Overview

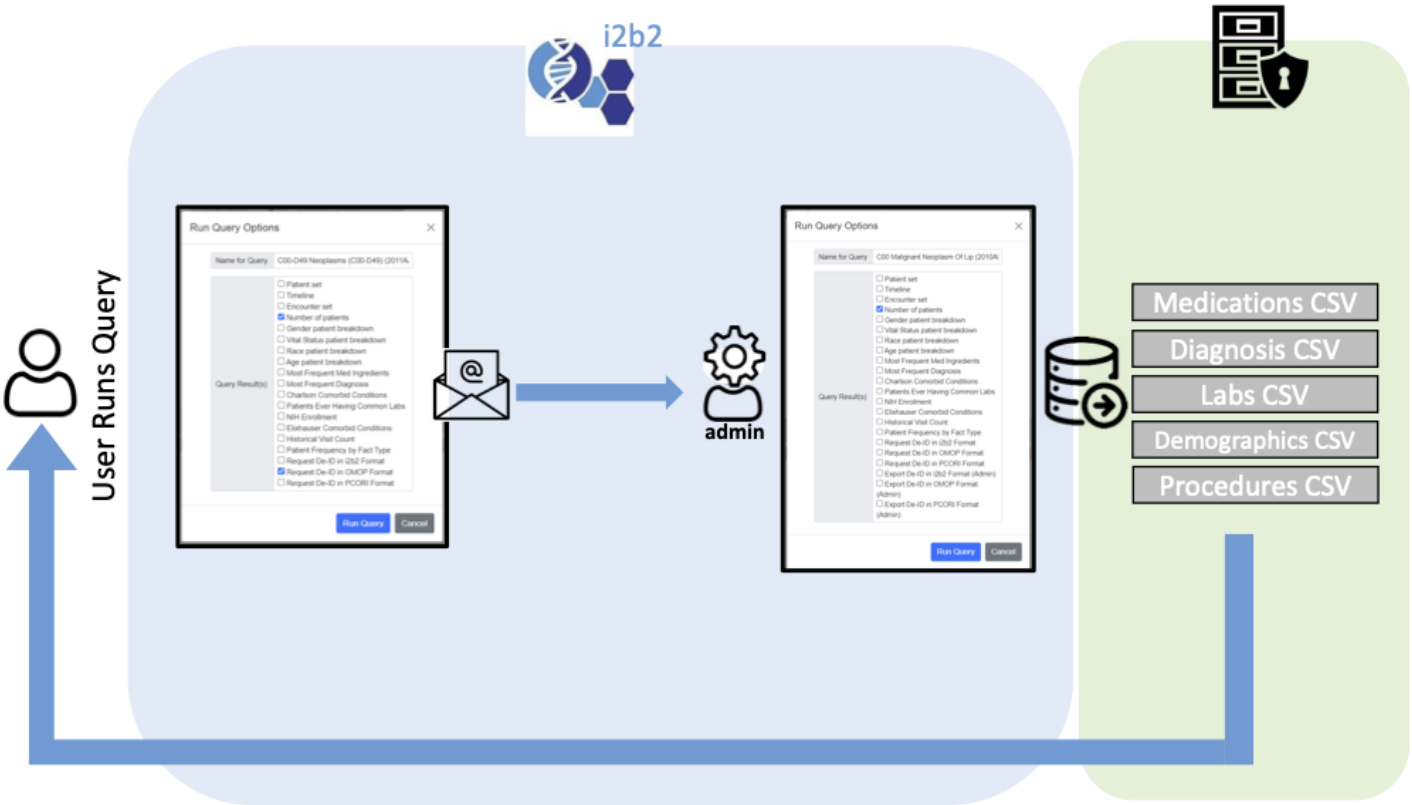
Data Exporter (v 1.8.1) · https://i2b2.org/docs/data-exporter-v-1-8-1/

- Overview
- Data Export Workflow
  - Data Request (User Process)
  - Data Export (Manager Process)
    - Sample Letter generated for Data Manager
  - Example Export files (all data are fake)
    - Demographics
    - Diagnosis
- Data Export Configuration
  - Design and Architecture
    - Database changes
    - Configuring Database Parameters
      - Request/Export query types: QT_QUERY_RESULT_TYPE
      - Request/Export file definition details: QT_BREAKDOWN_PATH
      - Export file format/location and Email parameters set-up: HIVE_CELL_PARAMS
      - Export query definition variables
      - Table of allowable characters in {{{DATE_xxx}}} strings:
    - Request and Export Results:
    - Software Changes:

