Lost or Found? Discovering Data Needed for Research: Supplementary Materials

Kathleen Gregory, Paul Groth, Andrea Scharnhorst, Sally Wyatt

Published on: Apr 30, 2020
Updated on: Apr 24, 2020
DOI: 10.1162/99608f92.30a5e938
Appendix A

Survey questionnaire

Introduction

This study investigates how participants locate and evaluate data they do not create themselves.

The survey consists of three main sections:

• Part 1: Data Needs
• Part 2: Finding Data
• Part 3: Evaluating Data

Our funding comes from the Netherlands Organization for Scientific Research (NWO). The study is part of a collaborative research project between researchers at the Data Archiving and Networked Services (DANS), the University of Amsterdam, the Vrije Universiteit Amsterdam and Elsevier.

By clicking on the below button to start the survey, you indicate your consent to participate in this research. You can read more about the survey and what will be done with the data here (this will launch a new window).

Thank you for your participation.

*Please click >> button to indicate consent to participate and to begin the survey.*

Survey Questions

Part 1: Data Needs

Q1: Which of the following best describes you?

Please select one answer

• Researcher
• Student
• Librarian, archivist or research/data support provider
• Manager
• Other. Please specify ___________

Q2: Please describe the secondary data that you (might) need. (We define secondary data as data that you do not create yourself).

Please write your answer in the box below:

Q3: Please select the options that describe the secondary data that you (might) need.

Please select all that apply

• Observational or empirical (e.g. sensor data, survey data, interview transcripts, sample data, neuroimages, ethnographic data, diaries)
• Experimental (e.g. gene sequences, chromatograms, toroid magnetic field data)
• Simulation (e.g. climate models, economic models)
• Derived or compiled (e.g. text and data mining, compiled database, 3D models)
• Other, Please specify ___________

Q4: Why do you use or need secondary data?

Please select all that apply

• As the basis for a new study
• To calibrate instruments or models
- For benchmarking
- To verify my own data
- As model, algorithm or system inputs
- To generate new ideas
- For teaching/training
- To prepare for a new project or proposal
- To experiment with new methods and techniques (e.g. to develop data science skills)
- To identify trends or make predictions
- To compare multiple datasets to find commonalities or differences
- To create summaries, visualizations, or analysis tools
- To integrate with other data to create a new dataset
- Other. Please specify ____________

Q5: Have you ever used data outside of your area of expertise?

*Please select one answer*

- Yes
- No

Q5a: How did you find this data?

*Please write your answer in the box below:

---

Part 2: Finding Data

Q6: When you need data, who finds it for you?

*Please select all that apply*
• I find it myself
• Graduate student
• Research support professional (e.g. librarian, archivist, data or literature manager)
• Someone else in my personal network (e.g. peers, collaborators, mentors)
• Other. Please specify ____________

Q7: How frequently do you use the following to find data?

Please select one answer per row

<table>
<thead>
<tr>
<th></th>
<th>Often</th>
<th>Occasionally</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multidisciplinary data repositories</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discipline-specific data repositories</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Governmental agencies and websites</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal networks (e.g. colleagues, peers)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic literature (e.g. journal articles, conference proceedings)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code repository (e.g. GitHub)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General search engines (e.g. Google)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional associations</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Q7_open: Please specify any other resources that you use to find data:

Please write your answer in the box below:

Q7a: Which statement(s) describe how you discover data using the academic literature?

Please select all that apply

- I search the academic literature with the goal of finding data.
- I find data serendipitously while reading articles or performing literature searches.
- I follow citations and references in the literature to datasets.
- I extract and use data from the literature directly (e.g. from tables, graphs, or instrument specifications and parameters)
- Other. Please specify ___________

Q7b How successful are you at finding data with a general search engine (e.g. Google)?

Please select one answer

- Very successful
- Successful
- Sometime successful, sometimes not
- Rarely successful
- Not successful

**Q8: How frequently do you find data in the following ways?**

*Please select one answer per row*

<table>
<thead>
<tr>
<th></th>
<th>Often</th>
<th>Occasionally</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>By actively searching for data in an online resource</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serendipitously, when searching for something else (e.g. when looking for journal articles or news)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serendipitously, when NOT actively looking for something else (e.g. via an email notice or interaction with a colleague)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the course of sharing or managing my own data</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Q9: Please indicate if you use the following to discover, access, or make sense of data.**

*Please select all that apply*

<table>
<thead>
<tr>
<th></th>
<th>Q10a - Discover</th>
<th>Q10b - Access</th>
<th>Q10c - Making sense of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversations with personal networks (e.g. colleagues, peers)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contacting the data creator</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Q10: Do you discover data differently than how you discover academic literature?

