



# Beancounter



*You stalking you*

Libby Miller, Vicky Buser, Yves Raimond, Dan Brickley  
NoTube

What is Beancounter?

# There's information out there about what you've been up to



enjoyed watching astronauts playing in space on bbc2 last night <http://bit.ly/PVw7O>

1:31 AM Jun 22nd from web



vickybuser watched a programme on BBC iPlayer

1 week ago



**Henry VIII: Patron or Plunderer?**

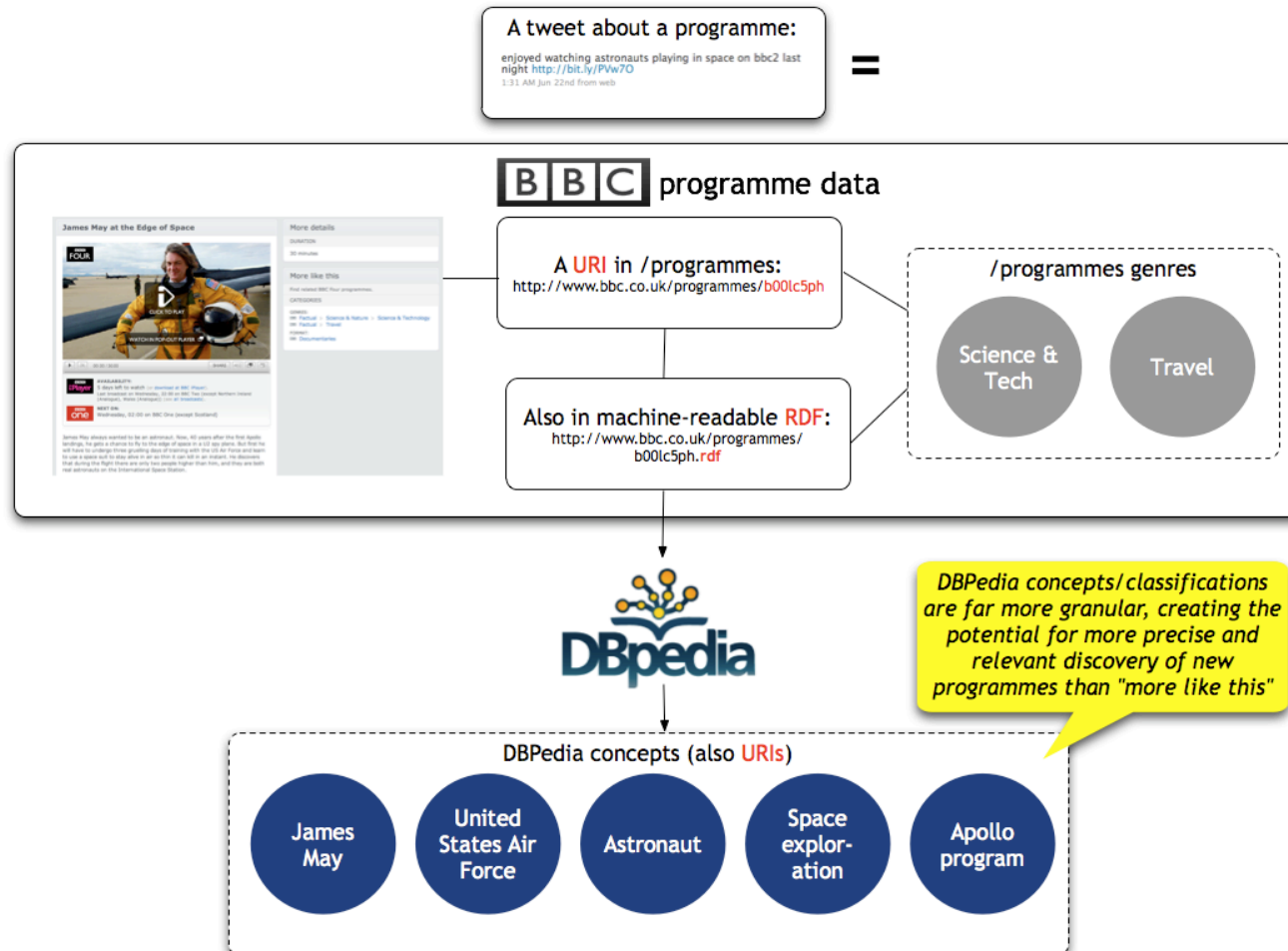
Episode 1



Vicky Buser Watching Wimbledon semi-finals

20 seconds ago · Comment · Like

# That can be enriched with common concepts



Over time these concepts build up to determine a user's interests

# And can be counted

Concepts (What categories of programme do you like?)

