Industrial Big Data – when Big Data meets Big Business

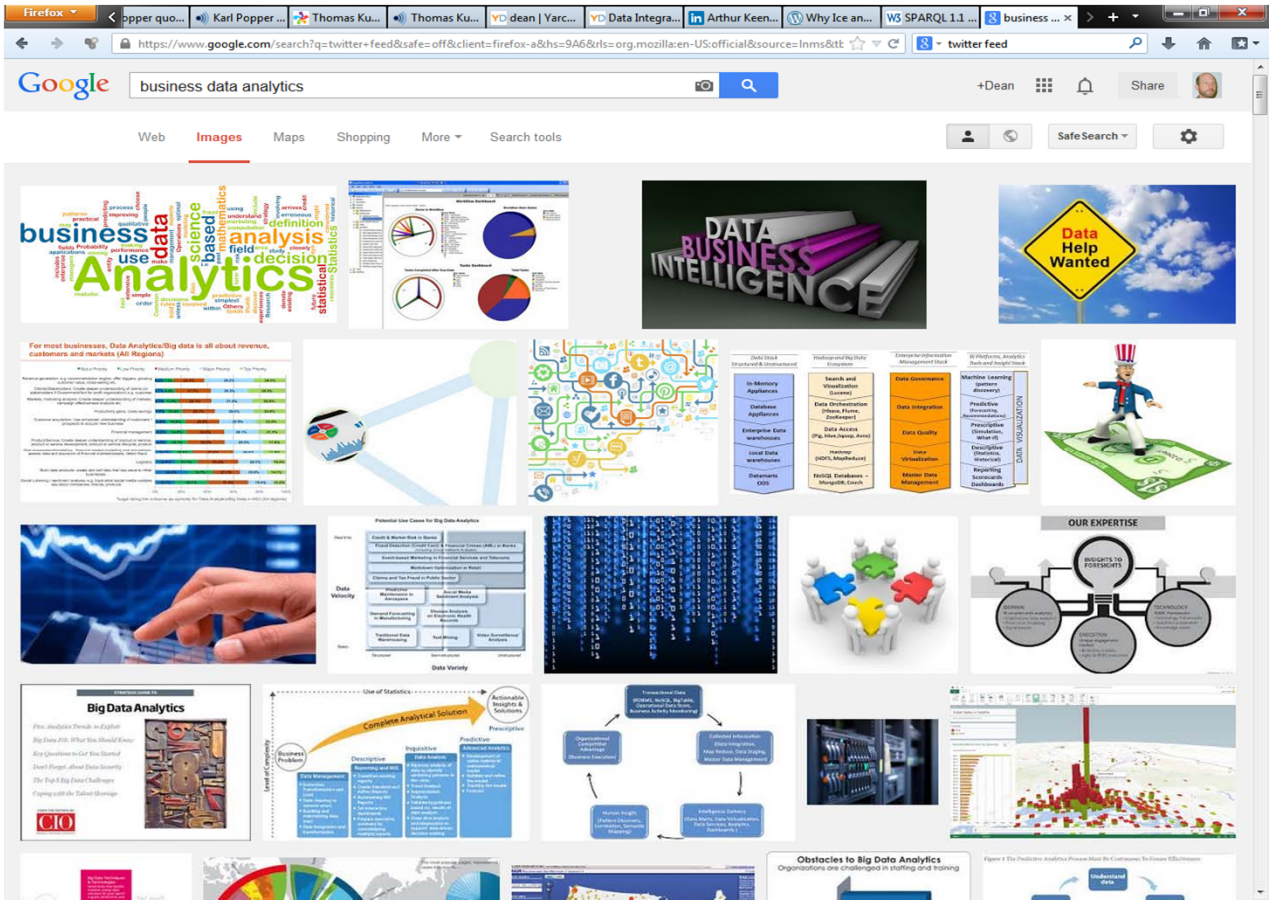
Dean Allemang
Working Ontologist, LLC
dallemang@workingontologist.com

