How to add/upload custom data using the "spreadsheet" upload function

David A Gutman MD PHD
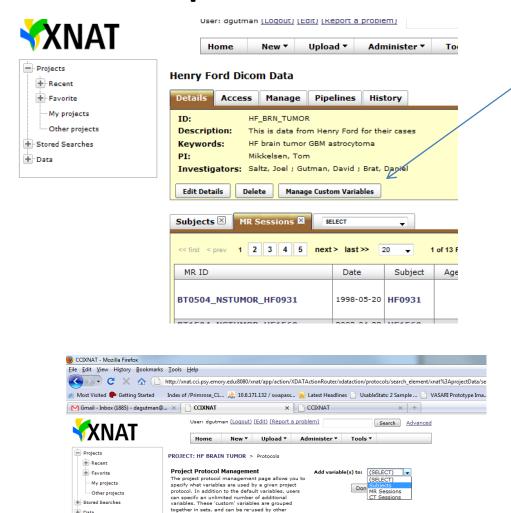
General Goal

- Want to be able to quickly add/update "custom" fields in XNAT using the spreadsheet interface--- in certain situations the complexity of deciphering XML may be daunting for certain scenarios

Use Case: Want to add a custom column to schema, view it in "main" data view, and automatically upload data in CSV format

Specifics: Want to add a column called TCGA_ID to my database..my current Subject_ID for my study also overlaps with another ongoing larger study, and want to be able to add the "other" study ID to my schema so I can search for it, also two other fields will be added in this example

Step #1: Add custom variable
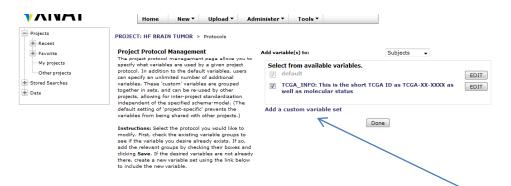
After selecting a given project that you want to add the custom set of variables to, click manage custom variables

A drop down box now comes up on the new menu--- in my case I want to ADD the variable "TCGA_ID" which provides add'l info for a given subject, so I select "ADD VARIABLES TO→SUBJECT"

So I want to add a custom variable set...

Which then opens another box, so I already created one called TCGA_INFO (you can also edit an older set if you want...)
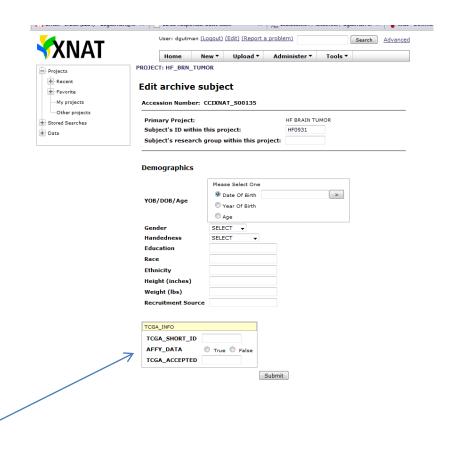
So now you add your custom variables…
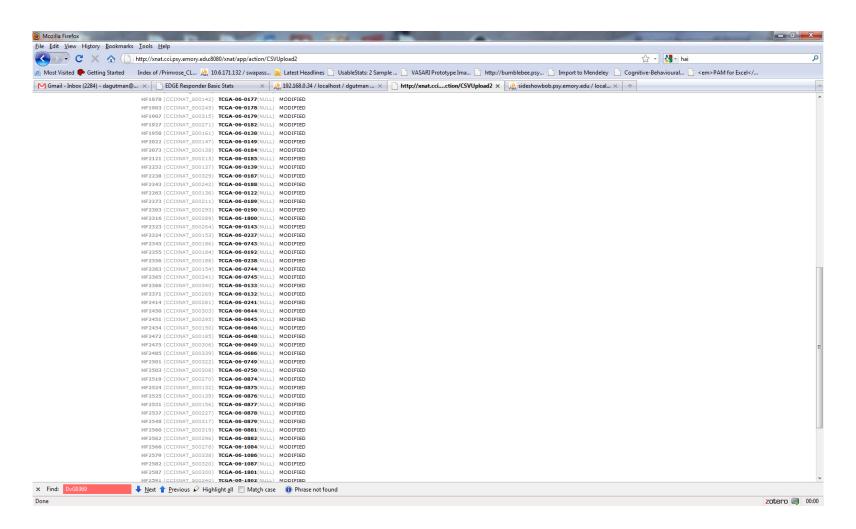
So this is specific to whatever you want to add… I added 3 variables related
To my specific project– also you should probably pick specific data types–
XNAT uses this data to validate your entries so that you don't accidentally put
characters in a field that's supposed to contain numbers…

Edit Variable Set

Name: TCGA_INFO   ☑ Project Specific?

