

Harvard Data Science Review

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](https://doi.org/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

	Often	Occasionally	Never
Multidisciplinary data repositories			
Discipline-specific data repositories			
Governmental agencies and websites			
Personal networks (e.g. colleagues, peers)			
Academic literature (e.g. journal articles, conference proceedings)			
Code repository (e.g. GitHub)			
General search engines (e.g. Google)			
Professional associations			

Data specific search engines			
Commercial sources			
Consultation with research support professionals (e.g. librarians, archivists or data managers)			

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

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

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

Please select all that apply

	Q10a - Discover	Q10b - Access	Q10c - Making sense of data
Conversations with personal networks (e.g. colleagues, peers)			
Contacting the data creator			

Developing new academic collaborations with data creators			
Attending conferences			
Disciplinary mailing lists or discussion forums			

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

	Extremely important	Important	Somewhat important	Less important	Not important
Data collection conditions and methodology					
How data has been processed and handled					
Reputation of data creator					
Personally knowing the data creator					
Reputation of data source (e.g. repository or journal)					

Detailed and complete metadata and documentation					
Data size					
Data format					
Licensing/copyright conditions					
Correct coverage (time, location, population, etc.)					
Original purpose of the data					
Ease of access					
Topic relevance					

Q12_open: Please specify any other information you consider when deciding whether to use or not secondary data.

Q13: How important are the following strategies in evaluating and making sense of data?

Please select one answer per row

	Extremely important	Important	Somewhat important	Less important	Not important
Consulting associated journal articles					

Consulting data documentation and codebooks					
Consulting the data creator					
Consulting personal networks (e.g. colleagues, peers)					
Exploratory data analysis (e.g. statistical checks, graphical analysis)					

Q13_open: Please specify any other strategies you consider to evaluate and make sense of data.

Please write your answer in the box below:

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

Please select one answer per row

	Extremely important	Important	Somewhat important	Less important	Not important
Others' prior usage of the data					
Reputation of source (e.g. repository, journal)					

Reputation of data creator					
Transparency in data collection methods					
Lack of errors					
Ease of access					
Personal relationship with the data creator					

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

	Extremely important	Important	Somewhat important	Less important	Not important
Lack of errors					
Ease of downloading and exploring data					
Data size					
Data completeness					

Reputation of source (e.g. repository, journal)					
Resolution or clarity					
Reputation of data creator					
Detail or amount of work done to prepare data					
Consistency of formatting					

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.

- | | |
|---|---|
| <ul style="list-style-type: none">• Agriculture• Arts and Humanities• Astronomy• Biochemistry, Genetics, and Molecular Biology• Biological Sciences• Business, Management and Accounting• Chemical Engineering• Chemistry• Computer Sciences / IT• Decision Sciences• Dentistry• Earth and Planetary Sciences• Economics, Econometrics and Finance• Energy• Engineering and Technology• Environmental Sciences | <ul style="list-style-type: none">• Health professions• Immunology and Microbiology• Materials Science• Mathematics• Medicine• Multidisciplinary• Neuroscience• Nursing• Pharmacology, Toxicology and Pharmaceutics• Physics• Psychology• Social Science• Veterinary• Information science• Other. Please specify_____ |
|---|---|

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

	Data sharing is strongly encouraged	Data sharing is somewhat encouraged	Data sharing is neither encouraged nor discouraged	Data sharing is somewhat discouraged	Data sharing is strongly discouraged	Don't know/ Not applicable
You						
The people you work with directly						
Your disciplinary community						
Your institution						

D6: Please indicate how the following people feel about reusing data produced by other people.