BIG DATA

Think of some data set you might be able to use if you had it

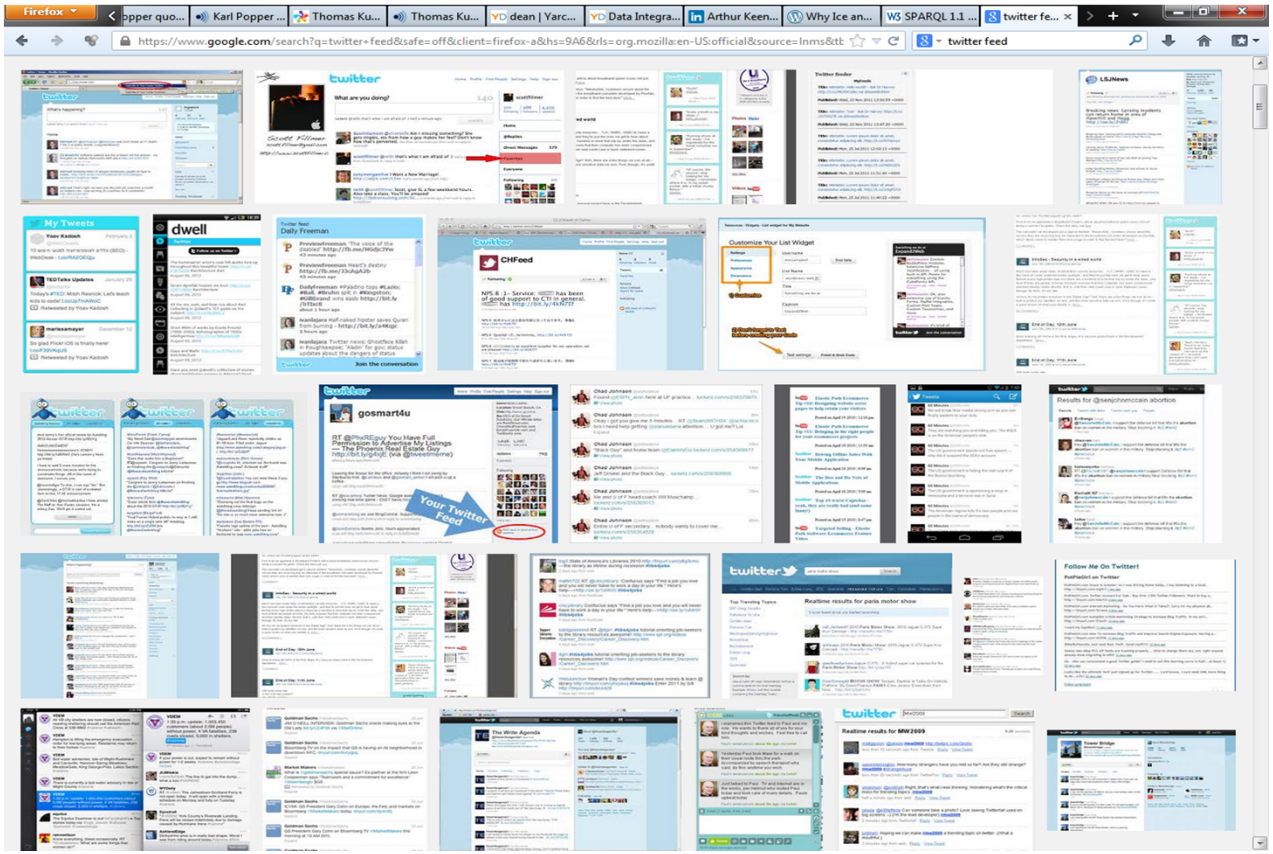
- What is it about?
- Where does it come from?
- What could you do with it?

What Data?



