DataSpace@HKUST
Research data repository and management service

K.T. Lam, HKUST Library (lblkt@ust.hk)
• Responses on **data preservation and management**
  • 65% of the respondents build up collections of research data in the course of their research.
  • 56% of the respondents build up collections of image or media
  • Most important data sources are those collected by the researchers themselves and those that are freely available online
  • Most important data management features are
    • Ability to track academic research citing my data
    • Ability to store multiple versions of my data in the same place
    • Ability to store my data with my corresponding published outputs

• Most of the researchers manage their data on their own computers
• More than one half of researchers manage their data on the cloud storage service (such as Google Drive, Dropbox, Flickr, etc.)
• Majority perceived data support provided by
  • Freely available software
  • University Library
• Majority preserve data by themselves, using commercially or freely available software or service
Library Research Support Faculty Survey 2015

DATA6 – Perceived Values of Data Support

The question lists 8 sources of data support. *Free software* was ranked the highest on average, followed by *library*.

### Average Scores of Data Support

<table>
<thead>
<tr>
<th>Source</th>
<th>Average Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freely available software</td>
<td>7.11</td>
</tr>
<tr>
<td>My university library</td>
<td>6.36</td>
</tr>
<tr>
<td>My university IT department</td>
<td>5.64</td>
</tr>
<tr>
<td>A scholarly society</td>
<td>4.87</td>
</tr>
<tr>
<td>A disciplinary or departmental repository at my institution</td>
<td>4.77</td>
</tr>
<tr>
<td>An AV or media support at my institution</td>
<td>4.49</td>
</tr>
<tr>
<td>A disciplinary repository at another institution</td>
<td>4.19</td>
</tr>
<tr>
<td>A publisher or a university press</td>
<td>3.94</td>
</tr>
</tbody>
</table>

[http://library.ust.hk/ithaka/results-data.html](http://library.ust.hk/ithaka/results-data.html)
DATA7 – Data Preservation Practices

There are 5 methods listed here. The majority (70%) of the respondents preserve data themselves; all other options have relatively few votes.

- 70%: I preserve these materials myself, using commercially or freely available software or services.
- 12%: I preserve these materials myself in a repository made available by my institution or another type of online repository.
- 10%: My campus or university library preserves these materials on my behalf.
- 1%: A publisher preserves these materials on my behalf alongside the final research output.
- 6%: These materials are generally not preserved following the conclusion of a project.

http://library.ust.hk/ithaka/results-data.html
Requirement on research data by publishers and funding agents

• Others can replicate your research findings
• Data broadly available for sharing
• Long term access is sustainable

Funding agents

**NSF Data Sharing Policy**

“Investigators are expected to **share** with other researchers, at no more than incremental cost and within a reasonable time, **the primary data**, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing...”

**NSF Data Management Plan Requirements**

“...Proposals must include a **supplementary document** of no more than two pages labeled "Data Management Plan”... and may include ... 5. **plans for archiving data**, samples, and other research products, and for **preservation of access** to them...”

http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#dmp

Publishers

Nature’s Authors Policy - Availability of data, material and methods

An inherent principle of publication is that others should be able to **replicate** and build upon the authors' published claims. A condition of publication in a Nature journal is that **authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications**....”

http://www.nature.com/authors/policies/availability.html
Data Repositories

• Specialist repositories
  • GenBank, NeuroMorpho.org, HepData, EarthChem, PubChem, ICPSR (openICPSR), Harvard Dataverse, etc.

• Generalist repositories
  • Dryad, figshare

• Institutional (project-based) repositories
  • DataSpace@HKUST

More examples: http://www.nature.com/sdata/data-policies/repositories
DataSpace – Research data portal

Dual Functions

Repository of published datasets

Workspace for data management
Launched in February 2016 with 4 datasets

Based on open source software from Harvard University. Over 14 installations worldwide.

Harvard Dataverse: https://dataverse.harvard.edu/

<table>
<thead>
<tr>
<th>Dataset Title</th>
<th>Authors</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HKUST Library Research Support Faculty Survey (2015)</td>
<td>Hong Kong University of Science and Technology Library</td>
<td>In March and April 2015, the Library conducted a survey with HKUST faculty on library research support services. This was a joint survey involving librarians at 4 UGC-funded institutions, using the Ithaka Faculty Survey Series.</td>
</tr>
<tr>
<td>Longitudinal Links to Construct the Korean Multi-Generational Panel Dataset</td>
<td>Campbell, Cameron Dougall, Loe, James; Dong, Hao</td>
<td>The longitudinal and interpersonal links provided here transform the cross-sectional Tansung Household Registers (THR) data to a multi-generational panel dataset which we refer to as Korean Multi-Generational Panel Dataset-Tansung (KMGPD-TS). This linkage work was carried out...</td>
</tr>
</tbody>
</table>
Dataset deposited (published) by the group

Research group

Mobile Real-time Air Monitoring Platform Dataverse

Street Level Air Quality Data Collected by the Mobile Real-time Air Monitoring Platform (MAP)

Dec 7, 2015

Fang, Ming; Yao, Xiaohong; Chan, Chak Keung; Lau, Ngai Ting; Lau, Arthur P. S., 2015, "Street Level Air Quality Data Collected by the Mobile Real-time Air Monitoring Platform (MAP)", http://dx.doi.org/10.14711/dataset/l5VZXZ, DataSpace@HKUST, V1 [UNF:6:DLd0PVcvPZR9n88lQpJc1g=-]

