

CD-ROM Redigitization Project (CDRiP)

Electronic Resources and Libraries Conference

February 22, 2007

Harrison Dekker

Coordinator of Data Services

Doe and Moffitt Libraries, UC Berkeley

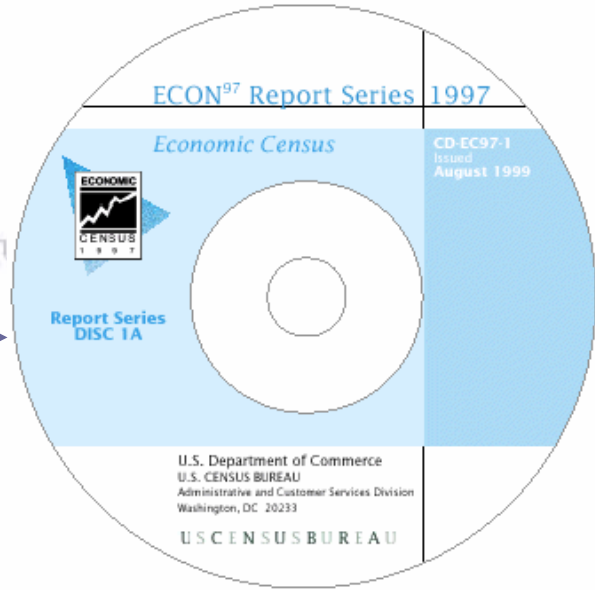
<codeBook>
</codeBook>



-1961 </titl>
h:edu/cocoon/ICPSR-STI

-1961 </titl>
(Kuhnie </AuthEnty>
Ent100>
t700>

stitute for Social Research
al and Social Research



File Edit View Go Bookmarks Tools Help

Go

Disable CSS+ Forms Images Information Miscellaneous Outliner Resizer View Source Options

ICPSR INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH [Need Help?](#)

DIRECT

About ICPSR Data Access & Analysis Data Use Tutorial Depositing Data Education & Instruction Research at ICPSR Special Topic Archives

MyData: Login/Account Info - Data Cart - Logout ICPSR Home - Contact Us - OR Site - Search

Data Access & Analysis

ICPSR supplies data files for use with statistical software, such as SAS, SPSS, and Stata. The links below provide various ways of finding the data. For instructions on accessing and using the data, visit our [Data Use Tutorial](#).

See Also

- Citing Electronic Data Files
- IRB Information
- Responsible Use Statement
- Access Non-Publicly Available Files
- Online Analysis Using the DAS

Search the Web Site & Data Holdings

Web Site & Data Holdings Related Literature

in all fields

Web Site Data Holdings Search

[Advanced Search](#)
[Search Tips](#)

The Searching FAQ provides search tips and detailed explanations on how our search engine works.

Done

Background: CD-ROMs and Library Data Services

<distDate date="1984-06-20">1984-06-20</distDate>

</distStmnt>

<serStmnt> </serStmnt>

<version date="1992-02-16">1992-02-16</version>

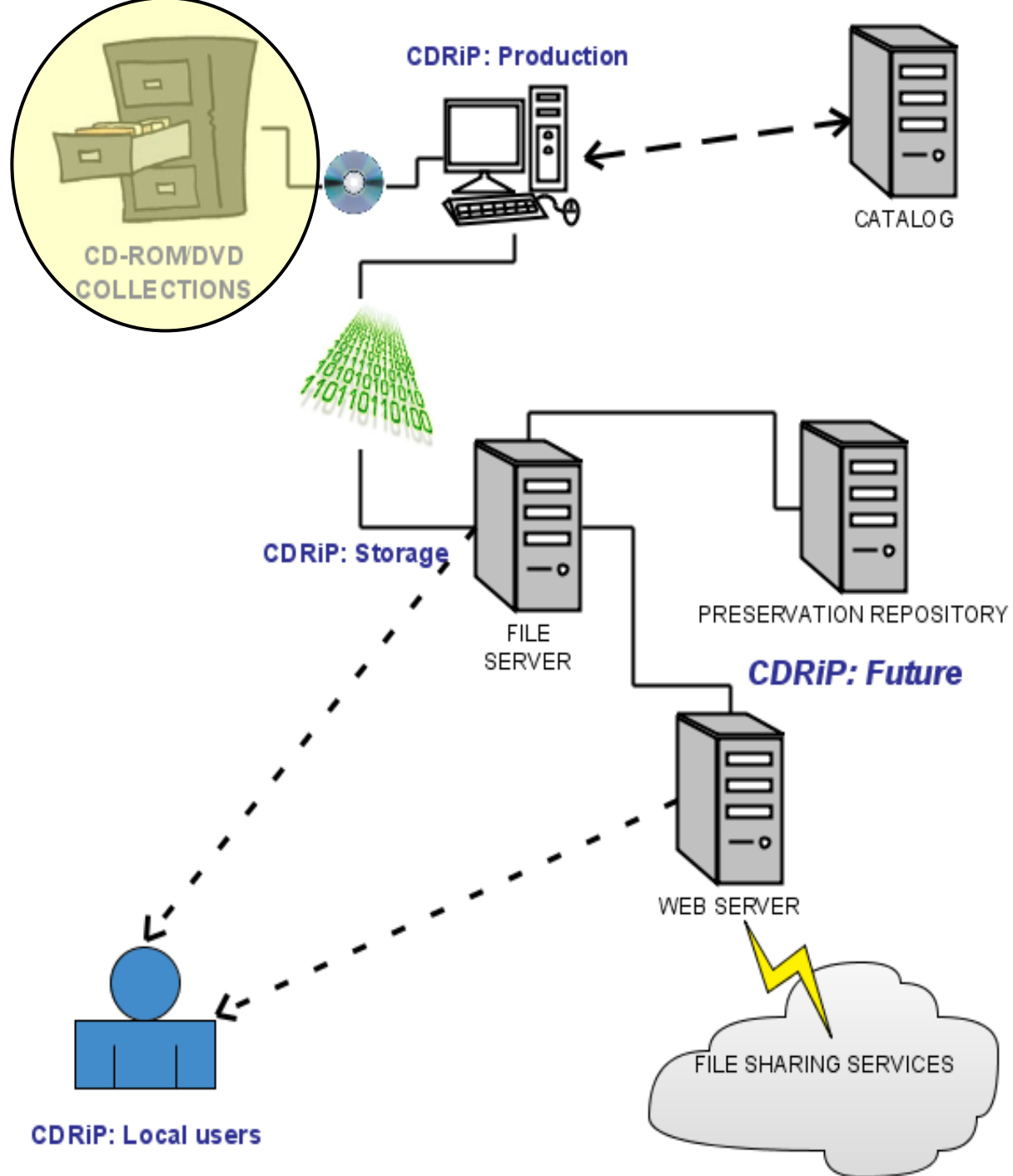
Aarebrot, Frank, and Stein Kuhnie. NORWEGIAN ECOLOGICAL
Bergen, Institute of Sociology. ICPSR ed. Ann Arbor, MI: Inter-univ
[distributor], 1998.

</bibliCit>

<notes>LRECL and Card Image</notes>

</citation>

<stdyinfo>



Why do it?

