

using the web of data the user

Alan Dix

InfoLab21, Lancaster University, UK

www.hcibook.com alan/papers/WOD-PD-2008/

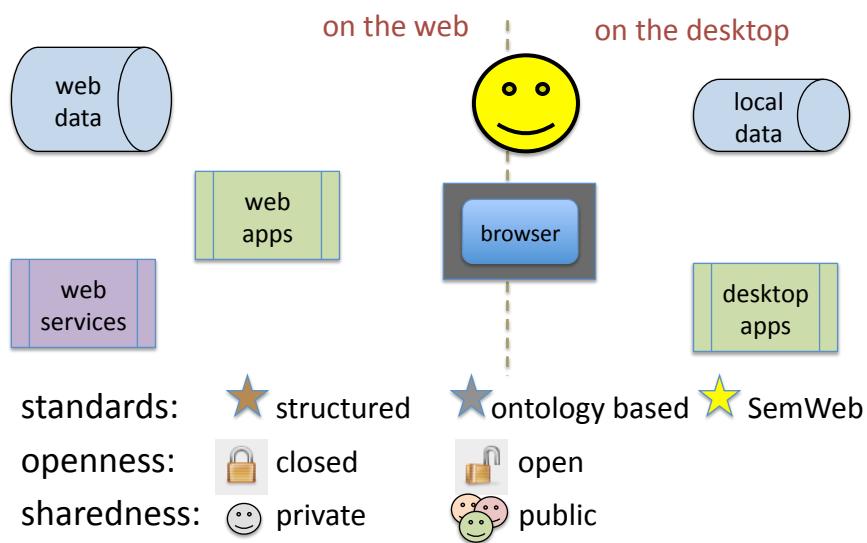
www.alandix.com/blog

in the end
it's not about data
it's about use

user and the web of data two sides

- generating the web of data
 - accessing the web of data
- and maybe linked through user interaction

the ecology of the web



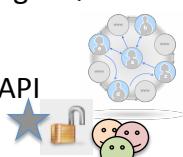
generating the web of data

top down
bottom up

top down

- from legacy sources
 - taking existing online data
 - from structured data, e.g. SQL -> RDF
 - web scraping
 - e.g. Google Social Graph API
- from scratch
 - building new RDF stores for new data
 - user contributed

? reverse ?
wrappers



other talks

top down – user creation

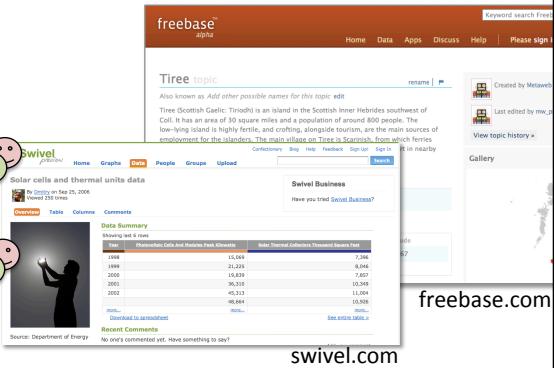
creating public data

- Freebase 

- ontology based

- Swivel 

- tables & graphs



swivel.com

- maybe sharing more private data

- Google spreadsheets 

- linking desktop ontologies



bottom up

- implicit

- folksonomies => structured data *



- explicit

- personal ontologies



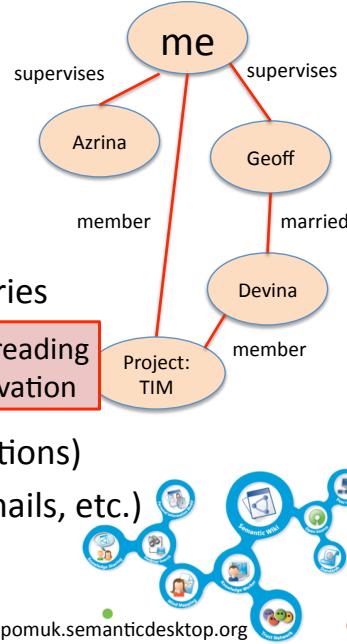
- automatic

- semantic desktop – PIM mining

* A. Dix, S. Levialdi and A. Malizia (2006). Semantic Halo for Collaboration Tagging Systems. www.hcibook.com/alan/papers/SemanticHalo-2006/

