1º Fórum W3C Brasil

Por uma Web Única

On the Way to the Semantic Web

Presented on 1° Fórum W3C Brasil, by Klaus Birkenbihl,
Coordinator World Offices, W3C

based on a slide set mostly created by Ivan Herman,
Semantic Web Activity Lead, W3C

Sept. 30th 2008

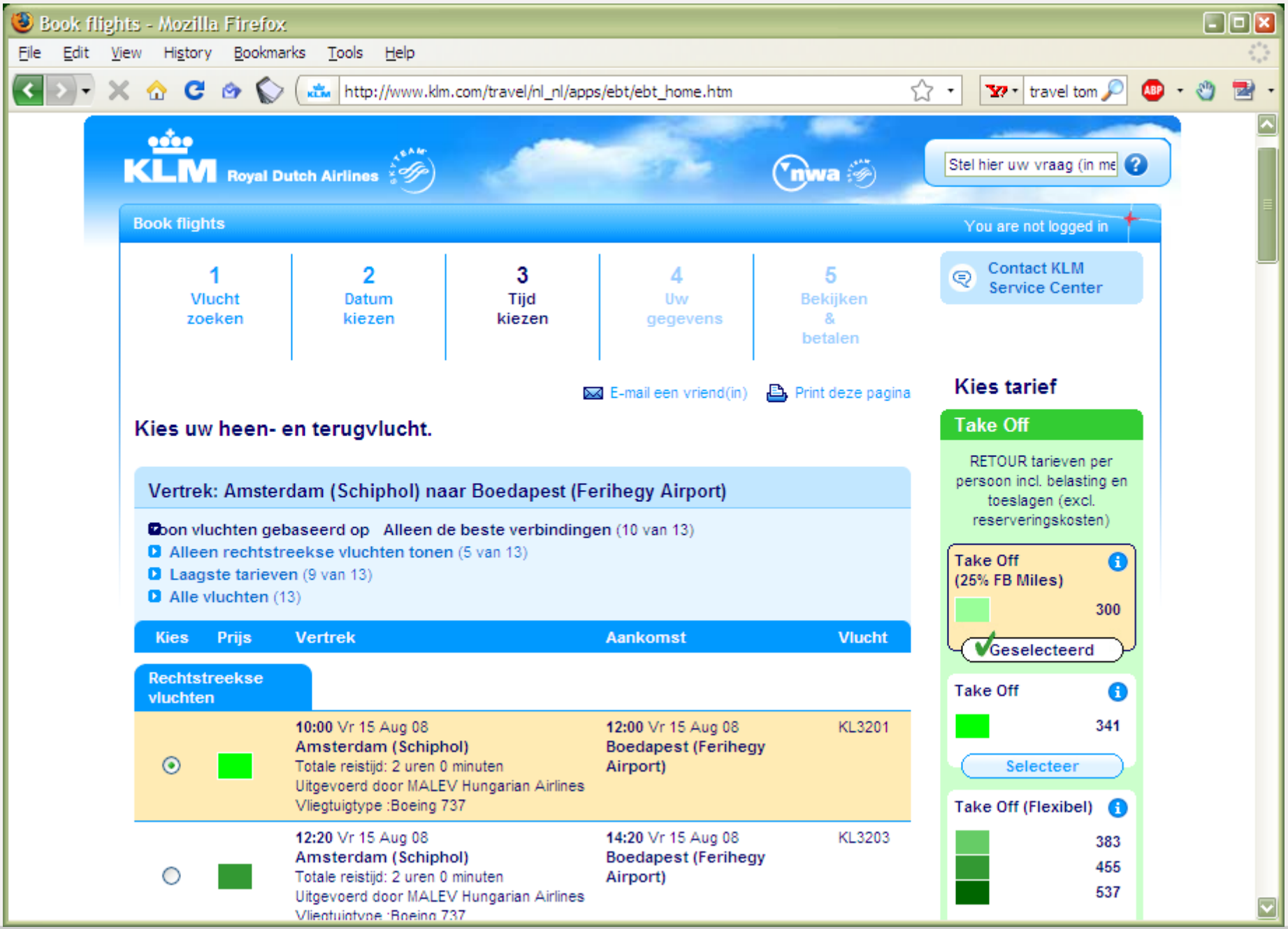
Let’s organize a trip to Budapest using the Web!

Copyright © 2008, W3C

You try to find a proper flight with …

Copyright © 2008, W3C

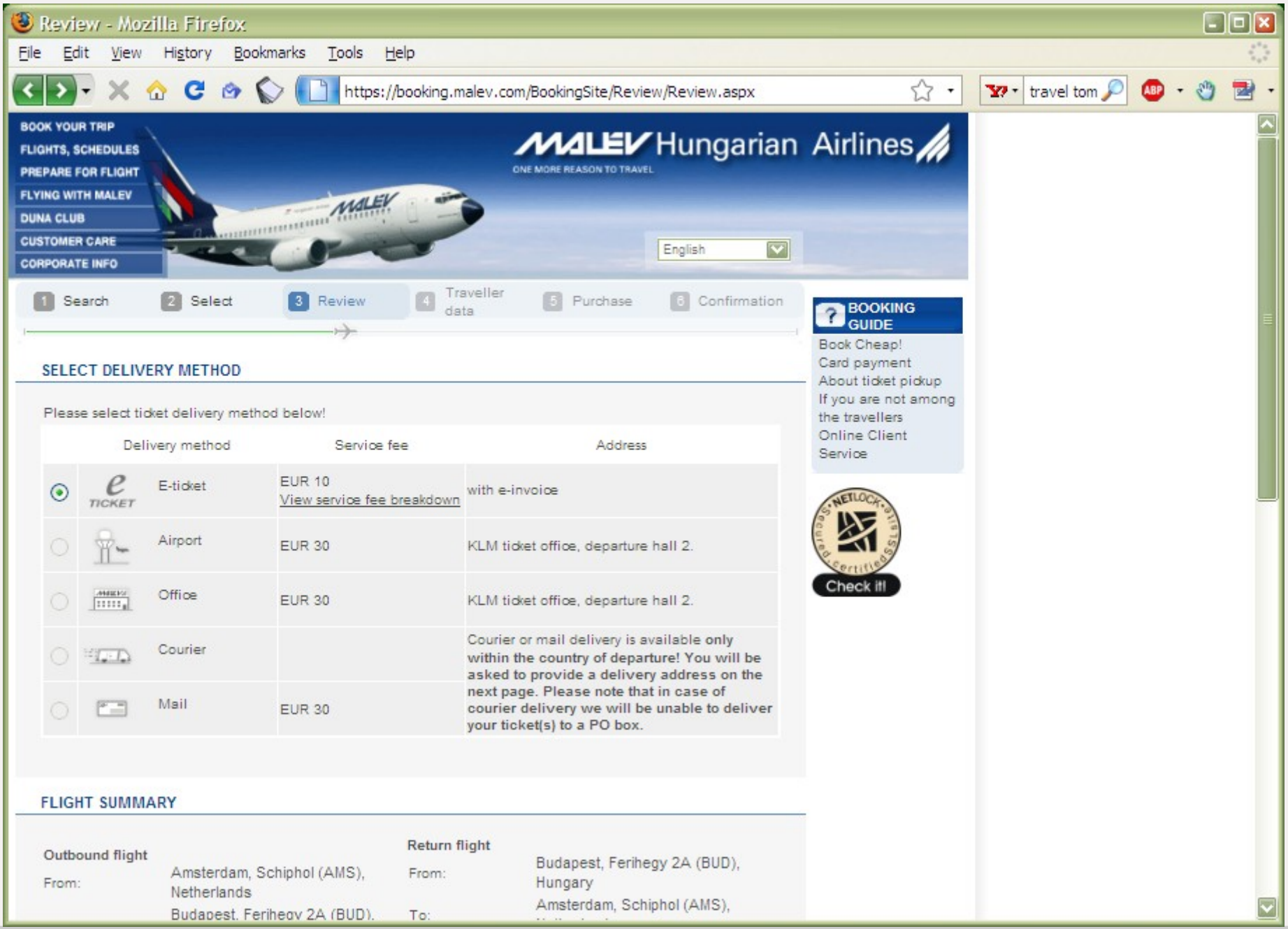
… a big, reputable airline, or …



Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

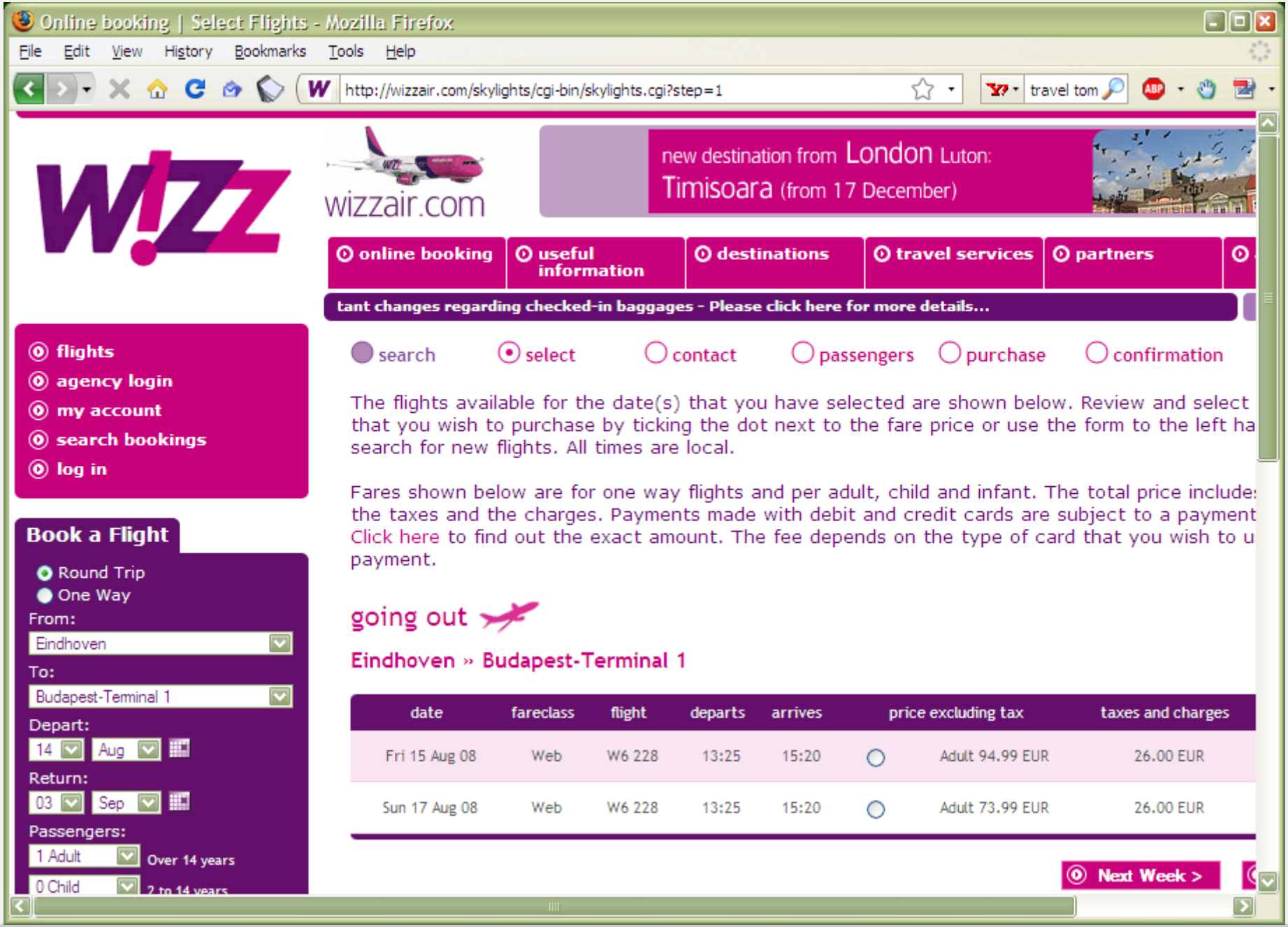
… the airline of the target country, or …



Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

… or a low cost one



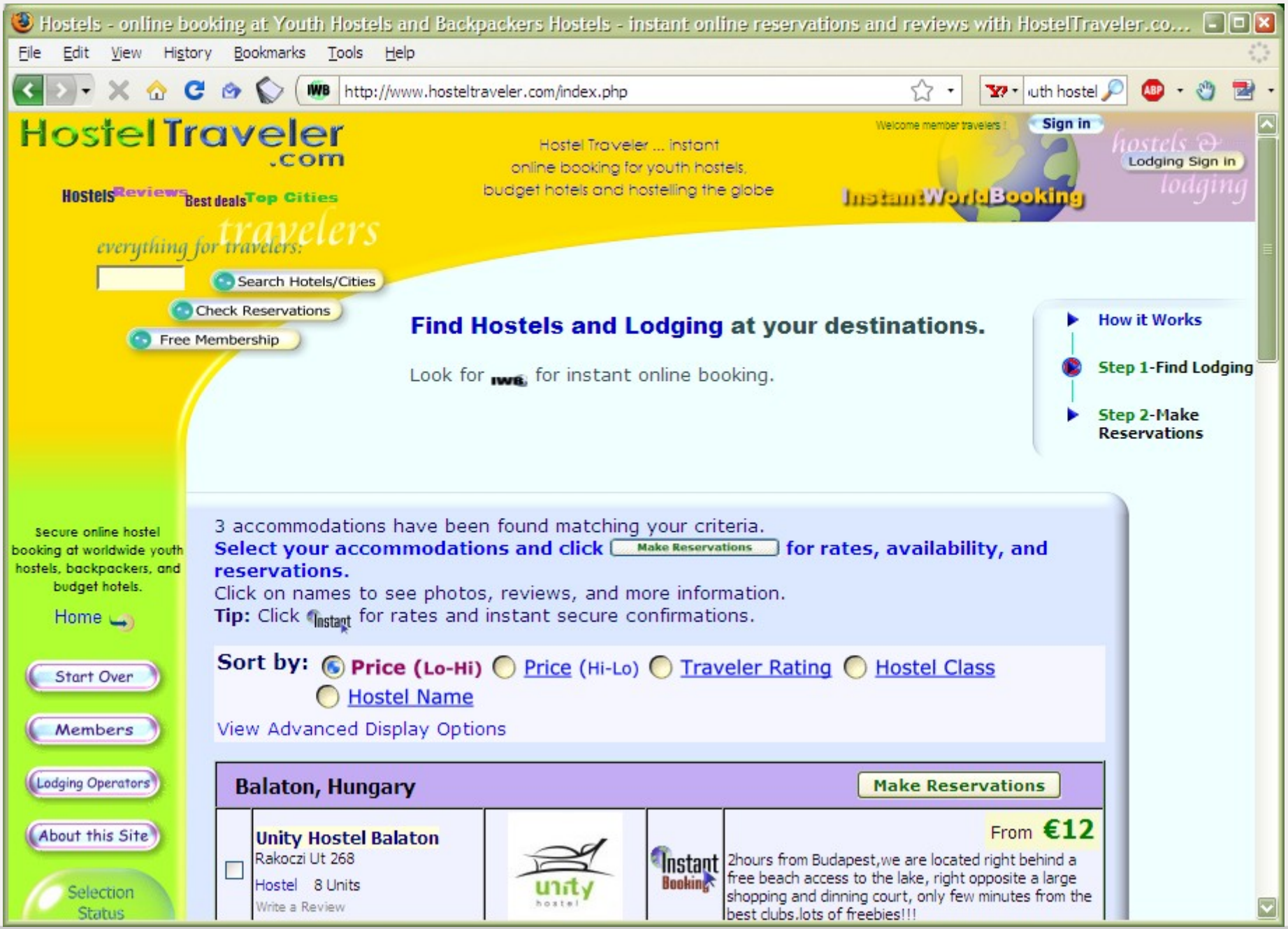
Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

You have to find a hotel, so you look for…

Copyright © 2008, W3C

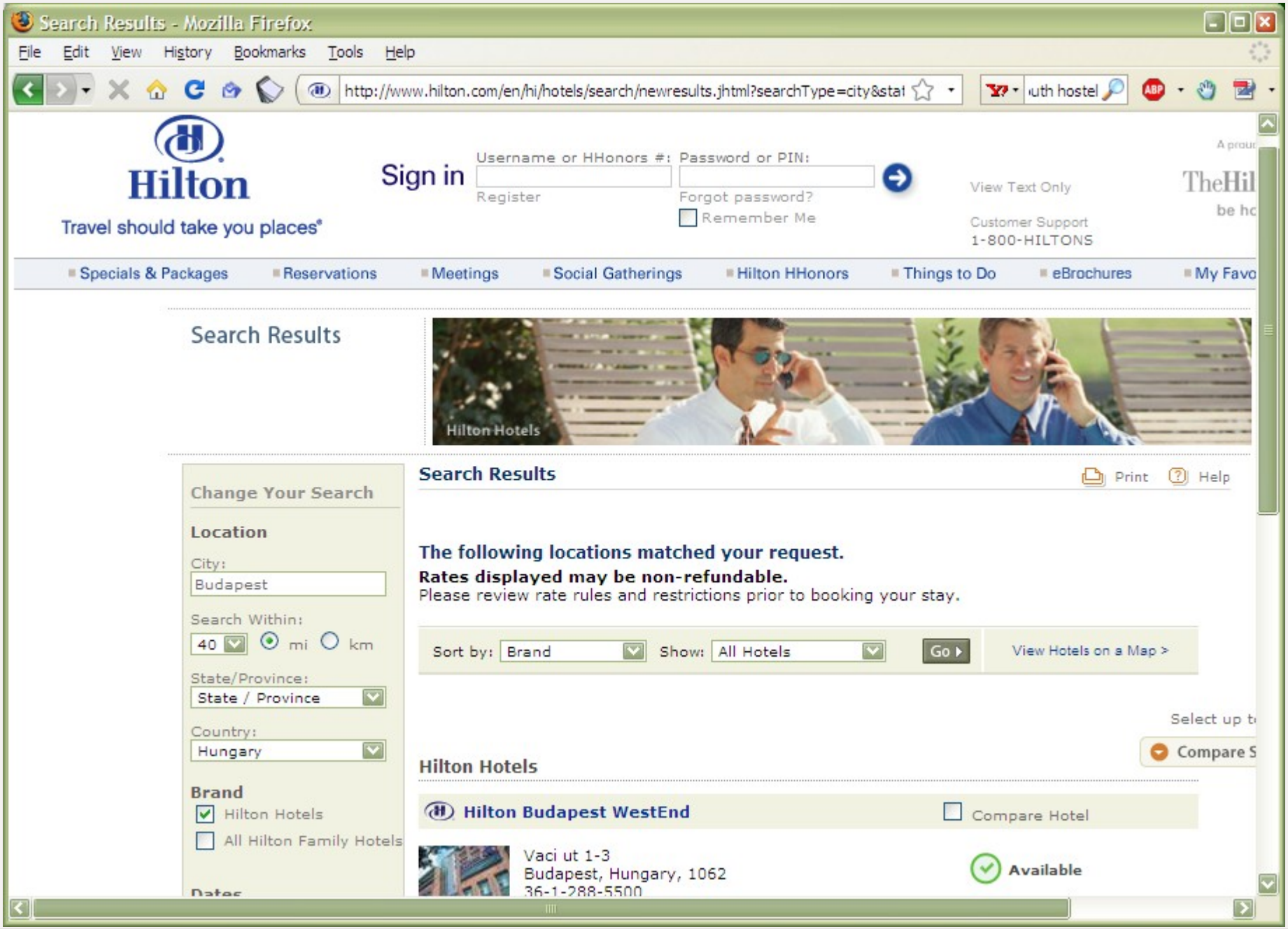
… a really cheap accommodation, or …



Copyright © 2008, W3C

1º Fórum W3C Brasil Por uma Web Única
W3C Semantic Web

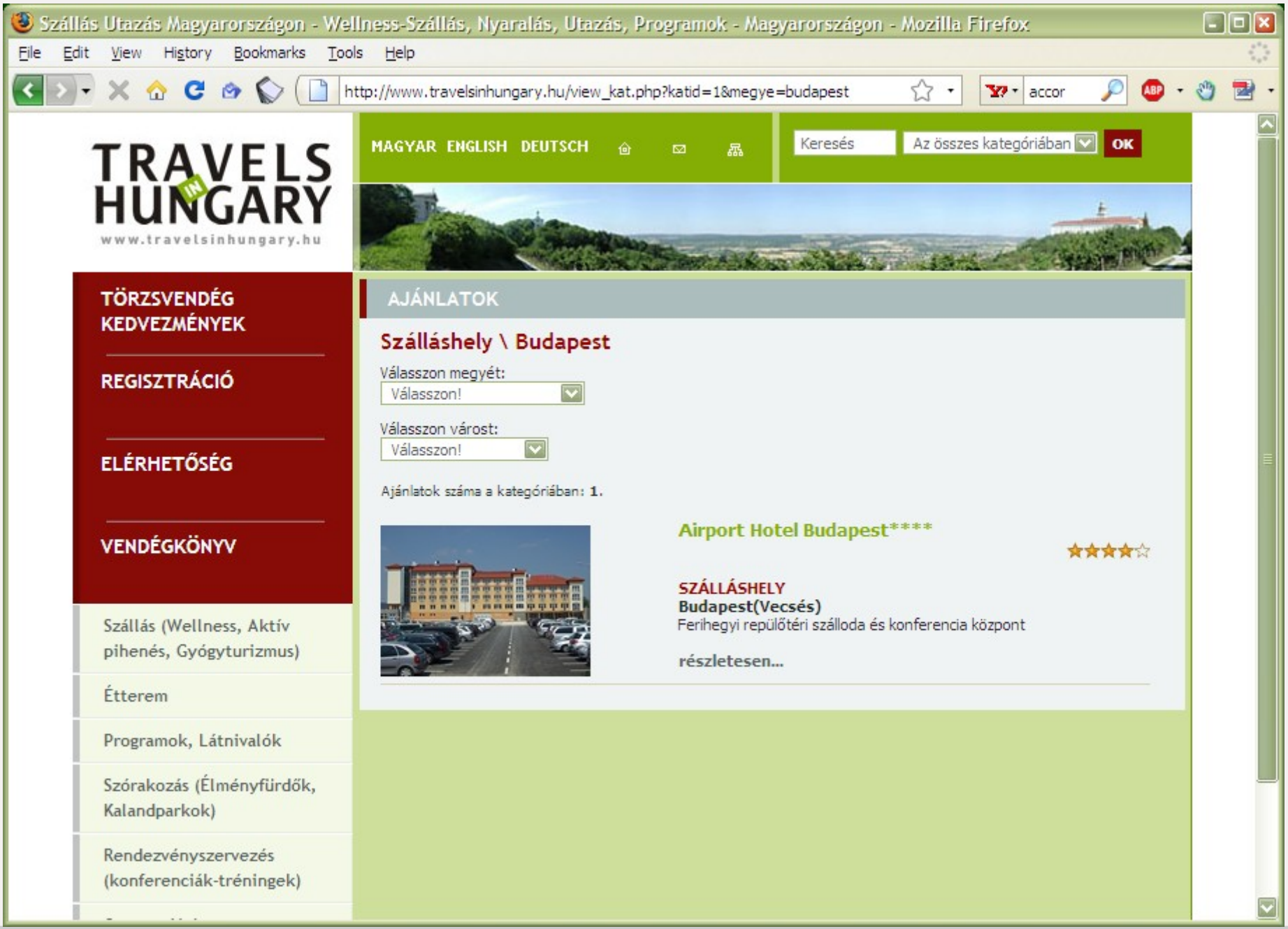
… or a really luxurious one, or …



Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

… and intermediate one …



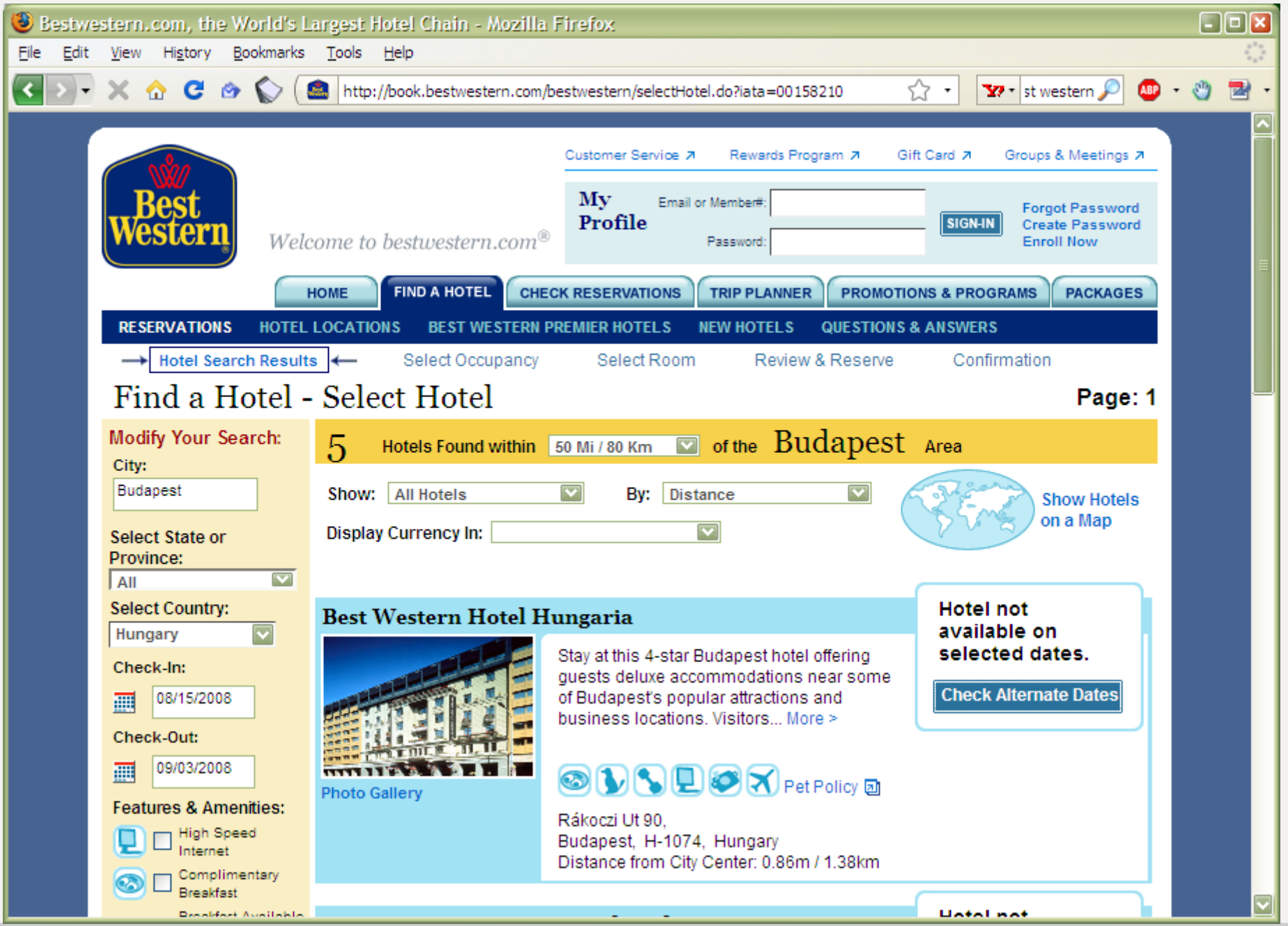
Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

oops, that is no good, the page is in Hungarian that almost nobody understands, but…

