



Data on the Web

Presented on

W3C Day + Webrebuild Anniversary,

2010 年 7 月 17 日在北京举

by Klaus Birkenbihl, Head World Offices, W3C

with some ideas from Ivan Herman,
Semantic Web Activity Lead, W3C

W3C on one slide

W3C <http://www.w3.org/> the home of (X)HTML, XML, CSS, RDF, the Web and Semantic Web ...

~330 members (the usual suspects, SMEs, users, grassroots ...) <http://www.w3.org/Consortium/Member/List>

65 groups doing the work

<http://www.w3.org/Consortium/activities>

18 world offices <http://www.w3.org/Consortium/Offices/>

a team of ~66 individuals - working and living distributed around the globe - coordinated by 3 hosts: MIT, US <http://www.csail.mit.edu/>, ERCIM, Europe <http://www.ercim.org/>, Keio University, Japan <http://www.keio.ac.jp/>

director: WWW inventor Tim Berners Lee

<http://www.w3.org/People/Berners-Lee/>

we reach the community (4Mio hits/day on <http://www.w3.org/>)



Contents

Something the Web of documents is missing

A use case example

A taste of data

Same use case, another example

A better way to provide and access data on the Web:
RDF

RDF (Data) within HTML: RDFa

RDFa in the wild

Some examples

Epilogue

For my trip to Beijing I had to:

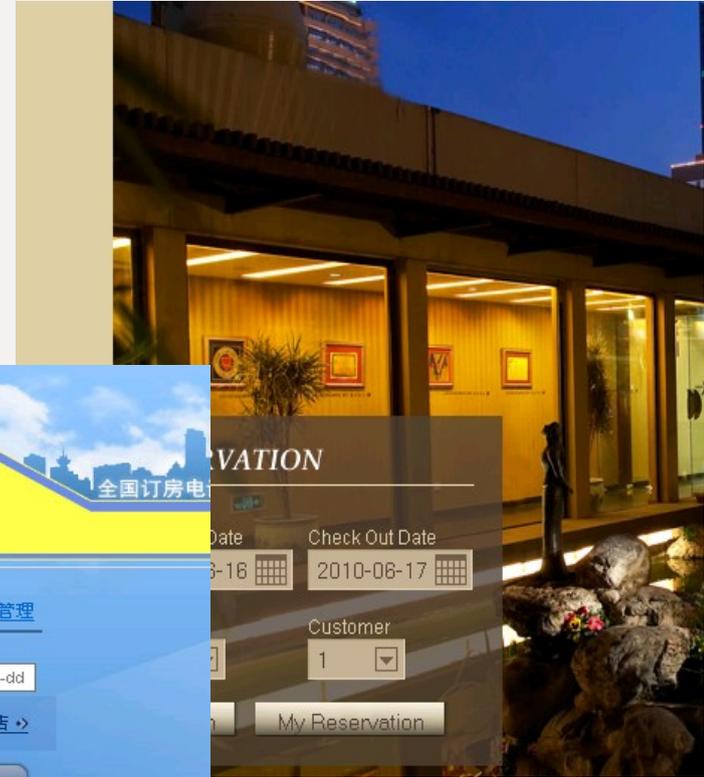
The image shows a collage of airline websites. On the left, the Lufthansa website is visible with the text "There's no better way to fly" and "Buchung" (Booking) buttons. In the center, the Air China website is shown with the text "AIR CHINA 中国国际航空公司" and "Buchung" (Booking) buttons. To the right, the China Southern website is visible with the text "中国南方航空 CHINA SOUTHERN" and "CSN News" section. Below that, the Emirates website is shown with the text "Emirates" and "Kinder reisen kostenlos nach" (Children travel free to). On the far right, the KLM website is shown with the text "KLM Royal Dutch Airlines" and "Ihren Flug buchen" (Book your flight) section. The KLM website also shows flight details: "Von Berlin - Tegel Airport (TXL)" and "Abflugdatum 23/6/2010".

find a flight

check with several airlines

remember their answers and offers

Same for hotels



[Well not really]

Virtual travel agencies

For flights

Or complete arrangements

Again it is up to me to

I only see what they want to show me

flug.ideal.de
Reise-Preisvergleich | Flug-Preisvergleich | Flughäfen | Flüge

Idealo.de > Reise > Flug-Preisvergleich

Top Shops

- GOVODO
- fluege.de
- fly.de
- opodo
- tripadvisor
- flugladen.de
- airline direct
- Elumbus

...mehr Flugschors

Idealo-Service

Live-Flugsuche
Idealo News

Top Flug-Routen

- München - Düsseldorf - ab 151.00 €
- Düsseldorf - Wien - ab 128.91 €
- Hannover - Wien - ab 92.68 €
- Frankfurt-Main - Budapest - ab 87.00 €
- München - Ankara - ab 56.00 €
- Stuttgart - Izmir - ab 307.08 €
- Düsseldorf - Dresden - ab 220.65 €
- Frankfurt-Main - Berlin - ab 138.60 €
- Frankfurt-Main - Bergamo - ab 87.00 €
- München - Johannesburg - ab 508.81 €

Top Fluglinien

- airberlin
- Lufthansa
- germanwings
- TURKISH AIRLINES
- AIR FRANCE
- BRITISH AIRWAYS
- KLM
- Austrian

fluege.de Alle Airlines und Veranstalter in

Flug Hotel Last Minute Reisen Mietwa

So einfach mit einem

Wirklich alle Flüge im Vergleich!

Schnellsuche

Flug Hotel Mietwagen Lastminute

flug.de

www.discount-flug.de
Online suchen und buchen mit Sofortbestätigung

Home Flug Flugspecials

- Europa
- Nordamerika
- Asien
- Australien & Ne
- Lateinamerika

Verfügbarkeit Persönliche Angaben Bestätig

gewünschten Reisertermin eingeben und weiter klicken - wir nennen Ihnen Flugpreis plus Steuern und Sie ONLINE buchen.

Expedia.de

Home Flug Hotels Click & Mix

opodo und die Reise beginnt

Schon länger nicht bei uns gebucht?

Erstellen Ihrer Reise

Wir vergleichen für Sie die günstigsten Reisen aller großen Veranstalter

Nur Flug
Nur Hotel
Mietwagen

Flug & Hotel
Flug & Hotel & Mietwagen

Zur Hotel-Suche

Ihre Vorteile bei Opodo

- 100.000 Hotels & Pensionen weltweit
- Urlaub, Business, Design, Strand...
- Landkarten & Umkreissuche

Top-Angebote Städtereisen

tripadvisor

BEST DESTINATIONS OF 2010

Home Hotels Flights Restaurants Vacation Rentals

City, hotel name, etc. Search

Plan the Perfect Trip

- Hotels
- Flights
- Restaurants
- Things to Do
- Cruises
- Vacation Rentals

Find Hotels Travelers Trust

City:

Check-in: 6/18/2010
mm/dd/yyyy

Check-out: 6/20/2010
mm/dd/yyyy

Adults: 2

Find Hotels

And 100s more ...

What happened here?

I had to consult a large number of sites, all different in style, purpose, possibly language...

I had to mentally *integrate* all this information to achieve my goals

We all know that, sometimes, this is a long and tedious process!

What would we like to have?

