Correlation Analysis Cell: Messaging Specifications

Version 1.0

Table of Contents

Correlation Analysis Cell Messaging version 1.0

About this Guide

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 guide provides a description of the common messaging used by the *Correlation Analysis Cell* of the i2b2 hive. This specialized analysis cell uses mutual information theory to calculate observed correlations within the data of the hive. This type of cell represents an important achievement of the hive.

Document Version History

Date	Revision	Description	Author(s)
June 25, 2008	version 1.0	Initial revision, 1.0	Vlad Valtchinov

1

Messaging Overview

This document reviews the common i2b2 messaging used in the *Correlation Analysis Cell*. Consult the *i2b2 Common Header Messaging* document from i2b2's web site for a complete set of definitions used in the messaging within the hive. We follow this document for a brief overview of the i2b2 messaging given next.

i2b2 Hive
The hive is comprised of a collection of cells representing unique functional units. Cells in the hive have an array of roles, such as data storage, data analysis, ontology or identity management, natural language processing, and data conversion, derivation or de-identification. Each cell is a self-contained modular application that communicates with other cells via XML web services. A common i2b2 messaging protocol has been defined to enable the cells to interact with each other, sharing business logic, processes and data.

i2b2 Messaging Structure Overview
All cells in the i2b2 hive communicate using standard, pre-defined i2b2 XML request and response messages.

A request message is sent from a client to a service and contains information, inside the top-level <request> tag, that allows the service to satisfy the request. The <request> tag contains a <message_header>, <request_header> and <message_body> as shown, below, in Figure 1.

The service sends back a response message, inside a top-level <response> tag, which informs the client about the status of the request and may also contain the actual results. The <response> tag contains it's own <message_header>, <response_header> and <message_body> and it may optionally echo the request's <request_header> as shown, below, in Figure 1.
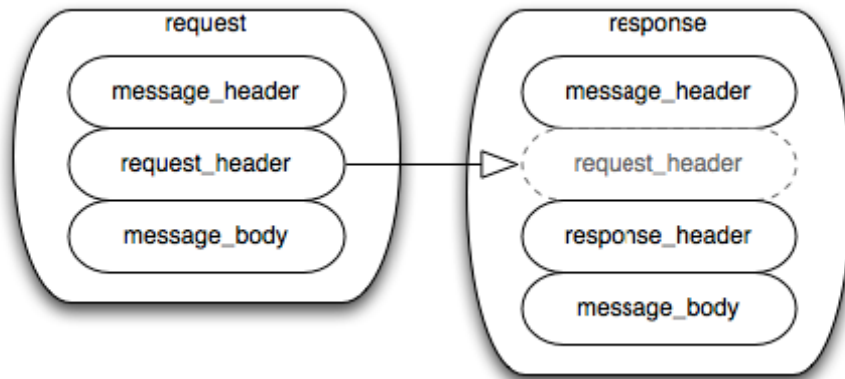
Correlation Analysis Cell Messaging version 1.0

Figure 1: The basic structure of a request and response message. The request_header in the request can be echoed in the response.

Message Header
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. Both request and response messages contain this <message_header> tag which has control information such as sending application, receiving application and message type.

Request Header
The request must contain a <request_header> tag which includes information about how to process a request such as the amount of time it is willing to wait for a response. The <request_header> tag may optionally be echoed back in the response.

Response Header
The response must include a <response_header> tag which includes general information about the response such as status and error messages or where to look for the results if they are not included with the response.

