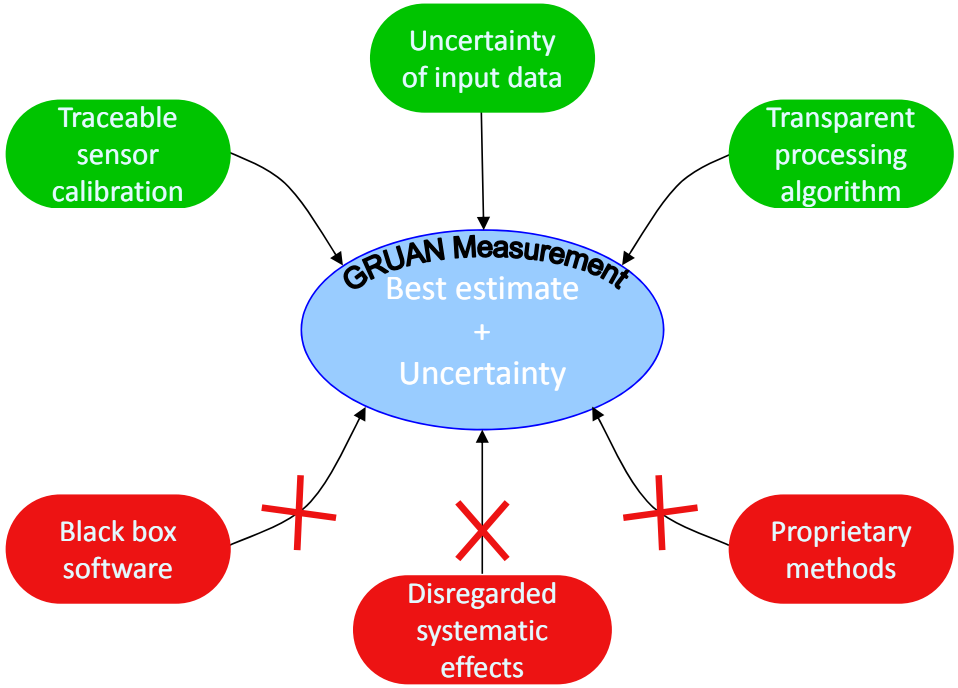
Data and meta-data exploration and data quality reporting for GCOS

Jared Lewis (jared@bodekerscientific.com)

Bodeker Scientific, New Zealand

GRUAN

- GCOS Reference Upper Air Network
- Continuous long-term records of high-quality reference data
- Best possible uncertainties



The problem

- Extracting value from meta-data is difficult
- Large number of variables
- Large amounts of data
- Need to know the question you want to answer