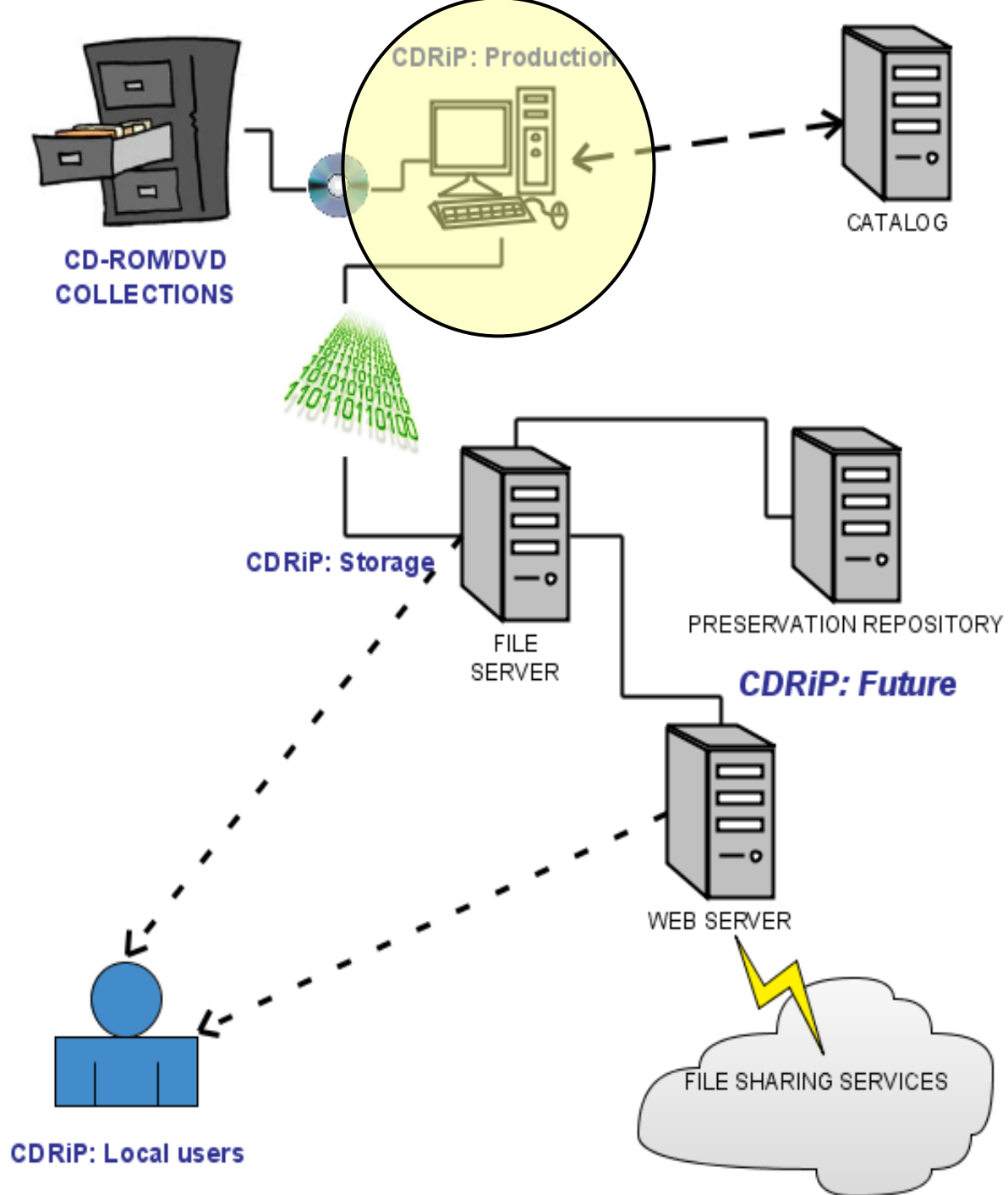
Pros:

- Convenience
- Better access
- Preservation
- Hardware availability

Cons:

- Copyright issues
- Low use (?)

```
<codeBook>
- <docDscr>
- <citation>
- <titlStmnt>
  <titl>Norwegian Ecological Data, 1949-1961 </titl>
</titlStmnt>
- <verStmnt>
  <version date="2003-12-18"/>
</verStmnt>
  <holdings URI="http://webapp.icpsr.umich.edu/cocoon/ICPSR-STUDY/00040.xml"/>
</citat>
</docDscr>
- <stdyDscr>
- <citation>
- <titlStmnt>
  <titl>Norwegian Ecological Data, 1949-1961 </titl>
  <IDNo>40</IDNo>
</titlStmnt>
- <rspStmnt>
  <AuthEnty>Aarebrot, Frank, and Stein Kuhnle </AuthEnty>
  <AuthEnt100>Aarebrot, Frank </AuthEnt100>
  <AuthEnt700> </AuthEnt700>
</rspStmnt>
<prodStmnt> </prodStmnt>
- <distStmnt>
- <distrbtr>
  <distrbtr>Inter-university Consortium for Political and Social Research, University of Michigan" URI="http://www.icpsr.umich.edu">
  Inter-university Consortium for Political and Social Research
</distrbtr>
- <distDate>
  <distDate date="1984-06-20">1984-06-20</distDate>
</distStmnt>
<serStmnt> </serStmnt>
- <verStmnt>
  <version>
  <version>0</version>
  <version>1</version>
  <version>2</version>
  <version>3</version>
  <version>4</version>
  <version>5</version>
  <version>6</version>
  <version>7</version>
  <version>8</version>
  <version>9</version>
  <version>10</version>
  <version>11</version>
  <version>12</version>
  <version>13</version>
  <version>14</version>
  <version>15</version>
  <version>16</version>
</version>
</verStmnt>
- <bibliCit>
  Aarebrot, Frank, and Stein Kuhnle. NORWEGIAN ECOLOGICAL DATA, 1949-1961 [Computer file]. Compiled by University of
  Bergen. Inter-university Consortium for Political and Social Research, ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and
  distributor], 1998.
</biblCit>
  <notes>LRECL and Card Image</notes>
</citation>
- <stdyInfo>
```



CDRiP: Production

- Student workers

- “Cheap” hardware

- Free software when possible

 - e.g. daemon tools, Icisocreator

- Some locally developed applications (Perl, XML, XSLT)

- Use existing standards (e.g. iso, mods, xml/xslt)

```
<codeBook>
- <docDscr>
- <citation>
- <titiStmnt>
  <titi>Norwegian Ecological Data, 1949-1961 </titi>
</titiStmnt>
- <verStmnt>
  <version date="2003-12-18"/>
</verStmnt>
  <holdings URI="http://webapp.icpsr.umich.edu/cocoon/ICPSR-STUDY/00040.xml"/>
</citation>
</docDscr>
- <stdyDscr>
- <citation>
- <titiStmnt>
  <titi>Norwegian Ecological Data, 1949-1961 </titi>
  <IDNo>40</IDNo>
</titiStmnt>
- <rspStmnt>
  <AuthEnt100>Aarebrot, Frank </AuthEnt100>
  <AuthEnt700>Kuhle, Stein </AuthEnt700>
</rspStmnt>
<prodStmnt> </prodStmnt>
- <distStmnt>
- <distData date="1998-02-16" />
</distStmnt>
<serStmnt> </serStmnt>
- <verStmnt>
  <version date="1992-02-16">1992-02-16</version>
</verStmnt>
- <bibliCit>
  Aarebrot, Frank, and Stein Kuhle. NORWEGIAN ECOLOGICAL DATA, 1949-1961 [Computer file]. Compiled by University of Bergen, Institute of Sociology. ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor], 1998.
</bibliCit>
  <notes>LRECL and Card Image</notes>
</citation>
- <stdyInfo>
```



