Secondary Data Sharing System

Eiji Ikoma
The Univ. of Tokyo
IMPAC-T Information Server Team
IMPAC-T Server Services

• Telemetry Data Crawling and Archive
  – Impact-t1,impact-t2@KU
  – Impact-r-t1,t2,t3,t4@RID
  – Impact-t-t1,t2@TMD

• Data Integration System (DIS)
  – Impact-di@KU
  – Impact-r-di@RID
  – Impact-t-di@TMD

• Web Server
  – Impact-www@KU

• Application Server (for models)
  – Impact-a1,ap,mr@KU
  – Impact-v1,v2,v3,v4@KU (virtual)

• Secondary Data Sharing System
What is “Secondary Data”?

↔ Primary Data
≠ Telemetry Data
≠ Observation Data
→ Secondary Data

= Data for Secondary Use?
Secondary data is collecting and possibly processing data by people other than the researcher.
Secondary data

Secondary data, is data collected by someone other than the user. Common sources of secondary data for social science include censuses, organisational records and data collected through qualitative

Data collected by someone other than the user

Secondary data analysis saves time that would otherwise be spent collecting data and, particularly in the case of quantitative data, provides larger and higher-quality databases that would be unfeasible for any individual researcher to collect on their own. In addition, analysts of social and economic change consider secondary data essential, since it is impossible to conduct a new survey that can adequately capture past change and/or developments.

Contents

1. Sources of secondary data
2. Secondary analysis or re-use of qualitative data
3. References
4. Further reading
5. External links
What is “Secondary Data”?

Anyway, Secondary data is important data which can be used for research.

→ necessary to develop easy-to-use environment
Secondary Data Sharing System

- System for **Sharing “Data”**
- Somebody upload to Server
  - → Store and Manage on Server
  - → Users can use/download data
Upload

• By http
  – Using Web-based Interface
• By ftp
  – Using FTP client software
• By direct-connection
  – Using USB HDD for connecting to Server
Management

• All Secondary data will be stored at IMPAC-T Data Integrated Server
• 1 Dataset = 1 Directory
• Dataset should include meta-information
  – Free style document file
  – Well-formatted text (XML etc..)
• Security/Permission Information
• Is Pre-processing necessary?
For Use

• Download
  – Web-based download Interface
  – Data Integrated Servers prepare “download data set” for each user

• Direct-use on server
  – Each user has their own environment/disk space on application server
  – Data Integrated Servers prepare data and copy to user’s home directory

• Direct-connected HDD and “carry”
Case 1

- ISI-MIP data (by Dr. Kiguchi)
  - Upload: Direct-connection (USB HDD)
  - Manage: original data and Thai-area data (pre-processing)
  - Use/Download: Prepare download interface

- All data are stored at impact-di server
  = 22TB (compressed, including pre-processed data)
Interface for Data Selection

Choose
- Dataset
- Area (global/thai)
- Model
- Element

Select Scenario, and Term
STEP 1 (Area, Model, Element) --> STEP 2 (Scenario, Term)
- Dataset: isimip
- Area: global/thai
- Model: cmip5, esm2, esm2a, esm2g
- Scenario: historical, rcp2p6, rcp4p5, rcp8p5

Choose
- Scenario
- Term

Download Copy to Your Home Directory @ AppServer
Confirmation
For Data Download

20496 MB (Total), Download from IMPAC-T Server

Is this correct? Yes
Case 2

- KAKUSHIN-MRI data (by Dr. Kiguchi)
  - Upload: Direct-connection (USB HDD)
  - Manage: original data
  - Use/Download: (under consideration)

- 70% of data are stored at impact-di server
  Finally, 13TB (original data) will be stored.

→ depend on the discussion after my talk
Case 3

- Any data
  - Upload: Direct-connection (USB HDD)
  - Manage: original data
  - Use/Download: Web(Directory View only)

→ Simplest Interface for download
## Index of /Secondary/isi-mip/data/global/gfdl-esm2m/rcp8p5

<table>
<thead>
<tr>
<th>Name</th>
<th>Last_modified</th>
<th>Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent Directory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LWdown/</td>
<td>28-Dec-2012 20:51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSurf/</td>
<td>29-Dec-2012 00:52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qtair/</td>
<td>29-Dec-2012 04:20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rainf/</td>
<td>29-Dec-2012 07:44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWdown/</td>
<td>29-Dec-2012 11:26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Snowf/</td>
<td>29-Dec-2012 14:53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tair/</td>
<td>16-Jan-2013 10:48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wind/</td>
<td>16-Jan-2013 14:08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Apache/2.2.3 (CentOS) Server at impact-diengku.ac.th Port 80

## Index of /Secondary/isi-mip/data/thai/gfdl-esm2m/historical/LWdown/1966

<table>
<thead>
<tr>
<th>Name</th>
<th>Last_modified</th>
<th>Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent Directory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LWdown_1966.0101.00.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
<tr>
<td>LWdown_1986.0101.03.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
<tr>
<td>LWdown_1986.0101.05.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
<tr>
<td>LWdown_1986.0101.09.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
<tr>
<td>LWdown_1986.0101.15.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
<tr>
<td>LWdown_1986.0101.18.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
<tr>
<td>LWdown_1986.0101.21.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
<tr>
<td>LWdown_1986.0102.00.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
<tr>
<td>LWdown_1986.0102.03.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
<tr>
<td>LWdown_1986.0102.06.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
<tr>
<td>LWdown_1986.0102.09.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
<tr>
<td>LWdown_1986.0102.12.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
<tr>
<td>LWdown_1986.0102.15.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
<tr>
<td>LWdown_1986.0102.18.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
<tr>
<td>LWdown_1986.0102.21.bin</td>
<td>28-Dec-2012 12:50</td>
<td>640</td>
<td></td>
</tr>
</tbody>
</table>
Discussion

• What kind of data can you prepare?

Depend on the data format, size, structure, permission…

I’ll prepare appropriate system for upload, direct-use/download Interface -software, hardware-

Please give rough information about your data!
Thank you for your attention