Caltech CODA

Eric F. Van de Velde
California Institute of Technology
evdv@library.caltech.edu

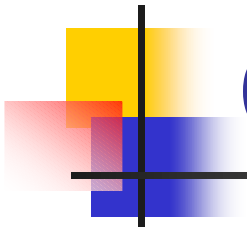
Open Archives Workshop
Oct. 17-19, 2002
CERN, Geneva

Caltech CODA

- Collection of Open Digital Archives: CODA

  Mus. A passage of more or less independent character introduced after the completion of the essential parts of a movement, so as to form a more definite and satisfactory conclusion. -- OED

  - http://coda.caltech.edu
- Over 1,100 documents available now!
  - Theses
  - Technical Reports
  - Conference Proceedings
  - Book
  - Non-Research Collections

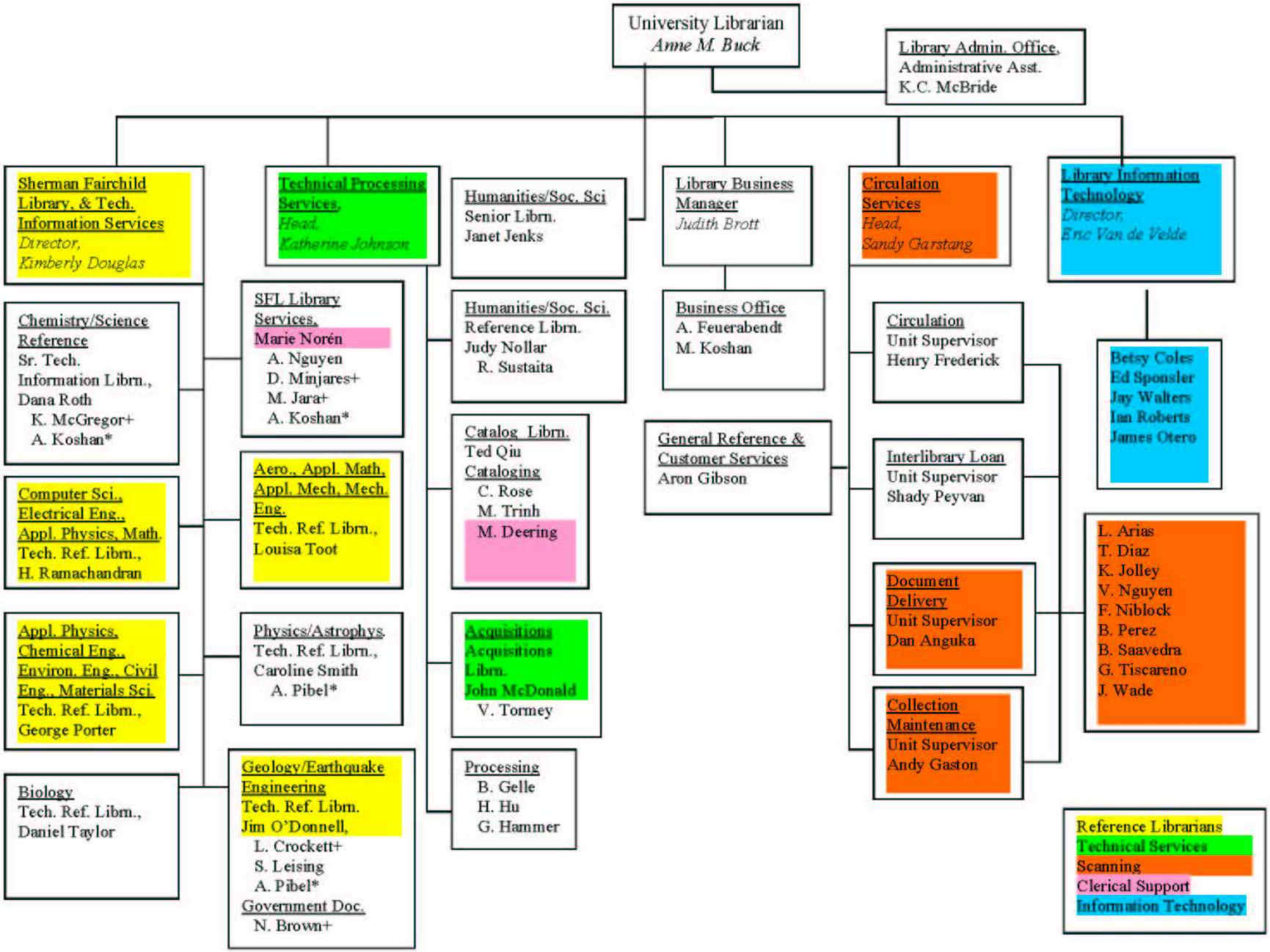


Caltech

- 285 Professorial faculty
- 900 Undergraduate students
- 1100 Graduate students
- 1000 Post-Doctoral researchers

Caltech Library System

- 55 FTEs, 14 Librarians, 6 IT Staff
- 500,000 Volumes
- 3556 Paid Print-Journal Subscriptions
- 2116 Paid E-Journal Subscriptions
- 88 Online Databases
- 160 Workstations, 100 Public
- 4 Major Libraries + 2 Locations
- $6,370,000 Annual Budget (FY 02)


University Librarian
Anne M. Buck
Library Admin. Office,
Administrative Asst.
K.C. McBride
Sherman Fairchild Library, & Tech. Information Services
Director,
Kimberly Douglas
Technical Processing Services,
Head,
Katherine Johnson
Humanities/Soc. Sci
Senior Librn.
Janet Jenks
Library Business Manager
Judith Brott
Circulation Services
Head,
Sandy Garstang
Library Information Technology
