IMPORTING, EXPORTING, AND UPDATING VIVO DATA



GOALS

- Complement existing data ingest tools
- Make it easier for both new and existing VIVO sites to manipulate their data



SOME DATA SOURCES AT DARTMOUTH

- HR database feed (Official data from HR)
- Grants (Local source, database view)
- Organizations, Departments (Local sources)
- Publications, Journals (PubMed API)
- Finding Publications (Profiles Disambiguation Engine API)



Our data sources all provide us with spreadsheet style data. Generally it is all two dimensional rows and columns and comes as a CSV file or a spreadsheet which can easily be saved in CSV format.



The challenge is to get from here:

```
"uid","NetId","PersonalId","DNDUID","Name_First","Name_Middle","Name
"d3891w9","d3891w9","68785","77283","Tracy","L","Onega","PhD","Tracy
```



to the triples...



```
<${namespace}/${col01}-vcard> <http://purl.obolibrary.org/obo/ARG_2000029> <http://vivo.dartmouth.edu/individual/${col01}> .
<${namespace}/${col01}-vcard> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2006/vcard/ns#Individual> .
<${namespace}/${col01}-vcard> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2006/vcard/ns#Kind> .
<${namespace}/${col01}-vcard> <http://www.w3.org/2006/vcard/ns#hasEmail> <${namespace}/${col01}-email> .
<${namespace}/${col01}-vcard> <http://www.w3.org/2006/vcard/ns#hasName> <${namespace}/${col01}-name> .
<${namespace}/${col01}-vcard> <http://www.w3.org/2006/vcard/ns#hasTitle> <${namespace}/${col01}-title> .
<${namespace}/${col01}-email> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2006/vcard/ns#Email> .
<${namespace}/${col01}-email> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2006/vcard/ns#Work> .
<${namespace}/${col01}-email> <http://www.w3.org/2006/vcard/ns#email> "${col09}"^^<http://www.w3.org/2001/XMLSchema#string> .
<${namespace}/${col01}-title> <http://www.w3.org/2006/vcard/ns#title> "${col11}"^^<http://www.w3.org/2001/XMLSchema#string> .
<${namespace}/${col01}-name> <http://vivoweb.org/ontology/core#middleName> "${col06}"^^<http://www.w3.org/2001/XMLSchema#string> .
<${namespace}/${col01}-name> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2006/vcard/ns#Name> .
<${namespace}/${col01}-name> <http://www.w3.org/2006/vcard/ns#familyName> "${col07}"^^<http://www.w3.org/2001/XMLSchema#string> .
<${namespace}/${col01}-name> <http://www.w3.org/2006/vcard/ns#givenName> "${col05}"^^<http://www.w3.org/2001/XMLSchema#string> .
<${namespace}/${col01}> <http://purl.obolibrary.org/obo/ARG_2000028> <${namespace}/${col01}-vcard> .
<${namespace}/${col01}> <http://vivo.dartmouth.edu/ontology/degrees> "${col08}" .
<${namespace}/${col01}> <http://vivo.dartmouth.edu/ontology/dndUid> "${col04}"^^<http://www.w3.org/2001/XMLSchema#int> .
<${namespace}/${col01}> <http://vivo.dartmouth.edu/ontology/netId> "${col02}" .
<${namespace}/${col01}> <http://vivo.dartmouth.edu/ontology/personalId> "${col03}"^^<http://www.w3.org/2001/XMLSchema#int> .
<${namespace}/${col01}> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://vivoweb.org/ontology/core#FacultyMember> .
<${namespace}/${col01}> <http://www.w3.org/2000/01/rdf-schema#label> "${col10}"^^<http://www.w3.org/2001/XMLSchema#string> .
```



File Edit View Search Terminal Help

PAST

- At Dartmouth we first created many of these tools in Python and Django.
- We have used these as command line tools and as a Django App made to appear to be inside our VIVO.