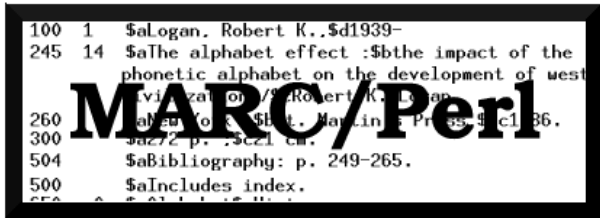
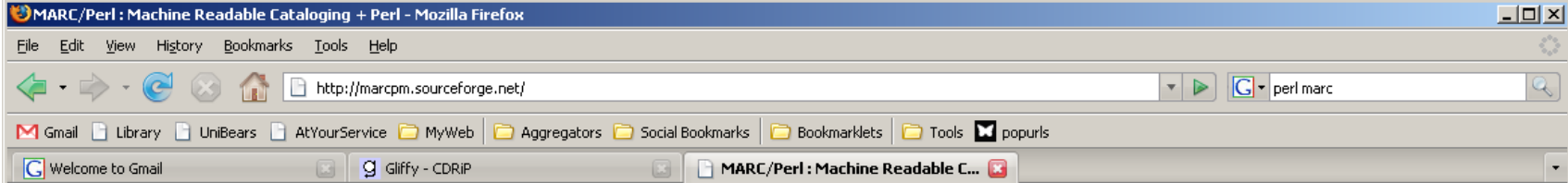
```
dir.xml (D:\My Documents\Scripts\Perl\reptools\trunk) - VIM
<directory name="DOE" depth="1">
  <path>\DOE</path>
  <modify-time epoch="1171492155">Wed Feb 14 14:29:15 2007</modify-time>
  <directory name="100701747" depth="2">
    <path>\DOE\100701747</path>
    <modify-time epoch="1170992573">Thu Feb 8 19:42:53 2007</modify-time>
    <file name="1988 (2).iso">
      <mode code="33206">rw-</mode>
      <size unit="bytes">99999744</size>
      <modify-time epoch="1170977570">Thu Feb 8 15:32:50 2007</modify-time>
    </file>
    <file name="1988 County & city data book.iso">
      <mode code="33206">rw-</mode>
      <size unit="bytes">100306944</size>
      <modify-time epoch="1151946720">Mon Jul 3 10:12:00 2006</modify-time>
    </file>
    <file name="1988.iso">
      <mode code="33206">rw-</mode>
      <size unit="bytes">99999744</size>
      <modify-time epoch="1170977490">Thu Feb 8 15:31:30 2007</modify-time>
    </file>
    <file name="1994 County & city data book.iso">
      <mode code="33206">rw-</mode>
      <size unit="bytes">35612672</size>
      <modify-time epoch="1151946996">Mon Jul 3 10:16:36 2006</modify-time>
    </file>
    <file name="1994.iso">
      <mode code="33206">rw-</mode>
      <size unit="bytes">34998272</size>
      <modify-time epoch="1170977640">Thu Feb 8 15:34:00 2007</modify-time>
    </file>
  </directory>
  <directory name="100938789" depth="2">
    <path>\DOE\100938789</path>
    <modify-time epoch="1171491834">Wed Feb 14 14:23:54 2007</modify-time>
    <file name="2006.iso">
      <mode code="33206">rw-</mode>
      <size unit="bytes">211894272</size>
      <modify-time epoch="1171479478">Wed Feb 14 10:57:58 2007</modify-time>
    </file>
  </directory>
  <directory name="117630412" depth="2">
    <path>\DOE\117630412</path>
    <modify-time epoch="1153181758">Mon Jul 17 17:15:58 2006</modify-time>
    <file name="2004 Yearbook.iso">

```

12,1 0%

XML:

- good for hierarchical data
- existing tools to save time
- easier coding for new tools
- platform independence



Description

MARC/Perl is a Perl 5 library for reading, manipulating, outputting and converting bibliographic records in the [MARC](#) format. The **M**achine **R**eadable **C**ataloging format was designed by the Library of Congress in the late 1960s in order to allow libraries to convert their card catalogs into a digital format. The MARC format was standardized as [ANSI/NISO Z39.2](#) in 1971 and is now being used by libraries around the world to provide computerized access to their collections.

It is common for corporate, public and academic libraries to purchase Integrated Library Systems to organize and manage their collections. However, it is often advantageous for libraries to be able to create or massage pre-existing MARC data on the way into or out-of such systems. This is the niche that MARC/Perl was designed to fill. If you can think of any other niches (ie. Open Source Software for libraries) please let us know!

Development of this software library began with MARC.pm in the summer of 1999. After much experimentation, enhancement and discussion the MARC::Record framework was adopted which provides a more scalable approach to processing MARC data with Perl. While you can still download and get the documentation for MARC.pm, it is recommended that new users download MARC::Record. Future development of the MARC namespace on CPAN will fit into the framework provided by MARC::Record. If you have ideas for new modules please join the [development team](#), or consider posting to the [perl4lib](#) list.

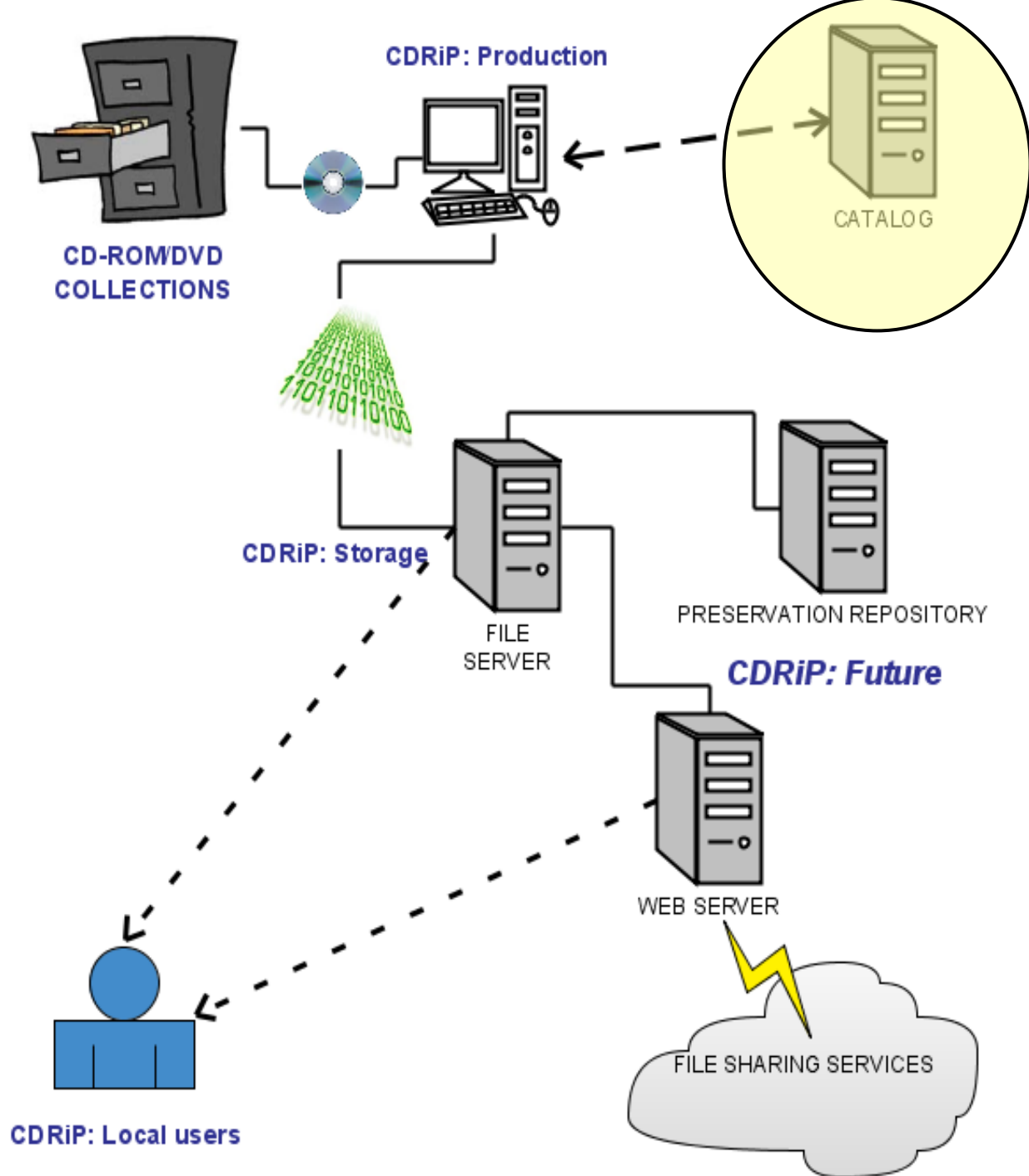
If you have any feature requests, or find any bugs (gasp) please report them via [RT](#), the CPAN request tracking system.

Downloads

The latest stable version of the MARC::Record package is available from [Comprehensive Perl Archive Network \(CPAN\)](#). Experimental versions can often be found in the [download area](#) here at Sourceforge. Although older versions of MARC.pm are available for download, we recommend you use the MARC::Record package since future development will use it as a foundation.

Documentation

- MARC::Record::Tutorial : [pdf](#)
- MARC::Record : [html](#)
- MARC::Lint : [html](#)



REST - “representational state transfer”

- Facilitates “mash-ups” other forms of re-use.