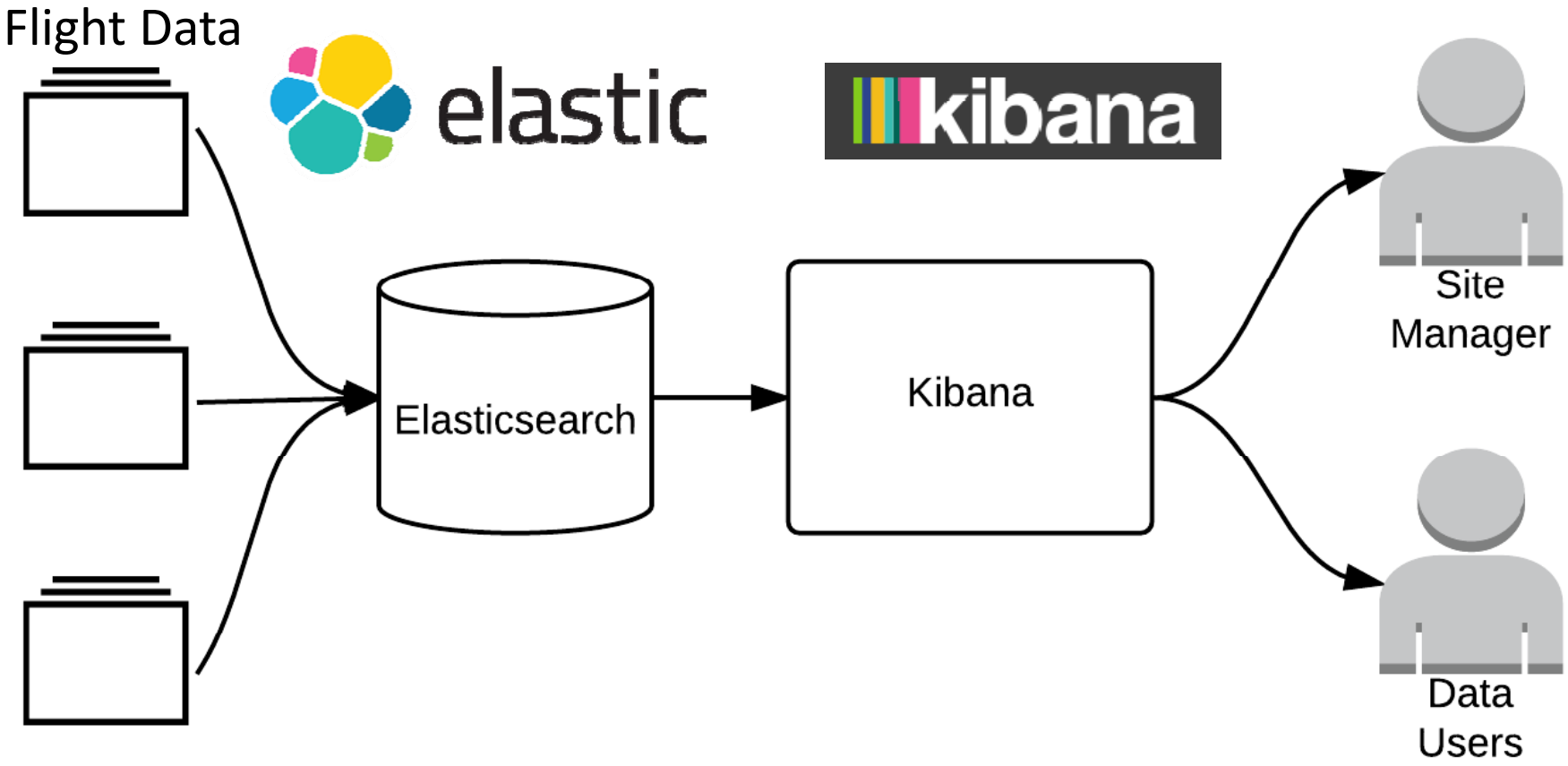
Similar problem

- Websites create large volumes of logs containing meta-data
- Meta-data provides information about customers
- Large incentive to extract value for decision making ($$$)
- Sparked the “Big Data” revolution
  - New industry using meta-data
  - New tools

