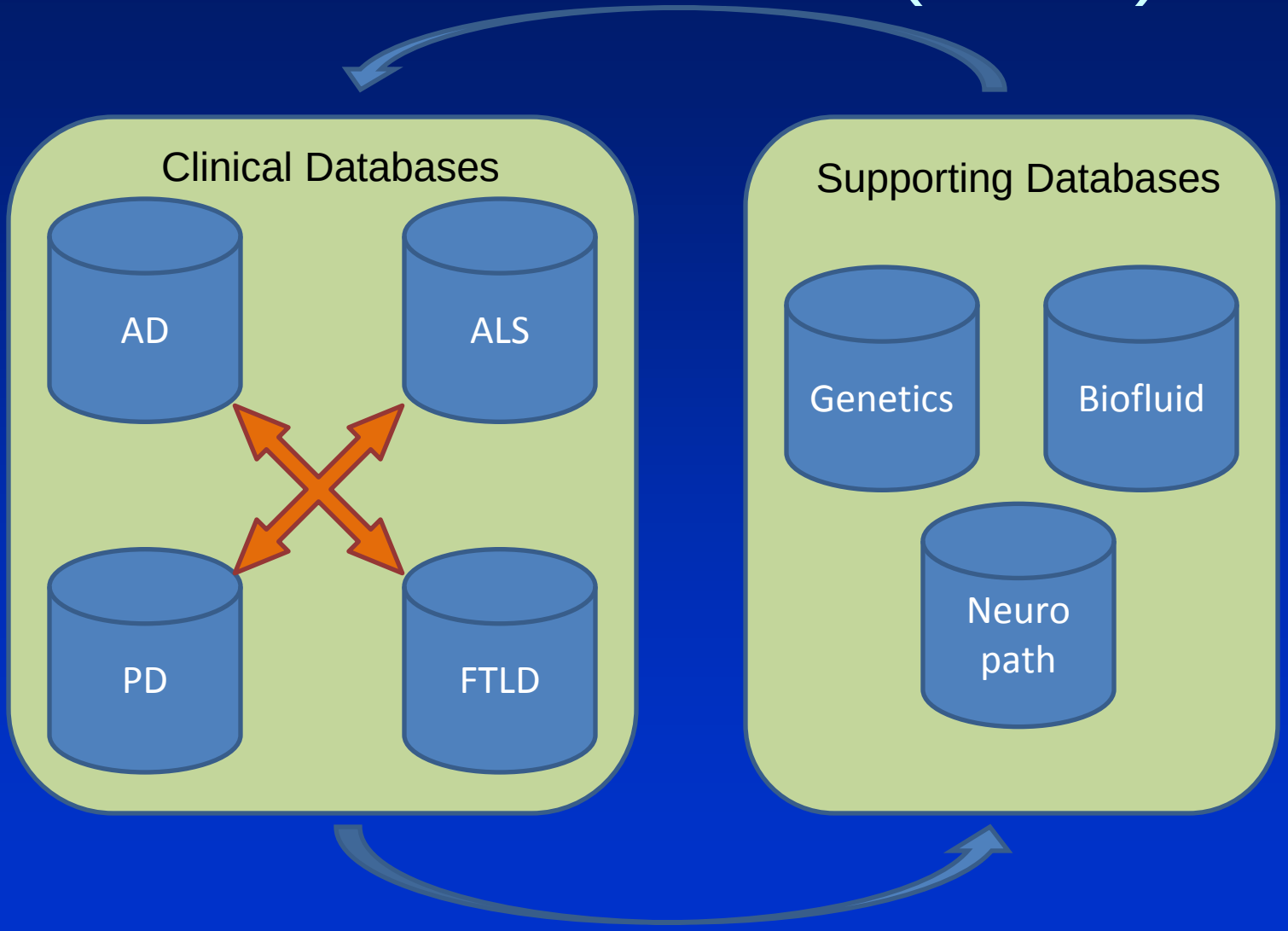
Penn Integrated NeuroDegenerative Disease Database (INDD)

