

Carmen Paul

Matthew Mayernik

LIS 545

2/4/23

Term Project - Data and Metadata Profile

Attempt 2

Database: [Figshare](#)

Data set: [If You're House Is Still Available, Send Me an Email: Personality Influences Reactions to Written Errors in Email Messages](#)

Data

What are the data?

The data are the collected results from research concerning personality and reactions to typos and grammatical mistakes (Boland & Queen, 2016b).

Where do the data come from, i.e. what is their origin?

The data set was shared by two listed authors: Julie E. Boland and Robin Queen. The data is from their original research. There is no further identifying information about the authors, but the acknowledgements section of the article, published in PLOS ONE, refers to the Psycholinguistics Group at the University of Michigan (Boland & Queen, 2016a). It looks like they are [Julie Boland](#), a Professor of Psychology and Linguistics,

and [Robin Queen](#), Professor of Linguistics, English Language and Literatures and Germanic Languages and Literatures (*Julie Boland*, n.d.; *Robin Queen*, n.d.).

Who are the key stakeholders for the data? Stakeholders could include the creators, users, managers, funders, or other relevant individuals/organizations.

The stakeholders for this data set include Boland and Queen and their colleagues at the University of Michigan, including not only the previously mentioned Psycholinguistics group, but also any staff in the psychology or linguistics departments. This data set could also be of interest to any other professionals in those fields or the fields of sociology, IT, etc. It could specifically be used by organizations social media apps or providing email services, as well as anyone utilizing either such apps and services. On the PLOS ONE article page, the authors state that they did not receive funding for this project (Boland & Queen, 2016a, Figures).

How many data files are there? What do they contain? What file format are they?

The data set includes six items:

- **S1 Appendix.docx**, a document that contains the email texts used in the research;
- **S2 Appendix.pdf**, a copy of an email from Richard Redman, of the Health Sciences and Behavioral Sciences Institutional Review Board at the University of Michigan;
- **S3 Appendix.csv**, a table of collated information resulting from the research;
- **S1 Table.docx**, **S2 Table.docx**, and **S3 Table.docx**, which each highlight excerpts of the data from **S3 Appendix.csv** (Boland & Queen, 2016b).

Does the data set come with any usage restrictions? If so, what are they?

This data set is published under a [CC BY 4.0](#) license, which means it can be shared and adapted with attribution (CC BY 4.0, n.d.).

Is there any specific software required to open or analyze the data files?

Technically, the thumbnail view provides nearly complete access to this data set, as the file formats are readable from the browser in that view. **S3 Appendix.csv** contains more than 50 rows, so it must be downloaded for complete access; CSV files can be opened by any text editor or compatible program. DOCX and PDF files, while proprietary file formats, can be opened by most word processing apps, web browsers, and a variety of other programs.

Metadata

What metadata does the data come with?

Each item in this data set comes with “file info” listing metadata, which includes the name of the file, the file format, the size of the file, and an MD5 checksum. The file info also includes a direct download link for the file.

Where is the metadata? E.g. is it in the data files, included in separate files, both?

Is the metadata structured according to any metadata standard? How comprehensive is the metadata provided?

One way that the metadata can be accessed is by clicking on an “information” button in the file view. As the metadata is very brief, there is no easily-identifiable metadata standard being employed.

How could the current data and/or metadata be enriched? In particular, what additional information could be provided to:

- *improve users' ability to discover the data set in the repository environment.*
- *assist somebody unfamiliar with the data to make use of the dataset for new purposes.*

At the most basic level, some sort of tagging at the data level could help users find individual data. For example, **S2 Appendix.pdf** could be tagged “Richard Redman”—the data set entry itself does not mention Redman. At the moment, searching for **Richard Redman** (or **Redman, Richard**) does not return results including this particular item. The data set as a whole includes more metadata than each item at the data level, including keyword and category tags, an abstract, and a link to the published article (Boland & Queen, 2016b). This should aid researchers in finding this data set, but not necessarily in finding individual figures.

Some of the keywords are very specific and, at least on Figshare, are used exclusively by Boland & Queen. For example, the keyword **language attitude questions** is only used for Boland & Queen’s own research. Incidentally, this keyword also revealed figures directly related to the *If You’re House Is Still Available* article and research; in my opinion, it would have been better to collect them all in one data set.

What publications have been written (if any) based on this dataset?

Are any publications listed/provided with the data set?

A link to the PLOS ONE article published by Boland & Queen is included in the metadata for the data set and can be found in an informational sidebar.

What publications can you find using a literature database or general web search engine that cite or otherwise reference the use of this dataset? How did you search for these publications (if any)?

The metadata for this data set includes usage metrics; they do not list any other Figshare entries that have cited this data set. The metadata also includes an Altmetric score and [a link to a detailed analysis of mentions of the PLOS ONE article](#).

Furthermore, Altmetric links to [a Dimension page](#), which seems to be tracking citations of the PLOS ONE article as well as figures used both in this particular Figshare data set and other related data sets. This was my first experience with Altmetric and Dimension and I hope to use these services again in the future.

Although the Altmetric and Dimension analysis pages list a number of references, for this question I performed a general search for **“If You’re House Is Still Available, Send Me an Email: Personality Influences Reactions to Written Errors in Email Messages”** on DuckDuckGo. Most of the results were database entries for the article itself, but a few results were for articles briefly discussing the research results. I didn’t find any sources that referred explicitly to the data set, as they instead cited the article.

References

- Boland, J. E., & Queen, R [Article]. (2016a). If You're House Is Still Available, Send Me an Email: Personality Influences Reactions to Written Errors in Email Messages . *PLOS ONE*, 11(3). <https://doi.org/10.1371/journal.pone.0149885>
- Boland, J. E., & Queen, R [Figshare dataset]. (2016b). *If You're House Is Still Available, Send Me an Email: Personality Influences Reactions to Written Errors in Email Messages* [Data set]. PLOS ONE. <https://doi.org/10.1371/journal.pone.0149885>
- CC BY 4.0. (n.d.). Creative Commons — Attribution 4.0 International. Retrieved February 4, 2023, from <https://creativecommons.org/licenses/by/4.0/>
- Julie Boland. (n.d.). U-M LSA Department of Psychology. Retrieved February 4, 2023, from <https://lsa.umich.edu/psych/people/faculty/jeboland.html>
- Robin Queen. (n.d.). U-M LSA Linguistics. Retrieved February 4, 2023, from <https://lsa.umich.edu/linguistics/people/faculty/core-faculty/rqueen.html>