Visualising Meta-data

Log files

Visualising Meta-data



Total valid flights

21,474

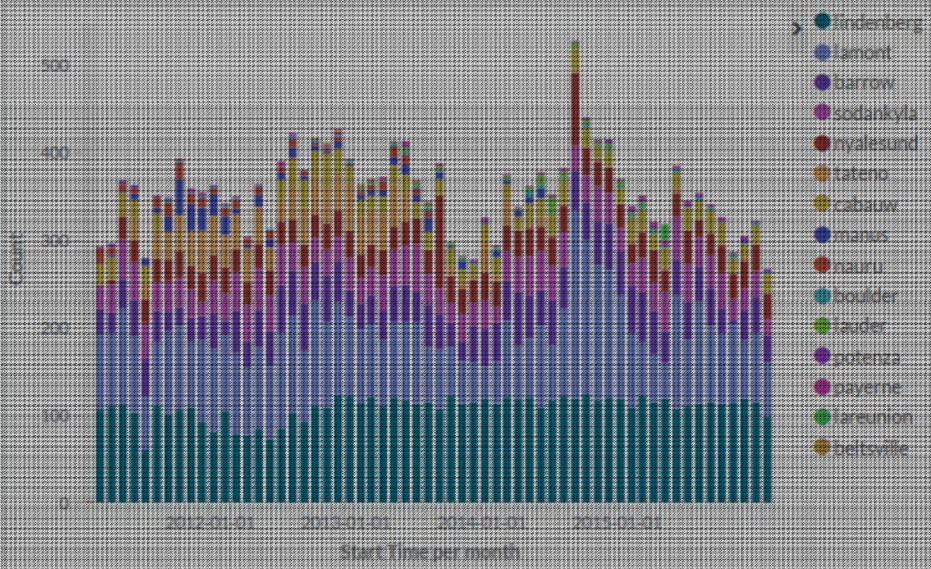
Count

Number of sites providing data

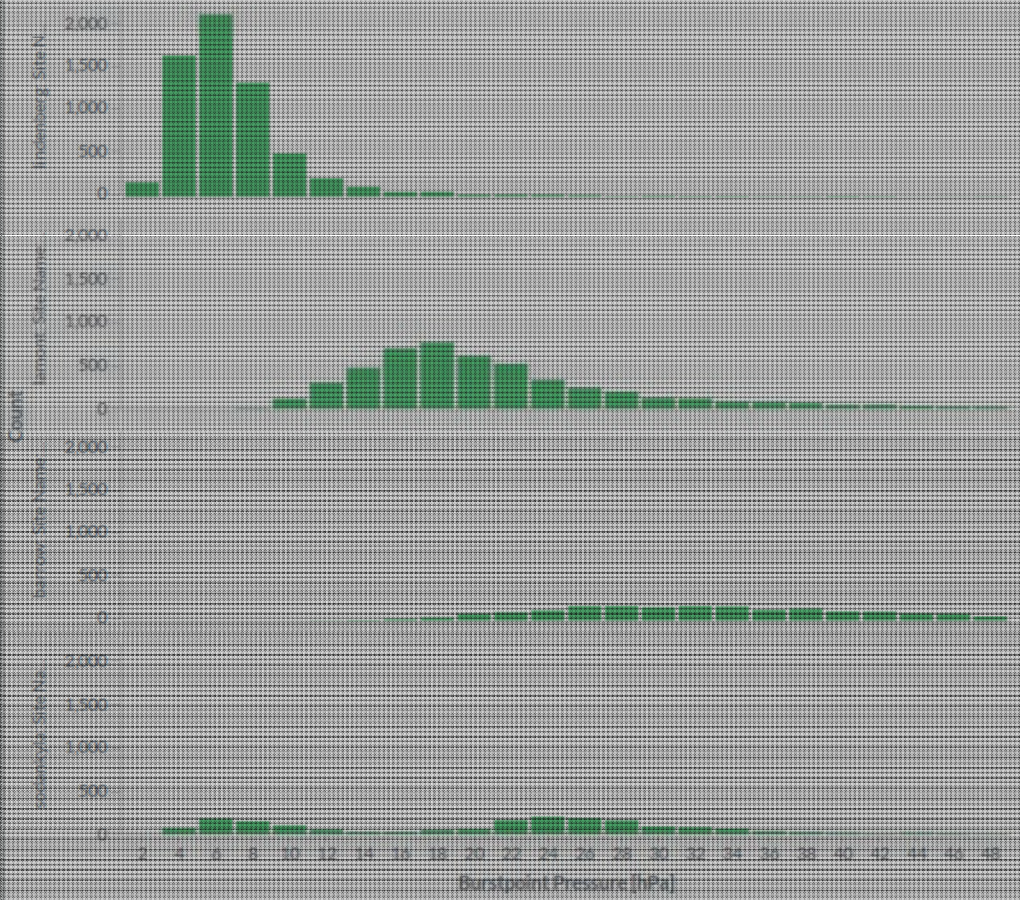
15

Unique count of Site Code

Launches per site per month



Burst-point pressure for top 4 sites by flight count



Create a new visualization

Step 1

Area chart	Great for stacked timelines in which the total of all series is more important than comparing any two or more series. Less useful for assessing the relative change of unrelated data points as changes in a series lower down the stack will have a difficult to gauge effect on the series above it.
Data table	The data table provides a detailed breakdown, in tabular format, of the results of a composed aggregation. Tip, a data table is available from many other charts by clicking grey bar at the bottom of the chart.
Line chart	Often the best chart for high density time series. Great for comparing one series to another. Be careful with sparse sets as the connection between points can be misleading.
Markdown widget	Useful for displaying explanations or instructions for dashboards.
Metric	One big number for all of your one big number needs. Perfect for showing a count of hits, or the exact average a numeric field.
Pie chart	Pie charts are ideal for displaying the parts of some whole. For example, sales percentages by department.Pro Tip: Pie charts are best used sparingly, and with no more than 7 slices per pie.
Tile map	Your source for geographic maps. Requires an elasticsearch geo_point field. More specifically, a field that is mapped as type:geo_point with latitude and longitude coordinates.
Vertical bar chart	The goto chart for oh-so-many needs. Great for time and non-time data. Stacked or grouped, exact numbers or percentages. If you are not sure which chart your need, you could do worse than to start here.

Potential Benefits

- Empowers users to answer their own questions
- Identify where the data may be falling short of satisfying user's needs
- Provide quantitative feedback for sites
- Benchmark existing sites
- Identify problems early

Outlook

- Kibana is being actively developed to be more user friendly
- Help the GRUAN working group to capture useful metrics
- Create site specific dashboards and metrics
- Need to find other datasets which could use these tools