

# The Field-Specificity of Open Data Practices

Velden, Theresa\* & Tcypina, Anastasiia\*\*

\**velden@dzhw.eu*, \*\**tcypina@dzhw.eu*

Department for Research System and Science Dynamics, DZHW, Germany

**Abstract:** Increasingly, researchers are expected to make their research data openly available. However, scientific fields differ in their research practices and norms for publicly sharing research data. We provide quantitative evidence of differences in data practices and the public sharing of research data at a granularity of field-specificity that is rarely reported in open data surveys. Based on a survey of 8,822 researchers at German Universities, we find considerable variation, within and between disciplines, of data practices and rates of open data sharing. For experimentally oriented subject areas we further observe a relationship between data self-sufficiency and public data sharing which likely reflects a link between data sharing and the epistemic specificity of data. Our findings underline that in order to quantitatively assess and evaluate rates of public data sharing, a better understanding of the embedding of public data sharing into field-specific research practices is needed.

## 1. Introduction

The public sharing of research data is not a new phenomenon. For example, star catalogues have been publicly shared by astronomers since ancient times, enabling scientific discoveries, such as the discovery of the precession of the equinox by Hipparch in the 2<sup>nd</sup> century B.C. (Goldstein & Bowen 1991). With digitization and the advance of the Internet, the technical capabilities to make research data widely available have vastly increased. This has moved the goal of making the public sharing of research data a new standard for publicly funded research to the forefront of science policy agendas (Kroes, 2012; Holdren 2013; OECD 2015; CNRS 2016; Wellcome 2017, Commission Recommendation (EU) 2018, G6 2021).

To inform research funders and other stakeholders about the state of public data sharing, a growing number of survey studies has sought to quantify the extent to which researchers embrace the idea of public sharing and follow suit in their practices. However, the coverage in terms of disciplines and countries as well as sample sizes differ widely between these survey studies, as do the ways in which the phenomenon of sharing research data is operationalized. Consequently, the empirical evidence about different forms of sharing, across disciplinary or (inter)national contexts, is still sketchy. Some general trends about the state of data sharing, however, seem fairly well supported: that there exists a gap between support for the idea of public data sharing, and actual practice (e.g. Ambrasat & Heger 2020, Zhu 2020, Tenopir 2020, Nicholas et al. 2020, Fecher 2017), and that a number of hindrances exist that either delay or entirely prevent the public sharing of data, such as the effort involved in preparing data for sharing, the sensitivity of data, or concerns about a lack of control over the (re)use of the data by others (see e.g. Campbell et al. 2002, Tenopir 2011, 2015, 2020, Zhu 2020, Nicholas et al. 2020, Alexandre-Benavent 2020).

The degree to which sharing practices differ between disciplines and research fields, however, have received only scant attention - even though understanding field-specific differences may be crucial for the responsible use of research assessment tools to avoid unintended detrimental effects that can occur when open science policies follow a one-size-fits-all approach to encourage public data sharing (see Gläser 2019, p. 436). Whereas most surveys apply some sort

of field classification, most multidisciplinary surveys use rather coarse classifications (e.g. Goodey 2022, Stuart et al. 2018, Fecher et al. 2017) or use this information merely to describe the overall composition of their sample rather than a systematic analysis of field differences (e.g. Aleixandre-Benavent et al. 2020; Choi & Lee 2020). A rare exception are the studies by Thursby et al. (2018) on pre-publication disclosure of results, and Kim & Stanton (2012) on public sharing of research data upon publication of results. Their statistical analyses confirm the existence of field differences in rates of sharing, and both studies suggest that the reported sharing, correlates, among other factors, with the strength of sharing-supporting norms in a field. However, neither study examines the underlying reasons why sharing supporting norms emerge more strongly in some fields than in others.

One hypothesis is that field-specific norms of sharing are the result of field specific epistemic practices (Steinhardt et al. 2022, Thursby et al. 2018, Velden 2013), i.e. the specific way of how knowledge is produced in a field: the empirical objects that are studied, the methods used, and the theories developed, which in turn influence data practices – the type of data and how they are produced and used (Borgman 2012). To further explore the field-specificity of public data sharing and how it is linked to the underlying research practices in a field, we examine the data provided by the *DZHW Science Survey* (WiBef), a large survey of over 8,000 researchers at German universities. It offers data on data practices and data sharing at a much finer resolution of fields than most multidisciplinary surveys provide, bringing us closer to the level of analysis required for investigating links between epistemic practices and data sharing.

## 2. Data and Methods

### 2.1. The data source

The data we use is a small subset of data from WiBef, a representative tri-annual survey-based trend study about the German science system, covering a variety of topics. The survey was conducted by the German Center for Higher Education and Science Studies (DZHW) between November 2019 and February 2020. The questionnaire was sent to a total 60,002 researcher from German universities, 52,769 of the emails were delivered. The 8,822 valid responses received represented a 16.72% response rate.

### 2.2. Survey instruments

In this study we focus on the information about *data practices* collected through an instrument included in the WiBef survey that asked a set of questions about the role that research data plays in a respondent's research practice (see Table 1). Specifically, we analysed whether researchers said they produce data themselves (alone or in a team), use data from third parties in their research, or do not work with data at all. Further, a field-specific *rate of sharing research data publicly* was determined by the proportion of respondents in a discipline or subject area who affirm that they make research data that they collect or produce 'publicly available to other scientists, regularly or occasionally.'

Information about the *field affiliation* of researchers was collected using a two-level classification covering humanities & social sciences, engineering, life sciences, and physical

Table 1. Question about data practices.

|   |
|---|
| fp2   |
| <p><b>What role do research data play in your research practice?</b></p> <p><i>Please select all that apply.</i></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> I analyze research data empirically.</li> <li><input type="checkbox"/> I collect/produce research data myself (alone or in a team).</li> <li><input type="checkbox"/> I use research data from third parties.</li> <li><input type="checkbox"/> I make data that I collect/produce publicly available to other scientists, regularly or occasionally.</li> <li><input type="checkbox"/> I prepare research data for empirical analysis by third parties.</li> <li><input type="checkbox"/> I don't work with research data.</li> </ul> |

sciences. The questionnaire offered respondents 10 disciplines to choose from, and within each discipline, one to eight different subject areas, such that in total 39 subject areas are distinguished. This relatively fine-grained classification is largely based on the field classification used in 2016-2019 by the German Research Foundation (DFG), with some small modifications. The subject areas correspond to review committees in the DFG grant review system (so called 'Fachkollegien').

