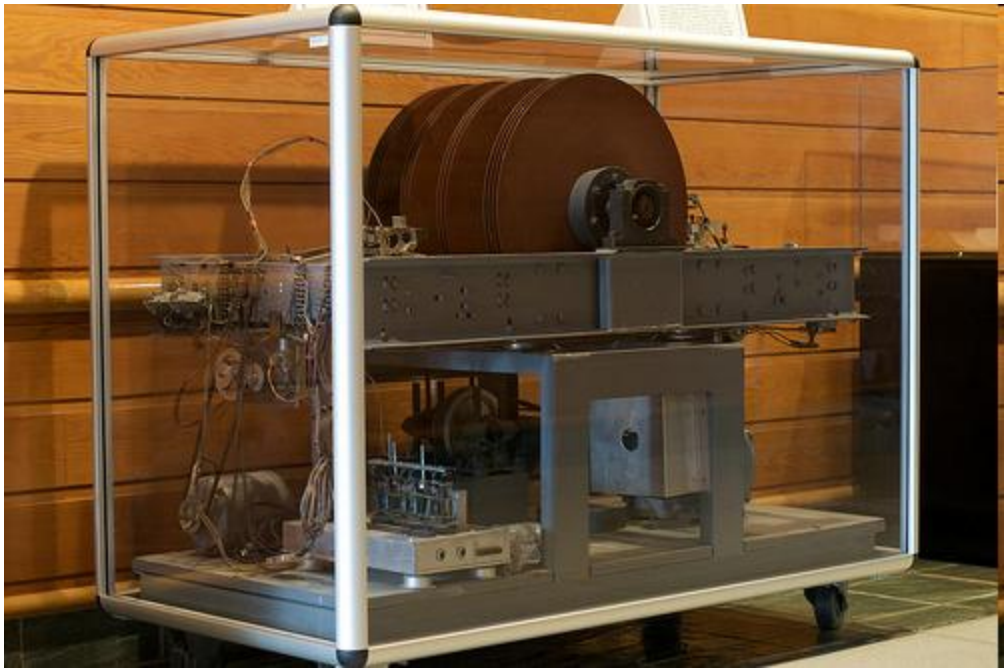
Bringing Research Data into the Library: Expanding the Horizons of Institutional Repositories

Outline

- Research data described
- Data management in libraries and Irs
- Issues with data management
- Cyberinfrastructure and the library

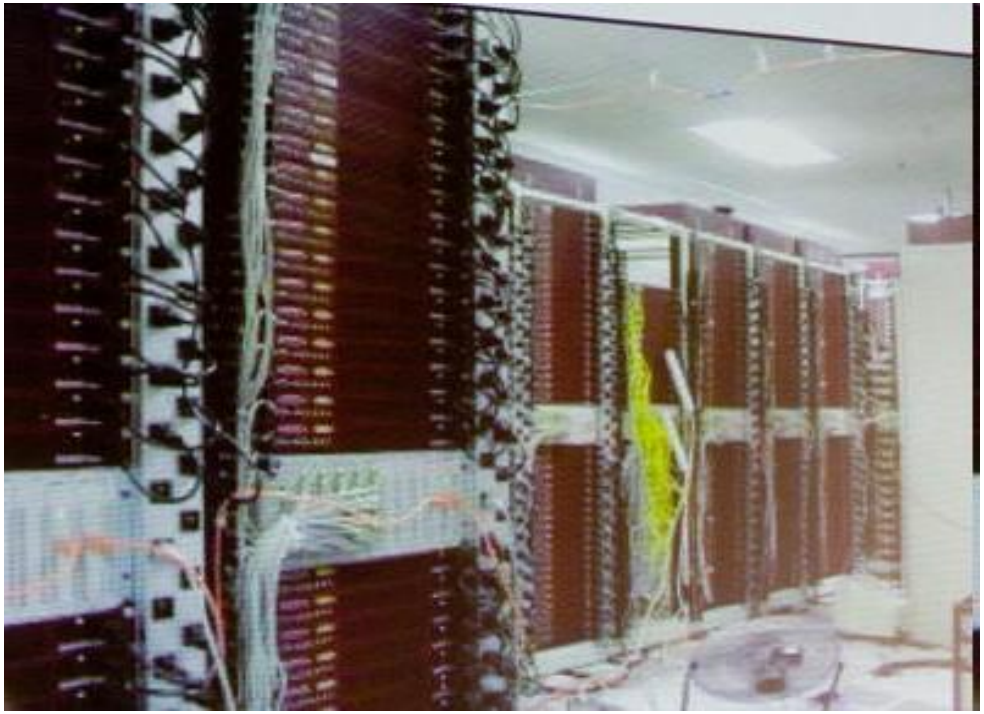

SONY
60
DVC
Digital Video Cassette
mini
x5

That Was Then



The world's first hard drive,
IBM Almaden Research Center in 1952-1954 (5Mb)

This is Now



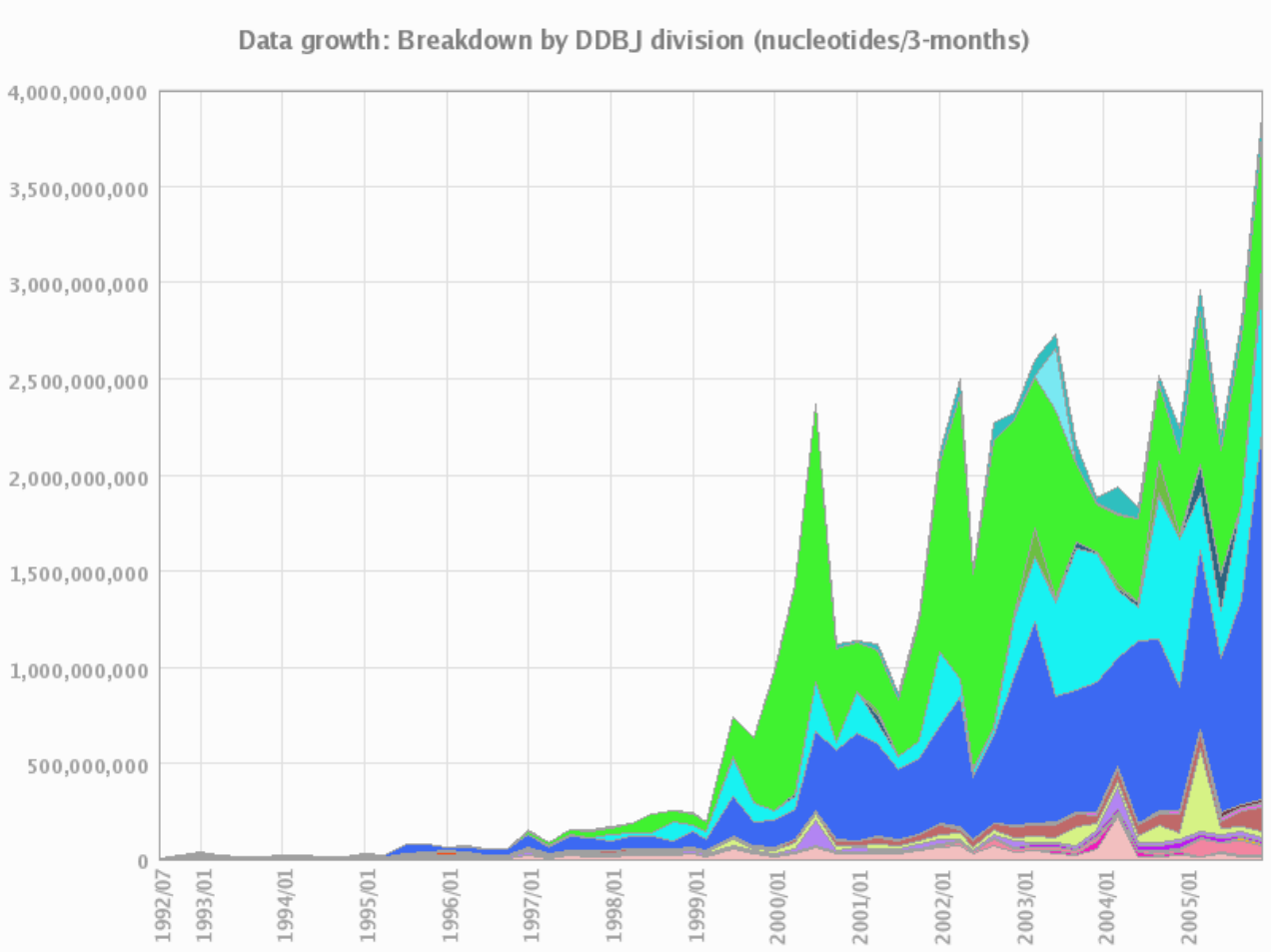
Google Data Center
(current capacity hard drives >2Tb)

How Much Information?

"IDC research shows that the digital universe —information that is either created, captured, or replicated in digital form — was 281 exabytes in 2007. In 2011, the amount of digital information produced in the year should equal nearly 1,800 exabytes, or 10 times that produced in 2006. ***The compound annual growth rate between now and 2011 is expected to be almost 60%***"

The Diverse and Exploding Digital Universe, 2008 IDC White Paper

How Much Information?



Sequence Submissions to DNA DataBank of Japan 1993-2005

What Is Research Data?

- **Observational** e.g. sensor, telemetry, survey, sample data
- **Experimental** e.g. genetic sequences, chromatograms
- **Simulation** e.g. climate models, economic models, 3D models
- **Derived/compiled** e.g. text/data mining, compiled databases
- **Multimedia** e.g. images, audio, video

Often very expensive or impossible to reproduce

What Is Research Data?

Formats include

- **Text** e.g. flat text files, Word, PDF
- **Numerical** e.g. SPSS, STATA, Excel, Access, MySQL
- **Multimedia** e.g. jpeg, tiff, dicom, mpeg, quicktime
- **Models** e.g. 3D, statistical
- **Software** e.g. Java, C
- **Domain-specific** e.g. FITS in astronomy, CIF in chemistry
- **Instrument-specific** e.g. Olympus Confocal Microscope Data Format