Please select one answer

• Yes
• Sometimes
• No

Q10a: How is your process for finding data different than your process for finding academic literature?

Please write your answer in the box below:

Q11: How easy is it to find data?

Please select one answer

• Easy
• Sometimes challenging
• Difficult

Q11a: Why is it challenging to find the data that you need?

Please select all that apply

• The data are not accessible (e.g. behind paywalls, held by industry).
• I don’t know where or how to best look for the data.
• The data are located in many different places.
• The data are not digital.
• Online search tools are inadequate.
• I do not have the personal network needed to find or access the data.
• Other. Please specify ____________

Part 3: Evaluating Data

Q12: Please indicate the importance of the following information when deciding whether or not to use secondary data.

*Please select one answer per row*

<table>
<thead>
<tr>
<th></th>
<th>Extremely important</th>
<th>Important</th>
<th>Somewhat important</th>
<th>Less important</th>
<th>Not important</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection conditions and methodology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How data has been processed and handled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reputation of data creator</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personally knowing the data creator</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reputation of data source (e.g. repository or journal)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q12_open: Please specify any other information you consider when deciding whether to use or not secondary data.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q13: How important are the following strategies in evaluating and making sense of data?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Please select one answer per row</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detailed and complete metadata and documentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data format</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Licensing/copyright conditions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct coverage (time, location, population, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original purpose of the data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ease of access</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic relevance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Consulting associated journal articles | Extremely important | Important | Somewhat important | Less important | Not important |
Q13_open: Please specify any other strategies you consider to evaluate and make sense of data.

Please write your answer in the box below:

<table>
<thead>
<tr>
<th>Question</th>
<th>Extremely important</th>
<th>Important</th>
<th>Somewhat important</th>
<th>Less important</th>
<th>Not important</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consulting data documentation and codebooks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consulting the data creator</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consulting personal networks (e.g. colleagues, peers)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exploratory data analysis (e.g. statistical checks, graphical analysis)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q14: Please indicate the importance of the following in helping you to establish trust in secondary data.

Please select one answer per row

<table>
<thead>
<tr>
<th>Question</th>
<th>Extremely important</th>
<th>Important</th>
<th>Somewhat important</th>
<th>Less important</th>
<th>Not important</th>
</tr>
</thead>
<tbody>
<tr>
<td>Others’ prior usage of the data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reputation of source (e.g. repository, journal)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Q14_open: Please specify any other important aspects you consider to help establish trust in secondary data.

Please write your answer in the box below:

### Q15: Please indicate the importance of the following in helping you to establish the quality of secondary data.

Please select one answer per row:

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Extremely important</th>
<th>Important</th>
<th>Somewhat important</th>
<th>Less important</th>
<th>Not important</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ease of downloading and exploring data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data completeness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reputatcion of source (e.g. repository, journal)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution or clarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reputation of data creator</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detail or amount of work done to prepare data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistency of formatting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Q15_open**: Please specify any other important aspects you consider to help establish the quality of secondary data.

*Please write your answer in the box below:*

---

**Part 4: Demographics**

You are nearly at the end of the survey. Below are some questions to help us classify your answers.

**D1**: In which subject discipline do you specialize?

*Please check all that apply.*
D2: How many years of professional experience do you have in your field?

Please select one answer

- 0-5
- 6-15
• 16-30
• 31+

D3: In which county do you currently work?