### 2.3. Sample design and weighting scheme

Different from surveys that work with convenience or snow-ball samples, the stratified, random sample of the WiBef can be considered representative of researchers at German Universities. The websites of 132 higher education institutions were systematically searched to collect the contact details of researchers employed by German Universities. From this sampling frame of 167,382 researchers, a disproportionally stratified sample was drawn, such that professors and postdocs were selected with a higher probability than researchers without doctoral degrees. This design was chosen to guarantee a satisfactory sample size for in-depth evaluations within status groups of interest (see the WiBef-survey method report by Ambrasat et al. 2020, in German).

The realized sample covers the spectrum of disciplines represented at German universities well and deviates only slightly from the official statistics on the distribution of scientific personnel by discipline (Ambrasat et al. 2020). An overview over the achieved sample sizes of the different disciplines and subject areas is given in Table 2.

Due to the stratification of the sample and the variance of response rates by status group, we apply a weighting scheme that adjusts observed proportions by respective weights in order to increase the representativeness of results for the targeted population (Ambrasat et al. 2020). The resulting sample sizes by status group are presented in Table 3. All results we report in the results section were calculated applying the weighting scheme.

Table 2. Sample sizes by discipline and by subject area (unweighted data).

| Discipline                           | Sample Size | Subject area  | Sample Size  |
|--------------------------------------|-------------|---|--------------|
| Social and Behavioural Sciences      | 1,968       | Educational Sciences  | 344          |
|                                      |             | Jurisprudence   | 139          |
|                                      |             | Psychology  | 473          |
|                                      |             | Social Sciences   | 621          |
|                                      |             | Economics   | 385          |
|                                      |             | Other   | 6            |
| Humanities                           | 1,443       | Fine Arts, Music, Theatre and Media Studies                               | 197          |
|                                      |             | Linguistics   | 304          |
|                                      |             | Literary Studies  | 276          |
|                                      |             | Philosophy  | 113          |
|                                      |             | History   | 294          |
|                                      |             | Religious Studies, Jewish Studies   | 24           |
|                                      |             | Social and Cultural Anthropology  | 73           |
|                                      |             | Theology  | 115          |
| Engineering Sciences                 | 1,383       | Engineering   | 1,381        |
|                                      |             | Other   | 2            |
| Medicine                             | 507         | Human Medicine  | 470          |
|                                      |             | Veterinary Medicine   | 35           |
|                                      |             | Other   | 2            |
| Biology                              | 1,025       | Zoology   | 117          |
|                                      |             | Basic Biological and Medical Research                                     | 408          |
|                                      |             | Microbiology, Virology and Immunology                                     | 187          |
|                                      |             | Biochemistry  | 146          |
|                                      |             | Plant Sciences  | 159          |
|                                      |             | Other   | 8            |
| Agricultural, Forestry, Horticulture | 127         | Agricultural, Forestry, Horticulture                                      | 127          |
| Physics                              | 623         | Astrophysics and Astronomy  | 65           |
|                                      |             | Condensed Matter Physics  | 203          |
|                                      |             | Optics, Quantum Optics and Physics of Atoms, Molecules and Plasma         | 148          |
|                                      |             | Particles, Nuclei and Fields  | 103          |
|                                      |             | Statistical Physics, Soft Matter, Biological Physics, Non-linear Dynamics | 90           |
|                                      |             | Other   | 14           |
| Mathematics                          | 397         | Mathematics   | 397          |
| Geology                              | 474         | Geology   | 474          |
| Chemistry                            | 576         | Chemical Solid State and Surface Research                                 | 62           |
|                                      |             | Molecular Chemistry   | 126          |
|                                      |             | Physical and Theoretical Chemistry  | 142          |
|                                      |             | Analytical Chemistry, Method Development (Chemistry)                      | 67           |
|                                      |             | Biochemistry  | 107          |
|                                      |             | Polymer Science   | 61           |
|                                      |             | Other   | 11           |
| No assignment                        |             |   | 299          |
| Total                                |             |   | <b>8,822</b> |

Table 3: Non-weighted and weighted sample sizes by status group.

| Status group                            | Professor | Postdoc | Predoc | Total    |
|---|-----------|---------|--------|----------|
| Sampling frame                          | 27,900    | 46,232  | 93,250 | 167,382  |
| Disproportionately stratified sample    | 13,927    | 22,950  | 23,125 | 60,002   |
| Drawing probability                     | 0.499     | 0.496   | 0.248  | 0.358    |
| Design weight                           | 2.003     | 2.014   | 4.032  |          |
| Net sample                              | 1,721     | 4,216   | 2,885  | 8,822    |
| Participation probability               | 0.124     | 0.184   | 0.125  | 0.147    |
| Non response weight                     | 8.092     | 5.444   | 8.016  |          |
| Combined weight                         | 16.212    | 10.966  | 32.322 | 18.973   |
| Combined probability weight, normalized | 0.854     | 0.578   | 1.704  | Mean = 1 |
| Sample after weighting                  | 1,470     | 2,437   | 4,915  | 8,822    |

Based on Table 7 from WiBef-survey method report (Ambrasat et al. 2020, in German)

### 3. Results

#### 3.1. Variation in data practices

Figure 1 compares data practices between Humanities & Social Sciences, Engineering, Life Sciences, and Natural Sciences. It shows that in each broad disciplinary grouping a large majority, over 75% of researchers, state that in their research practice they work with data. This orientation towards working with data is highest in the Life Sciences where only 2% of researchers state that they do not work with data. Next are Engineering and the Natural Sciences, with 11%, respectively 14% of researchers saying that they do not work with data. The highest proportion of researchers who do not work with research data is reported for Humanities and Social Science (approx. 20%).