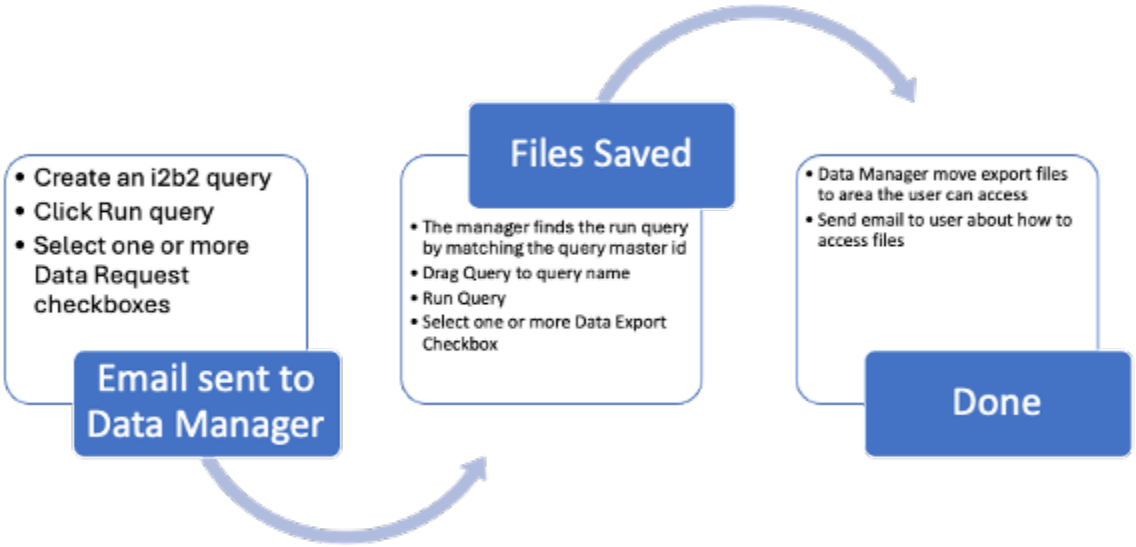
Overview

The Data Exporter functionality enables an i2b2 user to request patient data for the patients found by a desired query. In the configured workflow, data is requested and generated through new breakdown types that can be selected in the Run Query dialog. Users will see "data request" breakdowns, which send an e-mail to a predefined data manager requesting a type of data. The data manager (Manager user role) processes the request by rerunning the query with a special "data export" breakdown. The patient data is exported as a file and stored in a specified location for retrieval.

The patient data export types, request options, and file specifications are configured in the database using specific request type entries and configuration parameters in the database.

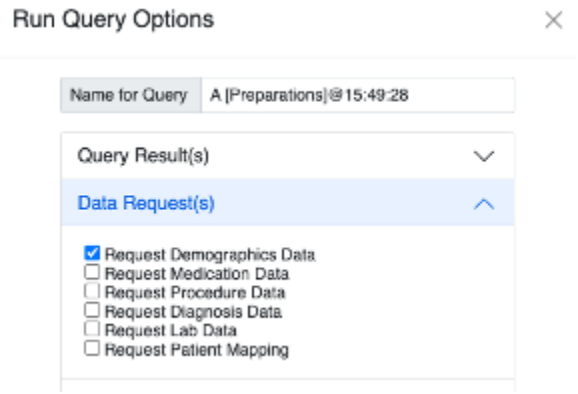


Data Export Workflow



Data Request (User Process)

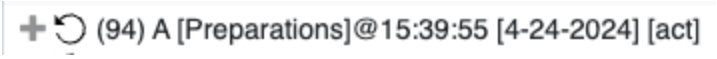
1. User logs into webclient and creates and runs a query.
2. In the Run Query dialog box, select one or more Data Request checkboxes



3. Emails are automatically sent - one to the requester's e-mail, to inform them the request has been made; and one to the specified manager user to inform them to run the user's export request. The query name (in previous queries) is tagged with the query master id, so it can be identified by the manager.

