

YORK OCT 2011

In praise of inconsistency the long tail of small data

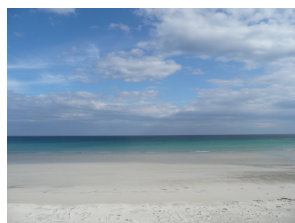


Alan Dix
Talis and Lancaster University



www.hcibook.com/alan/
alandix.com/blog

YORK OCT 2011



Tiree

Tiree Tech Wave
3-7 Nov



Lancaster
University



Talis



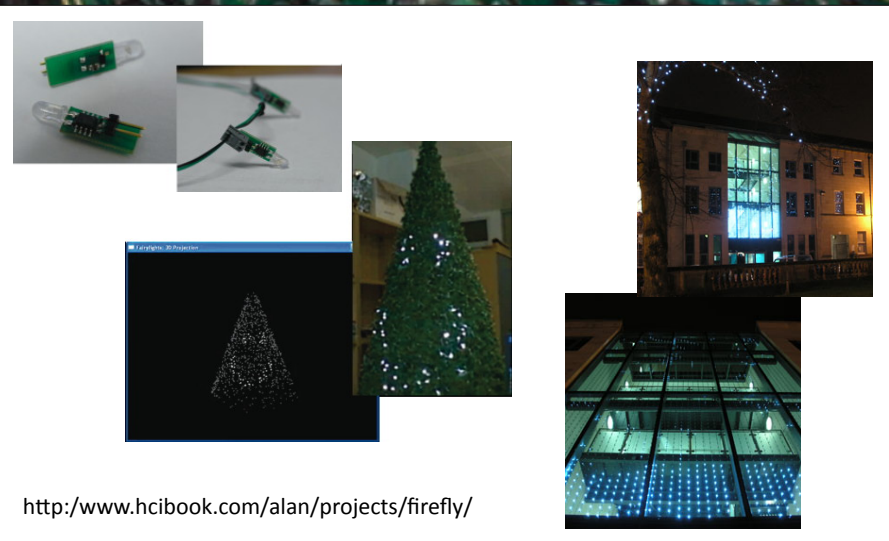
YORK OCT 2011

today I am not talking about ...

- intelligent internet interfaces
- visualisation and sampling
- situated displays, eCampus, small device – large display interactions
- fun and games, virtual crackers, artistic performance, slow time
- physicality and product design
- creativity and Bad Ideas
- modelling dreams and regret