- example:

<http://sunsite2.berkeley.edu:8000/WebZ/Authorize?sessionid=0:next=NEXTCMD%22/WebZ/QUERY:next=html/results.html:entitytoprecno=1:entitydbgroup=Glad:indexA=g1%3A:next=html/Cannedresultsframe.html:termA=84333509>

vs

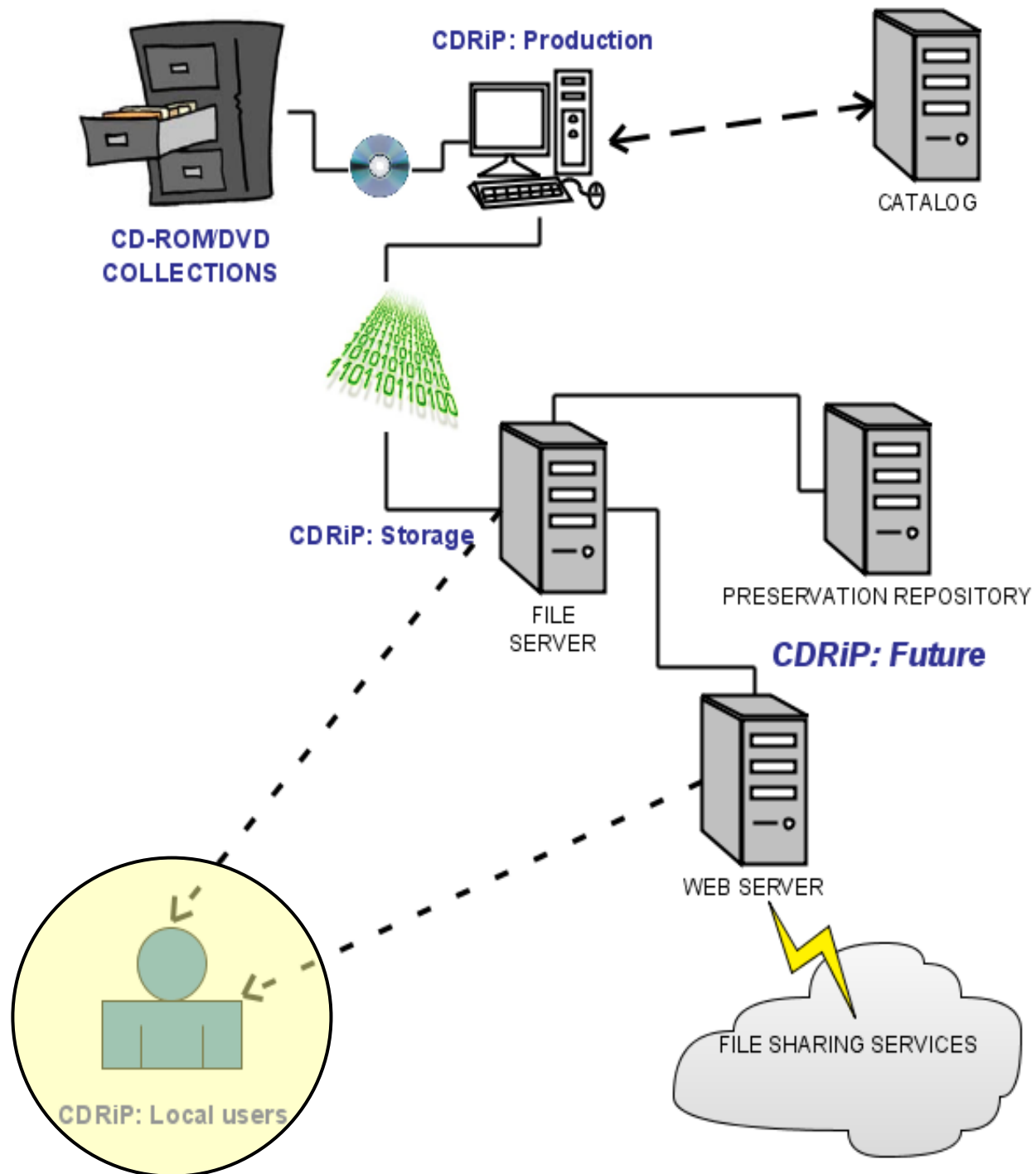
<http://librrc-rh.berkeley.edu/crax/find/g1/84333509>

```
<codeBook>
- <docDscr>
- <citation>
- <titlStmnt>
  <titl>Norwegian Ecological Data, 1949-1961 </titl>
</titlStmnt>
- <verStmnt>
  <version date="2003-10-10" />
</verStmnt>
  <holdings URI="http://webapp.icpsr.umich.edu/cocoon/ICPSR-STUDY/00040.xml"/>
</citation>
</docDscr>
- <stdyDscr>
- <cita>
- <titlStmnt>
  <titl>Norwegian Ecological Data, 1949-1961 </titl>
  <IDN>19-1961
</titlStmnt>
- <rspStmnt>
  <AuthEntA>Aarebrot, Frank and Stein Kuhnle
  <AuthEntC>Aarebrot, Frank
  <AuthEnt70>
</rspStmnt>
  <prodStmnt>
  <distStmnt>
- <distrbtr ab>
  Inter-university Consortium for Political and Social Research
  </distrbtr>
  <distDate date="1984-06-20">1984-06-20</distDate>
</distStmnt>
</serStmnt>
</verStmnt>
  <version date="1998-06-15" />
</verStmnt>
- <bibliCit>
  Aarebrot, Frank, and Stein Kuhnle. NORWEGIAN ECOLOGICAL DATA, 1949-1961 [Computer file]. Compiled by University of Bergen, Institute of Sociology. ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor], 1998.
</bibliCit>
  <notes>LRECL and Card Image</notes>
</citation>
- <stdyinfo>
```

```

- <modsCollection xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-1.xsd">
- <mods version="3.1">
- <titleInfo>
  <title>Governor's budget summary</title>
</titleInfo>
- <titleInfo type="alternative">
  <title>State of California, Governor's budget summary</title>
</titleInfo>
- <name type="corporate">
  <namePart>California</namePart>
  <namePart>Governor.</namePart>
</name>
<typeOfResource/>
- <originInfo>
- <place>
  <placeTerm type="text">Sacramento, Calif</placeTerm>
</place>
<publisher>California Office of State Printing</publisher>
<issuance/>
<frequency>Annual</frequency>
</originInfo>
- <physicalDescription>
  <extent>v. ill. 28 cm.</extent>
</physicalDescription>
- <note type="statement of responsibility">
  submitted by ... Governor, to the California Legislature
</note>
<note>Some issues accompanied by CD-ROM.</note>
<note>Report year ends June 30.</note>
<note>Also available via the World Wide Web.</note>
- <subject>
  <geographicCode authority="marcgcac">n-us-ca</geographicCode>
</subject>
- <subject authority="lcsch">
  <topic>Budget</topic>
<geographic>California</geographic>
<topic>Statistics</topic>
<topic>Periodicals</topic>
</subject>

```



CDRiP: User tools

- Wiki pages

- Virtual machine technology

 - allows users to safely install applications from otherwise “locked-down” workstations

- Search (to be determined)