We would like to have applications that can combine all the data in the different Web sites (or underlying databases) in a useful way.

This would require that the applications can access the data

This would require that the data can be linked like Web pages today

Or put in another way:

We would like to *extend* the current Web with a *“Web of data”*:

allow for applications to exploit the data directly

Contents

Something the Web of documents is missing

A use case example

A taste of data

Same use case, another example

A better way to provide and access data on the Web:
RDF

RDF (Data) within HTML: RDFa

RDFa in the wild

Some examples

Epilogue

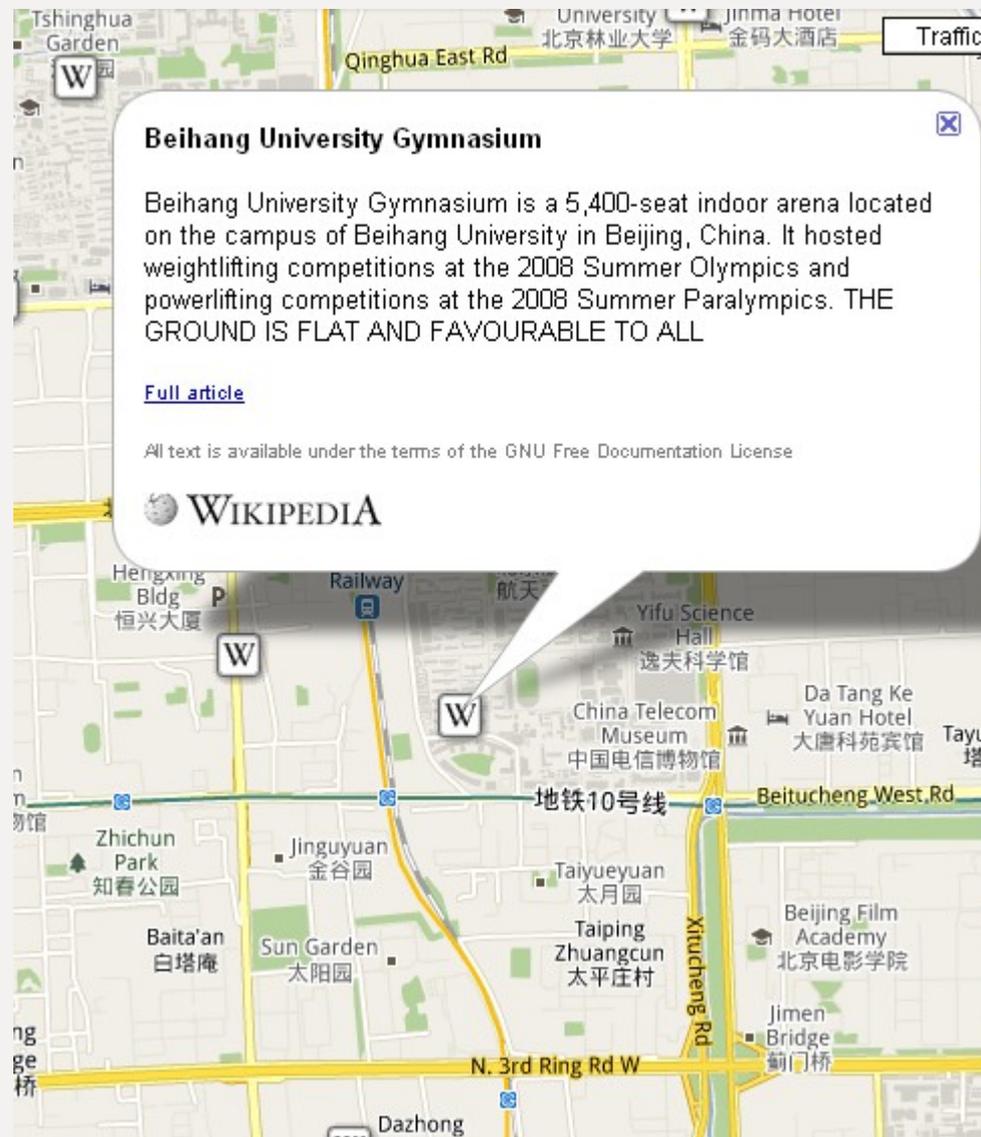
Example: Mash-up

Can't mash-ups solve do the integration?

Let's have a closer look on a mash-up application (no, not a map application 😊)

A trip organizer called *Tripit*

- Builds a nice itinerary from various sources
- Not much effort to use
- Includes some social networking capabilities



Map mash-ups are rather popular on the Web



China
Jul 8 - Jul 23, 2010 - Beijing, China
Beijing, China

Share Print
Export to calendar
Add plans

Itinerary: **Expand** | Collapse

Thu, Jul 8 Beijing, China - Avg: Hi 31°C / Lo 21°C + Add Plans

10:42 DB from Siegburg/Bonn to F-Flughafen Fernbf. Options
CST

DB 08 Conf # P1RWFD
F-Flughafen Fernbf., ICE 623

Depart: 8/7/2010, 10:42 CST, Siegburg/Bonn
Arrive: 8/7/2010, 11:34 CST, F-Flughafen Fernbf.

Booking Information
Booked on DB BAHN 20/5/2010
Reference #: P1RWFD
<http://www.bahn.de/>

Notes
Klasse: 2

14:50 Flight from Frankfurt am Main (FRA) to Beijing (PEK) Options
CEST

Check Flight Status
Track for refunds

Air China 966
Aircraft Airbus A330
nonstop 9h, 40m 7,786 km Economy Saver/Economy - [Get seating advice](#)

Depart: Frankfurt am Main (FRA), 14:50 CEST
Arrive: Beijing (PEK), 06:30 CST(+1 day)

Passenger
Birkenbihl Klaus

Fri, Jul 9 Beijing, China - Avg: Hi 31°C / Lo 21°C + Add Plans

06:40 Map of Beijing, China Options
CST



People and sharing

Who's close
No Connections close

People
Travelers: [Klaus Birkenbihl](#)
Non-travelers: None.

Social networks
[Tript Contacts](#), [Facebook](#) and

This trip is not private. [Change](#)

Comments

Sat, Jul 17 Beijing, China - Avg: Hi 31°C / Lo 21°C + Add Plans

09:00 W3C Day Options
CST

17/7/2010, 09:00 - 17:00

Sun, Jul 18 Beijing, China - Avg: Hi 31°C / Lo 21°C + Add Plans

Mon, Jul 19 Beijing, China - Avg: Hi 31°C / Lo 22°C + Add Plans

Tue, Jul 20 Beijing, China - Avg: Hi 31°C / Lo 22°C + Add Plans

Wed, Jul 21 Beijing, China - Avg: Hi 31°C / Lo 22°C + Add Plans

Thu, Jul 22 Beijing, China - Avg: Hi 31°C / Lo 22°C + Add Plans

Fri, Jul 23 Frankfurt Am Main, Germany - Avg: Hi 22°C / Lo 14°C + Add Plans

14:00 Flight from Beijing (PEK) to Frankfurt am Main (FRA) Options
CST

Check Flight Status

Air China 931
Aircraft Boeing 747-400 Combi
nonstop 10h, 10m 7,786 km Economy Saver/Economy - [Get seating advice](#)

Depart: Beijing (PEK), 14:00 CST
terminal 3
Arrive: Frankfurt am Main (FRA), 18:10 CEST
terminal 1

Passenger: Birkenbihl Klaus. See [related booking](#) on 8/7/2010 for booking information for this flight.