Copyright © 2008, W3C

… this one could work



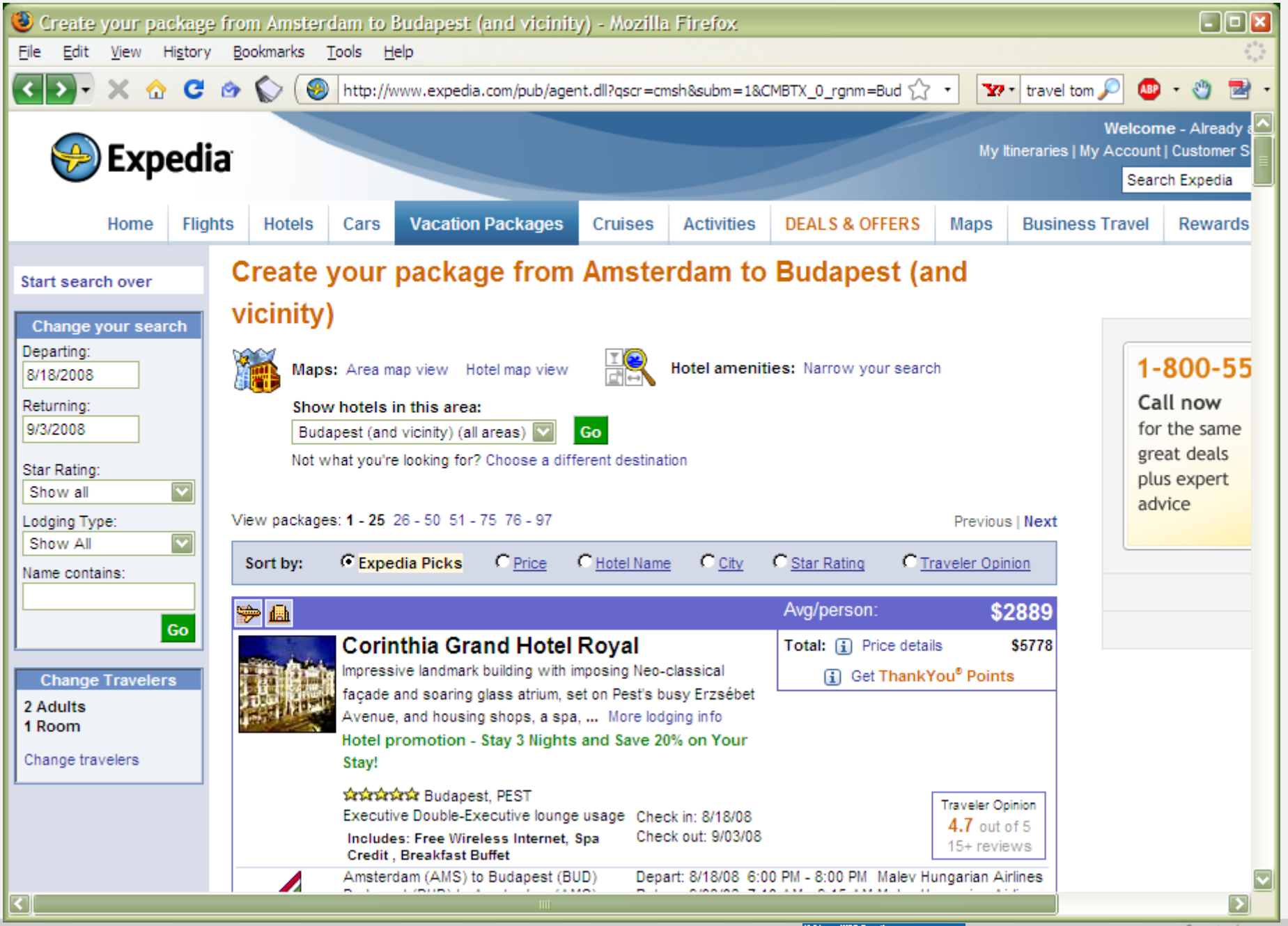
Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

Of course, you could decide to trust a specialized site…

Copyright © 2008, W3C

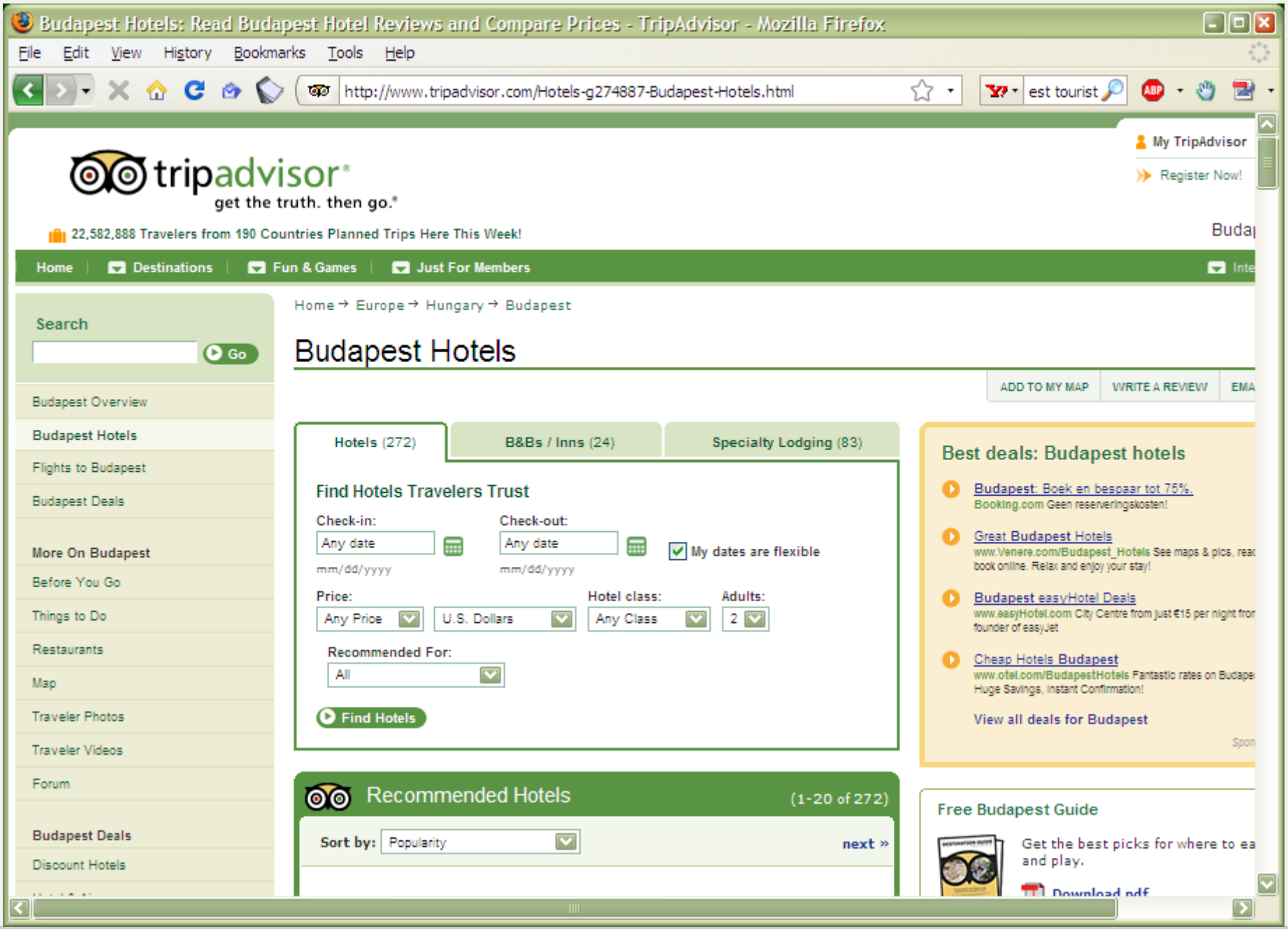
… like this one, or…



Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

… or this one



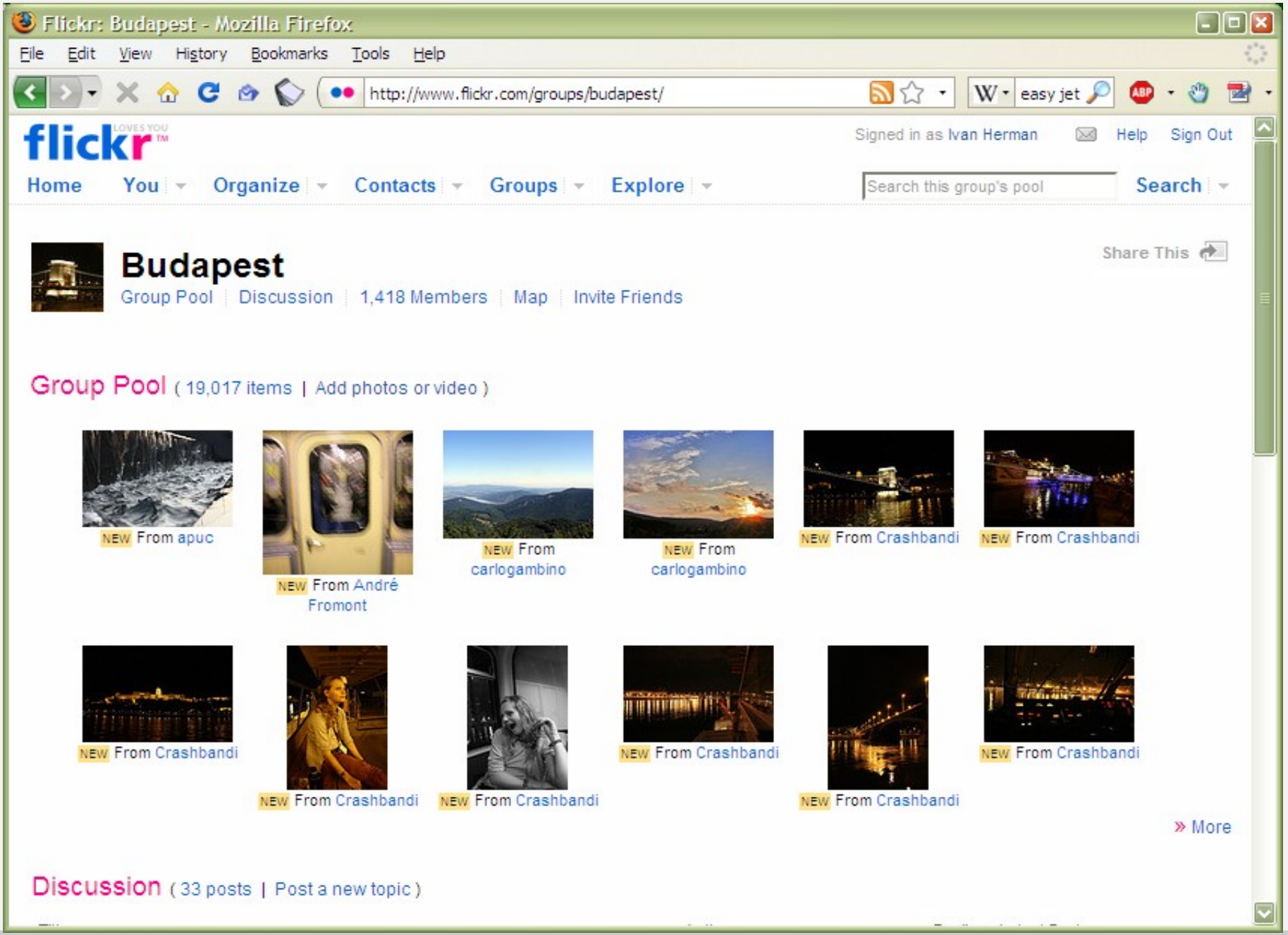
Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