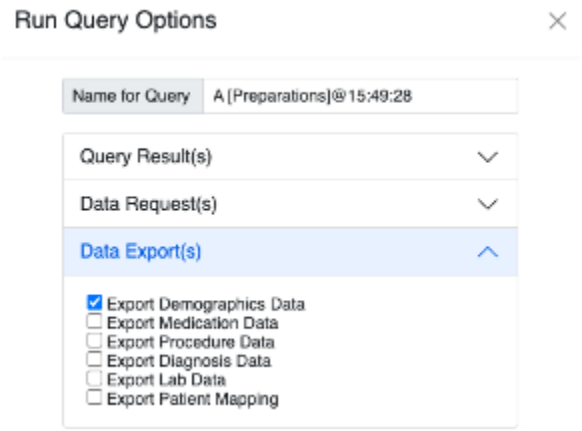
Data Export (Manager Process)

1. The manager logs into the webclient and finds the query from their e-mail by locating the query master id, which is now part of the title of the previous query.



2. The manager user then re-runs the previous query. In the Run Query dialog, the manager selects the requested export option in the Data export breakdown list.

   It is very important to put the original query master id (or some other identifier that links the export to the request) into the query name. The export letter will contain the query name, and this is the only way to trace the export back to the original request. (Currently there is no way to tie the export directory to the data requester query_master_id.) If the data manager runs the query as a query-in-query, the query name will automatically be populated with this identifier.



3. The data export runs in the background and the file is generated in a specified folder

4. An e-mail is sent to the manager that can be forwarded to the requesting user, to inform them that their export is available and to provide retrieval info.

Sample Letter generated for Data Manager

Results of the i2b2 request entitled - "Demographics_12_48_52", submitted on Apr 18, 2024 1:47:58 PM, are available.

Important notes about your data:
- Total number of patients returned in your data request: 133
- i2b2 reviewer:

Only persons specifically authorized and selected (as listed at the top of this letter) can download these files. If additional user access is needed, please ensure the person is listed on your project IRB protocol and contact the i2b2 team.

Specifically:
- Remove all PHI from computer, laptop, or mobile device after analysis is completed.
- Do NOT share PHI or PII with anyone who is not listed on the IRB protocol.

Your guideline for the storage of Protected Health Information can be found at:
https://www.site.com/guidelines_for_protecting_and_storing_phi.pdf

To download these files
- You must be logged onto your site

These results are the data that was requested under the authority of the Institutional Review Board. The query resulting in this identified patient data is included at the end of this letter. A copy of this letter is kept on file and is available to the IRB in the event of an audit.

Thank you,

The i2b2 Team

Example Export files (all data are fake)

Demographics

I2B2_PATIENT_NUMBER|BIRTH_DATE|DEATH_DATE|GENDER|AGE_IN_YEARS|PRIMARY_SPOKEN_LANGUAGE|RACE|MARTIAL_STATUS|RELIGION|ZIP_CODE|STATE_CITY_ZIP|INCOME|VITAL_STATUS
1000000023|1976-11-04 00:00:00||M|47|english|hispanic|married|jewish|02142|Zip codes\Massachusetts\Cambridge\02142\|Medium|
1000000025|1933-10-06 00:00:00||F|90|spanish|hispanic|widow|roman catholic|02026|Zip codes\Massachusetts\Dedham\02026\|Medium|
1000000128|1967-11-02 00:00:00||F|56|english|black|divorced|christian|02140|Zip codes\Massachusetts\Cambridge\02140\|High|

Diagnosis

```
I2B2_PATIENT_NUMBER|PATIENT_ID|START_DATE|END_DATE|DIAGNOSIS_NAME|DIAGNOSIS_CODE|MODIFIER|FACILITY|ENCOUNTER_TYPE|PROVIDER|ENCOUNTER_NUMBER
1000000128|3000001948|2005-08-26 00:00:00|2005-08-26 00:00:00|Tachycardia, unspecified|ICD9:785.0|1|@|Outpatient|Hansen, Sydney, MD|477825
1000000025|101809330|2005-05-05 00:00:00|2005-05-05 00:00:00|Other specified disorders of liver|ICD9:573.8|1|@|Inpatient|Kiley, Steven, MD|477561
1000000025|S500003075|2005-05-10 00:00:00|2005-05-10 00:00:00|Other specified disorders of liver|ICD9:573.8|2|@|Outpatient|Kiley, Steven, MD|477570
1000000066|1000000066|2008-09-08 00:00:00|2008-09-08 00:00:00|Unspecified disease of the jaws|ICD9:526.9|1|@|Outpatient|Casey, Ben, MD|481520
```

