i2b2 Cell Messaging

Patient Data Object (PDO)

1 Table of Contents

2 Document Version History

This section lists only major document versions and not detailed changes. Versions are indicated by sequential integers.

Date	Version	Description	Author(s)
12/01/2006	0.9	Pre-Release Version	Kristel Hackett

3 Introduction
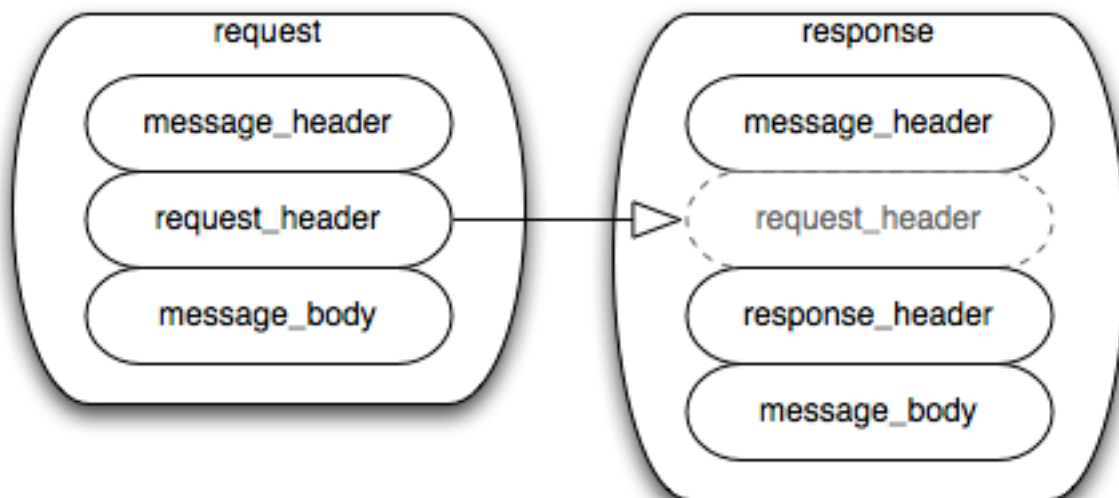
This document gives an overview of i2b2 cell messaging as well as a more detailed description of the Patient Data Object (PDO) xml data format.

3.1 The i2b2 Hive

The Informatics for Integrating Biology and the Bedside (i2b2) is one of the sponsored initiatives of the NIH Roadmap National Centers for Biomedical Computing (http://www.bisti.nih.gov/ncbc/). One of the goals of i2b2 is to provide clinical investigators broadly with the software tools necessary to collect and manage project-related clinical research data in the genomics age as a cohesive entity – a software suite to construct and manage the modern clinical research chart. The i2b2 hive is a set of cells or modules that have a common messaging protocol that allow the cells to interact using web services and XML messages.

3.2 i2b2 Cell Messaging

All cells in the i2b2 hive must communicate using standard i2b2 XML messages. This message specifies certain properties that are common to cells and essential to the administration tasks associated with sending, receiving and processing messages. All requests are sent using a <request> tag and responses are returned using a <response> tag. The same <message_header> tag is used for both. The <request_header> is used for requests but may optionally be echoed back in the response. The response must include a <response_header>. The XSD specification of the i2b2 message permits individual cells to add cell-specific XML in the <message_body> tag. This cell-specific XML need not extend the i2b2 message schema since the i2b2 schema will allow insertion of tags from any namespace into the <message_body> tag. The following table illustrates the basic top-level elements contained within the request and response messages.

The i2b2 XML schema consists of three XSD files:

- **i2b2.xsd**
  This schema is not used directly to create i2b2 messages, but is included in the i2b2_request.xsd and the i2b2_response.xsd. It defines the <message_header> tag.

- **i2b2_request.xsd**
  This schema is used for validating i2b2 request messages. It defines the <i2b2:request> tag, which includes the <message_header> tag.

- **i2b2_response.xsd**
  This schema is used for validating i2b2 response messages. It defines the <i2b2:response> tag, which includes the <message_header> tag.

4 Patient Data Object (PDO)

The Patient Data Object is a representation of patient data.

4.1 XML Schema Definitions

The Patient Data Object (PDO) XML schema consists of three XSD files:

- **i2b2_PDO.xsd**
  This schema is used for validating PDO and defines a <patient_data> tag.

- **i2b2_PDO_fields.xsd**
  This schema is not used directly to validate a PDO but it is included in the i2b2_PDO.xsd schema. It defines different PDO building blocks.

- **i2b2_PDO_types.xsd**
  This schema is not used directly to validate a PDO but it is included in the i2b2_PDO_fields.xsd schema. It defines basic data types that are reused.

