

University of Massachusetts Medical School

eScholarship@UMMS

University of Massachusetts and New England
Area Librarian e-Science Symposium

2010 e-Science Symposium

Apr 7th, 12:00 AM

Bringing Research Data into the Library: MIT's Experience

Mackenzie Smith

Massachusetts Institute of Technology

Follow this and additional works at: https://escholarship.umassmed.edu/escience_symposium



Part of the [Library and Information Science Commons](#)



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

Repository Citation

Smith, M. (2010). Bringing Research Data into the Library: MIT's Experience. *University of Massachusetts and New England Area Librarian e-Science Symposium*. <https://doi.org/10.13028/fbak-w002>. Retrieved from https://escholarship.umassmed.edu/escience_symposium/2010/program/6

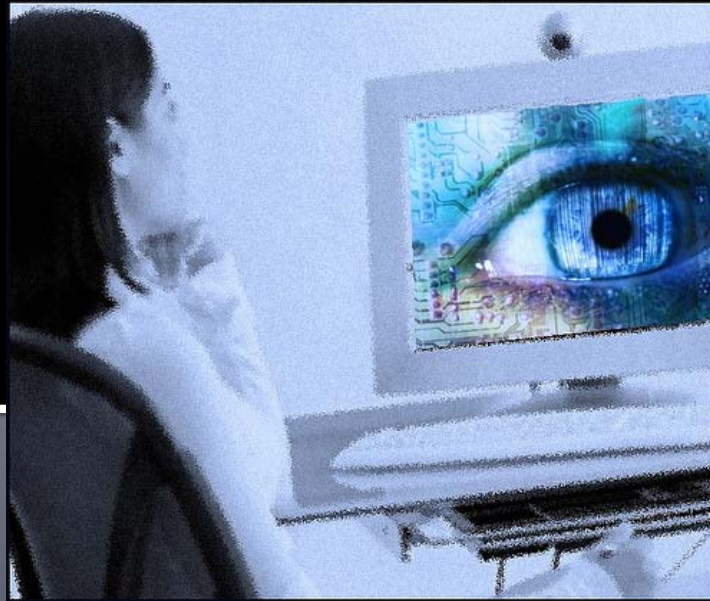
Creative Commons License



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

Bringing Research Data into the Library: MIT's Experience



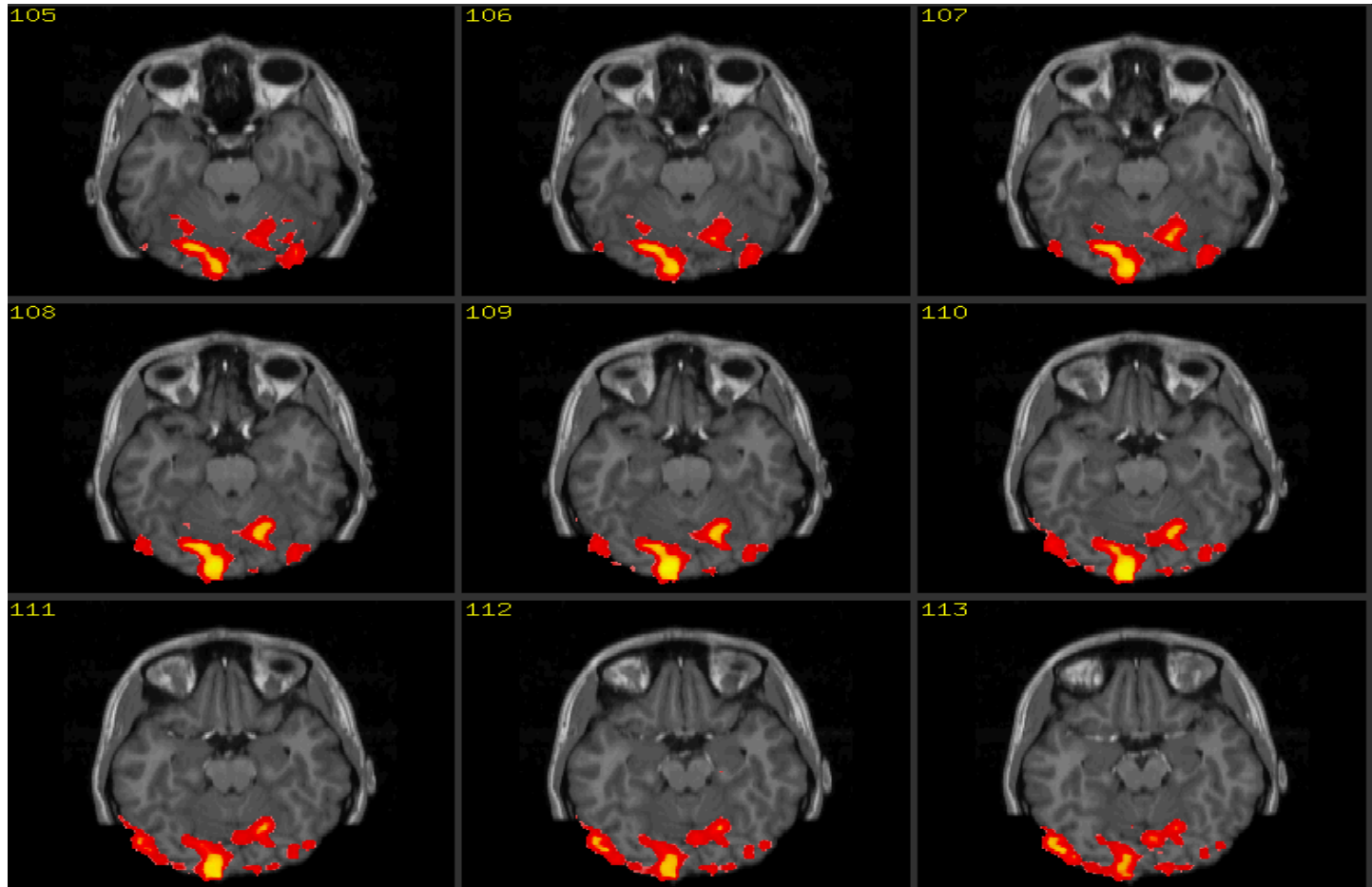
How Much Information? @MIT

Case studies for

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

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

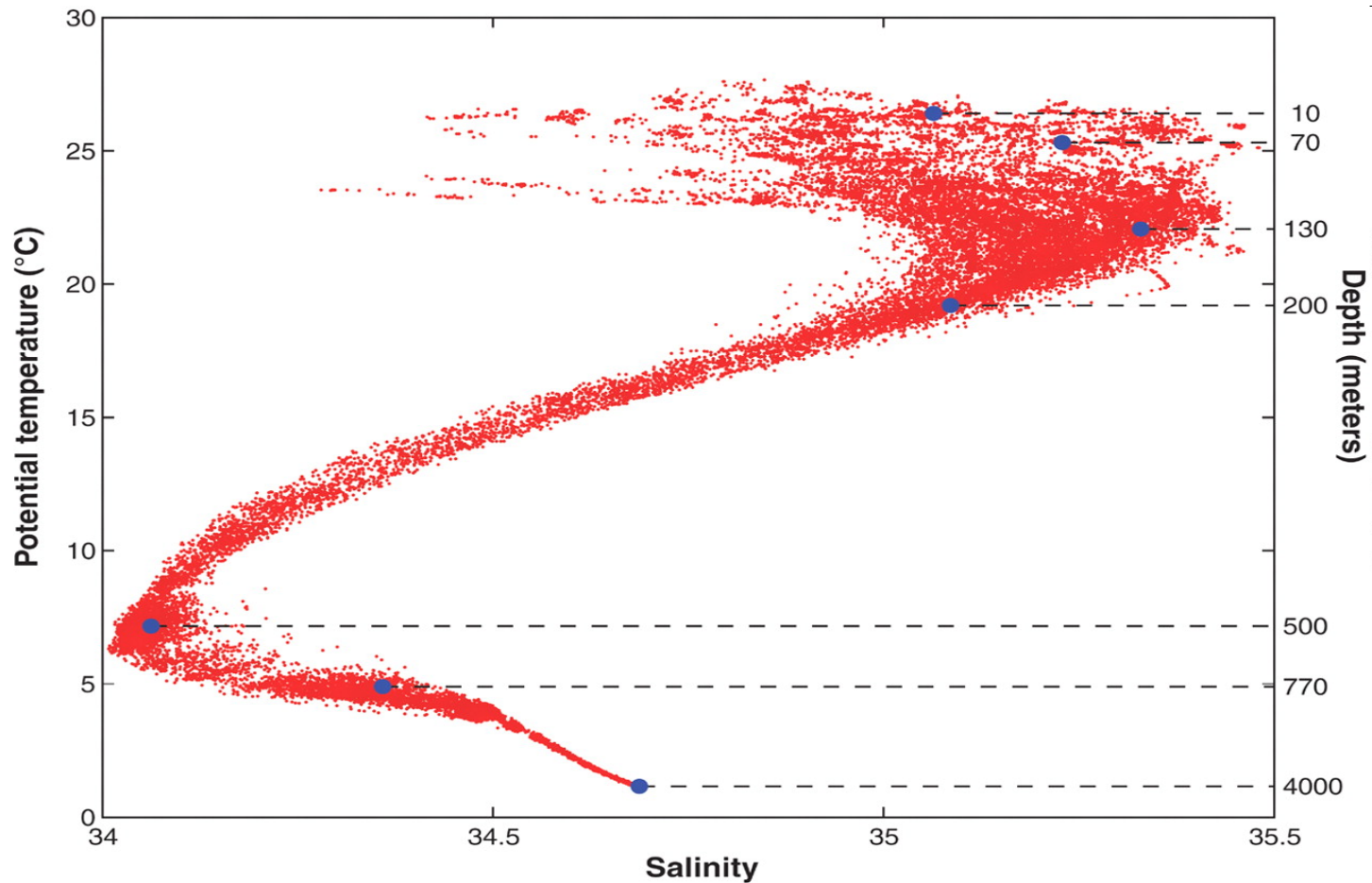
Neuroimaging Case Study



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. for the same subject
- LARGE-SCALE
>10Tb per year for one group of 4 faculty
- EXPENSIVE
each subject ~\$1000 (1500/year, per machine)
- HARD TO LOCATE, 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* metagenomic data *with* biochemical sensor data, including water chemistry, optical properties, physical data (e.g. sample location)