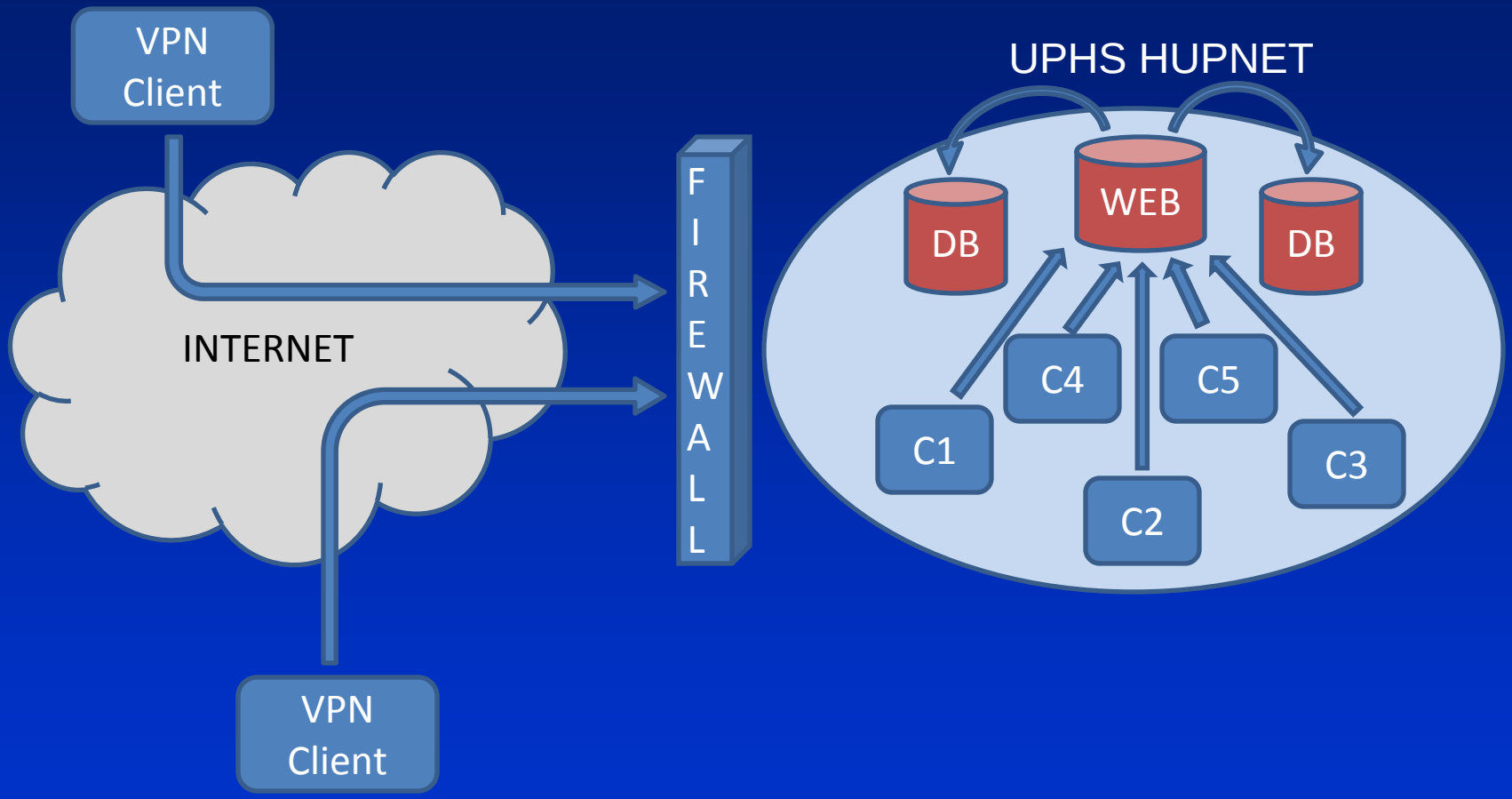
INQuery - Search and You Will Find

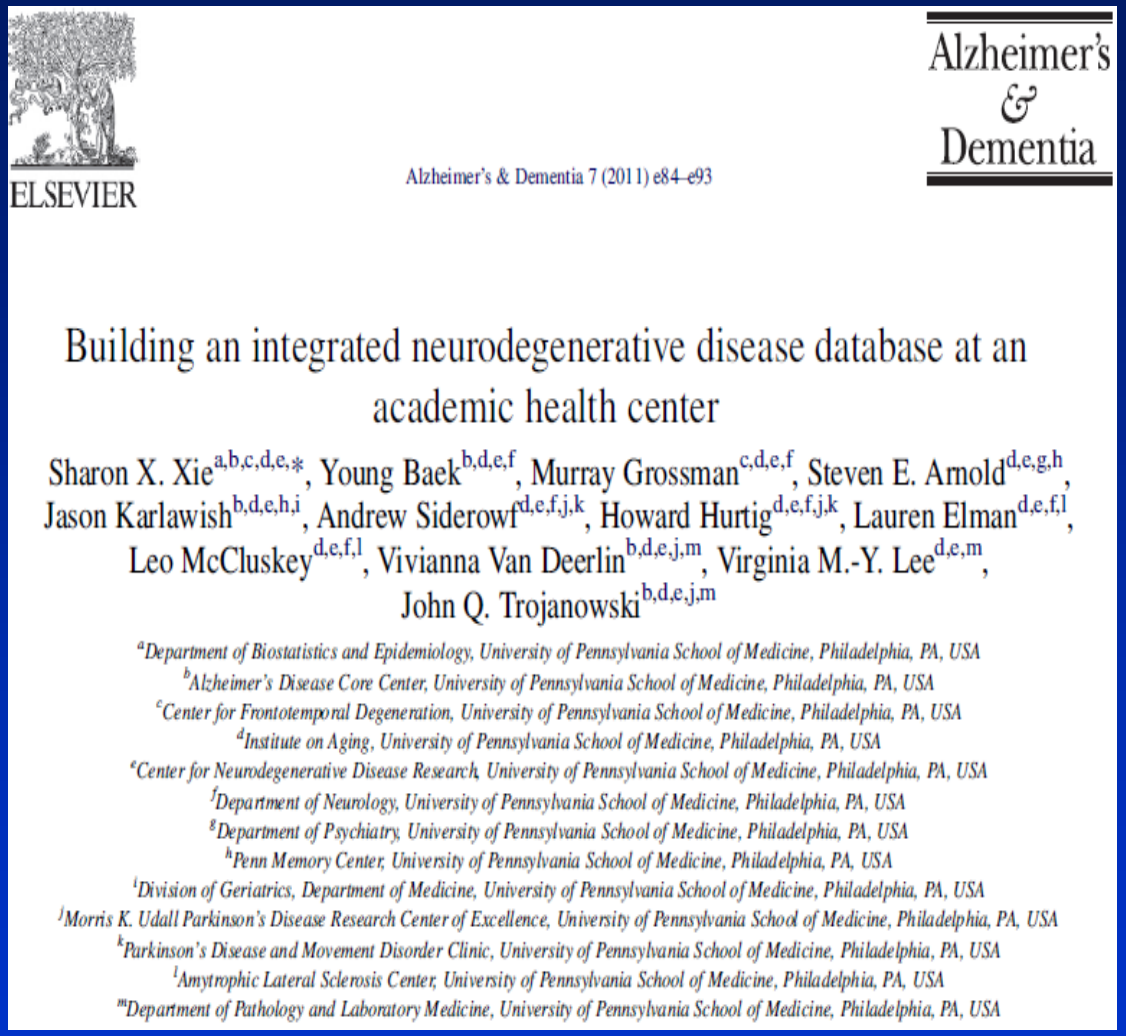
Rui Tong, Data Manager
University of Pennsylvania Alzheimer's Disease Core Center

Integrated Neurodegenerative Disease Database (INDD)



Integrated Neurodegenerative Disease Database (INDD)





ELSEVIER

Alzheimer's & Dementia 7 (2011) e84–e93

Alzheimer's & Dementia

Building an integrated neurodegenerative disease database at an academic health center

Sharon X. Xie[a,b,c,d,e,*], Young Baek[b,d,e,f], Murray Grossman[c,d,e,f], Steven E. Arnold[d,e,g,h], Jason Karlawish[b,d,e,h,i], Andrew Siderowf[d,e,f,j,k], Howard Hurtig[d,e,f,j,k], Lauren Elman[d,e,f,l], Leo McCluskey[d,e,f,l], Vivianna Van Deerlin[b,d,e,j,m], Virginia M.-Y. Lee[d,e,m], John Q. Trojanowski[b,d,e,j,m]

