Ideas generation

1. Looked at the options for reducing congestion on hasting street. This is a very complex problem and was rejected as a possibility
2. Generate a databasing website/application to increase usability and user to data interfacing .
3. Expand on data set usage

Idea created/expanded

1. To generate an interface for accessing data with a more statistical approach to allow the user to cross check the validity and credibility of the data set.
2. Application of search parameters to filter through different datasets (i.e. sorting through different government departments, title keywords, file types, date submitted etc.)
3. Creating a ranking algorithm to better separate data.
4. Integrate visualization suite for quick and easy representations of data sets and refined subsets of data.
5. Scraping existing data from current databases, as well as individual submissions all into one database
6. Incorperate API search queries and CKAN…

Datasets used:

http://qldspatial.information.qld.gov.au/catalogue/custom/detail.page?fid={AFB67F25-0571-4C7B-A06B-4EF1CB4A6DD3}

This dataset was utilised as an example for our DataStat search function. Mock data was generated by using the metadata as reference.

https://soil-chem.information.qld.gov.au/odata/SiteLabMethodResults

CHALLENGES APPLIED FOR:

Bounty: Making open data more open. {national}

How can open data be presented on search.data.gov.au to make it easier and friendlier to use? Does this mean making it more similar to using standard search engines, like Google, or something else entirely?

Searching for datasets on data.gov.au can be a long-winded process, due to the sheer amount of data stored on the site. By using a more refined search function, through introducing further parameters, the amount of datasets that a user desires can be narrowed down to a more manageable outcome.

This is the aim of DataStat, where the metadata of these datasets remains the same, however, the search function includes an additional parameter of selecting a category that may correspond to the organisation or department selected. Furthermore, it was identified that data.gov.au does not host the entirety of the some of the organisations catalogue. An intended purpose of DataStat is to centralise data from all organisations to ensure ease of access from users across a wide spectrum.

DataStat features a ranking system that uses an implemented AI to rank the data, factoring in credibility, availability and formatting of the data. This system can be compared to the current one on data.gov.au, however, a user that is unfamiliar with how the site operates may question why a government department dataset has a 2 or 3 star rating, as first impressions of this dataset may result in the user deeming it unsatisfactory. Additionally, we think the implementation of the number of citations the dataset has undertaken can prove beneficial in identifying data quality. We believe these are integral steps in selecting open data for further application, and therefore DataStat introduces a platform where users can rate the dataset according to similar parameters listed above, as well as how relevant it was to their processes.

FInally, we noticed that users can request data that may not be readily available on the site. It is safe to assume that data.gov.au may run into instances where they do not have the relevant data, and acquisition of such will be a lengthy period. Therefore, we believe that giving the power of uploading datasets to users, and even private groups, can assist in availability of data and overall improve the catalogue of data that can be distributed.

Science Data Challenge {state}

How might we make discovering and understanding scientific data for a location possible?

The current issue when searching for location data online is the vast majority of databases that have no connection to each other. This means that if you wish to find a dataset for a location, you may have to sift through 3-4 different websites before you find something you can use. In the field you don't have time to waste searching for information.

Our solution involves scraping data from any amount of existing databases, and uploading all that information onto one database. An A.I. implemented search engine will allow for a single search result page with all available datasets.

The process of scraping data has a secondary varifing process, that removes anomalies and rates the quality of the data. This helps to make the data clear and concise to the user, giving them a better understanding of what the data may be telling them.

Bounty: Visualise the Numbers {national}

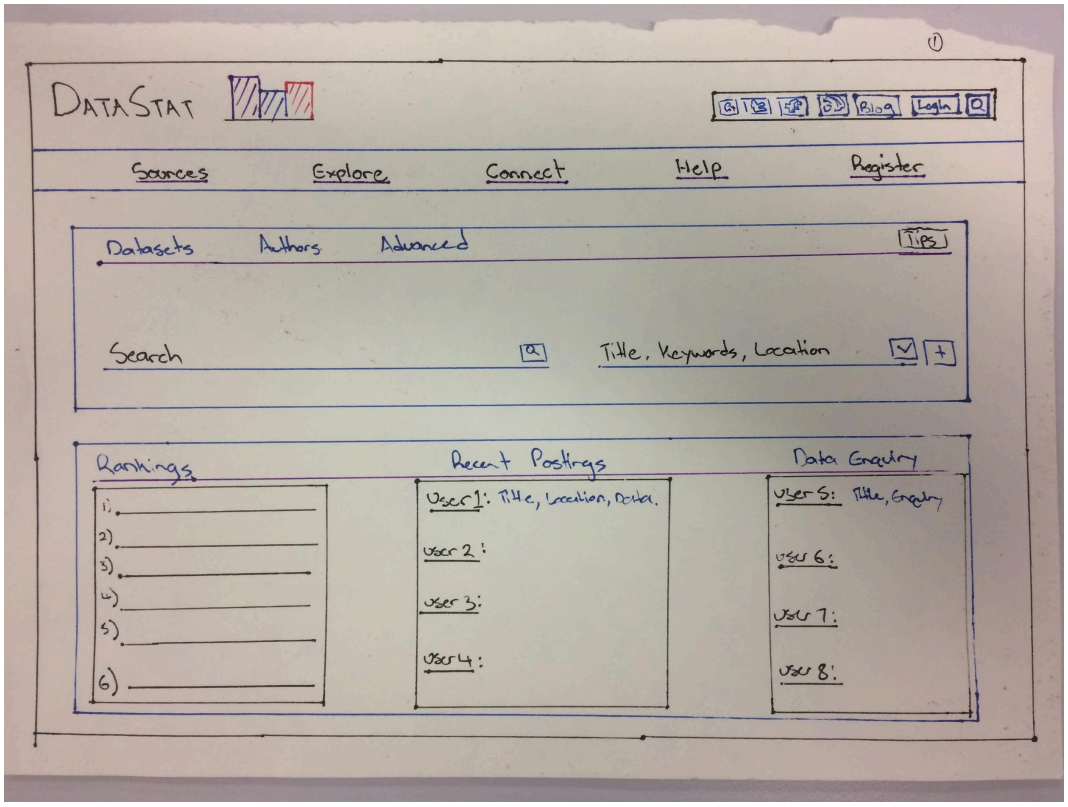
How can people better view data on GovCMS in visuals?