• Afghanistan
• Albania
• Algeria
• American Samoa
• Andorra
• Angola
• Anguilla
• Antarctica
• Antigua and Barbuda
• Argentina
• Armenia
• Aruba
• Australia
• Austria
• Azerbaijan
• Bahamas
• Bahrain
• Bangladesh
• Barbados
• Belarus
• Belgium
• Belize
• Benin
• Bermuda
• Bhutan
• Bolivia
• Bosnia and Herzegovina
• Botswana
• Brazil
• British Indian Ocean Territory
• Brunei
• Brunei Darussalam
• Bulgaria
• Burkina Faso
• Burundi
• Cambodia
• Cameroon
• Canada
• Cape Verde
• Cayman Islands
• Central African Republic
• Chad
• Chile
• China
• Christmas Island
• Cocos (Keeling) Islands
- Colombia
- Comoros
- Congo
- Cook Islands
- Costa Rica
- Cote d'Ivoire
- Croatia
- Cuba
- Cyprus
- Czech Republic
- Denmark
- Djibouti
- Dominica
- Dominican Republic
- East Timor
- Ecuador
- Egypt
- El Salvador
- Equatorial Guinea
- Eritrea
- Estonia
- Ethiopia
- Falkland Islands (Malvinas)
- Fiji
- Finland
• France
• French Guiana
• French Polynesia
• French Southern Territories
• Gambia
• Georgia
• Germany
• Ghana
• Gibraltar
• Greece
• Greenland
• Grenada
• Guadeloupe
• Guam
• Guatemala
• Guinea-Bissau
• Haiti
• Heard Island and McDonald Islands
• Holy See (Vatican City State)
• Honduras
• Hong Kong
• Hungary
• Iceland
• India
• Indonesia
- Iran (Islamic Republic of)
- Iraq
- Ireland
- Israel
- Italy
- Jamaica
- Japan
- Jordan
- Kazakhstan
- Kenya
- Kiribati
- North Korea
- Kuwait
- Kyrgyzstan
- Lao People's Democratic Republic
- Laos
- Latvia
- Lebanon
- Lesotho
- Liberia
- Libyan Arab Jamahiriya
- Lithuania
- Luxembourg
- Macau
- Madagascar
• Malawi
• Malaysia
• Maldives
• Mali
• Malta
• Martinique
• Mauritania
• Mauritius
• Mexico
• Micronesia (Federated States of)
• Monaco
• Mongolia
• Montserrat
• Morocco
• Mozambique
• Myanmar
• Namibia
• Nauru
• Nepal
• Netherlands
• Netherlands Antilles
• New Caledonia
• New Zealand
• Nicaragua
• Niger
- Nigeria
- Niue
- Norfolk Island
- Norway
- Oman
- Pakistan
- Palau
- Panama
- Papua New Guinea
- Paraguay
- Peru
- Philippines
- Pitcairn
- Poland
- Portugal
- Puerto Rico
- Qatar
- Reunion
- Romania
- RUSSIA
- Rwanda
- Saint Helena
- Saint Kitts and Nevis
- Saint Lucia
- Saint Vincent and the Grenadines
• Samoa
• Sao Tome and Principe
• Saudi Arabia
• Senegal
• Serbia and Montenegro
• Seychelles
• Sierra Leone
• Singapore
• Slovakia
• Slovenia
• Solomon Islands
• Somalia
• South Africa
• South Korea
• Spain
• Sri Lanka
• Sudan
• Suriname
• Swaziland
• Sweden
• Switzerland
• Syrian Arab Republic
• Taiwan
• Tajikistan
• TANZANIA
• Thailand
• Togo
• Tonga
• Trinidad and Tobago
• Tunisia
• Turkey
• Turkmenistan
• Turks and Caicos Islands
• Uganda
• Ukraine
• United Arab Emirates
• United Kingdom
• United States Minor Outlying Islands
• Uruguay
• USA
• Uzbekistan
• Vanuatu
• Venezuela
• Viet Nam
• Virgin Islands
• Virgin Islands (US)
• Virgin Islands, British
• Wallis and Futuna
• Yemen
• Zambia
• Zimbabwe
• Palestinian Territory, Occupied
• Moldova, Republic of
• Marshall Islands
• Macedonia, The Former Yugoslav Republic of
• Liechtenstein
• Korea, Republic of
• Guyana
• Guinea
• Gabon
• Faroe Islands
• Zanzibar
• Tokelau

**D4: What type of organization do you work for?**

*Please select one answer*

• University or college
• Research institution
• Government agency
• Corporate
• Independent archive or library
• m Other. Please specify ____________

**D5: Please indicate how the following people feel about sharing their research data.**

*Please select one answer per row*
# D6: Please indicate how the following people feel about reusing data produced by other people.

