DATA PAPER

JSON-Encoded Textual Variance Between the Published Versions of David Mitchell’s *Cloud Atlas* in “an Orison of Sonmi ~451”

Martin Paul Eve
All, Birkbeck, University of London, GB
martineve@bbk.ac.uk

This dataset represents my work to interpret the variances between the two published editions of David Mitchell’s *Cloud Atlas* as described in Martin Paul Eve, “‘You have to keep track of your changes’: The Version Variants and Publishing History of David Mitchell’s *Cloud Atlas*, *Open Library of Humanities*, 2(1), http://dx.doi.org/10.16995/olh.82. It was created through a hermeneutic process of cross-comparing the different texts and then marked-up in JSON by hand. Potential re-uses include corrections, further visualization, digital-humanities teaching, and structural analysis.

**Keywords:** Literature; Literary Forms and Genres; Literary Criticism; Fiction

---

(1) Overview

Repository location
https://dx.doi.org/10.6084/m9.figshare.3493103.

Context
This data was produced as part of the research article:
Martin Paul Eve, “‘You have to keep track of your changes’: The Version Variants and Publishing History of David Mitchell’s *Cloud Atlas*, *Open Library of Humanities*, 2(1), http://dx.doi.org/10.16995/olh.82.

(2) Methods
Steps
The dataset was created by close-reading the two versions of David Mitchell’s *Cloud Atlas* [1, 2] side-by-side and making subjective/hermeneutic/interpretational judgements about the functional equivalence, or otherwise, of passages. I then transcribed this into a JSON dictionary.

The methodology for this transcription was as follows:

1. Assign a name to each edition (I used “E” and “P” to represent “Electronic” and “Print” respectively).
2. For each “question” posed by the archivist, assign a sequentially incremented identifier. The first question in the P edition, then, becomes PQ1. The second PQ2 etc. For the Electronic edition, this would be EQ1 and EQ2.
3. For each “answer” or “response” provided by the Fabricant Sonmi ~451, assign a sequentially incremented identifier. The first answer in the P edition, then, becomes PR1, the second PR2 etc. In the E edition, this would be ER1 and ER2.

This methodology for identifying textual passages works well for this chapter of the novel, since the text is strictly divided into alternating question and answer sections.

The final part of the methodology was to match sections between the two editions and to map crossovers. This is the most contestable part of this dataset but a reference table for how this has been achieved can be found in the appendix of the associated article at: http://dx.doi.org/10.16995/olh.82.

Those who wish to move from this dataset back to the original text should count questions and responses from the start of the respective edition. Verification that the correct point has been reached can be made with reference to the aforementioned appendix.

Sampling strategy
The dataset covers the section of *Cloud Atlas* entitled “An Orison of Sonmi ~451” in its entirety. A side-by-side comparison of the texts indicated that this is where the primary differences lie between the versions.

Quality Control
The data was checked at multiple stages by the author.

(3) Dataset description

Object name
CloudAtlasVariants.json
Other scholars may wish to check and revise the interpretations of functional equivalence that I have pos-
ited in this dataset and re-issue future versions that have greater community consensus around the interpretation.
Perhaps most importantly, though, this dataset serves as a base on which others may draw in tracing the tex-
tual genetics of contemporary fiction. This dataset is also deliberately designed to be ingested by various D3.js tools 
for visualization and I hope that others will re-use this dataset within novel visualization paradigms for textual 
genetics. Digital humanities courses that wish to pursue visualization of textual genetics may also find this dataset 
of use as a relatively small and manageable corpus that can be set as in-class exercises.
There are no barriers to the re-use of this dataset as no copyright portions of the fictional work are here reproduced.

Acknowledgements
My thanks to David Mitchell for his help in ascertaining the cause of the version variants here documented and 
to Caroline Edwards for her patient and detailed editorial work on the accompanying article.

Competing Interests
The author has no competing interests to declare.

References

How to cite this article: Eve, M P 2016 JSON-Encoded Textual Variance Between the Published Versions of David 
org/10.5334/johd.6
Published: 25 November 2016
Copyright: © 2016 The Author(s). This is an open-access article distributed under the terms of the Creative Commons 
Attribution 4.0 Unported License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, 
provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.