

Data Treasure Hunting

URL: <https://2015.spaceappschallenge.org/challenge/data-treasure-hunting/>

Hashtags:

#humans, #datatreasurehunting, #advanced

Contact:

datatreasurehunting@spaceappschallenge.org

Tags:

Platform, Data Visualization

Background

In recent years, NASA and other government agencies worldwide have been publishing open data in machine-readable, non-proprietary, and no-cost format on the web (e.g., <http://data.nasa.gov/>). Everyone is interested in new ways to search that publicly available data and integrate these information assets into innovative databases and applications.

Inconsistent metadata (i.e., information such as keywords that empower search engines such as Google to discover these assets) is a consistent challenge across organizations. The challenge is to develop a new technique or application that would enable anyone to add meaningful keywords to the descriptions of our data – keywords that describe the hidden potential of these assets to better leverage our data beyond space applications to other data that may appear unrelated.

Challenge

Devise a clever way to discover good keywords to describe the potential, hidden, secondary uses of open data. For example, how might you discover that a particular information asset might be relevant to or benefit from other keywords, such as waste-processing or disaster-preparedness? Remember: without these additional and

seemingly unrelated keywords, entrepreneurs like you might not discover and use open data to solve your most perplexing problems.

You can use any technique that might help discover new keywords. For example, a crowdsourcing application could display information about these assets online and query people about how the assets can be used. You may want to consider predictive analytics or machine-learning techniques to compare the metadata and the data of one information asset to another in order to find new keywords. Or, you might use the unique identifiers of the published data-files to search on the web, discover who already used the data and for what purpose, then catalog it. In fact, you could even develop a clever solution to download the data itself and 'squeeze it' in order to generate new keywords.

Not only are we asking you to discover new keywords, but also to retain the log file that explains how these new keywords were discovered.

Considerations

A starter toolkit will be made available. This will include the complete existing metadata and download links for information assets that were published on Open Data websites or by other agencies worldwide.

Sample Resources (Participants do not have to use these resources, and NASA in no way endorses any particular entity listed).

- <https://project-open-data.cio.gov/schema> - Provides a dictionary of existing metadata-fields in the popular data.json catalog that agencies use to prepare and upload information assets to their Open Data portals.
- <http://www.engagedata.eu> - The European Engage project that uses a crowdsourcing technique to address a similar problem.
- <http://www.opendataresearch.org/project/2013/odb> - The Open Data Barometer project that uses an expert system to address a similar problem.