*Please select one answer per row*

<table>
<thead>
<tr>
<th></th>
<th>Data sharing is strongly encouraged</th>
<th>Data sharing is somewhat encouraged</th>
<th>Data sharing is neither encouraged nor discouraged</th>
<th>Data sharing is somewhat discouraged</th>
<th>Data sharing is strongly discouraged</th>
<th>Don't know/Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>You</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The people you work with directly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Your disciplinary community</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Your institution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Data reusing is strongly encouraged</th>
<th>Data reusing is somewhat encouraged</th>
<th>Data reusing is neither encouraged nor discouraged</th>
<th>Data reusing is somewhat discouraged</th>
<th>Data reusing is strongly discouraged</th>
<th>Don't know/Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>You</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The people you work with directly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Your disciplinary community</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Your institution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
D7: Have you ever shared your own research data?

*Please select one answer*

- Yes
- No

D8: Final comments: Do you have anything else that you would like us to know?

*Please write your comments in the box below:

Additional questions asked to participants selecting “Librarian, archivist or research/data support provider” as their role.

L3: Do you use or need secondary data for your own research or to support others?

*Please select one answer*

- For my own research
- To support others
- For both my own research and to support others

L4: Who are the people whom you support?

*Please select all that apply*

- Students
- Researchers
- Industry employees
- Other. Please specify ___________

L5: How do you support people with their data needs?
Please select all that apply

- I teach people about data management planning (e.g. through consultations, workshops, etc.).
- I teach people how to discover and evaluate data (e.g. through consultations, workshops, etc.).
- I find data for people.
- I help people to curate their data.
- I find literature for people.
- Other. Please specify ____________

