Introductory Guide of Open Data for Administrative Staff

Nami Hoshiko (Kyushu University Library)

2015.10.21 第2回 SPARC Japan セミナー2015（オープンアクセス・サミット2015）
Self Introduction

• Chief of the Digital Repository Section at Kyushu University Library

• A member of the working group of SPARC Japan seminar

• I am NOT an open data expert!
Current Situation

• Great impact of the report “Promoting Open Science in Japan”
• “Open Access” is the first priority.
• Administrative office in the university (including library) is not handling research data.

It is difficult for administrative staff to image their specific workflow.
Target of this presentation

• Introductory guide of open data for non-researchers by a non-researcher.

• Sharing basic information of open data to clarify the image of our work as an administrative staff.
(1) the data format must be machine-readable; and
(2) the data should be published under a usage rule that allows secondary use.

"Five stages of open data (Source: *)" and data formats

<table>
<thead>
<tr>
<th>Stage</th>
<th>State of release</th>
<th>Data format example</th>
<th>Ref.) Linked Open Data 5star</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>Release data under an open license</td>
<td>PDF, JPG</td>
<td>OL – Open License (Can be referenced by computer (readable))</td>
</tr>
<tr>
<td>Stage 2</td>
<td>In addition to Stage 1, publish data that can be processed by the computer</td>
<td>xls, doc</td>
<td>RE – Readable (Human &amp; Machine) (Data is editable on the computer)</td>
</tr>
<tr>
<td>Stage 3</td>
<td>In addition to Stage 2, release data in a format that can be openly used</td>
<td>XML, GSV</td>
<td>OF – Open Format (Formats that do not depend on application)</td>
</tr>
<tr>
<td>Stage 4</td>
<td>Release data in a web standard format (such as RDF)</td>
<td>RDF, XML</td>
<td>URI – Universal Resource Identifier (Make resources unique, web links)</td>
</tr>
<tr>
<td>Stage 5</td>
<td>Release data in a state where the four stages can be coordinated externally</td>
<td>LoD, RDF schema</td>
<td>LD – Linked Data (Establishes information fusion between data, Searchable)</td>
</tr>
</tbody>
</table>

Lists of the rare book collection is not downloadable.

The image is not published under a usage rule that allows secondary use.
Decision

Shareability
Case Study: Governments in Japan

“In the past, activities related to open data were frequently interpreted as public disclosures of administrative data and other data possessed by governments.”

Promoting Open Science in Japan — Opening up a new era for the advancement of science— (English version), p8
Case Study: Governments in Japan

• “Let’s Start Open Data”
  http://www.data.go.jp/data/dataset/cas_20150305_0002

オープンデータをはじめよう

〜 地方公共団体のための最初の手引書 〜

オープンデータにむけた6つのステップ（1）

この章では、データをオープンデータとして整備し、公開する作業の6つのステップを説明します。

1. 担当チームを決めよう
2. 現状を把握しよう
3. 公開データの準備をしよう
4. データ公開の仕組みを作ろう
5. データを公開しよう
6. 改善サイクルを回そう
Case Study: Governments in Japan

- Open Data of FUKUOKA CITY（福岡市オープンデータ）
  http://www.open-governmentdata.org/
Case Study: Governments in Japan

- Infrastructure for Multilayer Interoperability (IMI) Core Vocabulary 2.20
  http://imi.ipa.go.jp/ns/core/Core22.html
Differences between research and government data

“Thus, in establishing Japan’s basic policies, we must conduct our investigations with a clear understanding of the differences between the openness of research and government data. “

Promoting Open Science in Japan – Opening up a new era for the advancement of science – (English version), p8
Data Management Plan (DMP)

https://dmptool.org/

https://dmptool.org/plans/8273.pdf
## Digital Data Storage Options for QUT Researchers

<table>
<thead>
<tr>
<th>Decision point(s)</th>
<th>Storage options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are you working with very large datasets?</td>
<td>QUT’s Research Data Storage service</td>
</tr>
<tr>
<td>Are you using active data?</td>
<td>QUT High Performance Computing (HPC) &amp; Research Support</td>
</tr>
<tr>
<td>Are you sharing your data with researchers NOT at QUT?</td>
<td>IHIBI &amp; IFE Storage on U Drive (i.e. QUT institutes)</td>
</tr>
<tr>
<td>Do you require remote access?</td>
<td>U Drive (excluding IHIBI &amp; IFE (not suitable for research data)</td>
</tr>
<tr>
<td></td>
<td>H Drive (eStore) (not suitable for research data)</td>
</tr>
<tr>
<td></td>
<td>QUT Media Warehouse (and other archival options e.g. QUT Wiki)</td>
</tr>
<tr>
<td></td>
<td>QCIF QRIscloud</td>
</tr>
<tr>
<td></td>
<td>AARNet’s CloudStor service</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Yes</strong></th>
<th><strong>Yes/No</strong></th>
<th><strong>Yes/No</strong></th>
<th><strong>Yes/No</strong></th>
<th><strong>Yes/No</strong></th>
<th><strong>Yes/No</strong></th>
<th><strong>Yes/No</strong></th>
<th><strong>Yes/No</strong></th>
<th><strong>Yes/No</strong></th>
<th><strong>&lt;100GB</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>&lt;100GB</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>&lt;100GB</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>&lt;100GB</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>&lt;100GB</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>&lt;2GB</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>&lt;2GB</td>
<td>&lt;2GB</td>
<td>&lt;2GB</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>Yes/No</td>
<td>&lt;2GB</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>&lt;2GB</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>&lt;2GB</td>
</tr>
</tbody>
</table>

1. Dataset size (guide only): *small* = up to 30GB; *medium* = between 30GB and 100GB; *large* = greater than 100GB up to 600GB; and *very large* = greater than 600GB.

2. For the purposes of this document, active or working data are defined as data that require ongoing access for modification, analysis, compilation, etc. Archival storage solutions are more appropriate for "end state" or completed data.

3. The use of sensitive data includes any licensed, proprietary and commercially-produced data.

---

*This document can be downloaded from [http://www.library.qut.edu.au/services/research/data/Policies](http://www.library.qut.edu.au/services/research/data/Policies).

**For more information, please contact the QUT IT Helpdesk [www.ithelpdesk.qut.edu.au](http://www.ithelpdesk.qut.edu.au).**

---

QUT Digital Research Data Storage Options by Queensland University of Technology is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/). To view a copy of this license, visit [http://creativecommons.org/licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/). This work is based on Monash University's "Digital Data Storage Options at Monash."
Roles of Universities

文部科学省 科学技術・学術審議会学術分科会 学術情報委員会
「学術情報のオープン化の推進について（中間まとめ）」より

http://www.mext.go.jp/component/b_menu/shingi/toushin/__icsFiles/afieldfile/2015/10/06/1362565_1.pdf

• Data Management Plan
• Data Storage (Academic Cloud)
• Data Preservation (Digital Object Identifier)
• Assessment
• Data Scientist / Data Curator
Cooperative Relationship

Planning Department

Academics

Systems Department

Research Administrator

Library

Legal Department
Roles of Librarian

1. As a coordinator...

2. As a data curator...