specification from SDK team

**Deployment Model:**

caAdapter software distribution and user documentation will be provided.

## SDK Support:

The caCORE SDK 4.2.1 will be enhanced to support ISO 21090 data types. The code generator framework will be enhanced to leverage the ISO 21090 library enabled with the Hibernate support. The SDK’s middleware system (a.k.a. SDK generated system) will be enhanced to provide constrained QBE query support and serialization of the beans with ISO datatypes.

**Key Personnel:**

* Satish Patel
* Dan Dumitru

**Dependencies:**

* The ISO21090 Java Library developed above

**Deployment Model:**

The SDK distribution consisting of software, UML tag value specification for model mapping, and user documentation will be provided.

## Service Development Process Documentation:

The process of developing a grid service that leverages the ISO21090 in its operations, and queries the SDK, will be documented. The documentation will describe how to use the SDK to create the generated system, create a new grid service in Introduce, add the ISO21090 data types to the service via the Introduce support, and create and implement grid service operations which pass these data types. A step-by-step tutorial will be provided, as well as a reference implementation “example service.” The tutorial and example service will provide reusable code snippets that demonstrate service layer operations in the grid service interface implemented via the SDK-generated QBE APIs. It is expected this process will be refined over time, and will drive the requirements for later iterations of the infrastructure support (akin to the current caGrid tooling for creating Data Services).

**Key Personnel:**

* Scott Oster
* Shannon Hastings
* Steve Langella
* Satish Patel

**Dependencies:**

* The ISO21090 Java Library developed above
* caCORE SDK Support
* Introduce Support

**Deployment Model:**

A tutorial will be available on the caGrid website, with a downloadable “solution service.”

# Release Non-Functional Requirements

1. **Open Source**

The product will continue to use open source tools and technologies. If an appropriate open source software or tool is not available then prior approval from the product manager will be required before using the tool.

1. **Technology Stack compatibility**

The products will continue to adhere to the CBIIT technology stack when choosing the versions of the tools it’s using. Any deviation from the technology stack will be pre-approved by the product manager.

1. **Automated builds**

The products will continue to support their existing automated build processes.

# APPENDIX

## Appendix A – SDK Constraints on Database Table Configuration Options

The ISO datatype is a complex structure with multiple attributes within each datatype. The nested attributes within each datatype can be of simple string type or can be other complex ISO datatypes. As these datatypes are used in the persistence tier, the user has to map these datatypes in the database. In order to perform mapping of the datatypes from the object layer to the database, one has to prepare the database structure to store the information in each of the datatype. Due to the complex nature of the ISO 21090 datatype structure, one can map the datatype in more than one way in the database table.

**Example:**

AD is one of the commonly used ISO datatype. The UML diagram from AD is shown below:



The AD datatype can be mapped in more than one way to the relational database. Following are two of the possible alternatives to map the datatypes to the relational database.

1. Map attributes of the AD datatype in the columns of a separate table with foreign key reference to the table corresponding to the Person object. This scenario is illustrated by PERSON1 and ADDRESS1 tables.
2. Map attributes of the AD datatype in the columns of the table corresponding to the Person object. This scenario is illustrated by PERSON2 table.

These two database mapping options for AD datatype are illustrated in the diagram below:



**Advantage:**

The proposed approach of providing support for one way to perform database table mapping will limit the delivery scope for the development teams in following manner

1. **SDK:**
2. Reduced number of O/R configurations in the Hibernate user types library
3. Reduced number of configurations in the code generator
4. Reduced validation requirements
5. Simplified Query By Example translator
6. Simplified Query specifications for the end user
7. **caAdapter:**
8. Reduced number of O/R configurations
9. Reduced validation requirements
10. **caGrid:**
11. Simplified Query By Example translator
12. Simplified Query specifications for the end user

Given the short timeframe of the project, the reduced scope will help the development team to focus on the core and essential functionalities to make the solution work.

**Disadvantages:**

In the first release of ISO21090 implementation, the SDK will allow using one of the possible options to create the database table structure. Due to the proposed restriction, the user of the SDK will be required to create data model based on the specifications provided by the SDK.

**Resolution:**

The SDK’s proposed architecture to support the ISO21090 datatype can be extended in future to provide support for alternate database configurations not supported in the initial version. SDK team can incrementally add support for these additional database configurations in the future.

## Appendix B - Security Support in this Solution

### Application Security (caCORE SDK)

Security features in the caCORE SDK will not be supported in this release. User can secure the application using the grid security when creating the grid service.

### Grid Security

#### Authentication

Existing x.509/PKI-based authentication will work with ISO-type services as it does with standard data and analytical services.

#### Authorization

The use of CSM and/or Grid Grouper to perform authorization decisions on the invocation of service operations will work with ISO-type services as it does with standard data and analytical services.