Director,
Eric Van de Velde
Chemistry/Science Reference
Sr. Tech. Information Librn.,
Dana Roth
K. McGregor+
A. Koshan*
SFL Library Services,
Marie Norén
A. Nguyen
D. Minjares+
M. Jara+
A. Koshan*
Humanities/Soc. Sci.
Reference Librn.
Judy Nollar
R. Sustaita
Business Office
A. Feuerabendt
M. Koshan
Circulation
Unit Supervisor
Henry Frederick
Betsy Coles
Ed Sponsler
Jay Walters
Ian Roberts
James Otero
Computer Sci., Electrical Eng., Appl. Physics, Math.
Tech. Ref. Librn.,
H. Ramachandran
Aero., Appl. Math, Appl. Mech, Mech. Eng.
Tech. Ref. Librn.,
Louisa Toot
Catalog Librn.
Ted Qiu
Cataloging
C. Rose
M. Trinh
M. Deering
General Reference & Customer Services
Aron Gibson
Interlibrary Loan
Unit Supervisor
Shady Peyvan
Appl. Physics, Chemical Eng., Environ. Eng., Civil Eng., Materials Sci.
Tech. Ref. Librn.,
George Porter
Physics/Astrophys.
Tech. Ref. Librn.,
Caroline Smith
A. Pibel*
Acquisitions
Acquisitions Librn.
John McDonald
V. Tormey
Document Delivery
Unit Supervisor
Dan Anguka
L. Arias
T. Diaz
K. Jolley
V. Nguyen
F. Niblock
B. Perez
B. Saavedra
G. Tiscareno
J. Wade
Biology
Tech. Ref. Librn.,
Daniel Taylor
Geology/Earthquake Engineering
Tech. Ref. Librn.
Jim O'Donnell,
L. Crockett+
S. Leising
A. Pibel*
Government Doc.
N. Brown+
Processing
B. Gelle
H. Hu
G. Hammer
Collection Maintenance
Unit Supervisor
Andy Gaston
Reference Librarians
Technical Services
Scanning
Clerical Support
Information Technology