Figure 1: Data Practices by High-Level Disciplinary Grouping

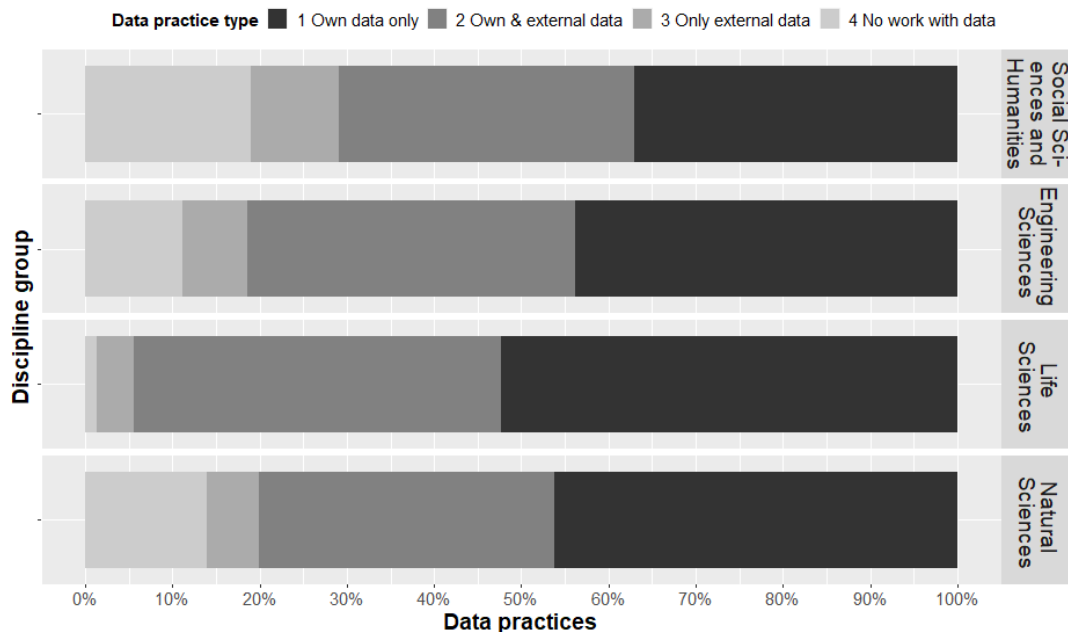
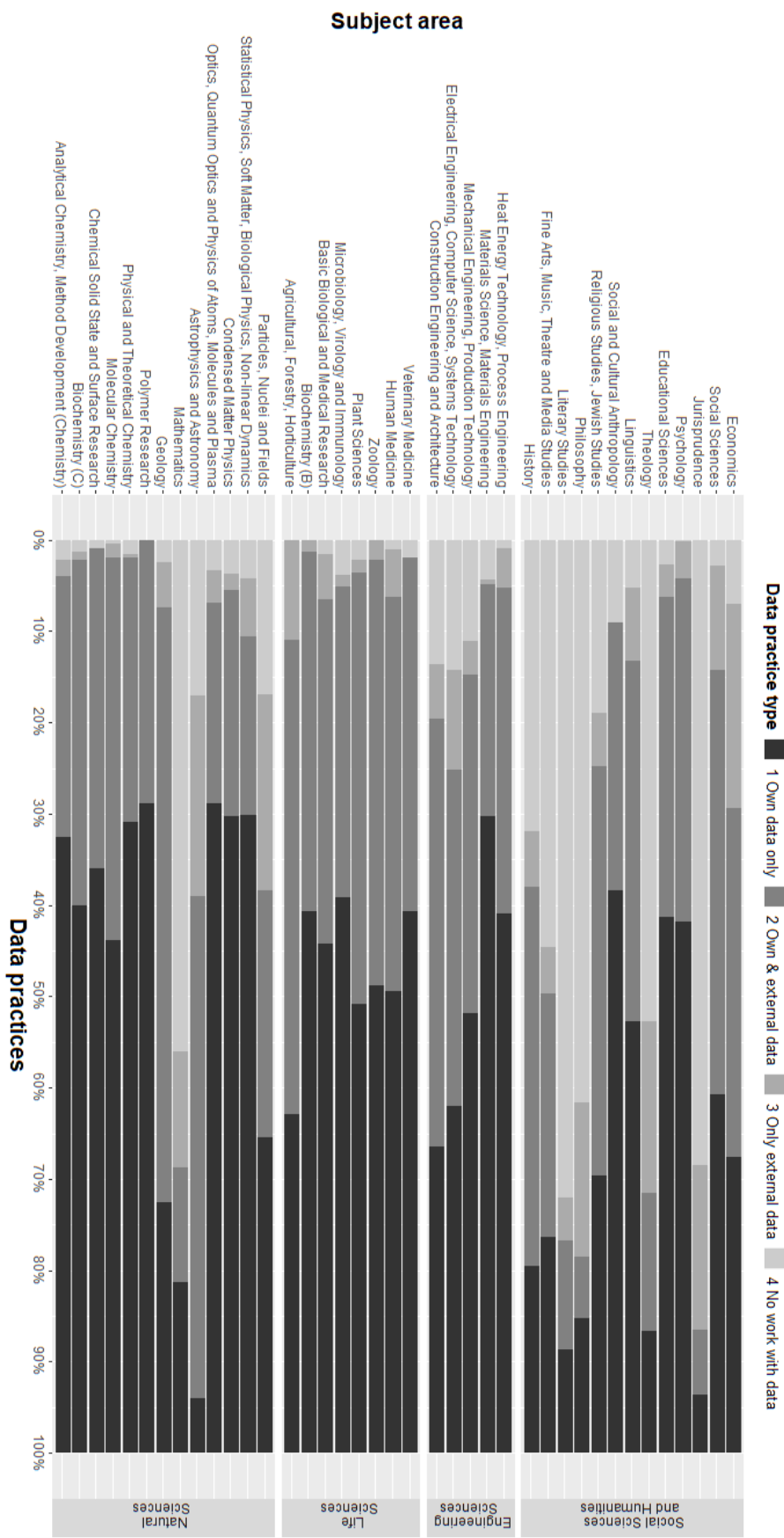


Figure 2: Data Practices at Subject Area Level



Moving to a more fine-grained resolution, at the level of subject areas (Figure 2), we find stark differences in data practices, even within the same discipline.