 - how to integrate with other databases

 - how to leverage additional metadata

```
<codeBook>
- <docDscr>
- <citation>
- <titiStmnt>
  <titi>Norwegian Ecological Data, 1949-1961 </titi>
</titiStmnt>
- <verStmnt>
  <version date="2003-12-18"/>
</verStmnt>
  <holdings URI="http://webapp.icpsr.umich.edu/cocoon/ICPSR-STUDY/00040.xml"/>
</citation>
</docDscr>
- <stdyDscr>
- <citation>
- <titiStmnt>
  <titi>Norwegian Ecological Data, 1949-1961 </titi>
  <IDNo>40</IDNo>
</titiStmnt>
- <rspStmnt>
  <AuthEnty>Aarebrot, Frank, and Stein Kuhnle </AuthEnty>
  <AuthEnt100>Aarebrot, Frank, and Stein Kuhnle </AuthEnt100>
  <AuthEnt700>Aarebrot, Frank, and Stein Kuhnle </AuthEnt700>
</rspStmnt>
- <prodStmnt> </prodStmnt>
- <distStmnt>
  <distbtr>Inter-university Consortium for Political and Social Research, University of Michigan" URI="http://www.icpsr.umich.edu">
  Inter-university Consortium for Political and Social Research
</distbtr>
  <distDate>1998-04-01</distDate>
</distStmnt>
- <serStmnt> </serStmnt>
- <verStmnt>
  <version date="1992-02-16">1992-02-16</version>
</verStmnt>
- <bibliCit>
  Aarebrot, Frank, and Stein Kuhnle. NORWEGIAN ECOLOGICAL DATA, 1949-1961 [Computer file]. Compiled by University of Bergen, Institute of Sociology. ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor], 1998.
</bibliCit>
  <notes>LRECL and Card Image</notes>
</citation>
- <stdyinfo>
```



Library Data Lab
 215 Doe Library (map)
 Berkeley CA 94720-6000
 510-642-8095

About

- Home
- Services
- Staff
- Hours

Computing

- Lab Software
- Campus Labs
- Buying Software
- Online Help

Finding Data

- Our Collection
- Archives
 - ICPSR
 - Roper Center
 - UC Berkeley
 - Harvard MIT Data Center
- Google data search

Data Analysis

- Learning Materials
- Workshops

UC Berkeley
 UC Berkeley Library
 Doe/Moffitt Libraries

staff only [edit SideBar](#)

Search:

View Edit History Print

DREF

The following is a partial inventory of the statistical titles in the Doe Reference Collection. Please consult the [Data Lab staff](#) or a [reference librarian](#) if you wish to use one of the following or are having difficulty locating the data you need.

Statisticka rocenka Slovenskej republiky = : Statistical yearbook of the Slovak Republic

Publisher Statisticky urad Slovenskej republiky : Bratislava ()
Access [see PathFinder](#)

100-percent summary file : Census 2000 dress rehearsal

Publisher U.S. Dept. of Commerce, Bureau of the Census : Washington, D.C.] ([1999])
Access [see PathFinder](#)

108th Congressional district summary files : census of population and housing, 2000

Publisher U.S. Dept. of Commerce, Economics and Statistics Administration, U.S. Census Bureau : Washington, DC ([2003])
Access [see PathFinder](#)

1987 census of agriculture

Publisher Bureau of the Census : Washington, D.C (1990)
Access [see PathFinder](#)

1987 economic censuses

Publisher U.S. Dept. of Commerce, Bureau of the Census, Data User Services Division : Washington, D.C (1989-)
Access [see PathFinder](#)

1989 USSR population census

Publisher East View Publications : Minneapolis, MN (c1996)
Access [see PathFinder](#)

1990 Census of population and housing : P.L. 94-171 redistricting data

Publisher U.S. Dept. of Commerce, Bureau of the Census, Data User Services Division : Washington, D.C (1991)
Access [see PathFinder](#)


1990 census of population and housing

Publisher U.S. Dept. of Commerce, Bureau of the Census, Data User Services Div. : Washington D.C (1991-)

Windows XP Professional - VMware Workstation

File Edit View VM Team Windows Help

Home Windows XP Professional



Microsoft®
Windows[™] xp

Copyright © Microsoft Corporation

Microsoft

bookmarks... 2wiki.txt dekker_insti...

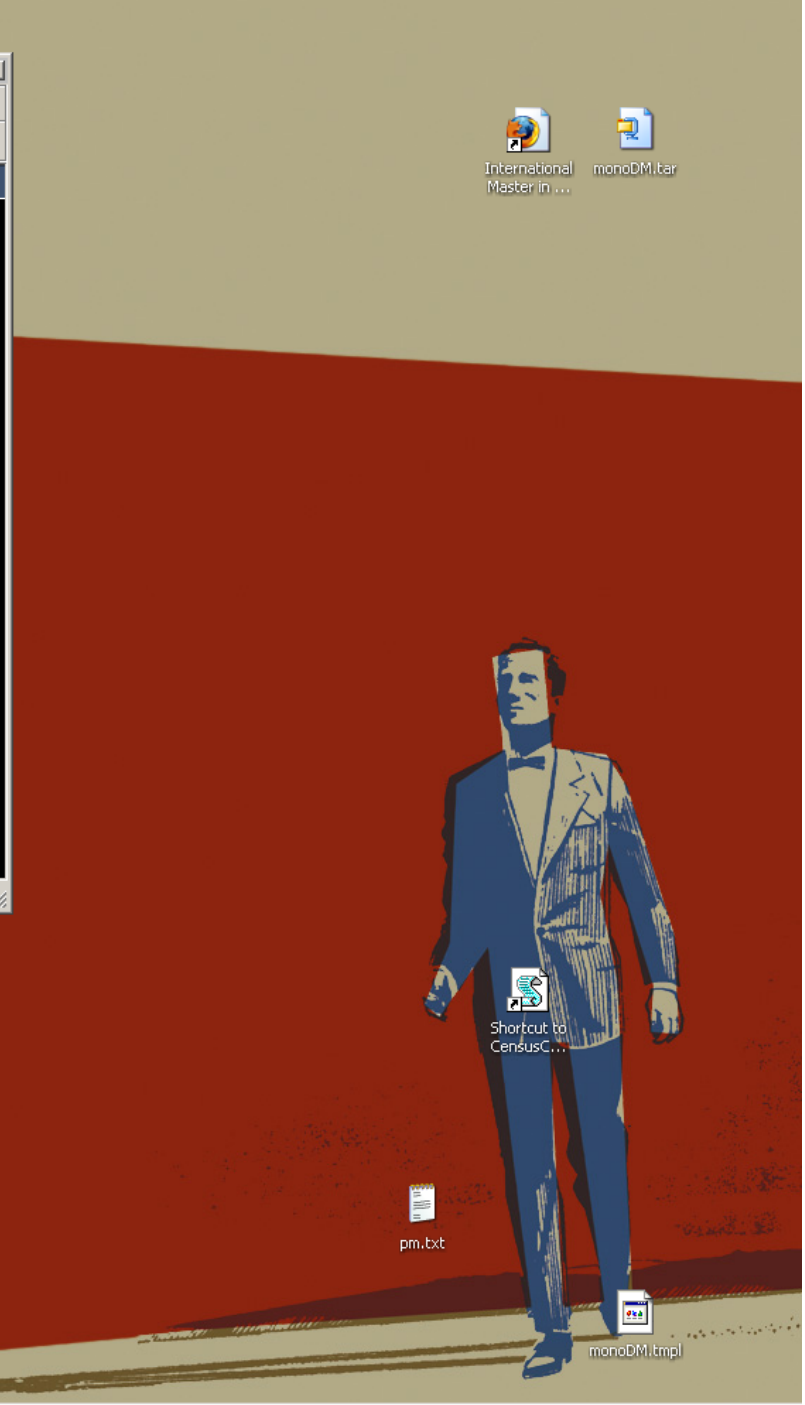
schizo Midterm II B.doc

DataFerrett README.txt 381214611...

leadoutput... Harrison Dekker CV.rtf 7046404.zip

mementest66... VMware-wo... CDRIP.png

Start VMware... librc-rh... ERL_CD... CDRIP P... Window... 51% of ... IrfanVi... EN