- LARGE-SCALE
Solexa sequencer produces 1Tb per run X 2-3 runs/week
- IRREPLACEABLE
time dependent, not fully analyzable today
- HARD TO LOCATE
no integrated DB exists (e.g. GenBank only takes sequence data)

MIT Libraries and Data

Building on *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](#)

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

Subject Expert

Katherine McNeill
Data Services and Economics Librarian
[Dewey Library](#)

mcneilh@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 \(MIT only\)](#)



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



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](mailto:gishelp(at)mit.edu)

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](mailto:gishelp(at)mit.edu) to check if a GIS specialist can meet with you at a different time A brief description of

Library Roles in Data Curation

- Acquire and catalog faculty-authored and externally-published datasets
- Consultation and training services
- Liaise with national archives (e.g. ICPSR)
- Develop standards (e.g. DDI, FGDC)
- Maintain website, data repositories

Data Management and Publishing

[Home](#)
[Why Manage Your Data?](#)
[What is Data?](#)
[Evaluate Your Data Needs](#)
[Documentation and Metadata](#)
[File Formats](#)
[Organizing Your Files](#)
[Backups and Security](#)
[Sharing Your Data](#)
[Citing Data](#)
[Data Integration](#)
[Funding Requirements](#)
[Ethical and Legal Issues](#)
[Managing Research Data 101 \(pdf\)](#)

Related Guides:

[Bioinformatics](#)