You may want to know something about Budapest; look for some photographs…

Copyright © 2008, W3C

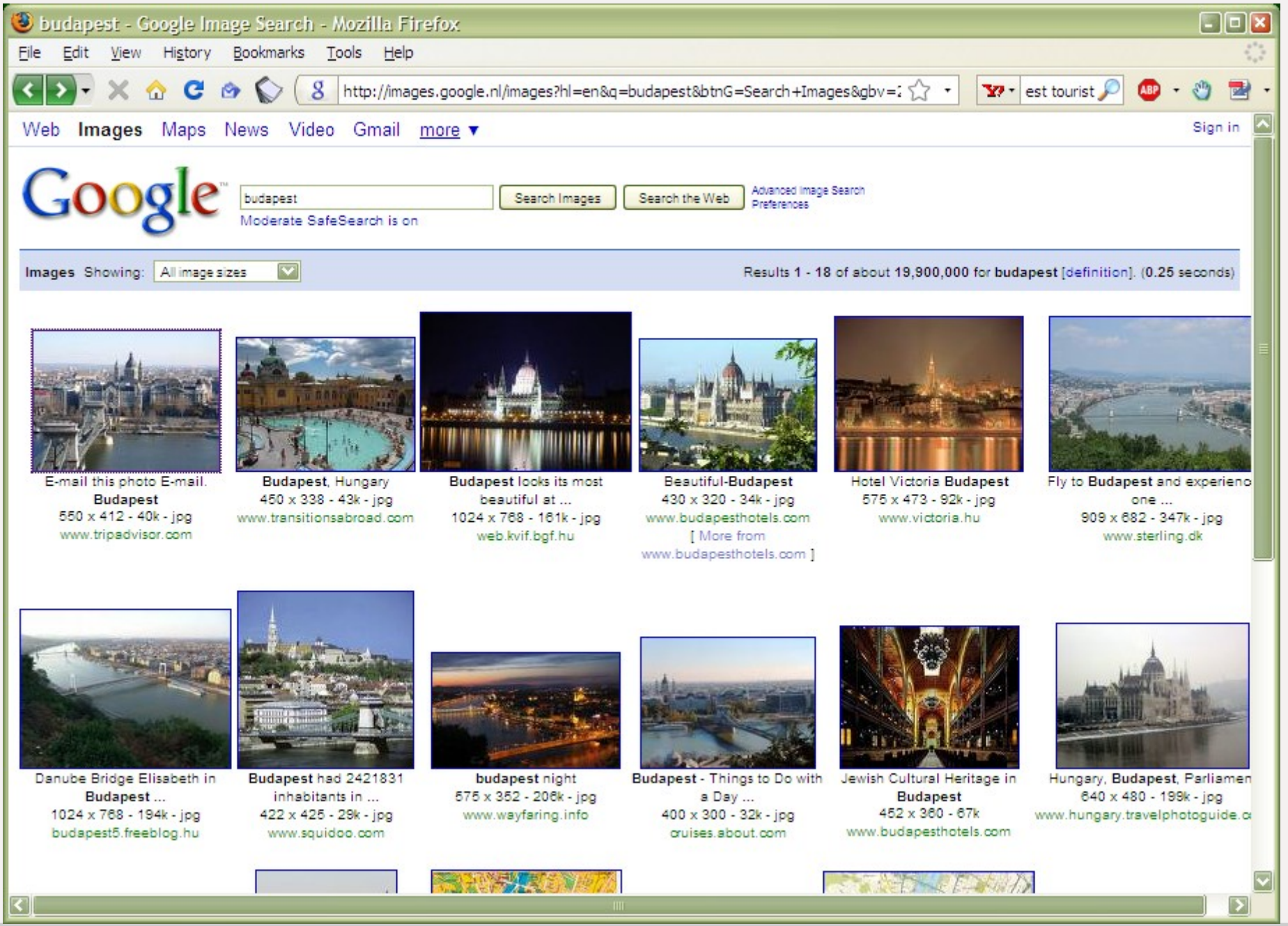
… on flickr …



Copyright © 2008, W3C

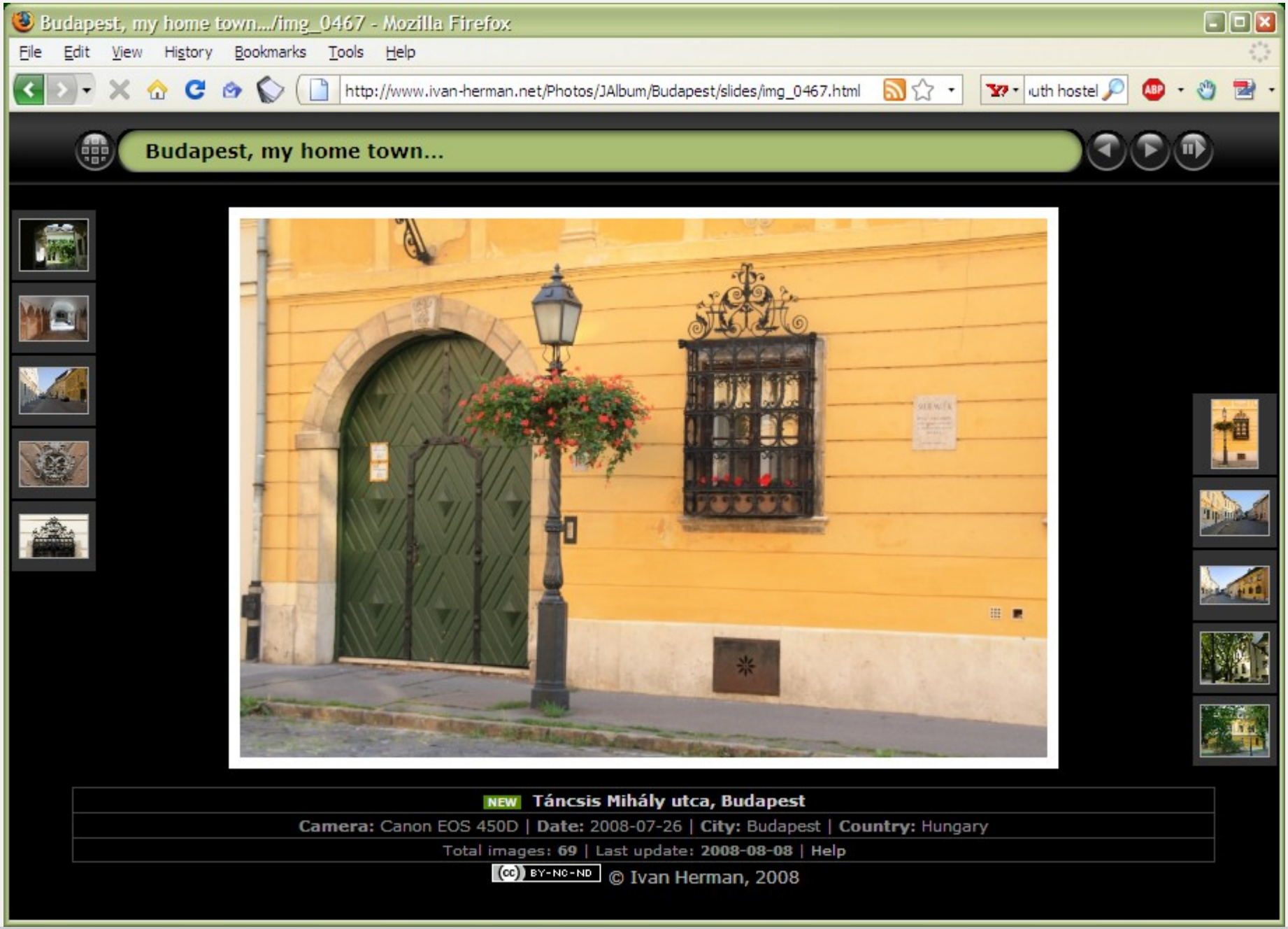
1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

… on Google …



Copyright © 2008, W3C

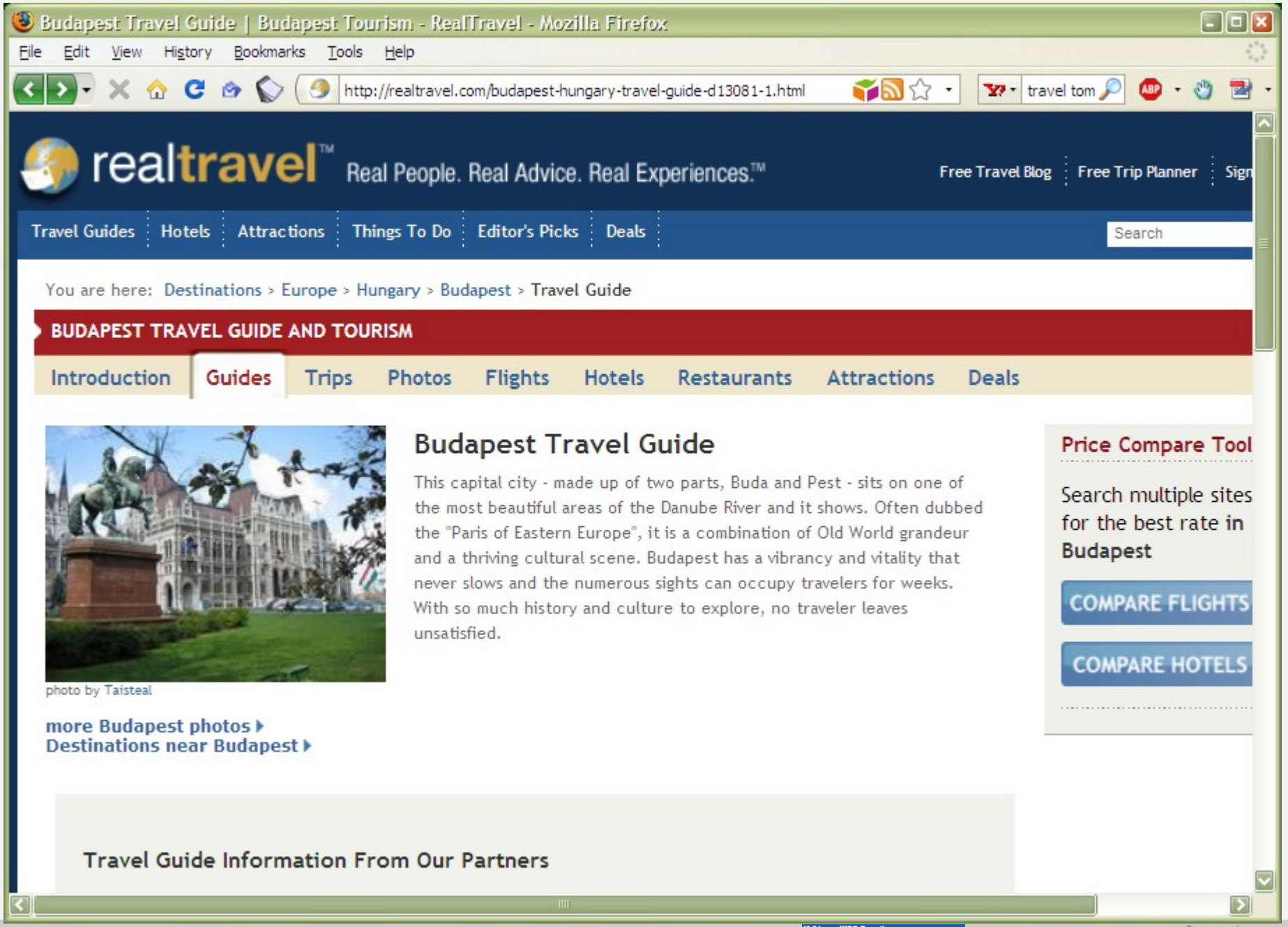
… or you can look at Ivan's



Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

but you can also look at a (social) travel site



Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

What happened here?

- You had to consult a large number of sites, all different in style, purpose, possibly language…
- You had to mentally *integrate* all those information to achieve your goals
- We all know that, sometimes, this is a long and tedious process!

Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

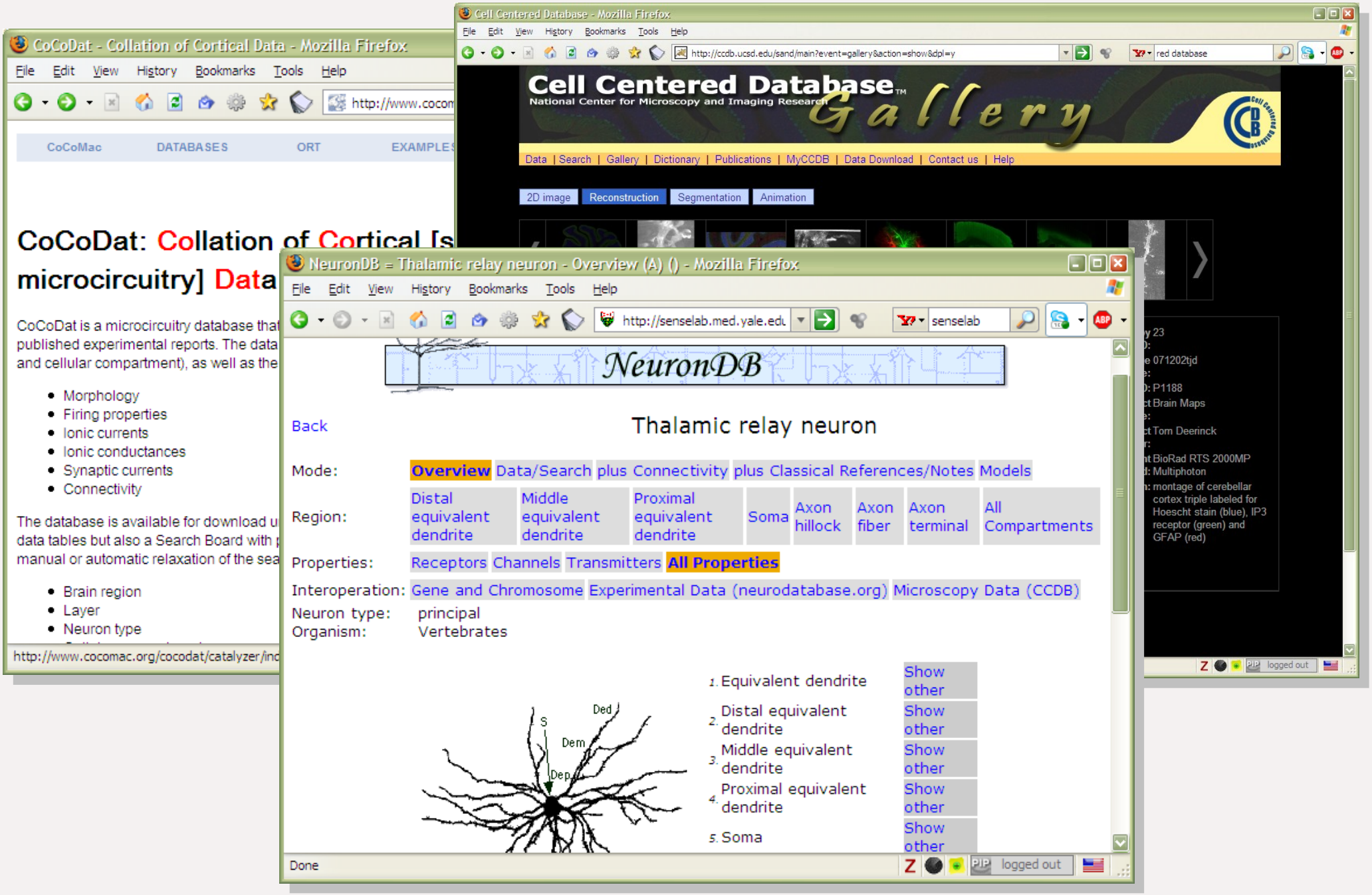
- All those pages are only tips of respective icebergs:
  - the real *data* is hidden somewhere in databases, XML files, Excel sheets, …
  - you have only access to what the Web page designers allow you to see
- Specialized sites (Expedia, TripAdvisor) do a bit more:
  - they gather and combine data from other sources (usually with the approval of the data owners)
  - but they still control how you see those sources

Copyright © 2008, W3C

- Sometimes you want more: you may want access to the original data and combine it yourself!