We can now identify distinctively ‘**data distant**’ fields, ‘where most researchers responded that they neither create data nor use data of others in their research practice’<sup>1</sup>. Examples are Philosophy, Literary Studies, Theology, Mathematics, and Jurisprudence, where most researchers said that they do not work with data. At the other end of the spectrum, we have strongly empirically oriented subject areas, such as Agriculture, Polymer Science, Zoology, Biochemistry (Biology), and Psychology, where almost everyone (>99%) reports to be working with research data.

Among empirically oriented subject areas where a large majority of researchers work with data, we find differences regarding whether data from third parties are used, or whether researchers rely exclusively on data they produce themselves.

If we label researchers who work exclusively with their own data (i.e. produce data and do not work with data from third parties) as ‘**data self-sufficient**’, we find in most disciplines subject areas where a majority of researchers is data self-sufficient: in Humanities (Social And Cultural Anthropology), Social and Behavioural Sciences (Psychology, Educational Sciences), Engineering (e.g. Materials Science & Materials Engineering), Medicine (Veterinary Medicine), Biology (e.g. Biochemistry, Microbiology, Virology & Immunology), Chemistry (All Subject Areas), and Physics (e.g. Condensed Matter Physics).

In other subject areas the majority of researchers are ‘**data combiners**’<sup>2</sup>, i.e. researchers who produce data themselves and use data provided by third parties. This is the case in Geosciences, Agricultural, Forestry & Horticulture, and Astronomy & Astrophysics.

Further, we note that some of the empirically oriented subject areas have a comparatively large fraction of researchers (15-25%) who we may refer to as ‘**data consumers**’ because they do not produce data themselves, but instead exclusively rely on data provided by third parties (Astronomy & Astrophysics, Particles, Nuclei & Fields, and Economics).

### *3.2 Variation in public data sharing*

To probe the connection between research practices and the public sharing of research data, we select four disciplines: Biology, Chemistry, Physics, and Social & Behavioural Sciences<sup>3</sup> for closer examination where most researchers produce data and hence face the question of whether to publicly share this data or not.

---

<sup>1</sup> Note that this categorization is dependent on how respondents interpreted the term ‘data’. The survey instrument did not make any suggestions of how to define the term. Hence certain types of empirical evidence, such as archival materials used in historical research, may not have been perceived by respondents as ‘data’.

<sup>2</sup> We have to interpret the label of ‘data combiners’ with caution, in so far as these are researchers who have specified that they produce data themselves (or in a team) and that in their research they use data that has been provided by a third party. This does not necessarily imply that they combine own and third-party data within the same research process or project.

<sup>3</sup> In the analysis that follows, we omit Jurisprudence (formally grouped into the social sciences by the DFG classification), due to its very low rate of data production. The reported public sharing rate among the data producers in this subject area is 0%.

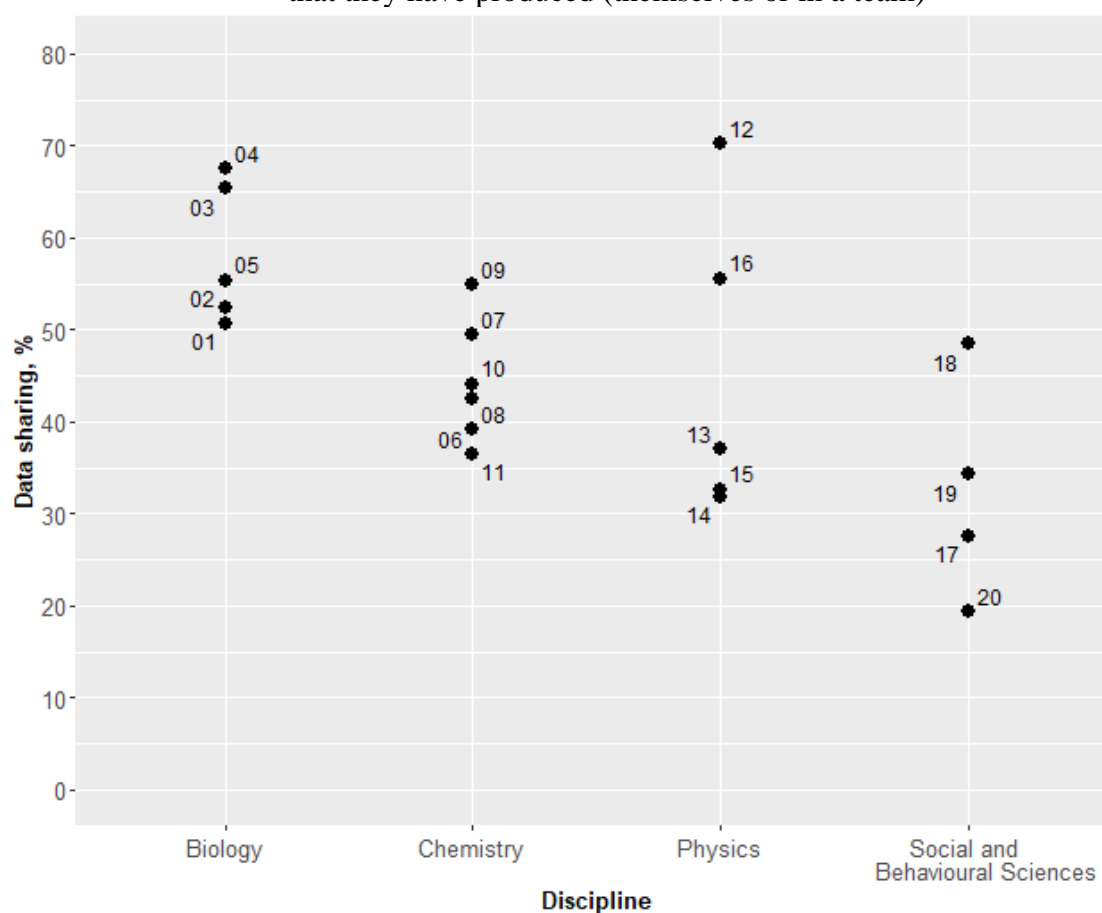
Table 3 and Figure 3 show the rates of public data sharing for the subject areas in the selected disciplines. They reveal considerable variation between subject areas across and within disciplines. The rates of public data sharing range from 70% in Astronomy & Astrophysics at the high end to 19% in Economics at the low end. The within discipline variation between subject areas is highest in Physics and Social Sciences, and lower in Chemistry and Biology.