Appendix B

P-Value Tables
<table>
<thead>
<tr>
<th>Field</th>
<th>Observ/empirical</th>
<th>Experimental</th>
<th>Deriv/compiled</th>
<th>Simulation</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricul</td>
<td>0.647</td>
<td>p &lt; .001*</td>
<td>0.281</td>
<td>0.066</td>
<td>0.59</td>
</tr>
<tr>
<td>Arts hum</td>
<td>1.000</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
</tr>
<tr>
<td>Astronom</td>
<td>0.103</td>
<td>0.92</td>
<td>0.008</td>
<td>p &lt; .001*</td>
<td>0.92</td>
</tr>
<tr>
<td>Biochem</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>0.92</td>
<td>0.003</td>
<td>0.322</td>
</tr>
<tr>
<td>Biolog</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>0.069</td>
<td>0.091</td>
<td>0.806</td>
</tr>
<tr>
<td>Busin</td>
<td>0.001</td>
<td>p &lt; .001*</td>
<td>0.054</td>
<td>0.718</td>
<td>0.267</td>
</tr>
<tr>
<td>Chem</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>0.647</td>
<td>0.624</td>
<td>0.446</td>
</tr>
<tr>
<td>Chem Eng</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>1.000</td>
<td>0.002</td>
<td>0.862</td>
</tr>
<tr>
<td>Comp Sci</td>
<td>0.92</td>
<td>0.032</td>
<td>p &lt; .001*</td>
<td>0.001</td>
<td>0.203</td>
</tr>
<tr>
<td>Decis Sci</td>
<td>0.275</td>
<td>0.138</td>
<td>p &lt; .001</td>
<td>0.001</td>
<td>0.752</td>
</tr>
<tr>
<td>Dentist</td>
<td>0.39</td>
<td>0.023</td>
<td>0.296</td>
<td>0.11</td>
<td>p &lt; .001*</td>
</tr>
<tr>
<td>Earth Plan</td>
<td>0.003</td>
<td>1.000</td>
<td>0.02</td>
<td>p &lt; .001*</td>
<td>0.92</td>
</tr>
<tr>
<td>Econ</td>
<td>p &lt; .001</td>
<td>p &lt; .001*</td>
<td>1.000</td>
<td>p &lt; .001*</td>
<td>0.232</td>
</tr>
<tr>
<td>Energy</td>
<td>0.087</td>
<td>0.001</td>
<td>0.862</td>
<td>p &lt; .001*</td>
<td>0.603</td>
</tr>
<tr>
<td>Eng Tech</td>
<td>0.009</td>
<td>0.022</td>
<td>0.187</td>
<td>p &lt; .001*</td>
<td>0.488</td>
</tr>
<tr>
<td>Environ</td>
<td>0.098</td>
<td>0.841</td>
<td>0.647</td>
<td>p &lt; .001*</td>
<td>0.791</td>
</tr>
<tr>
<td>Health Prof</td>
<td>p &lt; .001</td>
<td>0.002</td>
<td>0.458</td>
<td>0.001</td>
<td>0.841</td>
</tr>
<tr>
<td>Immun</td>
<td>0.03</td>
<td>p &lt; .001*</td>
<td>0.462</td>
<td>0.001</td>
<td>0.538</td>
</tr>
<tr>
<td>Info Sci</td>
<td>0.002</td>
<td>0.001</td>
<td>p &lt; .001*</td>
<td>0.031</td>
<td>0.004</td>
</tr>
<tr>
<td>Math</td>
<td>0.145</td>
<td>0.45</td>
<td>0.029</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
</tr>
<tr>
<td>Matl Sci</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>0.103</td>
<td>0.008</td>
<td>0.59</td>
</tr>
<tr>
<td>Med</td>
<td>p &lt; .001*</td>
<td>0.187</td>
<td>0.025</td>
<td>p &lt; .001*</td>
<td>0.475</td>
</tr>
<tr>
<td>Multi</td>
<td>0.005</td>
<td>0.92</td>
<td>0.013</td>
<td>0.002</td>
<td>0.823</td>
</tr>
<tr>
<td>Neuro</td>
<td>0.137</td>
<td>0.123</td>
<td>0.764</td>
<td>0.09</td>
<td>0.639</td>
</tr>
<tr>
<td>Nurs</td>
<td>0.187</td>
<td>0.037</td>
<td>0.252</td>
<td>0.144</td>
<td>0.729</td>
</tr>
<tr>
<td>Pharma</td>
<td>0.03</td>
<td>p &lt; .001*</td>
<td>0.639</td>
<td>0.079</td>
<td>0.538</td>
</tr>
<tr>
<td>Physics</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>0.806</td>
<td>p &lt; .001*</td>
<td>0.094</td>
</tr>
<tr>
<td>Psych</td>
<td>p &lt; .001*</td>
<td>0.006</td>
<td>0.458</td>
<td>0.001</td>
<td>0.862</td>
</tr>
<tr>
<td>Soc Sci</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>0.332</td>
<td>0.005</td>
<td>0.791</td>
</tr>
<tr>
<td>Vet</td>
<td>0.729</td>
<td>p &lt; .001*</td>
<td>0.092</td>
<td>0.046</td>
<td>0.21</td>
</tr>
<tr>
<td>Other</td>
<td>0.071</td>
<td>0.037</td>
<td>0.377</td>
<td>0.074</td>
<td>0.035</td>
</tr>
</tbody>
</table>

**Table B1. P-Value Table for Figure 6: Associations Between Disciplinary Domain and Needed Data**

*Note.* Significance was determined at the $p < .05$ level with a Bonferroni correction with $m = 155$. Significant associations are marked with an asterisk and colored in blue.
<table>
<thead>
<tr>
<th></th>
<th>Observ/empirical</th>
<th>Experimental</th>
<th>Deriv/compiled</th>
<th>Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>New study</td>
<td>p &lt; .001*</td>
<td>0.044</td>
<td>0.007</td>
<td>0.566</td>
</tr>
<tr>
<td>New projects</td>
<td>0.001</td>
<td>0.043</td>
<td>0.09</td>
<td>0.764</td>
</tr>
<tr>
<td>New ideas</td>
<td>0.124</td>
<td>0.005</td>
<td>0.017</td>
<td>0.153</td>
</tr>
<tr>
<td>Integration</td>
<td>p &lt; .001*</td>
<td>0.512</td>
<td>p &lt; .001*</td>
<td>0.003</td>
</tr>
<tr>
<td>New methods</td>
<td>0.33</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
</tr>
<tr>
<td>Inputs</td>
<td>0.135</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
</tr>
<tr>
<td>Calibration</td>
<td>0.242</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
</tr>
<tr>
<td>Benchmark</td>
<td>0.017</td>
<td>0.095</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
</tr>
<tr>
<td>Verification</td>
<td>0.251</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
</tr>
<tr>
<td>Trends</td>
<td>p &lt; .001*</td>
<td>0.888</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
</tr>
<tr>
<td>Comparison</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
<td>0.002</td>
</tr>
<tr>
<td>Summ/vis/tools</td>
<td>p &lt; .001*</td>
<td>0.292</td>
<td>p &lt; .001*</td>
<td>p &lt; .001*</td>
</tr>
<tr>
<td>Teaching</td>
<td>p &lt; .001*</td>
<td>0.624</td>
<td>0.151</td>
<td>0.01</td>
</tr>
<tr>
<td>Other</td>
<td>0.003</td>
<td>0.095</td>
<td>0.101</td>
<td>0.06</td>
</tr>
</tbody>
</table>