Copyright © 2008, W3C

Here is another example…



Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

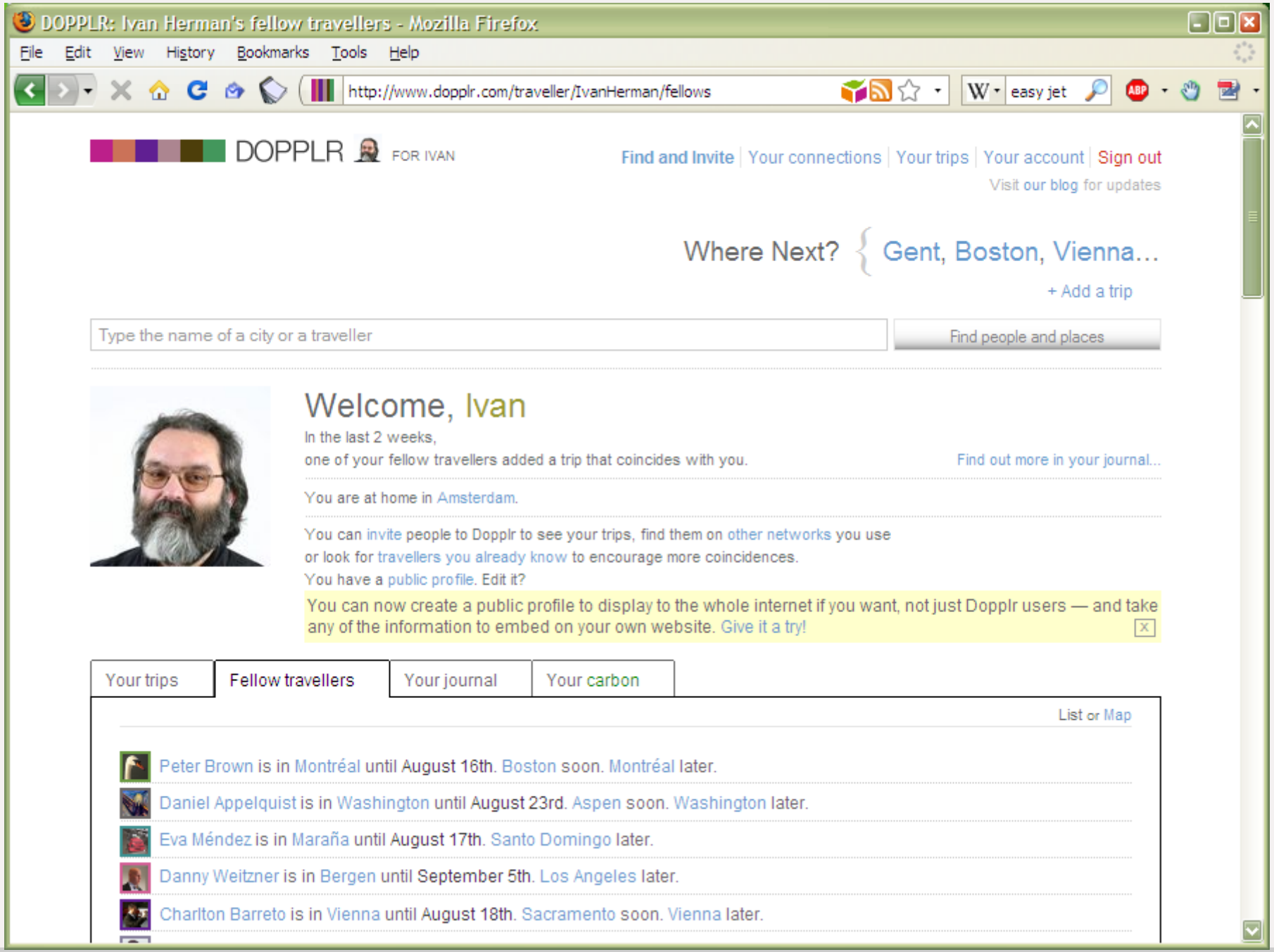
- Companies may have to hire a person to answer questions based on those (public!) databases!

Copyright © 2008, W3C

Another example: social sites. Ivan has a list of “friends” by…

Copyright © 2008, W3C

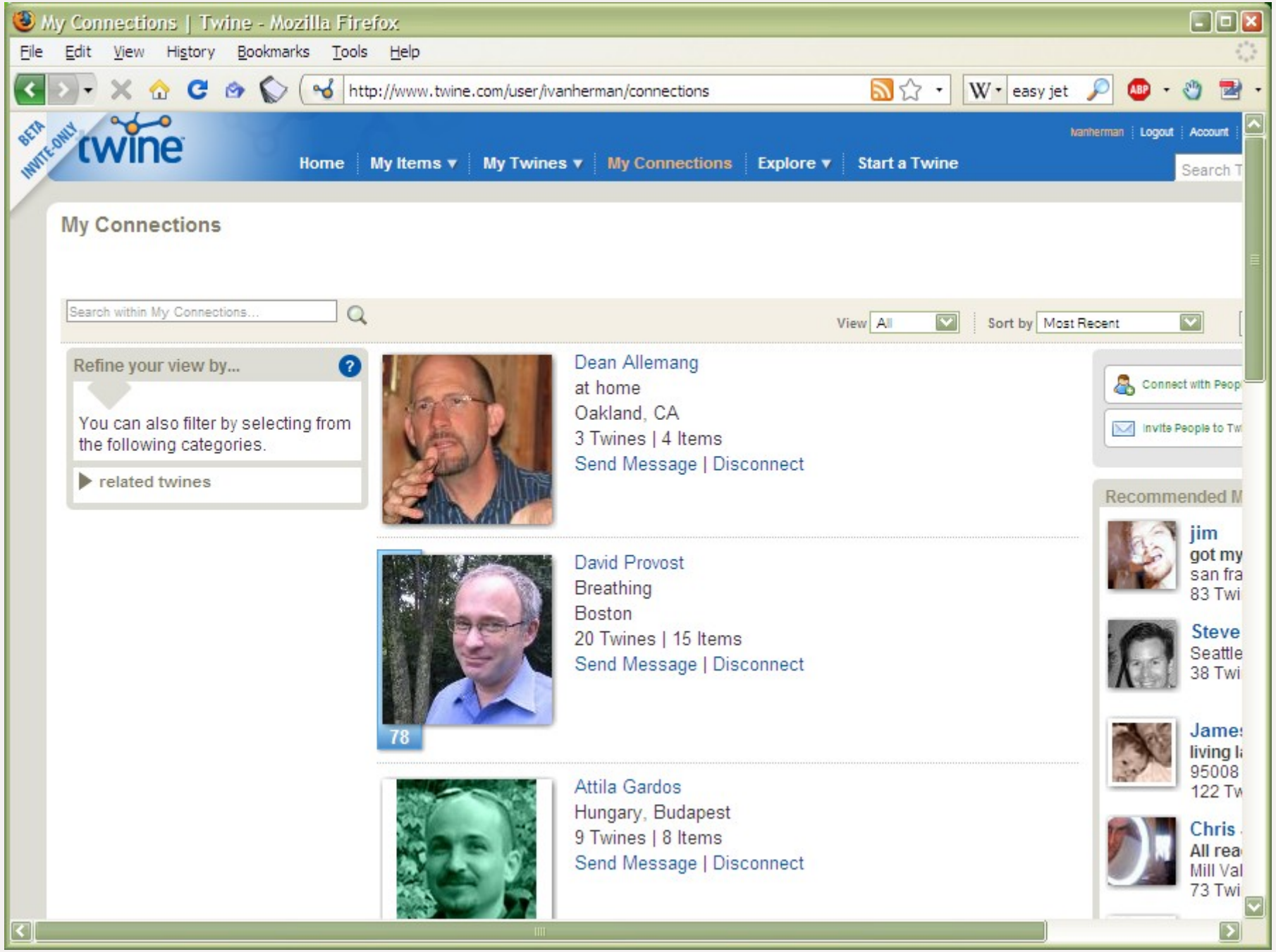
... Dopplr,



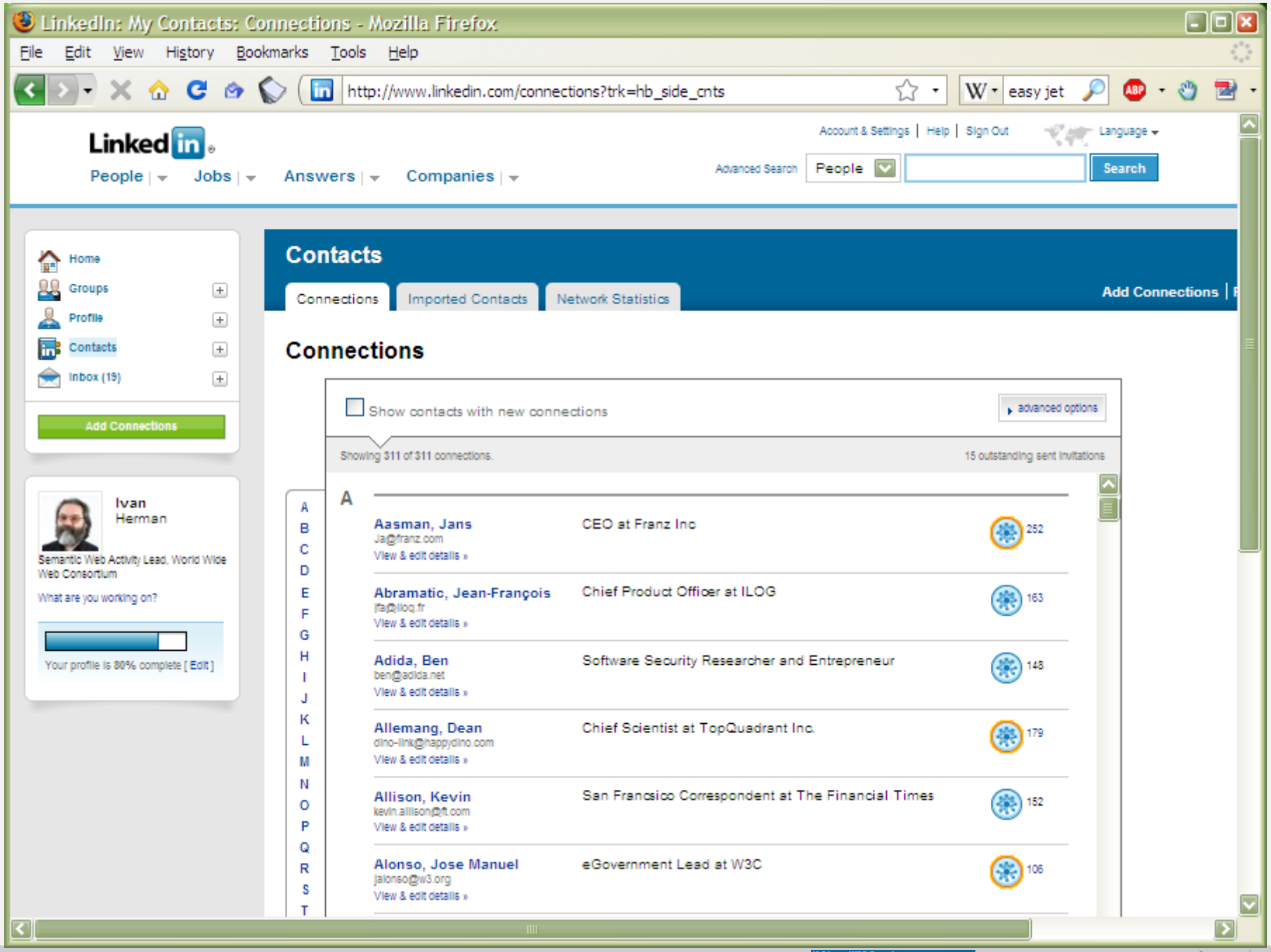
Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