19:09 DB from F-Flughafen Fernbf. to Siegburg/Bonn Options
CEST

DB 514 Conf # P1RWFD
ICE, Fern 6, Coach #33, Seat(s): 101

Depart: 23/7/2010, 19:09 CEST, F-Flughafen Fernbf.
Arrive: 23/7/2010, 19:47 CEST, Siegburg/Bonn

Passengers: See [related booking](#) on 8/7/2010 for booking information for this reservation.

What I do – what *Tripit* does

I open a trip with *Tripit* by giving it a name, a start and an end date

I book my flights and forward the confirmations to *Tripit*

I book my trains and forward the confirmations to *Tripit*, etc. ...

DB BAHN Online-Ticket
Bitte auf A4 ausdrucken

ICE Fahrkarte
Gültigkeit: **Hin- und Rückfahrt ab 08.07.2010, Rückfahrt ab 23.07.2010**
DB: Gilt nur in gebuchten Zügen an angegebenen Reisetagen. Gilt nur auf eingetragener Strecke im NV (S/RB/RE/IRE) vor/nach den gebuchten Zügen am 1. Tag und Folgetag bis 10 Uhr.

Sparpreis (Hin- und Rückfahrt)
Klasse: **2**
Erw. Preis: **1, mit 1 BC25**

Hinfahrt: **Siegburg/Bonn -> Frankfurt(M)Flugh.**, mit ICE, (SP-R0529)
Rückfahrt: **Frankfurt(M)Flugh. -> Siegburg/Bonn**, mit ICE, (SP-R0249)

Über: **H: Siegbg 10:42 ICE623 R: F-Flugh 19:09 ICE514**
Umtausch/Erstattung 15 EURO; ab dem 1. Geltungstag ausgeschlossen

Zahlungspositionen und Preis				
Positionen		Preis	MwSt (D) 19%	MwSt (D) 7%
ICE Fahrkarte	1	58,50€	58,50€	9,34€
Reservierung	1	2,50€	2,50€	0,40€
Summe		61,00€	61,00€	9,74€

Kreditkartenzahlung

Hinfahrt: 208H YEKT N48
Zertifikat: 08.07.2010
Gültig ab: 08.07.2010

Rückfahrt: 20MX JADT 5TB
Zertifikat: 23.07.2010
Gültig ab: 23.07.2010

AIR CHINA 中國國際航空公司

Buchungen | Meine Reise | Angebote | PhoenixMiles | Kundenservice

Schritte: Suche Daten Flüge Überprüfung Reisende Kauf Reservierung

Ihre Reisereservierung

Reservierungsnummer(n) der Fluggesellschaft: **Air China C9SSE**
Reisestatus: **Bestätigt**

- Ihre Reise wurde bestätigt, und die Reservierung ist garantiert.
- Wir empfehlen Ihnen, sich die Reservierungsnummer zu notieren oder diese Seite zu drucken.
- Die Bestätigung wird per E-Mail versandt.

Angaben zum Reisenden

Herr Birkenbihl Klaus
Vielflieger: CA 992004656029446

Informationen zum Reiseziel - Deutschland
Reisedokument: Deutschland Reisepass C79TPXG72: Gültig bis Sonntag, 14. April 2019.: Nationalität
Deutschland: Männlich: Geboren am Donnerstag, 22. April 1948
Zieladresse: XUEYUAN ROAD, BEIJING, CHINA, 100083
Heimatadresse: ZEDERNWEG 85, SANKT AUGUSTIN, DEUTSCHLAND, 53757

What I do – what *Tripit* does

I can add events and meetings to the itinerary manually

Either by default or on demand *Tripit* adds other useful information

Nearly no effort – nice itinerary – easy to modify – friends stay informed – *cool!*

Weather for Beijing, China - [Add to iGoogle](#)

31°C | °F
 Current: **Clear**
 Wind: W at 11 km/h
 Humidity: 38%

Mon	Tue	Wed	Thu
33°C 20°C	34°C 19°C	33°C 19°C	34°C 17°C

[Start search over](#)
[Back to search results](#)

[Hotel Information](#) >

[Expedia.co.uk](#) > [Hotels](#) > [China](#) > [Beijing \(region\)](#) > [Beijing Hotels](#) > [Vision Hotel](#)

Vision Hotel
 Hotel in Beijing's Haidian with an indoor pool
 ★★★★★
 Hotel Class

How it works

I forward to *Tripit* the documents with the information related to a trip. e.g.

Flight bookings

Hotel reservations

Meetings

Any time I have new documents I may add them (either by e-mail or by providing a URI)

Tripit tries to extract the data from the information

Based on the dates it associates the documents to a trip

It adds information from other sites about weather, directions, travel guides ...

It checks its own database for travel activities of my friends

It connects with social networking sites to inform my friends

It compiles a structured itinerary

As I said it is cool ... but

Sometimes *Tripit* sends me a message “**Problem with your TripIt submission**” and does nothing though the information was delivered

Sometimes some information within a document is not used

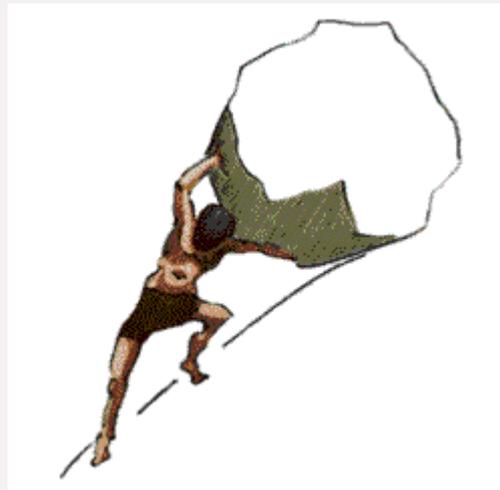
Sometimes I read: “**Please help us to improve! Let us know how good we captured your flight.**”

This gives a hint on what *Tripit* does: in case it does not know exactly how to find the data, it **guesses**

Because it has no standardized way to access the data, *Tripit* has to use proprietary interfaces or – even worse – has to scrape the text for data, and follow the all the changes – for all the many sources . *Greetings from Σίσυφος(Sysiphus) to the programmers!*

Excursion to greek mythology: *Σίσυφος*

Sinner condemned in Tartarus to an eternity of rolling a boulder uphill then watching it roll back down again. Sisyphus was founder and king of Corinth, or Ephyra as it was called in those days. He was notorious as the most cunning knave on earth. He even tricked Hades the god of death.



Sisyphus work: trying hard - but never finally succeeding. Moving in circles instead.

Lesson learned

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

Put it another way (again)...

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

Contents

Something the Web of documents is missing

A use case example

A taste of data

Same use case, another example

A better way to provide and access data on the Web:
RDF

RDF (Data) within HTML: RDFa

RDFa in the wild

Some examples

Epilogue

What does this mean?

What makes the current (document) Web work?

people create different documents

they give a globally unique address to it (i.e. a URI) and make it accessible to others on the 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...)

Can this be used 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 URIs to the data*

make it possible to “link” to that URI from *other* sources of data (not only Web pages) using standard approaches

- i.e., 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...