Message Body
Both request and response messages contain a <message_body> tag which may contain any well-formed xml. Individual cells may define cell-specific XML that will be put inside <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.
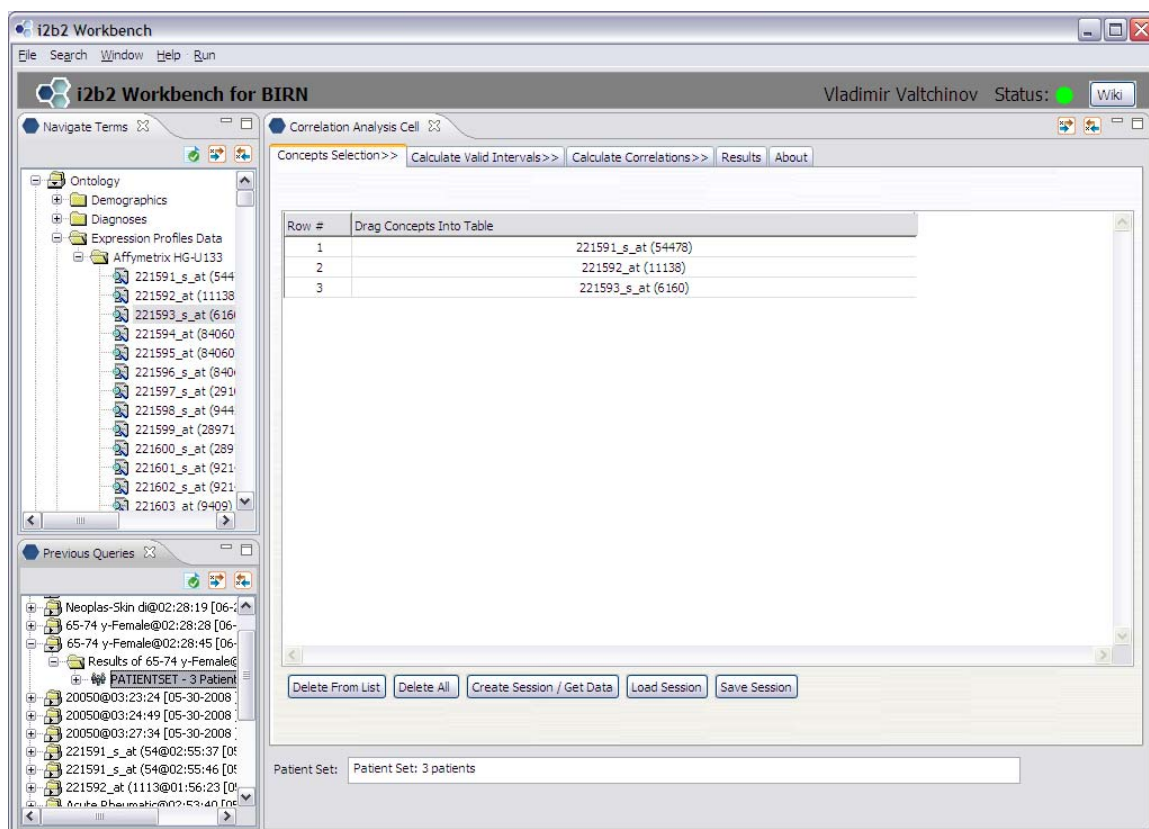
2

Correlation Cell: Messaging Workflow

The *Correlation Analysis Cell* uses the i2b2 common messaging to get the concept data it operates on.

The process is driven by the selection of a list of concepts for analysis. By default, data will be retrieved for all patients for a given concept.

Upon start, the user is prompted to select concepts for correlation.

After the selection is made (see above screen shot where 3 concepts of type expression profiles data were selected for analysis), one can use a specific Patient Set to restrict the data span to only these patients.

Correlation Analysis Cell Messaging version 1.0

A Patient Set can be most easily defined and formatted for use by the *Correlation Analysis Cell* using previous queries. In the *Previous Queries* plug-in window, navigate to the corresponding branch in the list of previous queries (see above). The selected Patient Set can be drag-and-dropped in the active area below the session and list management buttons on the *Results Selection>>* main area.

The *Cell* initiates the data retrieval when one clicks on the *Create Session / Get Data* button. At this point, an XML formatted Request message is sent out to the common data service. It follows the i2b2 hive common messaging specifications briefly covered in the last Chapter.

The *Correlation Analysis Cell* parses out the Response XML received back from the w3 services to retrieve the data for use in the computations.

3

Correlation Cell: Messaging Details

We describe in some details the structure of the Request and Response XML-s that form the basis of this *Cell*'s data communications with the i2b2 Hive.

Request XML

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<ns6:request xmlns:ns4="http://www.i2b2.org/xsd/cell/crc/psm/1.1/" xmlns:ns7="http://www.i2b2.org/xsd/cell/ont/1.1/"
xmlns:ns3="http://www.i2b2.org/xsd/cell/crc/pdo/1.1/" xmlns:ns5="http://www.i2b2.org/xsd/hive/plugin/"
xmlns:ns2="http://www.i2b2.org/xsd/hive/pdo/1.1/" xmlns:ns6="http://www.i2b2.org/xsd/hive/msg/1.1/"
xmlns:ns8="http://www.i2b2.org/xsd/cell/crc/psm/querydefinition/1.1/">
  <message_header>
    <sending_application>
      <application_name>i2b2_QueryTool</application_name>
      <application_version>0.2</application_version>
    </sending_application>
    <sending_facility>
      <facility_name>PHS</facility_name>
    </sending_facility>
    <receiving_application>
      <application_name>i2b2_DataRepositoryCell</application_name>
      <application_version>0.2</application_version>
    </receiving_application>
    <receiving_facility>
      <facility_name>PHS</facility_name>
    </receiving_facility>
    <security>
      <domain>i2b2test</domain>
      <username>vv05</username>
      <password>....</password>
    </security>
    <message_type>
      <message_code>Q04</message_code>
      <event_type>EQQ</event_type>
    </message_type>
    <message_control_id>
      <message_num>3EK0JDlW4ATIKLr3ZCoa</message_num>
      <instance_num>0</instance_num>
    </message_control_id>
    <processing_id>
      <processing_id>P</processing_id>
      <processing_mode>I</processing_mode>
    </processing_id>
    <accept_acknowledgement_type>messageId</accept_acknowledgement_type>
    <project_id>BIRN</project_id>
  </message_header>
  <request_header>
```