[a]*Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[b]*Alzheimer's Disease Core Center, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[c]*Center for Frontotemporal Degeneration, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[d]*Institute on Aging, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[e]*Center for Neurodegenerative Disease Research, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[f]*Department of Neurology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[g]*Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[h]*Penn Memory Center, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[i]*Division of Geriatrics, Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[j]*Morris K. Udall Parkinson's Disease Research Center of Excellence, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[k]*Parkinson's Disease and Movement Disorder Clinic, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[l]*Amytrophic Lateral Sclerosis Center, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[m]*Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*

INDD Database

- Relational Database
- Microsoft SQL Server 2008 R2 with ASP.NET MVC 3
- Contains >= 300 Tables
- Data from >= 17000 Research Subjects
- Millions of Data Records

INQuery - Search and you will find

Web-based database query tool, enabling end-users to query and extract data without the need of database administrators

Objectives

- Web-based: Universal access
- Easy to use with intuitive interface
- Includes advanced features to filter, merge, sort data

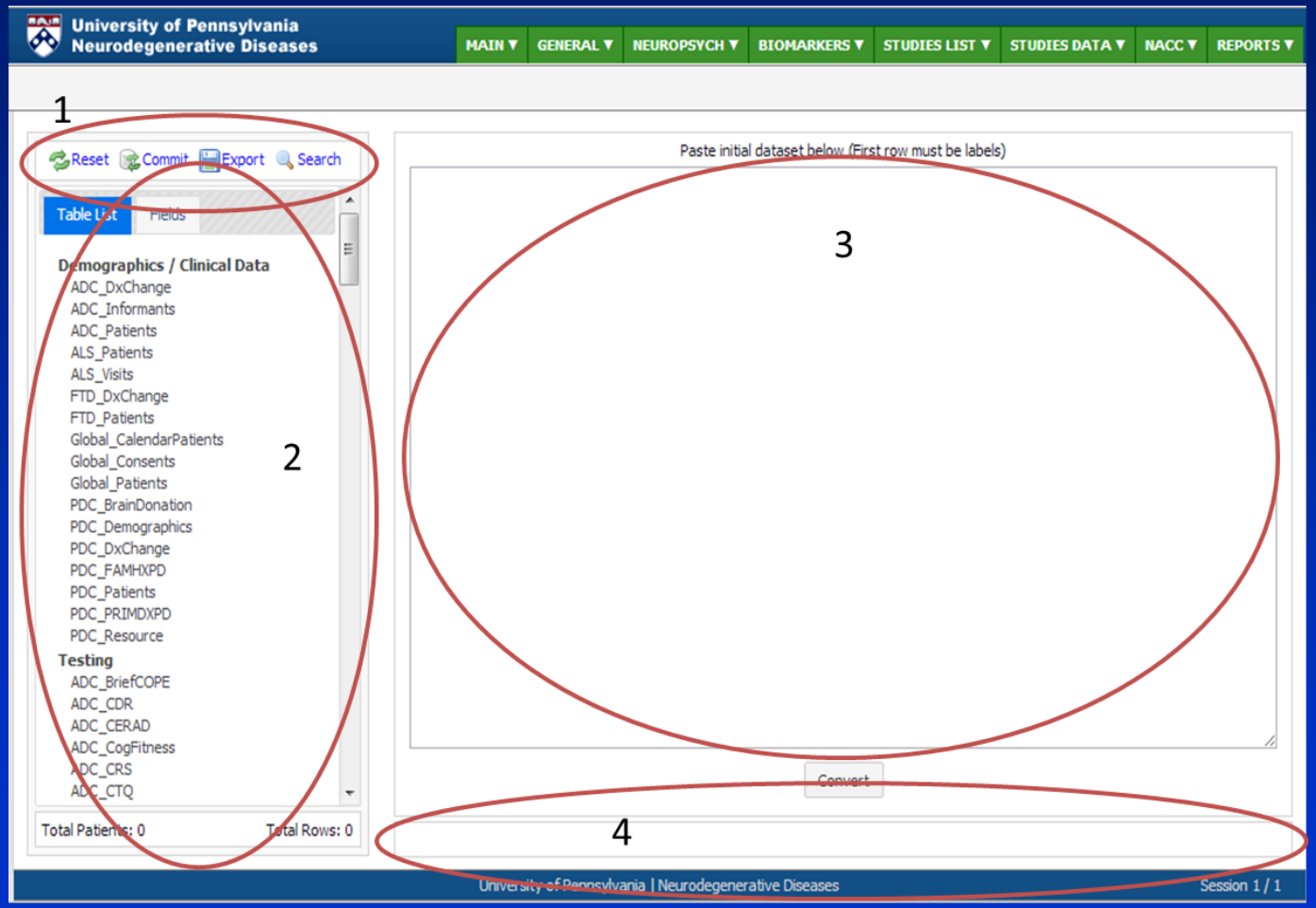
Methods

- Using stages to perform queries
- Filtering, sorting between stages
- Merge tables based on specific date ranges