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 e.g. if you read on a Web page

--- [Impressum](#) --- ...

you can guess that the link labelled “Impressum” leads you to the registration information of that Company (well some Latin required).

It all only works in a meaningful way if you, the human, can make correct and meaningful assumptions about the link.

Machines cannot interpret labels ...

Something is missing in our model for the web of data!

extra information (“label”) must be added to a link:
“this links to an Impressum information”.

this information should be machine readable

this label is a characterization (or “classification”) of
both the link *and* its target

in some cases, the classification should allow for
some limited “reasoning”

Summary: requirements for a Web of Data?

URI-s to publish data, not only full documents

data can 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

The RDF data model

Short introduction

RDF consists of labelled connections between two resources

Let's call the first resource *subject* (*s*), the second resource *object* (*o*) and the labelled connection *predicate* (*p*)

s and *p* are represented by URIs. Same URI denotes the same resource. *o* can be either a URI or a literal denoting a value

An RDF triple (*s,p,o*) can be represented by simple graph.

E.g.: the triple

`(<http://ex.com/fritz>, <http://xmlns.com/foaf/0.1/knows>, <http://ex.com/karl>)`

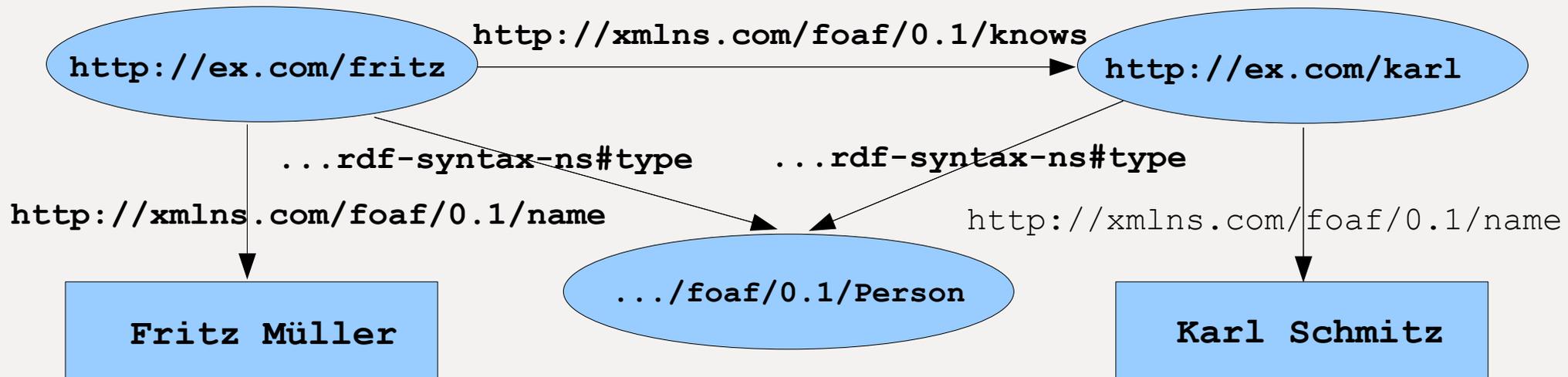


RDF (continued)

There are several way to serialize RDF. We will use a subset of turtle (<http://www.w3.org/TeamSubmission/turtle/>).

```
@prefix p: <http://ex.com/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
p:fritz rdf:type foaf:Person .
p:karl rdf:type foaf:Person .
p:fritz foaf:knows p:karl .
p:fritz foaf:name "Fritz Müller" .
p:karl foaf:name "Karl Schmitz" .
```

Denotes the following graph:



It is that simple...

but of course, the devil is in the details

the “classification” of the terms can become very complex for specific knowledge areas: this is where ontologies, thesauri, vocabularies, etc, enter the game...

W3C has developed a set of standards for this

- RDF – the Resource Description Framework
- OWL – the Web Ontology Language (based on RDF)
- SPARQL – a Query language for the Semantic Web

and a few more that make it easier to use

There are connected data on the Web

Connect Distributed Data across the Web

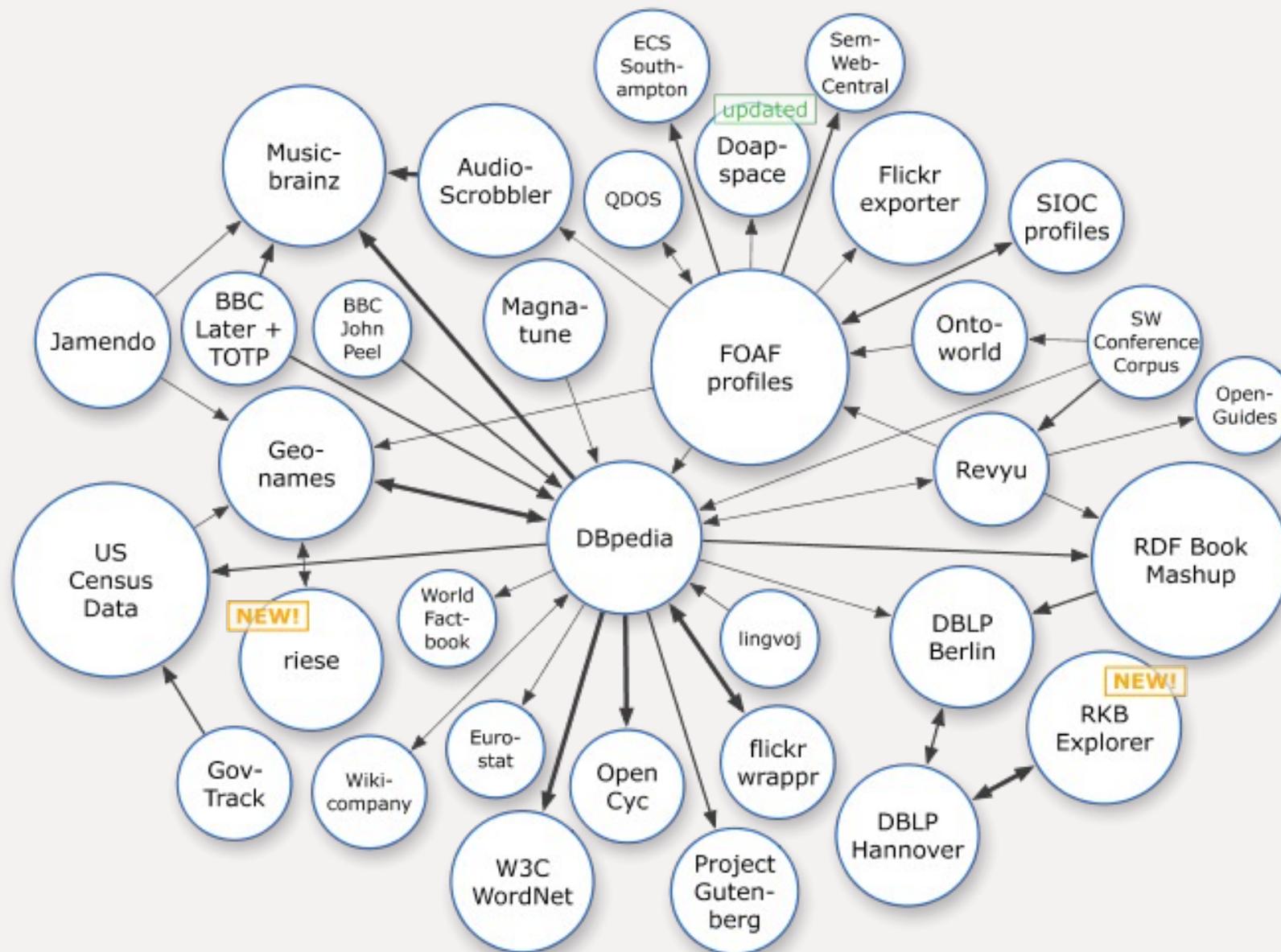
Linked Data

Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. More specifically, Wikipedia defines Linked Data as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF."