```
            <result_waittime_ms>180000</result_waittime_ms>
        </request_header>
        <message_body>
            <ns3:pdoheader>
                <patient_set_limit>0</patient_set_limit>
                <estimated_time>180000</estimated_time>
                <request_type>getPDO_fromInputList</request_type>
            </ns3:pdoheader>
            <ns3:request xsi:type="ns3:GetPDOFromInputList_requestType" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance">
                <input_list>
                    <patient_list max="30000001" min="1">
                        <patient_set_coll_id>143</patient_set_coll_id>
                    </patient_list>
                </input_list>
                <filter_list>
                    <panel name="\\i2b2\i2b2\Expression Profiles Data\Affymetrix HG-U133\221591_s_at (54478)">
                        <invert>0</invert>
                        <panel_accuracy_scale>0</panel_accuracy_scale>
                        <total_item_occurrences>0</total_item_occurrences>
                        <item>
                            <item_key>\\i2b2\i2b2\Expression Profiles Data\Affymetrix HG-U133\221591_s_at (54478)</item_key>
                            <dim_tablename>concept_dimension</dim_tablename>
                            <dim_dimcode>i2b2\Expression Profiles Data\Affymetrix HG-U133\221591_s_at (54478)</dim_dimcode>
                        </item>
                    </panel>
(...........deleted content...................)
                </filter_list>
                <output_option>
                    <observation_set blob="false" onlykeys="false"/>
                    <patient_set select="using_input_list" onlykeys="false"/>
                </output_option>
            </ns3:request>
        </message_body>
    </ns6:request>
```

The Patient List defined in the selection step is included in the <patient_list></patient_list> tag. It passes the internal ID of the list as creted and referenced by the *Previous Queries Cell* for the specific user (vv05 in this instance).

```
<patient_list max="30000001" min="1">
            <patient_set_coll_id>143</patient_set_coll_id>
</patient_list>
```

The identification and enumeration of the concepts that are included in the selection list is done via the <panel_name> tag. It defines the concept's name and its full path as included in the Concept_Dimension table of the i2b2 Data Schema.

```
<panel name="\\i2b2\i2b2\Expression Profiles Data\Affymetrix HG-U133\221591_s_at (54478)">
            <invert>0</invert>
            <panel_accuracy_scale>0</panel_accuracy_scale>
            <total_item_occurrences>0</total_item_occurrences>
            <item>
                <item_key>\\i2b2\i2b2\Expression Profiles Data\Affymetrix HG-U133\221591_s_at (54478)</item_key>
                <dim_tablename>concept_dimension</dim_tablename>
                <dim_dimcode>i2b2\Expression Profiles Data\Affymetrix HG-U133\221591_s_at (54478)</dim_dimcode>
            </item>
```

Correlation Analysis Cell Messaging version 1.0

```
</panel>
```

There is a separate `<panel_name>` tag for each of the concepts that are included in the list for correlation.

Response XML

```xml
<ns5:response xmlns:tns="http://axis2.crc.i2b2.harvard.edu" xmlns:ns9="http://www.i2b2.org/xsd/hive/msg/result/1.1/"
xmlns:ns4="http://www.i2b2.org/xsd/cell/crc/psm/1.1/" xmlns:ns7="http://www.i2b2.org/xsd/cell/pm/1.1/"
xmlns:ns3="http://www.i2b2.org/xsd/cell/crc/pdo/1.1/" xmlns:ns5="http://www.i2b2.org/xsd/hive/msg/1.1/"
xmlns:ns2="http://www.i2b2.org/xsd/hive/pdo/1.1/" xmlns:ns6="http://www.i2b2.org/xsd/cell/crc/psm/querydefinition/1.1/"
xmlns:ns8="http://www.i2b2.org/xsd/cell/ont/1.1/">
  <message_header>
    <i2b2_version_compatible>1.1</i2b2_version_compatible>
    <hl7_version_compatible>2.4</hl7_version_compatible>
    <sending_application>
      <application_name>edu.harvard.i2b2.crc</application_name>
      <application_version>1.0</application_version>
    </sending_application>
    <sending_facility>
      <facility_name>i2b2 Hive</facility_name>
    </sending_facility>
    <receiving_application>
      <application_name>i2b2_QueryTool</application_name>
      <application_version>0.2</application_version>
    </receiving_application>
    <receiving_facility>
      <facility_name>i2b2 Hive</facility_name>
    </receiving_facility>
    <message_control_id>
      <instance_num>1</instance_num>
    </message_control_id>
    <project_id>BIRN</project_id>
  </message_header>
  <response_header>
    <info>Log information</info>
    <result_status>
      <status type="DONE">DONE</status>
      <polling_url interval_ms="100" />
    </result_status>
  </response_header>
  <message_body>
    <ns3:response xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="ns3:patient_data_responseType">
      <ns2:patient_data>
        <ns2:patient_set>
          <patient>
            <patient_id source="HIVE">1000000025</patient_id>
            <param name="vital_status_cd" column="vital_status_cd">N</param>
            <param name="birth_date" column="birth_date">2033-10-06T00:00:00.000-04:00</param>
            <param name="sex_cd" column="sex_cd">F</param>
            <param name="age_in_years_num" column="age_in_years_num">73</param>
            <param name="language_cd" column="language_cd">spanish</param>
            <param name="race_cd" column="race_cd">hispanic</param>
            <param name="religion_cd" column="religion_cd">roman catholic</param>
            <param name="marital_status_cd" column="marital_status_cd">widow</param>
            <param name="zipcode_char" column="zipcode_char" />
            <param name="statecityzip_path_char" column="statecityzip_path_char" />
          </patient>
```

