Cochran Systematic Review Update Dataset Structure and Manual

by Amal Alharbi and Mark Stevenson

July 2019

The dataset consists of 25 Intervention Cochrane reviews. For each review, it contains information for both the original and updated version of the review. This include:

- Review title.

- Boolean query.

- PMID, title, abstract and MeSH for each study returned from search.

- PMIDs for included studies.

- Review keywords.

- Terms and MeSH execrated from the Boolean query.

The structure of the dataset is illustrated in Figure 1.

The dataset is available in `GitHub` repository as a `pickle` file format (`update_dataset.pkl`) which can be accessed using the following Python code:

```python
import pickle
reviews = pickle.load(open('reviews.pkl','rb'))
```

To access the list of PMIDs of included studies based on abstract level:

```python
#for the original review
abs_included = reviews['CD000155']['included']['original_ids']
#for the updated review
abs_included = reviews['CD000155']['included']['update_ids']
```

To access the full records of search results for the original review:

```python
search_results = {}
for review in reviews:
    search_results[review] = reviews[review]['search_results']['original_records']
```
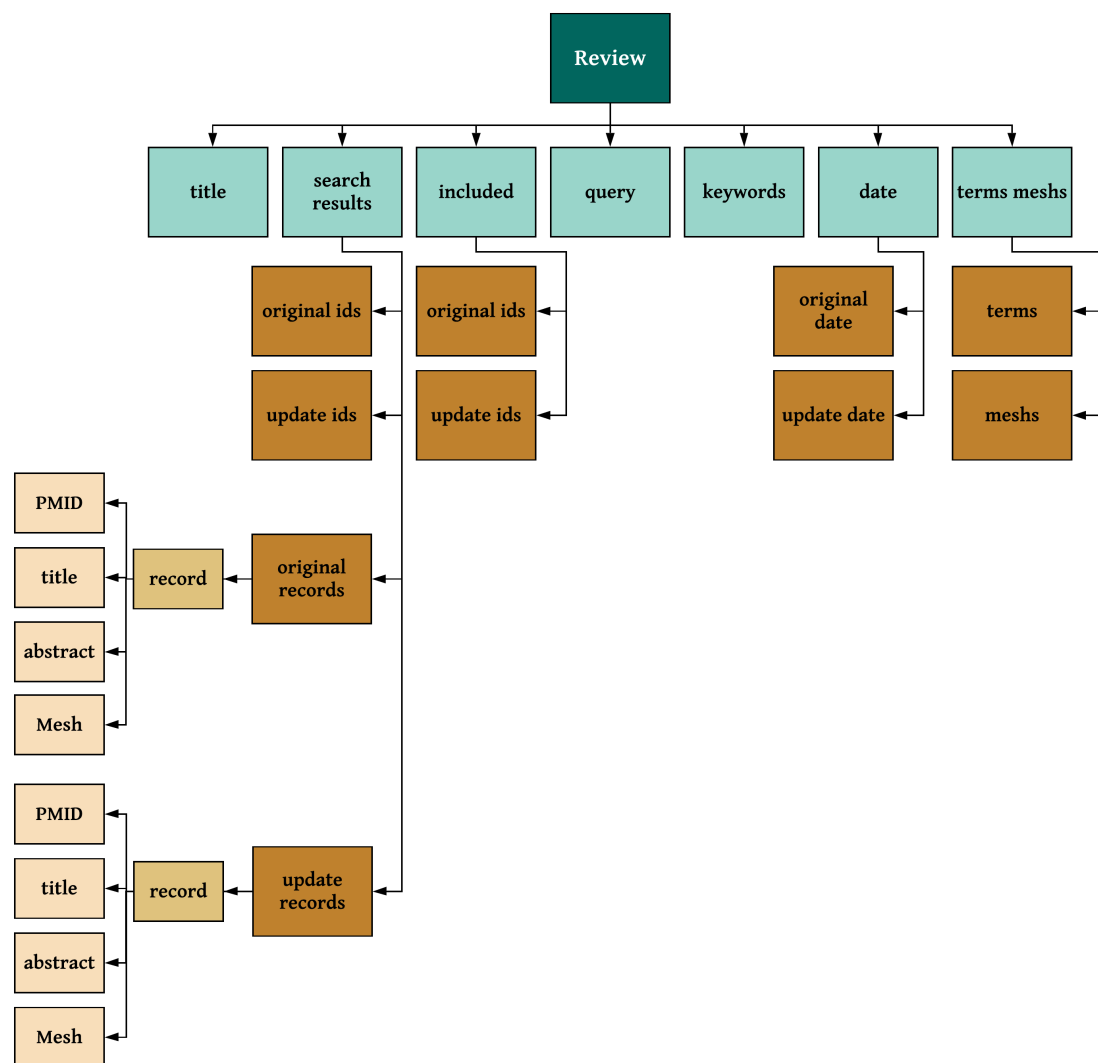
Figure 1: The Structure of the Cochran Update Dataset.