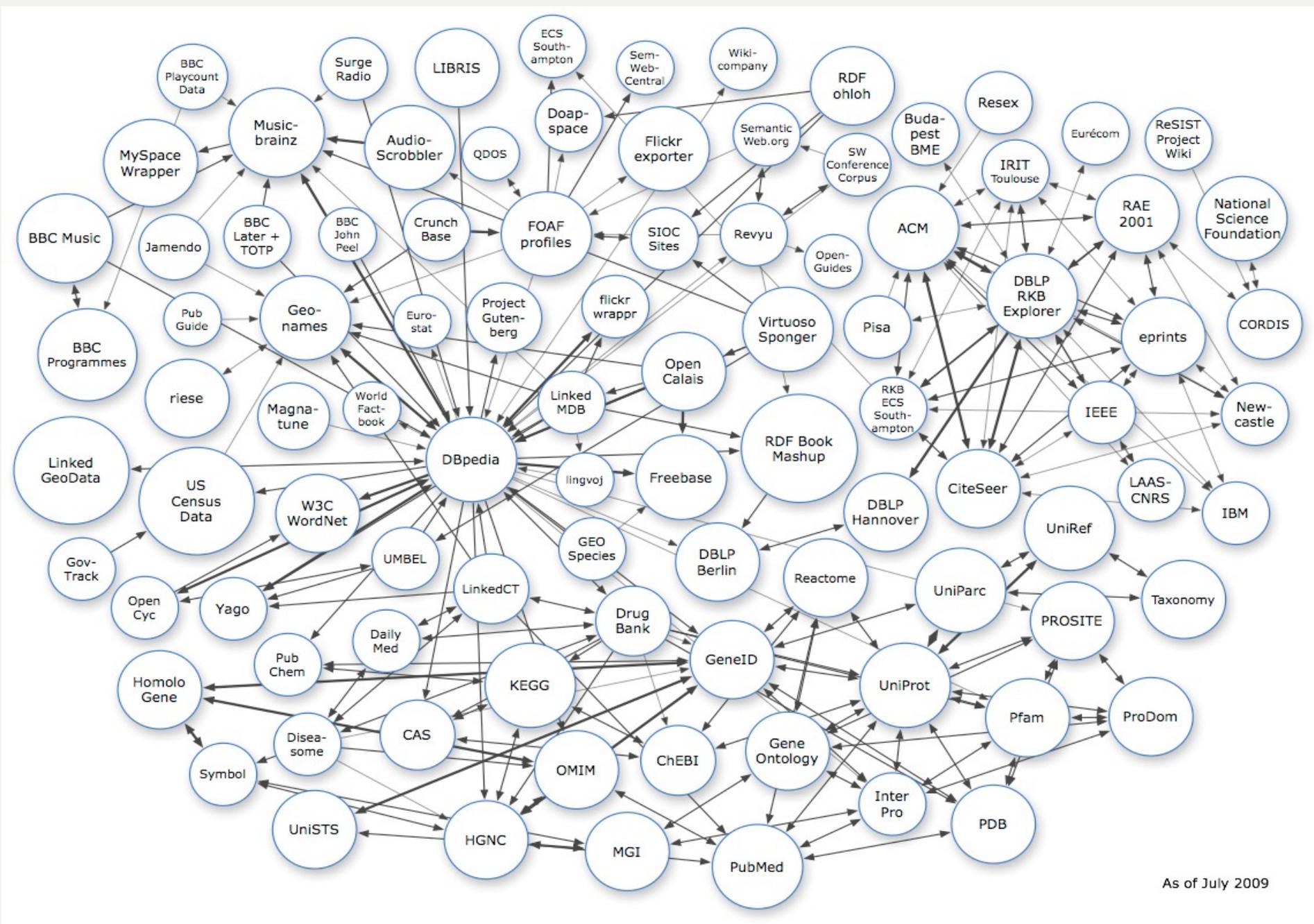
This site exists to provide a home for, or pointers to, resources from across the Linked Data community.



2007: A lot of people publish their data ...



... and link it in the LOD project: 2009



Contents

Something the Web of documents is missing

A use case example

A taste of data

Same use case, another example

A better way to provide and access data on the Web:
RDF

RDF (Data) within HTML: RDFa

RDFa in the wild

Some examples

Epilogue

So what does it mean for a Web developer? ⁽³¹⁾

Questions

Is there any benefit for a common Website today?

How can such the benefit be achieved?

How to combine my Web page with RDF?

Let's focus on embedding RDF into HTML first (the next chapter will show possible impact on searching or social networking)

RDFa

Straight forward way to add RDF to XHTML by reusing some existing attributes and adding a few new ones

RDFa attributes:

Existing ones:

- `rel`
- `rev`
- `content`
- `href`
- `src`

New ones

- `about`
- `property`
- `resource`
- `datatype`
- `typeof`

Suggested reading:

<http://www.w3.org/TR/xhtml-rdfa-primer/> or, in Chinese

<http://iws.seu.edu.cn/resource/Translations/RDFa-Primer-Simplified-Chinese.htm>

Other standards for data in HTML: eRDF, microformats, microdata. All can be easily transformed to RDF. GRDDL is another way to extract data from well structured HTML

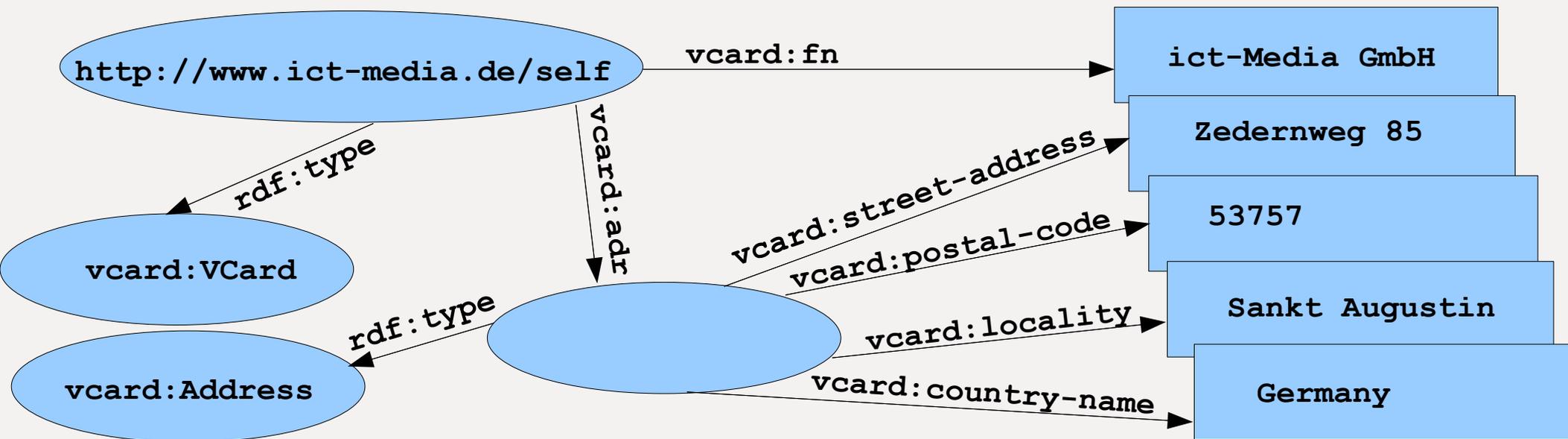
RDFa example

```

<span typeof="vcard:VCard" about="http://www.ict-media.de/self">
  <span property="vcard:fn">ict-Media GmbH</span><br />
  <span rel="vcard:adr">
    <span typeof="vcard:Address">
      <span property="vcard:street-address">Zedernweg 85</span><br />
      <span property="vcard:postal-code">53757</span>
      <span property="vcard:locality">Sankt Augustin</span><br />
      <span property="vcard:country-name" content="Germany"/>
    </span>
  </span>
</span>