- Sort by earliest or latest visit / test dates

INQuery - Terminology

INDD	Integrated NeuroDegenerative Disease (Database)
INQuery	INDD Querying System
INDDID	Unique identifier for individual patients in the INDD Database System
Dataset	2 dimentional table layed out in columns and rows (active / working)
Table	Collection of records stored in the database (database)
Record	A single row within a dataset, table
Field	Name of columns represented in a table
Data types	Int, float, decimal, string (varchar), bit (boolean), date
Primary key	A field or fields in a table that can uniquely identify all records in a table
Filter	Ability to remove records based on specified criteria
Logical [AND]	Combination of two or more filters operating together. Both must be true in order to result in true statement
Logical [OR]	Combination of two or more filters operating independently of each other. Either one of them can be true to result in true statement
Merge / Join	Process of combining dataset with data from a table
Left outer join	Process of leaving the current dataset post-merge regardless of a matching record is present in the joining table

INQuery - Layout



INQuery – Generating Queries

1. Start out with a question you want to ask the database.
2. Import the initial dataset if you have one.
3. Add a table from the table list to the working dataset.
4. Use field filters / field checkboxes to narrow down to desired dataset.
5. Add additional tables and merge them to the current active dataset.
6. Further refine the active dataset using filters / checkboxes.
7. Repeat Steps 5, 6.

INQuery – Live Demo

- Male Patients from ADC
- Diagnosis of AD or MCI
- MMSE of >= 20 at their latest visit
- CSF
- APOE Status