Account



Search

Organizations People Research Home

Events

Verify Publications for Tracy Onega

7 pending articles have been found.

Characteristics associated with requests by pathologists for second opinions on breast biopsies.

Published: 2017

Journal: Journal of clinical pathology

Authors: Geller BM, Nelson HD, Weaver DL, Frederick PD, Allison KH, Onega T, Carney PA, Tosteson

ANA, Elmore JG

Pubmed ID: 28465449

Claim as mine

Reject (Not mine)

Skip (Not sure)

Manually add PubMed article

Verify PubMed ID:

Return to Profile Page



PRESENT

- We have moved and extended most of the tools from Python into Java and integrated into VIVO 1.9.2.
- If you are interested in testing these tools, please contact: rodney.jacobson@dartmouth.edu



FUTURE

 We plan to release these tools on GitHub by the end of the month.



VIVO CSV TOOLS

- Train new file formats
- Export data from VIVO
- Import data into VIVO
- Publication Finder



Home

People

Organizations

Research Events

Capability Map

Site Administration

Data Input

Faculty Member (vivo)

Add individual of this class

Site Configuration

Institutional internal class

Manage profile editing

Page management

Menu ordering

Site information

User accounts

Ontology Editor

Ontology list

Class Management

Class hierarchy

Class groups

Property Management

Object property hierarchy

Data property hierarchy

Faux Property Listing

Property groups

Advanced Data Tools

Add/Remove RDF data

Ingest tools

RDF export

SPARQL query

CSV Data Tools

Import CSV Data **Export CSV Data** Training



TRAINING FOR NEW FILE FORMATS



The tool can be trained for a new file type by manually entering a sample row of data.

- No knowledge of the ontology is needed
- Sparql queries are created automatically



Home People Organizations Research Events Capability Map

CSV Data Tools - Training

Record







SYNERGY
Climical and Translational Sciencer Institute

Home People Organizations Research Events Capability Map

Site Administration

Data Input

Academic Department (vivo)

Add individual of this class

Site Configuration

Institutional internal class

Manage profile editing

Page management

Menu ordering

Site information

User accounts

Ontology Editor

Advanced Data Tools



Home People Organizations Research Events Capability Map

Create a new Academic Department

Name *

test22

Create Academic Department

or <u>Cancel</u>

* required fields



Tutorials Support Site Admin root Search

Research Events Capability Map People Organizations Home

CSV Data Tools - Training	
Recorded Triples:	
col01	
col02 test22	
col03 555-1212	
col04 deptemail@example.com	
http://vivoloc/vivo/individual/n10389	
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2006/vcard/ns#Email
http://www.w3.org/2006/vcard/ns#email	deptemail@example.com^^http://www.w3.org/2001/XMLSchema#string
\${namespace}/org{col01}	
http://vivoloc/vivo/individual/n21678	
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://vivoweb.org/ontology/core#AcademicDepartment
http://www.w3.org/2000/01/rdf-schema#label	test22^^http://www.w3.org/2001/XMLSchema#string
http://purl.obolibrary.org/obo/ARG_2000028	http://vivoloc/vivo/individual/n665
http://purl.obolibrary.org/obo/ARG_2000028	http://vivoloc/vivo/individual/n665
http://vivoloc/vivo/individual/n5464	
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2006/vcard/ns#Telephone
http://www.w3.org/2006/vcard/ns#telephone	555-1212^^http://www.w3.org/2001/XMLSchema#string
http://vivoloc/vivo/individual/n665	
http://www.w3.org/2006/vcard/ns#hasTelephone	http://vivoloc/vivo/individual/n5464
http://purl.obolibrary.org/obo/ARG_2000029	http://vivoloc/vivo/individual/n21678
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2006/vcard/ns#Individual

http://www.w3.org/1999/02/22-rdf-syntax-ns#type http://www.w3.org/2006/vcard/ns#Individual

