Global Alzheimer’s Association Interactive Network

Transforming the way researchers approach the study of Alzheimer’s disease

Arthur W. Toga

July 2014
### Neuroimaging Study Size (Typical)

<table>
<thead>
<tr>
<th>Year</th>
<th>Size</th>
<th>Equivalent to</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>54MB</td>
<td>20 copies of War and Peace</td>
</tr>
<tr>
<td>2005</td>
<td>67MB</td>
<td>24 copies of War and Peace</td>
</tr>
<tr>
<td>2012</td>
<td>531MB</td>
<td>193 copies of War and Peace</td>
</tr>
</tbody>
</table>
Image Data Expansion

Each neuroimaging scan can spawn many derived image leading to exponential growth

Typical Example:
One 22MB structural scan ➔
  Five preprocessed images (176 MB) ➔
  Eleven postprocessed images (222 MB)

22MB of raw data produces 420MB data for one scan
Genetic Data

- Circa 2010 GWAS Data (per sample)
  - 620,000+ rows of data
  - ~81MB

- 2012: Whole Genome Sequencing (per sample)
  - Standard output from Illumina – multiple files and formats
  - ~250GB per sample

  - Example
    - 800 subjects x 250GB = 195TB
    - Time to transfer 195TB:
      - High speed internet (90 Mbit/s): 26 days
      - DSL (45 Mbit/s): 59 days
      - Dial-up (56 kbit/s): 100+ years!
Image & Data Archive (IDA)

*Names scaled to amount of image data archived

60+ Projects
Image Data Uploaded

- 10 TB Total
- ~ 2 TB/Year

- 17,200 subjects
- 200,000 scans
- 121,000 processed scans
- 64 million image files
Image Data Downloaded

170 TB Total

~70 TB/Year
Projected for 2014

- 9,087 IDA users
- ~50 active users / day
- 46 countries
Sharing Alzheimer's Research Data with the World

The Alzheimer's Disease Neuroimaging Initiative (ADNI) unites researchers with study data as they work to define the progression of Alzheimer's disease. ADNI researchers collect, validate and utilize data such as MRI and PET images, genetics, cognitive tests, CSF and blood biomarkers as predictors for the disease. Data from the North American ADNI's study participants, including Alzheimer's disease patients, mild cognitive impairment subjects and elderly controls, are available from this site.
Number of subjects is consistent, but fewer raw MR scans were archived in 2013 due to a reduced frequency in collecting phantom scans.
Data Use Applications

Applications received to date: 5,614
Applications received in 2013: 1,191
(29% increase from 2012)
Images Downloaded 2006 - 2013

<table>
<thead>
<tr>
<th></th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR</td>
<td>34,575</td>
<td>127,341</td>
<td>191,914</td>
<td>234,253</td>
<td>354,848</td>
<td>519,121</td>
<td>806,083</td>
<td>1,342,381</td>
</tr>
<tr>
<td>PET</td>
<td>5,739</td>
<td>17,983</td>
<td>41,630</td>
<td>50,578</td>
<td>56,433</td>
<td>131,681</td>
<td>149,543</td>
<td>262,801</td>
</tr>
</tbody>
</table>

MRI: 3,610,516
PET: 716,388
Total Downloads: 4,326,904
The process of obtaining genetic data follow two paths. Smaller datasets such as SNPs are available from within the IDA and are available to all active ADNI investigators. Large datasets, the WGS BAM assembly files require investigators to send hardware on which LONI can place a copy of the files.
WGS SNPs
(PLINK & VCF formats)
THE FUTURE OF ALZHEIMER'S RESEARCH

The Global Alzheimer's Association Interactive Network (GAAIN) is a collaborative project that will provide researchers around the globe with access to a vast repository of Alzheimer's disease research data and the sophisticated analytical tools and computational power needed to work with that data. Our goal is to transform the way scientists work together to answer key questions related to understanding the causes, diagnosis, treatment and prevention of Alzheimer's and other neurodegenerative diseases.
Why GAAIN?

Imagine Venice without bridges

- Isolated islets
- Inaccessible to pedestrians
- Requires a boat
- Requires special navigation skills
Why GAAIN?

Create an infrastructure that provides bridges between islets of data to form an accessible global information network easily traversed.
Tripit.com aggregates a user’s travel and booking information (ie: airline tickets, vacation rentals).
Problems for Investigators

- Discovering data needed for investigation
- Obtaining data from study investigators
- Combining & interpreting data across studies

Problems for Data Holders

- Protecting subject privacy
- Maintaining data control
- Receiving credit
Research efforts in Alzheimer’s disease

Research efforts could be vastly expanded in scope and capabilities if data were linked to a global infrastructure that would enable scientists to access and utilize vast, interlinked repositories of data on thousands of subjects at risk for or already suffering from the ravages of Alzheimer’s disease.
GAAIN recognizes a user’s existing accounts for partnering data sources and allows him/her to analyze the data with our tools and/or apply for additional consortiums.

The dashboard indicates which data sources are unavailable to the user (ie: the user must apply for access, data source is currently offline).
Comprehensive Analytical Tool Stack

Bank of sophisticated imaging and genetic analytical tools available

Tools are supported by the LONI Pipeline
# Data Heterogeneity

## AD Data Consortium X

<table>
<thead>
<tr>
<th>XADC</th>
<th>XID</th>
<th>SEX</th>
<th>BIRTHYR</th>
<th>MMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## ADNI

<table>
<thead>
<tr>
<th>RID</th>
<th>..</th>
<th>EXAMDATE</th>
<th>GENDER</th>
<th>DOB</th>
<th>MMSCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## AIBL

<table>
<thead>
<tr>
<th>RID</th>
<th>..</th>
<th>PTGENDER</th>
<th>PTDOB</th>
<th>MMSETOT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## AD Data Consortium Y

<table>
<thead>
<tr>
<th>PTID</th>
<th>..</th>
<th>MF</th>
<th>BIRTHDATE</th>
<th>APOE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“Google with graphs” Search
A big solution for big data

• GAAIN serves as a benchmark for large data research efforts
• Provides seamless connections of a users’ existing Alzheimer’s disease consortium data accounts
• Allows researchers to narrow down a study population that relates to their work across multiple partner consortiums
• Provides tools capable of analyzing clinical, imaging and genetic data types via the LONI Pipeline