

Cloudbuilding

**semantic imprecision
in social science data
(*and how to live with it*)**

Phil Edwards

**Second International Conference on e-Social Science
June 2006**

Lies, damned lies...

- Clear numbers, fuzzy data
 - 10% of *exactly who said exactly what?*
 - Is my 10% the same as your 10%?
 - (Clue: Probably not)
- Clear data, cloudy concepts
 - *Imprecise*
 - *Contested*
 - *Mutable*

Prospects: cloudy

- Imprecise concepts
 - Vague definitions

Living with vagueness

- Sometimes resolvable
 - gender/sex; age-based categories
- Sometimes there is no ‘lower level’
 - subjective reporting
- Sometimes vagueness is built in
 - ‘Anti-social behaviour’

Prospects: cloudy

- Imprecise concepts
 - Vague definitions
- Contested concepts
 - Multiple definitions

Living with contestedness

- Sometimes resolvable
 - at the risk of displeasing everybody
- Sometimes there is no middle ground
 - drugs and harm
- Sometimes contestedness is the story
 - ‘Anti-social behaviour’ (again)

Prospects: cloudy

- Imprecise concepts
 - Vague definitions
- Contested concepts
 - Multiple definitions
- Mutable concepts
 - Changing definitions

Living with mutability

- Seldom resolvable
 - retroactive change tends to distort
- The time-lag problem: forward is as bad as back
 - ‘Head of Household’
- Ultimately, change is always the story
 - Ideas change, but do they progress?
 - No one right answer, not even ours

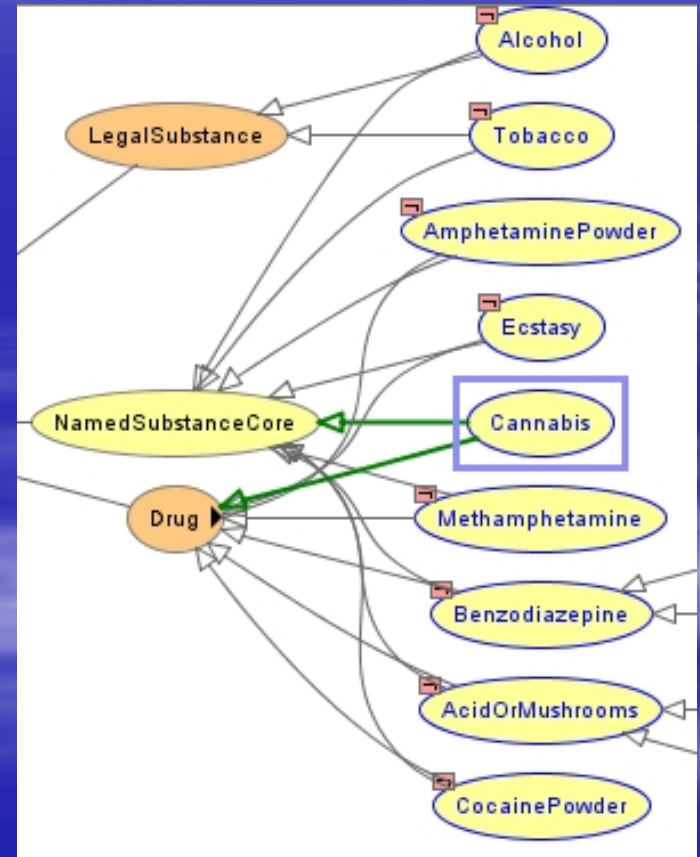
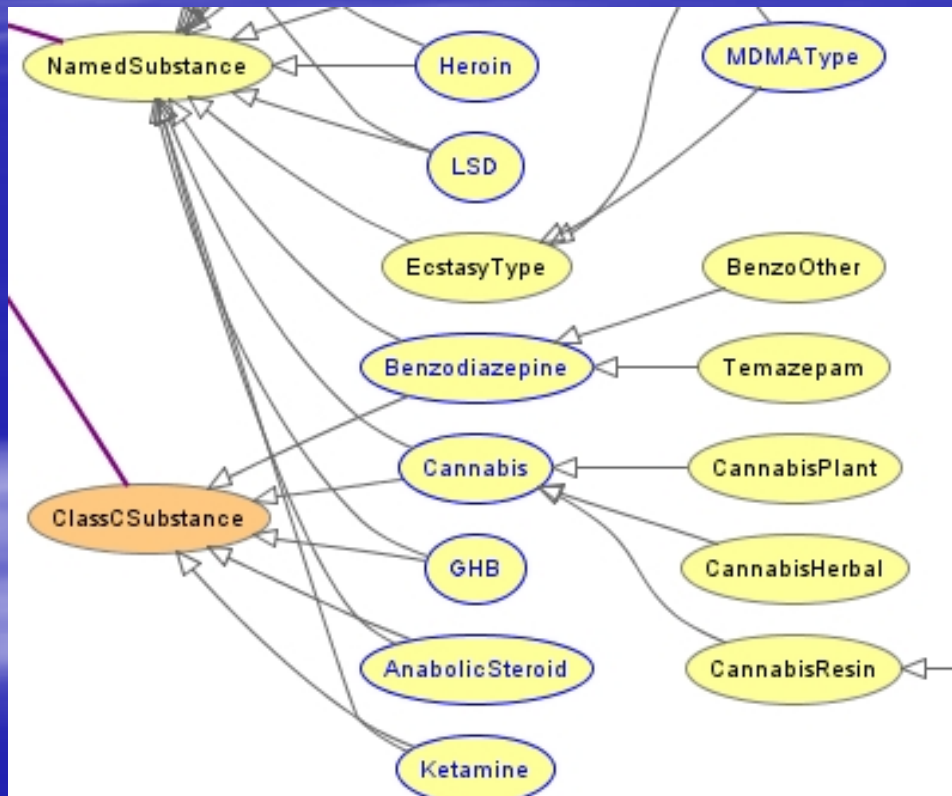
Cloudbusting

