Data Literacy Education Design
Based on Needs of Graduate Students
in University of Chinese Academy of Sciences

ECIL, Saint-Malo, France, Sep. 2017

Wu Ming\textsuperscript{1,2}  Hu Hui\textsuperscript{1}
1 National Science Library, Chinese Academy of Sciences
2 University of Chinese Academy of Sciences
Content

➢ Background
➢ Methodology
➢ Results
➢ Conclusions
1. In the context of e-science and data intensive environment, research data has become an important part of scientific findings. Researchers are encouraged or required to manage and share research data along with their research outputs by a series of research data management and sharing (RDMS) policies.

2. During the past few years, not only the researchers and graduate students have increased attention to RDMS around research lifecycle, but also services relating to research data management and data information literacy (DIL) education have attracted a great deal of attention of library.
Background

3. National Science Library (NSL) is an academic library affiliated to Chinese Academy of Sciences (CAS). In recent year, NSL has set up an subject librarian team with science discipline background. One important role of subject librarians is to expand library services to meet users’ needs.

4. Since graduate students are both as users and as future creators of research data. In this study, NSL cooperated with University of Chinese Academy of Sciences (UCAS), and chose graduate students in life Science College, UCAS to be participants. By means of questionnaires online, we conducted a data literacy survey. Based on a better understanding of their current situations and needs, data literacy education model is designed for library to improve graduate students knowledge and skills in DIL.
Methodology

1. Questionnaire

The 12 competencies of Data Information Literacy were used as a guide in designing the survey. Also further consideration was given to the specific situation of life science and graduate students of UCAS..

Part one: The status of graduates‘ data literacy
✓ Q: Basic idea about research data
✓ Q: Data management plan, collection and documentation
✓ Q: Data processing and analysis
✓ Q: Data preservation/ publication
✓ Q: Research data ethics

Part Two: The needs of graduates’ data literacy
✓ Q: Attitudes towards data literacy education
✓ Q: The contents data literacy education needs
Methodology

2. Participants
The survey was carried out in Sep 2016, 59 graduate students who were enrolling in course offered by life science college, UCAS.
The subject area included genetics, botany, biology, zoology, biochemistry, molecular biology, bioengineering, cell biology, etc.

3. Statistical method
We used excel for statistical analysis of each questions. The percentage was the number proportion of total respondents.
Results

The status of graduate students' data literacy

Basic idea about research data

Q. The importance of RDMS

- Promotes innovation and potential new data uses: 89.83%
- Reduces cost of duplicating data collection: 86.44%
- Encourages improvement and validation of research methods: 83.05%
- Maximizes transparency and accountability: 79.66%
- Leads to new collaborations between data users and creators: 77.97%
- Enables scrutiny of research findings: 76.27%
- Provides important resources for education and training: 69.49%
- Increases the impact and visibility of research: 66.10%
- Other: 5.08%

Q. Policies and requirements of RDMS

- Yes, I know the research data policies and requirements in my laboratory: 50.85%
- Yes, I know the research data policies and requirements in my working organization: 33.90%
- Yes, I know the research data policies and requirements of journal publishers: 25.42%
- Yes, I know the research data policies and requirements of data repository: 22.03%
- Yes, I know the national research data policies and requirements: 18.64%
- Yes, I know the research data policies and requirements of research funding agencies (NSF): 13.56%
- Yes, I know the research data policies and requirements of international organizations: 3.39%
- I don't know (uncertain): 25.42%

- Most of graduate students have a fully awareness about the importance of RDMS.
- But most of them lack knowledge about RDMS policies and requirements. Especially journal publishers, data repositories, research funding agencies data management policies and requirements.
Most of graduate students are clear about life science research data type. But most of them lack knowledge about concept of research data lifecycle.
Most of graduate students understand the components of DMP. But most all of them do not know how to use DMP tool.
Only 59.32% graduate students’ data sources are from data center, data warehouses.

And the most common way of data recording and storing are their own or team portable storage devices and computers.
Data processing and analysis

Q. The methods of recording and storing research data

- USB flash disk or mobile hard disk and other portable storage devices: 94.92%
- Personal Computer: 94.92%
- Paper or Laboratory Notebooks: 93.22%
- Laboratory or office computer: 86.44%
- Online storage space (e.g., Baidu Cloud etc.): 57.63%
- Other: 1.69%
- I don’t know (uncertain): 1.69%

Q. The quality evaluation of research data

- Data authenticity (to ensure the authenticity of the data and objectivity): 98.31%
- Data integrity (check the integrity of data): 91.53%
- Data reproducibility (the frequency of the same experimental results by different experimenters): 83.05%
- Data normative (check representation of data is normative): 88.14%
- Data repeatability (the frequency that the same experimenters do the experiment and get the same experimental results): 81.36%

Tools and software that graduate students mostly use for data processing and analysis are Excel, SPSS, and R and some life science special tools like Primer and SigmaPlot.