personal ontologies

- use ‘general’ categories:
 - post code, name, place
- linking to personal ontology
 - users own entities and categories
 - egocentric & ideocentric
- how to build?
 - by hand (during useful interactions)
 - automatically (mining files, emails, etc.)
 - e.g. Gnowsis, Neomuk and other semantic desktop projects



www.hcibook.com/alan/projects/TIM/

gnowsis.org

nepomuk.semanticdesktop.org

user data collection

- private
 - personal ontologies
 - semantic desktop
- public
 - freebase
 - folksonomies

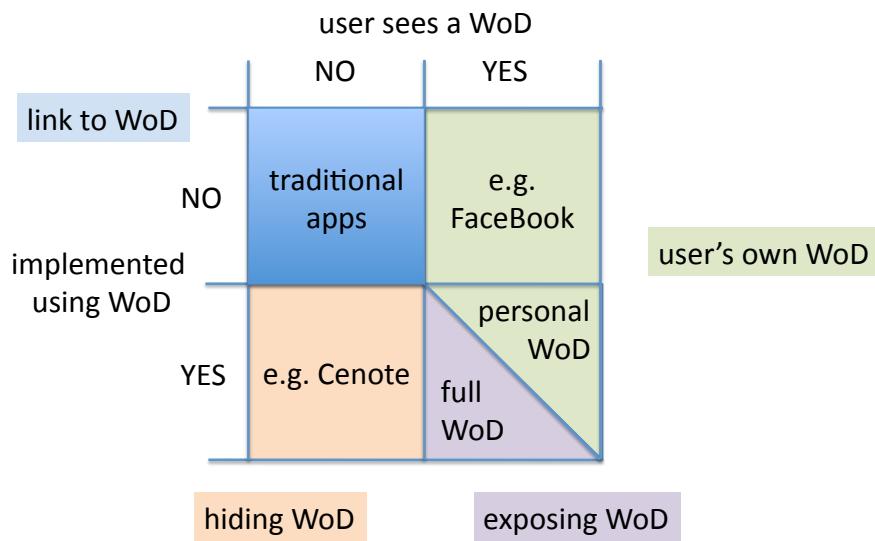
critical mass



accessing the web of data

exposing the WoD hiding the WoD
users own WoD linking to the WoD

visualising web of data

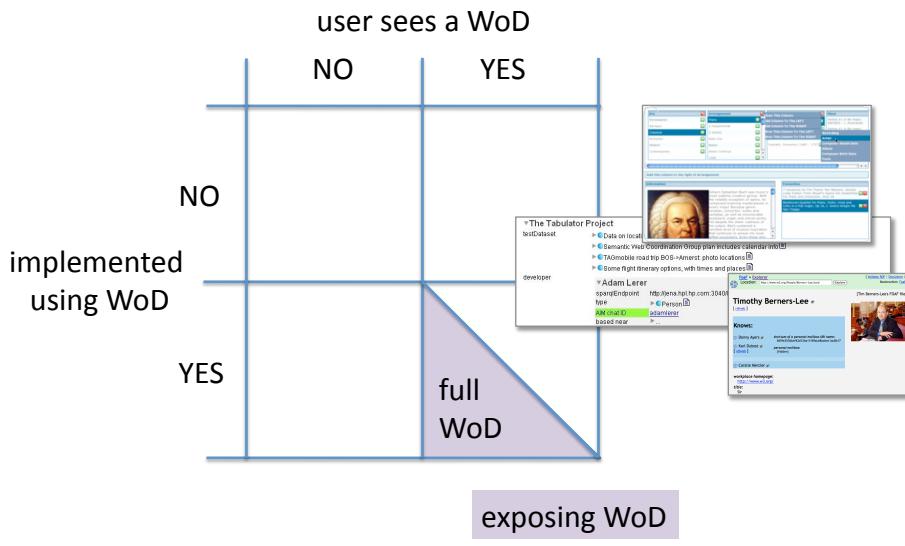


hiding vs. exposing

similar issues for RDMS

- exposing
 - MySQLadmin
 - mainly for administrators/developers
- hiding
 - most DB apps
 - for end users

exposing the web of data



exposing the web of data tabulator

- outliner style + query list views + editing

The screenshot shows the Tabulator interface with the following details:

- Session:** This Session
- brokenURI:** [Test] from a FOAF document to the person.
- n32:**
- http://www.ivan-herman.net/Ivan_Herman:**
- featureAndBug:** featureAndBug, redirection (303 and Other)
- Ability to add an inverse property:**
 - comment:** For example, one should be able to add a subclass to a class, such as Feature or bug. In the outline mode, when editing, we are prompted for a list of properties .. but they are all forward. We can't do a backward (inverse) link.
 - name:** Ability to add an inverse property.
 - Address:** n33
- is feature/bug:** The Tabulator Project
- of:**
 - TAMI Justification browser issues
 - Not pointed to by main page
 - Hypothesis management
 - Track whether data was obtained using credentials
 - Editing of literals doesn't work for some reason
 - metadata of a web page
 - rdfs:seeAlso failure??

<http://dig.csail.mit.edu/2005/ajar/release/tabulator/0.8/tab.html>

exposing the web of data tabulator

- outliner style + query list views + editing

The Tabulator Project

testDataset

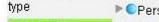
- ▶ Data on location of libraries mostly in the UK 
- ▶ Semantic Web Coordination Group plan includes calendar info 
- ▶ TAGmobile road trip BOS->Amerst: photo locations 
- ▶ Some flight itinerary options, with times and places 

developer

▼ Adam Lerer

sparqlEndpoint <http://jena.hpl.hp.com:3040/backstage>

type  Person 

AIM chat ID  adamlerer 

based near 



The screenshot shows the Tabulator Project interface. At the top, there's a toolbar with buttons for 'View and Save Current Query', 'Save Current Query', 'Help About', 'File type...', and 'Export'. Below the toolbar, there's a navigation bar with tabs: 'Table', 'Map', 'Calendar', 'SPARQL', and 'Debug'. A dropdown menu 'Select queries to display' is open, showing 'Query #1' which is selected (indicated by a green circle). The main area contains two tables. The first table, titled 'The Tabulator Project developer', has two rows: 'Adam Lerer' and 'David Sheets'. The second table, titled 'The Tabulator Project developer AIM chat ID', also has two rows: 'adamlerer' and 'silentOpen'.

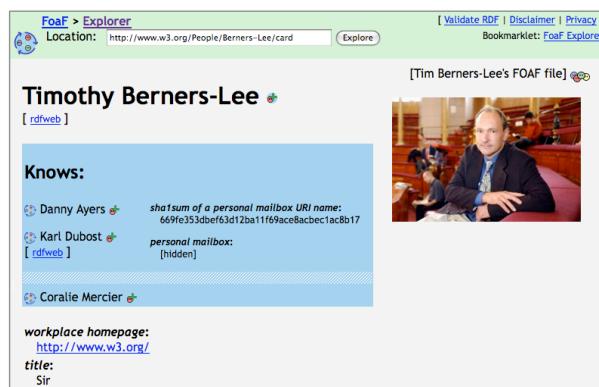
<http://dig.csail.mit.edu/2005/ajar/release/tabulator/0.8/tab.html>

exposing the web of data graph ...