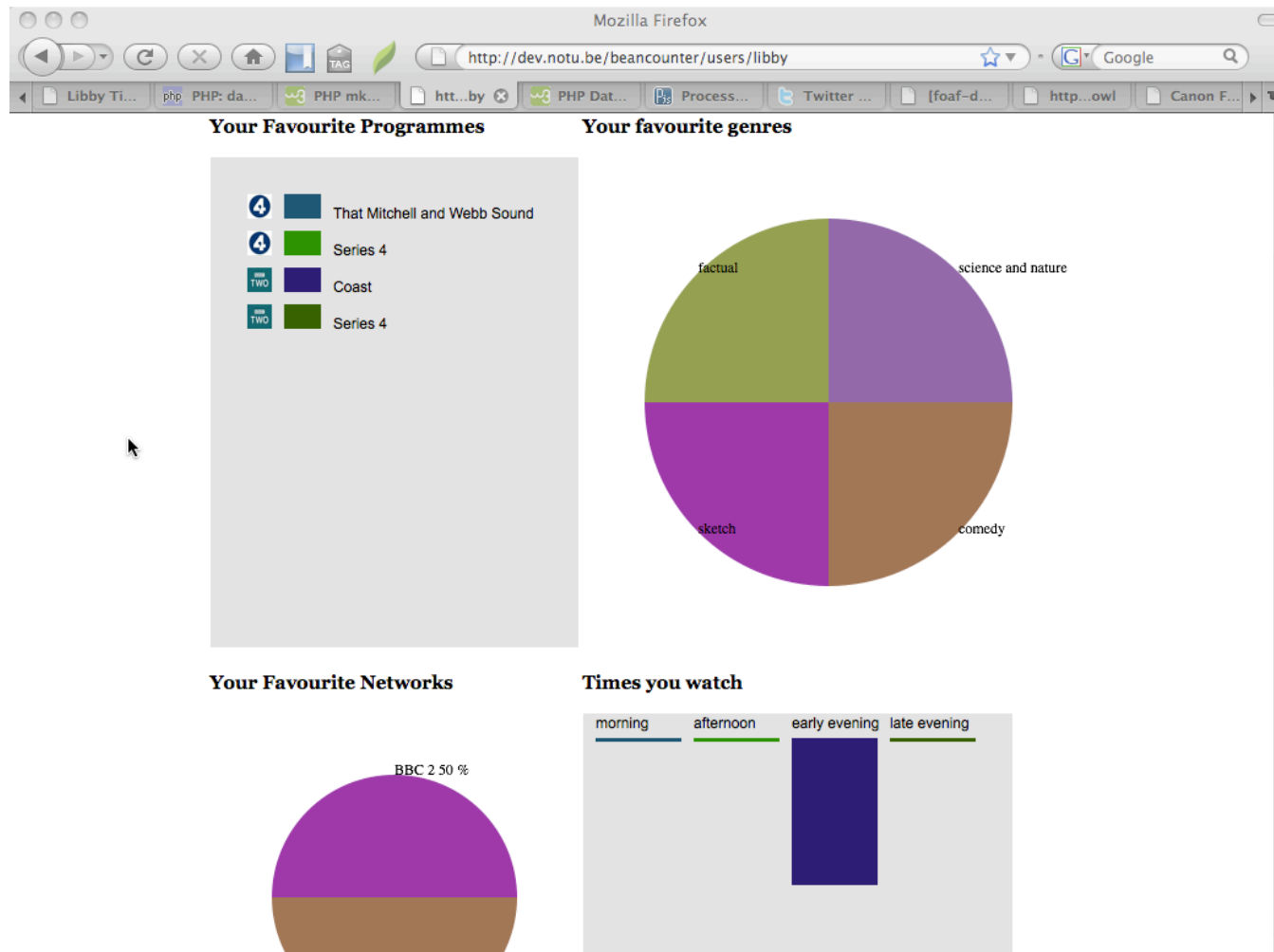
Series (What series have you watched the most?)

Location / context (Where do you usually watch TV?)

Time periods (When do you usually watch?)