… Twine,

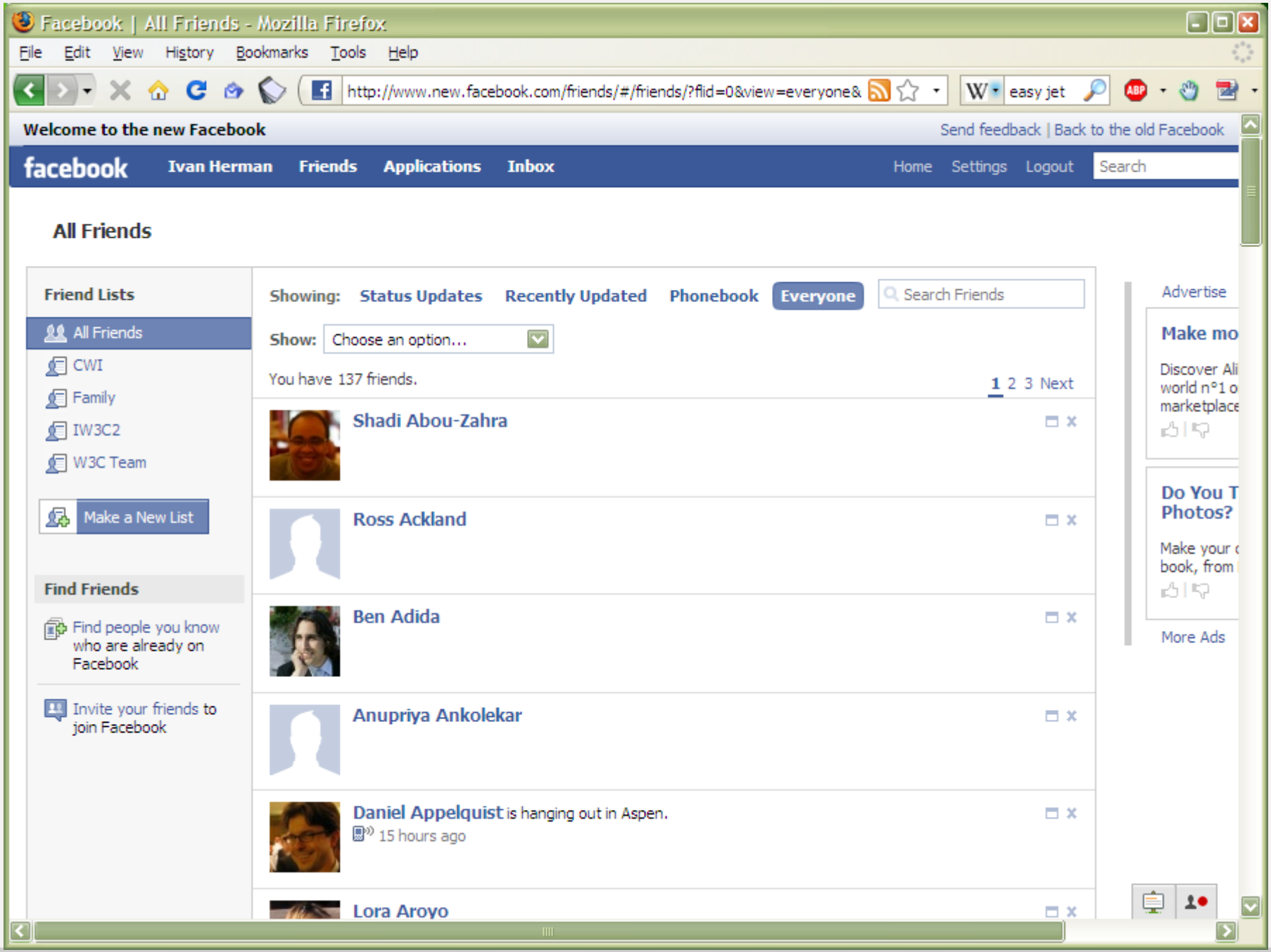

My Connections | Twine - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.twine.com/user/ivanherman/connections
easy jet
BETA INVITE-ONLY
twine
Home
My Items
My Twines
My Connections
Explore
Start a Twine
Logout
Account
My Connections
Search within My Connections...
View All
Sort by Most Recent
Refine your view by...
You can also filter by selecting from the following categories.
related twines
Dean Allemang
at home
Oakland, CA
3 Twines | 4 Items
Send Message | Disconnect
David Provost
Breathing
Boston
20 Twines | 15 Items
Send Message | Disconnect
78
Attila Gardos
Hungary, Budapest
9 Twines | 8 Items
Send Message | Disconnect
Recommended

… LinkedIn,



Copyright © 2008, W3C

… and, of course, the ubiquitous Facebook



Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

- He had to type in and connect with friends again and again for each site independently 
- This is even worse then before: he **feeds** the icebergs, but he still does not have an easy access to data…

Copyright © 2008, W3C

What would we like to have?

- Use the data on the Web the same way as we do with documents:
  - be able to link to data (independently of their presentation)
  - use that data the way I want (present it, mine it, etc.)
  - agents, programs, scripts, etc. should possibly be able to *interpret* part of that data
- This does not mean you have to do these presentations, scripts ... but it would provoke ideas and applications that make better use of existing data
- *"The bane of my existence is to do things that my computer could do for me"* — Dan Connolly

Copyright © 2008, W3C

Put it another way…

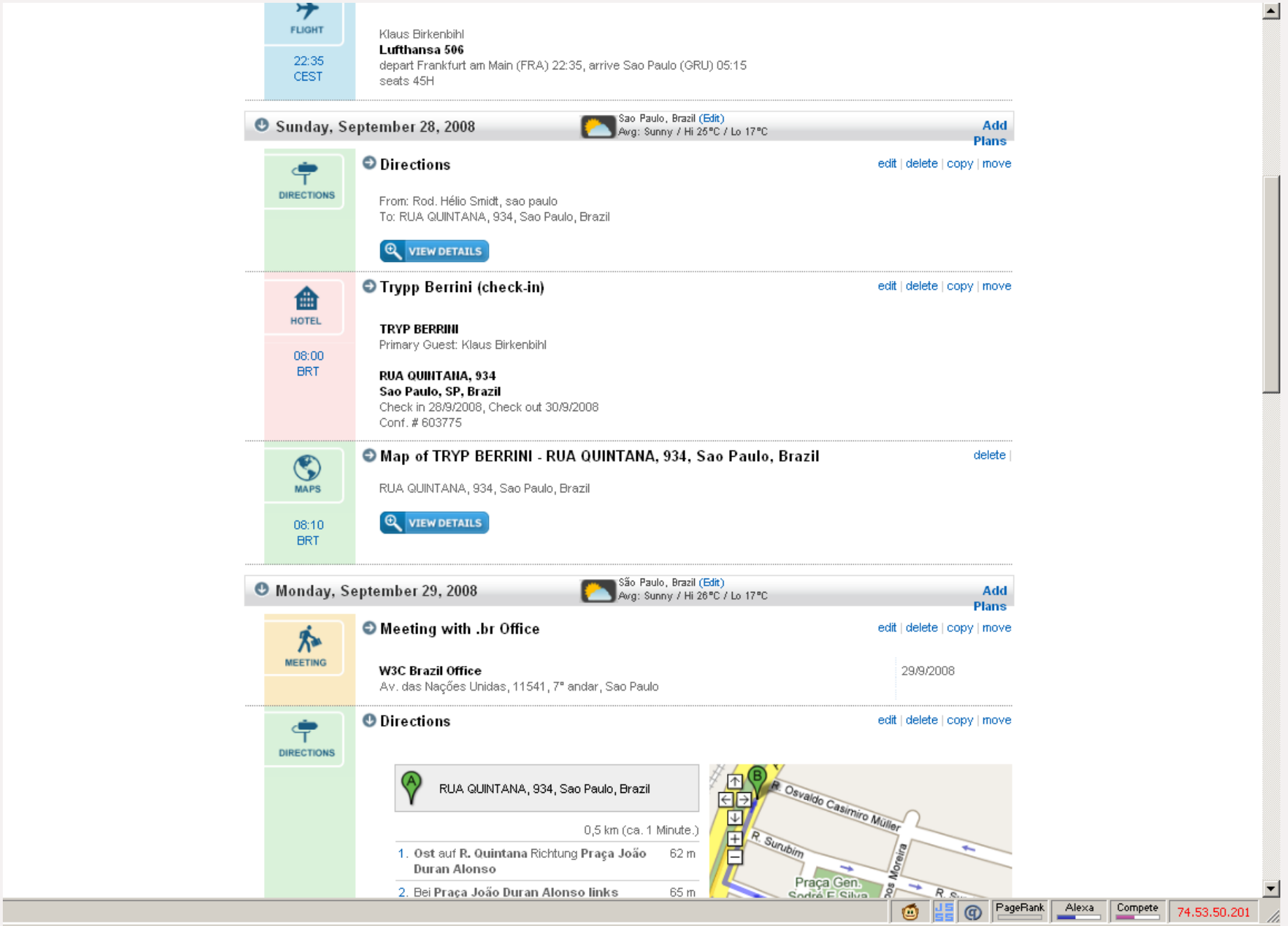
- We would like to *extend* the current Web with a "*Web of data*":
  - allow for applications to exploit the data directly

Copyright © 2008, W3C

But wait! Isn’t this what mash-up sites are already doing?

Copyright © 2008, W3C

Example: Klaus' trip to Brazil (tripit.com)



Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

How does it work

- Klaus forwards to Tripit the documents (mails, URIs) I have wrt a trip. e.g.
  - Flight bookings
  - Hotel reservations
  - Meetings
- Any time he has new documents he may add them

- Tripit tries to extract the relevant data from these documents
- It associates the documents to a trip
- It add information from other sites about weather, directions, travel guides ...
- It checks its own database for travel activities of friends
- It compiles a structured itinerary

Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

This is great ... but

- Sometimes Tripit sends Klaus a message “Problem with your TripIt submission”
- Sometimes some data from a document are not identified
- Sometimes Klaus reads: “Please help us to improve! Let us know how good we captured your flight.”
- This is a hint on what Tripit does: in case it does not know how to find the data it has to guess
- Because there is no standardized way to access the data it has to use proprietary interfaces to get them.

Copyright © 2008, W3C

- So in some ways, this shows the huge power of what a Web of data provides
- But mash-up sites are forced to do very ad-hoc jobs
  - various data sources expose their data via Web Services
  - each with a different API, a different logic, different structure
  - these sites are forced to reinvent the wheel many times because they don't use a standard way of doing things

Copyright © 2008, W3C

Put it another way (again)…

- We would like to *extend* to the current Web with a **standard** way for a "*Web of data*"

Copyright © 2008, W3C

But what does this mean?

- What makes the current (document) Web work?
  - people create different documents
  - they give an globally unique address to it (i.e. a URI) and make it accessible to others on the Web

Copyright © 2008, W3C

An example: Steven's site on Amsterdam

Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

Then some magic happens…

- Others discover the site and they link to it
  - So Search engines can find it and index it
- The more they link to it, the more important and well known the page becomes
  - remember, this is one criterion, Search engines use to rank pages.
- This is the “Network effect”: some pages become important, and others begin to rely on it (even if the author did not expect it…)

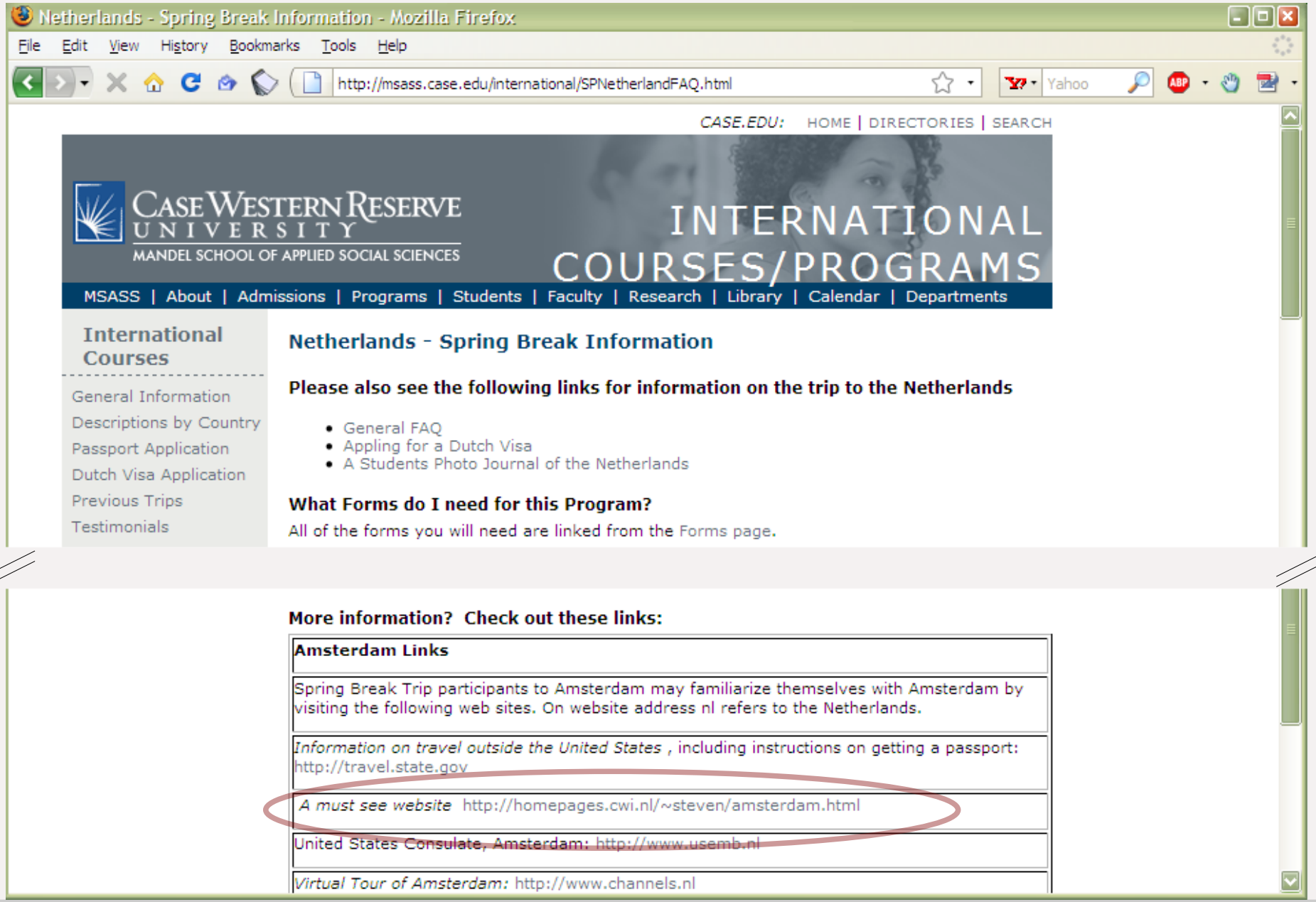
Copyright © 2008, W3C

This link to Steven's is, sort of, understandable…

Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

… but this one is on the other side of the Globe!