Most students think the quality evaluation of research data is very important, especially authenticity, integrity and normative.
Data preservation and publication

**Q. The description of research data (metadata)**

- 94.92% graduate students usually describe research data following rules of their research team or laboratory.
- Many of them know the way to provide DOI or submit to data repository.
Not so many of the graduates clear the research data ethics, such as research data ownership.

Most of graduates would cite the data source normatively and acknowledge to the authors of the data sources.
Results

- The needs of graduate students' data literacy

All of graduates think it necessary to have data literacy education before starting graduates' research and hope to improve their research data management skills. We also investigate the needs of data literacy education around data management lifecycle and education approaches. There are some differences in strength.

<table>
<thead>
<tr>
<th>Data management lifecycle</th>
<th>The needs of data literacy education</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data management plan</td>
<td>Tools of data management plan</td>
<td>89.83%</td>
</tr>
<tr>
<td></td>
<td>Policies and requirements about data management and sharing</td>
<td>81.36%</td>
</tr>
<tr>
<td></td>
<td>Research data life cycle</td>
<td>71.19%</td>
</tr>
<tr>
<td></td>
<td>Elements of data management plan</td>
<td>61.02%</td>
</tr>
<tr>
<td>Data collection and documentation</td>
<td>The type, format and data volume of research data</td>
<td>94.92%</td>
</tr>
<tr>
<td></td>
<td>Quality control and evaluation of research data</td>
<td>83.05%</td>
</tr>
<tr>
<td></td>
<td>How to document and describe research data in fields</td>
<td>81.36%</td>
</tr>
<tr>
<td>Data processing and analysis</td>
<td>Tools to process and analyze research data</td>
<td>98.31%</td>
</tr>
<tr>
<td></td>
<td>Data visualization</td>
<td>76.27%</td>
</tr>
<tr>
<td>Data management and preservation</td>
<td>Research data security</td>
<td>91.53%</td>
</tr>
<tr>
<td></td>
<td>Research data store and backup</td>
<td>88.14%</td>
</tr>
<tr>
<td></td>
<td>Metadata standards for describing research data in fields</td>
<td>81.36%</td>
</tr>
<tr>
<td></td>
<td>Research data naming rules</td>
<td>62.71%</td>
</tr>
<tr>
<td></td>
<td>Research data version control</td>
<td>47.46%</td>
</tr>
<tr>
<td>Data sharing and reuse</td>
<td>Retrieval and acquisition external research data</td>
<td>98.31%</td>
</tr>
<tr>
<td></td>
<td>ways to publish and share research data</td>
<td>88.14%</td>
</tr>
<tr>
<td></td>
<td>Standard for citation of research data</td>
<td>79.66%</td>
</tr>
<tr>
<td></td>
<td>Ethics related to research data and protection measures</td>
<td>77.97%</td>
</tr>
<tr>
<td></td>
<td>Research data sharing and license agreement</td>
<td>69.49%</td>
</tr>
<tr>
<td></td>
<td>Lecture</td>
<td>69.49%</td>
</tr>
<tr>
<td></td>
<td>Online course</td>
<td>66.10%</td>
</tr>
<tr>
<td></td>
<td>WeChat public, blog and other media</td>
<td>67.80%</td>
</tr>
<tr>
<td></td>
<td>Credit course</td>
<td>61.02%</td>
</tr>
<tr>
<td></td>
<td>RSS/Alerts</td>
<td>50.85%</td>
</tr>
<tr>
<td></td>
<td>Workshop</td>
<td>27.12%</td>
</tr>
</tbody>
</table>
Results

Data Literacy Education Design

Analysis results reveal the current situations and needs of graduate students in life science for research data literacy.

Our aim of data literacy education:
- Enhance the graduate students’ awareness of data information literacy (DIL)
- Cultivate the graduate student’s knowledge and skills of DIL

Three levels of learning around research data lifecycle are design:
- Level one - basic learning: develop basic knowledge of DIL
- Level two - advanced learning: learn research data management skills of DIL
- Level three - promotion learning: improve data literacy ethics

The librarian can implement education services though credit curriculum, training and online self-learning for graduate students.
Conclusions

◆ Analysis survey results reveal life science graduate students’ situation and need for data literacy.
◆ Design education three levels to help graduate students improve their data literacy.
◆ Implementation scenes of library services are designed, particularly in credit curriculum, training and online self-learning for graduate students.
◆ The conclusions provide a reference for designing a range of library data literacy education services for other disciplines.
Thanks for your attention!

Acknowledgements:
We would like to thank 59 graduate students of life science in University of Chinese Academy of Sciences (UCAS) who participated in the survey.