http://vivoloc/vivo/individual/n10389

http://vivoloc/vivo/individual/n21678



http://www.w3.org/2006/vcard/ns#hasEmail

http://purl.obolibrary.org/obo/ARG_2000029

Clear Save

The system examines the triples that were created by manually entering the row of data, and creates sparql templates for selecting and inserting rows of this data.



```
<${namespace}/${col01}-vcard> <http://purl.obolibrary.org/obo/ARG_2000029> <http://vivo.dartmouth.edu/individual/${col01}> .
<${namespace}/${col01}-vcard> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2006/vcard/ns#Individual> .
<${namespace}/${col01}-vcard> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2006/vcard/ns#Kind> .
<${namespace}/${col01}-vcard> <http://www.w3.org/2006/vcard/ns#hasEmail> <${namespace}/${col01}-email> .
<${namespace}/${col01}-vcard> <http://www.w3.org/2006/vcard/ns#hasName> <${namespace}/${col01}-name> .
<${namespace}/${col01}-vcard> <http://www.w3.org/2006/vcard/ns#hasTitle> <${namespace}/${col01}-title> .
<${namespace}/${col01}-email> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2006/vcard/ns#Email> .
<${namespace}/${col01}-email> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2006/vcard/ns#Work> .
<${namespace}/${col01}-email> <http://www.w3.org/2006/vcard/ns#email> "${col09}"^^<http://www.w3.org/2001/XMLSchema#string> .
<${namespace}/${col01}-title> <http://www.w3.org/2006/vcard/ns#title> "${col11}"^^<http://www.w3.org/2001/XMLSchema#string> .
<${namespace}/${col01}-name> <http://vivoweb.org/ontology/core#middleName> "${col06}"^^<http://www.w3.org/2001/XMLSchema#string> .
<${namespace}/${col01}-name> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2006/vcard/ns#Name> .
<${namespace}/${col01}-name> <http://www.w3.org/2006/vcard/ns#familyName> "${col07}"^^<http://www.w3.org/2001/XMLSchema#string> .
<${namespace}/${col01}-name> <http://www.w3.org/2006/vcard/ns#givenName> "${col05}"^^<http://www.w3.org/2001/XMLSchema#string> .
<${namespace}/${col01}> <http://purl.obolibrary.org/obo/ARG_2000028> <${namespace}/${col01}-vcard> .
<${namespace}/${col01}> <http://vivo.dartmouth.edu/ontology/degrees> "${col08}" .
<${namespace}/${col01}> <http://vivo.dartmouth.edu/ontology/dndUid> "${col04}"^^<http://www.w3.org/2001/XMLSchema#int> .
<${namespace}/${col01}> <http://vivo.dartmouth.edu/ontology/netId> "${col02}" .
<${namespace}/${col01}> <http://vivo.dartmouth.edu/ontology/personalId> "${col03}"^^<http://www.w3.org/2001/XMLSchema#int> .
<${namespace}/${col01}> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://vivoweb.org/ontology/core#FacultyMember> .
<${namespace}/${col01}> <http://www.w3.org/2000/01/rdf-schema#label> "${col10}"^^<http://www.w3.org/2001/XMLSchema#string> .
```



File Edit View Search Terminal Help

FILE CONFIGURATIONS CAN BE SHARED

Configurations supplied with the tools include:

- Faculty HR data
- Organizations
- Publications (and authorships)
- Grants (and roles)



EXPORT DATA FROM VIVO

 For each file format that has been configured, there are urls to download both CSV and JSON data.