Copyright © 2008, W3C

1º Fórum W3C Brasil Por uma Web Única
W3C Semantic Web

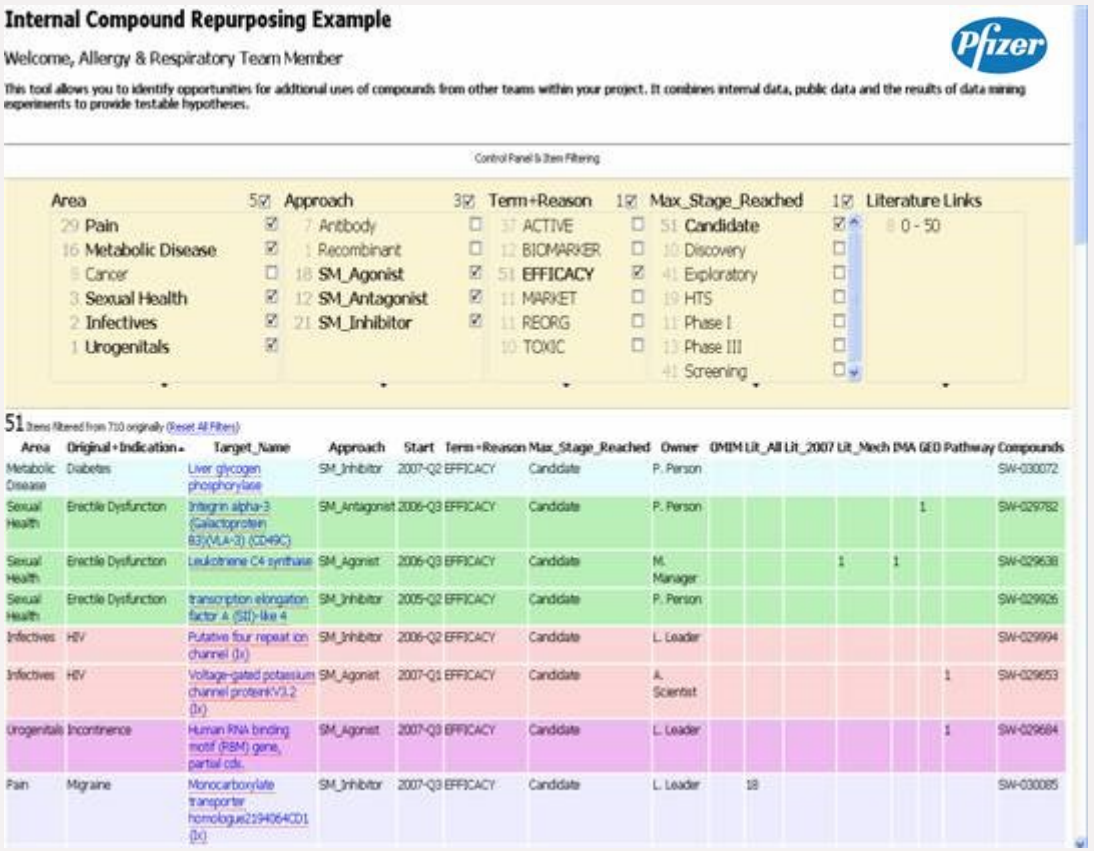
What would that mean for a Web of Data?

- Lessons learnt: we should be able to:
  - "publish" the data to make it known on the Web
    - standard ways should be used instead of ad-hoc approaches
    - the analogous approach to documents: *give URI-s to the data*
  - make it possible to "link" to that URI from *other* sources of data (not only Web pages) using standard approaches
    - ie, applications should not be forced to make targeted developments to access the data (as we saw with mash-ups)
    - generic, standard approaches should suffice
  - and let the network effect work its way…

Copyright © 2008, W3C

Example: combine data from experiments

- A drug company has huge amount of old experimental data on its Intranet
- Data in different formats (XML, databases, …)
- To reuse them:
  - make the important facts available on the Web via standards
  - use *off-the-shelf* tool to integrate, display, search



Courtesy of Nigel Wilkinson, Lee Harland, Pfizer Ltd, Melliyal Annamalai, Oracle (SWEO Case Study)

Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

But it is a little bit more complicated 

- On the traditional Web, humans are implicitly taken into account
- A Web link has a “context” that a person may use

Copyright © 2008, W3C

Eg: public encryption key on Ivan's page:

Ivan Herman - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.w3.org/People/Ivan/

Ivan Herman

My Work at W3C | Contact information | Short CV | Upcoming trips | Public presentations

My Work at W3C

I am Semantic Web Activity Lead; that is my main work at W3C. I am member of IW3C2 (International World Wide Web Conference Committee) (the committee coordinating the yearly WWW conference series), serving as a liaison for W3C, and of SWSA (Semantic Web Science Association), the committee responsible for the International Semantic Web Conferences series.

As part of my work, I also participate in lots of outreach activities, and I regularly make presentations, tutorials, etc. You can consult my list of presentations for further details.

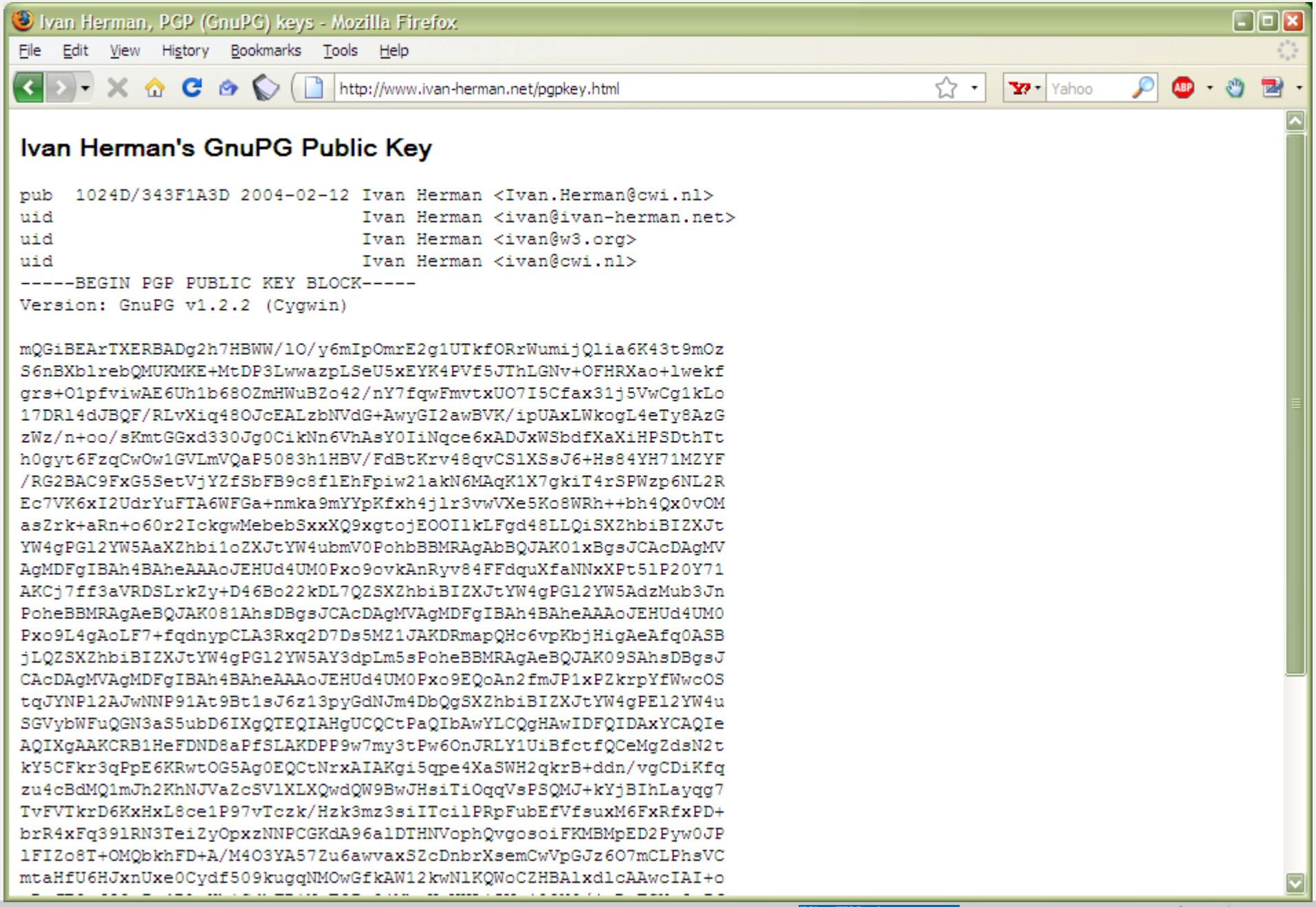
Contact information

Email:
ivan@w3.org
(sha1sum: 5ac8032d5f6012aa1775ea2f63e1676bafd5e80b)
Postal address:
C/o Centre for Mathematics and Computer Sciences (CWI)
Kruislaan 413, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands.
Phone numbers:
phone: +31-20-5924163
mobile phone: +31-641044153
fax: +31-20-5924312
PGP/GPG:
My GnuPGP key and signature is available on-line.
FOAF:
You can either extract a short FOAF information from this page, of consult my more complete, public FOAF file.
Misc:

Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

Eg: public encryption key on Ivan's page:



Ivan Herman's GnuPG Public Key

```
pub  1024D/343F1A3D 2004-02-12 Ivan Herman <Ivan.Herman@cwi.nl>
uid                            Ivan Herman <ivan@ivan-herman.net>
uid                            Ivan Herman <ivan@w3.org>
uid                            Ivan Herman <ivan@cwi.nl>
-----BEGIN PGP PUBLIC KEY BLOCK-----
Version: GnuPG v1.2.2 (Cygwin)

mQGiBEArTXERBADg2h7HBWW/lO/y6mIpOmrE2g1UTkfORrWumijQlia6K43t9mOz
S6nBXblrebQMUKMKE+MtDP3LwwazpLSeU5xEYK4PVf5JThLGNv+OFHRXao+lwekf
grs+O1pfviwAE6Uh1b68OZmHWuBZo42/nY7fqwFmvtxUO7I5Cfax31j5VwCg1kLo
17DR14dJBQF/RLvXiq48OJcEALzbNVdG+AwyGI2awBVK/ipUAxLWkogL4eTy8AzG
zWz/n+oo/sKmtGGxd330Jg0CikNn6VhAsY0IiNqce6xADJxWSbdfXaXiHPSDthTt
h0gyt6FzqCwOw1GVLmVQaP5083h1HBV/FdBtKrv48qvCS1XSsJ6+Hs84YH71MZYF
/RG2BAC9FxG5SetVjYZfSbFB9c8flEhFpiw21akN6MAqK1X7gkiT4rSPWzp6NL2R
Ec7VK6xI2UdrYuFTA6WFGa+nmka9mYYpKfxh4jlr3vwVXe5Ko8WRh++bh4Qx0vOM
asZrk+aRn+o60r2IckgwMebebSxxXQ9xgtojEOOIlkLFgd48LLQiSXZhbiBIZXJt
YW4gPGl2YW5AaXZhbi1oZXJtYW4ubmV0PohbBBMRAgAbBQJAK01xBgsJCAcDAgMV
AgMDFgIBAh4BAheAAAoJEHUd4UM0Pxo9ovkAnRyv84FFdquXfaNNxXPt5lP20Y71
AKCj7ff3aVRDSLrkZy+D46Bo22kDL7QZSXZhbiBIZXJtYW4gPGl2YW5AdzMub3Jn
PoheBBMRAgAeBQJAK081AhsDBgsJCAcDAgMVAgMDFgIBAh4BAheAAAoJEHUd4UM0
Pxo9L4gAoLF7+fqdnypCLA3Rxq2D7Ds5MZ1JAKDRmapQHc6vpKbjHigAeAfq0ASB
jLQZSXZhbiBIZXJtYW4gPGl2YW5AY3dpLm5sPoheBBMRAgAeBQJAK09SAhsDBgsJ
CAcDAgMVAgMDFgIBAh4BAheAAAoJEHUd4UM0Pxo9EQoAn2fmJP1xPZkrpYfWwcOS
tqJYNPl2AJwNNP91At9Bt1sJ6z13pyGdNJm4DbQgSXZhbiBIZXJtYW4gPEl2YW4u
SGVybWFuQGN3aS5ubD6IXgQTEQIAHgUCQCtPaQIbAwYLCQgHAwIDFQIDAxYCAQIe
AQIXgAAKCRB1HeFDND8aPfSLAKDPP9w7my3tPw6OnJRLY1UiBfctfQCeMgZdsN2t
kY5CFkr3qPpE6KRwtOG5Ag0EQCtNrxAIAKgi5qpe4XaSWH2qkrB+ddn/vgCDiKfq
zu4cBdMQ1mJh2KhNJVaZcSVlXLXQwdQW9BwJHsiTiOqqVsPSQMJ+kYjBIhLayqg7
TvFVTkrD6KxHxL8ce1P97vTczk/Hzk3mz3siITcilPRpFubEfVfsuxM6FxRfxPD+
brR4xFq391RN3TeiZyOpxzNNPCGKdA96alDTHNVophQvgosoiFKMBMpED2Pyw0JP
lFIZo8T+OMQbkhFD+A/M4O3YA57Zu6awvaxSZcDnbrXsemCwVpGJz6O7mCLPhsVC
mtaHfU6HJxnUxe0Cydf509kugqNMOwGfkAW12kwNlKQWoCZHBAlxdlcAAwcIAI+o
```

Humans can interpret labels ...

- A human understands that this is Ivan's encryption key (it is in the text!)
- She knows what she can do with it (e.g. add it to her keyring). Therefore to label a link “click here” is a usability and accessibility clash in most cases.

- On a Web of Data, something is missing; machines can make no sense of that link alone

Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

So for the Web of data ...

- extra information (“label”) must be added to a link: “this links to a GnuPG public key”.
- this information should be machine readable
- this is a characterization (or “classification”) of *both* the link *and* its target
- in some cases, the classification should allow for some limited “reasoning”

Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

Let us put together what we need for a Web of Data

- URI-s to publish data, not only full documents
- data can to link to other data
- the data and the links (the “terms”) should be characterized/classified to convey some extra meaning
- standards for all these to maintain interoperability

Copyright © 2008, W3C

Example: find the right experts at NASA

- NASA has nearly 70,000 civil servants over the whole of the US
- Their expertise is described in 6-7 databases, geographically distributed, with different data formats, access types…
- Task: find the right expert for a specific task within NASA!