```

Produces html rendering as usual:
 + the following RDF graph



RDFa support

Some content management systems (e.g. Drupal) support authoring of RDFa – others can embed it easily into the templates

Browser plug-in or extensions are available (e.g. Firefox Operator)

Online RDFa Checker

Quite some parsers



The screenshot shows the 'checkrdfa' website interface. At the top, there is a navigation bar with 'Home', 'Results', and 'FAQ' links. Below the navigation bar, the search results are displayed for the URL '<http://www.ict-media.de/>'. The 'Results' tab is active, and the 'Facebook' sub-tab is selected. The main content area shows the 'Facebook / Open Graph Protocol' results for the URL. The results are listed as follows:

- `og:country-name`: Germany
- `og:description`: ict-Media GmbH berät zu Web-Standards und Web-Technologien
- `og:email`: info@ict-media.de
- `og:fax_number`: (empty)

Contents

Something the Web of documents is missing

A use case example

A taste of data

Same use case, another example

A better way to provide and access data on the Web:
RDF

RDF (Data) within HTML: RDFa

RDFa in the wild

Some examples

Epilogue

RDFa and **facebook**

Facebook uses RDFa for its *Open Graph Protocol*

A very simple (and flat) vocabulary

Only one subject per Web page

All values (objects in RDF terminology) are literals (strings)

@prefix og: <<http://opengraphprotocol.org/schema/>>

Properties of *OGP*:

og:title, og:type, og:image, og:url,
og:description, og:site_name, og:latitude,
og:longitude, og:street-address, og:locality,
og:postal-code, og:country-name, og:email,
og:phone_number, og:fax_number, og:upc, og:isbn

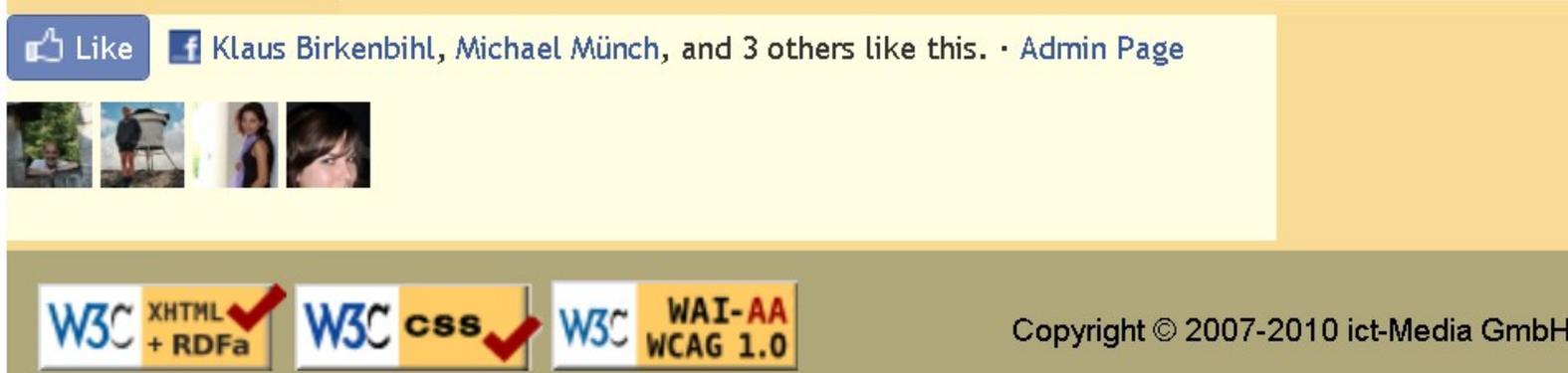
RDFa and facebook

Allowed values for og:type:

activity, sport, bar, company, cafe, hotel, restaurant, cause, sports_league, sports_team, band, government, non_profit, school, university, actor, athlete, author, director, musician, politician, public_figure, city, country, landmark, state_province, album, book, drink, food, game, movie, product, song, tv_show, article, blog, website

Websites within the *Open Graph* will be treated like social objects within Facebook when registered

A *Like* button allows you to indicate that you like the subject represented by the Web site. Your *friends* will be informed



RDFa and **facebook**

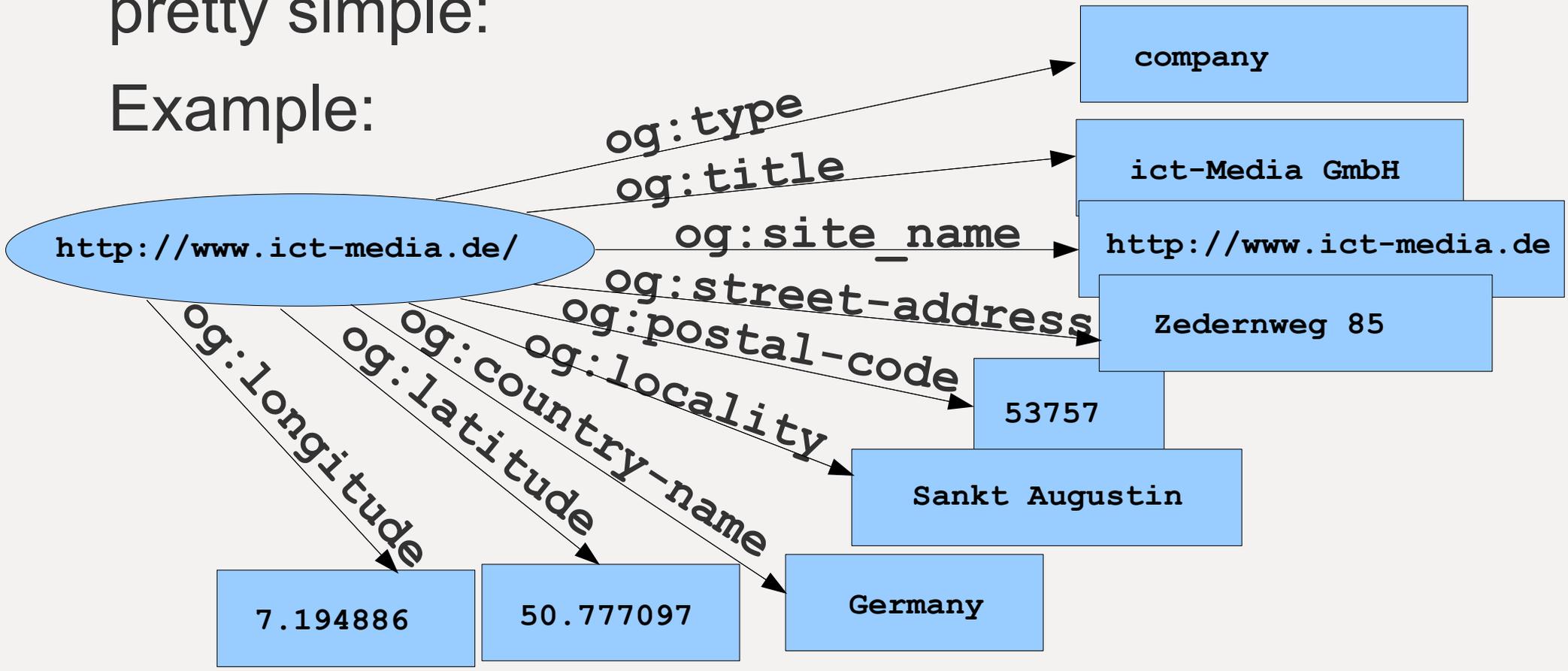
Data provided on a Web site can be used in Facebook applications in the same way as data from Facebook internal social objects is used