Description: This is the short TCGA ID as
TCGA-XX-XXXX as well as
molecular status

Custom Variables
Please list the variables to be included in this set. (Variable names should be a one
word description.)

Name	Type
TCGA_SHORT_ID	string
AFFY_DATA	boolean
TCGA_ACCEPTED	string

Add Variable

Save   Cancel

Add a custom variable set

Done
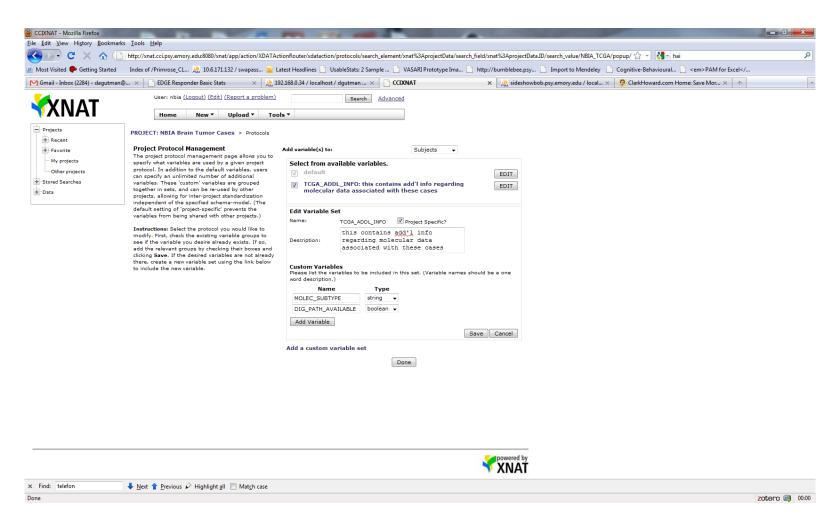
Note: Now that this variable set is added.. It shows up when I go edit a specific subject

Since I already have 200 subjects, I do NOT want to manually enter the new fields for each subject...
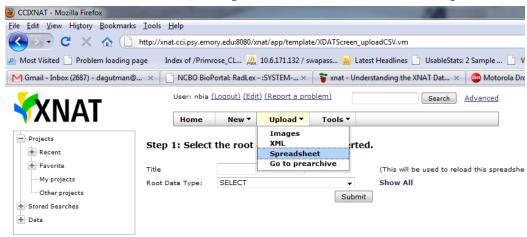
Added two more fields.. Logged on as user NBIA...

Now what..
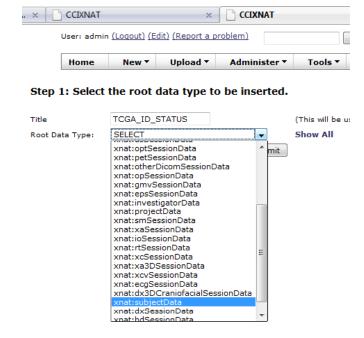
- My projects now has 3 new custom fields available for upload… I want to be able to upload data for each patient to populate these fields… and I want to go low tech and use the upload spreadsheet option

Upload→Spreadsheet

I am going to create a template where I can upload "CSV" type data..the fields I want to upload are associated with a SUBJECT..

So I created a "title" and then under root data type selected xnat:subjectData
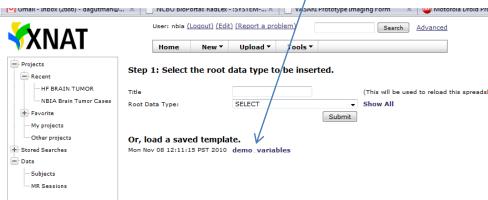
So added a subject descriptor

- Could also add add'l data , but I only cared about the subject_ID.. THIS however now however shows up as a

  - "SAVED TEMPLATE"... unfortunately I can't
  - automatically add custom variables, so I now need to manually edit my template to make it useful

So now what..