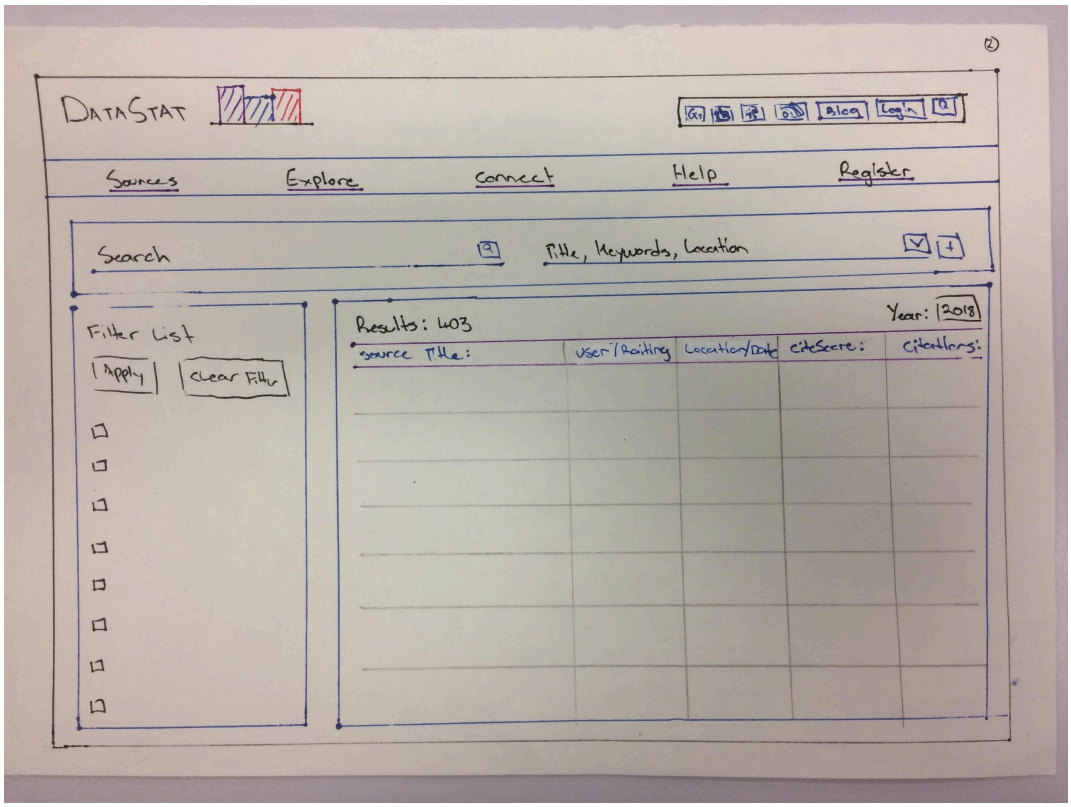
Most Commercially Viable Project {state}

