Standardizing for Open Data

Ivan Herman, W3C

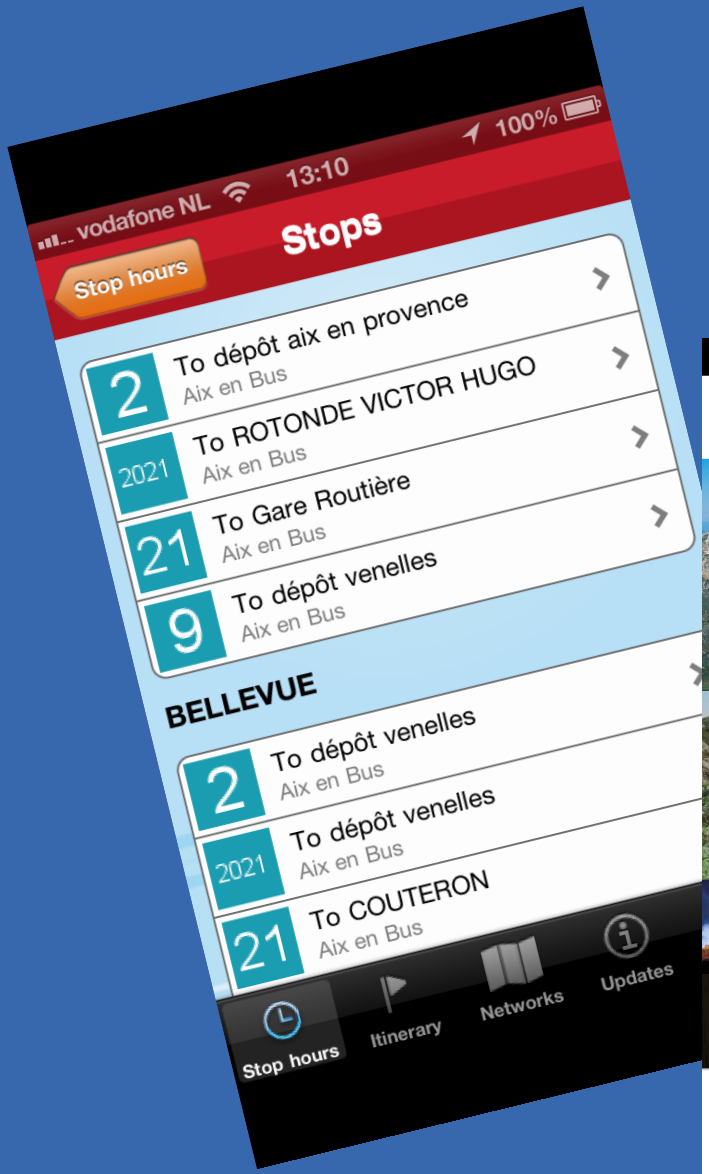
Open Data Week

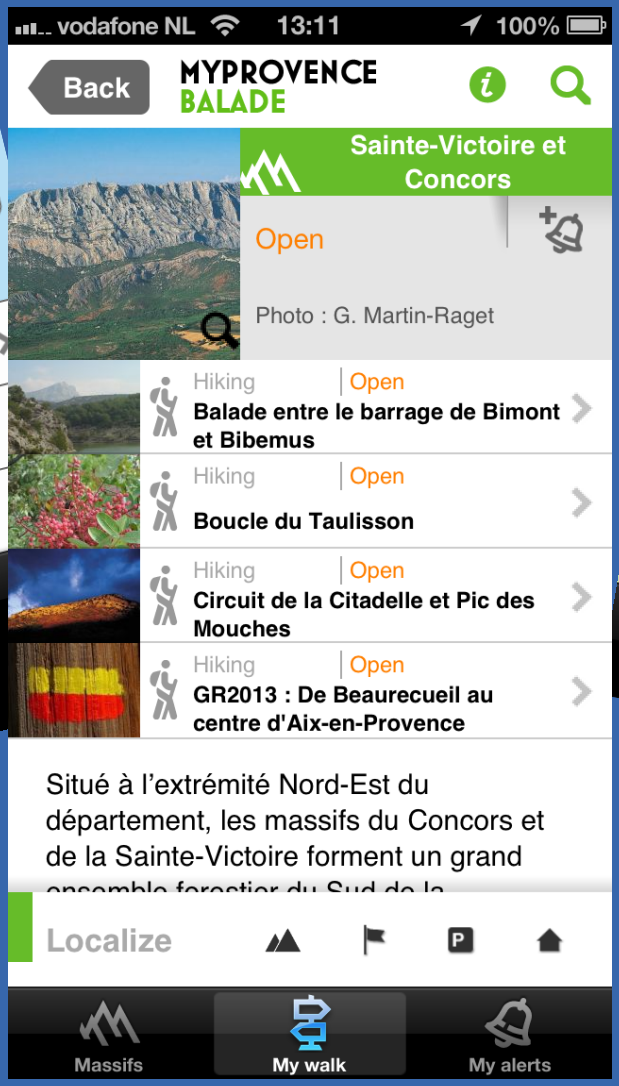
Marseille, France, June 26 2013

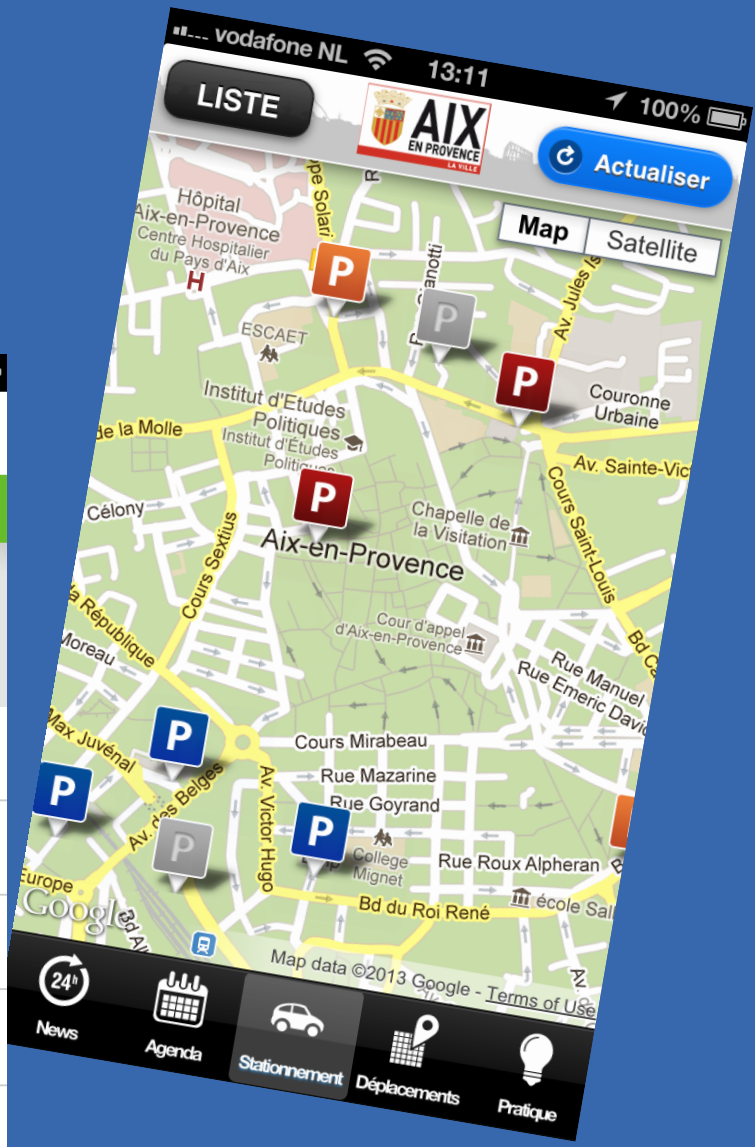
Slides at: http://www.w3.org/2013/Talks/0626-Marseille-IH/

Data is everywhere on the Web!

- Public, private, behind enterprise firewalls
- Ranges from informal to highly curated
- Ranges from machine readable to human readable
  - HTML tables, twitter feeds, local vocabularies, spreadsheets, …
- Expressed in diverse models
  - tree, graph, table, …
- Serialized in many ways
  - XML, CSV, RDF, PDF, HTML Tables, microdata,…


vodafone NL
13:10
100%
Stops
Stop hours
2
To dépôt aix en provence
Aix en Bus
2021
To ROTONDE VICTOR HUGO
Aix en Bus
21
To Gare Routière
Aix en Bus
9
To dépôt venelles
Aix en Bus
BELLEVUE
2
To dépôt venelles
Aix en Bus
2021
To dépôt venelles
Aix en Bus
21
To COUTERON
Aix en Bus
Stop hours
Itinerary
Networks
Updates


vodafone NL
13:11
100%
Back
MYPROVENCE
BALADE
Sainte-Victoire et Concors
Open
Photo : G. Martin-Raget
Hiking
Open
Balade entre le barrage de Bimont et Bibemus
Hiking
Open
Boucle du Taulisson
Hiking
Open
Circuit de la Citadelle et Pic des Mouches
Hiking
Open
GR2013 : De Beaurecueil au centre d'Aix-en-Provence