- visualising graph ... example

exposing the web of data foaf explorer

- drill-down / link style
- part
bespoke

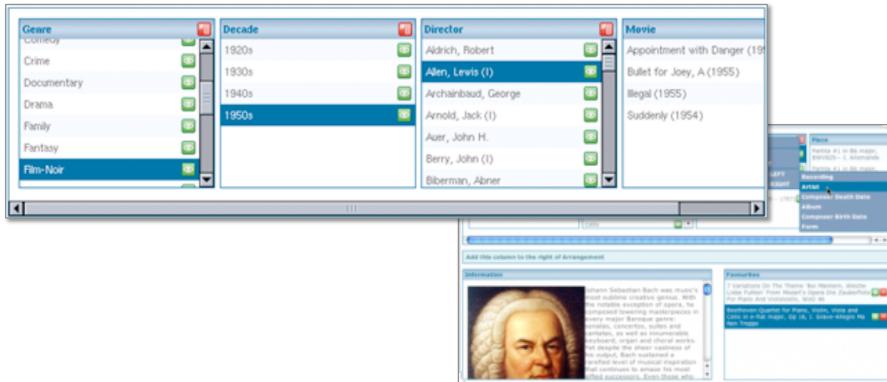


The screenshot shows a web page titled "Timothy Berners-Lee". At the top right are links for "Validate RDF", "Disclaimer", and "Privacy". Below that is a "Bookmarklet: Foaf Explorer" button. The main content area has a blue header "Knows:" containing three entries: "Danny Ayers", "Karl Dubost", and "Coralie Mercier". To the right of the header is a small photo of Tim Berners-Lee sitting at a desk. Below the header, there is a section for "workplace homepage" with a link to "http://www.w3.org/" and a "title" entry "Sir".

<http://xml.mfd-consult.dk/foaf/explorer/>

exposing the web of data mspace

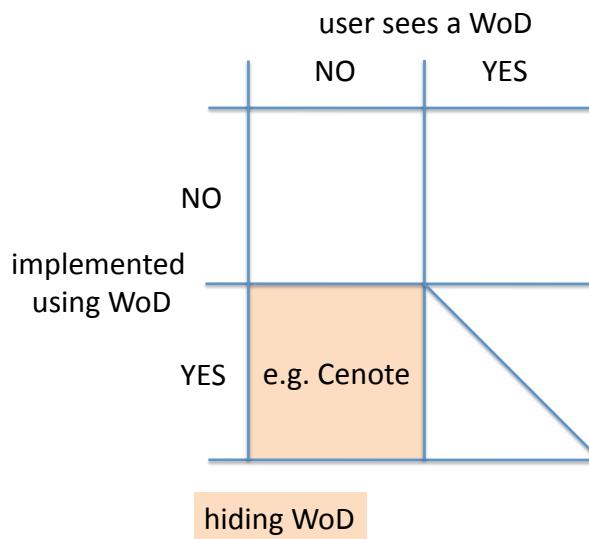
- faceted browsing



The screenshot displays a faceted browsing interface for movies. On the left, there are four vertical facets: "Genre" (Comedy, Crime, Documentary, Drama, Family, Fantasy, Film-Noir), "Decade" (1920s, 1930s, 1940s, 1950s), "Director" (Aldrich, Robert; Allen, Lewis (I); Archibald, George; Arnold, Jack (I); Auer, John H.; Berry, John (I); Biberman, Abner), and "Movie" (Appointment with Danger (1937), Bullet for Joey, A (1955), Illegal (1955), Suddenly (1954)). In the center, there is a large image of Alfred Hitchcock. Below the facets, there are sections for "Information" (a bio about Hitchcock) and "Footnotes" (links to external sources).