Table 3. Rates of public sharing by subject area.

| Discipline / Subject area  | Rate of public sharing |
|--|------------------------|
| <b>Biology</b>   |                        |
| 01 Biochemistry (B)  | 51%                    |
| 02 Basic Biological and Medical Research                                     | 52%                    |
| 03 Microbiology, Virology and Immunology                                     | 65%                    |
| 04 Plant Sciences  | 67%                    |
| 05 Zoology   | 55%                    |
| <b>Chemistry</b>   |                        |
| 06 Analytical Chemistry, Method Development (Chemistry)                      | 39%                    |
| 07 Biochemistry (C)  | 49%                    |
| 08 Chemical Solid State and Surface Research                                 | 42%                    |
| 09 Molecular Chemistry   | 55%                    |
| 10 Physical and Theoretical Chemistry  | 44%                    |
| 11 Polymer Research  | 37%                    |
| <b>Physics</b>   |                        |
| 12 Astrophysics and Astronomy  | 70%                    |
| 13 Optics, Quantum Optics and Physics of Atoms, Molecules and Plasma         | 37%                    |
| 14 Condensed Matter Physics  | 32%                    |
| 15 Statistical Physics, Soft Matter, Biological Physics, Non-linear Dynamics | 33%                    |
| 16 Particles, Nuclei and Fields  | 56%                    |
| <b>Social and Behavioural Sciences</b>                                       |                        |
| 17 Educational Science   | 28%                    |
| 18 Psychology  | 48%                    |
| 19 Social Sciences   | 34%                    |
| 20 Economics   | 19%                    |



Figure 3: Proportion of researchers who occasionally or regularly share research data publicly that they have produced (themselves or in a team)