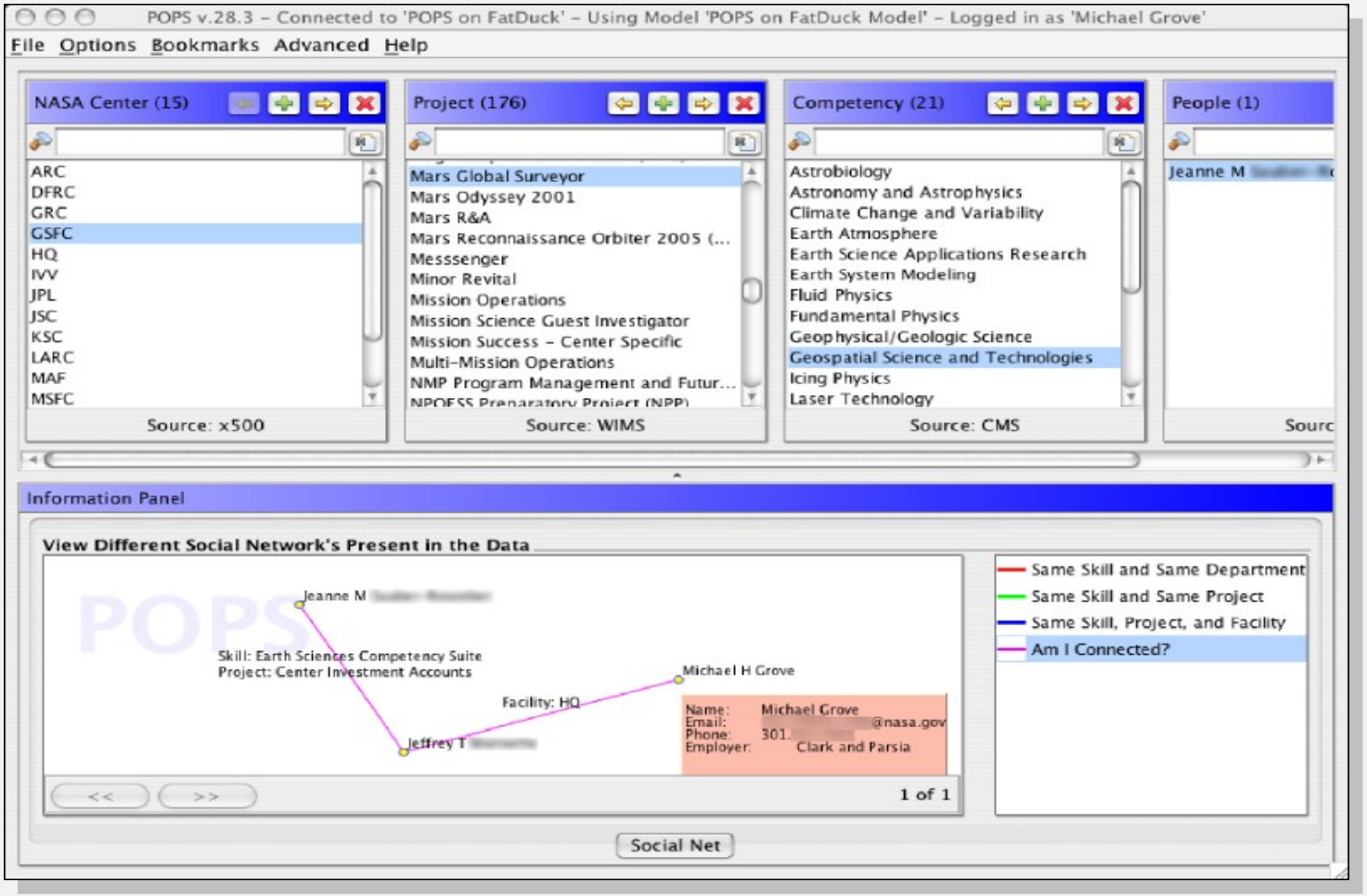
Michael Grove, Clark & Parsia, LLC, and Andrew Schain, NASA, (SWEO Case Study)

Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

Example: find the right experts at NASA

- Approach: integrate all the data with standard means, and describe the data and links using generic vocabularies



Michael Grove, Clark & Parsia, LLC, and Andrew Schain, NASA, (SWEO Case Study)

Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

So What *is* the Semantic Web?

Copyright © 2008, W3C

It is a collection of standard technologies to realize a Web of Data

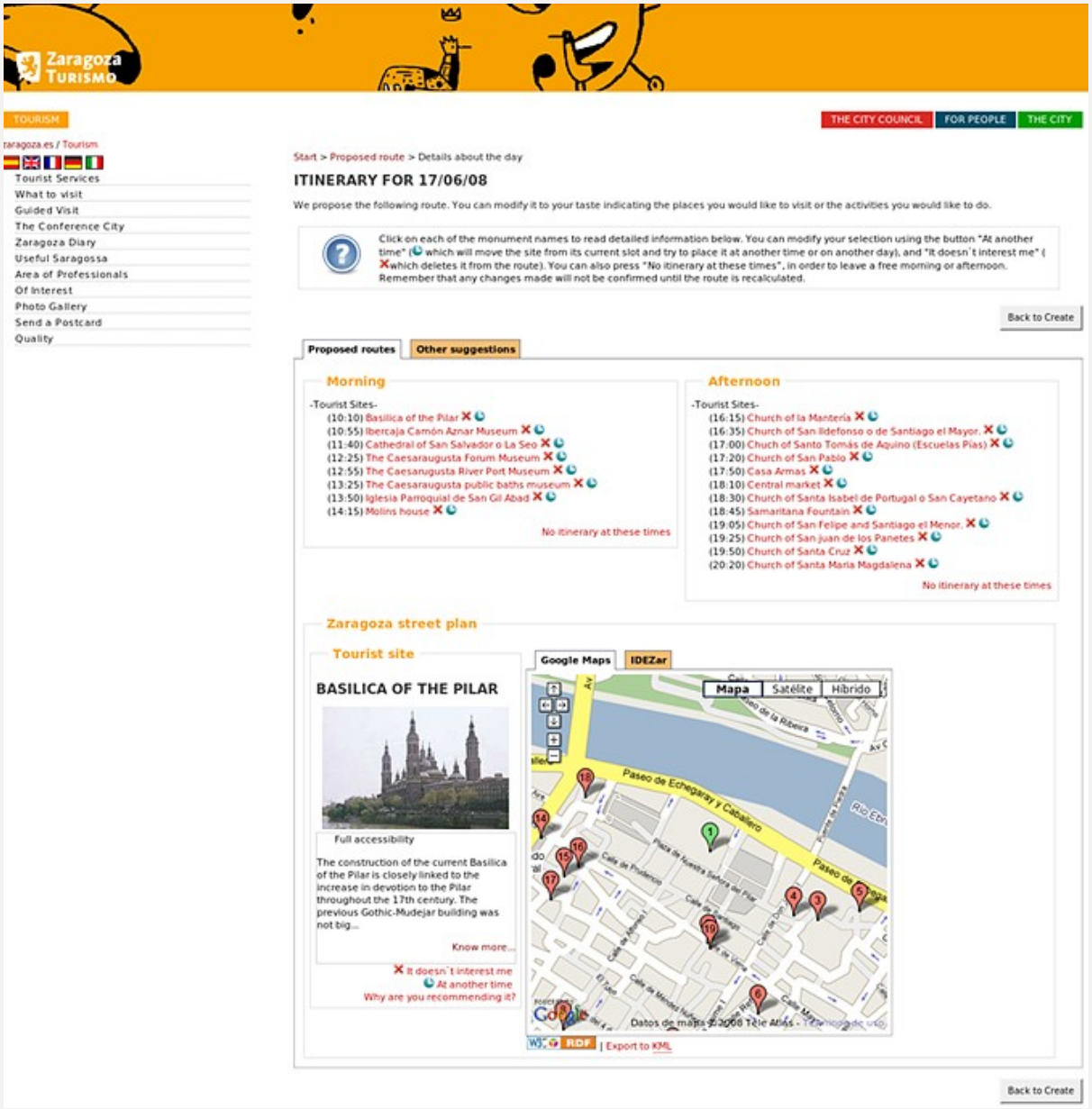
Copyright © 2008, W3C

It is that simple… but of f course, the devil is in the details

- a common model has to be provided for machines to understand the “labels” and draw some conclusions from that info
- the “classification” of the terms can become *very* complex for specific knowledge areas: this is where ontologies, thesauri, vocabularies, etc, enter the game…

Copyright © 2008, W3C

Example: eTourism in Zaragoza



- Provide personalized itinerary service
- Integration of different databases in Zaragoza (using targeted ontologies)
- Use rules on the data to provide a proper itinerary

Courtesy of Jesús Fernández, Municipality of Zaragoza, and Antonio Campos, CTIC (SWEO Use Case)

Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

Wait! Does it mean that you have to convert all your data in some way?

Copyright © 2008, W3C

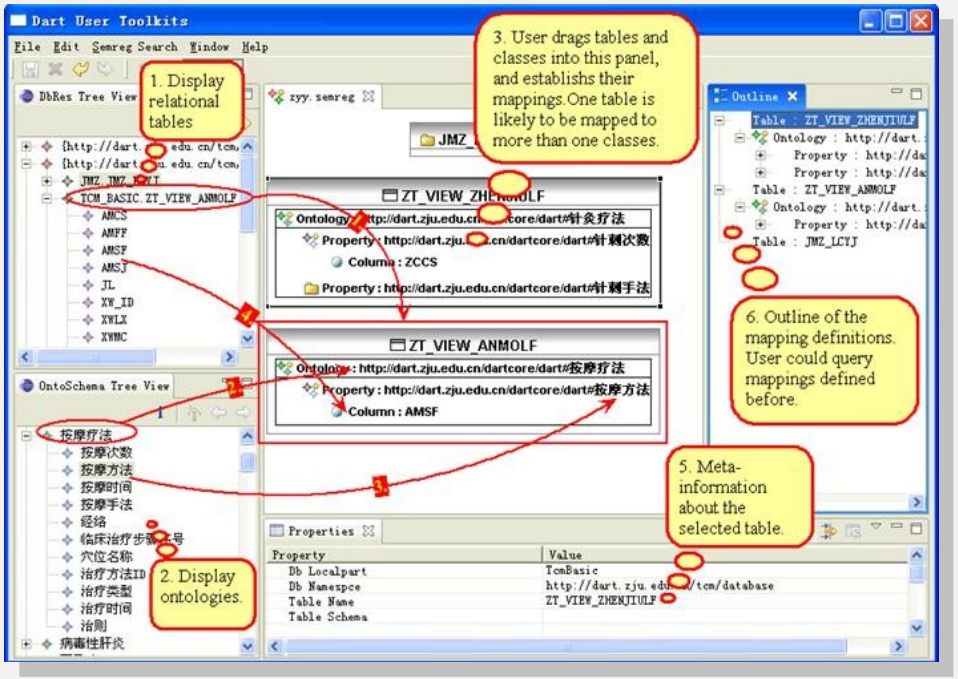
Convert your data?

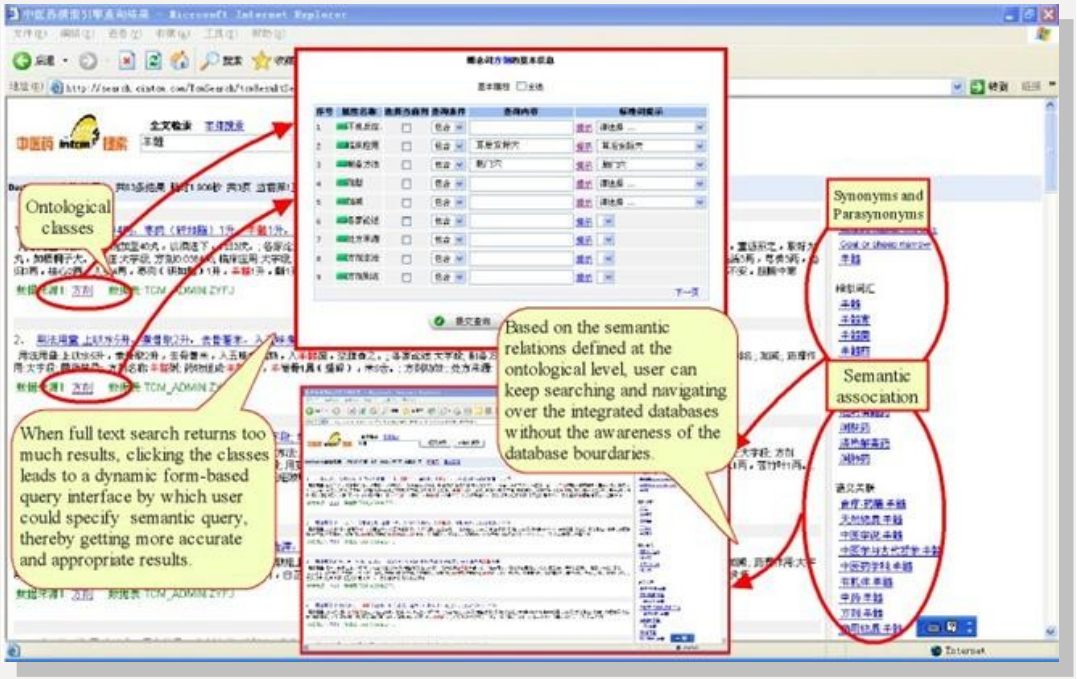
- Not necessarily; this would not always be feasible
- There are technologies to make your data accessible to standard means without converting it
  - run-time “bridges” (e.g. rewriting queries on the fly)
  - generate only entry points in separate datasets
  - annotate existing data (e.g. XHTML pages)
  - etc

Copyright © 2008, W3C

Example: integrate knowledge for Chinese Medicine

- Integration of a large number of TCM databases
  - around 80 databases, around 200,000 records each
- Uses specialized ontologies
- Queries are converted on-the-fly to the databases





Courtesy of Huajun Chen, Zhejiang University, (SWEO Case Study)

Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

In the end ...

- There is a huge potential for useful applications once we have a web of data
- Mash-ups give good examples for applications based on linking data
- When (re)designing a web site some thoughts on publishing data along with the documents might pay in the future.

Copyright © 2008, W3C

1º Fórum W3C Brasil
Por uma Web Única
W3C Semantic Web

- Slides are available at:
  http://www.w3.org/2008/Talks/0930SaoPaulo-KB-IH/
  in OpenDocument Presentation Format and PDF.

Copyright © 2008, W3C