Example: ASCII data file

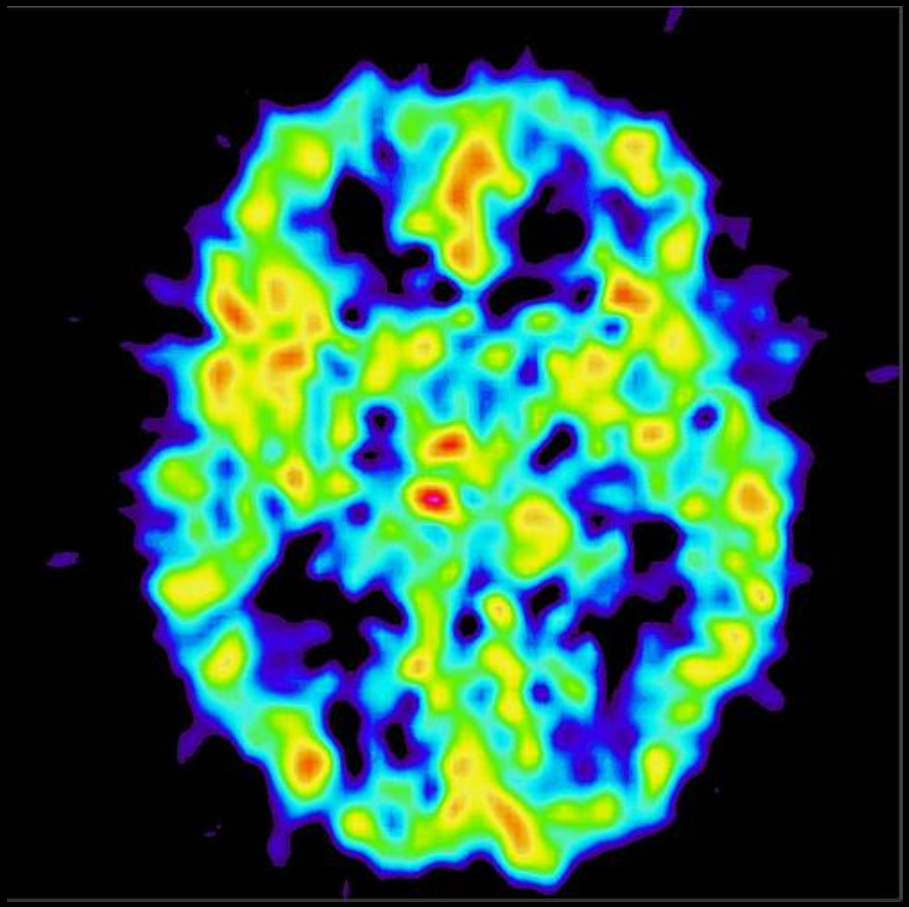
ATOT185.txt - Notepad

File Edit Format View Help

```
85032562137  1047 17139  1 2 11001  1611 146 154 353      0 1178 1532    0 0 0    0 0 0    0 1 0    0 8374    0 11350
85032563524  2537 14229  2 2 11002  2562 170 210 353  90385 1445    0    0 0 0    0 0 0    0 0 0    0 0 0
85032561034  3657 11141  3 2 11003  3045 148 140 353      0 1288 1881    0 2 0    0 8379    0 14405    0 20414    0 27421
85032562002  6456 12227  7 2 11004  6235 161 161 353      0 1162    0    0 0 0    0 0 0    0 0 0    0 1 0    0 7356
85032561002  7559 15302 11 2 11005  7785 163 297 353  90585 1428    0    0 0 0    0 0 0    0 0 0    0 0 0
85032563111  8621 17239 16 2 11006  8681 154 147 353      0 1251 1468    0 0 0    0 2 0    0 2 0    0 4450    0 10510
85032562527  1195 16243  1 2 11007  1499 146 150 353      0 1510 1765    0 0 0    0 0 0    0 3331    0 8396    0 12403
85032563215  2027 16427  2 2 11008  2669 170 147 353      0 1659 2094    0 0 0    0 1 0    0 4 0    0 6 0    0 7864
85032561027  3184 12035  3 2 11009  3099 148 161 353      0 1754 1853    0 0 0    0 0 0    0 0 0    0 3 0    0 5 0
85032562301  6267 15160  7 2 11010  6678 161 171 353      0 1246 1508    0 0 0    0 0 0    0 0 0    0 0 0    0 0 0
85032561129  7601 14405 11 2 11011  7525 163 175 353      0 1186 1380    0 0 0    0 0 0    0 0 0    0 0 0    0 0 0
85032563515  8657 11257 16 2 11012  8086 154 164 353      0 1107 1067    0 0 0    0 0 0    0 0 0    0 0 0    0 4392
85032562537  1099 17306  1 2 11013  1556 146 140 353      0 1103 1272    0 1 0    0 7290    0 13317    0 19312    0 26321
85032563207  2263 13361  2 2 11014  2148 170 171 353      0 1296 1754    0 0 0    0 0 0    0 0 0    0 0 0    0 0 0
85032562535  1216 18420  1 2 11015  1672 146 140 353      0 1162 1636    0 1 0    0 7334    0 14372    0 20378    0 27383
85032563126  2394 14440  2 2 11016  2504 170 168 353      0 1205 1617    0 0 0    0 0 0    0 0 0    0 0 0    0 1 0
85032561304  3381 12347  3 2 11017  3077 148 140 353      0 1536 1861    0 1 0    0 7458    0 13479    0 19492    0 25499
85032562329  6503 16309  7 2 11018  6751 161 147 353      0 1097 1280    0 0 0    0 2 0    0 8353    0 14362    0 21377
85032561402  7392 12202 11 2 11019  7360 163 161 353      0 1355 1458    0 0 0    0 0 0    0 0 0    0 1 0    0 5397
85032563103  8021 11124 16 2 11020  8171 154 154 353      0 1051 1435    0 0 0    0 0 0    0 2 0    0 9373    0 15398
85032562526  1704 17337  1 2 11021  1540 146 143 353      0 1229 1546    0 0 0    0 2336    0 2336    0 7348    0 12346
85032563524  2537 15509  2 2 11022  2630 170 185 353      0 1385 2213    0 0 0    0 0 0    0 0 0    0 0 0    0 0 0
85032561027  3184 12035  3 2 11023  3098 148 150 353      0 1448 1777    0 0 0    0 0 0    0 6437    0 12459    0 19473
85032566524 26244 21412  6 2 11024  5135 164 161 353      0  987 1393    0 0 0    0 0 0    0 0 0    0 1 0    0 7379
85032562501  6443 12504  7 2 11025  6220 161 168 353      0 1251 1325    0 0 0    0 0 0    0 0 0    0 0 0    0 3 0
85032561419  7150 12537 11 2 11026  7332 163 161 353      0 1227 1547    0 0 0    0 0 0    0 0 0    0 3 0    0 8457
85032563502  8553 17106 16 2 11027  8678 154 164 353      0 1276 1646    0 0 0    0 0 0    0 0 0    0 0 0    0 6456
85032562526  1704 18349  1 2 11028  1664 146 147 353      0  891 1295    0 0 0    0 1 0    0 5282    0 9297    0 14321
85032563215  2027 15323  2 2 11029  2605 170 171 353      0 1114 1732    0 0 0    0 0 0    0 0 0    0 0 0    0 0 0
85032561433  3516 12135  3 2 11030  3138 148 203 353      0 1149 1841    0 0 0    0 0 0    0 0 0    0 0 0    0 0 0
85032562426  6692 12142  7 2 11031  6249 161 161 353      0 1279 1580    0 0 0    0 0 0    0 1 0    0 4502
85032561337  7519 12161 11 2 11032  7365 163 178 353      0  806 1372    0 0 0    0 0 0    0 0 0    0 0 0    0 0 0
85032563101  8016 17148 16 2 11033  8695 154 150 353      0 1055 1273    0 0 0    0 0 0    0 5420    0 12422    0 19423
85032562136  1001 13126  1 2 11034  1236 146 140 353      0 1110 1426    0 1 0    0 7335    0 14423    0 19417    0 26403
85032563120  2349 14127  2 2 11035  2541 170 182 353      0 1073 1837    0 0 0    0 0 0    0 0 0    0 0 0    0 0 0
85032562122  1323 16008  1 2 11036  1395 146 140 353      0 1243 1394    0 2 0    0 8320    0 15343    0 21346    0 27350
85032563205  2717 13034  2 2 11037  2182 170 164 353      0  934 1623    0 0 0    0 0 0    0 0 0    0 0 0    0 3479
85032562526  1704 17337  1 2 11038  1537 146 147 353      0 1281 1366    0 0 0    0 3 0    0 6329    0 11358    0 18370
85032563219  2553 14027  2 2 11039  2256 170 168 353      0 1285 1819    0 0 0    0 0 0    0 0 0    0 0 0    0 1 0
85032561433  3516 12135  3 2 11040  3137 148 154 353      0 1396 1878    0 0 0    0 0 0    0 2 0    0 7440    0 12462
85032562329  6503 16309  7 2 11041  6753 161 147 353      0 1193 1463    0 0 0    0 1 0    0 6384    0 13390    0 16395
85032561319  7509 13461 11 2 11042  7433 163 150 353      0 1167 1347    0 0 0    0 0 0    0 3438    0 5483    0 10385
85032563102  8024 15149 16 2 11043  8544 154 161 353      0 1442 1675    0 0 0    0 0 0    0 0 0    0 1 0    0 6428
85032562137  1047 16249  1 2 11044  1495 146 140 353      0 1441 1732    0 1 0    0 7370    0 14410    0 22405    0 28409
85032563123  2034 13223  2 2 11045  2222 170 175 353      0 1361 1771    0 0 0    0 0 0    0 0 0    0 0 0    0 0 0
85032561036  3335 11260  3 2 11046  3028 148 140 353      0 1525 1973    0 3 0    0 6502    0 10495    0 16490    0 22499
85032565406 16272 12256  6 2 11047  5091 164 140 353      0 1537 1799    0 2 0    0 8364    0 15381    0 22449    0 27447
85032562027  6805 16017  7 2 11048  6693 161 154 353      0 1233 1535    0 0 0    0 0 0    0 1 0    0 6449    0 11444
85032561318  7160 13061 11 2 11049  7437 163 147 353      0 1215 1421    0 0 0    0 1 0    0 3 0    0 9414    0 15424
85032563110  8597 16357 16 2 11050  8570 154 147 353      0 1431 1674    0 0 0    0 1 0    0 4566    0 11501    0 16464
85032561202  9115 11450 20 2 11051  9033 161 161 353      0 1182 1386    0 0 0    0 0 0    0 0 0    0 1 0    0 4 0
```

start | PoDS | python | ATOT185.txt - Notepad | 3:09 PM

Example: Scientific Image



Metadata is Mandatory

To interpret data, you need *metadata*, e.g.

- Code books for statistical data
- Protocol metadata for images
- Schemas for structured ASCII data files
- Schemas for structured databases

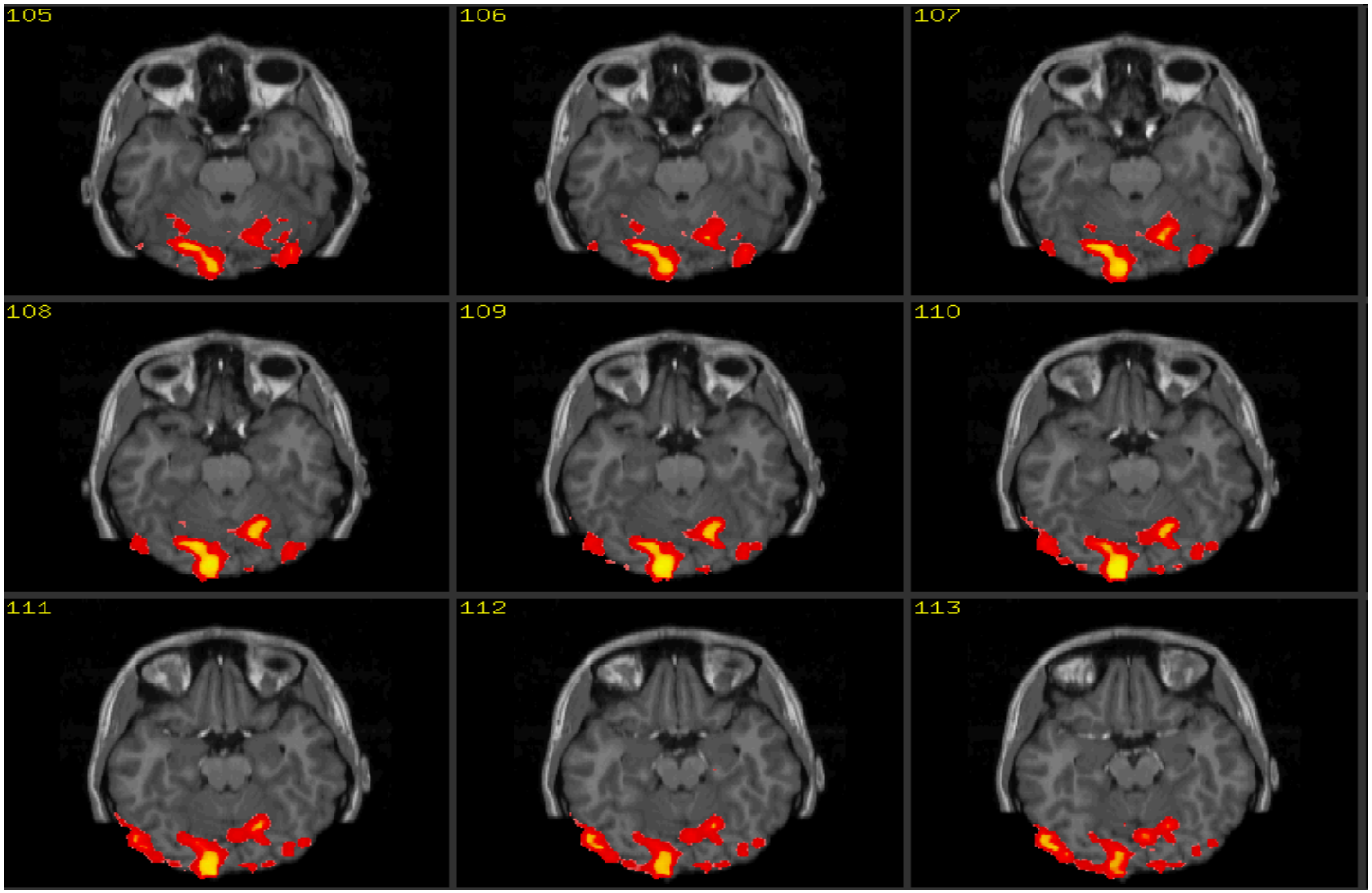
Research Data at MIT

MIT case studies for How Much Information?

- Biological Oceanography
- Climate Change
- Chemistry and Chemical Engineering
- Materials Science and Engineering
- Neuroimaging
- High Energy Physics

http://hmi.ucsd.edu/howmuchinfo.php

Neuroimaging Case Study

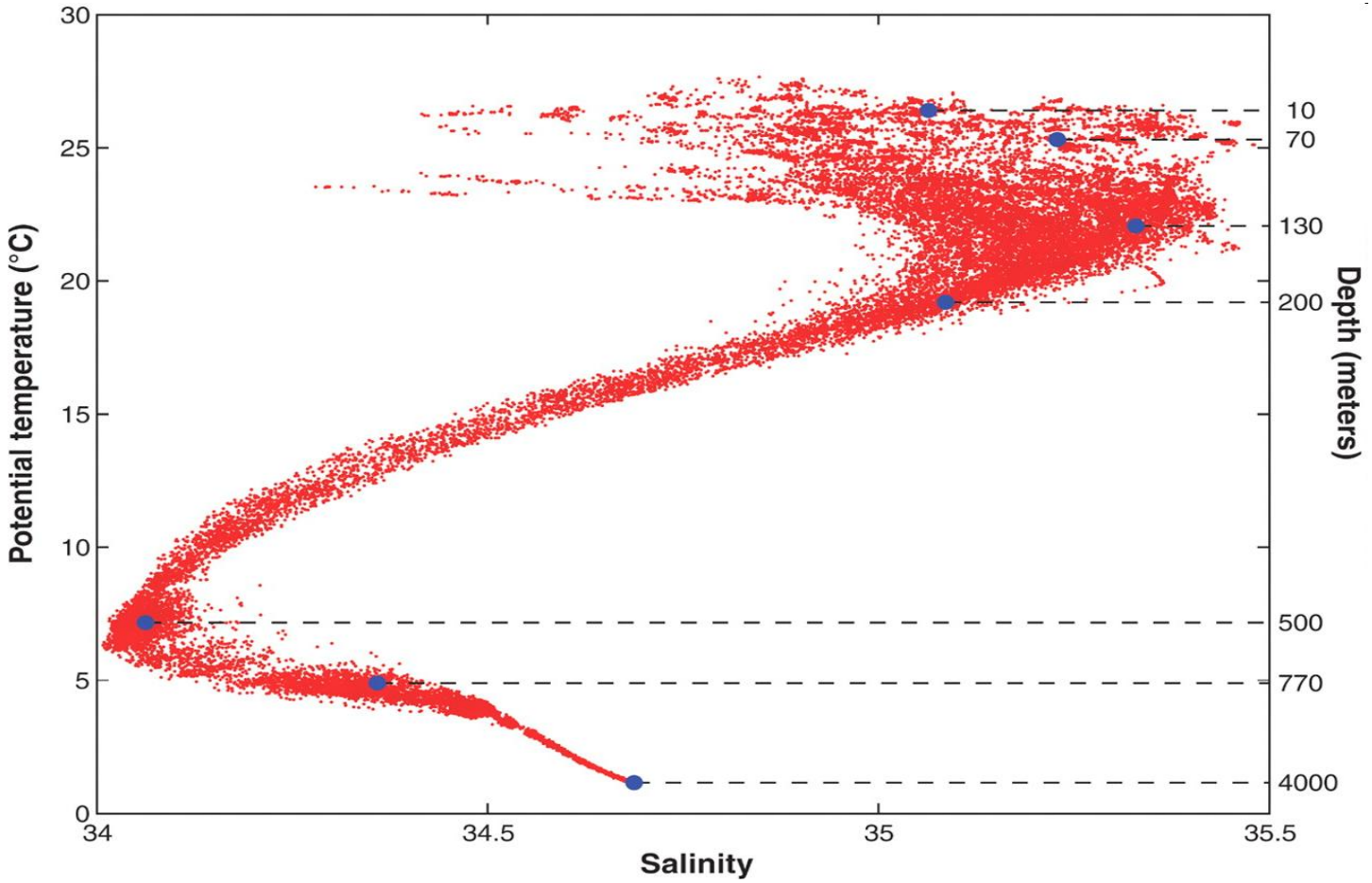


Neuroimaging Case Study

- Sources: Brain & Cognitive Science Department; McGovern Institute for Brain Research; Martinos Imaging Center; Research Lab of Electronics
- Need to *combine* digital images (MRIs, DTIs, VBM, etc.) *with* protocol data, phenotype and genomic data, EEGs, etc. from the same subject
- LARGE-SCALE
  >10Tb per year for one group of 4 faculty
- EXPENSIVE

  each subject ~$1000 (1500/year, per machine)
- HARD TO FIND, INTERPRET
  no standard way to annotate images for sharing, reuse