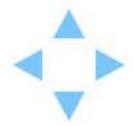
Home People Organizations Research Events Capability Map

CSV Data Tools - Export

authors <u>CSV JSON</u>
departments <u>CSV JSON</u>
faculty <u>CSV JSON</u>
articles <u>CSV JSON</u>



```
🖰 vivoloc:8080/vivo/dat ×
← → C ① vivoloc:8080/vivo/dataTools?action=export&what=departments&format=json
{"departments": [
         "DeptID": "1001",
         "Name": "Medicine"
         "DeptID": "1011",
         "Name": "Psychiatry"
         "DeptID": "2222",
         "Name": "Data Science"
         "DeptID": "1044",
         "Name": " Community and Family Medicine"
         "DeptID": "1002",
         "Name": "Pharmacology and Toxicology"
```



Q 🕁 🦁 :

```
vivoloc:8080/vivo/dataTools?action=export&what=faculty&format=json - Chromium
□ vivoloc:8080/vivo/dat ×
← → C ① vivoloc:8080/vivo/dataTools?action=export&what=faculty&format=json
                                                                                      Q 🕁 🦁
{"faculty": [
      "PersonalId":
      "Name First":
      "uid": "
      "Name Middle":
      "Name Last": "
      "NetId": "
      "DNDUID": """
      "Email":
      "Degree": "PhD"
      "FullName": "
      "PrefTitle": "Professor of Microbiology and Immunology"
      "PersonalId":
      "Name First":
      "Name Middle": 🖜
      "Name Last": "
      "NetId": "IIII
      "DNDUID":
      "Email": "
      "Degree": "PhD",
      "FullName":
      "PersonalId": ""
      "Name First": "
      "uid": "
      "Name Middle":
```



CSV Data Tools - Export

authors <u>CSV JSON</u>
departments <u>CSV JSON</u>
faculty <u>CSV JSON</u>
articles <u>CSV JSON</u>



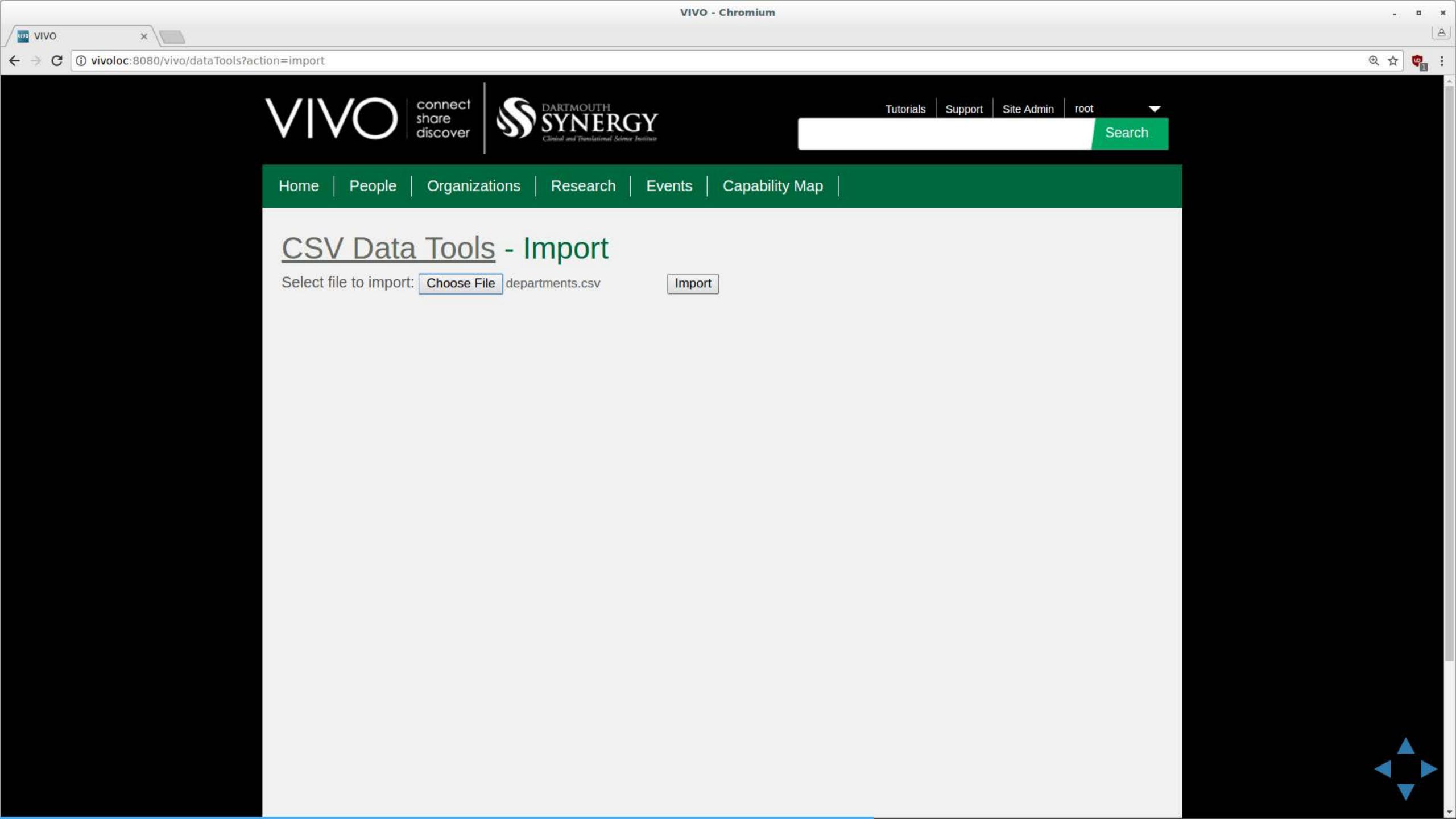
```
e Edit View Search Terminal Help
```

```
"DeptID", "Name"
"1001", "Medicine", "http://vivoloc/vivo/individual/org1001"
"1011", "Psychiatry", "http://vivoloc/vivo/individual/org1011"
"2222", "Data Science", "http://vivoloc/vivo/individual/org2222"
"1044", " Community and Family Medicine", "http://vivoloc/vivo/individual/org1044"
"1002", "Pharmacology and Toxicology", "http://vivoloc/vivo/individual/org1002"
```