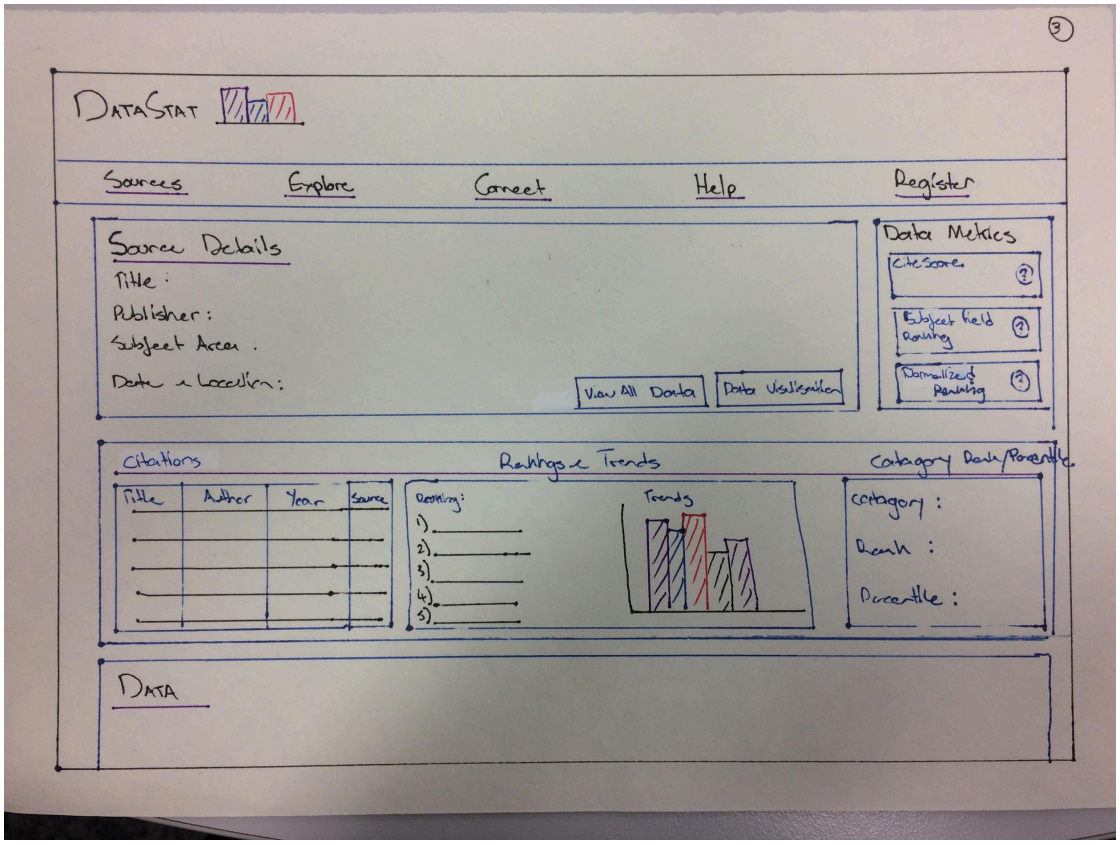
Most Commercially Viable Project

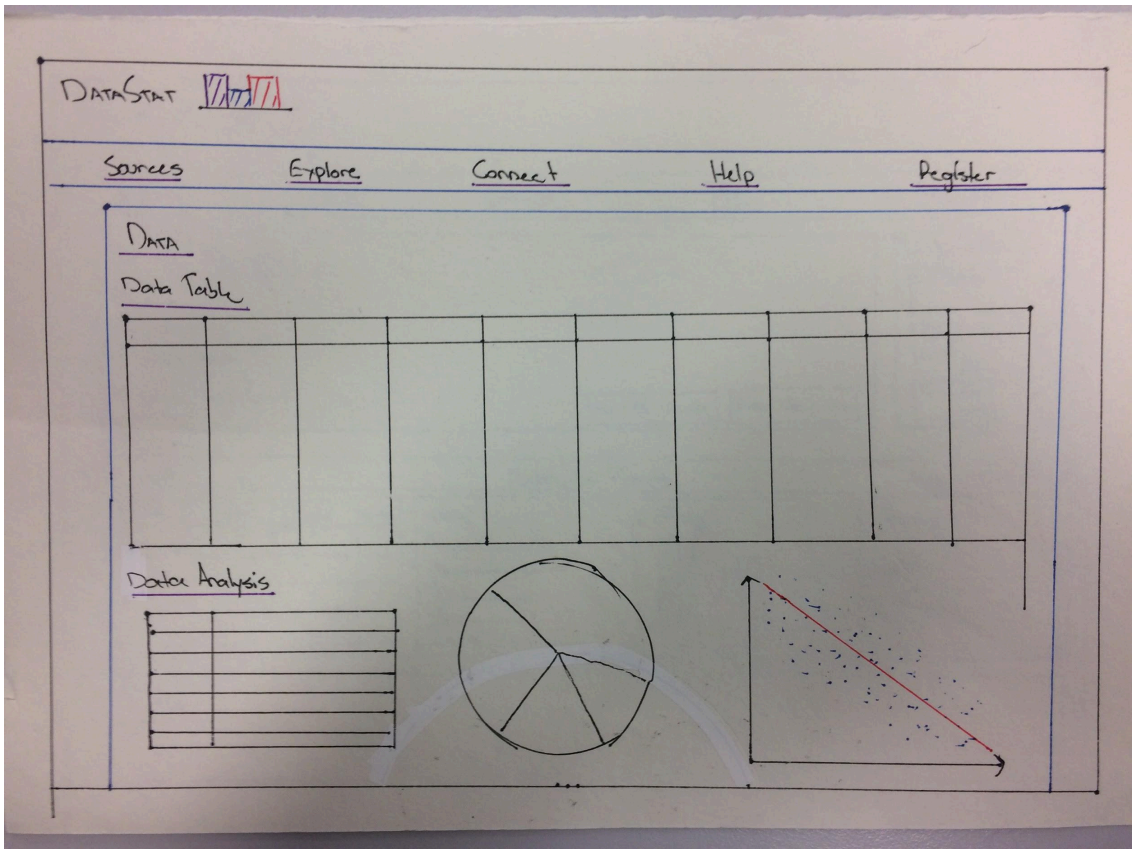
DataStat allows for streamline searching for datasets, saving time and money. Our concept can give companies quality data that can be instantly assessed to ensure they are using the best available information.