[GIS Services](#)
[Metadata Services](#)
[Social Science Data Services](#)

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](#)

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

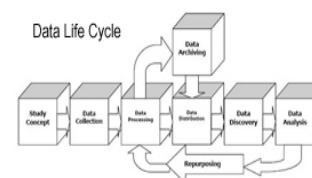
Planning for your data management needs ahead of time will save you time and resources in the long run and ensure that your data will be usable in the future. A formal plan can be valuable to you and may be required by your funding agency. Topics to cover in a written plan include:

- 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



Source: DD Structural Reform Group. "DD Version 3.0 Conceptual Model." DOI Alliance 2004. Accessed on 11 August 2008. <http://www.igpp.uic.edu/DD/committee/etd/ConceptModel%30.pdf>.

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
 - Led by liaisons for Civil and Environmental Engineering, Computer Science, GIS, Social Science
 - Broader group includes Bioinformatics, Chemistry, Humanities, others
- Consultation from Metadata Services specialists, IT 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

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
- Loaded data from previous repository and added CC.o licenses
- Non-MIT submissions reviewed by community owner or library staff

Search DSpace@MIT

Home ♦ [Robotics Research Datasets](#)

Radish: The Robotics Data Set Repository

Search within this collection:

[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)



All Items in DSpace@MIT are protected by original copyright, with all rights reserved, unless otherwise indicated.