Please select one answer per row

	Data reusing is strongly encouraged	Data reusing is somewhat encouraged	Data reusing is neither encouraged nor discouraged	Data reusing is somewhat discouraged	Data reusing is strongly discouraged	Don't know/ Not applicable
You						
The people you work with directly						
Your disciplinary community						
Your institution						

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**

	Observ/empirical	Experimental	Deriv/compiled	Simulation	Other
Agricul	0.647	p < .001*	0.281	0.066	0.59
Arts hum	1.000	p < .001*	p < .001*	p < .001*	p < .001*
Astronom	0.103	0.92	0.008	p < .001*	0.92
Biochem	p < .001*	p < .001*	0.92	0.003	0.322
Biolog	p < .001*	p < .001*	0.069	0.091	0.806
Busin	0.001	p < .001*	0.054	0.718	0.267
Chem	p < .001*	p < .001*	0.647	0.624	0.446
Chem Eng	p < .001*	p < .001*	1.000	0.002	0.862
Comp Sci	0.92	0.032	p < .001*	0.001	0.203
Decis Sci	0.275	0.138	p < .001	0.001	0.752
Dentist	0.39	0.023	0.296	0.11	p < .001*
Earth Plan	0.003	1.000	0.02	p < .001*	0.92
Econ	p < .001	p < .001*	1,000	p < .001*	0.232
Energy	0.087	0.001	0.862	p < .001*	0.603
Eng Tech	0.009	0.022	0.187	p < .001*	0.488
Environ	0.098	0.841	0.647	p < .001*	0.791
Health Prof	p < .001	0.002	0.458	0.001	0.841
Immun	0.03	p < .001*	0.462	0.001	0.538
Info Sci	0.002	0.001	p < .001*	0.031	0.004
Math	0.145	0.45	0.029	p < .001*	p < .001*
Matl Sci	p < .001*	p < .001*	0.103	0.008	0.59
Med	p < .001*	0.187	0.025	p < .001*	0.475
Multi	0.005	0.92	0.013	0.002	0.823
Neuro	0.137	0.123	0.764	0.09	0.639
Nurs	0.187	0.037	0.252	0.144	0.729
Pharma	0.03	p < .001*	0.639	0.079	0.538
Physics	p < .001*	p < .001*	0.806	p < .001*	0.094
Psych	p < .001*	0.006	0.458	0.001	0.862
Soc Sci	p < .001*	p < .001*	0.332	0.005	0.791
Vet	0.729	p < .001*	0.092	0.046	0.21
Other	0.071	0.037	0.377	0.074	0.035

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.

	Observ/empirical	Experimental	Deriv/compiled	Simulation
New study	p < .001*	0.044	0.007	0.566
New projects	0.001	0.043	0.09	0.764
New ideas	0.124	0.005	0.017	0.153
Integration	p < .001*	0.512	p < .001*	0.003
New methods	0.33	p < .001*	p < .001*	p < .001*
Inputs	0.135	p < .001	p < .001*	p < .001*
Calibration	0.242	p < .001*	p < .001*	p < .001*
Benchmark	0.017	0.095	p < .001*	p < .001*
Verification	0.251	p < .001*	p < .001*	p < .001*
Trends	p < .001*	0.888	p < .001*	p < .001*
Comparison	p < .001*	p < .001*	p < .001*	0.002
Summ/vis/tools	p < .001*	0.292	p < .001*	p < .001*
Teaching	p < .001*	0.624	0.151	0.01
Other	0.003	0.095	0.101	0.06

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 study	New project	New ideas	Integration	New methods	Inputs	Calibration	Benchmark	Verification	Trends	Comparison	Summ/vis/tools
New project	p < .001*											
New ideas	p < .001*	p < .001*										
Integration	p < .001*	p < .001*	p < .001*									
New methods	p < .001	p < .001*	p < .001*	p < .001*								
Inputs	0.097	0.245	0.028	p < .001*	p < .001*							
Calibration	0.31	0.157	1.000	0.001	p < .001*	p < .001*						
Benchmark	0.148	0.133	0.764	0.003	p < .001*	p < .001*	p < .001*					
Verification	0.001	p < .001*	p < .001*	p < .001	p < .001*	0.92	p < .001*	p < .001*				
Trends	p < .001*	p < .001*	p < .001*	p < .001*	p < .001*	p < .001*	0.003	p < .001*	0.001			
Comparison	p < .001*	p < .001*	p < .001*	p < .001*	p < .001*	p < .001*	0.017	0.001	p < .001*	p < .001*		
Summ/vis/tools	0.013	p < .001*	p < .001*	p < .001*	p < .001*	p < .001*	0.001	p < .001*	0.072	p < .001*	p < .001*	
Teaching	p < .001*	p < .001*	p < .001*	0.005	p < .001	0.063	0.699	0.045	p < .001*	0.016	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.