Conceptual design for Datastat interface

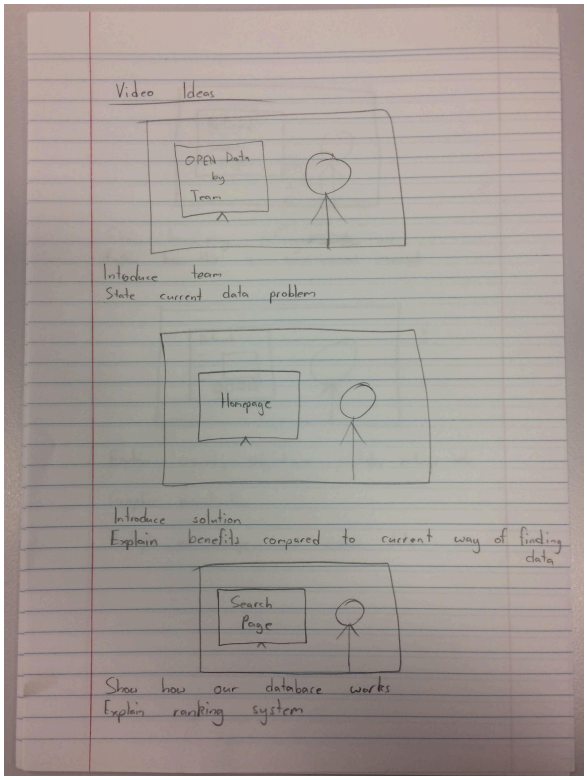

①
DataStat
Blog
Login
Sources
Explore
Connect
Help
Register
Datasets
Authors
Advanced
Tips
Search
Title, Keywords, Location
Rankings
1)
2)
3)
4)
5)
6)
Recent Postings
User 1: Title, Location, Date.
User 2:
User 3:
User 4:
Data Enquiry
User 5: Title, Enquiry
User 6:
User 7:
User 8:

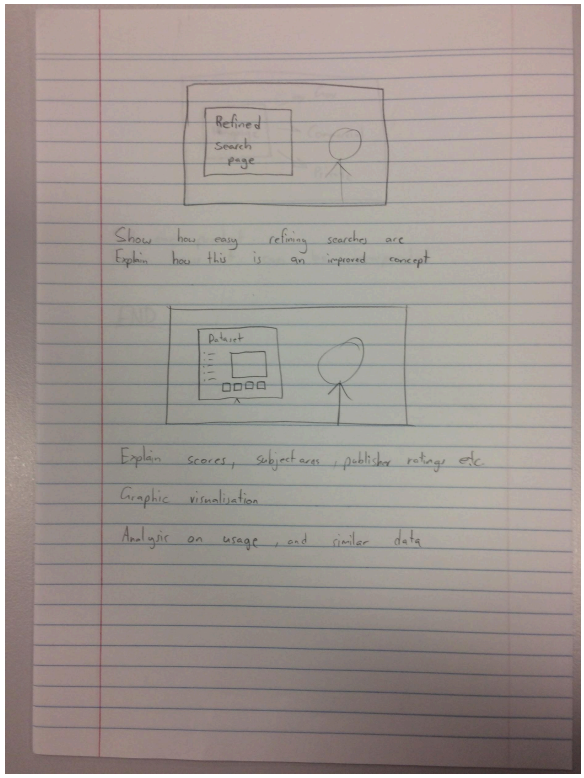

②
DataStat
Blog
Login
Sources
Explore
Connect
Help
Register
Search
Title, Keywords, Location
Filter List
Apply
Clear Filter
Results: 403
Year: 2018
Source Title:
User Rating
Location/Date
citeScore:
citations:

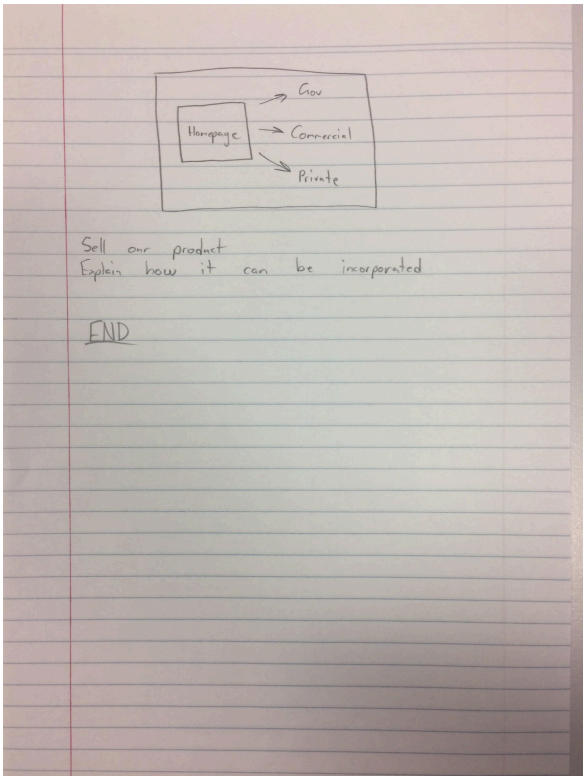

3
DataStat
Sources
Explore
Connect
Help
Register
Source Details
Title:
Publisher:
Subject Area:
Data Location:
View All Data
Data Visualization
Data Metrics
CiteScore
Subject field Ranking
Normalized Ranking
Citations
Rankings Trends
Category Rank/Percentile
Title
Author
Year
Source
Ranking:
1)
2)
3)
4)
5)
Trends
category:
Rank:
Percentile:
Data


DataStat
Sources
Explore
Connect
Help
Register
Data
Data Table
Data Analysis

Concept of initial storyboard for pitch video







Script for pitch video

First Slide

1. (show govhack image) Hello and welcome to GovHack 2018

2. {show data stat logo} During this video, our Team will outline the conceptual design of an improved meta search engine for raw data.

