Introduction

The OpenTrialsFDA team is one of the six finalists for the Open Science Prize. This is a collaboration between Drs. Erick Turner, Ben Goldacre and the OpenTrials team at Open Knowledge International. The work is closely related to the OpenTrials project, which aims to locate, match, and share all publicly accessible data and documents, on all trials conducted, on all medicines and other treatments, globally.

OpenTrialsFDA works on making clinical trial data from the FDA (the US Food and Drug Administration) more easily accessible and searchable. Until now, this information has been hidden in the user-unfriendly Drug Approval Packages that the FDA publishes via its dataportal Drugs@FDA. OpenTrialsFDA will extract the relevant data from the FDA documents, link it to other clinical trial data and present it through a new user-friendly web interface. Any user will be able to type in a drug name, and see all the places where this drug is mentioned in an FDA document. Users will also be able to access, search and present this information through the application programming interfaces (APIs) the team will produce.

Unlocking the trove of FDA clinical trials data

The FDA Drug Approval Packages report the methods and results of all premarketing clinical trials, both published and unpublished. Interestingly, the FDA data is unbiased, compared to reports of clinical trials in academic journals. This is because FDA reviewers, unlike most academic journal reviewers and editors, have access to the original trial protocol and can spot and reject any attempts at post hoc changes to planned outcomes and statistical analyses. Therefore, a systematic review (SR) or meta-analysis (MA) based on FDA Drug Approval Packages, which includes unpublished and “unspun” trial results, will often conclude that the drug in question is less safe and/or less effective than what one would conclude from a conventional SR or MA based solely on the published literature.

Unfortunately, despite their high value, these FDA documents are notoriously difficult to access, aggregate, and search. The Drugs@FDA website itself is not intuitive to navigate, and the reviews are stored as individual PDFs which, except for relatively new drugs, cannot be searched. As a consequence, they are rarely used by clinicians and researchers. OpenTrialsFDA will improve this situation, so that valuable information that is currently hidden away can be discovered, presented, and used to properly inform evidence-based treatment decisions.

The OpenTrialsFDA prototype

The team will scrape the FDA website and extract the relevant information from the PDFs through a process of OCR (optical character recognition). Through the new OpenTrialsFDA interface, users will be able to explore and discover the FDA data. In addition, the information will be integrated into the OpenTrials database, so that the FDA report can be linked to reports from other sources, such as ClinicalTrials.gov, EU CTR, HRA, WHO ICTRP, and PubMed.

The six finalists of the Open Science Prize will showcase their prototypes at the BD2K Open Data Science Symposium on 1 December 2016, when public voting will begin. The public is asked to help select the most promising, innovative and impactful prototype from among the six finalists - of which one will receive the grand prize of $230,000.