Industrial Big Data – when Big Data meets Big Business

Dean Allemang
Working Ontologist, LLC
dallemang@workingontologist.com
Think of some data set you might be able to use if you had it

- What is it about?
- Where does it come from?
- What could you do with it?
What Data?

Business Data
What Data?

Textual Data
What Data?

**Scientific Data**
Value from Corporate Big Data

Gather Data

Improve sales

Advertising
Product Placement
Product Development
Production
The superior man understands what is right; the inferior man understands what will sell.

- Confucius
Industrial Big Data
Business Value of Sharing Data (at an industrial level)

• Science as a Service

• Pre-competitive Research
Challenges of Science as a Service

What’s easier?

Perform the experiment (again)

Find the results of an experiment
**Finding Experimental Results**

What procedure did we perform?
What were the inputs?
What were the control variables?
What was measured?
What were the results?

...  

**Well-known search techniques won’t be very helpful**

<table>
<thead>
<tr>
<th>Trial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sprite</td>
<td>11.4</td>
<td>11.9</td>
<td>12.6</td>
<td>12.8</td>
<td>13.1</td>
</tr>
<tr>
<td>7-up</td>
<td>11.9</td>
<td>12.4</td>
<td>12.8</td>
<td>13.1</td>
<td>13.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sprite</td>
<td>13.7</td>
<td>14.1</td>
<td>14.5</td>
<td>15.0</td>
<td>15.5</td>
</tr>
<tr>
<td>7-up</td>
<td>14.1</td>
<td>14.5</td>
<td>15.0</td>
<td>15.5</td>
<td>15.8</td>
</tr>
</tbody>
</table>

*All values +/- 0.3 mL.

**Observations:**
- Both acids were clear after indicator was added
- Turned pink for a longer period of time after each drop added
- Stopped tightly pink after a certain amount of drops

**Calculations:**
**Average of the amount of NaOH used**
First volume minus the initial volume for trials first through last to find the differences between final and initial volumes. Find the sum of the differences, which is then divided by a number of trials performed to find the average.

Finding differences: Trial 1: 11.9 - 11.4 = 0.5 mL; Trial 2: 11.9 - 12.4 = -0.5 mL; etc.
Finding sum of the differences: Sprite: 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 = 3 mL.
Dividing by the amount of trials: Sprite & 7-up: 3 / 6 = 0.5 mL.

**Error:**
Find the difference between the average and the number that is farther away from the average.
Sprite & 7-up: 0.4 - 0.3 = 0.1 mL.

**Concentration:**
Write a balance equation.

Base concentration * base liters = moles of base * mole ratio of acid/mole ratio of base = moles of acid/liters of acid = concentration of acid

Sprite and 7-up both contained citric acid:

\[ \text{NaOH} + C_6H_8O_7 \rightarrow H_2O + NaC_6H_7O_7 \]
Data Challenges for Science as a Service

- Representing experiments
- Inputs
- Controls
- Measurements
- Matching different descriptions
Pre-competitive Research

Repeatability means that anyone can perform the experiments. They’ll get the same results.
A rising tide lifts all boats.

- John F. Kennedy
Industrial Big Data Challenges – Tower of Babel
Industrial Big Data Challenges – Financing

Who should fund the basic research that produces the data?
Industrial Big Data Challenges – Control and Ownership

Private funding leads to infighting

Public funding doesn’t respond to the market
IMI / OpenPHACTS
Control Solution

The Private Sector poses the questions, the Public Sector selects the answers.
OIL AND GAS
Integrated Operations in the High North

- hydraulics
- geology
- electronics
- navigation

Semantic Model
MATERIALS ENGINEERING
Integration Computational Materials Engineering

External Linkage

Flexible Queries

Provenance

Extensions

Future-Proofing

ICME
US National Research Council
FINANCE
Financial Industry Business Ontology

http://www.edmcouncil.org/financialbusiness
If I have seen further it is by standing on the shoulders of giants.

- Sir Isaac Newton