INQuery – Data Security

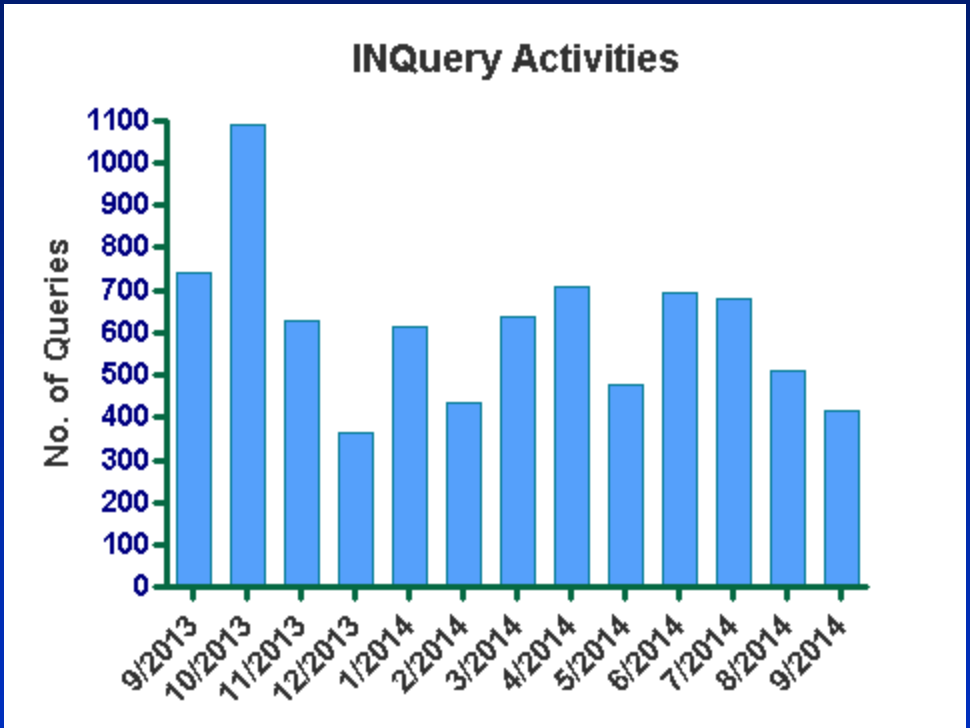
Data Security

- Enforced by levels of access rights
- Different users have access to different sets of tables
- We can choose which table or field to be published to which group.
- Logical tables are published using Views to separate data by their sensitivity

Views

- Views are virtual tables.
- Result set of a stored query on data.
- Can represent a subset of data.
- Can join and simplify multiple tables.
- Can act as aggregated tables.
- Do not take up actual storage space.

INQuery Activities



- Pre-INQuery: ~3 queries per day
- Post-INQuery: ~Average of 153 queries per week / 21 queries per day

Some Studies Inspired by INQuery

Acta Neuropathol
DOI 10.1007/s00401-014-1328-5

ORIGINAL PAPER

Abnormal serine phosphorylation of insulin receptor substrate 1 is associated with tau pathology in Alzheimer's disease and tauopathies

Mark Yarchoan · Jon B. Toledo · Edward B. Lee · Zoe Arvanitakis · Hala Kazi · Li-Ying Han · Natalia Louneva · Virginia M.-Y. Lee · Sangwon F. Kim · John Q. Trojanowski · Steven E. Arnold

Brain Advance Access published November 30, 2012
doi:10.1093/brain/aws271
Brain 2012: Page 1 of 8 | 1
BRAIN
A JOURNAL OF NEUROLOGY

Cerebrovascular atherosclerosis correlates with Alzheimer pathology in neurodegenerative dementias

Mark Yarchoan,[1] Sharon X. Xie,[2] Mitchel A. Kling,[3,4] Jon B. Toledo,[1] David A. Wolk,[5] Edward B. Lee,[1] Vivianna Van Deerlin,[1] Virginia M.-Y. Lee,[1] John Q. Trojanowski[1] and Steven E. Arnold[1,3,5]

ELSEVIER
Alzheimer's & Dementia ■ (2013) 1–9
Alzheimer's & Dementia

A platform for discovery: The University of Pennsylvania Integrated Neurodegenerative Disease Biobank

Jon B. Toledo[a], Vivianna M. Van Deerlin[a], Edward B. Lee[a], EunRan Suh[a], Young Baek[a], John L. Robinson[a], Sharon X. Xie[b], Jennifer McBride[a], Elisabeth M. Wood[a], Theresa Schuck[a], David J. Irwin[a], Rachel G. Gross[c], Howard Hurtig[c], Leo McCluskey[c], Lauren Elman[c], Jason Karlawish[c], Gerard Schellenberg[a], Alice Chen-Plotkin[c], David Wolk[c], Murray Grossman[c], Steven E. Arnold[c,d], Leslie M. Shaw[a], Virginia M.-Y. Lee[a], John Q. Trojanowski[a,*]

RESEARCH ARTICLE

Comparative Survey of the Topographical Distribution of Signature Molecular Lesions in Major Neurodegenerative Diseases

Steven E. Arnold,[1]* Jon B. Toledo,[2] Dina H. Appleby,[1] Sharon X. Xie,[3] Li-San Wang,[2] Young Baek,[1] David A. Wolk,[1] Edward B. Lee,[2] Bruce L. Miller,[4] Virginia M.-Y. Lee,[2] and John Q. Trojanowski[2]

Acknowledgements

- Young Baek (Former Data Manager)
- John Trojanowski
- Steve Arnold
- Sharon Xie
- All Colleagues of Penn ADCC
- NIA Funding (AG 101024)
- NACC, an inspiration of current work
- Patients and Their Families

Thank You!
Q & A