The page becomes a node on Facebook's social graph

RDFa and **facebook**

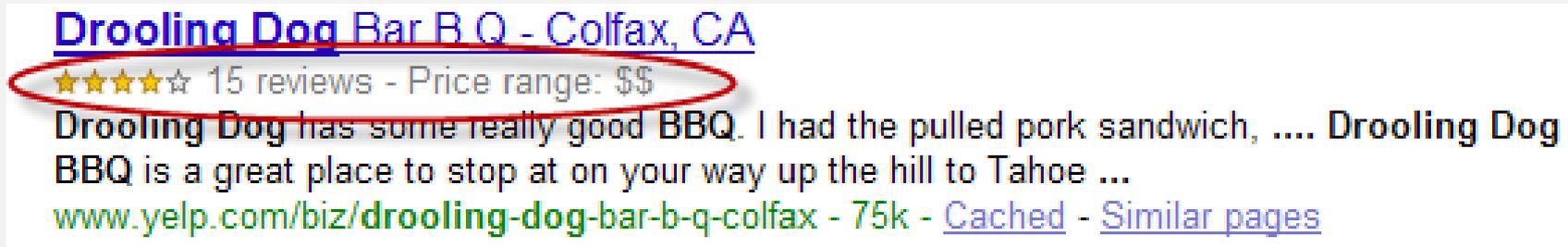
The RDFa graph for Facebook looks (today) pretty simple:

Example:



RDFa and Google

Google is introducing what they call *rich snippets*.



[Drooling Dog Bar B Q - Colfax, CA](#)
★★★★☆ 15 reviews - Price range: \$\$
Drooling Dog has some really good BBQ. I had the pulled pork sandwich, Drooling Dog BBQ is a great place to stop at on your way up the hill to Tahoe ...
www.yelp.com/biz/drooling-dog-bar-b-q-colfax - 75k - [Cached](#) - [Similar pages](#)

“With rich snippets, webmasters with sites containing structured content—for example, review sites or business listings—can label their content to make it clear that each labeled piece of text represents a certain type of data ...”

Currently only for review sites and social networking/people profile sites. Other types of content in the future

Moving forward with caution

RDFa and Google

Formats: microdata, microformats, or RDFa.

Started with its own vocabulary

<http://rdf.data-vocabulary.org/>

Meanwhile supports other popular vocabularies like foaf or vCard

Still sort of experimental – but with a high potential

Google search preview

The following errors were found during preview generation:

- Insufficient data to generate the preview.

Extracted Rich Snippet data from the page

Organization

Address

street-address = Zedernweg 85
 locality = 53757 Sankt Augustin
 region = Germany

Geo

latitude = 50.777097
 longitude = 7.194886

url = <http://www.ict-media.de/>

name = ict-Media GmbH

tel = +49 2241 396415

VCard

Warnings : RDFa tree has unknown type 'VCard'.

Address

street-address = Zedernweg 85
 locality = Sankt Augustin
 postal-code = 53757
 country-name = Germany

geo

latitude = 50.777097
 longitude = 7.194886

url = <http://www.ict-media.de/>

tel = +49 2241 396415

SearchMonkey roles:

Developers develop search applications

Site owners provide data (e.g. using RDFa)

Users can register for certain SearchMonkey applications

All applications are sort of Open Source (Yahoo licence)

The SearchMonkey application gallery holds a lot of examples.

Developers can submit their applications for the Search Gallery

Developers can also offer the use of their application on their Website

	<p>Flickr Developed by: Yahoo! Search Flickr loves Search! View fun, interesting photos from other Yahoo! users for image-related queries.</p>
<p>Official</p>	
	<p>Babelfish Translations Developed by: Marco V This app displays links to translate the given page to multiple languages using Babelfish.</p>
	<p>Feed Discovery Developed by: 老朱 Extract posts from the first embedded feed in every search results if any</p>
	<p>Blogspot Recent Posts Developed by: Marco V This app displays links to the 10 most recent posts on a Blogspot blog. For each post, it shows the date and title.</p>
	<p>deliciousify Developed by: anand kishore enrich your search results with metadata from delicious. - View number of bookmarks - View top tags as a color coded tag cloud</p>
	<p>Best Answer Developed by: Marco V This app displays the best answer and rating from the Yahoo Answers question, as well as the user's name and profile link.</p>

YAHOO! SearchMonkey

SearchMonkey applications are pieces of PHP that define how – for a set of URIs – the search result are displayed.

SearchMonkey gives special support for some data “Objects”.

Data can be specified using popular vocabularies like *vcard*, *Dublin Core*, *foaf* and private ones like

`http://search.yahoo.com/searchmonkey/`

SearchMonkey applications go beyond use of data within the page (e.g. you can enhance a result with URLs for machine translations etc.)

Example search result with custom application:

[ict-Media GmbH](#)

[万维网 \(cn\)](#) | [Das Web \(de\)](#) | [The Web \(en\)](#) | [La Web \(es\)](#)

- Straße: Zedernweg 85
- Ort: 53757 Sankt Augustin
- Land: Germany
- Telefon: +49-2241-296415

 www.ict-media.de/index.html - [Cached](#)



SearchMonkey > Start Overview > Object Finder

Testing Markup



DATA NOT CACHED YET

The data stored in the Yahoo! Search Index doesn't match the data just retrieved from the live version. This is normal if you just changed your page — simply wait for us to recrawl and reindex your page. The average update time is two weeks for most URL. Here is how your page [currently appears in Yahoo! search results](#).



MULTIPLE OBJECTS

You have multiple objects on your page. Unfortunately, SearchMonkey currently only supports displaying one object per page, so no enhanced result will display for this URL. You can remove all objects except for the one you wish to display, or wait for SearchMonkey to support displaying multiple objects.