- Requires a bit of hacking... but it's fairly straightforward
- Basically the "template" is.. Not surprisingly, an XML document...
- So to add these custom fields to a template you need to
- → Find (and edit) the template XML document
- → Figure out what the "name" of the new custom fields you've just added and insert them into the XML template

SO... I had to find the XML document

- They seem to live in the /home/xnat/xnat_data/cache directory
- The specific directory depends on which userid added the xml template...
- So:

```
xnat@ubuntu:~/xnat_data/cache$ ls -l */csv/*
-rw-r--r-- 1 xnat xnat  181 2010-11-08 12:11 10/csv/1289247075450.xml
-rw-r--r-- 1 xnat xnat  247 2010-09-13 12:47 2/csv/1283260737762.xml
-rw-r--r-- 1 xnat xnat  182 2010-08-31 07:07 2/csv/1283263715322.xml
-rw-r--r-- 1 xnat xnat  218 2010-06-02 13:03 4/csv/1275508974084.xml
-rw-r--r-- 1 xnat xnat  179 2010-06-02 13:20 4/csv/1275509955166.xml
-rw-r--r-- 1 xnat xnat  144 2010-06-08 05:28 4/csv/1276000110319.xml
-rw-r--r-- 1 xnat xnat  254 2010-09-14 07:44 4/csv/1283263764729.xml
-rw-r--r-- 1 xnat xnat 1213 2010-10-23 11:24 4/csv/1287857541187.xml
-rw-r--r-- 1 xnat xnat  405 2010-10-27 08:02 4/csv/1288192346741.xml
xnat@ubuntu:~/xnat_data/cache$ █
```

Connected to xnat.cci.psy.emory.edu

- chdir /home/xnat/xnat_data/cache
- ls –l */csv/*
- Find the xml document that was most recently created... and then do
- cd 4/csv/
- Then do nano/gedit/vm/whatever the long file is
- Like "gedit 1288192346741.xml" or alternatively just do
- nano 4/csv/1288192346741.xml
- SINCE I ONLY ADDED ONE FIELD IN MY TEMPLATE.. IT LOOKS UNINTERESTINGly like below
- Just has <field>xnat:SubjectData/ID</field>.....

```
Access documents, folders and network places
xnat@ubuntu: ~/xnat_data/cache/2/csv
File  Edit  View  Terminal  Help
  GNU nano 2.0.9                    File: 1283260737762.xml

<FieldMapping data-type="xnat:subjectData" create-date="Tue Aug 31 06:18:57 PDT 2010" title="TCGA_ID_STATUS" ID="1283260737762">
        <field>xnat:subjectData/ID</field>
</FieldMapping>
```

Figuring out what my new custom fields are called

- Download the XML document..
- Or click on "View XML" under the options..

Not the most fun thing in the world to read..
But for low-tech solution I just opened it up
In WORDPAD

Basically I just looked for the name of the variable I just added….

<xnat:field name="CORE_sample" type="custom" datatype="string" required="0" xmlPath="xnat:subjectData/fields/field[name=core_sample]/field"/>
<xnat:field name="TUMORTISSUESITE" type="custom" datatype="string" required="0" xmlPath="xnat:subjectData/fields/field[name=tumortissuesite]/field"/>
<xnat:field name="VITALSTATUS" type="custom" datatype="string" required="0" xmlPath="xnat:subjectData/fields/field[name=vitalstatus]/field"/>
<xnat:field name="KarnScore" type="custom" datatype="string" required="0" xmlPath="xnat:subjectData/fields/field[name=karnscore]/field"/>

THEN I basically cut/and pasted the name of the custom variables into my XML file.. So I now have 4 fields that I can upload using my template

The page at http://xnat.cci.psy.emory.edu:8080 says:

<?xml version="1.0" encoding="UTF-8"?>
<xdat:bundle ID="HF_BRN_TUMOR_xnat:subjectData" allow-diff-columns="0" secure="0" brief-description="Subjects" tag="HF_BRN_TUMOR" xmlns:cat="http://nrg.wustl.edu/catalog" xmlns:xnat_a="http://nrg.wustl.edu/xnat_assessments" xmlns:xnat="http://nrg.wustl.edu/xnat" xmlns:arc="http://nrg.wustl.edu/arc" xmlns:pipe="http://nrg.wustl.edu/pipe" xmlns:prov="http://www.nbirn.net/prov" xmlns:wrk="http://nrg.wustl.edu/workflow" xmlns:xdat="http://nrg.wustl.edu/security" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://nrg.wustl.edu/security http://xnat.cci.psy.emory.edu:8080/xnat/schemas/security/security.xsd http://nrg.wustl.edu/catalog http://xnat.cci.psy.emory.edu:8080/xnat/schemas/catalog/catalog.xsd http://nrg.wustl.edu/pipe http://xnat.cci.psy.emory.edu:8080/xnat/schemas/pipeline/repository.xsd http://nrg.wustl.edu/xnat http://xnat.cci.psy.emory.edu:8080/xnat/schemas/xnat/xnat.xsd http://www.nbirn.net/prov http://xnat.cci.psy.emory.edu:8080/xnat/schemas/birn/birnprov.xsd http://nrg.wustl.edu/xnat_assessments http://xnat.cci.psy.emory.edu:8080/xnat/schemas/assessments/assessments.xsd http://nrg.wustl.edu/arc http://xnat.cci.psy.emory.edu:8080/xnat/schemas/project/project.xsd http://nrg.wustl.edu/workflow http://xnat.cci.psy.emory.edu:8080/xnat/schemas/pipeline/workflow.xsd">
<xdat:root_element_name>xnat:subjectData</xdat:root_element_name>
<xdat:search_field>
<!--hidden_fields[xdat_search_field_id="null",xdat_stored_search_id="null"]-->
<xdat:element_name>xnat:subjectData</xdat:element_name>
<xdat:field_ID>SUB_PROJECT_IDENTIFIER=HF_BRN_TUMOR</xdat:field_ID>
<xdat:sequence>0</xdat:sequence>
<xdat:type>string</xdat:type>
<xdat:header>Subject</xdat:header>
<xdat:value>HF_BRN_TUMOR</xdat:value>
</xdat:search_field>
<xdat:search_field>
<!--hidden_fields[xdat_search_field_id="null",xdat_stored_search_id="null"]-->
<xdat:element_name>xnat:subjectData</xdat:element_name>
<xdat:field_ID>GENDER_TEXT</xdat:field_ID>
<xdat:sequence>1</xdat:sequence>
<xdat:type>string</xdat:type>
<xdat:header>M/F</xdat:header>
</xdat:search_field>
<xdat:search_field>
<!--hidden_fields[xdat_search_field_id="null",xdat_stored_search_id="null"]-->
<xdat:element_name>xnat:subjectData</xdat:element_name>
<xdat:field_ID>HANDEDNESS_TEXT</xdat:field_ID>
<xdat:sequence>2</xdat:sequence>
<xdat:type>string</xdat:type>
<xdat:header>Hand</xdat:header>
</xdat:search_field>
<xdat:search_field>
<!--hidden_fields[xdat_search_field_id="null",xdat_stored_search_id="null"]-->
<xdat:element_name>xnat:subjectData</xdat:element_name>
<xdat:field_ID>DOB</xdat:field_ID>
<xdat:sequence>3</xdat:sequence>
<xdat:type>UNKNOWN</xdat:type>
<xdat:header>DOB</xdat:header>
</xdat:search_field>
<xdat:search_field>
<!--hidden_fields[xdat_search_field_id="null",xdat_stored_search_id="null"]-->
<xdat:element_name>xnat:subjectData</xdat:element_name>
<xdat:field_ID>XNAT_SUBJECTDATA_FIELD_MAP=tcga_short_id</xdat:field_ID>
<xdat:sequence>4</xdat:sequence>
<xdat:type>string</xdat:type>
<xdat:header>TCGA_SHORT_ID</xdat:header>
<xdat:value>tcga_short_id</xdat:value>
</xdat:search_field>
<xdat:search_field>
<!--hidden_fields[xdat_search_field_id="null",xdat_stored_search_id="null"]-->
<xdat:element_name>xnat:subjectData</xdat:element_name>
<xdat:field_ID>XNAT_SUBJECTDATA_FIELD_MAP=affy_data</xdat:field_ID>
<xdat:sequence>5</xdat:sequence>
<xdat:type>boolean</xdat:type>
<xdat:header>AFFY_DATA</xdat:header>
<xdat:value>affy_data</xdat:value>
</xdat:search_field>
<xdat:search_field>
<!--hidden_fields[xdat_search_field_id="null",xdat_stored_search_id="null"]-->
<xdat:element_name>xnat:subjectData</xdat:element_name>
<xdat:field_ID>SUB_EXPT_COUNT=xnat:mrSessionData</xdat:field_ID>
<xdat:sequence>6</xdat:sequence>
<xdat:type>integer</xdat:type>
<xdat:header>MR Sessions</xdat:header>
<xdat:value>xnat:mrSessionData</xdat:value>
</xdat:search_field>
<xdat:search_where method="AND">