... or even lots of lights



<http://www.hcibook.com/alan/projects/firefly/>

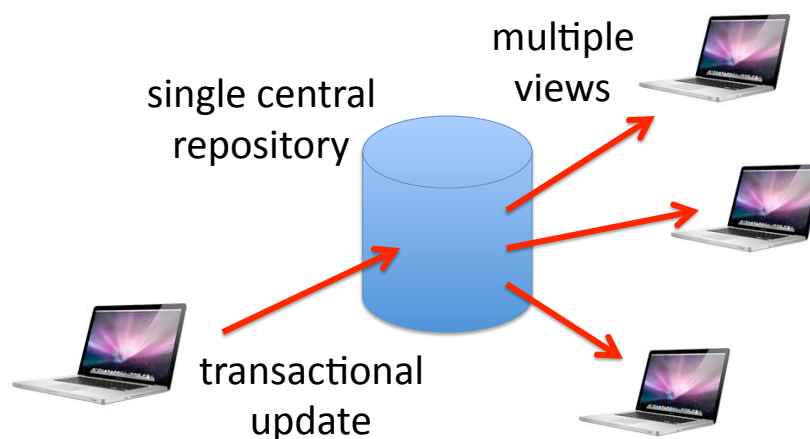
YORK OCT 2011

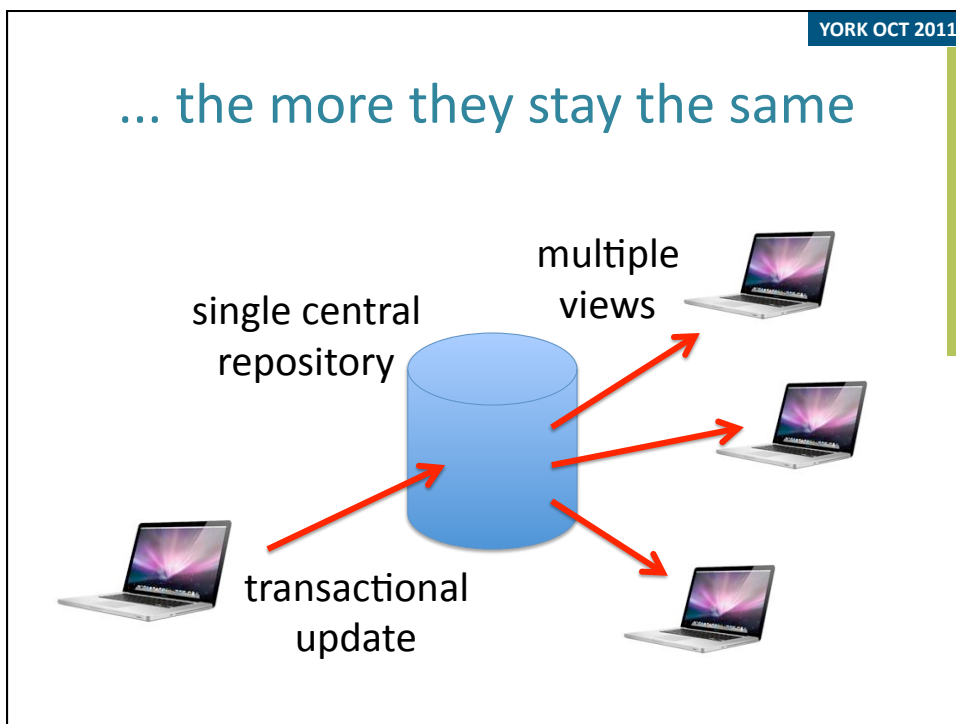
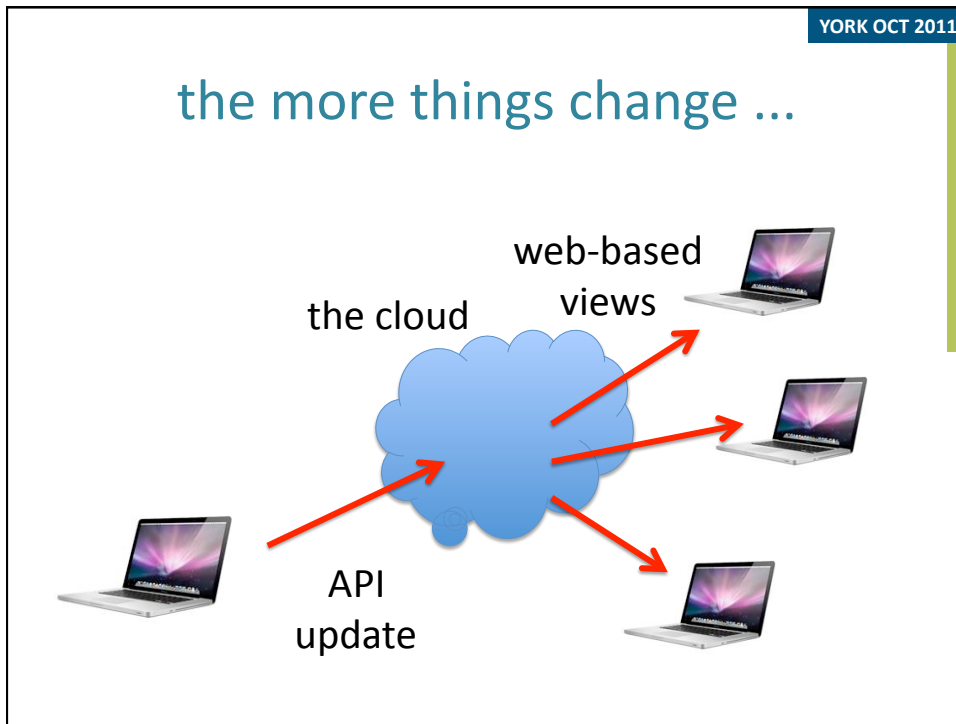
back in the 1980s ... Codd and all that