<http://www.iam.ecs.soton.ac.uk/projects/292.html>
<http://mspace.fm/>

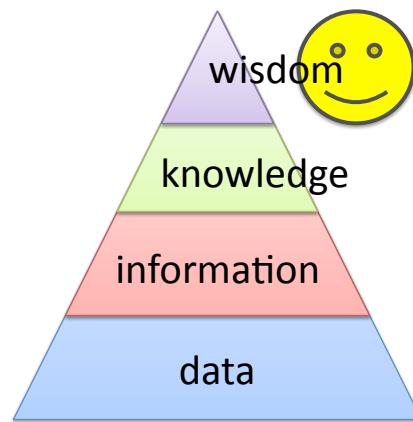
hiding the web of data



hiding the web of data

users don't want data

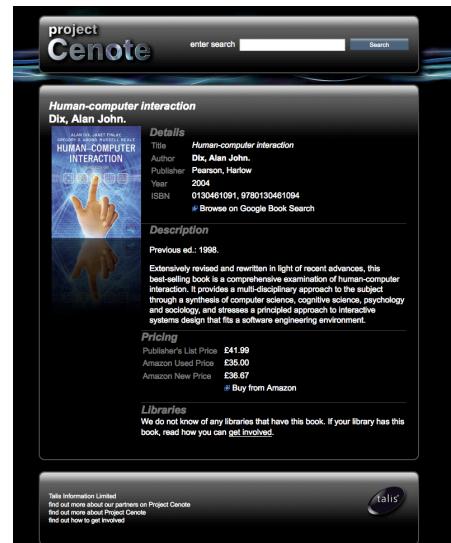
they want to do things



hiding the web of data

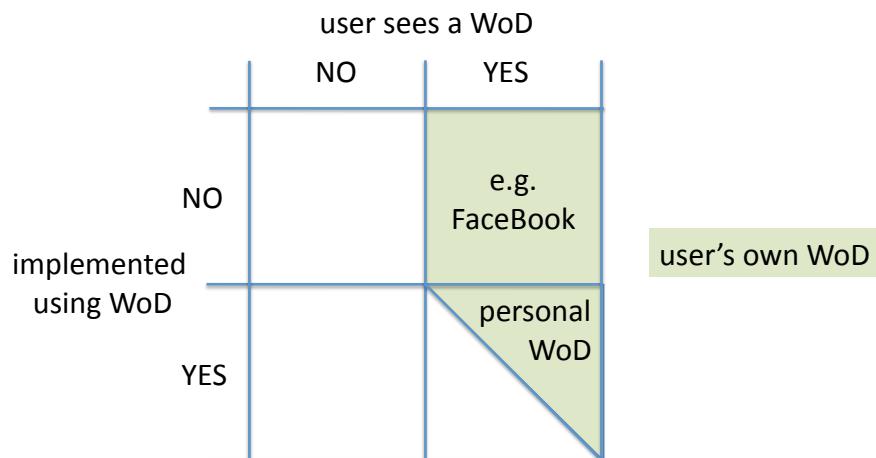
Cenote 

- dedicated view for books
- draws together several linked data sources
- semantic web structure not apparent to user



<http://cenote.talis.com/>

user's own web of data

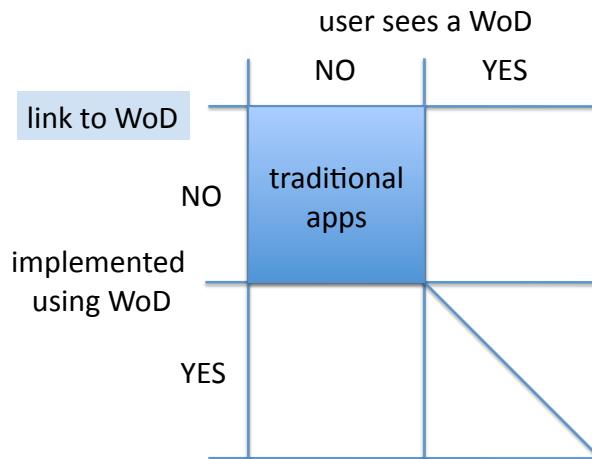


user's own web of data

- FaceBook 
 - apps add relationships
 - but insulated from one another
 - NOT RDF
- Developer API
 - add own relationships



linking to the web of data



linking to the web of data

some rich semantic web services ...

... but most web pages and apps are not!

can we make them loci of entry to WoD?

- at point of creation:
 - microformats, zLinks
- at point of use:
 - data detectors and intelligent stuff

linking at point of creation

offer alternative RDF or XML formats

or use markup in web page

– dedicated markup

e.g. zLinks 

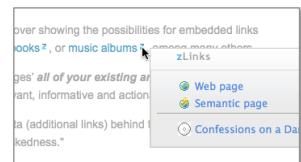
– microformats 

human readable text ...