	New study	New project	New ideas	Integration	New methods	Inputs	Calibration	Benchmark	Verification	Trends	Comparison	Summ/vis/tools	Teaching	Other
Agricul	0.689	0.377	0.225	0.114	0.034	0.374	0.862	0.862	0.007	0.139	0.517	0.365	0.175	0.549
Arts hum	0.083	0.186	0.086	0.045	0.024	0.008	0.08	0.185	1.000	0.532	0.42	0.442	0.001	0.92
Astronom	0.31	0.264	1.000	0.067	0.417	0.007	p < .001*	0.112	0.327	0.008	0.016	0.348	0.104	0.403
Biochem	0.68	0.052	p < .001	0.371	p < .001	0.806	0.192	0.21	p < .001*	0.48	0.001	0.549	0.862	0.371
Biology	1.000	0.029	0.038	p < .001*	0.065	0.114	0.104	0.002	p < .001	0.04	p < .001*	0.699	0.488	0.699
Busin	0.025	0.806	0.359	0.175	0.888	0.23	0.823	0.004	0.202	p < .001*	0.56	0.357	0.68	0.671
Chem	0.699	0.699	0.435	0.124	0.071	0.081	0.06	0.777	0.001	0.374	0.345	0.086	0.269	0.439
Chem Eng	0.467	0.462	0.578	0.484	0.001	0.051	p < .001*	0.383	0.002	0.216	0.129	0.92	0.708	0.308
Comp Sci	p < .001*	0.001	0.271	0.699	p < .001*	p < .001*	0.076	p < .001*	0.038	0.315	0.154	0.007	0.02	0.225
Decis Sci	0.343	0.038	0.286	0.01	0.028	p < .001*	0.004	0.038	0.043	0.004	0.129	0.036	0.386	0.332
Dentist	0.752	0.041	0.92	0.124	0.708	0.022	1.000	0.41	0.271	0.119	0.92	0.343	0.578	0.12
Earth Plan	0.091	0.718	1.000	0.001	0.063	p < .001*	p < .001*	0.337	0.791	0.003	p < .001*	0.001	0.92	0.156
Econ	0.006	1.000	0.475	p < .001	0.247	0.001	0.624	0.005	p < .001	p < .001	1.000	0.267	0.532	0.632
Energy	0.841	0.522	0.168	0.791	0.017	p < .001*	p < .001	p < .001*	0.027	0.502	0.502	0.708	0.162	0.216
Eng Tech	0.032	0.028	0.043	0.002	0.001	p < .001*	p < .001*	p < .001*	0.299	0.862	0.191	0.299	0.007	0.462
Environ	0.603	0.256	0.11	p < .001*	0.049	p < .001*	0.003	0.252	0.025	0.001	0.003	p < .001*	0.699	0.471
Health Prof	0.699	0.003	p < .001	0.003	1.000	0.108	0.68	0.841	1.000	0.118	0.639	0.125	0.617	0.075
Immun	0.08	0.006	p < .001	0.584	0.021	0.317	0.044	0.252	0.002	1.000	p < .001	0.48	0.185	0.284
Info Sci	0.343	0.371	0.322	p < .001	0.089	p < .001	0.121	0.001	0.708	p < .001	0.104	p < .001*	1.000	0.532
Math	0.093	0.343	0.325	0.198	0.005	p < .001*	0.292	0.301	0.047	0.275	0.74	0.357	0.584	0.035
Matl Sci	0.647	0.374	0.34	0.012	0.006	0.488	0.179	0.493	p < .001*	0.24	0.048	0.271	0.427	0.168
Med	0.313	0.045	0.002	0.597	0.014	0.005	0.004	0.012	1.000	0.791	0.427	0.256	0.036	0.151
Multi	1.000	0.002	0.044	p < .001*	0.034	0.002	0.027	0.031	0.699	p < .001*	p < .001*	p < .001*	0.022	1.000
Neuro	0.823	0.306	0.238	1.000	0.071	0.791	0.752	0.279	0.045	0.806	0.362	0.021	0.174	0.301
Nurs	0.45	0.026	0.126	0.671	0.13	0.046	1.000	0.584	0.357	0.566	0.777	0.224	0.17	0.48
Pharma	0.823	0.431	0.017	0.493	0.009	0.192	0.92	0.68	0.039	0.137	0.035	0.105	0.439	0.284
Physics	0.427	0.02	0.003	0.002	0.862	p < .001*	p < .001*	0.002	0.024	0.296	0.517	0.084	0.049	0.262
Psych	0.064	0.383	0.088	0.064	0.399	0.006	0.92	0.332	1.000	0.841	0.003	0.507	0.014	0.125
Soc Sci	p < .001	0.001	0.11	0.004	0.001	p < .001*	p < .001	0.512	0.023	0.043	0.036	0.037	0.015	0.41
Vet	0.538	0.68	0.191	0.888	0.427	0.097	0.777	0.173	0.632	1.000	0.399	0.841	0.252	0.458
Other	0.202	0.92	0.56	0.235	0.791	0.195	0.45	0.888	0.68	0.663	0.219	0.165	0.148	0.002