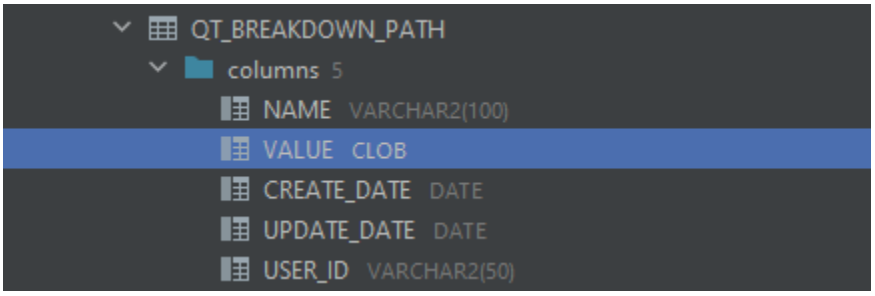
Data Export Configuration

Design and Architecture

The i2b2 breakdown architecture has been modified to support new breakdown types for data request and export. HIVE_CELL_PARAMS has new parameters for global configuration, the QT_RESULT_TYPE table has new entries for query export types, and the QT_BREAKDOWN_PATH defines the request and export. These are XML documents containing e-mails to send and SQL to execute. These breakdown types display in the run query dialog. User e-mails are sent to the e-mail address of the currently logged-in use. Data Manager e-mails are sent to an address configured in HIVE_CELL_PARAMS, and the data manager e-mail in an export is also saved as a file in the export location. The export file generation location, also configured in HIVE_CELL_PARAMS, is on the local drive hosting the i2b2 application.

Database changes

Currently QT_BREAKDOWN_PATH .value column has datatype of varchar(2000). In order to support the XML structure needed for the Data Exporter, the value column has been modified to **CLOB** datatype (ORACLE) and varchar(max) for (SQLServer).



Configuring Database Parameters

Request/Export query types: QT_QUERY_RESULT_TYPE

The data request and data export query types are configured as breakdowns in the QT_QUERY_RESULT_TYPE table. VISUAL_ATTRIBUTE_TYPE_ID should be set to **LR** for Data request and **LX** for Data export

USER_ROLE_CD should be set to **DATA_LDS** for Data Request and USER_ROLE_CD should be **MANAGER** for Data Export.

NAME	DESCRIPTION	DISPLAYTYPE_ID	VISUAL_ATTRIBUTE_TYPE_ID	USER_ROLE_CD
PATIENT_DEMOGRAPHIC_REQUEST	Request Demographics Data	CATNUM	LR	DATA_LDS
PATIENT_DEMOGRAPHIC_CSV	Export Demographics Data	CATNUM	LX	MANAGER

We recommend data requests be limited to DATA_LDS users (irrespective of Admin path role - admin, user, manager) and data exports be limited to MANAGER users (irrespective of data path role - data_lds, data_agg,data_obf, data_deid). The user role determines if exports/requests are visible and runnable.

QT_QUERY_RESULT_TYPE	**user role**	**data request**	**data export**
DATA_LDS	DATA_LDS	Success	non-display
MANAGER	MANAGER	non-display	Success
MANAGER with DATA_LDS	MANAGER		Success

MANAGER with DATA_LDS	DATA_LDS	Success	

Request/Export file definition details: QT_BREAKDOWN_PATH

QT_BREAKDOWN_PATH defines e-mail content (for request and export) and SQL code (for export) in the following XML format. Any parameters defined in an export XML override the default in HIVE_CELL_PARAMS (e.g., field separator character.)

Name	Value
• **_REQUEST** defines the request email generation parameters. • Request email can be configured in PM_USER_DATA parameters to define the email address of the User. • User email content configurable in />xml value parameter • Data Manager gets the data manager email (email address configured in the HIVE_CELL_PARAMS) • Data manager email content configurable in the /xml value parameter	[...version number...] 08/09/2024 12:00:00 [...details sent to User on data request...] [...details sent to data Manager on user data request...]

• **_CSV** defines the data export email parameters and the SQL to for the export file generation. • The data export email has detailed info on the file location and access info and can be customized according to your local institution's policies. • The exported file(s) are generated in a secure location on the server's file system along with the data export email file in a zipped folder • Email content and export SQL are configurable in the XML entries for the request/export in the Value parameters. • Multiple SQL statements can be configured in the Value tags as one SQL statement per file tag	[...version number...] details sent to data Manager on data request [... filename specification..] [... SQL query to generate the data export content...] [...file format seperator...]