OK

```
GNU nano 2.0.9                    File: 4/csv/1288192346741.xml

<FieldMapping data-type="xnat:subjectData" create-date="Wed Oct 27 08:12:26 PDT 2010" title="tcga_surg_info" ID="1288192346741">
        <field>xnat:subjectData/ID</field>
<field>xnat:subjectData/fields/field[name=bcr_surgery_barcode]/field</field>
<field>xnat:subjectData/fields/field[name=days_to_procedure]/field</field>
<field>xnat:subjectData/fields/field[name=procedure_type]/field</field>
</FieldMapping>
```

So now what….

1) Click on the template.. And now (see left vs right) instead of just having xnat:subjectData/ID   THREE new add'l fields exist… so I can either type them in manually and "Add row" (which would be painful)

2) OR!! (see next page)

Step 3: Enter the data

After entering your data (via upload or manual entry), you will need to click the 'Save' button at the bottom of the page to complete the process.

Option 1: Upload the data

File [_____] [Browse...] Download Template

[Upload]

Option 2: Enter the data manually.

[_____]   xnat:subjectData/ID
[_____]   xnat:subjectData/fields/field[name=bcr_surgery_barcode]/field
[_____]   xnat:subjectData/fields/field[name=days_to_procedure]/field
[_____]   xnat:subjectData/fields/field[name=procedure_type]/field

[Add row]

Project (SELECT) ▼
[Save] [Cancel]

Create a spreadsheet with the columns.. IN THE SAME EXACT ORDER as on
The previous page… and do SAVE AS → CSV

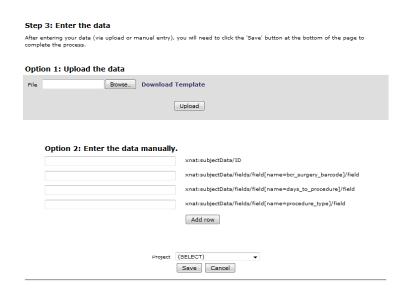|	A	B	C	
	A1		f_x	bcr_patient_barcode
	A	B	C	
1	bcr_patient_barcode	bcr_surgery_barcode	days_to_procedure	procedure_type
2	TCGA-02-0001	TCGA-02-0001-S1	0	Surgical Resection
3	TCGA-02-0003	TCGA-02-0003-S1	0	Surgical Resection
4	TCGA-02-0004	TCGA-02-0004-S1	0	Surgical Resection
5	TCGA-02-0006	TCGA-02-0006-S2	0	Surgical Resection
6	TCGA-02-0006	TCGA-02-0006-S1	0	Surgical Resection
7	TCGA-02-0007	TCGA-02-0007-S1	0	Surgical Resection
8	TCGA-02-0009	TCGA-02-0009-S2	0	Surgical Resection
9	TCGA-02-0009	TCGA-02-0009-S1	0	Surgical Resection
10	TCGA-02-0010	TCGA-02-0010-S1	0	Surgical Resection
11	TCGA-02-0011	TCGA-02-0011-S1	0	Surgical Resection
12	TCGA-02-0014	TCGA-02-0014-S1	0	Surgical Resection
13	TCGA-02-0015	TCGA-02-0015-S2	null	Secondary Surgery for tu
14	TCGA-02-0015	TCGA-02-0015-S1	0	Surgical Resection
15	TCGA-02-0016	TCGA-02-0016-S1	0	Surgical Resection
16	TCGA-02-0021	TCGA-02-0021-S1	0	Surgical Resection
17	TCGA-02-0023	TCGA-02-0023-S2	null	Secondary Surgery for tu
18	TCGA-02-0023	TCGA-02-0023-S1	0	Surgical Resection
19	TCGA-02-0024	TCGA-02-0024-S1	0	Surgical Resection
20	TCGA-02-0025	TCGA-02-0025-S1	0	Surgical Resection
21	TCGA-02-0025	TCGA-02-0025-S2	null	Secondary Surgery for tu
22	TCGA-02-0026	TCGA-02-0026-S1	0	Surgical Resection