- *“The ones who drove within half an hour of smoking cannabis **must** have been under the influence...”*
- *“The British Crime Survey asks about sniffing glue and Mixmag asks about alcohol and tobacco, but those aren’t really **drugs**...”*
- *“Class B drug use dropped in 2005, but that’s only because of the reclassification so it didn’t **really**...”*
- You don’t want to do this!
 - ...that is, researchers may want to do this, but you don’t want to do it for them

The trees and the forest

- Every source has its own ontology
 - ‘In vivo’ concepts
- No one ontology is correct
- No two ontologies are identical
 - ...but*
- In vivo ontologies can be described logically
 - as can the correspondences between them

Two trees (or two clouds)



Cloud logic: the uses of OWL

- Conceptual hierarchy
 - Cannabis is a type of Drug
- Horizontal links (properties)
 - Cannabis Use [is use of] Cannabis
- Multiple hierarchies
 - Cannabis is a type of Class C Substance
- Multiple ontologies
 - Cannabis [corresponds to] Dope

Forest management

- Ontology-to-ontology correspondences
 - N:M
 - N:1
- Central 'organising concept' ontology
 - Not the 'correct' view
 - ...more of a lowest common denominator

Frequently-asked questions

- What are we actually building?
- Can we make something useful out of it?
- Who cares what we think anyway?
(ontologically speaking)
- Watch this space!
- “If you go to an acupuncturist with toothache he sticks a needle instead into your thigh. Do you know why he does that, Mrs Rawlinson? No, neither do I, Mrs Rawlinson, but I intend to find out.”
 - Douglas Adams, *Dirk Gently’s Holistic Detective Agency*

Summary

- Social science data is built on inherently fuzzy concepts
 - Imprecise, contested, mutable
- Imposing a single unifying structure is not not satisfactory
 - Nor is doing without structure
- Wanted: logical documentation of illogical illogical concepts
 - A forest full of OWL(s)

Any more questions?