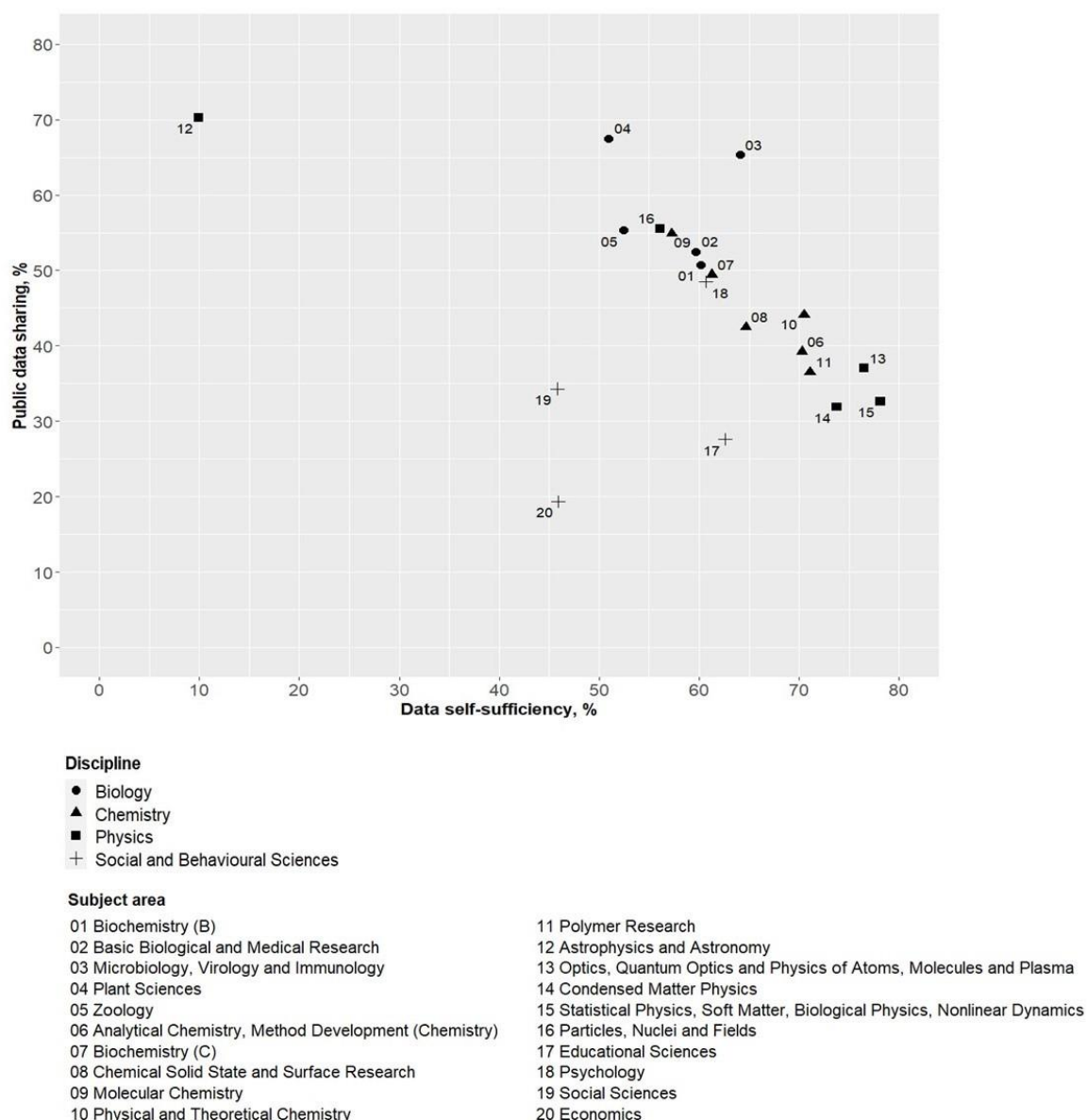
#### Subject area

- 01 Biochemistry (B)
- 02 Basic Biological and Medical Research
- 03 Microbiology, Virology and Immunology
- 04 Plant Sciences
- 05 Zoology
- 06 Analytical Chemistry, Method Development (Chemistry)
- 07 Biochemistry (C)
- 08 Chemical Solid State and Surface Research
- 09 Molecular Chemistry
- 10 Physical and Theoretical Chemistry
- 11 Polymer Research
- 12 Astrophysics and Astronomy
- 13 Optics, Quantum Optics and Physics of Atoms, Molecules and Plasma
- 14 Condensed Matter Physics
- 15 Statistical Physics, Soft Matter, Biological Physics, Nonlinear Dynamics
- 16 Particles, Nuclei and Fields
- 17 Educational Sciences
- 18 Psychology
- 19 Social Sciences
- 20 Economics

However, the disciplinary affiliation of subject areas tells only part of the story of their epistemic practices. To learn more about the dependency of rates of public sharing on epistemic practices, we juxtapose in Figure 4 the public data sharing rate among data producing scientists in a subject area with the rate of data self-sufficiency among them. Data self-sufficiency may serve – in first approximation – as indicative of an experimental research orientation. Research that is experimentally oriented generates data under controlled conditions to address a specific research question. Such data are epistemically highly specific, produced by the researchers themselves in custom made experimental set-ups, rather than obtained from a third party (Borgman 2012).

In Figure 4 we can identify a large group of subject areas (**Group I**) that displays a linear association between rate of data self-sufficiency and rate of public sharing. Then there is a second group of six subject areas that are outliers in so far as that they deviate in one or the other way from group 1 (**Group II: subject areas 03, 04, 12, 17, 19, 20**).

Figure 4: Rate of data self-sufficiency versus rate of public sharing among data producers



**Observations regarding Group I (linear trend):** All subject areas in this group have a pronounced rate of data self-sufficiency ( $> 50\%$ ), suggesting a high relevancy of experimental methods. We observe a linear trend for this group: the more data self-sufficient, the lower the rate of public sharing of research data. Three subject areas in Physics occupy one end of this spectrum, with high data self-sufficiency and public data sharing below 40%. At the other end of the spectrum, we find subject areas with higher rates of public sharing and lower rates of data self-sufficiency. Among them are subject areas such as Molecular Chemistry, and Basic Biological & Medical Research, which produce (and use) data such as genetic sequences and molecular structures, which are stored and made available in global databases. Whereas experimental data from custom made experiments often is scientifically exploited after its original analysis and considered of limited re-use value (see e.g. Akmon 2014), molecular structures and genetic sequences are forms of data that are highly standardized and have great re-use value for scientific research (Borgman 2012). The lower rate of self-sufficiency along with the higher rate of public sharing could be indicative of a mutual interdependency of researchers in these subject areas regarding data they produce and use.

**Observations regarding Group II ('outliers'):** This group of subject areas shows a disparate pattern. We can distinguish two subgroups: three Natural Science subject areas with an extremely high rate of public data sharing ( $\geq 65\%$ ). The second subgroup consists of three subject areas in the Behavioural & Social sciences that expose a low rate of public data sharing ( $< 35\%$ ).

Among the three natural science subject areas, Astronomy & Astrophysics is set apart by an extremely low rate of data self-sufficiency. It is a distinctively observational field, unified by a common type of empirical data: electromagnetic radiation reaching Earth from the sky. Combining digital observational data from various sources is a frequent practice (Genova 2018, Hoeppe 2014), which would align with the low rate of data self-sufficiency reported here. In Plant Sciences a great diversity of sources of empirical data exists, such as genetic analyses, lab based or field-based experiments, and observational fieldwork. The data self-sufficiency and the rate of public data sharing reported for this subject area likely represent an aggregate that conflates a wide range of epistemic practices and types of data shared. Finally, Microbiology, Virology & Immunology combines a relatively high rate of data self-sufficiency of  $> 60\%$  with a high rate of public data sharing. This suggests a proliferation of research data, possibly through high-throughput genome sequencing (Connor et al. 2016, Loman & Pallen 2015). A high rate of data self-sufficiency combined with high rate of public sharing could be indicative of a division of work in this subject area in that a section of researchers is primarily engaged in the production of research data that is then publicly shared and re-used by others.

The three Social & Behavioural Science subject areas in group II, by contrast, are characterized by a low rate of public sharing, below 35%. The subject area of Educational Sciences is an interesting case because it is characterized by a higher level of data self-sufficiency, above 60%, than the other two subject areas in this subgroup, Social Sciences and Economics. In these latter two subject areas most data producers also use data provided by third parties. Major sources of data used in these fields are social and economic statistical data provided by state authorities or large-scale survey projects (Hessel et al. 2019; Einav & Levin 2014). In the subject area of Social Sciences, the sharing of survey data is rather institutionalized, with institutions collecting and making survey data available under a range of access conditions, including public sharing. This applies mostly to quantitative data however, as the public sharing of qualitative data through publicly accessible repositories is only in its infancy. Notably, researchers in Economics who produce data, report a lower rate of public sharing compared to Social Sciences,

which may have to do with an increasing role of private sector data that is acquired for research through data sharing agreements that limit distribution (Einav & Levin 2014).