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

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.

	New study	New project	New ideas	Integration	New methods	Inputs	Calibration	Benchmark	Verification	Trends	Comparison	Summ/vis/tools	Teaching
Conditions/methodology	0.001*	p < .001*	0.001*	0.138	0.225	0.125	0.089	0.218	p < .001*	0.07	0.016	0.1	0.016
Processing/handling	0.001*	p < .001*	p < .001*	0.04	0.013	0.371	0.006	0.527	0.004*	0.004*	p < .001*	0.014	0.007
Topic relevance	0.197	p < .001*	0.001*	0.074	0.009	0.481	0.358	0.458	0.613	0.038	0.623	0.031	0.017
Coverage	0.022	0.167	0.331	p < .001*	0.093	0.012	0.517	0.139	0.666	p < .001*	p < .001*	p < .001*	0.041
Original purpose	0.385	0.002*	p < .001*	p < .001*	0.602	0.001*	0.151	0.438	0.021	0.02	0.627	0.71	0.006
Metadata/documentation	0.313	0.338	0.046	p < .001*	p < .001*	0.012	0.318	0.561	0.506	0.044	p < .001*	0.07	0.018
Format	0.798	0.58	0.351	0.229	0.103	0.323	0.263	0.714	0.692	0.215	0.279	0.22	0.002*
Size	0.949	0.42	0.238	0.58	0.159	0.082	0.486	0.738	0.706	0.113	0.281	0.69	0.013
Ease of access	0.622	0.037	0.002*	0.278	0.765	0.917	0.136	0.661	0.12	0.074	0.334	0.571	0.005
Licensing	0.606	0.188	0.202	0.100	0.078	0.146	0.088	0.087	0.916	0.297	0.369	0.004	0.001*
Source reputation	0.383	0.007	p < .001*	0.332	0.179	0.582	0.013	0.499	0.011	0.076	0.886	0.387	0.001*
Creator reputation	0.049	0.017	0.035	0.368	0.654	0.565	0.383	0.475	0.029	0.01	0.523	0.358	0.042
Knowing creator	0.377	0.427	0.117	0.201	0.051	0.226	0.666	0.835	0.329	0.015	0.634	0.772	0.717