NEW SLIDE #3

3. Currently a significant issue exists when searching for raw data. Numerous databases exist, and searching through them one by one can be very tedious.

NEW SLIDE #4

Even a detailed search on google or any open data host site will still give a broad range of different results or datasets.

Existing databases are often cluttered with information, and don't provide the tools needed to quickly and effectively refine a search for specific data. Additionally comparing credibility of individual datasets is at the discretion of the user.

NEW SLIDE #5

4. Our team has focused on addressing this issue in several ways by proposing our idea for a web based platform "DATASTAT".

Our platform will act as an international hub for data and will utilise an AI based meta search engine and statistical analysis software to locate and carry out a quality check to determine its validity.

NEW SLIDE #6

DataStat will gather data sets from either individual sources, or existing databases using web APIs to query existing **Comprehensive Knowledge Archive Networks** (**CKAN**), this will consolidate data sets and metadata.

NEW SLIDE #7

5. When comparing “DataStat” to the current methods of locating data, it is our objective to make searching efficient and user-friendly.

   By consolidating all types of data onto one site allows for streamline searching between different disciplines and fields.

   This idea is fortified by including AI based filtering system to refine a search using a large number of parameters, ranging from selecting data from a specific department, to searching for a single file type. This allows the user to narrow the search down to a minimal and desirable result.

NEW SLIDE #8

6. Our algorithms and statistical software will look at the source of data and produce a rating that corresponds the credibility, formatting and metrics of the dataset, as well as its effectiveness through feedback from previous users.

   This will help the user save time when deciding between a number of datasets by using the comparative ranking system.

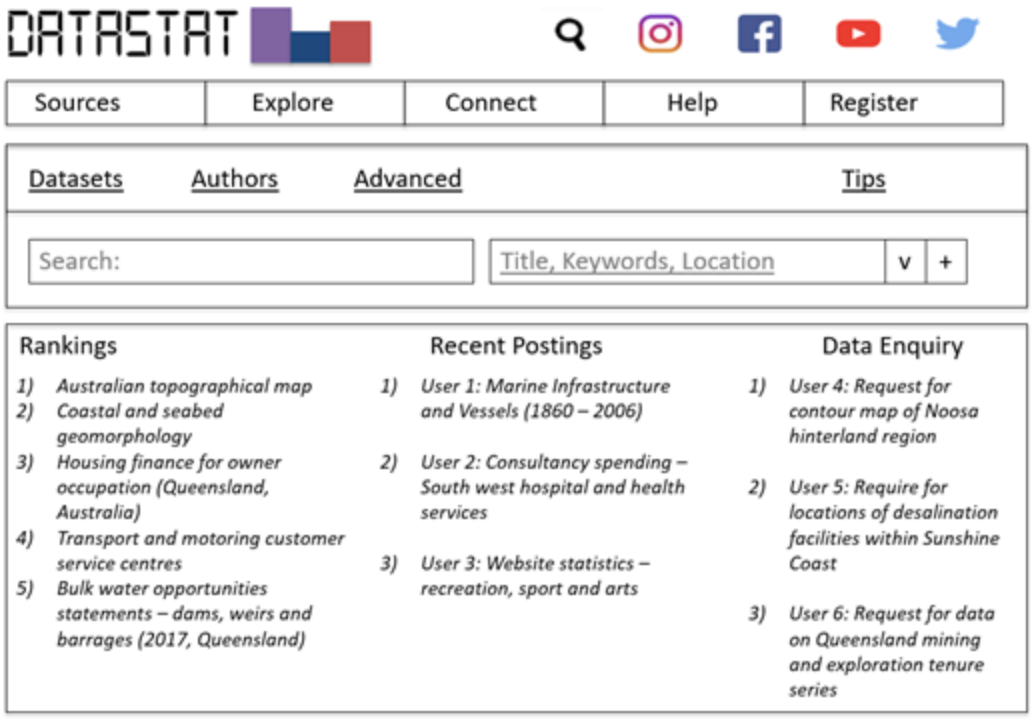
NEW SLIDE #9

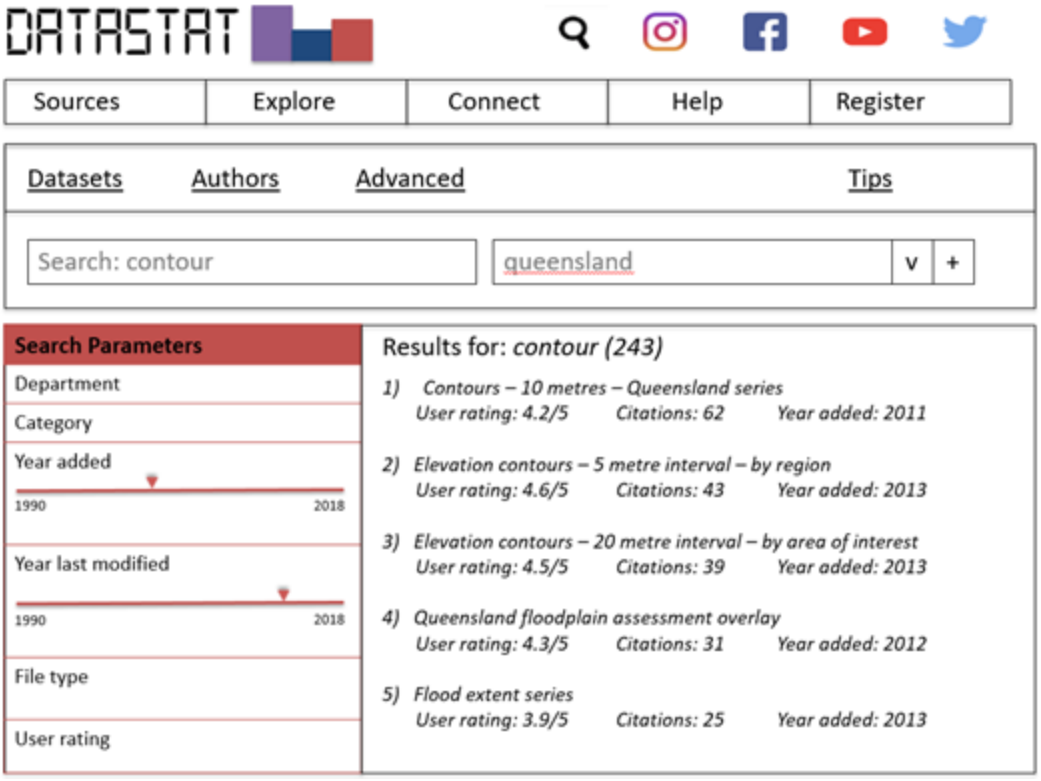
7. unlike many other sites out there, DataStat compiles with raw data and allows the user to directly output data in a graphical format which can give users a visual representation of the data before conducting further analysis.

