2.5 Sample Datasets

Meaningful analysis requires carefully collected data. The Sci2 Tool can import many varieties of scientometric data, outlined in section 4.2 Data Acquisition and Preparation, but the tool also comes bundled with several sample datasets. This sample data will be used throughout sections 4 Workflow Design and 5 Sample Workflows, and includes:

- Geo
- Network
- Scientometrics
  - Bibtex
  - CSV
  - Endnote
  - ISI
- Models
  - TARL
- NIH
- NSF
- Properties
- Scopus
- SDB
  - RNAi
- IVMOOC Sample Data

Files available:

Full Sample Dataset

Geo

usptoinfluenza.csv – Heavily pre-processed and geocoded data covering USPTO patents awarded containing the keyword ‘Influenza’.

Network

convertGraph_v0.8.0.graphml

convertGraph_v0.8.0.graphml

convertGraph_v0.8.0.graphmlkidscontest.net

kidscontest.netnetscience06-conference.net

netscience06-conference.netseiyu.graphml.xml

seiyu.graphml

Scientometrics

Bibtex

Bibsonomy.bib

LaszloBarabasi.bib

CSV

LaszloBarabasi.csv

20publications.csv

Medline_master_table.csv (most recent version is always available from the Scholarly Database site)

Endnote


LaszloBarabasi.enw

ISI

AlessandroVespignani – 101 publications authored or co-authored by Alessandro Vespignani from 1990-2006.

EugeneGarfield

FourNetSciResearchers – 361 publications spanning 52 years and four network scientists: Albert-László Barabási, Eugene Garfield, Alessandro Vespignani, & Stanley Wasserman. This data was collected in 2007 from Web of Science.
LaszloBarabasi
StanleyWasserman
Test5Papers

Scientometrics.isi – All articles (2,126) published in the journal Scientometrics from 1978-2008.

Models
TARL
Agingfunction
inscript.tarl

NIH

NSF
• Refer to the CShell Datasets wiki page

Properties
• Refer to the CShell Datasets wiki page

Scopus
BrainCancer.scopus

SDB
RNAI
Medline_co-author_table_(nwb_format)
USPTO_citation_table_(nwb_format)

IVMOOC Sample Data
AlessandroVespignani – 101 publications authored or co-authored by Alessandro Vespignani from 1990-2006.

Authors.csv
dynamic_data_example1.gexf
dynamic_data_example2.gexf
Florentine.pdf
PNAS_top50_words.net
SciSIFunding.csv