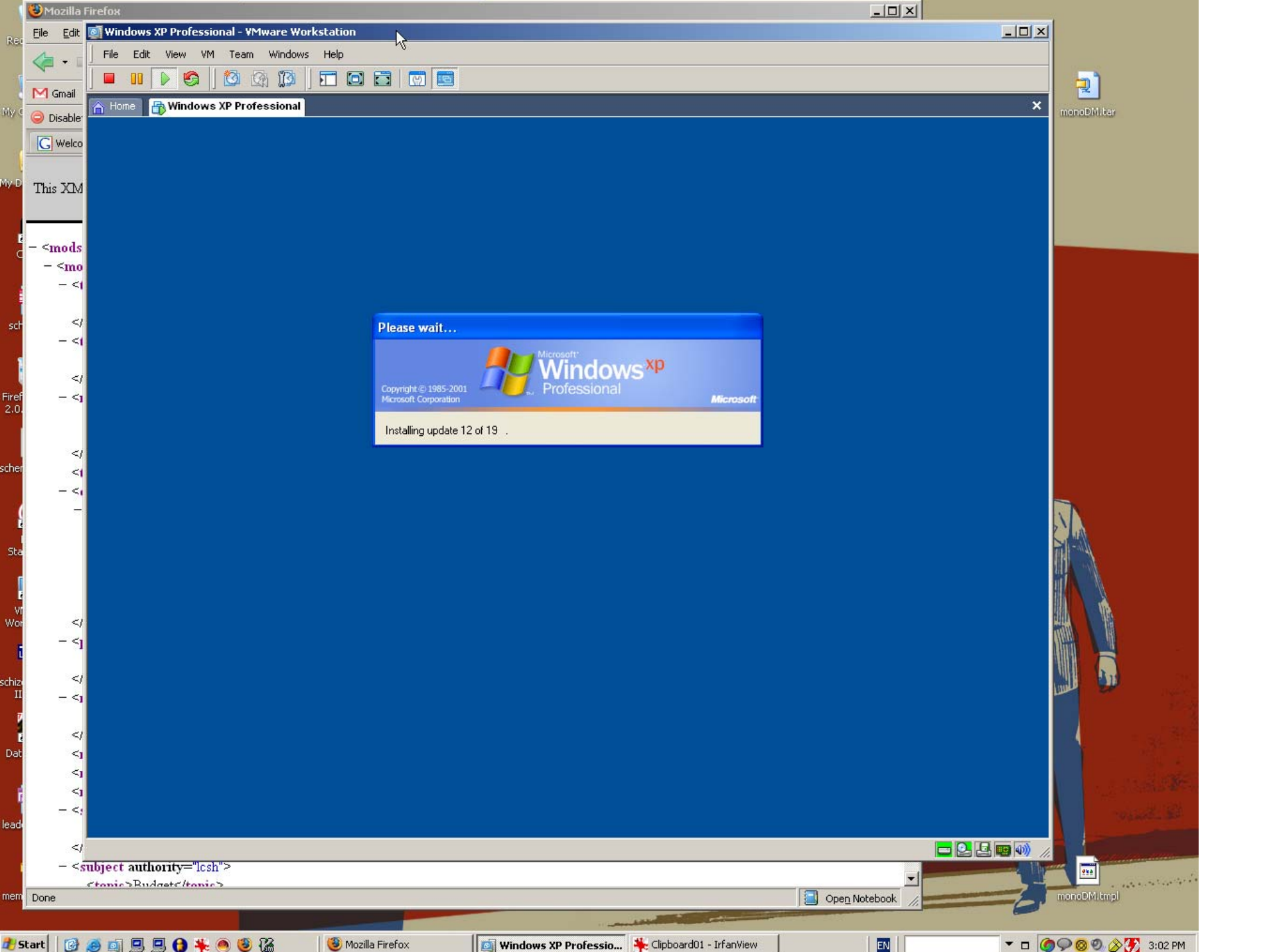
International Master in ... monoDM.tar

Shortcut to CensusC...

pm.txt

monoDM.tmpl

2:38 PM



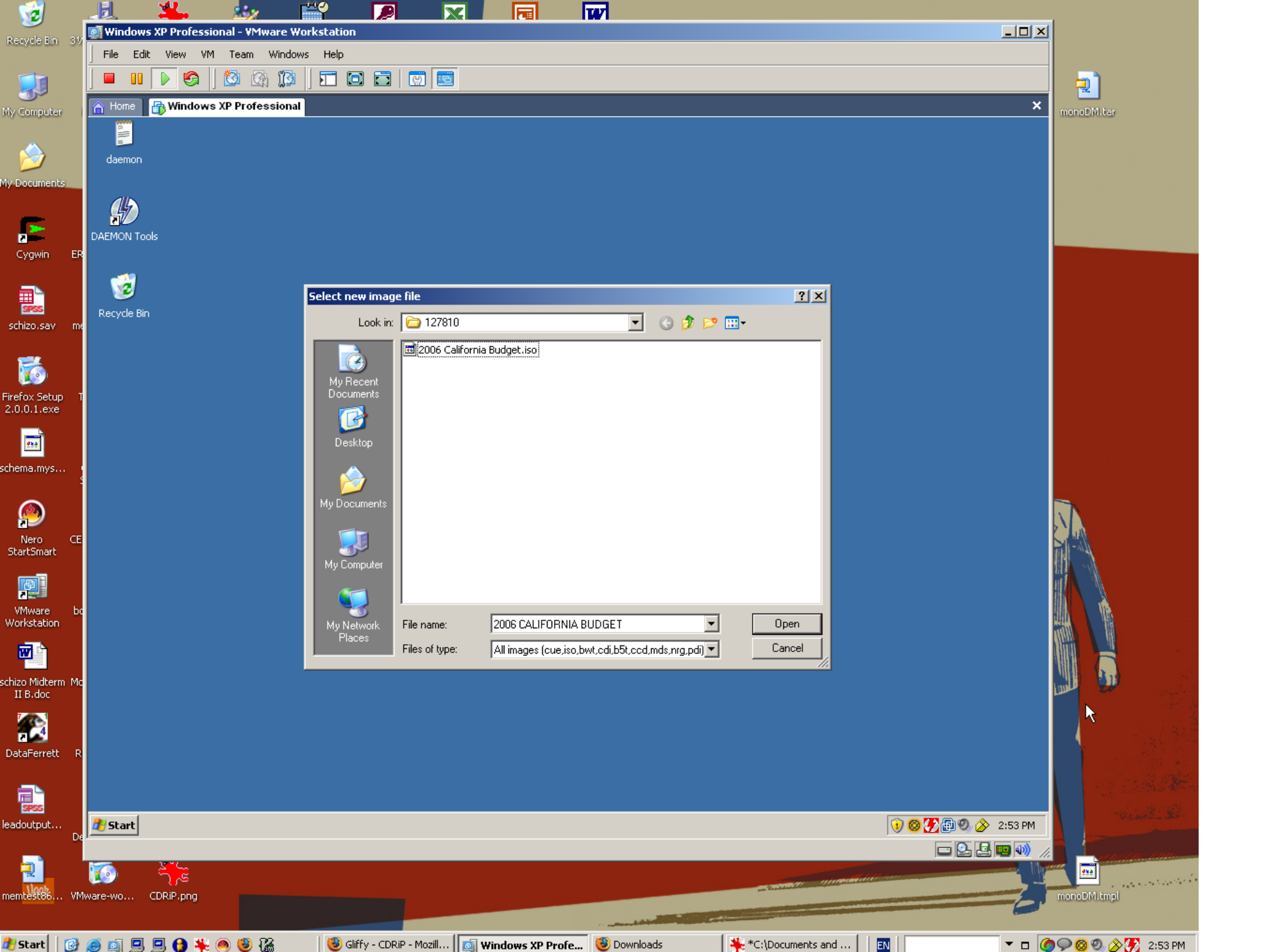
Please wait...



Copyright © 1985-2001
Microsoft Corporation

Microsoft

Installing update 12 of 19 .



Windows XP Professional - VMware Workstation

File Edit View VM Team Windows Help

Home Windows XP Professional

daemon

DAEMON Tools

Recycle Bin

Select new image file

Look in: 127810

2006 California Budget.iso

My Recent Documents

Desktop

My Documents

My Computer

My Network Places

File name: 2006 CALIFORNIA BUDGET

Open

Files of type: All images (cue.iso,bwt,cdi,b5t,ccd,mds,nrg,pdi)

Cancel

Start

2:53 PM

Start

Gliffy - CDRIP - Mozill...