Search DSpace@MIT

[Home](#) ♦ [Robotics Research Datasets](#) ♦ [Radish: The Robotics Data Set Repository](#)

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](#)

All Items in DSpace@MIT are protected by original copyright, with all rights reserved, unless otherwise indicated.



MASSACHUSETTS INSTITUTE OF TECHNOLOGY



Contact MIT Libraries

Search DSpace@MIT

 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](#)

DSPACE@MIT

Search DSPACE@MIT

Home > Robotics Research

Edmonton_Conve

Show full item record

Citable URI: <http://hdl.handle.net/1721.1/11111>

Title: Edmonton

Author: Roy, Nick

Description: This is the file, and is

URI: <http://hdl.handle.net/1721.1/11111>

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



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 18.70 51.15 51.15 51.15 15.30 51.10 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 21.60 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 24.1
05 13.05 51.15 51.15 13.25 13.15 51.15 51.15 51.15 51.15 13.50 51.15 13.15 13.15 13.15 13.15 13.25 12.25 11.85 11.75 11.90 13.15 15.12 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 15.13 13.30 51.15 51.15 51.15 51.15 51.15
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 3.33 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 5.40 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 21.45 18.70 18.40 18.40 18.20 18.30 18.25 18.70 51.15 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 1.6945410DOM -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 3.23 3.18 -25.75
.20 18.25 18.85 51.15 51.15 51.15 15.30 51.15 51.15 51.10 51.10 51.15 51.15 10.60 10.00 9.50 9.05 8.45 8.10 7.70 7.40 7.15 6.90 6.65 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 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
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 1.9932240DOM -25.759414 -10.540073 -3.491340 0.000000 0.000000 1028099446.636618 xavier 2.079755F
4.85 4.80 4.70 4.45 4.40 4.35 4.25 4.20 4.05 3.90 3.85 3.85 3.75 3.70 3.63 3.55 3.58 3.53 3.48 3.43 3.38 3.33 3.28 3.18 -25.759414
45 18.95 51.15 51.15 51.15 15.40 51.15 51.15 51.15 51.15 51.15 51.15 51.15 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 51.15 51.15
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 51.15 51.15 11.80 11.65 11.95 12.30 12.00 12.00 11.65 11.35 11.50 11.55 51.15 11.85 12.05 12.15 12.15 12.95 51.15 13.21 51.15 13.0
8099447.264384 xavier 2.7079240DOM -25.759414 -10.540073 -3.491340 0.000000 0.000000 1028099447.268331 xavier 2.711468FLASER 180 11.10 12.40
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 3.28 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 51.15 12.15 12.15 12.15 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 3.4352880DOM -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
20 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 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 13.10 13.20 51.15 51.15 51.15 51.15 13
6510DOM -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 3.28 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 10.45 10.60 10.80 11.05 4.50 4.40 4.40 4.40 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 13.50 51.15 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 11.90 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 5.2415170DOM -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 3.38 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 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 3.33 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 8.88 8.38 8.09 7.69 7.39 7.14 6.89 6.64 6.29 6.09 5.94 5.79 5.54 5.44 5.24 5.09 4.94 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
```

om the log

T Libraries

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: software accompanying
Master's Thesis

*"Analysis of Targeted and Combinatorial
Approaches to Phage T7 Genome Generation"*

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

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

- Local (library, campus, consortia)
or
- Cloud (Amazon S3, Azure, Atmos, etc.)
- Also for backup and replication (for digital preservation)
- Ideally policy-driven (e.g. # copies, locations, access guarantees)

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	Mandatory?	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 Description (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

Preserving Data

- Similar to other digital archives

neutral, standard formats are best; keep lots of copies, lots of documentation

- Special concern for *software*

- e.g. SPSS, netCDF, CATIA

- Archiving software is difficult

(need source code, compilers, sometimes hardware)

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 বাংলা Azərbaycanca 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 (latinica) Sotho svenska বাংলা 中文 中文(香港) 華語 (台灣) isiZulu



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.



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.

The Library's Role

Summing up

Data organization and annotation
e.g. ontologies and metadata

Data curation
e.g. long-term storage, preservation, publishing

Outreach and support to local researchers

