



DATA DE-IDENTIFICATION OVERVIEW

OER in and as MOOCs

Laura Czerniewicz, Andrew Deacon, Sukaina Walji and Michael Glover

The authors have administered a de-identification process in which any information which was deemed to have the potential to identify research subjects, whether in isolation or combination with other data, has been either omitted or reworded. In all cases, the authors have attempted to preserve as much of the original data as possible; and have tried to ensure that the integrity of original response was maintained.

The de-identification process utilised two techniques:

- **Omission:** Whenever disclosive information was included in a response, certain elements of that response (such as a word or phrase) have been omitted so as to obscure identity, while preserving as much of the intended meaning and salience of the response as possible. To make the data as seamless and interpretable as possible, there are no indications included in the data signalling where the omissions have taken place.
- **Revision:** In cases where disclosive information was given but could not be omitted without jeopardising the integrity of the response data, the authors have revised words or phrases to ensure anonymity, while retaining the essential meaning. In most cases, these instances entail changing specific information to something less particular, such as replacing a departmental identifier with a faculty-level identifier (see the “Text revised” section below for an example).

Throughout the data, [brackets] were used to indicate intervention mechanisms for de-identification and clarification.

1. **Text revised []:** Brackets indicate that a word or series of words have been altered in a way that preserves original meaning as far as possible. For example, if the original response was: “I work in the Sociology Department”; this was altered to: “I work in the [Faculty of Humanities].”
 2. **Unclear to transcriber [unclear]:** In cases where the transcriber was unable to interpret components of a response, “[unclear]” was inserted as an indicator in the text.
 3. **Redaction [redacted/section redacted]:** In cases where answers to certain questions are judged to be too disclosive (particularly when combined with ancillary responses), the authors have redacted certain key responses. Areas where paragraphs or larger sections of text were redacted are marked with “[section redacted]”.
-