Biological Oceanography Case Study

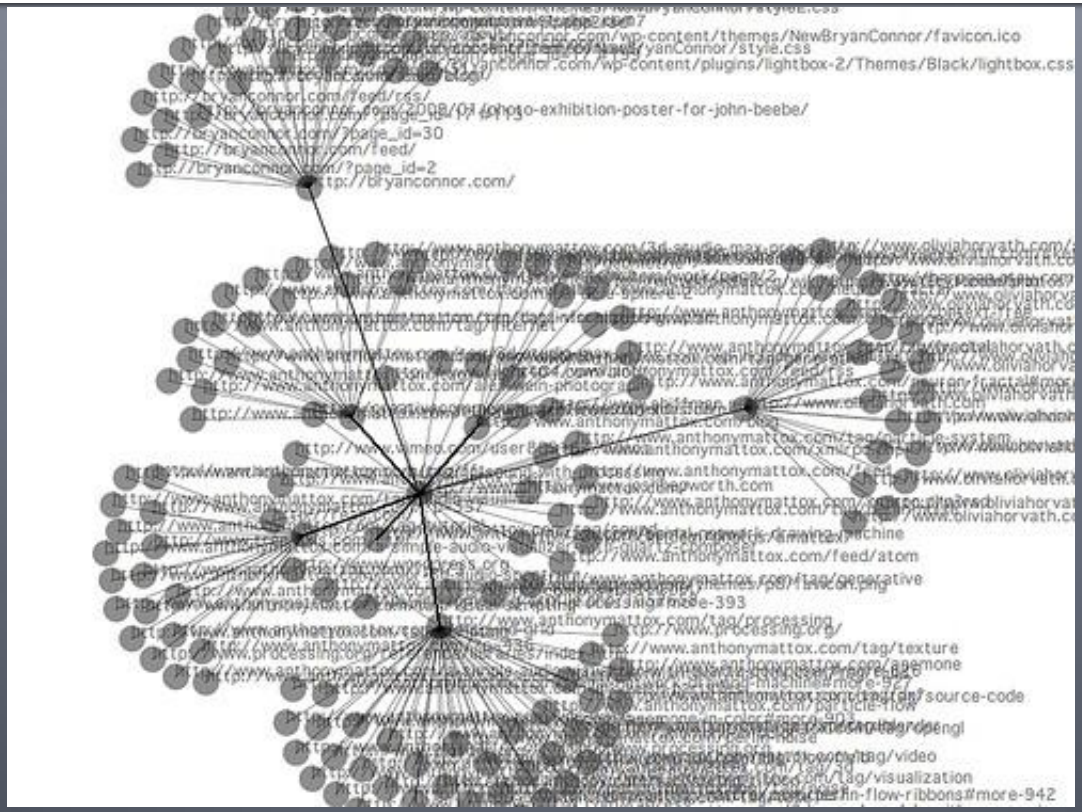


Temperature versus salinity (T-S) relations for the North Pacific Subtropical Gyre at station ALOHA

Biological Oceanography Case Study

- Sources: Earth, Atmospheric and Planetary Sciences; Civil Engineering; Biological Engineering
- Need to *combine* metagenomics data *with* biochemical sensor data, including water chemistry, optical properties, physical data (e.g. location)
- LARGE-SCALE
  Solexa sequencer produces 1Tb per run X 2-3 runs/week
- IRREPLACEABLE
  time dependent, not fully analyzable today
- NEED TO COLLABORATE
  no integrated DB exists (e.g. GenBank only takes sequences)

Libraries and Data



Libraries and Data

Already established data curation for

statistical (Harvard-MIT Data Center)

geospatial (Geodata Repository)

bioinformatics (via NCBI)

digital library collections (e.g. images, videos)

general datasets (DSpace digital archive)

Social Science Data Services

Finding and Managing Data for Research

Home

Social Science Data Services
- FAQ

Data Access
- By Subject
- Data Centers
- Locating Data
- Responsible Use
- Citing Data
- Restricted Data
- Suggest a Purchase

Training
- Tutorials
- Workshops

Software & Hardware
- Software
- Hardware & Facilities

Archiving & Disseminating Your Data
- Overview
- Metadata Standard: DDI
- Metadata Services Unit
- Data Management and Publishing

Welcome to Social Science Data Services

Social Science Data Services provides assistance with finding, understanding, and managing statistics or numeric or tabular data in the social sciences, management, and related areas. Browse the lefthand navigation bar for information about accessing data, training, software and hardware, and archiving and disseminating your data. See also how Social Science Data Services relates to the Geographic Information Systems (GIS).

If you're searching for data, see our resources by subject and consider searching in one of several data centers such as the Harvard-MIT Data Center and ICPSR.

News

- IAP Workshop: Finding Research Datasets
- ICPSR Summer Internship Program
- ICPSR Undergraduate Research Paper Competition
- New Office Hours for Statistical Consultant: Mondays 9am-12pm; NOTE: No office hours held in December

Subject Expert

Katherine McNeill
Data Services and Economics Librarian
Dewey Library

mcneillh@mit.edu
617-253-0787

Need help? Ask Us!

Quick Links

- Harvard-MIT Data Center
- HMDC Statistical Consultant
- ICPSR

New Resources

- China Data Online **(MIT only)**
- Historical Statistics of the United States **(MIT only)**

This page was last updated on Thursday, 11-Dec-2008 13:54:05 EST

Katherine McNeill, Data Services and Economics Librarian, mcneillh@mit.edu
MIT Libraries - Ask Us!
Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, MA 02139-4397 USA

Social Science Data Management

Both faculty authored and acquired datasets

- Consultation services
- Liaise with national archives (e.g. ICPSR)
- Develop standards (e.g. DDI)
- Maintain website

geographic information systems services @ mit

GIS Home
Hours

Spatial Data
GeoWeb
Hardware + Software
Teaching + Learning
Examples

Maps + Mapping
Social Science Data Services
US Census
Data Management & Publishing Guide

GIS Lab
Rotch Library, building 7-238

Contact Us
gishelp(at)mit.edu

Search GIS Services

GIS

Geographic Information Systems (GIS) enable users to visualize and analyze spatial information in a dynamic, digital environment. It provides tools for integrating, querying and analyzing a wide variety of data types, such as scientific and cultural data, satellite imagery and aerial photography, as well as data collected by individuals, into projects, with geographic locations providing the integral link between all the data.

Mission

The GIS Laboratory is a collaboration between the MIT Libraries, Information Services & Technology (IS&T), and the Office of Educational Innovation and Technology (OEIT). Our goal is to support the MIT community's research and academic activities, as they relate to Geographic Information Systems (GIS). This support not only includes the collection of literature, data, software, and hardware; but also, training and assistance in the use of GIS. We are available to work with faculty to implement and assist in teaching GIS-based labs in individual subjects.

MIT does NOT offer a GIS certificate or degree program.

Hot Topics:

- **New Workshop:** Introduction to GIS for Architecture
- How do I get started working with GIS at MIT?
  - How do I obtain ESRI GIS software for my personal or lab computer?
- What GIS data is available at MIT?
- ESRI Virtual Campus Classes
- **Data updates:**
  - Manhattan, NY 3D buildings from 2009 in MIT Geodata Repository - (MIT web certificates required for viewing and accessing)
  - New administrative boundaries for countries all around the world - including municipal, counties, postal codes, etc. (MIT web certificates required for viewing and accessing)
  - Energy data from Platts in the MIT Geodata Repository- (MIT web certificates required for viewing and accessing)
  - City of Boston Land Parcels- (MIT web certificates required for viewing and accessing)
  - New York City Tax Lots- (MIT web certificates required for viewing and accessing)
  - South Africa: Cape Town | Eastern Cape | Western Cape
  - India GIS datasets
  - China GIS datasets
  - data on the web by state webpage update: NYC Data Mine
  - Find out what's new in the MIT libraries through RSS feeds
- **Services updates:**
  - Download the latest version of the MIT Geodata Repository Search Tool - released October 2009
    - Learn more about it with the new MIT Geodata Repository Search Tool for ArcGIS Help Guide
  - ESRI ArcGIS 9.3.1 available to the MIT community
  - MIT GeoWeb - GIS data access with a web browser
  - GPS units available for checkout from Rotch Library

Hours

The GIS laboratory, housed in Rotch Library, is available for use during Rotch operating hours.

GIS Lab Assistance: September 14 - December 10

A GIS specialist will be available in the GIS lab for walk-in assistance from
12:30-4:00 pm, Monday-Thursday.

If you have a GIS question and are not available during walk-in lab hours please email gishelp(at)mit.edu to check if a GIS specialist can meet with you at a different time A brief description of

GIS Data Management

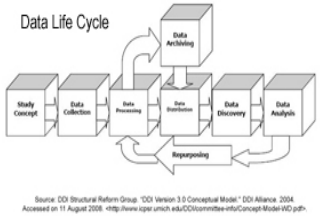
- Consultation and training services
- Acquire and catalog datasets
- Develop standards (e.g. FGDC)
- Maintain website, Geodata Repository

Data Management and Publishing

Manage Your Data

Managing your data before you begin your research and throughout its life cycle is essential to ensure its current usability and long-run preservation and access. Data management activities include:

A Data Planning Checklist | Create a Data Management Plan | Guides to Data Management



Source: DDI Structural Reform Group. "DDI Version 3.0 Conceptual Model." DDI Alliance. 2004. Accessed on 11 August 2008. <http://www.icpsr.umich.edu/DDI/committee-info/Concept-Model-WD.pdf>.

A Data Planning Checklist:

1. What type of data will be produced? Will it be reproducible? What would happen if it got lost or became unusable later?
2. How much data will it be, and at what growth rate? How often will it change?
3. Who will use it now, and later?
4. Who controls it (PI, student, lab, MIT, funder)?
5. How long should it be retained? e.g. 3-5 years, 10-20 years, permanently
6. Are there tools or software needed to create/process/visualize the data?
7. Any special privacy or security requirements? e.g. personal data, high-security data
8. Any sharing requirements? e.g. funder data sharing policy
9. Any other funder requirements? e.g. data management plan in proposal
10. Is there good project and data documentation?
11. What directory and file naming convention will be used?
12. What project and data identifiers will be assigned?
13. What file formats? Are they long-lived?
14. Storage and backup strategy?
15. When will I publish it and where?
16. Is there an ontology or other community standard for data sharing/integration?

Create a Data Management Plan

Planning for your data management needs ahead of time will save you time and resources in the long run and ensure that your data will be usable in the future. A formal plan can be valuable to you and may be required by your funding agency. Topics to cover in a written plan include:

- name of the person responsible for data management within your research project
- description of data to be collected and the methodology
- how data will be documented throughout the research project
- data quality issues
- backup procedures
- how data will be made available for public use and potential secondary uses
- preservation plans
- any exceptional arrangements that might be needed to protect participant confidentiality or intellectual property

For tips on creating a data sharing or management plan, see the:

- Australian National University Data Management Manual
- NIH examples of data sharing plans
- RELU-DSS Data Management Plan

Guides to Data Management

Faculty Successes:

"I've had thousands of downloads of my published data--I am impressed that it's been so useful to others!"

Esther Duflo, Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics, MIT

For advice on a data management project, contact:

data-management@mit.edu

Anne Graham
Civil and Environmental Engineering Librarian

Katherine McNeill
Data Services and Economics Librarian

Amy Stout
Computer Science Librarian

Lisa Sweeney
Head of GIS Services

General Data Management

- Website averages 650 hits/month
- Team of Public Services librarians
  - liaisons for Civil and Environmental Engineering, Computer Science, GIS, Social Science
  - Broader group includes Bioinformatics, Chemistry, Humanities, others
- Consultation from Metadata Services specialists

Data Archiving in DSpace@MIT

- Excel spreadsheets [microarray data, survey data]
- Zip files [gene sequences in .gb, code in .py]
- XML files [chemical molecules in CML]
- Media files [images, audio/video in .wav]
- Software-specific [Sequest DTA format for spectral data]
- Java Web Start files
- Lab e-notebooks
- Thesis supplementary data

Data Archiving in DSpace@MIT

Abdul Latif Jameel Poverty Action Lab (J-PAL)
J-PAL Datasets

- stata files (.dta)
- associated codebooks (.doc)

Data Archiving in DSpace@MIT

Research Laboratory for Electronics (RLE)
Speech Communication Group: MIT American English Map Task

- sound files (.wav) and scanned map files (.pdf)
- README file

Data Archiving in DSpace@MIT

Synthetic Biology: Code accompanying "Analysis of Targeted and Combinatorial Approaches to Phage T7 Genome Generation" Master's Thesis

- Software files (.py, .h, .lib, etc.)
- Input data files (.txt)
- README (.doc)

Case Study: RADISH

Robotics Datasets in DSpace@MIT

http://dspace-demo.mit.edu/handle/1234567890/41939

Robotics Data in DSpace@MIT

What we did:

- Defined local taxonomy for metadata values
- Changed item record display
- Changed submission form, simplified workflow
- Non-MIT submissions are reviewed by community owner or library staff
- Loaded data from previous repository and added CC.o licenses

Radish: The Robotics Data Set Repository

Search within this collection:

Go

Advanced Search

The Robotics Data Set Repository (Radish for short) provides a collection of standard robotics data sets. Contained here-in you will find:

- Logs of odometry, laser and sonar data taken from real robots.
- Logs of all sorts of sensor data taken from simulated robots.
- Environment maps generated by robots.
- Environment maps generated by hand (i.e., re-touched floor-plans).

By making these data sets available to the community, we aim to facilitate the development, evaluation and comparison of robotics algorithms. While the current focus is clearly on localization and mapping, we expect that Radish will ultimately expand to reflect the interests of the broader robotics community.

Radish is a community effort. Researchers are invited to download and make use of the data sets contained here-in, and, in return, to make their own contributions to the repository.

Good data sets are too precious to keep to ourselves!

Submitting to Radish and FAQ:

To register as a new data contributor, email radish@mit.edu.

FAQ: http://libraries.mit.edu/dspace-mit/about/radish.html

Software:

Player/Stage: http://playerstage.sourceforge.net/

Carmen: http://carmen.sourceforge.net/

OpenSLAM: http://openslam.org/

Please note: When you put your data in Radish, you are putting it in the public domain. Please contact radish@mit.edu if you have questions about this.