... but also machine readable

for search engines or plug-ins

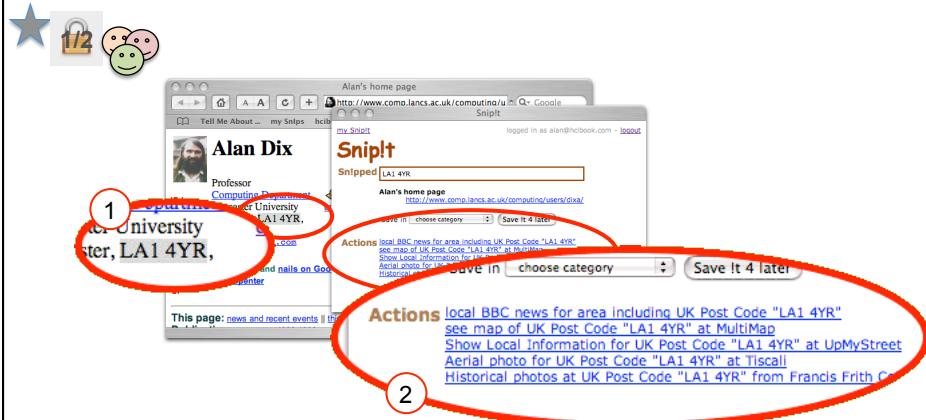
```
<p class="vcard">Hi, my name is <span class="fn">Jamie  
Jones</span> and I dig microformats!</p>
```



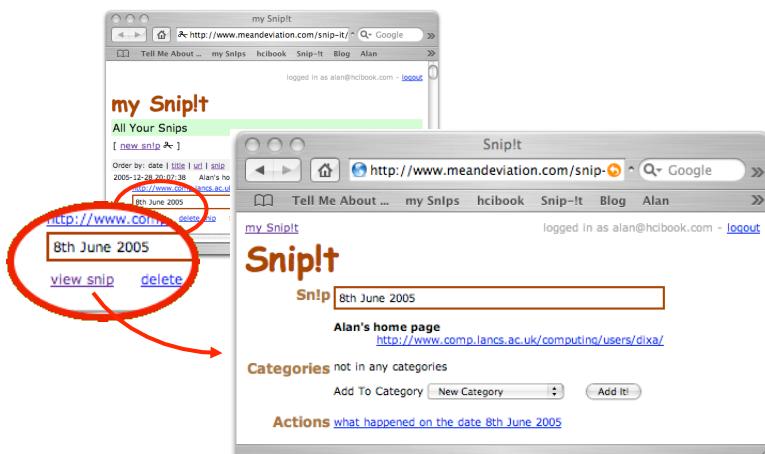
zitgist.com/products/zlinks

microformats.org

linking at point of use Snip!it ✈

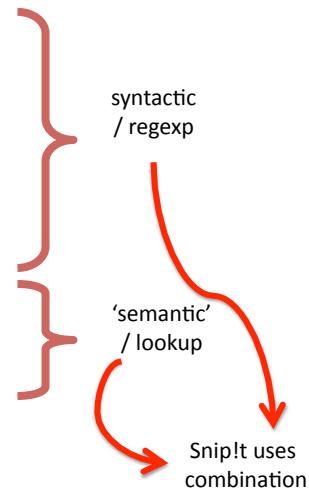


linking at point of use Snip!it ✈



class of systems 'data detectors'

- late 1990s
 - Intel selection recognition agent
 - Apple Data Detectors (Bonnie Nardi)
 - CyberDesk (Andy Wood led to onCue)
- recently
 - Microsoft SmartTags
 - Google extensions
 - Citrine – clipboard converter
 - CREO system (Faaberg, 2006)
- way back
 - Microcosm (Hypertext external linkage)



do it yourself!

cf. Calais yesterday

- tellmeabout web service
- XML version of Snip!t detection and actions
- Tag the Net
- used in WordPress

appears to use context words

<http://tagthe.net/api/?text=Alan+Dix+is+at+lancaster+not+London-on-29-Oct-2008>

```
<?xml version="1.0" encoding="UTF-8"?>
<memes>
  <dim type="topic">          <item>London</item>          </dim>
  <dim type="person">          <item>Alan Dix</item>          </dim>
  <dim type="location">        <item>Lancaster</item>        </dim>
  <dim type="language">        <item>english</item>         </dim>
</meme>
</memes>
```