   Our product will draw in users with its user friendly interface, ease of use, and ranking system that allow the users to receive accurate results faster for users from all sectors.

   MAKING DATA FOR THE END USER EASIER TO FIND and ANALYSE.

First Interface examples for web design





DATASTAT

Sources | Explore | Connect | Help | Register

Contours – 5 metres – Queensland series

Publisher: Department of Natural resources, mines and energy
Subject Area: Elevation contours
Date and location: 2016-08-16, Brisbane, Queensland

View all metadata | Data visualisation

Data Metrics

Cite score: 4.6

Subject field ranking: 4

Normalized ranking: 4.5

Rankings & Trends

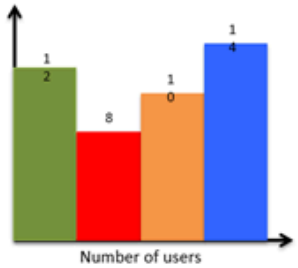
Rankings:

Daniel Smith: 5/5

Luke Batista: 4/5

Lee Shewchenko: 4/5

Jono Jordan 5/5



View all rankings | View all citations

Category Rank

Category: Department of natural resources, mines and energy

Rank: 4

Percentile: 88%

Data

DATASTAT

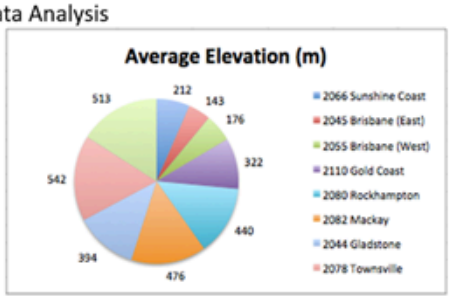
Sources | Explore | Connect | Help | Register