- in theory:
 - normalisation, atomicity
 - illusion of single use & strong internal consistency
- in practice
 - de-normalise for efficiency
 - maintain consistency through controlled transactions
 - business logic, APIs

YORK OCT 2011

the IS ideal





limits of consistency

consistency not always possible

- distribution and caching
- multi-user update (Alison and Brian)
- view-based updates



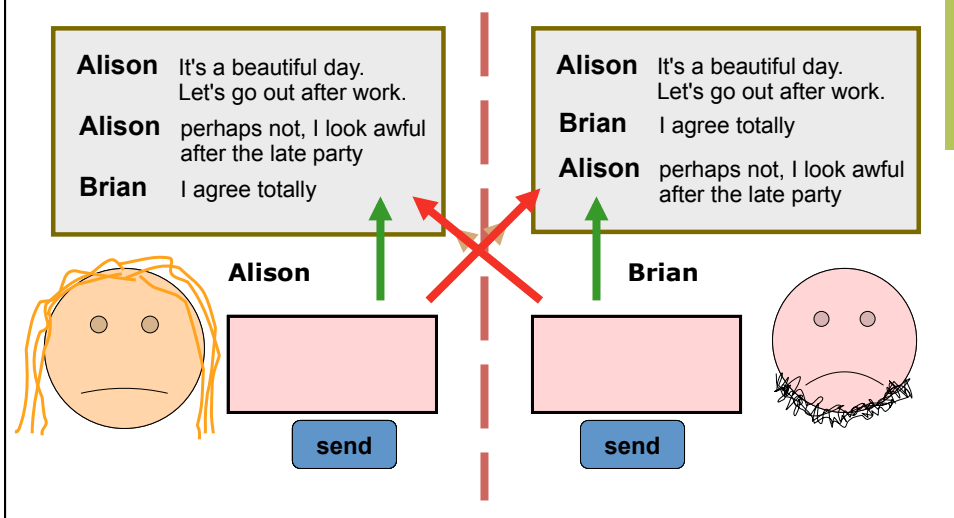
<http://www.vfridge.com/>

Virtual Fridge



<http://www.perryslingsbysystems.com/trenchers.html>

ordering problems (race conditions)



YORK OCT 2011

limits of consistency

consistency not always possible

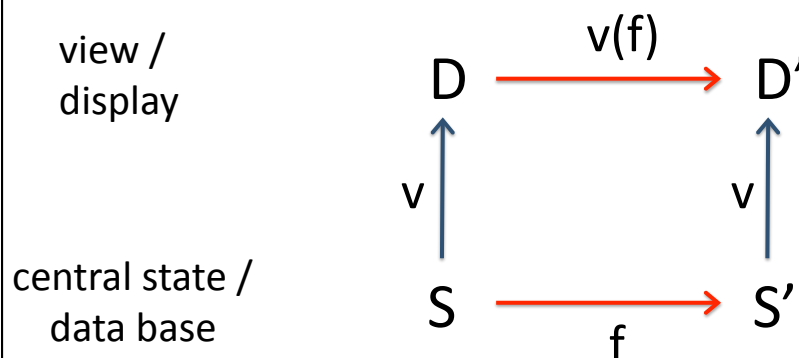
- distribution and caching
- multi-user update (Alison and Brian)
- view-based updates