```xml
(………………deleted content……………………………………)
        </ns2:patient_set>
        <ns2:observation_set panel_name="\\i2b2\i2b2\Expression Profiles Data\Affymetrix HG-U133\221591_s_at
(54478)">
            <observation>
                <event_id>1000000025</event_id>
                <patient_id source="HIVE">1000000025</patient_id>
                <concept_cd>Affy:221591_s_at</concept_cd>
                <observer_cd>@</observer_cd>
                <start_date>2005-04-05T00:00:00.000-04:00</start_date>
                <modifier_cd>@</modifier_cd>
                <valuetype_cd>N</valuetype_cd>
                <tval_char>E</tval_char>
                <nval_num>426.83057</nval_num>
                <valueflag_cd>@</valueflag_cd>
                <end_date>2005-04-05T00:00:00.000-04:00</end_date>
                <location_cd>@</location_cd>
            </observation>
            <observation>
                <event_id>11079</event_id>
                <patient_id source="HIVE">1000000096</patient_id>
                <concept_cd>Affy:221591_s_at</concept_cd>
                <observer_cd>@</observer_cd>
                <start_date>2005-04-05T00:00:00.000-04:00</start_date>
                <modifier_cd>@</modifier_cd>
                <valuetype_cd>N</valuetype_cd>
                <tval_char>E</tval_char>
                <nval_num>551.8</nval_num>
                <valueflag_cd>@</valueflag_cd>
                <end_date>2005-04-05T00:00:00.000-04:00</end_date>
                <location_cd>@</location_cd>
            </observation>
(………………….deleted content………………………………….)
        </ns2:observation_set>
      </ns2:patient_data>
    </ns3:response>
  </message_body>
</ns5:response>
```

The response XML first lists the Patient Set for which the data query was performed and data retrieved. This is done via the

```xml
<ns2:patient_set>
    <patient>…</patient>
</ns2:patient_set>
```

construct and it is highlighted in yellow for one of the patients with an internal ID of 1000000025.

The actual data points are returned via the < ns2:observation_set panel_name> tag that contains an <observation> tag with more or one <event_id> tags. The <event_id> tag gives lists all the data fields and other attributes know about this particular measurement.

```xml
        <ns2:observation_set panel_name="\\i2b2\i2b2\Expression Profiles Data\Affymetrix HG-U133\221591_s_at
(54478)">
            <observation>
                <event_id>1000000025</event_id>
                <patient_id source="HIVE">1000000025</patient_id>
```

Correlation Analysis Cell Messaging version 1.0

```
        <concept_cd>Affy:221591_s_at</concept_cd>
        <observer_cd>@</observer_cd>
        <start_date>2005-04-05T00:00:00.000-04:00</start_date>
        <modifier_cd>@</modifier_cd>
        <valuetype_cd>N</valuetype_cd>
        <tval_char>E</tval_char>
        <nval_num>426.83057</nval_num>
        <valueflag_cd>@</valueflag_cd>
        <end_date>2005-04-05T00:00:00.000-04:00</end_date>
        <location_cd>@</location_cd>
    </observation>
```

The data is parsed out in a specific way and stored in the internal data structures for further processing in finding the *valid intervals* and performing the pair-wise correlation computation with or without the *time-overlapping* step.

This data workflow and the structure of the Request and Response XML messages are very similar to the ones used by the *Timeline View Cell*.

Correlation Analysis Cell Messaging version 1.0