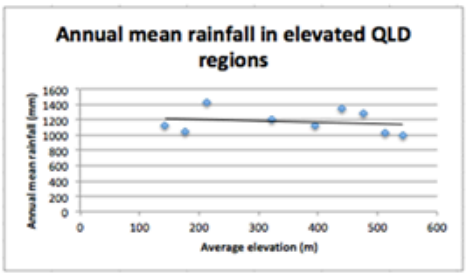
Data

Data Table

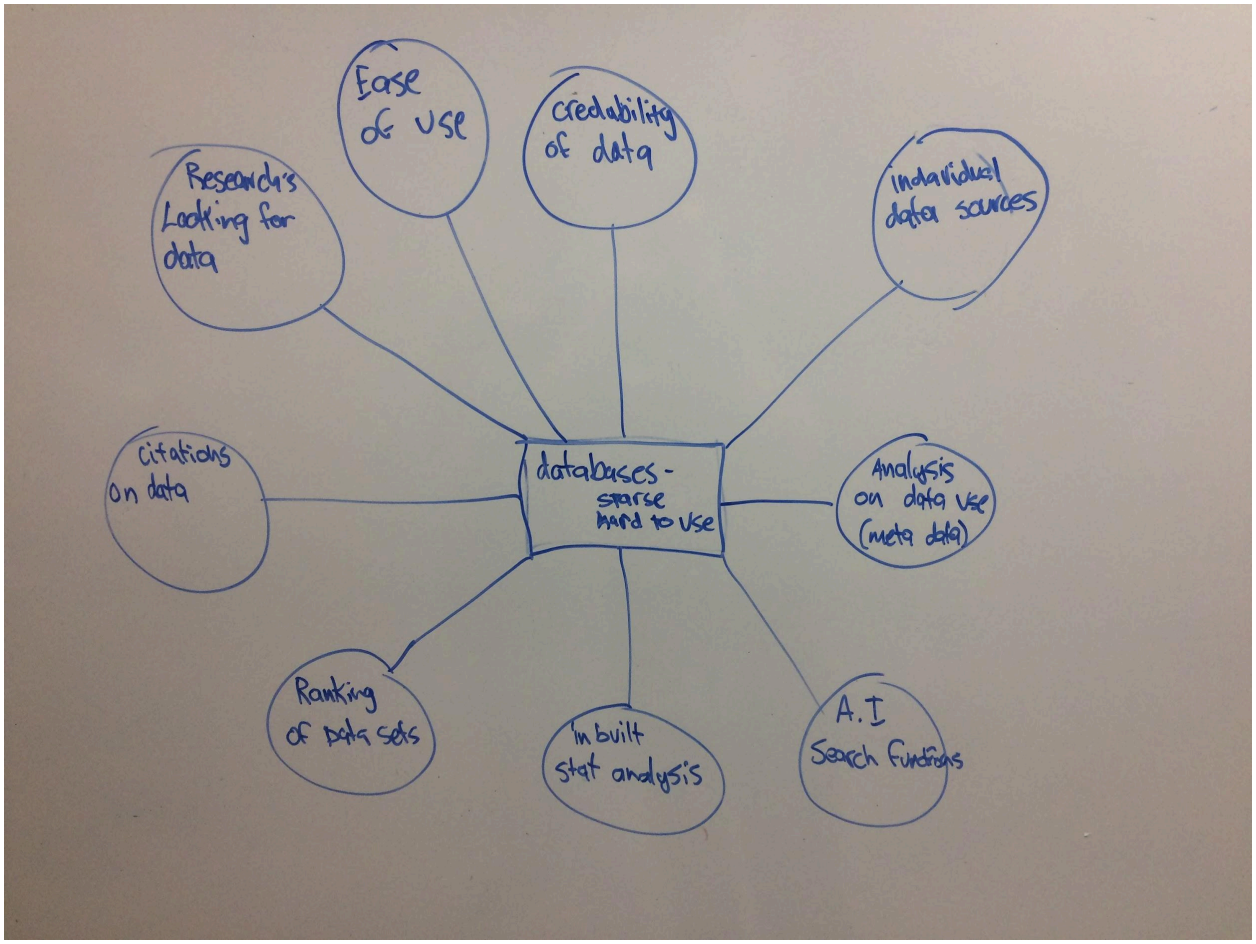
Application ID	Region	Average Elevation (m)	Annual mean ranfall (mm)
2066	Sunshine Coast	212	1432
2045	Brisbane (East)	143	1123
2055	Brisbane (West)	176	1045
2110	Gold Coast	322	1200
2080	Rockhampton	440	1343
2082	Mackay	476	1284
2044	Gladstone	394	1119
2078	Townsville	542	996
2094	Cairns	513	1026