IMPORT CSV DATA INTO VIVO

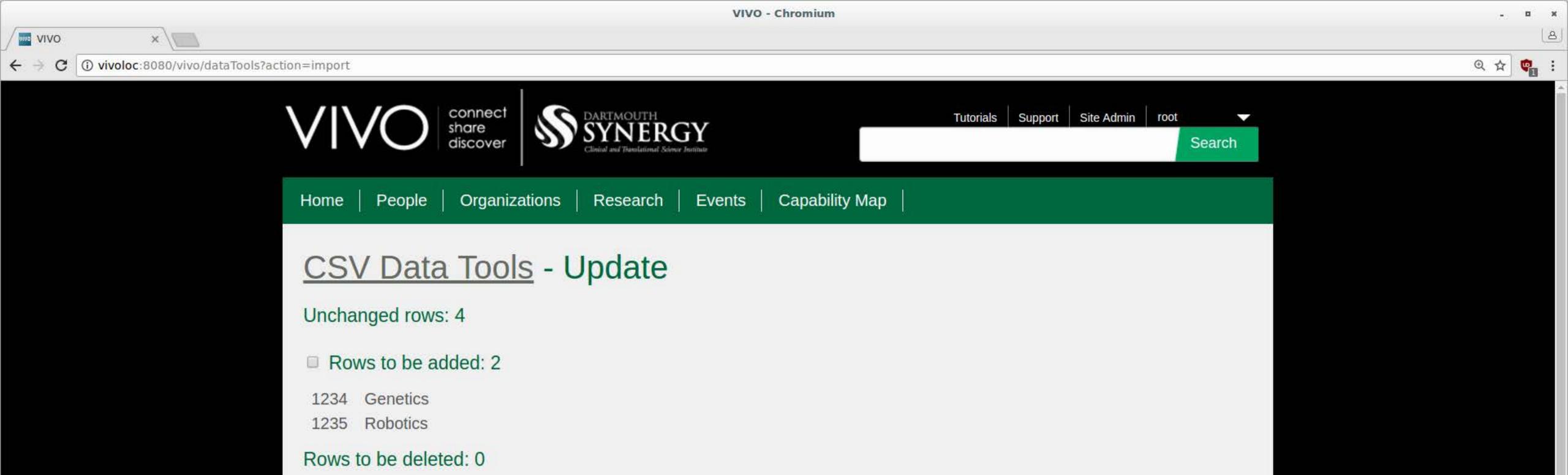




Imported data is compared to the data currently in VIVO to see what has been added, deleted, or updated.