#### **4. Discussion**

Our analysis shows strong variation of data practices and rates of public sharing at the subject area level, across and within disciplines. The observed association between public data sharing rates and data-self-sufficiency suggests the relevance of epistemic practices, such as experimental orientation, the specificity of data produced, and the need to combine data from diverse sources, for explaining rates of public data sharing.

This study is only a starting point for quantifying differences in data sharing between research fields. Research is highly diverse in its epistemic practices, and what constitutes data, how these data can be made available and (re)used, differs widely (Steinhardt et al. 2022, Leonelli & Tempini 2020, Kurata et al. 2017, Borgman 2012). Underlying this diversity is a variety of different ‘data economies’ – social systems of the production and use of research data that get supported by different forms of exchange and accompanying norms for data sharing. Therefore, to evaluate and compare rates of public sharing we need to consider the underlying epistemic practices, which are insufficiently operationalized by the discipline classifications systems used in most open data surveys (Gläser 2018). Even a field classification at the level of granularity used here, is bound to conflate different methods of data production and different types of data that come along with rather different opportunities and limitations for re-use.

#### **5. Conclusions**

A field-comparative analysis of data from a survey of researchers at German Universities in 2019/2020 shows variation of data practices and rates of public sharing of research data between subject areas across and within disciplines. Among empirically oriented sciences, we find a large group of subject areas with an association between data self-sufficiency and rates of public data sharing. The variation of sharing rates and its interdependency with epistemic practices suggests that to collect meaningful data for comparative and evaluative assessments of the state of open data, we need to go beyond mere field classifications and develop more sophisticated instruments to capture relevant dimensions of epistemic practices and the data economies that support them.

#### **Open science practices**

The survey instrument of the DZHW WiBef survey can be downloaded from: <https://doi.org/10.21249/DZHW:scs2019:2.0.0>. This way the way the survey data has been generated is more transparent. Further a method report (in German) that gives detailed insight in how the survey was designed and conducted is available (Ambrasat et al. 2020). The actual survey data itself is also available for further research, via the FDZ-DZHW data repository. It and can be accessed upon application for analysis via remote desktop free of charge (see Ambrasat, et al. 2022). The limitation of being restricted to analysis via a remote desktop is deemed necessary by the creators of the survey to protect the anonymity of the survey participants who shared extensive and detailed information in the survey.

## Acknowledgments

We are indebted to the 2019/20 WiBef team, Jens Ambrasat and Christoph Heger, for sharing the data from the core questionnaire that provided the basis for the field-comparative analysis in this study. We thank Nathalie Schwichtenberg for thoughtful feedback in various stages of instrument design and data analysis. We thank Laura Mages, who supported us as a student assistant in the initial, exploratory analysis of the survey data.

## Author contributions

Theresa Velden conceptualized the paper, participated in the formal data analysis, and in the writing of the paper. Anastasiia Tcypina has participated in the formal data analysis, prepared the visualization of results, and participated in the writing of the paper.

## Competing interests

No competing interests.

## References

- Akmon, D. R. (2014). *The role of conceptions of value in data practices: A multi-case study of three small teams of ecological scientists* (Doctoral dissertation, University of Michigan).
- Aleixandre-Benavent, R., Vidal-Infer, A., Alonso-Arroyo, A., Peset, F., & Ferrer Sapena, A. (2020). Research data sharing in Spain: Exploring determinants, practices, and perceptions. *Data*, 5(2), 29.
- Ambrasat, J., Heger, C., & Rucker, A. (2020) Wissenschaftsbefragung 2019/20 - Methoden und Fragebogen. DZHW Methodenbericht, Hannover: DZHW. Online at: [https://www.wb.dzhw.eu/downloads/WiBef\\_Methodenbericht2019-20.pdf](https://www.wb.dzhw.eu/downloads/WiBef_Methodenbericht2019-20.pdf)
- Ambrasat, J., & Heger, C. (2020). Barometer für die Wissenschaft. Ergebnisse der Wissenschaftsbefragung 2019/20. Berlin: DZHW.
- Ambrasat, J., Heger, C., Fabian, G. & Rucker, A. (2022). DZHW-Wissenschaftsbefragung 2019. Datenerhebung: 2019/2020. Version: 2.0.0. Datenpaketzugangsweg: Remote-Desktop-SUF. Hannover: FDZ-DZHW. Datenkuratierung: Weber, A., Daniel, A. & Schmidtchen, H. <https://doi.org/10.21249/DZHW:scs2019:2.0.0>
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078.
- Campbell, E. G., Clarridge, B. R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N. A., & Blumenthal, D. (2002). Data withholding in academic genetics: evidence from a national survey. *JAMA*, 287(4), 473-480.
- Choi, M. S., & Lee, S. (2020). Research data management status of science and technology research Institutes in Korea. *Data Science Journal*, 19(1).
- CNRS - Scientific and Technical Information Department. 2016. *White Paper — Open Science in a Digital Republic*. Laboratoire d'idées. Marseille: OpenEdition Press.

Commission Recommendation (EU) 2018/790 of 25 April 2018 on Access to and Preservation of Scientific Information, C/2018/2375'. <https://eur-lex.europa.eu/eli/reco/2018/790/oj> Accessed 2 December 2021.

Connor TR, Loman NJ, Thompson S, Smith A, Southgate J, Poplawski R, Bull MJ, Richardson E, Ismail M, Thompson SE, Kitchen C, Guest M, Bakke M, Sheppard SK, & Pallen MJ. (2016). CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microbial Genomics*, 2(9).

Einav, L., and J. Levin. 2014. Economics in the age of big data. *Science* 346 (6210): 1243089.