Compare with last week, last month, all time, yesterday

# Beancounter can profile you



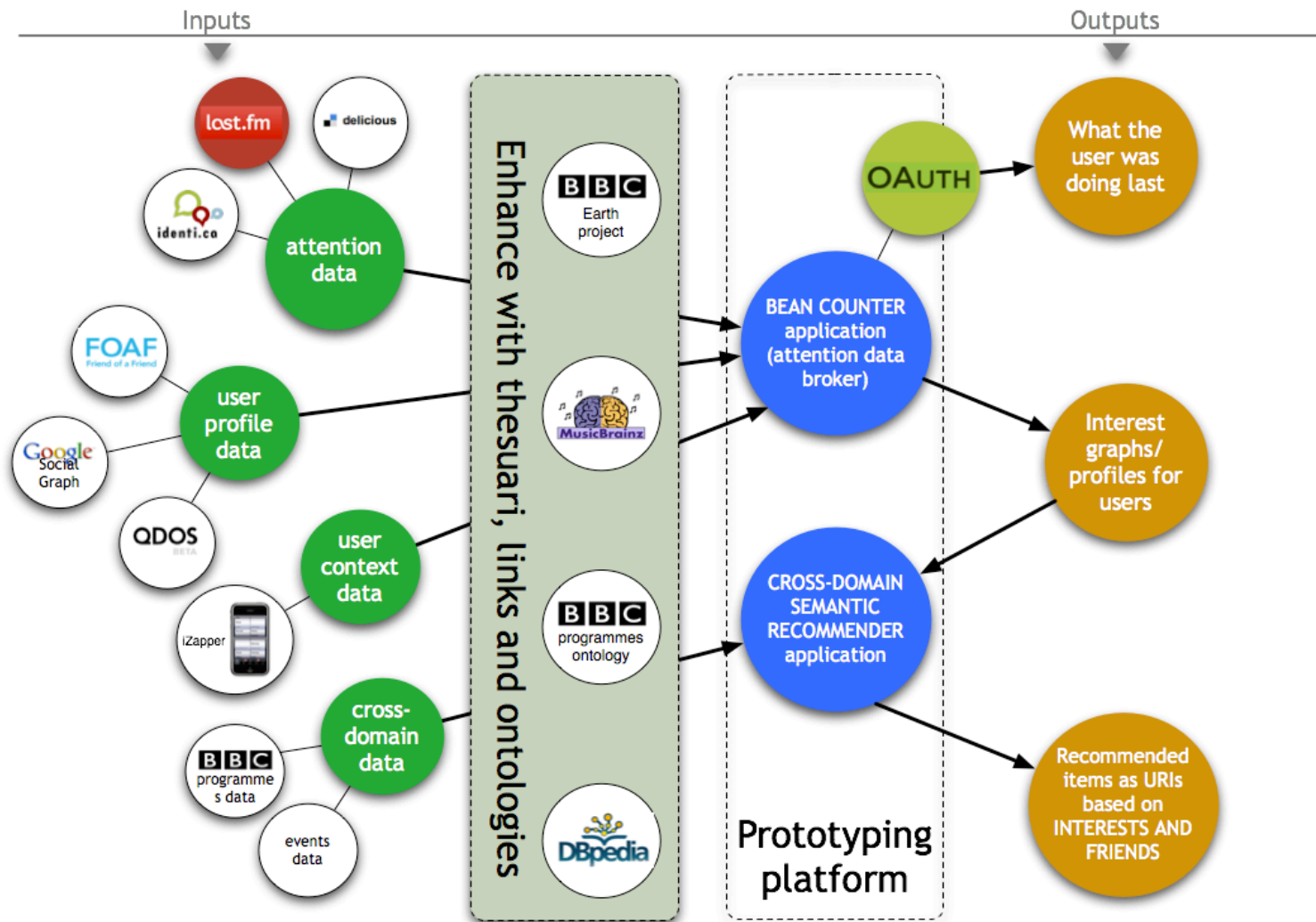
And more importantly

Output a *machine-processable* profile

Beancounter is about dumb accounting - and exporting an aggregated picture for other applications to query via an API

Why?