Data collected by the Mobile Real-time Air Monitoring platform (MAP) from 2003-2006. This project was done by the research team of the Institute for Environment and Sustainable Development and the Department of Chemical Engineering of the Hong Kong University of Science and Techn...

https://dataspac...
Service for all HKUST researchers
- Openly accessible perpetually
- DOI assigned – persistent identifier facilitating discovery
- Standardized citation and metadata – created by the Library
- Trusted repository - with HKUST branding

Repository - for you to publish datasets

Citation

http://dx.doi.org/10.14711/dataset/I5VZXZ

Usage

Tabs – Files, Metadata, Terms, Versions

Data files

Open access

http://dx.doi.org/10.14711/dataset/I5VZXZ
<table>
<thead>
<tr>
<th>Dataset Persistent ID</th>
<th>doi:10.14711/dataset/5VZXZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication Date</td>
<td>2015-12-07</td>
</tr>
<tr>
<td>Title</td>
<td>Street Level Air Quality Data Collected by the Mobile Real-time Air Monitoring Platform (MAP)</td>
</tr>
</tbody>
</table>
| Author               | Fang, Ming (HKUST)          
|                      | Yao, Xiaohong (HKUST)       
|                      | Chan, Chak Keung (HKUST) - ORCID: 0000-0001-9687-8771 
|                      | Lau, Ngai Ting (HKUST)      
|                      | Lau, Arthur P. S. (HKUST)   |
| Contact              | Use email button above to contact. 
|                      | Chan, Chak-Keung (HKUST)    |
| Description          | Data collected by the Mobile Real-time Air Monitoring platform (MAP) from 2003-2006. This project was done by the research team of the Institute for Environment and Sustainable Development and the Department of Chemical Engineering of the Hong Kong University of Science and Technology. See the readme file for information about the data. The data can be visualized by the GIS software developed by the data contributors. Their contacts are available in the readme file. |
| Subject              | Earth and Environmental Sciences; Medicine, Health and Life Sciences |
| Keyword              | Air quality 
|                      | Hong Kong 
|                      | Street level air quality 
|                      | Pollution 
|                      | Mobile Real-time Air Monitoring Platform |
| Language             | English |
| Depositor            | DataSpace Team |
| Deposit Date         | 2009-05-06 |
| Kind of Data         | Geographic information data; Meteorological data; Air quality data |
Workspace for data management

• For faculty members and their research project groups (including their research postgraduate students)
• Store, manage and share data within research group in a restricted environment

• Manage and share your data internally
  • Data storage throughout the research life cycle (50GB per dataset)
  • Share findings with collaborators
  • Student assignments
  • Fulfill data management plan specified in grant proposals
  • Replication of data readily available to go with the manuscript submission
Creating a dataverse for your research project data

Get Started by creating your workspace (dataverse)

Provide information about your research project, and assign roles to your team members: Contributors can add, update or delete datasets and change dataset access permission; Managers have the authority of contributors, plus the permission to edit your dataverse.

- **Project Information**
  - **Project Title**: China economic history research group
  - **Alias**: econ-han
    - The Alias will be part of the URL of the dataverse. Use all or without space; e.g. alias “project5” will produce the URL "https://dataspase.ust.hk/ds/requestdv/project5"
  - **Project Description**: Datasets collected and generated by the CEHR Group

- **Manager(s)**
  - A manager can add, update or delete datasets, change dataset access permission, and edit the dataverse.
  - **ITSC Account**
    - Franky Cheung
    - Email: frankycheung@ust.hk
  - **Last name**: Lam
  - **First name**: Ki Tat
  - **Department**: LIB
  - **Institution**: HKUST

- **Contributor(s)**
  - A contributor can add, update or delete datasets, change dataset access permission.
  - **ITSC Account**
    - Franky Cheung
    - Email: frankycheung@ust.hk
  - **Last name**: Cheung
  - **First name**: Wing Hong
  - **Department**: LIB
  - **Institution**: HKUST

- **Specify the project and team members**
- **Managers – manage the workspace**
- **Contributors**
  - Can upload and view datasets
  - Allow your non-HKUST collaborators to join
Managing dataset in workspace

1. Tabs - Files, Metadata, Terms of use, Versions
2. Manage permission – i.e. determine who can access this dataset
3. Citation – readily available [Note: no DOI, as it has not yet been published]
4. Upload, update, delete and set restrictions to files
5. Download files – restricted to members you specified in 2.
6. Request Library to publish the dataset
Data services provided by the DataSpace Team

- Help HKUST researchers to deposit (publish) datasets
  - Create standard metadata
  - Assign DOI
  - Format / normalize data files
  - Upload big data files
  - Make datasets accessible via SPD (Scholarly Publications Database)

- Have questions or request a demonstration, contact the Team at lbds@ust.hk
Library Services Supporting Research Output

- Electronic Theses: http://lbezone.ust.hk/rse/electronic-theses
- Scholarly Publications Database: http://repository.ust.hk/ir/sp
- DataSpace: http://dataspace.ust.hk
- Institutional Repository: http://repository.ust.hk/ir
Thank you!

Questions?