Recent Submissions

ualberta-csc-flr3-vision
Klippenstein, Jonathan (2009-04-07)

comparison_of_self-localization_methods_continued
Gutmann, Steffen (2009-04-07)

albert-b-laser-vision
Stachniss, Cyrill (2009-04-07)

isr-fctuc_lrf1_lrf2_cam_imu_carmen_dataset
Davim, Luis; Dias, Jorge Manuel Miranda; Ferreira, Filipe; Prado, Jose (2009-04-07)

fr101-explored
Stachniss, Cyrill (2009-03-11)

Edmonton_Convention_Centre

Show full item record

Citable URI: http://hdl.handle.net/1234567890/41975

Title:	Edmonton_Convention_Centre
Author:	Roy, Nick
Description:	This is the map used by GRACE to navigate from the registration desk to the site of the robot talk. The two data files consist of a CARMEN log file, and a CARMEN map file. The map file was constructed from the log file, and is 10cm resolution.
URI:	http://hdl.handle.net/1234567890/41975
Issue Date:	2009-03-11
Keywords:	2D sensor data, Indoor sensor data, SICK LMS laser range finder

Files in this item

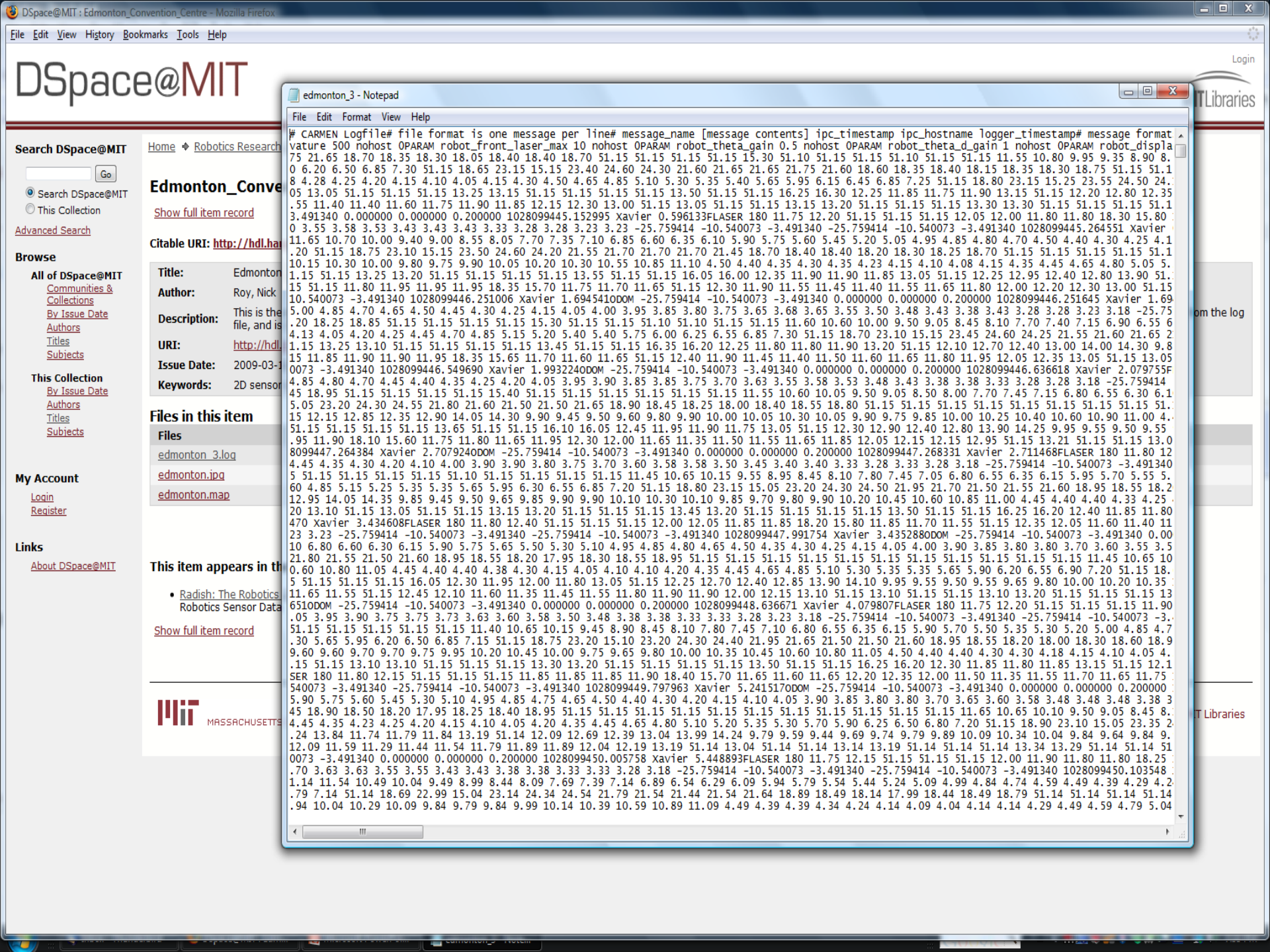
Files	Size	Format
edmonton_3.log	7.024Mb	application/octet-stream
edmonton.jpg	69.26Kb	image/jpeg
edmonton.map	4.071Mb	application/octet-stream

Download all files as a METS zip archive.

This item appears in the following Collection(s)

- Radish: The Robotics Data Set Repository
  Robotics Sensor Data

Show full item record


DSpace@MIT : Edmonton_Convention_Centre - Mozilla Firefox
File Edit View History Bookmarks Tools Help
DSpace@MIT
Login
Search DSpace@MIT
Go
Search DSpace@MIT
This Collection
Advanced Search
Browse
All of DSpace@MIT
Communities & Collections
By Issue Date
Authors
Titles
Subjects
This Collection
By Issue Date
Authors
Titles
Subjects
My Account
Login
Register
Links
About DSpace@MIT
Home
Robotics Research
Edmonton_Conve
Show full item record
Citable URI: http://hdl.ha
Title: Edmonton
Author: Roy, Nick
Description: This is the file, and is
URI: http://hdl
Issue Date: 2009-03-1
Keywords: 2D sensor
Files in this item
Files
edmonton_3.log
edmonton.jpg
edmonton.map
This item appears in th
Radish: The Robotics
Robotics Sensor Data
Show full item record
MASSACHUSETTS
IT Libraries
edmonton_3 - Notepad
File Edit Format View Help
CARMEN Logfile# file format is one message per line# message_name [message contents] ipc_timestamp ipc_hostname logger_timestamp# message format
vature 500 nohost OPARAM robot_front_laser_max 10 nohost OPARAM robot_theta_gain 0.5 nohost OPARAM robot_theta_d_gain 1 nohost OPARAM robot_displa
75 21.65 18.70 18.35 18.30 18.05 18.40 18.40 18.70 51.15 51.15 51.15 51.15 15.30 51.10 51.15 51.15 51.10 51.15 51.15 11.55 10.80 9.95 9.35 8.90 8.
0 6.20 6.50 6.85 7.30 51.15 18.65 23.15 15.15 23.40 24.60 24.30 21.60 21.65 21.65 21.75 21.60 18.60 18.35 18.40 18.15 18.35 18.30 18.75 51.15 51.1
8 4.28 4.25 4.20 4.15 4.10 4.05 4.15 4.30 4.50 4.65 4.85 5.10 5.30 5.35 5.40 5.65 5.95 6.15 6.45 6.85 7.25 51.15 18.80 23.15 15.25 23.55 24.50 24.
05 13.05 51.15 51.15 51.15 13.25 13.15 51.15 51.15 51.15 51.15 51.15 13.50 51.15 51.15 16.25 16.30 12.25 11.85 11.75 11.90 13.15 51.15 12.20 12.80 12.35
.55 11.40 11.40 11.60 11.75 11.90 11.85 12.15 12.30 13.00 51.15 13.05 51.15 51.15 13.15 13.20 51.15 51.15 51.15 13.30 13.30 51.15 51.15 51.15 51.1
3.491340 0.000000 0.000000 0.200000 1028099445.152995 Xavier 0.596133FLASER 180 11.75 12.20 51.15 51.15 51.15 12.05 12.00 11.80 11.80 18.30 15.80
0 3.55 3.58 3.53 3.43 3.43 3.43 3.33 3.28 3.28 3.23 3.23 -25.759414 -10.540073 -3.491340 -25.759414 -10.540073 -3.491340 1028099445.264551 Xavier
11.65 10.70 10.00 9.40 9.00 8.55 8.05 7.70 7.35 7.10 6.85 6.60 6.35 6.10 5.90 5.75 5.60 5.45 5.20 5.05 4.95 4.85 4.80 4.70 4.50 4.40 4.30 4.25 4.1
.20 51.15 18.75 23.10 15.15 23.50 24.60 24.20 21.55 21.70 21.70 21.70 21.45 18.70 18.40 18.40 18.20 18.30 18.25 18.70 51.15 51.15 51.15 51.15 51.1
10.15 10.30 10.00 9.80 9.75 9.90 10.05 10.20 10.30 10.55 10.85 11.10 4.50 4.40 4.35 4.30 4.35 4.23 4.15 4.10 4.08 4.15 4.35 4.45 4.65 4.80 5.05 5.
1.15 51.15 13.25 13.20 51.15 51.15 51.15 51.15 13.55 51.15 51.15 16.05 16.00 12.35 11.90 11.90 11.85 13.05 51.15 12.25 12.95 12.40 12.80 13.90 51.
15 51.15 11.80 11.95 11.95 11.95 18.35 15.70 11.75 11.70 11.65 51.15 12.30 11.90 11.55 11.45 11.40 11.55 11.65 11.80 12.00 12.20 12.30 13.00 51.15
10.540073 -3.491340 1028099446.251006 Xavier 1.694541ODOM -25.759414 -10.540073 -3.491340 0.000000 0.000000 0.200000 1028099446.251645 Xavier 1.69
5.00 4.85 4.70 4.65 4.50 4.45 4.30 4.25 4.15 4.05 4.00 3.95 3.85 3.80 3.75 3.65 3.68 3.65 3.55 3.50 3.48 3.43 3.38 3.43 3.28 3.28 3.23 3.18 -25.75
.20 18.25 18.85 51.15 51.15 51.15 51.15 15.30 51.15 51.15 51.10 51.10 51.15 51.15 11.60 10.60 10.00 9.50 9.05 8.45 8.10 7.70 7.40 7.15 6.90 6.55 6
4.13 4.05 4.20 4.25 4.45 4.70 4.85 5.15 5.20 5.40 5.40 5.75 6.00 6.25 6.55 6.85 7.30 51.15 18.70 23.10 15.15 23.45 24.60 24.25 21.55 21.60 21.65 2
1.15 13.25 13.10 51.15 51.15 51.15 51.15 13.45 51.15 51.15 16.35 16.20 12.25 11.80 11.80 11.90 13.20 51.15 12.10 12.70 12.40 13.00 14.00 14.30 9.8
15 11.85 11.90 11.90 11.95 18.35 15.65 11.70 11.60 11.65 51.15 12.40 11.90 11.45 11.40 11.50 11.60 11.65 11.80 11.95 12.05 12.35 13.05 51.15 13.05
0073 -3.491340 1028099446.549690 Xavier 1.993224ODOM -25.759414 -10.540073 -3.491340 0.000000 0.000000 0.200000 1028099446.636618 Xavier 2.079755F
4.85 4.80 4.70 4.45 4.40 4.35 4.25 4.20 4.05 3.95 3.90 3.85 3.85 3.75 3.70 3.63 3.55 3.58 3.53 3.48 3.43 3.38 3.38 3.33 3.28 3.28 3.18 -25.759414
45 18.95 51.15 51.15 51.15 51.15 15.40 51.15 51.15 51.15 51.15 51.15 51.15 11.55 10.60 10.05 9.50 9.05 8.50 8.00 7.70 7.45 7.15 6.80 6.55 6.30 6.1
5.05 23.20 24.30 24.55 21.80 21.60 21.50 21.50 21.65 18.90 18.45 18.25 18.00 18.40 18.55 18.80 51.15 51.15 51.15 51.15 51.15 51.15 51.15 51.1
15 12.15 12.85 12.35 12.90 14.05 14.30 9.90 9.45 9.50 9.60 9.80 9.90 10.00 10.05 10.30 10.05 9.90 9.75 9.85 10.00 10.25 10.40 10.60 10.90 11.00 4.
51.15 51.15 51.15 51.15 13.65 51.15 51.15 16.10 16.05 12.45 11.95 11.90 11.75 13.05 51.15 12.30 12.90 12.40 12.80 13.90 14.25 9.95 9.55 9.50 9.55
.95 11.90 18.10 15.60 11.75 11.80 11.65 11.95 12.30 12.00 11.65 11.35 11.50 11.55 11.65 11.85 12.05 12.15 12.15 12.95 51.15 13.21 51.15 51.15 13.0
8099447.264384 Xavier 2.707924ODOM -25.759414 -10.540073 -3.491340 0.000000 0.000000 0.200000 1028099447.268331 Xavier 2.711468FLASER 180 11.80 12
4.45 4.35 4.30 4.20 4.10 4.00 3.90 3.90 3.80 3.75 3.70 3.60 3.58 3.58 3.50 3.45 3.40 3.40 3.33 3.28 3.33 3.28 3.18 -25.759414 -10.540073 -3.491340
5 51.15 51.15 51.15 51.15 51.10 51.15 51.15 51.15 51.15 11.45 10.65 10.15 9.55 8.95 8.45 8.10 7.80 7.45 7.05 6.80 6.55 6.35 6.15 5.95 5.70 5.55 5.
60 4.85 5.15 5.25 5.35 5.35 5.65 5.95 6.30 6.55 6.85 7.20 51.15 18.80 23.15 15.05 23.20 24.30 24.50 21.95 21.70 21.50 21.55 21.60 18.95 18.55 18.2
12.95 14.05 14.35 9.85 9.45 9.50 9.65 9.85 9.90 9.90 10.10 10.30 10.10 9.85 9.70 9.80 9.90 10.20 10.45 10.60 10.85 11.00 4.45 4.40 4.40 4.33 4.25
20 13.10 51.15 13.05 51.15 51.15 13.15 13.20 51.15 51.15 51.15 13.45 13.20 51.15 51.15 51.15 51.15 13.50 51.15 51.15 16.25 16.20 12.40 11.85 11.80
470 Xavier 3.434608FLASER 180 11.80 12.40 51.15 51.15 51.15 12.00 12.05 11.85 11.85 18.20 15.80 11.85 11.70 11.55 51.15 12.35 12.05 11.60 11.40 11
23 3.23 -25.759414 -10.540073 -3.491340 -25.759414 -10.540073 -3.491340 1028099447.991754 Xavier 3.435288ODOM -25.759414 -10.540073 -3.491340 0.00
10 6.80 6.60 6.30 6.15 5.90 5.75 5.65 5.50 5.30 5.10 4.95 4.85 4.80 4.65 4.50 4.35 4.30 4.25 4.15 4.05 4.00 3.90 3.85 3.80 3.80 3.70 3.60 3.55 3.5
21.80 21.55 21.50 21.60 18.95 18.55 18.20 17.95 18.30 18.55 18.95 51.15 51.15 51.15 51.15 51.15 51.15 51.15 51.15 51.15 51.15 51.15 11.45 10.65 10
0.60 10.80 11.05 4.45 4.40 4.40 4.38 4.30 4.15 4.05 4.10 4.10 4.20 4.35 4.45 4.65 4.85 5.10 5.30 5.35 5.35 5.65 5.90 6.20 6.55 6.90 7.20 51.15 18.
5 51.15 51.15 51.15 16.05 12.30 11.95 12.00 11.80 13.05 51.15 12.25 12.70 12.40 12.85 13.90 14.10 9.95 9.55 9.50 9.55 9.65 9.80 10.00 10.20 10.35
11.65 11.55 51.15 12.45 12.10 11.60 11.35 11.45 11.55 11.80 11.90 11.90 12.00 12.15 13.10 51.15 13.10 51.15 51.15 13.10 13.20 51.15 51.15 51.15 13
651ODOM -25.759414 -10.540073 -3.491340 0.000000 0.000000 0.200000 1028099448.636671 Xavier 4.079807FLASER 180 11.75 12.20 51.15 51.15 51.15 11.90
.05 3.95 3.90 3.75 3.75 3.73 3.63 3.60 3.58 3.50 3.48 3.38 3.38 3.33 3.33 3.28 3.23 3.18 -25.759414 -10.540073 -3.491340 -25.759414 -10.540073 -3.
51.15 51.15 51.15 51.15 51.15 11.40 10.65 10.15 9.45 8.90 8.45 8.10 7.80 7.45 7.10 6.80 6.55 6.35 6.15 5.90 5.70 5.50 5.35 5.30 5.20 5.00 4.85 4.7
.30 5.65 5.95 6.20 6.50 6.85 7.15 51.15 18.75 23.20 15.10 23.20 24.30 24.40 21.95 21.65 21.50 21.50 21.60 18.95 18.55 18.20 18.00 18.30 18.60 18.9
9.60 9.60 9.70 9.70 9.75 9.95 10.20 10.45 10.00 9.75 9.65 9.80 10.00 10.35 10.60 10.80 11.05 4.50 4.40 4.40 4.30 4.30 4.18 4.15 4.10 4.05 4.
.15 51.15 13.10 13.10 51.15 51.15 51.15 13.30 13.20 51.15 51.15 51.15 51.15 13.50 51.15 51.15 16.25 16.20 12.30 11.85 11.80 11.85 13.15 51.15 12.1
SER 180 11.80 12.15 51.15 51.15 51.15 11.85 11.85 11.85 11.90 18.40 15.70 11.65 11.60 11.65 12.20 12.35 12.00 11.50 11.35 11.55 11.70 11.65 11.75
540073 -3.491340 -25.759414 -10.540073 -3.491340 1028099449.797963 Xavier 5.241517ODOM -25.759414 -10.540073 -3.491340 0.000000 0.000000 0.200000
5.90 5.75 5.60 5.45 5.30 5.10 4.95 4.85 4.75 4.65 4.50 4.40 4.30 4.20 4.15 4.10 4.05 3.90 3.85 3.80 3.80 3.70 3.65 3.60 3.58 3.48 3.48 3.48 3.38 3
45 18.90 18.50 18.20 17.95 18.25 18.40 18.95 51.15 51.15 51.15 51.15 51.15 51.15 51.15 51.15 51.15 51.15 11.65 10.65 10.10 9.50 9.05 8.45 8.
4.45 4.35 4.23 4.25 4.20 4.15 4.10 4.05 4.20 4.35 4.45 4.65 4.80 5.10 5.20 5.35 5.30 5.70 5.90 6.25 6.50 6.80 7.20 51.15 18.90 23.10 15.05 23.35 2
.24 13.84 11.74 11.79 11.84 13.19 51.14 12.09 12.69 12.39 13.04 13.99 14.24 9.79 9.59 9.44 9.69 9.74 9.79 9.89 10.09 10.34 10.04 9.84 9.64 9.84 9.
12.09 11.59 11.29 11.44 11.54 11.79 11.89 11.89 12.04 12.19 13.19 51.14 13.04 51.14 51.14 13.14 13.19 51.14 51.14 51.14 13.34 13.29 51.14 51.14 51
0073 -3.491340 0.000000 0.000000 0.200000 1028099450.005758 Xavier 5.448893FLASER 180 11.75 12.15 51.15 51.15 51.15 12.00 11.90 11.80 11.80 18.25
.70 3.63 3.63 3.55 3.55 3.43 3.43 3.38 3.38 3.33 3.33 3.28 3.18 -25.759414 -10.540073 -3.491340 -25.759414 -10.540073 -3.491340 1028099450.103548
1.14 11.54 10.49 10.04 9.49 8.99 8.44 8.09 7.69 7.39 7.14 6.89 6.54 6.29 6.09 5.94 5.79 5.54 5.44 5.24 5.09 4.99 4.84 4.74 4.59 4.49 4.39 4.29 4.2
.79 7.14 51.14 18.69 22.99 15.04 23.14 24.34 24.54 21.79 21.54 21.44 21.54 21.64 18.89 18.49 18.14 17.99 18.44 18.49 18.79 51.14 51.14 51.14 51.14
.94 10.04 10.29 10.09 9.84 9.79 9.84 9.99 10.14 10.39 10.59 10.89 11.09 4.49 4.39 4.39 4.34 4.24 4.14 4.09 4.04 4.14 4.14 4.29 4.49 4.59 4.79 5.04