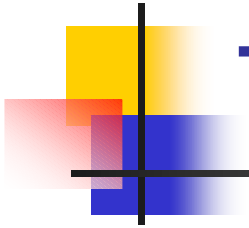
CLS OAI Team

- Info. Tech.:
  - B. Coles
  - E. Sponsler
- Scanning:
  - S. Garstang
- Technical Services:
  - K. Johnson
- Librarians:
  - K. Douglas
  - J. O'Donnell
  - G. Porter
  - H. Ramachandran
  - L. Toot*
- Support Staff

Theses

Theses

- NDLTD: Networked Digital Library of Theses and Dissertations
  - Caltech involved since 1998
  - http://www.ndltd.org/
  - ETD 2003, May 21-24, Berlin, Germany
- **E-submission compulsory July, 2002**
- http://etd.caltech.edu/ETD-db/ETD-browse/browse?first_letter=all
- 294+ Theses available (49 restricted)



Theses

- Recruitment
  - Administration: early supporters
  - Faculty: concerned about publication restrictions
  - Students: very interested
- Implementation
  - Install and maintain NDLTD software
  - Set up procedures with Graduate Office
  - Check metadata, document formats
  - Seminars on Copyright, Authoring tools, Metadata

Scanning Theses

- Flood damaged 200 circulating theses
- Scan archival copies
- Gradual retrospective conversion
- Issues:
  - Training: work flow similar to e-delivery
  - Software and hardware compatibility
  - Author permission (1978, ©,…) and open access
- Cost:
  - Scan during e-delivery off-peak hours
  - Metadata, abstracts (Librarian Time)

Technical Reports

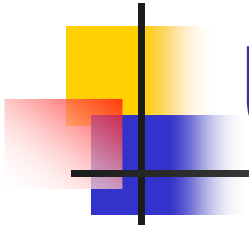
Faculty Self-Publishing (1999)

- 44 Engineering and Applied Science
- 3 Chemistry and Chemical Engineering
- 3 Physics, Mathematics, Astrophysics
- 3 Biology
- 2 Geology and Planetary Science
- 1 Humanities and Social Sciences

- Supplement Self-Publishing with Institutional Repositories

Technical Reports: Policies

- Voluntary participation
- Repository for any Research Unit
- RU = Professor, Option, Division,...
- RU controls policy on who may submit
- RU sets up internal approval process
- Metadata openly available
- Submission is permanent
- Non-exclusive permission to distribute



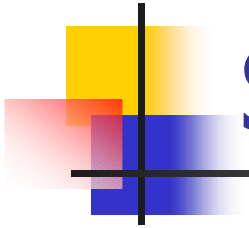
Use License

I hereby grant to [Caltech] the **irrevocable, non-exclusive royalty free right to reproduce, distribute, display, and perform this work in any format** including electronic formats throughout the world **for educational, research and scientific non-profit uses** during the full term of copyright including renewals and extensions via the Digital Collections mechanisms maintained by the Caltech Library System. I also hereby grant to Caltech the non-exclusive right to sub-license these rights to others should the Institute forego the ability to maintain distribution. I warrant that I have the copyright to make this grant to Caltech unencumbered and complete.

Once this paper is so published, it may not be withdrawn. With the approval of the repository administration revisions to available documents within this service will be accepted.

The following Notice Concerning Terms and Conditions of Use will be included with the electronic distribution copies of the work: *You are granted permission for individual, educational, research and non-commercial reproduction, distribution, display and performance of this work in any format.*

http://resolver.library.caltech.edu/caltechCSTR:2001.000a



Software: http://www.eprints.org/

- S. Harnad, R. Tansley, and C. Gutteridge
- Open Archives Initiative
- Free
  - Open Source: GNU General Public License (GPL)
  - Built on GPL software: Linux, Apache, MySQL, Perl,…
- Easy
  - Browsing and searching (readers)
  - Submission procedure (authors)
  - Document management (librarians)
  - System administration
- Getting better
  - EPRINTS-2-ALPHA-2 released Jan. 9th, 2002

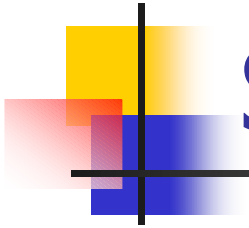
Computer Science

- http://caltechcstr.library.caltech.edu/
- Converted the NCSTRL database
- Scanned older print reports
- Submission process for new reports
- Report numbers
- Add, correct metadata
- 368 Reports available