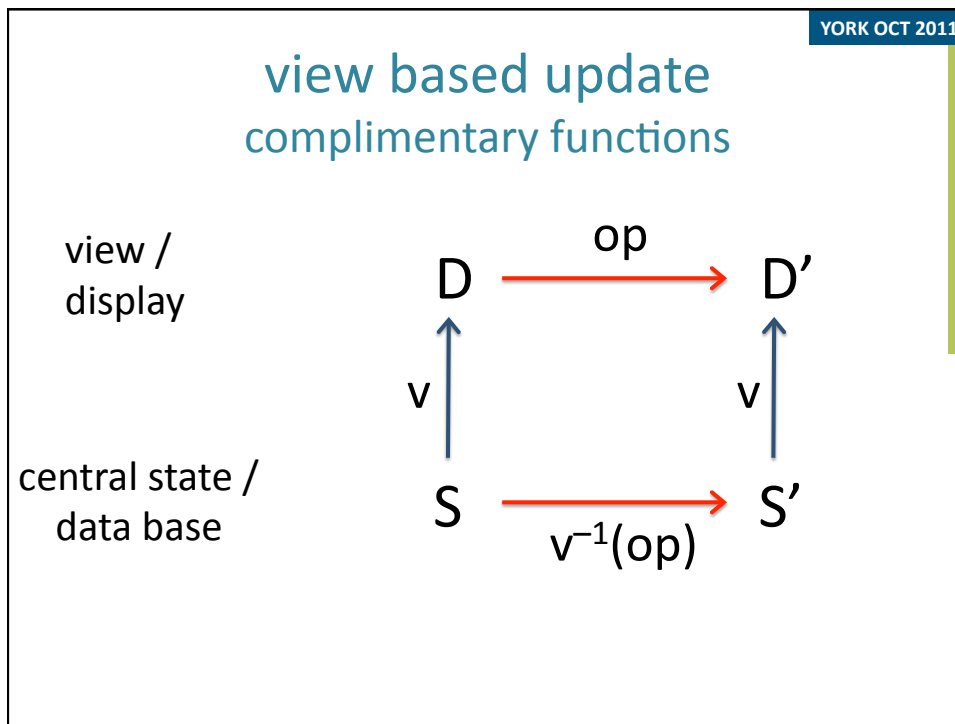


<http://www.perryslingsbysystems.com/trenchers.html>

YORK OCT 2011

view based update complimentary functions





... always

goal is eventual consistency

YORK OCT 2011

YORK OCT 2011

sometime not possible

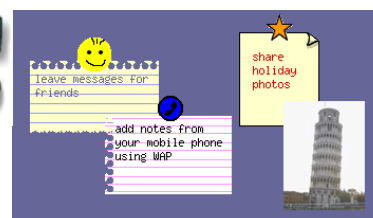
distributed garbage collection

- various algorithms ...
aim to make sure referenced items not lost
- but always storage node can die
- options:
 - prevent loss of referenced item
 - accept loss of referenced item
 - leases or “ref not found” exceptions

YORK OCT 2011

what is consistent?

- conflicting updates
 - long-term transactions
 - synchronisation
- ... and Apple still can't get it right!!

The logo for Virtual Fridge, featuring the words "Virtual" and "Fridge" in a stylized, bubbly font.

<http://www.vfridge.com/>

YORK OCT 2011

internal and external consistency

- the exam board

YORK OCT 2011

is the world consistent anyway?