# To use the data for TV recommendations

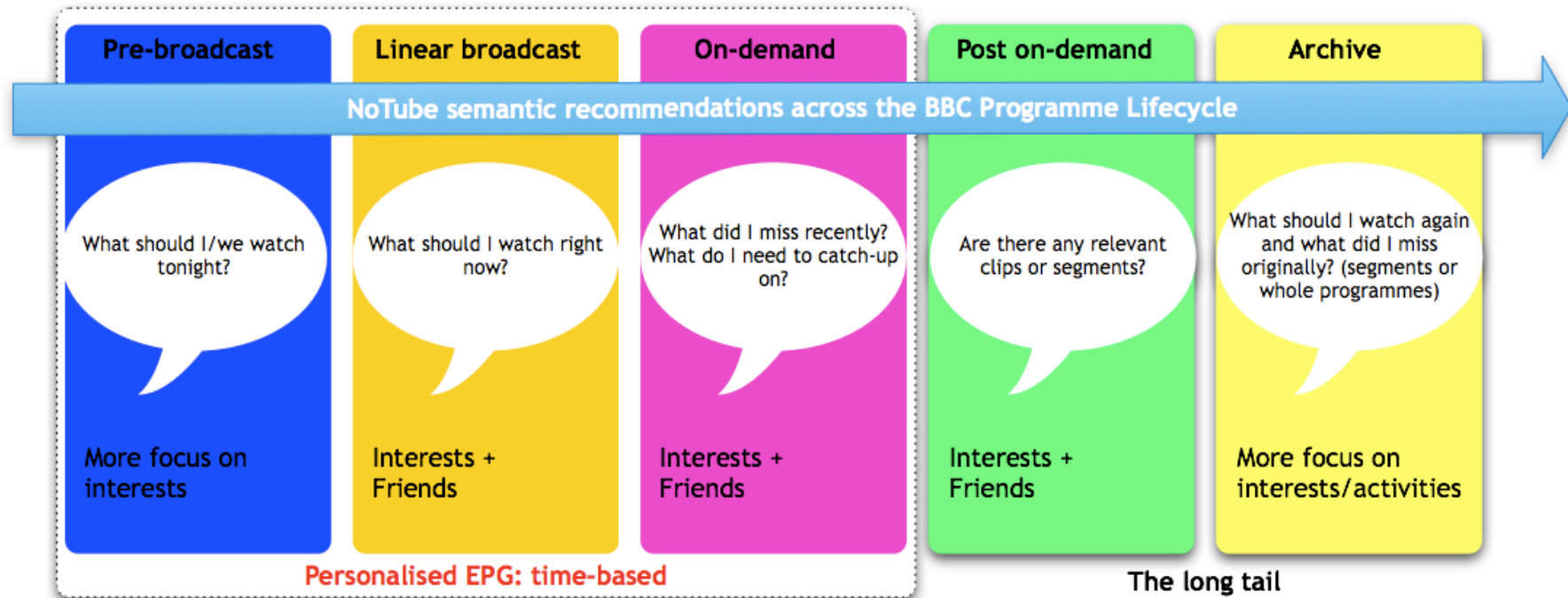


“The bane of my existence...”

“... is doing things that I know the computer could do for me.” (Dan Connolly)

The web knows what you've been watching and doing – we should be able to use this data to suggest things for you to watch

We want to find recommendations across the programmes lifecycle (including archival content)



## Or, if that's too hard...trivial recommendations

Combine 'series I like' or 'series my friends like'  
with

Broadcast and on-demand availability data  
(‘when is the Wire on next’?)

To say “the Wire is on tomorrow at 23:20 – I  
think you might be interested”

Some of those ‘friends’ could be professional  
reviewers (e.g. Charlie Brooker, Guardian  
Guide)

# Data Enrichment

# Enhancing with links

- Sometimes we will need to generate links because there's no URL in the data
- Perhaps using time as a basis 'I'm watching Channel 4' and matching the time of the activity to the schedule, to get a URL for the programme
- In the future we might expect TVs that output this data directly

# Enhancing with Linked Data

- ‘Pre-enhancement’ - things that you know you’ll need later such as genre names
  - On-demand enhancement – following links to related data or querying a sparql store
  - Generating DBpedia links where none exist
- > Enabling us to use common identifiers for someone’s interests – and match those to watchable programmes

# Linked Data issues

- Some patterns are only complete when you have multiple files crawled
- Data can be gappy
- Not sure what you'll be getting
- Hard to test

Easy(ish) with BBC /programmes data – get information about people, series, genres, showings; known schema

# Privacy and user control

# User control

- Download and run it on your own machine
- Choose which accounts you want to add
- Pick time periods to control
- Write to the web, keep the data private, or make suggestions just to you

# Cross-querying user data

- How many people watched Torchwood, and when?
- What else did they watch?
- The Tesco clubcard equivalent point of view

How might this impact privacy?

# Privacy issues can be fraught

Beancounter puts the user  
in control of what is  
stored about them  
(mostly)

Some things people might  
not to want to know  
about themselves

And definitely not want  
others to know

## Bob's Bean Counter Profile

Machine-readable versions: [FOAF](#) | [APML](#)

Bean Counter has analysed your  
watching and listening activity to  
generate a profile of your  
interests and friends:

### Your interests

**cars** [why?](#) **cookery** [why?](#) **food** [why?](#)

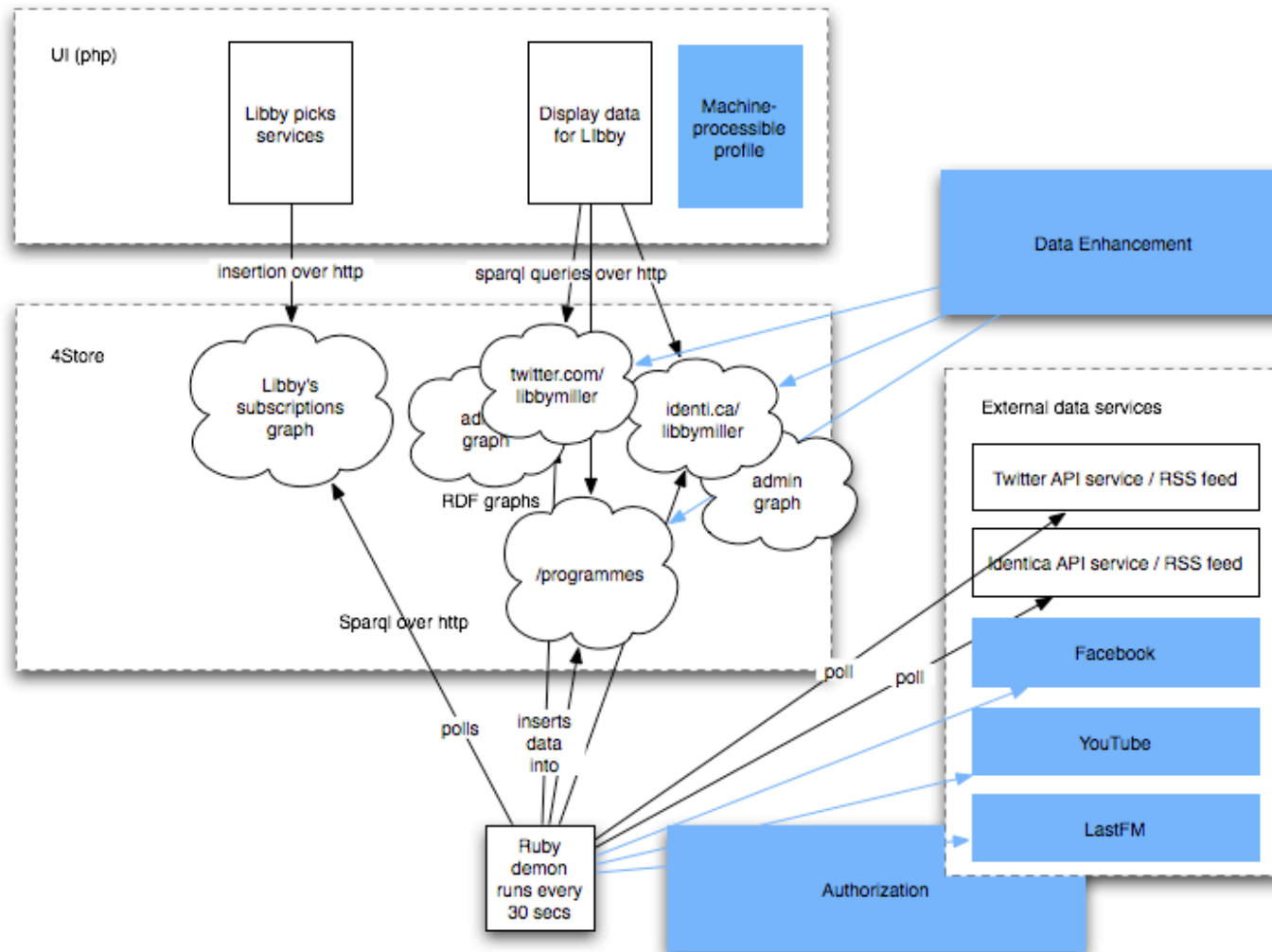
**jazz** [why?](#) **malaysia** [why?](#)

**ornithology** [why?](#)

**porn** [why?](#) **tennis** [why?](#)

Currently

# A simple working prototype for BBC programmes only



Next

# Interesting problems to address

- Can we effectively distinguish mentions and ‘watchings’?
- Can we use OAuth to create adaptors to private data from different systems?
- Can we create and describe acceptable privacy policies?
- Can we generalise to different channels and providers with less detailed data?
- Can we make interesting recommendations using the profiles?
- Can we get anyone to use it? What are the incentives?
- What else can we count? Could it help with recommendations?

# Thanks to:

- Notube project members, especially Asemantics and Vrije Universiteit Amsterdam
- Tom Scott and Dan Brickley for the Beancounter concept (and to Dan for the logo)
- Radio Pop for inspiration
- Damian Steer for helpful comments

<http://blog.notu.be/>

libby.miller@bbc.co.uk