Then will upload... MOST also select project you are uploading too

And voila…
can review/remove/ whatever data..
But SUBMIT button is at the bottom

x://xnat.cci.psy.emory.edu:8080/xnat/app/action/CSVUpload2

Index of /Primrose_CL...   10.6.171.132 / swapass...   Latest Headlines   UsableStats: 2 Sample ...   VASARI Prototype Ima...   http://bumblebee.psy...   In

Motorola Droid Pro preorders start t... × | CCIXNAT   × | CCIXNAT   ×

xnat:subjectData/ID

xnat:subjectData/fields/field[name=bcr_surgery_barcode]/field

xnat:subjectData/fields/field[name=days_to_procedure]/field

xnat:subjectData/fields/field[name=procedure_type]/field

Add row

	COLUMN 0	COLUMN 1	COLUMN 2	COLUMN 3	
0	bcr_patient_barcode	bcr_surgery_barcode	days_to_procedure	procedure_type	REMOVE
1	TCGA-02-0001	TCGA-02-0001-S1	0	Surgical Resection	REMOVE
2	TCGA-02-0003	TCGA-02-0003-S1	0	Surgical Resection	REMOVE
3	TCGA-02-0004	TCGA-02-0004-S1	0	Surgical Resection	REMOVE
4	TCGA-02-0006	TCGA-02-0006-S2	0	Surgical Resection	REMOVE
5	TCGA-02-0006	TCGA-02-0006-S1	0	Surgical Resection	REMOVE
6	TCGA-02-0007	TCGA-02-0007-S1	0	Surgical Resection	REMOVE
7	TCGA-02-0009	TCGA-02-0009-S2	0	Surgical Resection	REMOVE
8	TCGA-02-0009	TCGA-02-0009-S1	0	Surgical Resection	REMOVE
9	TCGA-02-0010	TCGA-02-0010-S1	0	Surgical Resection	REMOVE
10	TCGA-02-0011	TCGA-02-0011-S1	0	Surgical Resection	REMOVE
11	TCGA-02-0014	TCGA-02-0014-S1	0	Surgical Resection	REMOVE
12	TCGA-02-0015	TCGA-02-0015-S2	null	Secondary Surgery for tumor recurrence/progression: Locoregional procedure	REMOVE
13	TCGA-02-0015	TCGA-02-0015-S1	0	Surgical Resection	REMOVE
14	TCGA-02-0016	TCGA-02-0016-S1	0	Surgical Resection	REMOVE
15	TCGA-02-0021	TCGA-02-0021-S1	0	Surgical Resection	REMOVE
16	TCGA-02-0023	TCGA-02-0023-S2	null	Secondary Surgery for tumor recurrence/progression: Locoregional procedure	REMOVE
17	TCGA-02-0023	TCGA-02-0023-S1	0	Surgical Resection	REMOVE
18	TCGA-02-0024	TCGA-02-0024-S1	0	Surgical Resection	REMOVE
19	TCGA-02-0025	TCGA-02-0025-S1	0	Surgical Resection	REMOVE
20	TCGA-02-0025	TCGA-02-0025-S2	null	Secondary Surgery for tumor recurrence/progression: Locoregional procedure	REMOVE
21	TCGA-02-0026	TCGA-02-0026-S1	0	Surgical Resection	REMOVE
22	TCGA-02-0027	TCGA-02-0027-S1	0	Surgical Resection	REMOVE
23	TCGA-02-0028	TCGA-02-0028-S1	0	Surgical Resection	REMOVE
24	TCGA-02-0033	TCGA-02-0033-S1	0	Surgical Resection	REMOVE
25	TCGA-02-0034	TCGA-02-0034-S1	0	Surgical Resection	REMOVE