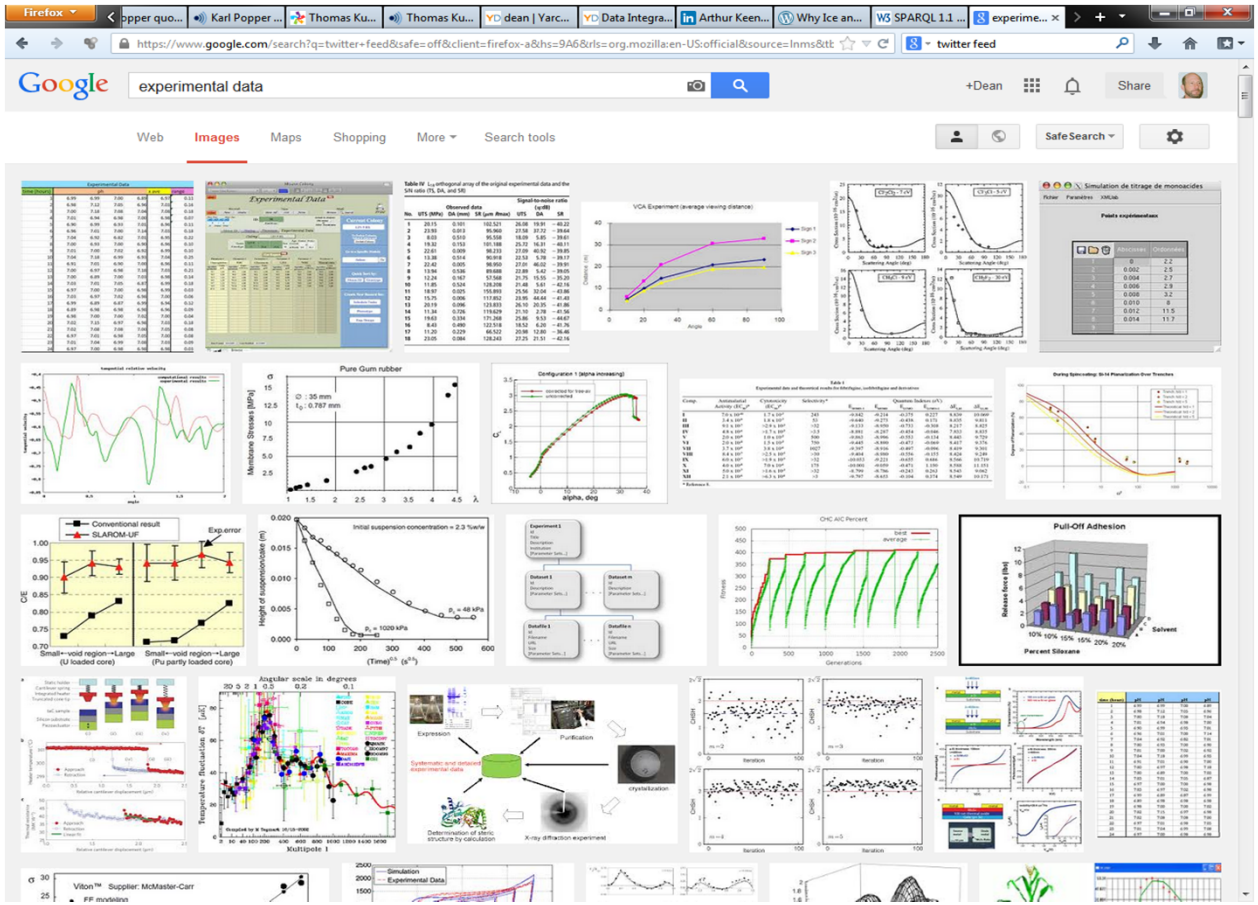
Business Data

What Data?



Textual Data

What Data?