The administrator can choose which actions to perform.



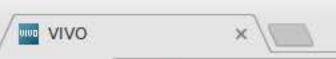


Rows to be updated: 1

Perform Selected Actions

Biomedical Data Science http://vivoloc/vivo/individual/org2222

Search



(i) vivoloc:8080/vivo/dataTools?action=import



Capability Map Organizations Research Events People

CSV Data Tools - Update

Unchanged rows: 4

Rows to be added: 2

Genetics

Robotics 1235

Rows to be deleted: 0

Rows to be updated: 1

Biomedical Data Science http://vivoloc/vivo/individual/org2222

old: http://www.w3.org/2000/01/rdf-schema#label>"Data">nttp://vivoloc/vivo/individual/org2222 http://www.w3.org/2000/01/rdf-schema#label "Data"

Science"^^...

new: http://www.w3.org/2000/01/rdf-schema#label">new: http://www.w3.org/2000/01/rdf-schema#label "Biomedical Data

Science"^^.

Perform Selected Actions



Q & Q

PUBLICATION FINDER



The Profiles RNS Author Disambiguation Engine

http://profiles.catalyst.harvard.edu/



The Profiles RNS Author Disambiguation Engine We send information about the author including:

- Name
- Email
- Institution

•••



and importantly:

- Known publications
- Excluded publications

Providing PubMed IDs for known publications, and excluded publications is extremely valuable to get accurate results.



PUBMED API

Given the IDs of a set of suggested publication, we can query the PubMed API to get information about the publications.

The information we get back for a publication can be treated as a row of CSV data.



BRINGING THIS DATA INTO VIVO

Two provided configurations can import this data into VIVO.

One for the articles:



```
□ vivoloc:8080/vivo/dat × \
← → C ① vivoloc:8080/vivo/dataTools?action=export&what=articles&format=json
                                                                                                                               ⊕ ☆ 😇 :
{"articles": [
         "ENDPAGE": "78",
         "ISSUE": "6",
         "YEAR": "",
         "PMC": "PMC4198523",
         "VOLUME": "24",
         "MID": "NIHMS595146",
         "TITLE": "",
         "STARTPAGE": "365",
         "PMID": "24733560",
         "DOI": "10.1002/rmv.1791"
         "ENDPAGE": "33",
         "ISSUE": "8",
         "YEAR": "",
         "PMC": "PMC4551341",
         "VOLUME": "7",
         "MID": ""
         "TITLE": "Haemophilus influenzae responds to glucocorticoids used in asthma therapy by modulation of biofilm
formation and antibiotic resistance.",
         "STARTPAGE": "1018",
         "PMID": "25995336",
         "DOI": "10.15252/emmm.201505088"
    },
         "ENDPAGE": "8",
         "ISSUE": "11",
         "YEAR": "",
```

And one for the authorships:



```
🖰 vivoloc:8080/vivo/dat ×
← → C ① vivoloc:8080/vivo/dataTools?action=export&what=authors&format=json
{"authors": [
      "NetId":
      "PMID": "
      "NetId":
      "PMID": "
      "NetId": "IIII"
      "PMID": "
      "NetId":
      "PMID": "
      "NetId":
      "PMID": "
      "NetId":
      "PMID": "
      "NetId":
"PMID":
      "NetId":
```



Q &

0

We display the publications for the author to confirm or reject:



Search

People

Organizations

Research Events Capability Map

Verify Publications

122 pending articles have been found:

Claim as mine Reject (Not mine) Skip (Not sure)

Oral Anticoagulation for Stroke Prevention in Canadian Practice: Stroke Prevention and Rhythm Interventions in Atrial Fibrillation (SPRINT-AF) Registry(.).