Case Study: Unilever Centre for Molecular Informatics

Molecular Data in DSpace@Cambridge

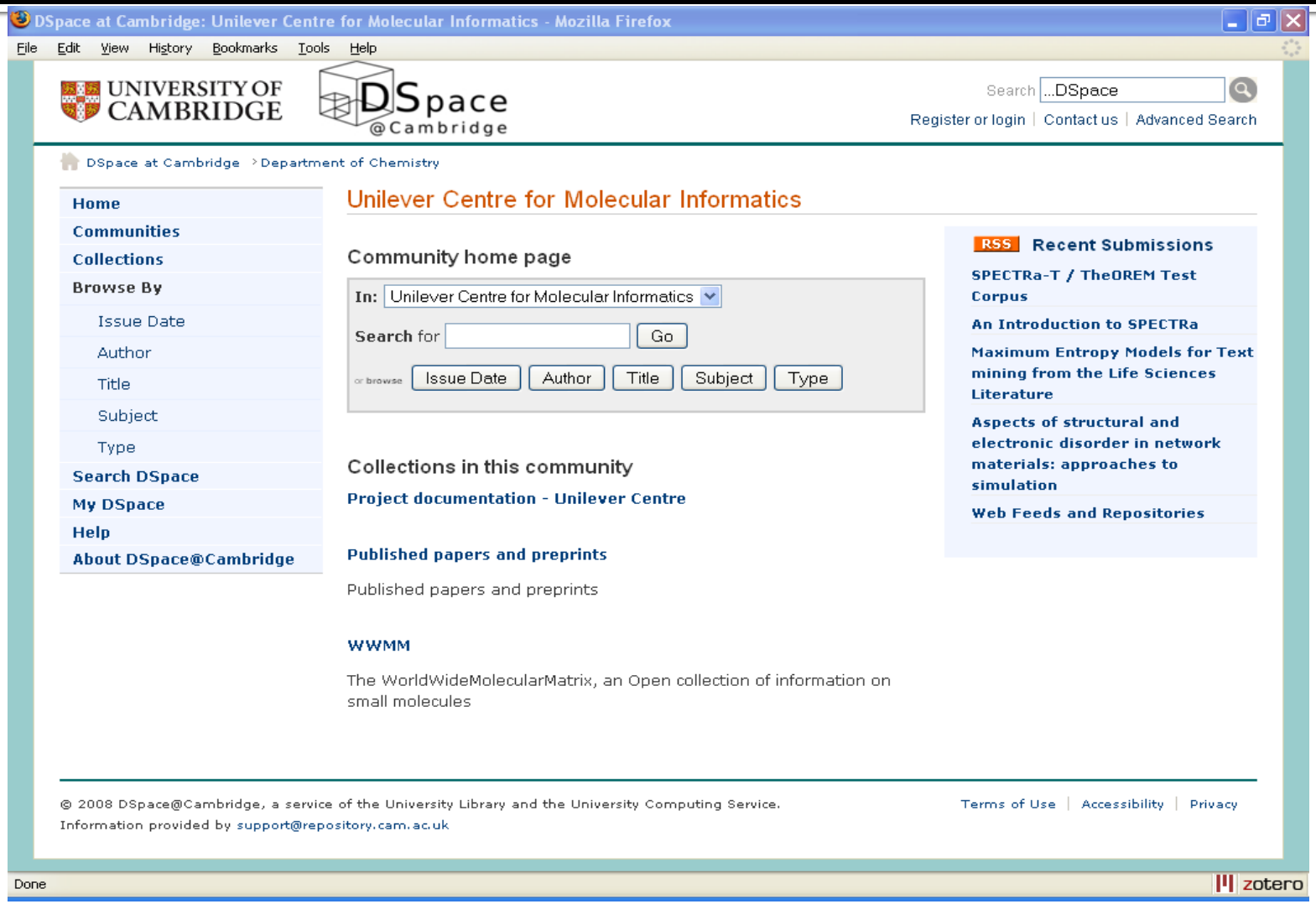
http://www.dspace.cam.ac.uk/handle/1810/723

University of Cambridge Datasets

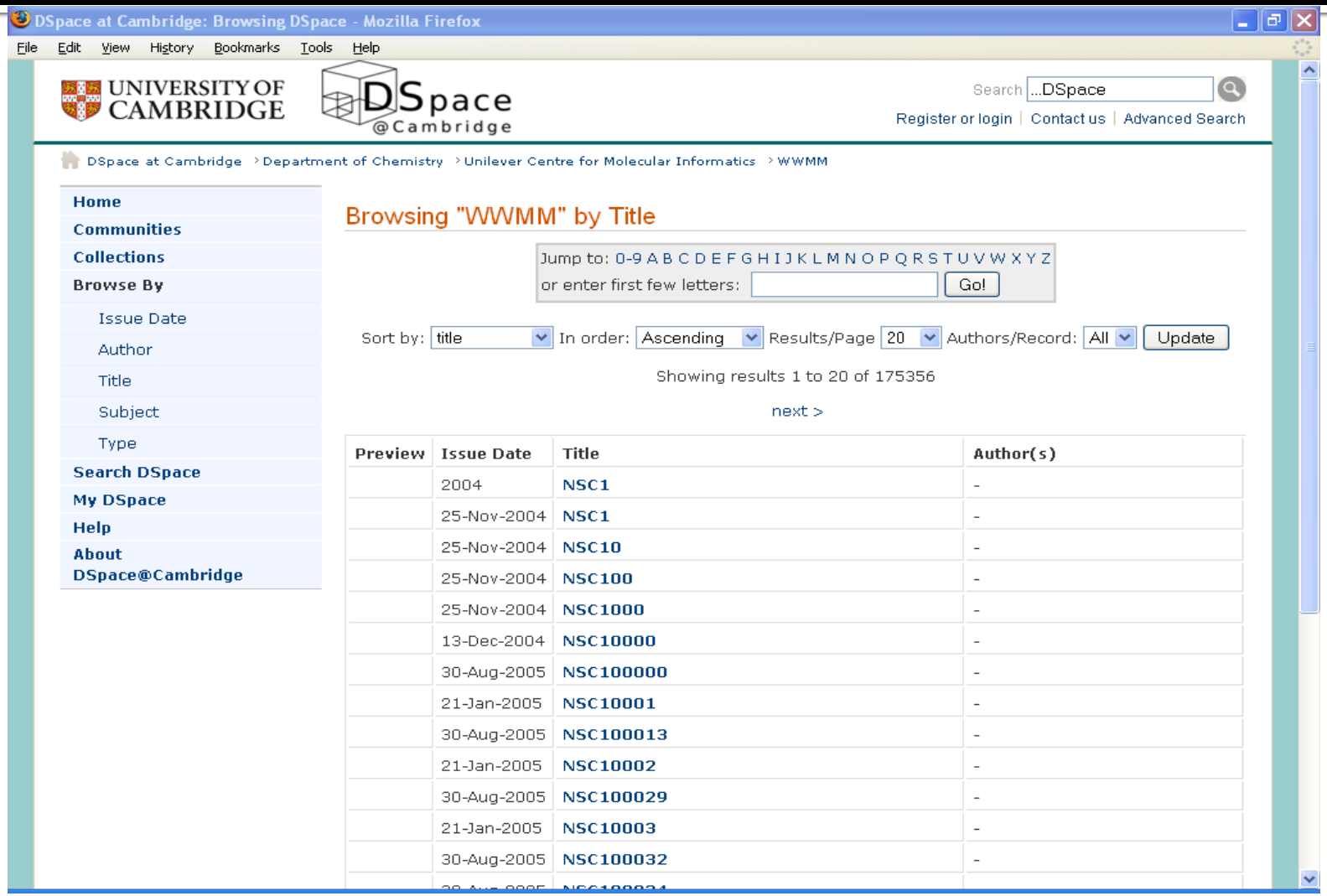
"Data sharing and data archiving have a long tradition within the academic community. Repositories offer new tools for data sharing within a complex information environment. DSpace@Cambridge accepts data deposit thus offering stewardship of institutional knowledge assets of all types and facilitating compliance with the new research council policies and mandates."