5 Example Object

```xml
<repository:patient_data
            xmlns:repository="http://i2b2.mgh.harvard.edu/repository_cell"

xsi:schemaLocation="http://i2b2.mgh.harvard.edu/repository_cell/patient_data.
xsd">
            <visit_dimension_set>
                <visit_dimension>
                    <encounter_ide source="PULENCTMAP">XDFD</encounter_ide>
                    <patient_num>12345</patient_num>
                    <inout_cd>Inpatient</inout_cd>
                    <location_cd>MGH</location_cd>
                    <location_path>Clinics\MGH\Oral Surgury</location_path>
                    <start_date>1999-02-28T13:59:00</start_date>
                    <end_date>1999-02-28T13:59:00</end_date>
                    <visit_blob/>
                    <update_date>1999-02-28T13:59:00</update_date>
                    <download_date>1999-02-28T13:59:00</download_date>
                    <import_date>1999-02-28T13:59:00</import_date>
                    <sourcesystem_cd>RPDRASTHMA</sourcesystem_cd>
                    <upload_id>123</upload_id>
                </visit_dimension>
            </visit_dimension_set

            <concept_dimension_set>
                <concept_dimension>
                    <concept_path>i2b2\Diagnoses\C0004096</concept_path>
                    <concept_cd>C0004096</concept_cd>
                    <name_char>Asthma</name_char>
                    <concept_blob/>
                    <update_date>1999-02-28T13:59:00</update_date>
                    <download_date>1999-02-28T13:59:00</download_date>
                    <import_date>1999-02-28T13:59:00</import_date>
                    <sourcesystem_cd>ASTHMADICT</sourcesystem_cd>
                    <upload_id>234</upload_id>
                </concept_dimension>
            </concept_dimension_set>

            <provider_dimension_set>
                <provider_dimension>
<provider_path>BWH\Pathology\B00123</provider_path>
                    <provider_id>B00123</provider_id>
                    <name_char>Doctor, John A., MD</name_char>
                    <provider_blob/>
                    <update_date>1999-02-28T13:59:00</update_date>
                    <download_date>1999-02-28T13:59:00</download_date>
                    <import_date>1999-02-28T13:59:00</import_date>
                    <sourcesystem_cd>token</sourcesystem_cd>
                    <upload_id>345</upload_id>
                </provider_dimension>
            </provider_dimension_set>

            <patient_dimension_set>
```

```xml
<patient_dimension>
    <patient_id source="EMP">123456789</patient_id>
    <vital_status_cd>Y</vital_status_cd>
    <birth_date>1930-02-28T13:59:00</birth_date>
    <death_date>2001-02-28T13:59:00</death_date>
    <sex_cd>F</sex_cd>
    <age_in_years_num>71</age_in_years_num>
    <language_cd>ENGL</language_cd>
    <race_cd>B</race_cd>
    <marital_status_cd>M</marital_status_cd>
    <religion_cd>CA</religion_cd>
    <zip_cd>12345-1234</zip_cd>
    <statecityzip_path>Zip
codes\Massachusetts\Boston\02120\</statecityzip_path>
    <patient_blob/>
    <update_date>1999-02-28T13:59:00</update_date>
    <download_date>1999-02-28T13:59:00</download_date>
    <import_date>1999-02-28T13:59:00</import_date>
    <sourcesystem_cd>token</sourcesystem_cd>
    <upload_id>456</upload_id>
</patient_dimension>
</patient_dimension_set>

<observation_fact_set>
    <observation_fact>
        <encounter_num>1234567</encounter_num>
        <patient_ide
source="EMP">XRTU8GT9FKIS35DSVB</patient_ide>
        <concept_cd>V450</concept_cd>
        <provider_id>12345</provider_id>
        <start_date>1999-02-28T13:59:00</start_date>
        <modifier_cd>2</modifier_cd>
        <valtype_cd>N</valtype_cd>
        <tval_char>=</tval_char>
        <nval_num>1.0</nval_num>
        <valueflag_cd>H</valueflag_cd>
        <quantity_num>1</quantity_num>
        <units_cd>ml</units_cd>
        <end_date>1999-02-28T13:59:00</end_date>
        <location_cd>MGH</location_cd>
        <confidence_num>95</confidence_num>
        <observation_blob/>
        <update_date>1999-02-28T13:59:00</update_date>
        <download_date>1999-02-28T13:59:00</download_date>
        <import_date>1999-02-28T13:59:00</import_date>
        <sourcesystem_cd>RPDRASTHMA</sourcesystem_cd>
        <upload_id>567</upload_id>
    </observation_fact>
</observation_fact_set>

</repository:patient_data>
```

6 Object Explanations

PatientData: The root element that holds data from the patient data tables. May contain any number of visit_dimension, concept_dimension, provider_dimension, patient_dimension and observation_fact elements. They can occur in any order.

visit_dimension: One row of data from the visit_dimension table.

encounter_idChoice: A choice between Encounter_Num, Encounter_Id, Encounter_Ide. See Special Groups, below.

patient_idChoice: A choice between Patient_Num, Patient_Id, Patient_Ide. See Special Groups, below.

InOut_Cd: A code classifying a vist by inpatient/outpatient status.

Location_Cd: A code representing the hospital associated with this visit.

Location_Path: A path that represents the hierarchical specification of the location.

Start_Date: The date that the observation was made, or that the observation started. If the data is derived or calculated from another observation (like a report) then the start date is the same as the observation it was derived or calculated from.

End_Date: The date that the observation was made, or that the observation ended. If the data is derived or calculated from another observation (like a report) then the end_date is the same as the observation it was derived or calculated from.