Published in: The Canadian journal of cardiology

PMID: 26195228 Date: 2016 Feb Authors: Ha AC, Singh N, Cox JL, Mancini GB, Dorian P, Fournier C, Gladstone DJ, Lockwood E, Shuaib A, Kajil M, Tsigoulis M, Gupta MK

Claim as mine
 Reject (Not mine)
 Skip (Not sure)

Cardiac monitoring for detection of atrial fibrillation after TIA: A systematic review and meta-analysis.

Published in: International journal of stroke: official journal of the International Stroke Society

PMID: 27681890 Date: 2017 Jan Authors: Korompoki E, Del Giudice A, Hillmann S, Malzahn U, Gladstone DJ, Heuschmann P, Veltkamp R

Claim as mine Reject (Not mine) Skip (Not sure)

Unattended Hospital and Home Sleep Apnea Testing Following Cerebrovascular Events.

Published in: Journal of stroke and cerebrovascular diseases: the official journal of National Stroke Association

PMID: 27717683 Date: 2017 Jan Authors: Boulos MI, Elias S, Wan A, Im J, Frankul F, Atalla M, Black SE, Basile VS, Sundaram A, Hopyan JJ, Boyle K, Gladstone DJ, Swartz RH, Murray BJ

Claim as mine Reject (Not mine) Skip (Not sure)

Quantitative analysis of brass compensators for commissioning of the Pinnacle planning system for IMRT.

Published in: Journal of applied clinical medical physics

PMID: 28297548 Date: 2015 Nov Authors: Gates LL, Gladstone DJ

Claim as mine Reject (Not mine) Skip (Not sure)

Screening for Atrial Fibrillation: A Report of the AF-SCREEN International Collaboration.

Published in: Circulation



Published in: IEEE transactions on medical imaging

PMID: 28644800 Date: 2017 Jun 20 Authors: Bruza P, Andreozzi JM, Gladstone DJ, Jarvis LA, Rottmann J, Pogue BW

Claim as mine
 Reject (Not mine)
 Skip (Not sure)

Use of Evidence-Based Therapy for Cardiovascular Risk Factors in Canadian Outpatients With Atrial Fibrillation: From the Facilitating Review and Education to Optimize Stroke Prevention in Atrial Fibrillation (FREEDOM AF) and Co-ordinated National Network to Engage Physicians in the Care and Treatment of Patients With Atrial Fibrillation (CONNECT AF).

Published in: The American journal of cardiology

PMID: 28666577 Date: 2017 Aug 15 Authors: Silberberg A, Tan MK, Yan AT, Angaran P, Dorian P, Bucci C, Gregoire JC, Bell AD, Gladstone DJ, Green MS, Gross PL, Skanes A, Demchuk AM, Kerr CR, Mitchell LB, Cox JL, Talajic M, Essebag V, Heilbron B, Ramanathan K, Fournier C, Wheeler BH, Lin PJ, Berall M, Langer A, Goldin L, Goodman SG, FREEDOM AF and CONNECT AF Investigators.

Claim as mine Reject (Not mine) Skip (Not sure)

Atrial premature beats predict atrial fibrillation in cryptogenic stroke: results from the EMBRACE trial.

Published in: Stroke

PMID: 25700289 Date: 2015 Apr Authors: Gladstone DJ, Dorian P, Spring M, Panzov V, Mamdani M, Healey JS, Thorpe KE, EMBRACE Steering Committee and Investigators.

Claim as mine Reject (Not mine) Skip (Not sure)

Cherenkov-excited luminescence scanned imaging.

