Discovery: Preliminary Outcome/Capability Analysis for Open FDA Data

Analysts: Robert Damashek, Gail Chen

Sources: Open Web Search on Challenges, Risks and Disruptions between FDA and Regulated Industry; Discovery Interviews with FDA and Industry SMEs

Date: June 18, 2015

Articles Reviewed:

1. “Makers of Generic Drugs Challenge F.D.A. Plan for Updated Warnings”, New York Times, March 27, 2015. <http://www.nytimes.com/2015/03/28/science/makersofgenericdrugschallengefdaplanforupdatedwarnings.html?_r=0>
2. “Challenges and Opportunities for the Generic Drug Industry”, Remarks at the GPhA Annual Meeting as Delivered by Margaret A. Hamburg, M.D., FDA Commissioner, February 23, 2012. <http://www.fda.gov/NewsEvents/Speeches/ucm294978.htm>
3. “Regulation in the Medical Device Industry: FDA insider lays out the issues”, Metropolitan Corporate Counsel, March 19, 2015. <http://www.metrocorpcounsel.com/articles/31944/regulationmedicaldeviceindustryfdainsiderlaysoutissues>
4. “Device Tax Repeal, FDA’s Actions Affecting Industry Remain Top Issues”, Bloomberg Bureau of National Affairs, January 1, 2015. <http://www.fr.com/wp-content/uploads/2015/01/Bloomberg-BNA.Medical-Device-Law-Industry-Report.January21.2015.pdf>
5. “Project FDA Report: The Digital Future of Molecular Medicine: Rethinking FDA Regulation”, Manhattan Institute, May 2013. http://www.manhattan-institute.org/pdf/fda\_06.pdf

Candidate Capability: The FDA Administrator needs the ability to readily understand and be able to intuitively demonstrate the impact of enhancing product labels on adverse drug events.

Why Important: FDA has been an advocate of more stringent drug labelling, and is now involved with industry in a controversial effort to impose greater labelling requirements on manufacturers of generic drugs.  The value/cost of this is the subject of debate in the news, so an aid in researching the value of labelling may be of interest to key stakeholders.

Candidate User Story: An FDA Analyst wants to determine if there is any correlation between changes in drug labelling and a reduction in adverse events across all drugs.

Related User Story: A Pharmaceutical Researcher wants to determine if there is any correlation between changes in drug labelling and a reduction in adverse events for a specific drug or related drugs.

Candidate Tasks:

* Identify events where a product label changed for a drug, and extract the drug and date of change.
  + The drug labels are accessible [here](http://dailymed.nlm.nih.gov/dailymed/spl-resources-all-drug-labels.cfm) on an HHS site.  The Open FDA data on SPL is all about a standardized data structure (XML) format for the data, but the data itself in that format doesn’t seem to be on the Open FDA site.  Look at the SPL structure to determine the field that contains the date of change.  On the label site, there appear to be a series of monthly updates to the labels.  These files are huge also.  There is also a huge (6+ GB) aggregate file, but not the historical archives.
  + Drug safety label changes by month are listed by the FDA [here](http://www.fda.gov/safety/medwatch/safetyinformation/safety-relateddruglabelingchanges/default.htm) for the past 7.5 years. This will likely need to be scraped to get it into usable form (e.g. EXCEL) for data analysis.
* Identify adverse events on the same drug, and attempt to determine if there is a correlation between the date and aggregate number of adverse events (positive or negative) and the date of the label change. One might expect that the label change has some positive effect (e.g. reduction in adverse events) going forward, but you never know!
* Visualize the results in aggregate and allow drill down to specific drugs.  If there is a drug category, this might be used as a filter to select or omit certain categories.
* Filter based on other key fields like generic/non-generic, drug manufacturer, state/location of the adverse event, type of correlation, magnitude of correlation.
* More complex might be to aggregate a drug with its generic and manufacturer- specific brand variants, assuming that data is readily available.