- departmental lists

YORK OCT 2011

a different approach

do not enforce consistency

but highlight inconsistency

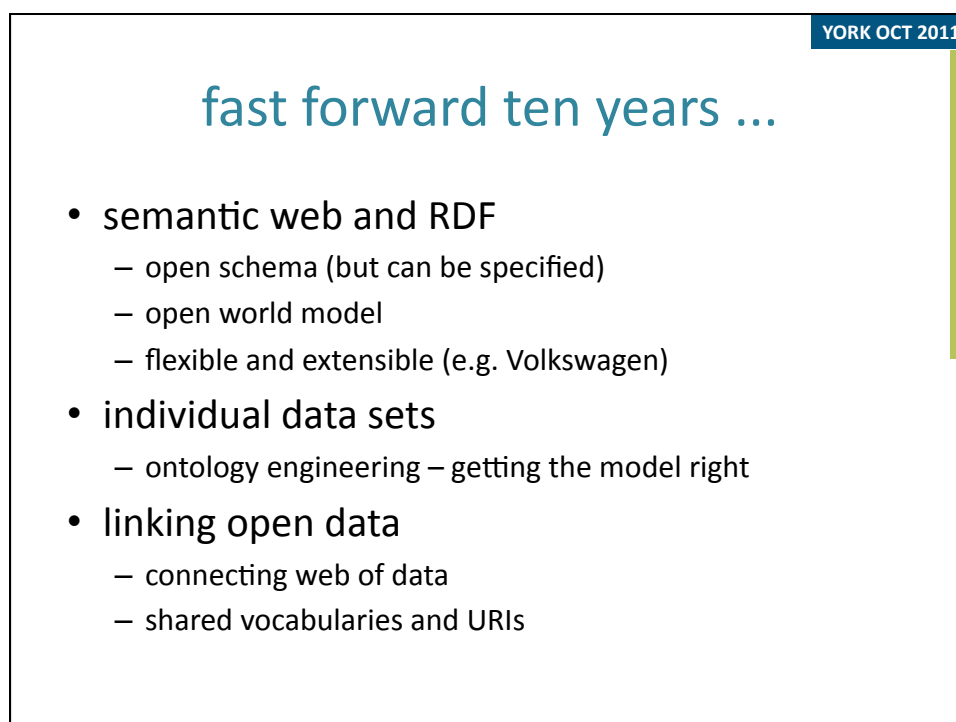
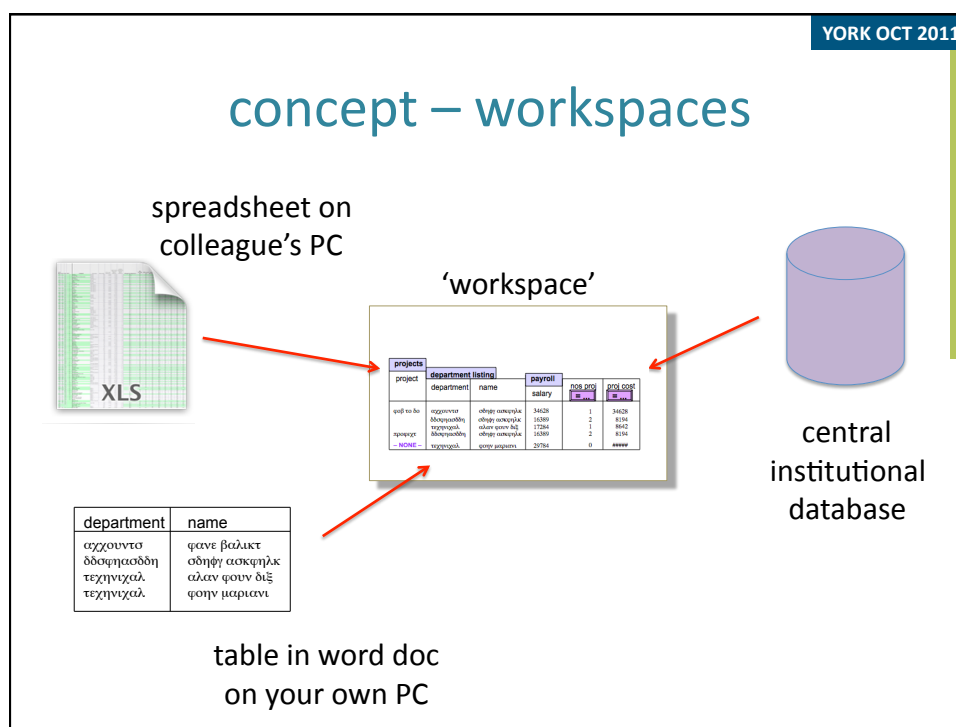
YORK OCT 2011

a different approach

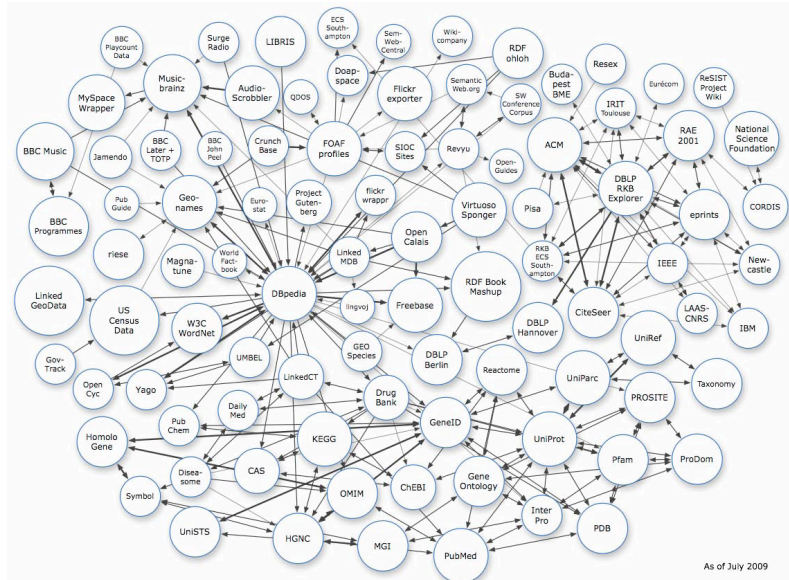
do not enforce consistency

but highlight inconsistency

- instead of views of central data, related yet different sources
- specify connections and automatically check inform of updates, highlight discrepancies but allow divergence