Event

RECOMMENDED

vcal:summary	Some Steps from the Web to a Semantic Web
vcal:dtstart	2010 -05-07

OPTIONAL

vcal:location	New Delhi, India
----------------------	------------------

MESSAGES

Info	Data Not Cached Yet This object data is not cached in the Yahoo! Search Index. This is normal if you just changed your page — simply wait for us to recrawl and reindex your page.
-------------	---



Event

RECOMMENDED

vcal:summary	The "what" and "how" of W3C
---------------------	-----------------------------

Example from <http://www.ict-Media.de/talks.html>

Included from: <http://www.w3.org/Talks/>



Local

RECOMMENDED

vcards:tel	+49 2241 396415
vcards:fn	ict-Media GmbH

OPTIONAL

vcards:street-address	Zedernweg 85
vcards:locality	Sankt Augustin
vcards:postal-code	53757
vcards:country-name	Germany

from <http://www.ict-Media.de/>

Also ...

Not only data users (like search engines) but also information providers use more and more RDFa on their Web pages

Famous RDFa providers: NewsWeek, TESCO, O'Reilly Catalog, Best Buy, Public Library of Science ...

There are many tools around for developers. (RDFa checker and RDFa parser)

Many Wikis, Blogs or CMSs can easily be configured to support RDFa

Contents

Something the Web of documents is missing

A use case example

A taste of data

Same use case, another example

A better way to provide and access data on the Web:
RDF

RDF (Data) within HTML: RDFa

RDFa in the wild

Some examples

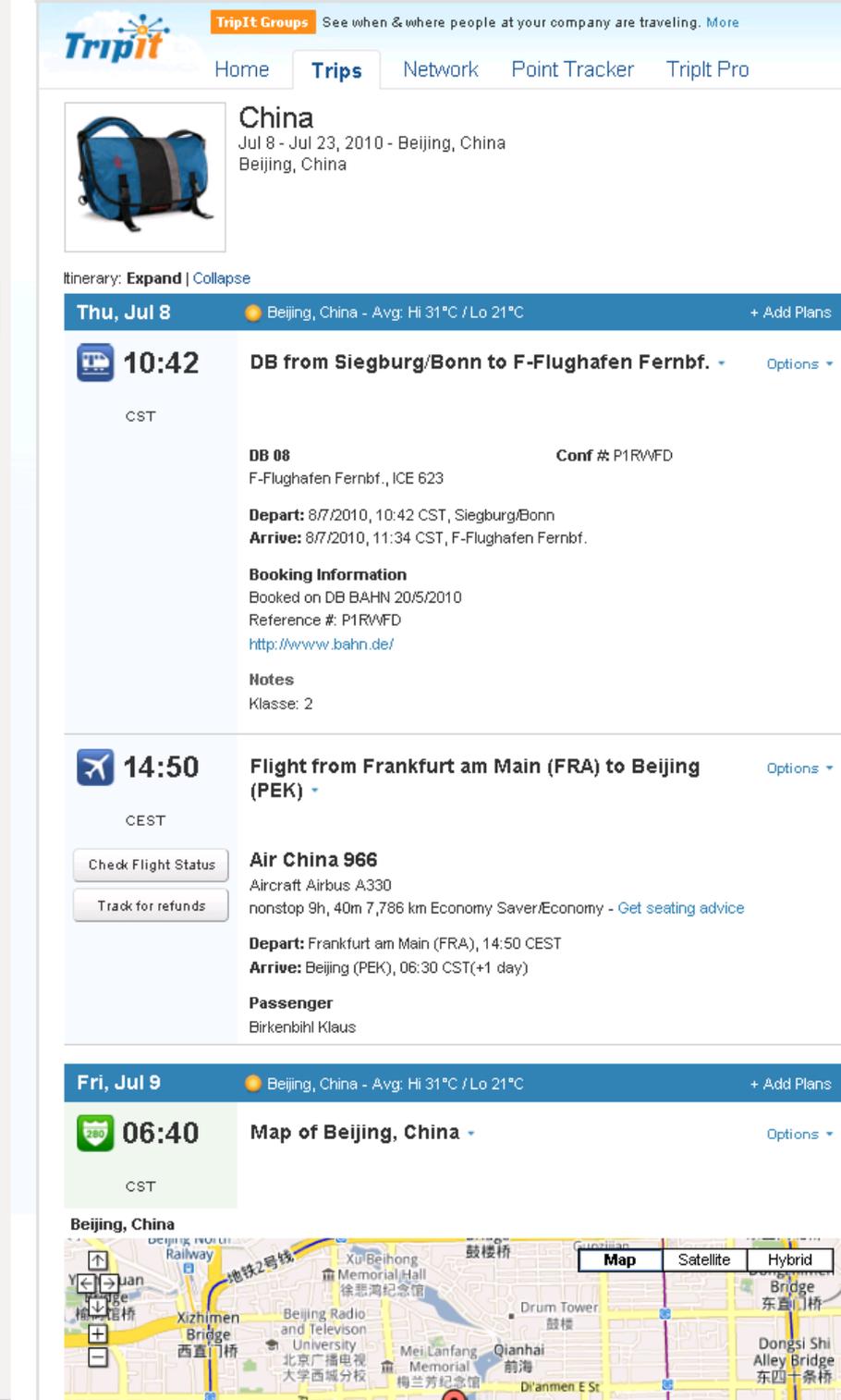
Epilogue

Imagine ...

... all Airlines, Hotels, Car rental companies, travel and conference organizers etc. would use RDFa and adequate Ontologies to mark-up data in their documents

An application like *TripIt* would be a child's play

The same holds for many other areas as well



The screenshot displays the TripIt website interface. At the top, there's a navigation bar with 'Home', 'Trips', 'Network', 'Point Tracker', and 'Tript Pro'. The main content area shows a trip titled 'China' for the dates 'Jul 8 - Jul 23, 2010 - Beijing, China'. Below this, there's a section for 'Thu, Jul 8' with a temperature indicator 'Beijing, China - Avg: Hi 31°C / Lo 21°C'. The itinerary includes a train departure at 10:42 CST from Siegburg/Bonn to F-Flughafen Fernbf. (DB 08) and a flight departure at 14:50 CEST from Frankfurt am Main (FRA) to Beijing (PEK) on Air China 966. The flight details include aircraft type (Airbus A330), nonstop duration (9h, 40m), and economy class. A map of Beijing, China is shown for Friday, July 9, with a departure time of 06:40 CST. The map includes labels for various landmarks and streets in Beijing.

Readings ...

To learn about RDF

<http://www.w3.org/TR/rdf-primer/>

<http://zh.transwiki.org/cn/rdfprimer.htm>

To learn about RDFa

<http://www.w3.org/TR/xhtml-rdfa-primer/>

<http://iws.seu.edu.cn/resource/Translations/RDFa-Primer-Simplified-Chinese.htm>

To stay tuned to RDFa news

<http://rdfa.info/>

Also holds a lot of references to tools and readings

RDFa for HTML authors

<http://www.w3.org/MarkUp/2009/rdfa-for-html-authors>





Slides are available at:
<http://www.w3.org/2010/Talks/0717Beijing-KB/>
in OpenDocument Presentation Format and PDF
in English and Chinese