**Table B2. P-Value Table for Table 4: Associations Between Types of Data Use and Needed Data Type**

*Note.* Significance was determined at the \( p < .05 \) level with a Bonferroni correction with \( m = 70 \). Significant associations are marked with an asterisk and colored in blue. “Other” options are not shown as there were no significant associations present.

| New project | New project | New ideas | Integration | New methods | Inputs | Calibration | Benchmark | Verification | Trends | Comparison | Summ/vis/tools |
|-------------|-------------|-----------|-------------|-------------|--------|-------------|-----------|--------------|--------|------------|----------------|}
| p < .001*   | p < .001*   | p < .001* | p < .001*   | p < .001*   | p < .001* | p < .001*   | 0.001    | p < .001*    | 0.001  | p < .001*   | p < .001*       |
| 0.148       | 0.133       | 0.764     | 0.003       | p < .001*   | p < .001* | p < .001*   | 0.001    | p < .001*    | 0.007  | p < .001*   | p < .001*       |
| 0.148       | 0.133       | 0.764     | 0.003       | p < .001*   | p < .001* | p < .001*   | 0.001    | p < .001*    | 0.007  | p < .001*   | p < .001*       |

**Table B3. P-Value Table for Table 4: Associations Between Types of Data Use and Other Data Uses**

*Note.* Significance was determined at the \( p < .05 \) level with a Bonferroni correction with \( m = 196 \). Significant associations are marked with an asterisk and colored in blue. “Other” options are not shown as there were no significant associations present; duplicate values were removed.
Note. Significance was determined at the $p < .05$ level with a Bonferroni correction with $m = 434$. Significant associations are marked with an asterisk and colored in blue. “Other” options are not shown as there were no significant associations present.

Table B4. $P$-Value Table for Figure 8: Associations Between Disciplinary Domain and Data Use