Published in: Optics letters

PMID: 25723443 Date: 2015 Mar 1 Authors: Zhang R, D'souza AV, Gunn JR, Esipova TV, Vinogradov SA, Glaser AK, Jarvis LA, Gladstone DJ, Pogue BW

Claim as mine
 Reject (Not mine)
 Skip (Not sure)

How to Monitor Patients Receiving Direct Oral Anticoagulants for Stroke Prevention in Atrial Fibrillation: A Practice Tool Endorsed by Thrombosis Canada, the Canadian Stroke Consortium, the Canadian Cardiovascular Pharmacists Network, and the Canadian Cardiovascular Society.

Published in: Annals of internal medicine

PMID: 26121536 Date: 2015 Sep 1 Authors: Gladstone DJ, Geerts WH, Douketis J, Ivers N, Healey JS, Leblanc K

Claim All



We run a new query of the Profiles Disambiguation Engine when the author:

- Confirms all of the suggested publications
- or rejects one or more publications



EXPORT DATA WITH FILTERS



```
vivoloc:8080/vivo/dataTools?action=export&what=faculty&format=json&Degree=MD - Chromium
□ vivoloc:8080/vivo/dat × \
← → C ① vivoloc:8080/vivo/dataTools?action=export&what=faculty&format=json&Degree=MD
                                                                                                      ⊕ ☆ 😇
{"faculty": [
       "PersonalId": "
       "Name First": "L
       "uid":
       "Name Middle":
       "Name Last":
       "NetId": "
       "DNDUID":
       "Email": "",
       "Degree": "MD",
       "FullName": "
       "PrefTitle": "Associate Professor of Medicine"
       "PersonalId": "
       "Name First": "
       "uid": "
       "Name Middle":
       "Name Last": "
       "NetId": "
       "DNDUID": "
       "Email": ""
       "Degree": "MD",
       "PrefTitle": "Professor of Surgery"
       "PersonalId": "
       "Name First": "
       "uid": "
       "Name Middle":
```

```
← → C ① vivoloc:8080/vivo/dataTools?action=export&what=faculty&format=json&Name_First=Tracy
{"faculty": [
       "PersonalId": """,
       "Name First": "Tracy",
       "uid": """,
       "Name Middle": "",
       "Name Last": " ,
       "NetId": "
       "DNDUID": """,
       "Email": ""
       "Degree": "PhD".
       "FullName": "Tracy, PhD",
       "PrefTitle": "Adjunct Assistant Professor of Psychiatry"
       "PersonalId": """,
       "Name First": "Tracy",
       "uid": """",
       "Name Middle": ".".
       "Name Last": ",
       "DNDUID": "".
       "Email": "Trans.",
       "Degree": "PhD",
       "FullName": "Tracy L, PhD",
       "PrefTitle": "Associate Professor of Biomedical Data Science"
```

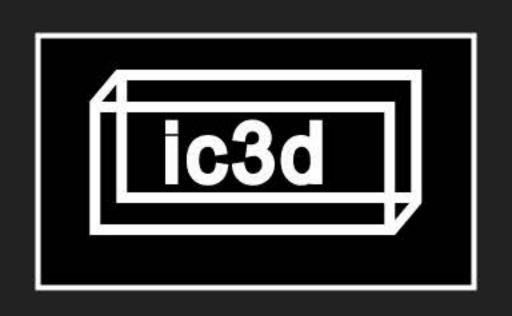
🖰 vivoloc:8080/vivo/dat x

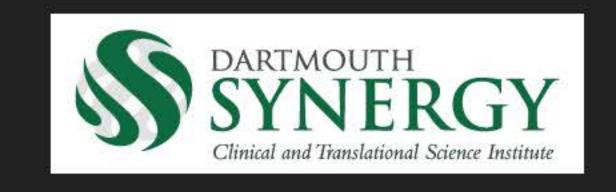


⊕ ☆ 😇 :

Rodney.Jacobson@dartmouth.edu Sukdith.Punjasthitkul@dartmouth.edu







ic3d.dartmouth.edu