[DSpace@Cambridge website]

University of Cambridge Datasets



World Wide Molecular Matrix



Chemical Markup Language (CML)

CML file for Triphenylphosphine

nsc000010_original.cml - WordPad

File Edit View Insert Format Help

```
<?xml version="1.0" encoding="UTF-8"?>
<entry>
   <identifier version="0.932Beta" tautomeric="0">
      <basic>C18H15P,1H-4H-10H-16(11H-5H-1)19(17-12H-6H-2H-7H-13H-17)18-14H-8H-3H-9H-15H-18</basic>
      <charge/>
   </identifier>
   <cas>603-35-0</cas>
   <smiles>P(c1ccccc1)(c2ccccc2)c3ccccc3</smiles>
   <name>Phosphine, triphenyl- (8CI9CI)</name>
   <name>Triphenylphosphine</name>
   <name>Triphenylphosphorus</name>
   <name>WLN: RPR&amp;amp;R</name>
   <molecule id="NSC10">
      <atomArray>
         <atom id="a1" elementType="P" x3="-0.0225" y3="1.8097" z3="0.012"/>
         <atom id="a2" elementType="C" x3="-0.1845" y3="-2.7573" z3="0.3483"/>
         <atom id="a3" elementType="C" x3="-1.3539" y3="-2.02" z3="0.4036"/>
         <atom id="a4" elementType="C" x3="1.0338" y3="-2.1202" z3="0.1924"/>
         <atom id="a5" elementType="C" x3="1.0879" y3="-0.7451" z3="0.0903"/>
         <atom id="a6" elementType="C" x3="-1.3106" y3="-0.6444" z3="0.3025"/>
         <atom id="a7" elementType="C" x3="-0.0867" y3="0.0018" z3="0.1449"/>
         <atom id="a8" elementType="C" x3="2.2271" y3="1.8604" z3="1.656"/>
         <atom id="a9" elementType="C" x3="3.5813" y3="1.9519" z3="1.9054"/>
         <atom id="a10" elementType="C" x3="1.7562" y3="1.9299" z3="0.3469"/>
         <atom id="a11" elementType="C" x3="4.4707" y3="2.1128" z3="0.8579"/>
         <atom id="a12" elementType="C" x3="2.6559" y3="2.0915" z3="-0.7042"/>
         <atom id="a13" elementType="C" x3="4.0082" y3="2.1819" z3="-0.4442"/>
         <atom id="a14" elementType="C" x3="0.2555" y3="3.0413" z3="-2.4724"/>
         <atom id="a15" elementType="C" x3="0.3907" y3="3.0508" z3="-3.8457"/>
         <atom id="a16" elementType="C" x3="0.1548" y3="1.8293" z3="-1.7932"/>
         <atom id="a17" elementType="C" x3="0.1911" y3="0.6324" z3="-2.5049"/>
         <atom id="a18" elementType="C" x3="0.4263" y3="1.8592" z3="-4.5479"/>
         <atom id="a19" elementType="C" x3="0.3266" y3="0.6528" z3="-3.8781"/>
         <atom id="a20" elementType="H" x3="-0.2222" y3="-3.8333" z3="0.4323"/>
         <atom id="a21" elementType="H" x3="-2.3024" y3="-2.522" z3="0.5254"/>
         <atom id="a22" elementType="H" x3="1.9439" y3="-2.7002" z3="0.1502"/>
         <atom id="a23" elementType="H" x3="2.0393" y3="-0.2487" z3="-0.0314"/>
         <atom id="a24" elementType="H" x3="-2.2241" y3="-0.0698" z3="0.345"/>
         <atom id="a25" elementType="H" x3="1.534" y3="1.7304" z3="2.474"/>
         <atom id="a26" elementType="H" x3="3.9474" y3="1.8985" z3="2.9201"/>
         <atom id="a27" elementType="H" x3="5.5298" y3="2.184" z3="1.0572"/>
```

For Help, press F1

Case Study: Ecological Data

Watershed Datasets in Cornell's eCommons Institutional Repository

http://ecommons.library.cornell.edu/

SEARCH
Go
Advanced Search

- Home
- About
- Help
- Browse All
- Titles
- Authors
- Subjects
- By Date
- My Account
  authorized users

eCommons@Cornell >
College of Agriculture and Life Sciences >

Browsing "Agricultural Ecology Program: Upper Susquehanna River Basin" by Title

Jump to: 0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
or enter first few letters: Go!

Sort by: title In order: Ascending Results/Page 20 Authors/Record: All Update

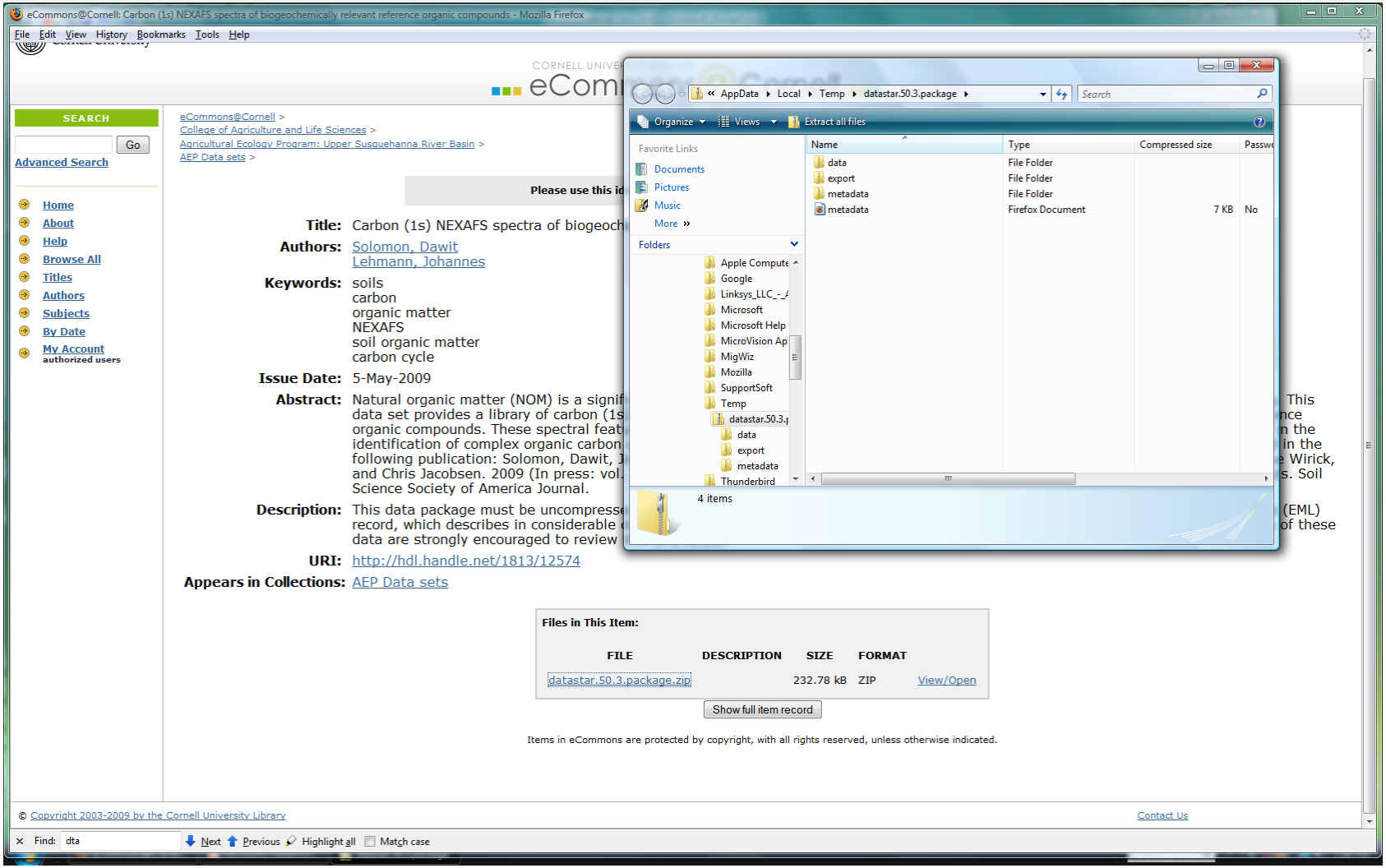
Showing results 1 to 10 of 10

PREVIEW	ISSUE DATE	TITLE	AUTHOR(S)
	5-May-2009	Carbon (1s) NEXAFS spectra of biogeochemically relevant reference organic compounds	Solomon, Dawit; Lehmann, Johannes
	26-Jun-2007	Distribution of Cs-137 in stream sediments and stream banks in the upper Susquehanna basin - 2006	Nagle, Greg; Fahey, Tim; Woodbury, Peter
	15-May-2007	Hearing on Non-point Source Pollution: The Impacts of Agriculture on Water Quality	Howarth, Robert W.
	15-May-2007	Progress Report: Understanding Sources and Sinks of Nutrients and Sediment in the Upper Susquehanna River Basin	Woodbury, Peter; Porter, Mary Jane
	2-Aug-2007	Water quality data for Fall Creek (Tompkins County, NY) sampling sites: 1972-1995	Bouldin, David
	29-Oct-2007	Water quality data for Kashong Creek Watershed (Ontario County and Yates County, NY) sampling sites: 1977-1979	Bouldin, David
	12-Nov-2007	Water quality data for southern tributaries to Cayuga Lake (Tompkins County, NY): 1987-1989	Bouldin, David
	17-Oct-2007	Water quality data for well, stream, and seep samples from the Harford Teaching and Research Farm (Cortland County, NY): 1974-1994	Bouldin, David
	2-Aug-2007	Well Logs for Wells at the Cornell Department of Animal Science Harford Teaching and Research Center	Bouldin, David
	6-Feb-2009	Workshop on Atmospheric Deposition of Nitrogen - Chesapeake Bay Program, Science and Technical Advisory Committee	Entringer, Ron; Howarth, Robert