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

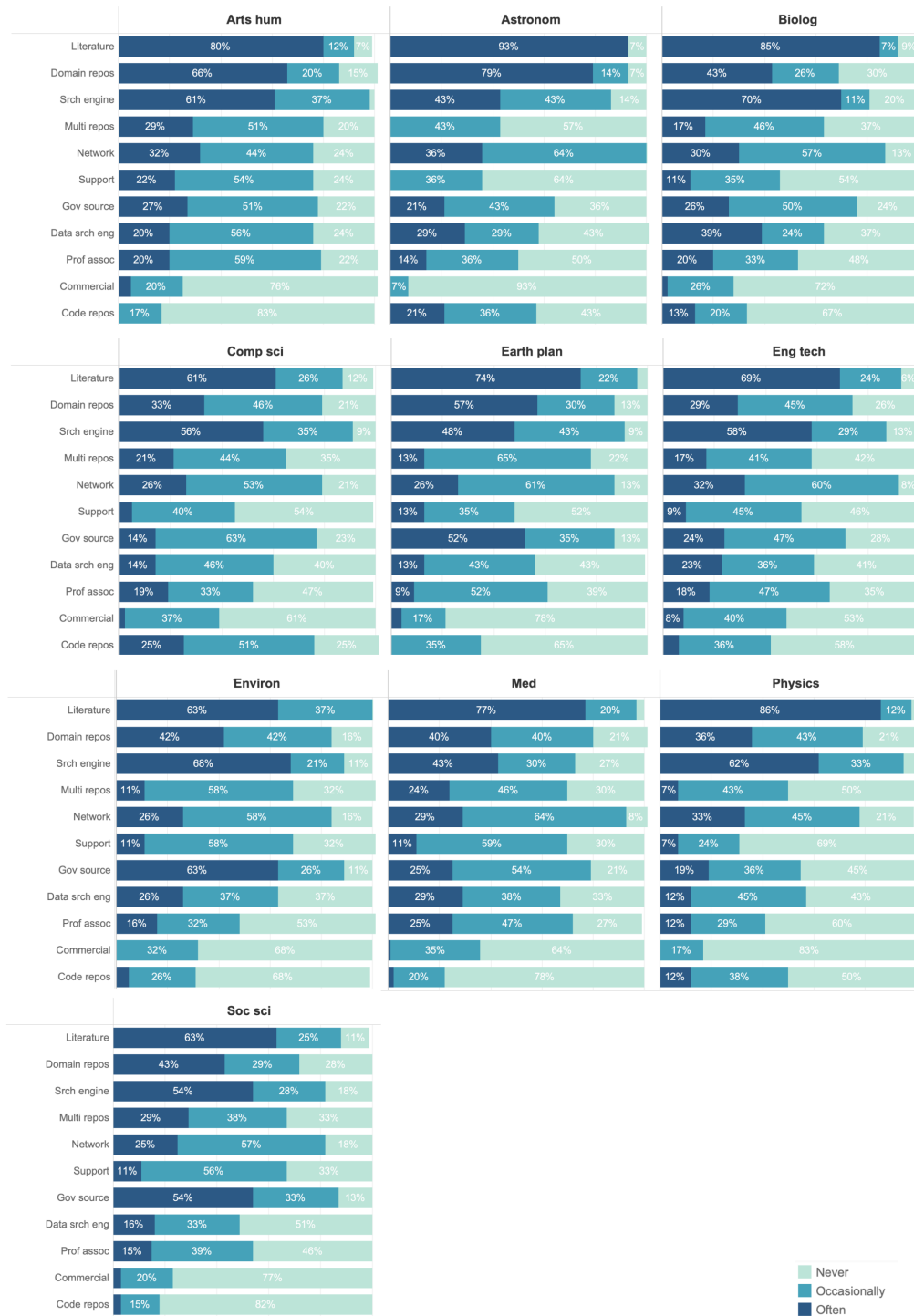


Figure C1. Sources used in the disciplinary subset for respondents selecting only one discipline. Percents are percent respondents. Arts & humanities ($n = 43$); astronomy ($n=14$); biological science ($n = 46$); computer science ($n = 57$); earth & planetary science ($n = 24$); engineering & technology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