Scientific Data

Value from Corporate Big Data

Gather Data

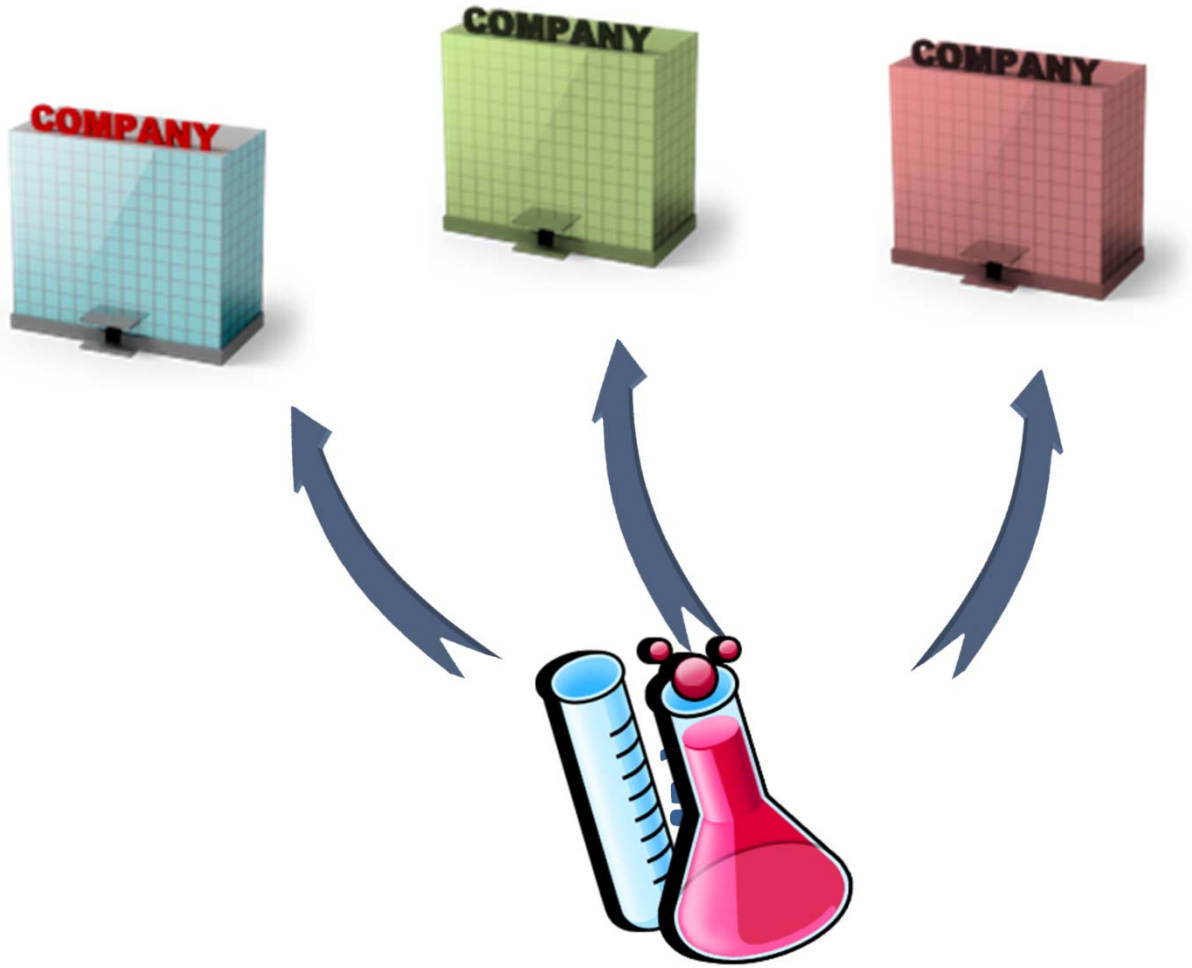
Advertising
Product Placement
Product Development
Production

Improve sales

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

Science as a Service

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

Titration of 7-up and Sprite using 1.0 M NaOH

		trial 1	trial 2	trial 3	trial 4	trial 5
Sprite	initial vol. of NaOH (mL)*	11.4	11.9	12.4	12.8	13.1
	final vol. of NaOH (mL)*	11.9	12.4	12.8	13.1	13.5
7-up	initial vol. of NaOH (mL)*	13.7	14.1	14.5	15.0	15.5
	final vol. of NaOH (mL)*	14.1	14.5	15.0	15.5	15.8

*All values +/- 0.1 mL

Volume of NaOH used (mL)*	trial 1	trial 2	trial 3	trial 4	trial 5	average	error
Sprite	0.5	0.5	0.4	0.3	0.4	0.4	0.1
7-up	0.4	0.4	0.5	0.5	0.3	0.4	0.1

*All values +/- 0.1 mL

Observations:

- Both acids were clear after indicator was added
- Turned pink for longer period of time after each drop added
- Stayed lightly pink after a certain amount of drops

Calculations:

Average of the amount of NaOH used

Final 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 amount 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 + 0.4 + 0.3 + 0.4 = 2.1 mL ; 7-up: 0.4 + 0.4 + 0.5 + 0.5 + 0.3 = 2.1 mL
Dividing by the amount of trials: Sprite & 7-up: 2.1/5=0.4 mL

Error:

Find the difference between the average and the number that is further away from 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:
NaOH + C6H8O7 = H2O + NaC6H7O7

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

LIFE SCIENCES

IMI / OpenPHACTS

Innovative Medicines Initiative

http://www.imi.europa.eu/

http://www.openphacts.org/

IMI / OpenPHACTS

EMBL-EBI

ChemSpider
The free chemical database

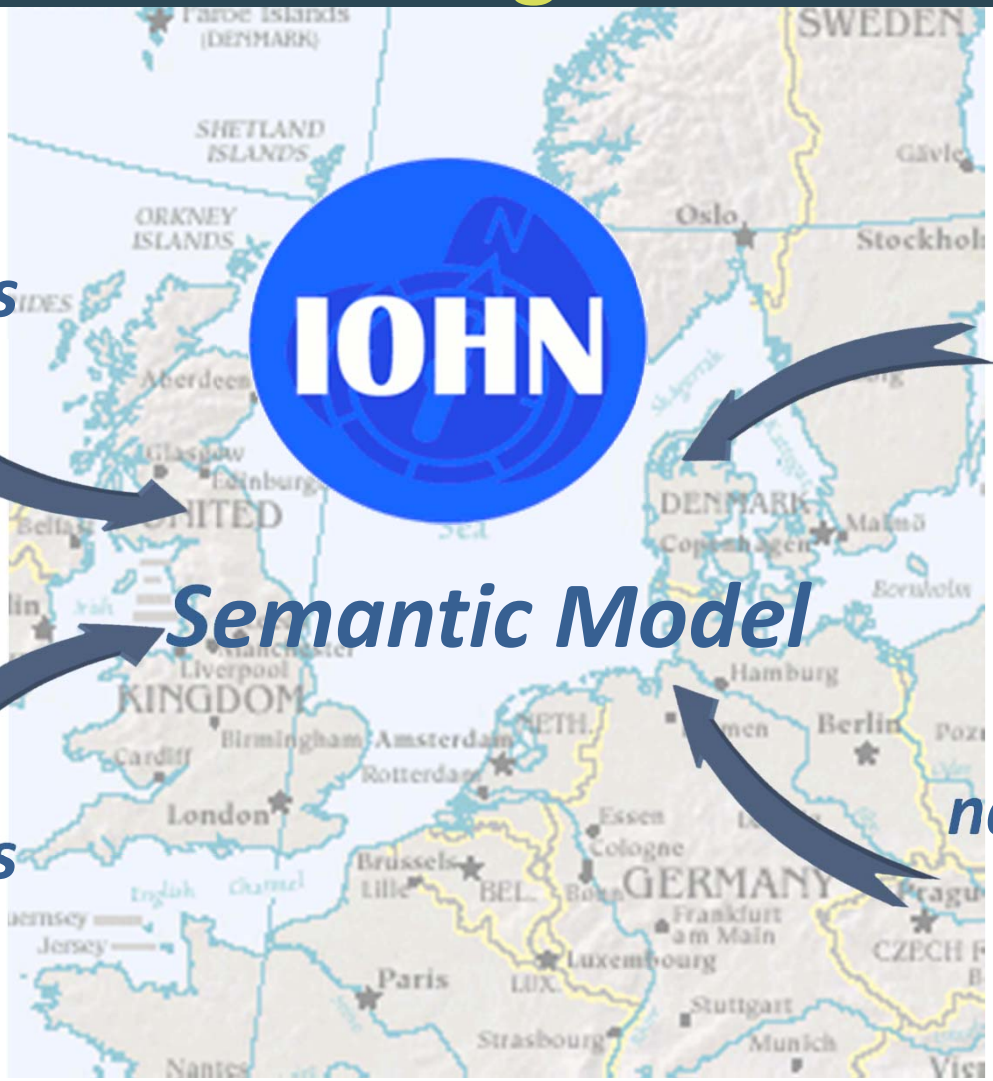
ChEBI

Control Solution

The Private Sector poses the questions, the Public Sector selects the answers.