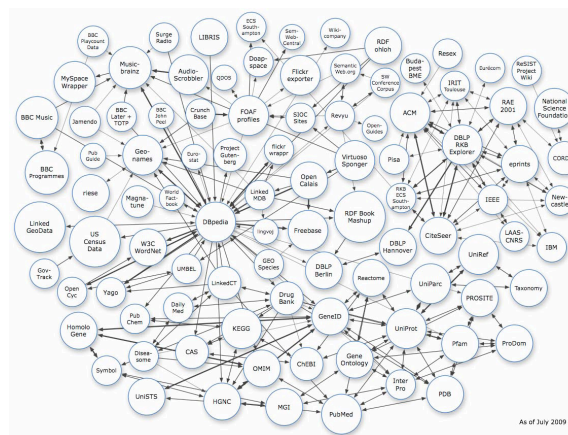
linking open data



linking open data

linking through:

- shared
- dereferencable
- URIs



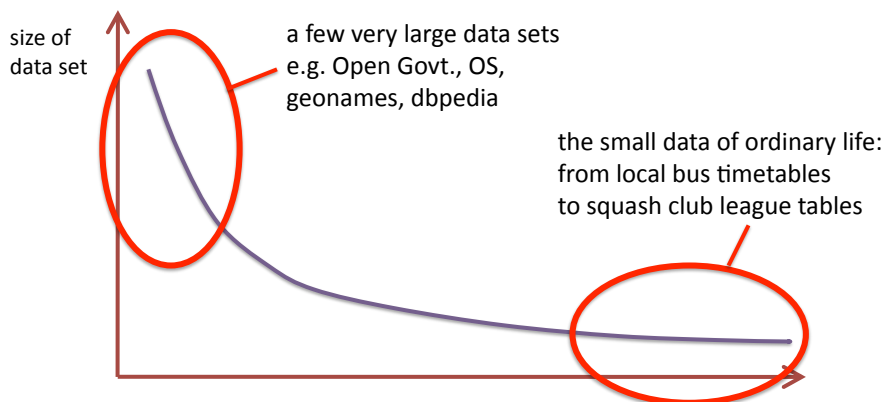
YORK OCT 2011

fast forward ten years ...

- semantic web and RDF
 - open schema (but can be specified)
 - open world model
 - flexible and extensible (e.g. Volkswagen)
 - individual data sets
 - ontology engineering – getting the model right
 - linking open data
 - connecting web of data
 - shared vocabularies and URIs
- sounds familiar?
→

YORK OCT 2011

the long tail



YORK OCT 2011

supporting small data?

FreebaseGoogle fusion tables
beta

google-refine

KASABI BETA

- mostly for 'middle' sized data

YORK OCT 2011

really small?

personal, but also Govt.

often tables

describe semantics rather than 'converting'

- explicit – simple description
- implicit – semantics through interaction

YORK OCT 2011

explicit – 3 levels

- the table as it is:
 - there are 5 columns
 - col 1 is called “name”, col 2 is ‘population
- internal semantics of table (in its dataset)
 - each row is the properties of a country entity defined by the ‘name’ column
- external linkage to standard data/vocab
 - rules + exceptions
 - country is ‘sameAs’ geoname country by matching name
except ‘Wales’ is geoname administrative region ...

YORK OCT 2011

implicit

action is specification:

- view a table and give it a name
- link items/columns from different data sources
- perform calculation

semantics are emergent through use

YORK OCT 2011

SO ...

long history of consistency
... but not always possible or desirable

do not **enforce** consistency
but highlight inconsistency

exploit the long tail of small data

YORK OCT 2011

plus ...

come to
Tiree Tech Wave
3-7 Nov 2011