Earthquake Engineering

- http://caltecheerl.library.caltech.edu/
- A historical collection
- All scanned
- OCR abstracts from TIFF files
- Few additional submissions anticipated
- 287 Reports available

Parallel and Distributed Systems Group



- http://caltechparadise.library.caltech.edu/
- Took over existing digital collection
- Anticipate ongoing submission
- 45 Reports available

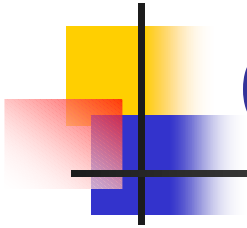
Library Reports

- http://caltechlib.library.caltech.edu/
- New submissions only
- 15 Reports available

In Development

- Graduate Aeronautical Labs.
  - Fluid-mechanics
  - Solid-mechanics
- Environmental Engineering
- Control and Dynamical Systems
- Mechanical Engineering

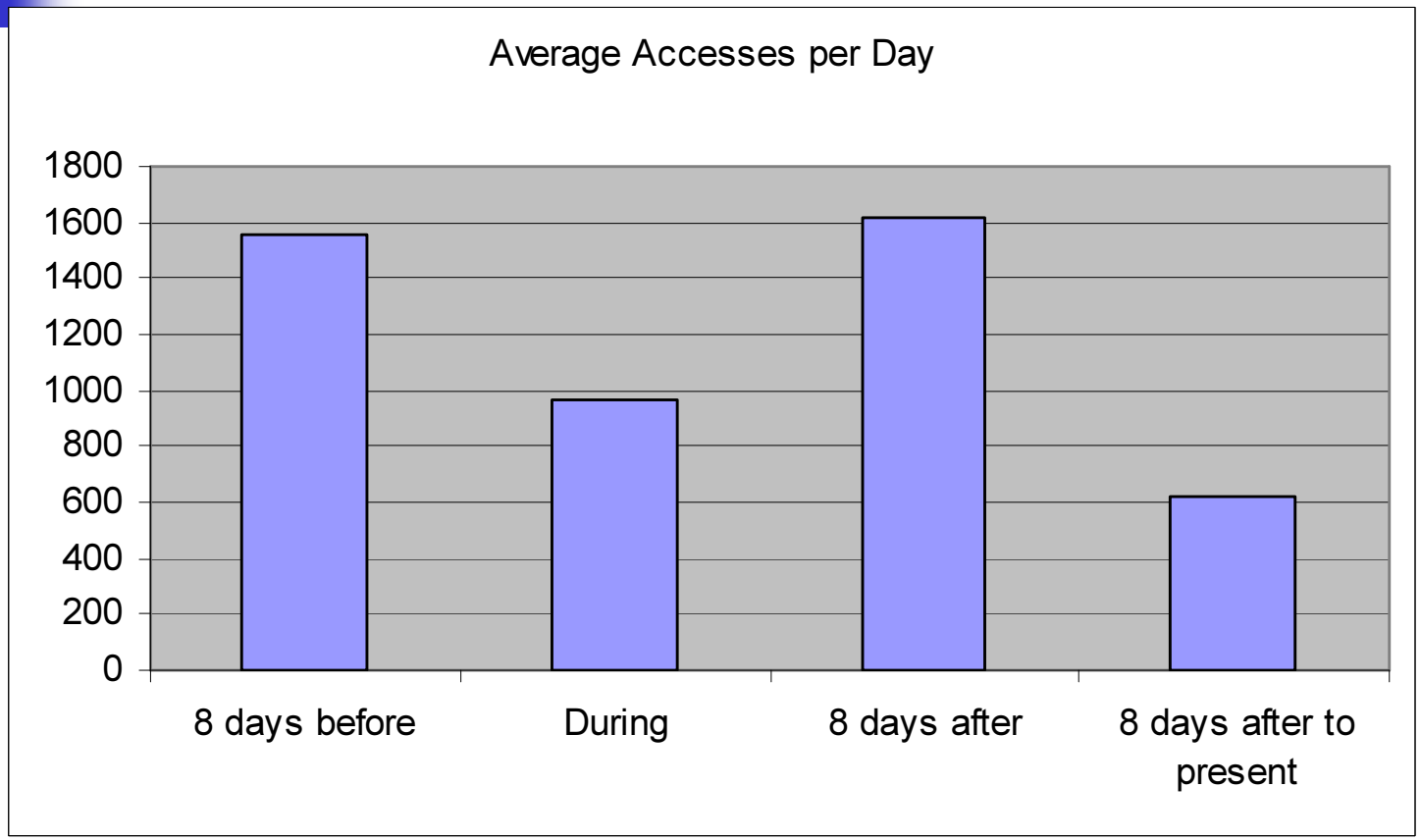
Conference Proceedings



Conference Proceedings

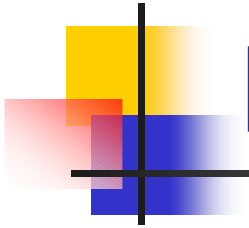
- Cavitation 2001
  - http://cav2001.library.caltech.edu/
- A one-time favor
- Very successful
- CAV 2003 considering similar model
- Next time, we would charge
- 110 Reports available

CAV 2001 Usage



Roughly half of the accesses represent PDF downloads. Human users comprise 88% and search engine robots 12% of total accesses based on browser type.

Books



Books

- Archive contains 1 book:
  - Brennen, Christopher Earls. Cavitation and Bubble Dynamics. New York : Oxford University Press, 1995.
  - Copyright waiver from publisher under special arrangement with author
- Ready for more books

Non-Research Repositories

Non-Research Repositories

- Oral Histories
  - A project of the Caltech Archives
  - Archives control the repository
  - Library maintains the repository
- Engineering and Applied Science
  - PR journal for friends and alumni
  - PR controls the repository