<table>
<thead>
<tr>
<th>Conditions/methodology</th>
<th>New study</th>
<th>New project</th>
<th>New ideas</th>
<th>Integration</th>
<th>New methods</th>
<th>Calibrati</th>
<th>Benchmark</th>
<th>Verification</th>
<th>Trends</th>
<th>Comparison</th>
<th>Summ/vis/tools</th>
<th>Teaching</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p &lt; .001$</td>
<td>$p &lt; .05$</td>
<td>$p &lt; .01$</td>
<td>0.138</td>
<td>0.225</td>
<td>0.125</td>
<td>0.138</td>
<td>0.218</td>
<td>$p &lt; .001$</td>
<td>0.07</td>
<td>0.016</td>
<td>0.1</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>Processing/handling</td>
<td>$p &lt; .001$</td>
<td>$p &lt; .05$</td>
<td>$p &lt; .01$</td>
<td>0.04</td>
<td>0.013</td>
<td>0.371</td>
<td>0.006</td>
<td>0.527</td>
<td>$0.004^*$</td>
<td>$p &lt; .001^*$</td>
<td>0.04</td>
<td>0.016</td>
<td>0.1</td>
</tr>
<tr>
<td>Topic relevance</td>
<td>$p &lt; .001$</td>
<td>$p &lt; .05$</td>
<td>$p &lt; .01$</td>
<td>0.074</td>
<td>0.009</td>
<td>0.481</td>
<td>0.358</td>
<td>0.458</td>
<td>0.613</td>
<td>0.038</td>
<td>0.623</td>
<td>0.031</td>
<td>0.17</td>
</tr>
<tr>
<td>Coverage</td>
<td>$p &lt; .022$</td>
<td>0.167</td>
<td>0.331</td>
<td>$p &lt; .001$</td>
<td>0.093</td>
<td>0.012</td>
<td>0.517</td>
<td>0.199</td>
<td>0.866</td>
<td>$p &lt; .001^*$</td>
<td>0.011</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td>Original purpose</td>
<td>$0.028$</td>
<td>0.431</td>
<td>0.493</td>
<td>0.009</td>
<td>0.162</td>
<td>0.92</td>
<td>0.68</td>
<td>0.039</td>
<td>0.137</td>
<td>0.035</td>
<td>0.105</td>
<td>0.439</td>
<td>0.284</td>
</tr>
<tr>
<td>Metadata/documentation</td>
<td>0.313</td>
<td>0.338</td>
<td>0.046</td>
<td>$p &lt; .001^*$</td>
<td>0.012</td>
<td>0.318</td>
<td>0.561</td>
<td>0.506</td>
<td>0.044</td>
<td>$p &lt; .001^*$</td>
<td>0.67</td>
<td>0.018</td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td>0.798</td>
<td>0.58</td>
<td>0.351</td>
<td>0.229</td>
<td>0.103</td>
<td>0.323</td>
<td>0.263</td>
<td>0.714</td>
<td>0.692</td>
<td>0.215</td>
<td>0.279</td>
<td>0.22</td>
<td>0.002</td>
</tr>
<tr>
<td>Size</td>
<td>0.949</td>
<td>0.42</td>
<td>0.238</td>
<td>0.58</td>
<td>0.159</td>
<td>0.082</td>
<td>0.486</td>
<td>0.738</td>
<td>0.706</td>
<td>0.113</td>
<td>0.281</td>
<td>0.69</td>
<td>0.013</td>
</tr>
<tr>
<td>Ease of access</td>
<td>0.622</td>
<td>0.037</td>
<td>0.002</td>
<td>0.278</td>
<td>0.765</td>
<td>0.917</td>
<td>0.136</td>
<td>0.661</td>
<td>0.12</td>
<td>0.074</td>
<td>0.334</td>
<td>0.571</td>
<td>0.005</td>
</tr>
<tr>
<td>Licensing</td>
<td>0.606</td>
<td>0.182</td>
<td>0.202</td>
<td>0.100</td>
<td>0.078</td>
<td>0.146</td>
<td>0.088</td>
<td>0.087</td>
<td>0.916</td>
<td>0.297</td>
<td>0.369</td>
<td>0.004</td>
<td>0.001</td>
</tr>
<tr>
<td>Source reputation</td>
<td>$p &lt; .001$</td>
<td>0.002</td>
<td>0.002</td>
<td>0.332</td>
<td>0.179</td>
<td>0.582</td>
<td>0.013</td>
<td>0.499</td>
<td>0.011</td>
<td>0.076</td>
<td>0.886</td>
<td>0.387</td>
<td>0.001</td>
</tr>
<tr>
<td>Creator reputation</td>
<td>0.049</td>
<td>0.017</td>
<td>0.035</td>
<td>0.368</td>
<td>0.654</td>
<td>0.565</td>
<td>0.383</td>
<td>0.475</td>
<td>0.029</td>
<td>0.01</td>
<td>0.523</td>
<td>0.358</td>
<td>0.042</td>
</tr>
<tr>
<td>Knowing creator</td>
<td>0.377</td>
<td>0.427</td>
<td>0.117</td>
<td>0.201</td>
<td>0.051</td>
<td>0.226</td>
<td>0.666</td>
<td>0.835</td>
<td>0.329</td>
<td>0.015</td>
<td>0.634</td>
<td>0.772</td>
<td>0.717</td>
</tr>
</tbody>
</table>

Table B5. $P$-Value Table for Figure 15: Associations Between Data Use and Evaluation Criteria

Note. Significance was determined at the $p < .05$ level with a Bonferroni correction with $m = 196$. Significant associations are marked with an asterisk and colored in blue. “Other” options are not shown as there were no significant associations present.
Appendix C

Sources Used in Disciplinary Subset
Figure C1. Sources used in the disciplinary subset for respondents selecting only one discipline. Percents are percent respondents. Arts & humanities ($n=43$); astronomy ($n=14$); biological science ($n=46$); computer science ($n=57$); earth & planetary science ($n=24$); engineering & technology ($n=80$); environmental science ($n=22$); medicine ($n=91$); physics ($n=42$); social science ($n=81$).