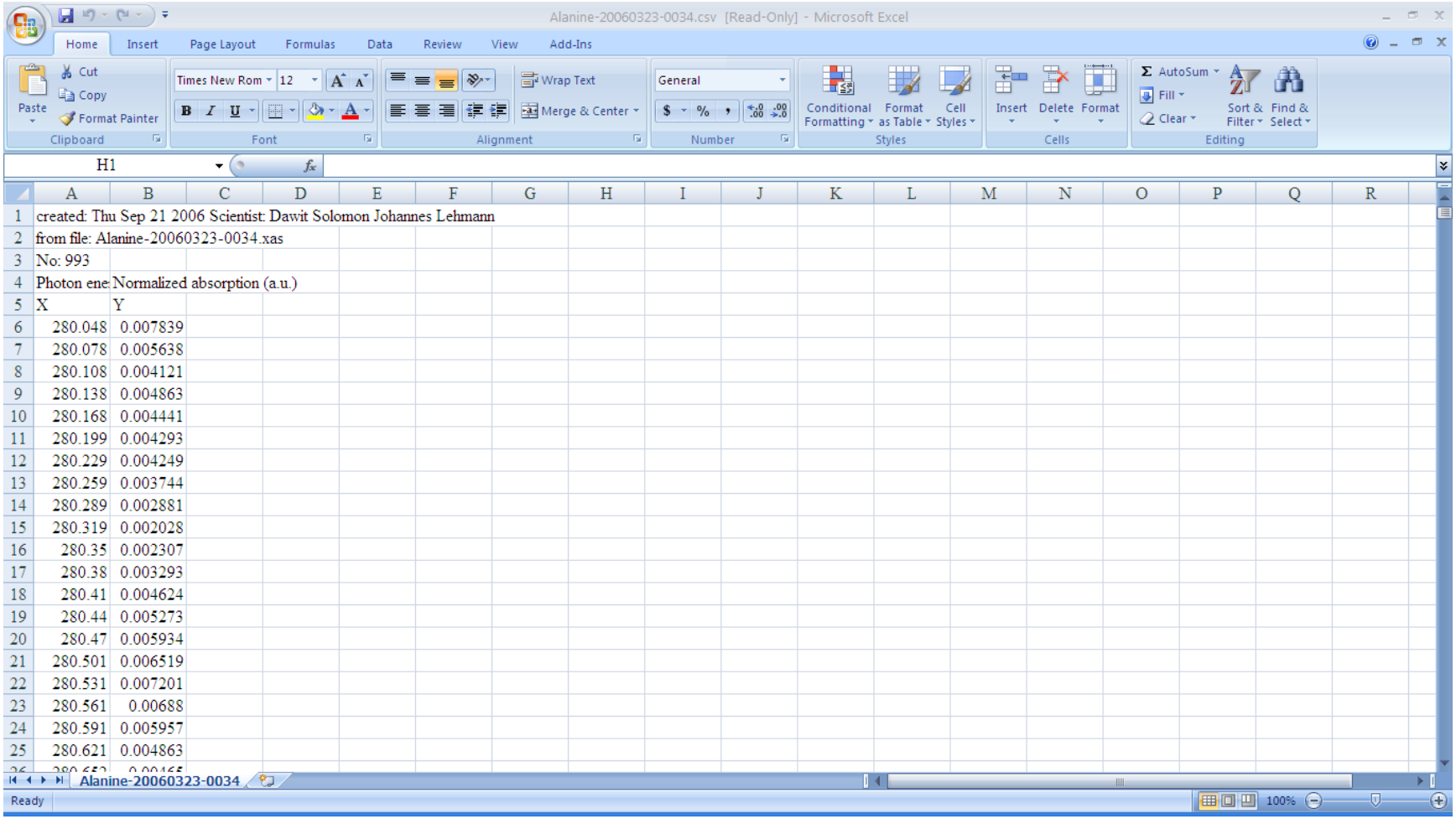
Showing results 1 to 10 of 10

Item Record for Dataset

ZIP file of Standard Data



Excel Data files



ASCII Data Files

StreamSedCs137UpperSusqTribs2006.txt - Notepad

File Edit Format View Help

```
SampleType      ID      SiteID  Sample  Notes
River   CA1     Cayuta 1        2.63    Bridge at lowman
River   CA2     Cayuta 2        1.96    250 yards downstream of 1
River   CH1     Chenango= Nan 1 6.15    Near head, below bridge, couple of miles below Eaton
River   CH2     Nan 2   3.09    Actually on Sangerfield river just above confluence with chenango
River   CH3     Nan 3   5.59    Pleasant brook, village of Smyrna.
River   CH4     Nan 4   5.44    Bridge on Rt. 12 at N Norwich, state fishing access site
River   CH5     Nan 5   2.67    Bridge downstream of Norwich, fishing access site. RT 220 bridge just N of Oxford has large cutbank that can be sampled for heavy
metals etc.
River   CH6     Nan 6   19.96   Geneganslet creek 2 miles west of green on 206.
River   CH7     Nan 7   21.26   Geneganslet creek, Smithville milldam site
River   CK2     Catatonk 2      5.11    Catatonk creek just below bridge, on catatonk hill road, 3 miles downstream of Candor.
River   CM1     Chemung 2       2.32    Mainstem about 2 miles below 1, , state fishing area, just downstream of bridge on rt 44
River   CM2     Chemung 3       1.95    Bridge at wellsburg, lowman
River   CO1     Cohocton 1 CH1  12.40   Fishing access by bridge on rt 371 north of cohocton
River   CO2     Cohocton 2 CH2  7.93    Wentworth road bridge, 5 mile south of cohocoton. Took pics from bridge
River   CO3     Cohocton 3 CH 3 7.53    Right below 415 bridge at Wallace, fishing access point
River   CO4     Cohocton 4 CH 4 5.06    About a mile upstream of Bath, bridge at junction with county rt 15
River   CO5     Cohoc 5 3.71    no notes
River   CO6     Cohoc 6 4.25    no notes
River   CO7     Cohoc 7 1.94    no notes
River   ET1     E fk Tiough 1   6.38    Loring crossing one mile above cortland
River   ET2     E fk Tiough 2   2.76    1 mile upstream of #1 at East river rd.
River   ET3     E fk Tiough 3   4.81    Bridge (closed) on cheningo creek with little bank erosion of any kind .
River   ET4     E fk Tiough 4   3.80    Mainstem, on 13 bridge 3 miles upstream of truxton
River   ME1     Meads  M 1      1.26    Lower bridge rt 26, lots of sediment running off of piles taken from below bridge, but definitely stream carried deposits, went
downstream but still pretty sandy stuff.
River   ME2     Meads M2        1.36    No notes on this but taken at fishing access site about 10 miles above M 1
River   ME3     Meads M 3       2.29    Near Monterrey, pine creek rod bridge, sandy stuff lot of bank erosion above here, cobbles etc.
River   ME4     Meads 4 1.89    Wixon Road bridge, fine stuff. Lots of bank erosion upstream of here,
River   ME5     Meads 5 2.24    Recent overbank deposit ¼ mile above bridge at DEC site.
River   ME6     M6      1.27    Recent overbank deposit 300 yards above lower highway bridge and M1.
River   ME7     M7      0.00    Dry run overbank, backwater  deposit very recent, below lower bridge.
River   ME8     M-8     0.63    Lower end near airfield, recent overbank, backwater deposits
River   NT1     Newtown 3       1.07    Above SWC dam, backwater area, stuff may have been deposited in high banks when lake full which has been there since 1988.
River   OT1     Otselic 1       1.97    Bridge on north road, 1 mile upstream of Otselic
River   OT2     Otselic 2       7.66    Fishing access site on main road bridge near 7th day hollow
River   OT3     Otselic 3       8.29    Just south of south otselic, fishing access
River   OT5     Otselic 5 ( no 4)       3.50    Bridge just below cincinnatus
River   OT6     Otselic 6       5.84    Head of whitney point reservoir
River   OW1     Owego 1 2.04    Actually owego creek at bridge on 17 c just upstream of Owego, next to ball field,
River   SE1     Seeley S1       4.64    Upstream 200 m from bridge to webber mills, taken from multiple places inside bankful but lots of stuff could be just eroded
from low floodplain which is inside bankful for the most part
River   SE2     Seeley 2        4.18    Just downsteam of webber  village, ¼ mile upstream of  #1,same floodplain sites, eroded banks 5 feet high by  corn field
River   SE3     Seeley 3        4.62    Hammond creek, just upstream of bridge at rts 328/549 just inside PA, really big bank across creek, took pics
River   SE4     Seeley 4        5.38    taken from many places along floodplain, below huge eroding banks
River   SE5     Seeley 5        5.34    Just upstream of 328/549 junction in PA. right next to place where I took a lot of pics of large bank erosion. This stuff may
be a bit sandy but lots of glacial erosion upstream
River   SE6     Seeley 6        0.00    Upstream and just below 328 bridge above webber mills. Most is from beaver dam on floodplain
River   SE7     Seeley 7        5.84    Upper seeley, at bridge on Birch creek road. Much area of wetland upstream of here and little  bank erosion.seen from main
road.
River   SE8     S 8     0.00    Floodplain backwater, 1 mile below bridge to Webb Mills
River   TB1     Trout 1 3.50    Just downstream of village of McGraw, from just below bridge and downstream. Sandy  sediment, dubious about this stuff
River   TB2     Trout 2 4.28    Just upstream of town park in McGraw,. Fine overbank materials.
```

Ecological Metadata Language (EML)

Ecological Metadata Language (EML) is a metadata specification developed by the ecology discipline and for the ecology discipline. It is based on prior work done by the Ecological Society of America and associated efforts (Michener et al., 1997, Ecological Applications). EML is implemented as a series of XML document types that can by used in a modular and extensible manner to document ecological data. Each EML module is designed to describe one logical part of the total metadata that should be included with any ecological dataset.

Send any comments, errors, or suggestions to **eml-dev@ecoinformatics.org** or through the **EML Bug Tracking system**. The preferred way to submit problems with EML or feature requests is the bug tracking system.

EML Version 2.1.0

You can access the EML specification online by reading it in HTML format, or you can download the entire specification, including both the HTML documentation and the XML Schema files.

- **EML 2.1.0 Specification** -- Read it online

  OR

- **Download EML**
  The download consists of the EML modules, described in the **XML Schema** language. In addition, the full documentation on the modules is provided in HTML format.
- **Changes to EML in version 2.1.0**
- **EML Frequently Asked Questions (FAQ)**
- **Validation service for EML**
  In addition to the online service found at the previous link, the EML distribution itself contains the validation software for your use (see "lib/runEMLParser" for details on how to run it).

About the EML Project

The EML project is an open source, community oriented project dedicated to providing a high-quality metadata specification for describing data relevant to the ecological discipline. The project is completely comprised of **voluntary project members** who donate their time and experience in order to advance information management for ecology. Project decisions are made by consensus according to the voting procedures described in the **ecoinformatics.org Charter**.

We welcome contributions to this work in any form. Individuals who invest substantial amounts of time and make valuable contributions to the development and maintenance of EML (in the opinion of

```
<?xml version="1.0" ?>
- <eml:eml packageId="gss1.6.23" system="knb" xmlns:eml="eml://ecoinformatics.org/eml-2.0.1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="eml://ecoinformatics.org/eml-2.0.1 eml.xsd">
  - <dataset>
      <title>Distribution of Cs-137 in stream sediments and stream banks in the upper Susquehanna basin - 2006</title>
    - <creator id="1181657625791">
      - <individualName>
          <givenName>Greg</givenName>
          <surName>Nagle</surName>
        </individualName>
        <organizationName>Cornell University, Department of Natural Resources</organizationName>
        <positionName>Researcher</positionName>
      - <address>
          <deliveryPoint>Fernow Hall</deliveryPoint>
          <deliveryPoint>Cornell University</deliveryPoint>
          <city>Ithaca</city>
          <administrativeArea>NY</administrativeArea>
          <postalCode>14853</postalCode>
        </address>
      </creator>
    - <creator id="1181657704416">
      - <individualName>
          <givenName>Tim</givenName>
          <surName>Fahey</surName>
        </individualName>
        <organizationName>Cornell University, Department of Natural Resources</organizationName>
        <positionName>Liberty Hyde Bailey Professor</positionName>
      - <address>
          <deliveryPoint>12 Fernow Hall</deliveryPoint>
          <deliveryPoint>Cornell University</deliveryPoint>
          <city>Ithaca</city>
          <administrativeArea>NY</administrativeArea>
          <postalCode>14853</postalCode>
        </address>
        <electronicMailAddress>tjf5@cornell.edu</electronicMailAddress>
```

EML metadata files

Mozilla Firefox

File Edit View History Bookmarks Tools Help

eCommons@Cornell: Carbon (1s) NEXA... | file:///C:/DOCU...p/metadata.html

Type real

Online Distribution Info:

ecogrid://knb/datastar.23.1

Data Table:

Name: **Spectral data: Alanine**

Description: Baseline corrected and normalized data output from WinXAS (version 3.1) software. Additional data included in the first 5 rows of table: date of analysis (date created), scientists, source file name, column headings.

Physical Structure Description:

Object Name: Alanine-20060323-0034.csv

Text Format: Maximum Record Length: column

Simple Delimited: Field Delimeter: ,

Attribute(s) Info:

Attribute Name	Column Label	Definition	Type of Value	Measurement Type	Measurement Domain	Missing Value Code	Accuracy Report	Accuracy Assessment	CoverageMethod
X		Photon energy (eV)		interval	**Unit** electronVolt **Precision** 0.001 **Type** real				
Y		Normalized absorption, in arbitrary units (a.u.)		interval	**Unit** dimensionless **Precision** 0.000001 **Type** real				

Online Distribution Info:

ecogrid://knb/datastar.24.1

Data Table:

Name: **Spectral data: Arabinose**

Description: Baseline corrected and normalized data output from WinXAS (version 3.1) software. Additional data included in the first 5 rows of table: date of analysis (date created), scientists, source file name, column headings.

Physical Structure Description:

Object Name: Arabinose-20060324-0171.csv

Text Format: Maximum Record Length: column

Simple Delimited: Field Delimeter: ,

Attribute(s) Info:

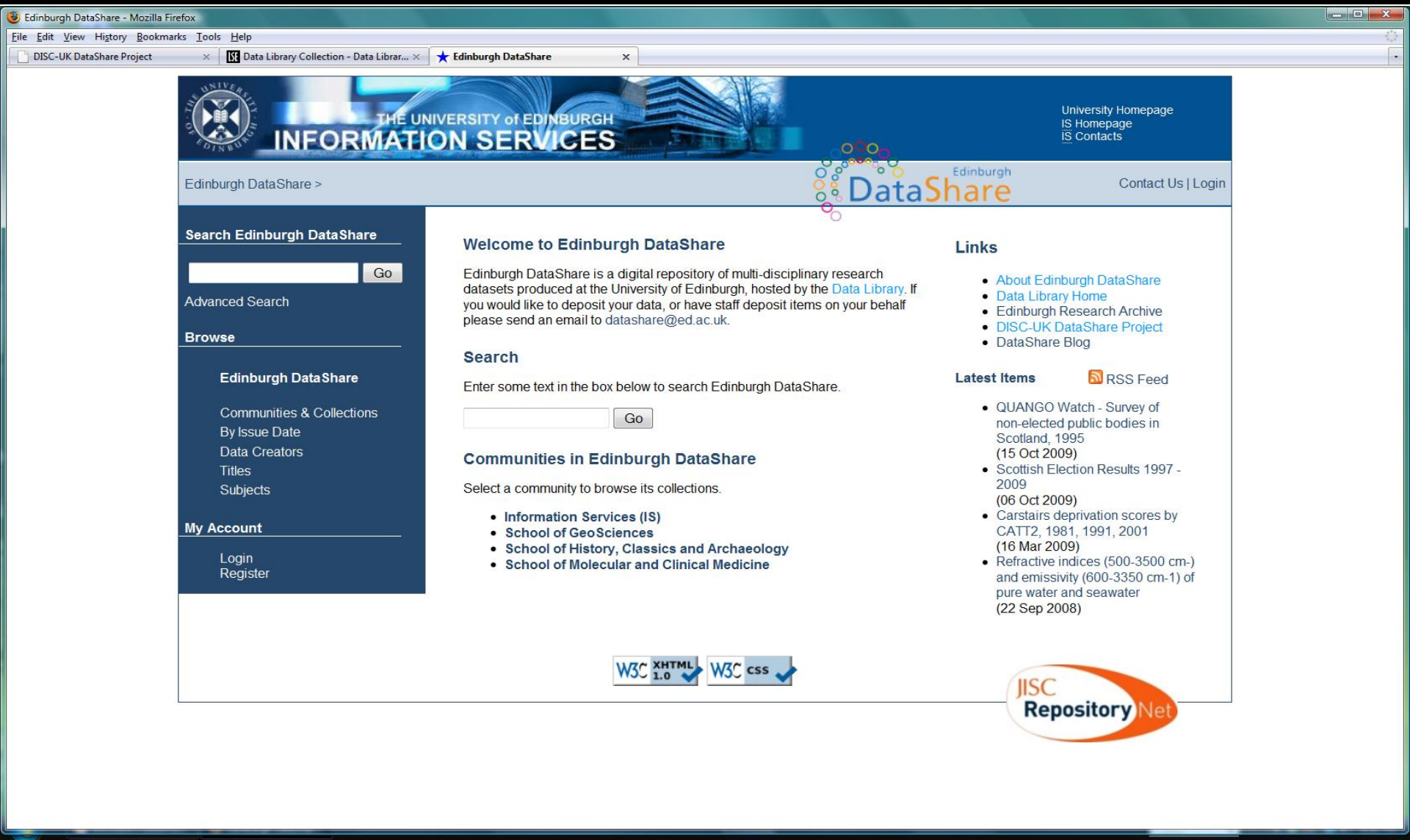
Attribute Name	Column Label	Definition	Type of Value	Measurement Type	Measurement Domain	Missing Value Code	Accuracy Report	Accuracy Assessment	CoverageMethod

Case Study: Disk-UK DataShare

Research Datasets in UK Data Repositories

http://www.disc-uk.org/datashare.html

DISC-UK DataShare project



Search Edinburgh DataShare

Go

Search Edinburgh DataShare
This Collection

Advanced Search

Browse

Edinburgh DataShare

Communities & Collections
By Issue Date
Data Creators
Titles
Subjects

This Collection

By Issue Date
Data Creators
Titles
Subjects

My Account

Login
Register

Refractive indices (500-3500 cm-) and emissivity (600-3350 cm-1) of pure water and seawater

Download Dataset

Show full item record

Title:	Refractive indices (500-3500 cm-) and emissivity (600-3350 cm-1) of pure water and seawater
Data Creator:	Owen Embury; Christopher Merchant; Mark Filipiak
Date Available:	2008-09-22
Citation:	Filipiak, Mark. (2008). Refractive indices (500-3500 cm-) and emissivity (600-3350 cm-1) of pure water and seawater [Dataset].
Permanent Identifier:	http://hdl.handle.net/10283/17
Dataset Description (abstract):	Tables of infrared refractive indices have been compiled and calculations made of emissivity for pure water and seawater (35 PSU), suitable for radiative transfer simulation of observations of thermal imagers, such as the series of Along Track Scanning Radiometers. The refractive indices are tabulated as a function of wave number (500–3500 cm-1) and temperature (274, 287, and 300 K). The emissivities are tabulated as a function of wave number (600–3350 cm-1), view angle (0–85°), temperature (270–310 K), and wind speed (0–25 m s-1 at 12.5 m).

Files	Description	Size	Format	View
ARCForwardEmissivityPureWater.nc		258.2Kb	application/x-netcdf	Download
ARCForwardEmissivitySeawater.nc		258.2Kb	application/x-netcdf	Download
ARCNadirEmissivityPureWater.nc		140.0Kb	application/x-netcdf	Download
ARCNadirEmissivitySeawater.nc		140.0Kb	application/x-netcdf	Download
ARCWideangleEmissivityPureWater.nc		356.6Kb	application/x-netcdf	Download
ARCWideangleEmissivitySeaWater.nc		356.6Kb	application/x-netcdf	Download
nk274_287_300_purewater.txt		198.2Kb	text/plain	Download
nk274_287_300_seawater.txt		198.2Kb	text/plain	Download
nk3Sigma.txt		78.16Kb	text/plain	Download
readme		1.937Kb	text/plain	Download

The following license files are associated with this item:

Depositor Agreement

This item appears in the following Collection(s)

Institute of Atmospheric and Environmental Science
Institute of Atmospheric and Environmental Science

CommunityCorner
- Director's Page
- Strategic Plan
- Unidata Leaflet
- Unidata Metrics Assessment
- 2009 Users Workshop
- Unidata Seminar Series
- CommunitE-letter
- Unidata Events
- Community Announcements
- Job Opportunities
- Acronyms List

ToolBox
- Downloads
- Data
- Software Tools
- Data-flow Quick Help
- Support
- Unidata Developer's Blog
- Mailing Lists
- RSS Feeds

DisplayAnalysis
- GEMPAK
- McIDAS
- IDV

DataAccess
- LDM
- IDD
- THREDDS
- OPeNDAP/DODS

DataManagement
- libcf
- netCDF
- netCDF Java
- UDUNITS

NetCDF

NetCDF (network Common Data Form) is a set of software libraries and machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data.