Fecher, B., Friesike, S., Hebing, M. *et al.* A reputation economy: how individual reward considerations trump systemic arguments for open access to data. *Palgrave Commun* 3, 17051 (2017). <https://doi.org/10.1057/palcomms.2017.51>

G6 statement on Open Science Brussels, December 2021. Online at: [https://os.helmholtz.de/fileadmin/user\\_upload/os.helmholtz.de/Dokumente/G6\\_statement\\_on\\_Open\\_Science.pdf](https://os.helmholtz.de/fileadmin/user_upload/os.helmholtz.de/Dokumente/G6_statement_on_Open_Science.pdf)

Genova, F. (2018). Data as a research infrastructure CDS, the Virtual Observatory, astronomy, and beyond. In *EPJ Web of Conferences* (Vol. 186, p. 01001). EDP Sciences.

Gläser, J. (2019). "Chapter 21: How can governance change research content? Linking science policy studies to the sociology of science". In *Handbook on Science and Public Policy*. Cheltenham, UK: Edward Elgar Publishing. Retrieved Jul 7, 2023, from <https://doi.org/10.4337/9781784715946.00033>

Gläser, J. (2018, September). Accounting for field-specific research practices in surveys. In *STI 2018 Conference Proceedings* (pp. 1364-1370). Centre for Science and Technology Studies (CWTS).

Goldstein, B. R., & Bowen, A. C. (1991). The Introduction of Dated Observations and Precise Measurement in Greek Astronomy. *Archive for History of Exact Sciences*, 43(2), 93–132.

Goodey, G., Hahnel, M., Zhou, Y., Jiang, L., Chandramouliswaran, I., Hafez, A , Taunton Paine, Susan Gregurick, Samuel Simango, Juan Miguel Palma Peña, Holly Murray, Matt Cannon, Rebecca Grant, Kate McKellar, Laura Day (2022). The state of open data 2022. Report by Digital Science, Springer Nature and Figshare. Accessed online (on April 13,2023) at: <https://apo.org.au/node/319974>

Hessels, L. K., Franssen, T., Scholten, W., & De Rijcke, S. (2019). Variation in valuation: How research groups accumulate credibility in four epistemic cultures. *Minerva*, 57, 127-149.

Hoeppe, G. (2014). Working data together: The accountability and reflexivity of digital astronomical practice. *Social Studies of Science*, 44(2), 243-270.

Holdren, John P. 2013. 'Memorandum for the Heads of Executive Departments and Agencies: Increasing Access to the Results of Federally Funded Scientific Research'. Edited by Executive Office of the President and Office of Science and Technology Policy. Washington, DC.

Kim, Y., & Stanton, J. (2012). Institutional and Individual Influences on Scientists' Data Sharing Practices. *The Journal of Computational Science Education*, 3(1), 47–56. <https://doi.org/10.22369/issn.2153-4136/3/1/6>

Kroes, Neelie. 2012. 'Scientific Data: Open Access to Research Results Will Boost Europe's Innovation Capacity'. Press Release. European Commission. 2012. [https://ec.europa.eu/commission/presscorner/detail/en/IP\\_12\\_790](https://ec.europa.eu/commission/presscorner/detail/en/IP_12_790).

Kurata, K., Matsubayashi, M., & Mine, S. (2017). Identifying the complex position of research data and data sharing among researchers in natural science. *Sage Open*, 7(3), 2158244017717301.

Leonelli, S., & Tempini, N. (2020). *Data journeys in the sciences* (p. 412). Springer Nature.

Loman, N. J., & Pallen, M. J. (2015). Twenty years of bacterial genome sequencing. *Nature Reviews Microbiology*, 13(12), 787-794.

Nicholas, D., Jamali, H. R., Herman, E., Watkinson, A., Abrizah, A., Rodríguez-Bravo, B., ... & Polezhaeva, T. (2020). A global questionnaire survey of the scholarly communication attitudes and behaviours of early career researchers. *Learned Publishing*, 33(3), 198-211.

OECD. 2015. 'Making Open Science a Reality'. 25. OECD Science, Technology and Industry Policy Papers. Paris: OECD Publishing. <https://doi.org/10.1787/5jrs2f963zs1-en>.

Steinhardt, I., Bauer, M., Wünsche, H., & Schimmler, S. (2022). The connection of open science practices and the methodological approach of researchers. *Quality & Quantity*. <https://doi.org/10.1007/s11135-022-01524-4>

Stuart, D., Baynes, G., Hrynaszkiewicz, I., Allin, K., Penny, D., Lucraft, M., & Astell, M. (2018, March 21). Whitepaper: Practical challenges for researchers in data sharing (Version 1) [figshare]. <https://doi.org/10.6084/m9.figshare.5975011.v1>

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. 2011. 'Data Sharing by Scientists: Practices and Perceptions'. *PLOS ONE* 6 (6): e21101. <https://doi.org/10.1371/journal.pone.0021101>.

Tenopir, Carol, Elizabeth D. Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, Danielle Pollock, and Kristina Dorsett. 2015. 'Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide'. *PLOS ONE* 10 (8): e0134826. <https://doi.org/10.1371/journal.pone.0134826>.

Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R., & Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLOS ONE*, 15(3), e0229003. <https://doi.org/10.1371/journal.pone.0229003>

.

Thursby, J. G., Haeussler, C., Thursby, M. C., & Jiang, L. (2018). Prepublication disclosure of scientific results: Norms, competition, and commercial orientation. *Science advances*, 4(5), eaar2133.

Velden, T. (2013, February). Explaining field differences in openness and sharing in scientific communities. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW)* (pp. 445-458).

Wellcome. 'Data, Software and Materials Management and Sharing Policy'. Wellcome. Accessed 27 February 2020. <https://wellcome.org/grant-funding/guidance/data-software-materials-management-and-sharing-policy>.

Zhu, Y. (2020) 'Open-access policy and data-sharing practice in UK academia', *Journal of Information Science*, 46(1), pp. 41–52.

Zinner, D. E., Pham-Kanter, G., & Campbell, E. G. (2016). *The Changing Nature of Scientific Sharing and Withholding in Academic Life Sciences Research: Trends From National Surveys in 2000 and 2013*. 18.