Visit_Blob: XML data that includes partially structured and unstructured data about a visit.

annotationGroup: A group of fields that store annotation or administrative information. (Update_Date, Download_Date, Import_Date, Sourcesystem_Cd, Upload_Id) See Special Groups below.

concept_dimension: One row of data from the concept_dimension table.

Concept_Cd: A unique code that represents a concept.

Name_Char: A string name that represents this concept, idea or person.

Concept_Blob: XML data that includes partially structured and unstructured data about a concept.

annotationGroup: A group of fields that store annotation or administrative information. (Update_Date, Download_Date, Import_Date, Sourcesystem_Cd, Upload_Id) See Special Groups below.

provider_dimension: One row of data from the provider_dimension table.

Provider_Path: A string representing the unique position of a provider in a hierarchical department tree.

Provider_Id: An ID that represents the provider, which could be a physician or a machine such as an MRI machine.

Name_Char: A string name that represents this concept, idea or person.

Provider_Blob: XML data that includes partially structured and unstructured data about a provider.

annotationGroup: A group of fields that store annotation or administrative information. (Update_Date, Download_Date, Import_Date, Sourcesystem_Cd, Upload_Id) See Special Groups, below.

patient_dimension: One row of data from the patient_dimension table.

patient_idChoice: A choice between Patient_Num, Patient_Id, Patient_Ide. See Special Groups, below.

Vital_Status_Cd: A code representing the vital status of a patient.

Birth_Date: The day, month and year of birth, time can default to midnight.

Death_Date: The day, month and year of death, time can default to midnight.

Sex_Cd: A code representing the sex of an individual.

Age_In_Years_Num: Age of patient in years within 1 year of their actual age.

Language_Cd: A code representing a language of communication.

Race_Cd: A code representing race.

Marital_Status_Cd: A code representing marital status.

Religion_Cd: A code representing religious affiliation.

Zip_Cd: A zipcode with 5 base digits, optionally including a hyphen and 4 extended digits.

StateCityZip_Path: A path type that includes the state, city and zipcode.

Patient_Blob: XML data that includes partially structured and unstructured data about a patient.

annotationGroup: A group of fields that store annotation or administrative information. (Update_Date, Download_Date, Import_Date, Sourcesystem_Cd, Upload_Id) See Special Groups, below

observation_fact: One row of data from the observation_fact table.

encounter_idChoice: A choice between Encounter_Num, Encounter_Id, Encounter_Ide. See Special Groups, below.

patient_idChoice: A choice between Patient_Num, Patient_Id, Patient_Ide. See Special Groups, below.

Concept_Cd: A unique code that represents a concept.

Provider_Id: An ID that represents the provider, which could be a physician or a machine such as an MRI machine.

Start_Date: The date that the observation was made, or that the observation started. If the data is derived or calculated from another observation (like a report) then the start date is the same as the observation it was derived or calculated from.

Modifier_Cd: A code representing whether a diagnosis is an admitting or whether a diagnosis or procedure is primary or secondary.

ValType_Cd: A code representing whether a value is stored in the TVal column or NVal column.

TVal_Char: A text value.

NVal_Num: A numerical value.

ValueFlag_Cd: A code that represents the type of value present in the NVal_Num or the TVal_Char.

Quantity_Num: The number of observations represented by this fact.

Units_Cd: A textual description of the units associated with a value.

End_Date: The date that the observation was made, or that the observation ended. If the data is derived or calculated from another observation (like a report) then the end_date is the same as the observation it was derived or calculated from.

Location_Cd: A code representing the hospital associated with this visit.

Confidence_Num: A code or number representing the confidence in the accuracy of the data.

Observation_Blob: XML data that includes partially structured and unstructured data about an observation.

annotationGroup: A group of fields that store annotation or administrative information. (Update_Date, Download_Date, Import_Date, Sourcesystem_Cd, Upload_Id) See Special Groups, below.

encounter_idChoice: A choice between Encounter_Num, Encounter_Id, Encounter_Ide. See Special Groups, below.

Encounter_Num: An integer that is unique to each Encounter.

Encounter_Id: A code and a sourcesystem that is unique to each Encounter.

Encounter_Ide: An encrypted code and a sourcesystem that is unique to each Encounter.

patient_idChoice: A choice between Patient_Num, Patient_Id, Patient_Ide. See Special Groups, below.

Patient_Num: An integer that is unique to each Patient.

Patient_Id: A code and a sourcesystem that is unique to each Patient.

Patient_Ide: An encrypted code and a sourcesystem that is unique to each Patient.

annotationGroup: A group of fields that always appear together at the end of a table and store annotation or administrative information.

Update_Date: The date the data was last updated according to the source system from which the data was obtained. If the source system does not supply this data, it defaults to the download_date.

Download_Date: The date that the data was obtained from the source system. If the data is derived or calculated from other data, then the download_date is the date of the calculation.

Import_Date: The date the data is placed into the table of the data mart.

SourceSystem_Cd: A code representing the source system that provided the data.

Upload_Id: Tracking number assigned to any file uploaded.