OIL AND GAS

Integrated Operations in the High North



MATERIALS ENGINEERING

Integration Computational Materials Engineering

External Linkage

ICME
US National Research Council

Flexible Queries

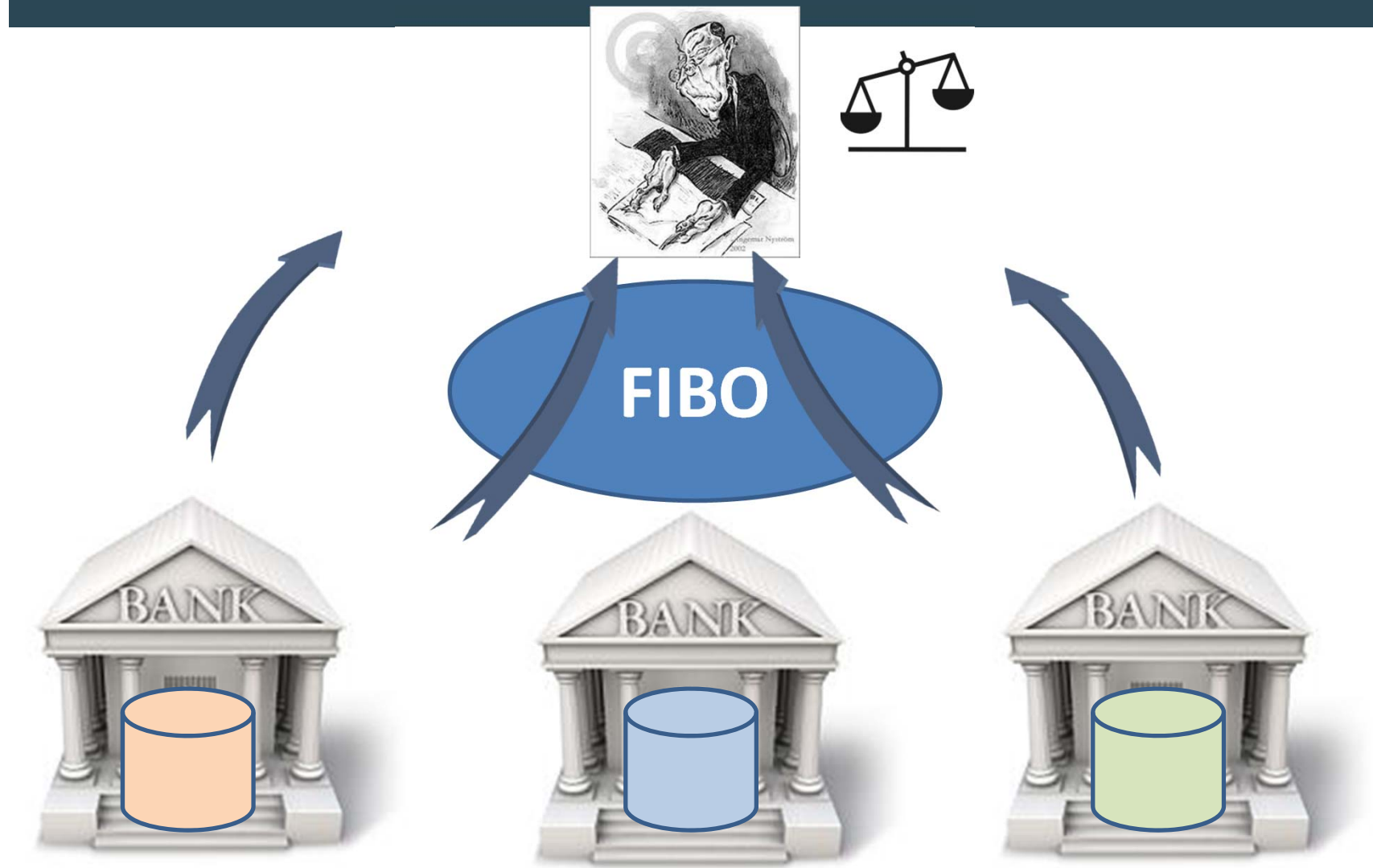
Provenance

Extensions

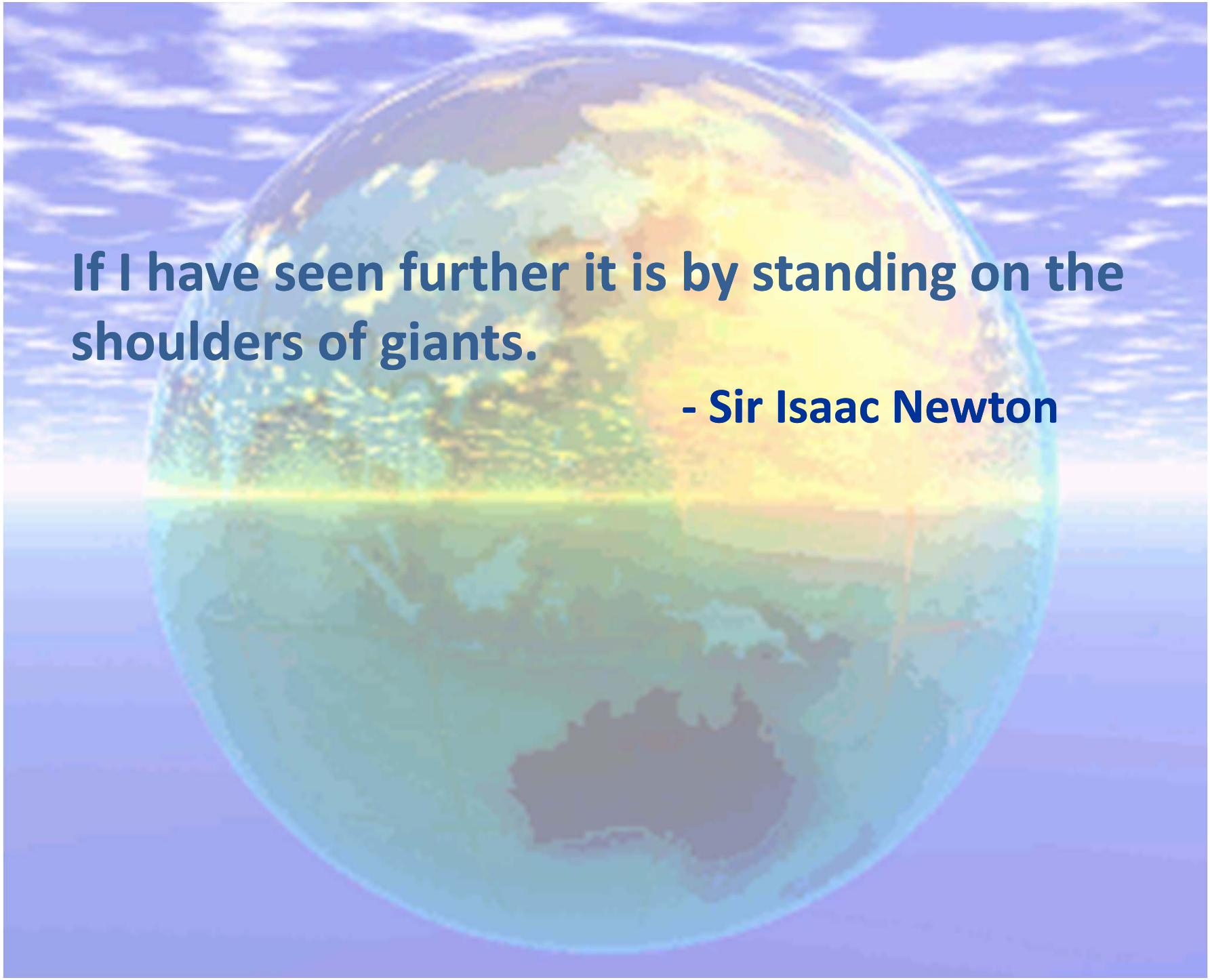
Future-Proofing

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