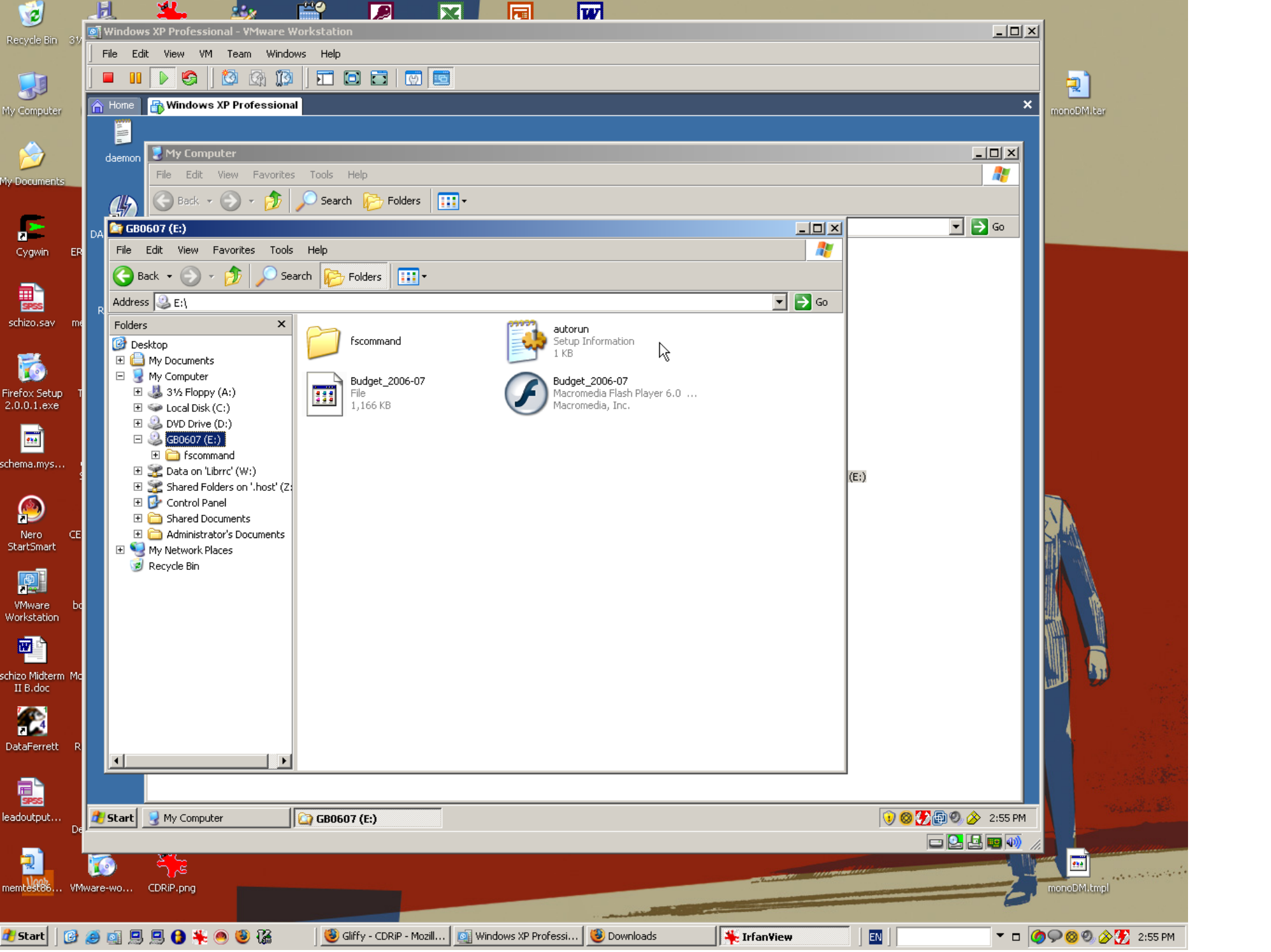
Windows XP Profe...

Downloads

*C:\Documents and ...

EN

2:53 PM



Windows XP Professional - VMware Workstation

File Edit View VM Team Windows Help



Home Windows XP Professional

monoDM.tar

My Computer

File Edit View Favorites Tools Help



GB0607 (E:)

File Edit View Favorites Tools Help



Address E:\

- Desktop
- My Documents
- My Computer
- 3 1/2 Floppy (A:)
- Local Disk (C:)
- DVD Drive (D:)
- GB0607 (E:)
 - fscommand
- Data on 'Librrc' (W:)
- Shared Folders on '.host' (Z:)
- Control Panel
- Shared Documents
- Administrator's Documents
- My Network Places
- Recycle Bin

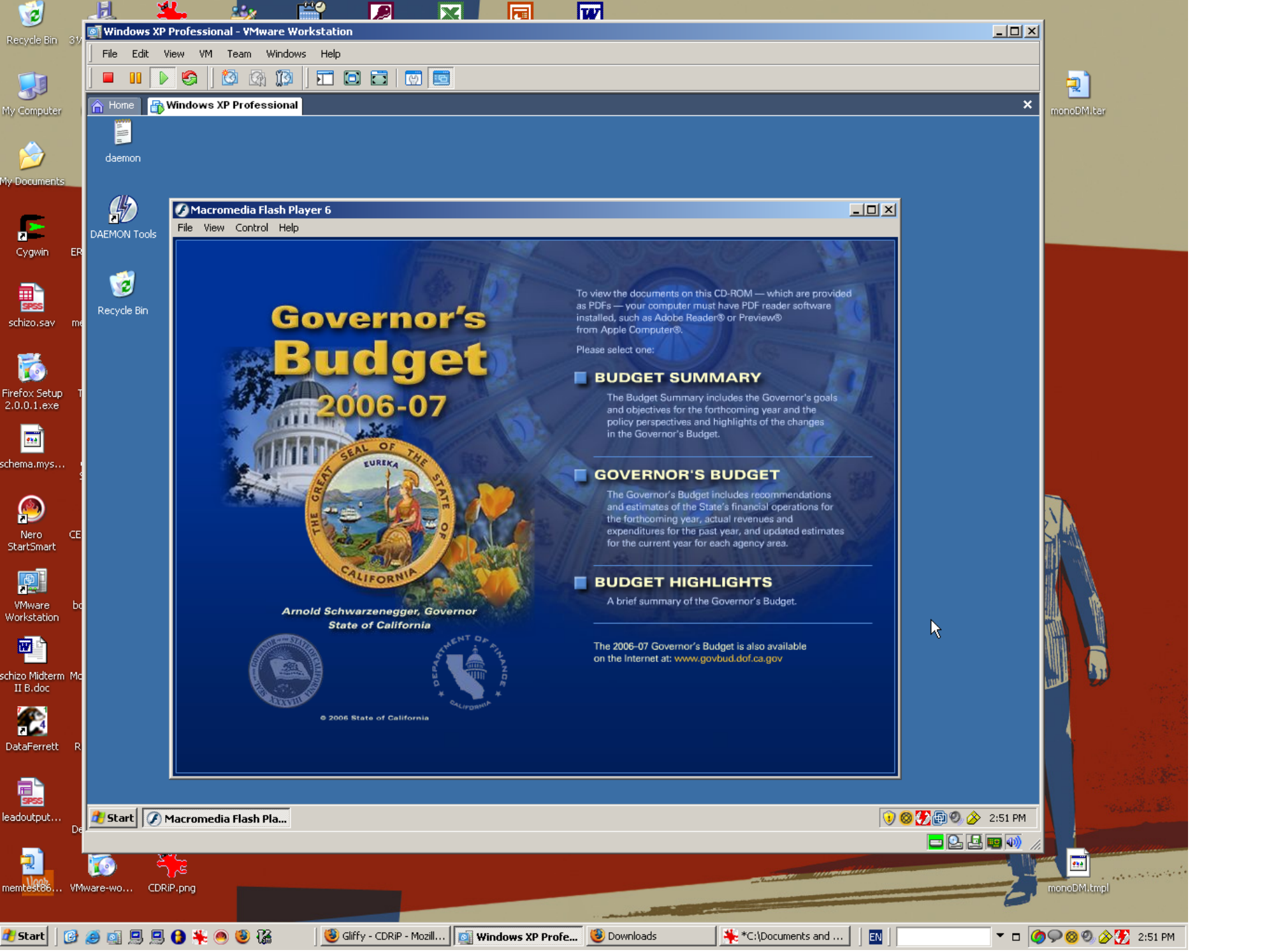
- fscommand
- Budget_2006-07 File 1,166 KB
- autorun Setup Information 1 KB
- Budget_2006-07 Macromedia Flash Player 6.0 ... Macromedia, Inc.