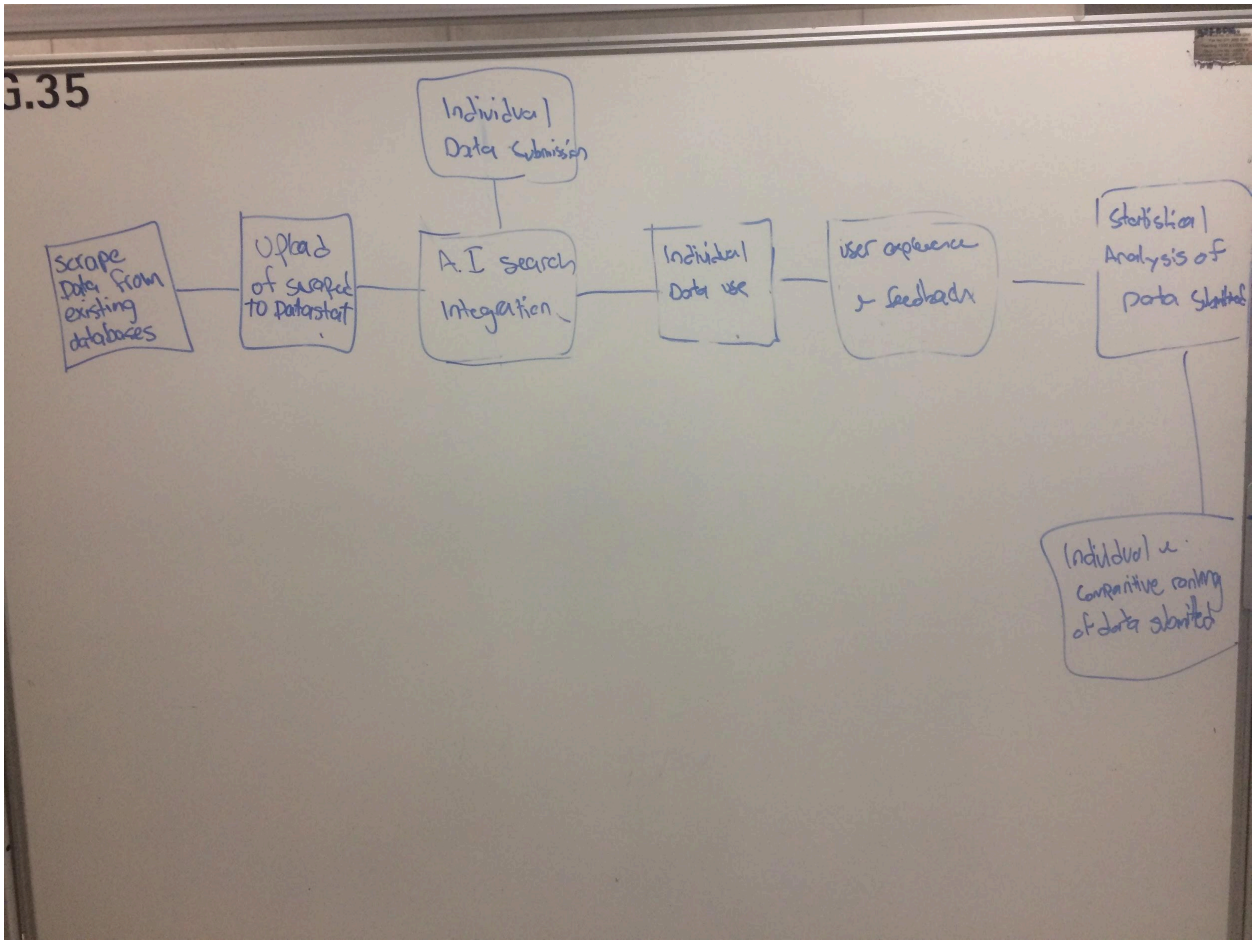
Data Analysis

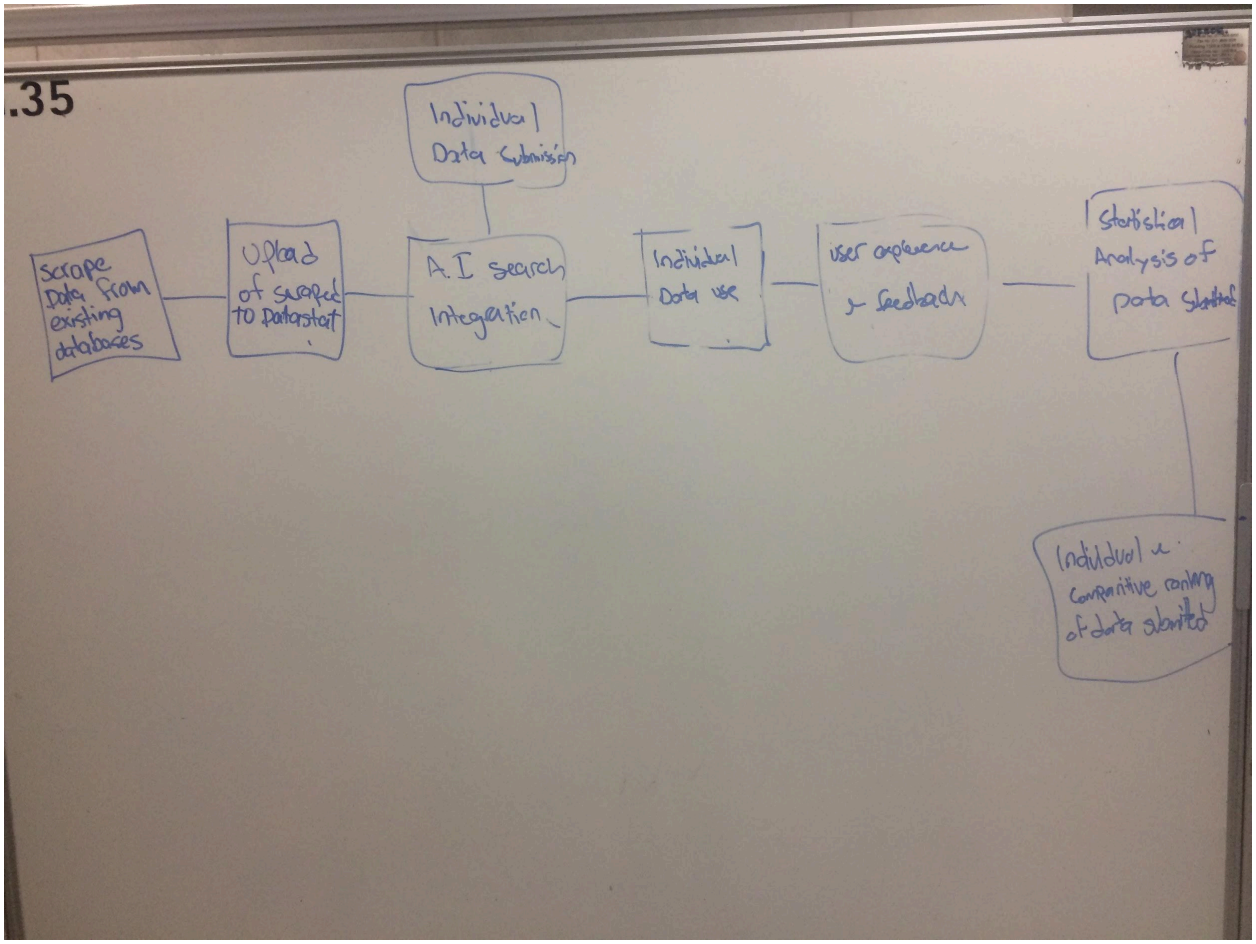




Mind-mapping introductory ideas


Ease of use
credability of data
Researchers Looking for data
individual data sources
citations on data
databases - sparse hard to use
Analysis on data use (meta data)
Ranking of data sets
In built stat analysis
A.I Search functions


G.35
Individual Data Submission
Scrape Data from existing databases
Upload of scraped to Datastart
A.I search Integration
Individual Data use
user experience & feedback
Statistical Analysis of data started
Individual comparative ranking of data started



https://soil-chem.information.qld.gov.au/odata/SiteLabMethodResults

Url for drupal: https://www.drupal.org/wproject/govcms_ckan

Url for pitch video: https://youtu.be/Jlxwt6T4G40