tagthe.net

linking in the hidden web

hidden web / deep web

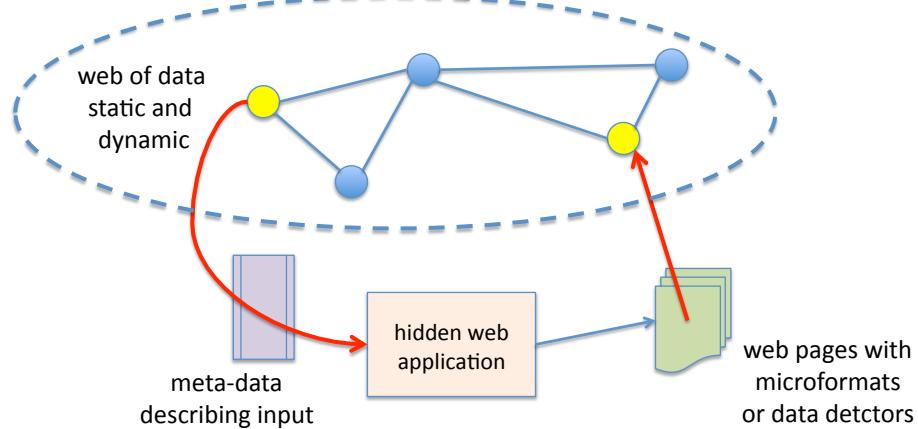
- searchable data bases etc.
 - 90% of web accessible data not searchable
- add meta-information**
- internal (page markup) or external (e.g. Snipit)

```
<urlservice>
  <type>name</type>
  <title>switchboard.com</title>
  <descpattern>lookup US person $name at switchboard.com</descpattern>
  <urlpattern>http://www.switchboard.com/bin/cgiqa.dll?
    SR=&MEM=1&LNK=33:4&F=${forename}&L=${surname}</urlpattern>
</urlservice>
```



www.snipit.org

reweaving the hidden web through user interaction





from use to data

Amazon don't ask people what they like
... people just buy what they like

action is data

from use to data

using interaction to generate semantics

- selection:
 - user selects data and uses it in semantic field
- confirmation
 - if user uses inferred data assume correct
- web forms
 - type annotation from use

context in forms



Hotels R Us

Name	Alan Dix
Org.	Lancaster Univ.

entry of first field sets
context for rest of form

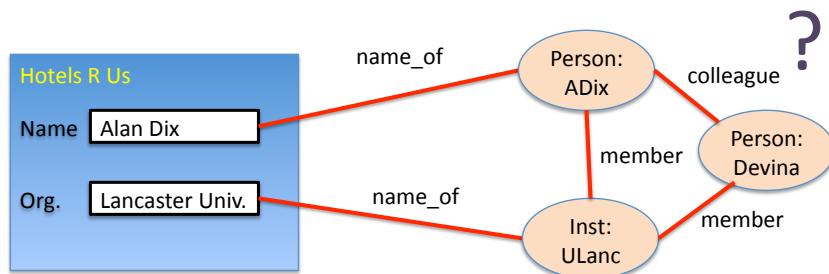
but what is the relationship?

maybe semantic markup on form

– good SemWeb style ... but rare

... or more inference ...

context in forms - inference



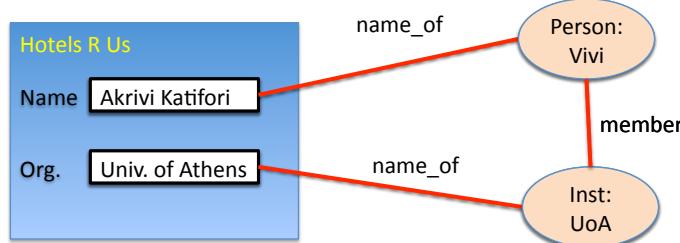
match terms in form to ontology

look for 'least cost paths'

- number of relationships traversed, fan-out

Dix, A., Katifori, A., Poggi, A., Catarci, T., Ioannidis, Y., Lepouras, G., Mora, M. (2007). From Information to Interaction: in Pursuit of Task-centred Information Management.
<http://www.hcibook.com/alan/papers/DELOS-TIM2-2007/>

context in forms - inference



match terms in form to ontology

look for 'least cost paths'

- number of relationships traversed, fan-out

later suggest based on rules