Start My Computer GB0607 (E:)

2:55 PM



Start Gliffy - CDRIP - Mozill... Windows XP Professi... Downloads IrfanView EN 2:55 PM



Governor's Budget 2006-07



Arnold Schwarzenegger, Governor
State of California



© 2006 State of California

To view the documents on this CD-ROM — which are provided as PDFs — your computer must have PDF reader software installed, such as Adobe Reader® or Preview® from Apple Computer®.

Please select one:

BUDGET SUMMARY

The Budget Summary includes the Governor's goals and objectives for the forthcoming year and the policy perspectives and highlights of the changes in the Governor's Budget.

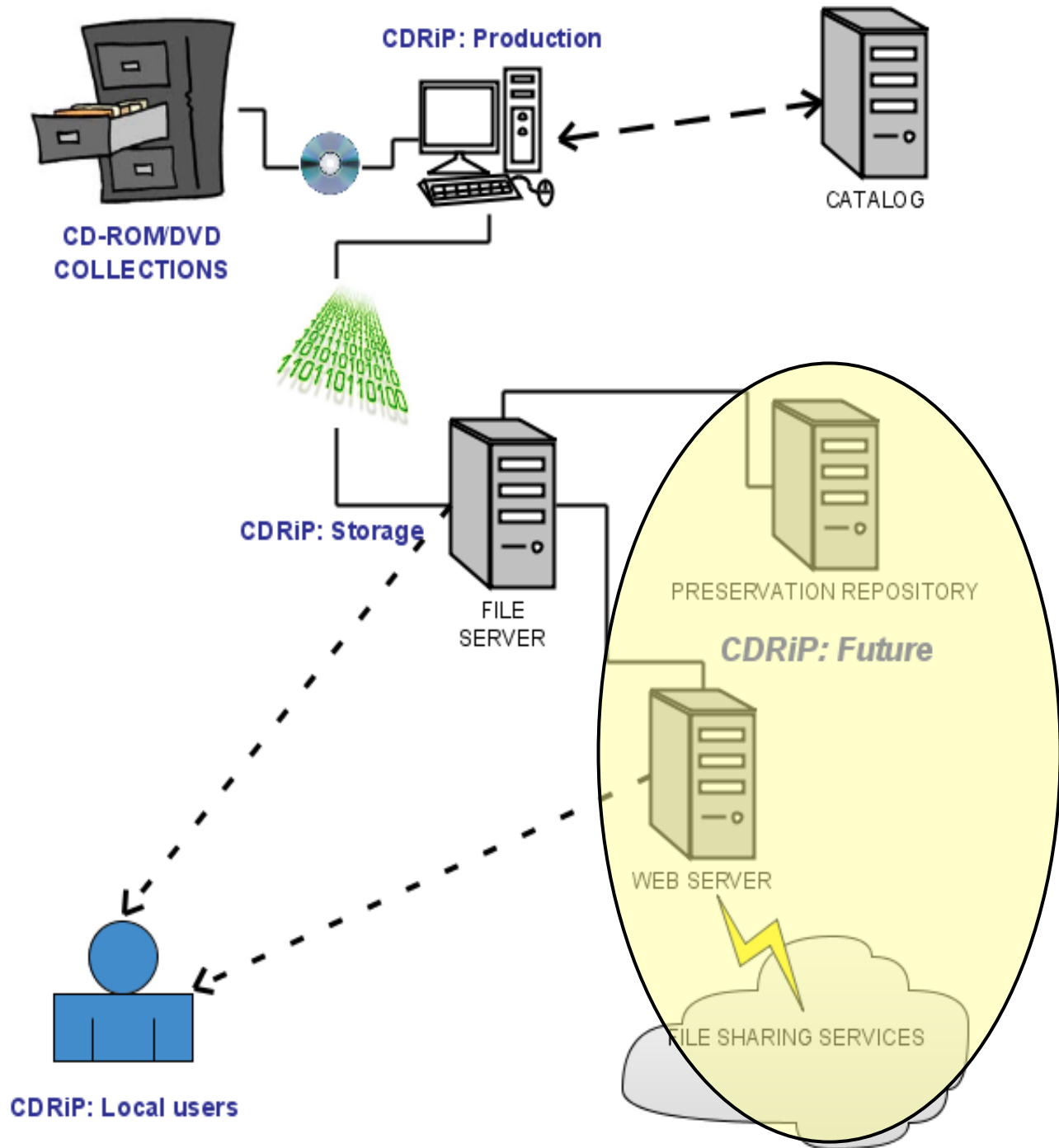
GOVERNOR'S BUDGET

The Governor's Budget includes recommendations and estimates of the State's financial operations for the forthcoming year, actual revenues and expenditures for the past year, and updated estimates for the current year for each agency area.

BUDGET HIGHLIGHTS

A brief summary of the Governor's Budget.

The 2006-07 Governor's Budget is also available on the Internet at: www.govbud.dof.ca.gov



CDRiP Future

```
<codeBook>
- <docDscr>
  - <citation>
    - <titiStmnt>
      <titi>Norwegian Ecological Data, 1949-1961 </titi>
    </titiStmnt>
    - <verStmnt>
      <version date="2003-12-18"/>
    </verStmnt>
    <holdings URI="http://webapp.icpsr.umich.edu/cocoon/ICPSR-STUDY/00040.xml"/>
  </citation>
</docDscr>
- <stdyDscr>
  - <citation>
    • Preservation pipeline
    • Publish to web
    • Data mining
    • File sharing
    • Virtualization projects
    - <titiStmnt>
      <titi>Norwegian Ecological Data, 1949-1961 </titi>
    </titiStmnt>
    - <rspStmnt>
      <AuthEnty>Aarebrot, Frank, and Stein Kuhnle </AuthEnty>
      <AuthEnt100>Frank </AuthEnt100>
      <AuthEnt700> </AuthEnt700>
    </rspStmnt>
    <prodStmnt> </prodStmnt>
  </citation>
  - <distStmnt>
    - <distr abbr="ICPSR" affiliation="Institute for Social Research, University of Michigan" URI="http://www.icpsr.umich.edu">
      Inter-university Consortium for Political and Social Research
    </distr>
    <distDate> </distDate>
  </distStmnt>
  <serStmnt> </serStmnt>
  - <verStmnt>
    <version date="1992-02-16">1992-02-16</version>
  </verStmnt>
  - <bibliCit>
    Aarebrot, Frank, and Stein Kuhnle. NORWEGIAN ECOLOGICAL DATA, 1949-1961 [Computer file]. Compiled by University of Bergen, Institute of Sociology. ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor], 1998.
  </bibliCit>
  <notes>LRECL and Card Image</notes>
</citation>
- <stdyinfo>
```

<http://cgi.cs.indiana.edu/~geobrown/svp>
see also: <http://freegovinfo.info/blog/124>

[Home Page](#)

[Contacts](#)

[Database](#)

[Search](#)

[Help](#)

[RSS](#)

SUDOC Virtualization Project

This site provides access to an online database containing government documents originally published on CDROM. There are two interfaces, a [Database](#) interface that enables browsing the entire collection and a [Search](#) interface that allows searching for specific documents using [CQL](#) queries.

Through either interface it is possible to access the original CDROM contents exported as [WebDAV](#). These WebDAV directories can be directly "mounted" on some operating systems (OS X, Linux). In Windows they can be accessed through the network neighborhood or mounted using Novell Netdrive. WebDAV access does require a password. Please contact Geoffrey Brown for access.

On the Indiana University Bloomington campus, the CDROMs can be imported using windows file sharing. Interested users may contact Geoffrey Brown for details.

Summary

CDRiP rationale:

- Better long-term access

CDRiP technologies:

- ISO files and tools
- Virtual machine software

REST

- Perl, XML, XSLT

Questions/comments: hdekker@library.berkeley.edu