Getting Started with NetCDF

NetCDF is freely available (LICENSE). To build netCDF download the netCDF source distribution. The distribution contains the C/C++/F77/F90 libraries, and netCDF utilities ncgen and ncdump. See the 4.0.1 downloads page or the 3.6.3 downloads page for precompiled binaries.

- Installation instructions for C, Fortran, and C++ libraries
- NetCDF for Java
- Other interfaces to netCDF data: MATLAB, Objective-C, Perl, Python, R, Ruby, Tcl/Tk,
- Software for manipulating or displaying netCDF data
- Who uses netCDF?
- Developers may wish to download daily netCDF snapshot release, or see output from netCDF testing.
- Developers may wish to download the stable netCDF netcdf-4.1-beta2 release; this release will change periodically, but is more stable than the daily snapshots. It may also be useful to obtain the release notes. Note particularly that the 4.1 beta2 release now contains an internal OPeNDAP client.

NetCDF Documentation

- Frequently Asked Questions about netCDF
- Full NetCDF Documentation
- Writing NetCDF Files: Best Practices
- Conventions, example files and programs
- NetCDF Papers and Presentations
- NetCDF Credits

NetCDF Support

- NetCDF mailing list
- Subscribe to the netcdfgroup or netcdf-porting mailing lists
- Search or browse the netCDF support archives
- Search or browse the netcdfgroup mailing list archives
- Search or browse the netcdf-porting mailing list archives

Questions or comments can be sent to Unidata netCDF Support

NetCDF Build Troubleshooter

- Special instructions for Intel and Portland Group compilers.
- Current release known problems/workarounds
- Successful build output for tested platforms
- Successful builds on other platforms
- The usual build problems
- Build failure symptoms and resolution
- Troubleshooting build problems
- Reporting problems

Note: the netCDF-4.x releases will only build with the netCDF-4 enhanced features if "--enable-netcdf-4" is provided as a configure option. By default the distribution will build the classic netCDF-3 library. We believe that most users will not need the enhanced netCDF-4 features at this time.

NetCDF News and Announcements

Posted: 2009-03-30
NetCDF Workshop Materials Available: The materials from the 2009 NetCDF User Workshop are now available on line: 2009 NetCDF Workshop.

Posted: 2009-03-30
NetCDF 4.0.1 Release: We are pleased to announce the release of version 4.0.1 of the netCDF C/Fortran/C++ libraries. This release includes bug fixes and portability and performance enhancements. See the release notes for more information. Please send any feedback to support-netcdf@unidata.ucar.edu.

Posted: 2008-12-26
NetCDF User Survey: We would like to learn how you use netCDF and what features you would like to see in the future. Please take the netCDF user survey.

Posted: 2008-12-01
NetCDF Workshop On-line: The web pages from the Unidata 2008 workshop NetCDF for Data Providers and Developers are now available.

more news items >

Calculation of emissivity of pure water and seawater, 600-3350 cm-1

M. J. Filipiak, C. J. Merchant and O. Embury

International Journal of Remote Sensing, Vol. X, No. X, Month 2008, xxx-xxx

Refractive indices (n, k) in the wavenumber range 500 - 3500 cm-1 are in

nk274_287_300_purewater.txt
nk274_287_300_seawater.txt

with 3-sigma uncertainties in

nk3Sigma.txt

Emissivities for the ATSR/ATSR view angles are in

ARCNadirEmissivityPureWater.nc
ARCNadirEmissivitySeawater.nc
ARCForwardEmissivityPureWater.nc
ARCForwardEmissivitySeawater.nc

and for angles 0 - 90 at coarse resolution in

ARCWideangleEmissivityPureWater.nc
ARCWideangleEmissivitySeawater.nc

tabulated as follows (use ncdump -c <file name> to view the netCDF information)

```
 nadir view_angle = 0, 8, 16, 19, 21, 22, 23 ; (degrees)
 forward view_angle = 51, 51.5, 52, 52.5, 53, 53.5, 54, 54.5, 55, 55.5, 56, 56.5, 57 ; (degrees)
 wideangle view_angle = 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85 ; (degrees)

 wavenumber = 600, 700, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830,
    840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970,
    980, 990, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900,
    2000, 2100, 2200, 2300, 2360, 2370, 2380, 2390, 2400, 2410, 2420, 2430,
    2440, 2450, 2460, 2470, 2480, 2490, 2500, 2510, 2520, 2530, 2540, 2550,
    2560, 2570, 2580, 2590, 2600, 2610, 2620, 2630, 2640, 2650, 2660, 2670,
    2680, 2690, 2700, 2710, 2720, 2730, 2740, 2750, 2760, 2770, 2780, 2790,
    2800, 2810, 2820, 2830, 2840, 2850, 2860, 2870, 2880, 2890, 2900, 2910,
    2920, 2930, 2940, 2950, 2960, 2970, 2980, 2990, 3000, 3010, 3020, 3030,
    3040, 3050, 3060, 3070, 3080, 3090, 3100, 3110, 3120, 3130, 3140, 3240,
    3350 ; (cm-1)
 wind_speed = 0, 1, 3, 5, 10, 15, 20, 25 ; (m s-1)
 temperature = 270, 280, 290, 300, 310 ; (K)
```

There are no error estimates for the emissivity data.

Data Issues to Think About

Challenges of Access and Preservation for Research Data

Persistent identifiers
Storage space
Metadata and documentation
Long-term preservation
Data sharing

Identifiers for Datasets

Citable data identifiers enable linking data to publications

- e.g. DSpace uses Handles
- New DOI registration agency
  - German National Library of Science and Technology
  - For research datasets from technology/science and medicine
  - Also, the British Library, the Library of the ETH Zurich, the French Institute for Scientific and Technical Information (INIST), the Technical Information Center of Denmark and the Dutch TU Delft Library

Dataset Storage

- Locally-based (library, campus, consortia)
  or
- Cloud-based (S3, Azure, Atmos, etc.)

- Also for backup and replication (for digital preservation)

- Ideally policy-driven (e.g. # copies, locations, access guarantees)

Dataset Storage

eXtensible Access Method (XAM)

- Storage Network Industry Association of EMC, IBM, HP, Sun, Microsoft, etc.
- Defines standard access method (API) between "Consumers" (application and management software) and "Providers" (storage systems
- Provides industry-standard approach to storage-related policy metadata

Dataset Storage

Sun Open Archive Framework

e.g. Fedora + Sun Storage 7000 Unified Storage Systems and J4000 series as well as Solid State Disk technology to enhance performance as part of a Hybrid Storage Pool.

Features include *data integrity verification and repair* (19x9s), *checksumming and protection* (Raid, snapshot, clone), *analytics, predictive self-healing sensors, policy based data migration* and simplified management.

Metadata and Documentation

DSpace Metadata Schema for *Edinburgh DataShare*

Robin Rice, Stuart Macdonald & George Hamilton — Ver.1 (3. 7. 2008)

DSpace Form	DSpace ID	Dublin Core Element	DC Qualifier	Field Label	Input Type	Manda-tory?	Notes
1.1	1	contributor		**Depositor**	name	false	Hint: Enter the name of the person entering this record.
1.2	9	creator		**Data Creator**	onebox	true	Hint: Enter the names of the data creators / principal investigators of this item below. If the name of the data creator is unknown use the name of the responsible organisation. Example: University of Edinburgh. School of GeoSciences. Institute of Geography Required: You must enter a Data Creator for this dataset. Repeatable: true
1.3	64	title		**Title**	onebox	true	Hint: Enter the main title of the dataset. Required: You must enter a main title for this dataset. Repeatable: false
1.4	65	title	Alternative	**Alternative Title**	onebox	false	Hint: If the dataset has any alternative titles, please enter them here. Repeatable: true Initial Question: 1
1.5	27	description	abstract	**Dataset Descrip-tion (abstract)**	textarea	false	Hint: Please provide a summary description of the study in which the data was generated. Repeatable: false
1.6	66	type		**Type**	list	true	Hint: Select the type(s) of content of the item. To select more than one value in the list, you may have to hold down the "CTRL" or "Shift" key. Repeatable: true

http://www.disc-uk.org/docs/Edinburgh_DataShare_DC-schema1.pdf

Project Documentation

- **Title**

  name of the dataset or research project that produced it

- **Creator**

  names and addresses of the organization or people who created the data, including all significant contributors

- **Identifier**

  A unique identification number used to identify the data

- **Subject**

  keywords or phrases describing the subject or content of the data

Project Documentation

- **Dates**

  key dates associated with the data, including:

  project start and end date; release date;

  other dates associated with the data lifespan, e.g. maintenance cycle, update schedule

- **Funders**

  organizations or agencies who funded the research

- **Language**

  language(s) of the intellectual content of the resource, when relevant

Project Documentation

- **Location**

  where the data relates to a physical location, record information about its spatial coverage

- **Rights**

  description of any known intellectual property rights held for the data

- **List of file names and relationships**

  list of all digital files in the archive, with their names and file extensions (e.g. 'NWPalaceTR.WRL', 'stone.mov')

Additional Metadata

- **Formats**

  format(s) of the data, e.g. FITS, SPSS, HTML, JPEG

- **Methodology**

  how the data was generated, including equipment or software used, experimental protocol, other things you would include in your lab notebook. Can reference a published article, if it covers everything

- **Sources**

  references to source material for data derived from other sources, including details of *where* the source data is held, how *identified* and *accessed*

Additional Metadata

- **Versions**

  date/time stamped; use a separate identifier for each version

- **Checksums**

  to test if your file has changed over time (see backups)

Additional Metadata

- **Explanation of codes used in file names**

  brief explanation of any naming conventions or abbreviations used to label the files

- **List of codes used in files**

  list of any special values used in the data (e.g. '999 indicates a "dummy" value in the data')

Store metadata in a readme.txt file together with the data

Preserving Research Data

- Similar to other digitally archived information – best to use neutral, standard formats, keep lots of copies, lots of documentation
- Special concern for software to edit, process, render data
  - e.g. SPSS, netCDF, CATIA
  - Archiving software is hard (need source code, compilers, sometimes hardware)

Data Sharing

"The NIH expects and supports the *timely release* and sharing of *final research data* from NIH-supported studies *for use by other researchers*.

Starting with the October 1, 2003 receipt date, investigators submitting an NIH application seeking $500,000 or more in direct costs in any single year are expected to *include a plan for data sharing* or state why data sharing is not possible."

Data Sharing

IPR and data licenses

- Most data NOT copyrightable in the U.S.
  facts cannot be copyrighted
- Licenses (e.g. CC licenses) usually DO NOT APPLY and are not enforceable
- Should be placed in the *public domain* or not shared at all

This page is available in the following languages:
Afrikaans български Català Česko Dansk Deutsch Ελληνικά English English (CA) English (GB) English (Hong Kong) English (Singapore) English (US) Esperanto Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Español (Guatemala) Castellano (MX) Castellano (PE) Euskara Suomeksi français français (CA) Galego עברית hrvatski Magyar Italiano 日本語 한국어 Macedonian Melayu Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски srpski (latinica) Sotho svenska ไทย 中文 中文(香港) 華語 (台灣) isiZulu

creative commons

CC0 1.0 Universal

No Copyright

The person who associated a work with this document has dedicated the work to the Commons by waiving all of his or her rights to the work worldwide under copyright law and all related or neighboring legal rights he or she had in the work, to the extent allowable by law.

Works under CC0 do not require attribution. When citing the work, you should not imply endorsement by the author.

Free Cultural APPROVED FOR Works

Other Rights — In no way are any of the following rights affected by CC0:

- Patent or trademark rights held by the person who associated this document with a work.
- Rights other persons may have either in the work itself or in how the work is used, such as **publicity** or privacy rights.

This is a human-readable summary of the Legal Code (read the full text). Disclaimer

Use this tool for your own work.

Data Sharing

Library-managed Institutional Repositories are a great way to share data. They support:

- Citation
- Open Access or embargoes, as needed
- Long-term access
- Discovery (via metadata, in IR or Google, etc.)

Cyberinfrastructure

And the Role of Libraries

Research Cyberinfrastructure

"Cyberinfrastructure includes computing cycles, but also broadband networking, massive storage, and *managed information*"

"These data can be observational inputs, experimental values, or results of calculations, images, or videos."

2003 Atkins Report to the NSF

Research Cyberinfrastructure

2005 National Science Board published report

"*Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*"

NSF creates Office of Cyberinfrastructure

"The Office of Cyberinfrastructure coordinates and supports the acquisition, development and provision of state-of-the-art cyberinfrastructure resources, tools and services essential to the conduct of 21st century science and engineering research and education."

Digital data curation is a major theme

Research Cyberinfrastructure

2007

ARL Agenda for Developing E-Science in Research Libraries

2008

NSF Blue Ribbon Task Force on Sustainable Digital Preservation and Access

DISC-UK DataShare: Data Sharing Continuum

Distributed high performance computing; analysis tools applied to data over secure international network; M2M interfaces	Data Grid
Peer review of datasets; seamless link to publications; role-based layers of access; data overlay journals	Data publishing
Graphs, charts, maps configurable online	Data visualisation
"Actionable" marked up dataset installed in a data browser tool subsetting capability	Data manipulation online
Original format plus XML markup of data or XML database; open standards used appropriate to domain; metadata or setup files may be bundled with dataset for importing elsewhere	Data enhanced for re-use
Quality assured metadata; guidance available for depositors; suitably anonymised/consent for sharing obtained from subjects; thorough documentation about data creation and methodology included; permanent IDs; formats validated and suitable for distribution; migration-based preservation commitment	Network of distributed repositories: subject and/or institutionally based
Data files with minimal documentation (e.g. readme file describing each data file) downloadable from Internet	Zip and ship Open access
Metadata record of dataset on website or in repository; possibly with embargo and contact information to request access	Search and discovery enabled; restricted access
Networked drive, available to research group, version control	Email dissemination by request
Password protected, networked drive (backup procedures)	Privileged access
Personal hard drives, un-networked	Simple data storage

Robin Rice, September 2007

Library's Role

- Data organization and annotation

  e.g. ontologies and metadata

- Data curation

  e.g. long-term storage and preservation, publishing for discovery and access

- *Outreach and support to local researchers*

Library's Role

- Libraries are well-positioned
- Need to start now, build over time
- Build expectation that libraries will play a role

"Even if you're on the right track,
you'll get run over if you just sit there."

Will Rogers