  - Library maintains the repository

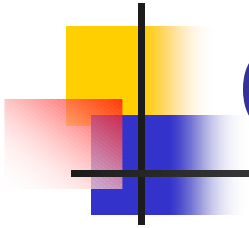
Technical/Managerial Issues

Persistent URLs

- http://resolver.library.caltech.edu/CaltechBOOK:1995.001
- Our resolver keeps track of location
- Insurance against future software changes
- Cheaper than DOIs…

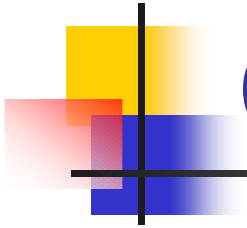
Many Repositories

- More work to set up and maintain, BUT:
- Isolated administrative control:
  - Who vets documents for content and format?
  - Who assigns report numbers?
  - Who sets access and harvesting policies?
  - Who decides on metadata?
- No endless “big-picture” debates
- Software choice (Eprints, NDLTD)
- Customized look and feel
- Other views through federation



Conclusions (1)

- Acquire technology
  - Need Linux applications developer
  - Begin small (obsolete workstation)
  - Be ready to grow fast (acquire server)
- Train library staff
  - Find in-house projects (NCSTRL✹ Eprints)
- Public Relations:
  - Publicize the service
  - Recruit leaders in departments
  - Remind on a continuing basis
  - Exploit all informal conversations



Conclusions (2)

- Educate
  - Copyright (fear is the enemy)
  - Effective use of multimedia
  - Authoring tools
- Maintain stable and reliable systems
  - Win or lose an advocate on first use!
- Choose projects
  - Initially, accept whatever comes your way
  - Later, set reasonable acceptance criteria.

Conclusions (3)

- Struggling
  - All faculty-initiated submissions
  - Voluntary student-initiated submissions
  - Fee for service
- Very successful
  - Compulsory student submissions (theses)
  - Library-administered collections
  - Scanning printed material (unexpected)
- Recommendation: Get all theses online! Join NDLTD!