Example Requests

NAME	VALUE

PATIENT_MEDICATION_CSV	3.02 08/09/2024 12:00:00 Your request on {{{PROJECT_ID}}} requested i2b2 request entitled - "{{{QUERY_NAME}}}", submitted on {{{QUERY_STARTDATE}}}, with the query master of {{{QUERY_ID}}} This user {{{USER_NAME}}} in project {{{PROJECT_ID}}} requested i2b2 request entitled - "{{{QUERY_NAME}}}", submitted on {{{QUERY_STARTDATE}}}, with the query master of {{{QUERY_ID}}}. /{{{USER_NAME}}}/{{{QUERY_MASTER_ID}}}/Readme.txt Results of the i2b2 request entitled - "{{{QUERY_NAME}}}", submitted on {{{QUERY_STARTDATE}}}, are available. Important notes about your data: - Total number of patients returned in your data request: {{{PATIENT_COUNT}}} - i2b2 reviewer: Only persons specifically authorized and selected (as listed at the top of this letter) can download these files. If additional user access is needed, please ensure the person is listed on your project IRB protocol and contact the i2b2 team. Specifically: - Remove all PHI from computer, laptop, or mobile device after analysis is completed. - Do NOT share PHI or PII with anyone who is not listed on the IRB protocol. Your guideline for the storage of Protected Health Information can be found at: https://www.site.com/guidelines_for_protecting_and_storing_phi.pdf *To download these files* - You must be logged onto your site These results are the data that was requested under the authority of the Institutional Review Board. The query resulting in this identified patient data is included at the end of this letter. A copy of this letter is kept on file and is available to the IRB in the event of an audit.

PATIENT_MEDICATION_REQUEST	3.02 08/09/2024 12:00:00 Your request on {{{PROJECT_ID}}} requested i2b2 request entitled - "{{{QUERY_NAME}}}", submitted on {{{QUERY_STARTDATE}}}, with the query master of {{{QUERY_MASTER_ID}}} This user {{{USER_NAME}}} in project {{{PROJECT_ID}}} requested i2b2 request entitled - "{{{QUERY_NAME}}}", submitted on {{{QUERY_STARTDATE}}}, with the query master of {{{QUERY_MASTER_ID}}}.

Export file format/location and Email parameters set-up: HIVE_CELL_PARAMS

Export File format, location and email server parameters are configurable in the HIVE_CELL_PARAMS.

The parameter in the QT_BREAKDOWN_PATH definition will override the default value specified in the HIVE_CELL_PARAMS.

Wildfly must be restarted for changes to the HIVE_CELL_PARAMs to take effect.

Cell ID	Parameter Name	Example Value
CRC	edu.harvard.i2b2.crc.exportcsv.datamanageremail	userid@partners.org
CRC	edu.harvard.i2b2.crc.smtp.host	smtp.partners.org
CRC	edu.harvard.i2b2.crc.smtp.port	25
CRC	edu.harvard.i2b2.crc.smtp.ssl.enabled	FALSE

CRC	edu.harvard.i2b2.crc.smtp.auth	FALSE
CRC	edu.harvard.i2b2.crc.smtp.username	none
CRC	edu.harvard.i2b2.crc.smtp.password	none
CRC	edu.harvard.i2b2.crc.smtp.enabled	FALSE
CRC	edu.harvard.i2b2.crc.smtp.from.fullname	Data Manager
CRC	edu.harvard.i2b2.crc.smtp.from.email	datamanager@site.org
CC	edu.harvard.i2b2.crc.smtp.subject	i2b2 Data Request
CRC	edu.harvard.i2b2.crc.exportcsv.defaultescapecharacter	"

CRC	edu.harvard.i2b2.crc.exportcsv.maxfetchrows	-1
CRC	edu.harvard.i2b2.crc.exportcsv.defaultlineend	\n
CRC	edu.harvard.i2b2.crc.exportcsv.defaultseperator	\t
CRC	edu.harvard.i2b2.crc.exportcsv.resultfetchsize	50000
CRC	edu.harvard.i2b2.crc.exportcsv.filename	{{{PROJECT_ID}}}/{{{DATE_yyyyMMdd}}}_{{{FULL
CRC	edu.harvard.i2b2.crc.exportcsv.defaultquotechar	"
CRC	edu.harvard.i2b2.crc.exportcsv.workfolder	/tmp/i2b2

CRC	edu.harvard.i2b2.crc.exportcsv.zipencryptmethod	none

Export query definition variables

These variables can be used in HIVE_CELL_PARAMS entries and in query export definitions:

Value	Description
{{{USER_NAME}}}	Current user running query
{{{PROJECT_ID}}}	The project name
{{{RANDOM_xxx}}}	A random integer where xxx is the max integer size
{{{DATE_xxx}}}	Date/Time format, where xxx can be from the table below. The string yyyy-MM-dd would output dates that look like 2022-07-26
{{{QUERY_NAME}}}	Query Name
{{{QUERY_STARTDATE}}}	Query Start date/time
{{{QUERY_ENDDATE}}}	Query End date/time
{{{PATIENT_COUNT}}}	Number of patients.
{{{FULL_NAME}}}	Fullname of user who ran query
{{{QUERY_MASTER_ID}}}	The query Master ID
{{{QUERY_RUNTIME}}}	Time to run the query

`{{{FULL_SCHEMA }}}`	Used in SQL export definitions. Inserts the full schema name (e.g., 'i2b2.dbo.'). Needed to be sure i2b2 accesses the correct data tables!

Table of allowable characters in {{{DATE_xxx}}} strings:

Value	Description	Value	Description
G	era	K	hour-of-am-pm (0-11)
u	year	k	clock-hour-of-am-pm (1-24)
yyyy	4-digit year	H	hour-of-day (0-23)
D	day-of-year	m	minute-of-hour
M/L	month-of-year	s	second-of-minute
d	day-of-month	S	fraction-of-second
Q/q	quarter-of-year	A	milli-of-day
Y	week-based-year	n	nano-of-second
w	week-of-week-based-year	N	nano-of-day
W	week-of-month	V	time-zone ID
E	day-of-week	z	time-zone name
e/c	localized day-of-week	O	localized zone-offset
F	week-of-month	X	zone-offset 'Z' for zero
a	am-pm-of-day	x	zone-offset
h	clock-hour-of-am-pm (1-12)	Z	zone-offset
		p	pad next

Request and Export Results:

Metadata about data request and export queries are stored in the QT_XML_RESULT table, in the XML_VALUE field:

Query Type	XML_VALUE example
Data Request	http://www.i2b2.org/xsd/cell/crc/psm/querydefinition/1.1/" xmlns:ns5=" http://www.i2b2.org/xsd/hive/msg/1.1/" xmlns:ns8="http://www.i2b2.org/xsd/cell/pm/1.1/" xmlns:ns7="http://www.i2b2.org/xsd/cell/crc/psm/analysisdefinition/1.1/" xmlns:ns9=" http://www.i2b2.org/xsd/cell/ont/1.1/" xmlns:ns10="http://www.i2b2.org/xsd/hive/msg/result/1.1/" xmlns:ns2="http://www.i2b2.org/xsd/hive/pdo/1.1/" xmlns:ns4=" http://www.i2b2.org/xsd/cell/crc/psm/1.1/" xmlns:ns3="http://www.i2b2.org/xsd/cell/crc/pdo/1.1/"> 1 This user act in project NODE9 requested i2b2 request entitled - "ACT_Vaccination_14_53_14", submitted on May 22, 2024 2:53:13 PM, with the query master of 1492. datamanager@site.org

Data Export	http://www.i2b2.org/xsd/cell/crc/psm/querydefinition/1.1/" xmlns:ns5=" http://www.i2b2.org/xsd/hive/msg/1.1/" xmlns:ns8="http://www.i2b2.org/xsd/cell/pm/1.1/" xmlns:ns7="http://www.i2b2.org/xsd/cell/crc/psm/analysisdefinition/1.1/" xmlns:ns9=" http://www.i2b2.org/xsd/cell/ont/1.1/" xmlns:ns10="http://www.i2b2.org/xsd/hive/msg/result/1.1/" xmlns:ns2="http://www.i2b2.org/xsd/hive/pdo/1.1/" xmlns:ns4=" http://www.i2b2.org/xsd/cell/crc/psm/1.1/" xmlns:ns3="http://www.i2b2.org/xsd/cell/crc/pdo/1.1/"> __ ____ _____3 _____Environment Settings:---------- _ Machine name: localhost Operating system: Linux 3.10.0-1160.99.1.el7.x86_64 Login user: wildfly Version number: [6.3.2.1] Config title: 2. PRODUCTION Run directory: /opt/dataexport/act/1491/Lab.csv File Name: /opt/dataexport/act/1491/Lab.csv md5 hash: d41d8cd98f00b204e9800998ecf8427e _____1491 ____ __

Software Changes:

- Data: New entries in HIVE_CELL_PARAMS, QT_QUERY_RESULT_TYPE, and QT_BREAKDOWN_PATH define the exporter configuration.
- Java code: New breakdown classes and updates to existing java classes to support the data exporter functionality.