Situé à l'extrémité Nord-Est du département, les massifs du Concors et de la Sainte-Victoire forment un grand
Localize
Massifs
My walk
My alerts


vodafone NL
13:11
100%
LISTE
AIX EN PROVENCE
Actualiser
Map
Satellite
Hôpital Aix-en-Provence Centre Hospitalier du Pays d'Aix
ESCAET
Institut d'Etudes Politiques
Couronne Urbaine
Aix-en-Provence
Chapelle de la Visitation
Cours Sextius
Cours Saint-Louis
Cour d'appel d'Aix-en-Provence
Cours Mirabeau
Rue Mazarine
Rue Goyrand
Rue Roux Alpheran
Bd du Roi René
Av. Victor Hugo
Map data ©2013 Google - Terms of Use
News
Agenda
Stationnement
Déplacements
Pratique

the artist movie - Google Search

https://www.google.nl/#safe=off&biw=1080&bih=800&sclient=psy-ab&q=the artist movie&oq=&

+Ivan Search Images Maps Play YouTube Gmail Drive Calendar Translate More

Google the artist movie

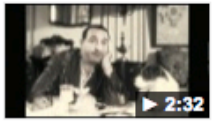
Ivan Herman 0

Web Images Maps Shopping Videos More Search tools SafeSearch

About 354,000,000 results (0.32 seconds)

The Artist (2011) - IMDb
www.imdb.com/title/tt1655442/
Rating: 8.1/10 - 118783 votes
A silent **movie** star meets a young dancer, but the arrival of talking pictures sends ... **'The Artist'** Star Bérénice Bejo On Returning to Cannes With 'The Past' and ...
Full cast and crew - Awards - Trivia - Bérénice Bejo

The Artist (film) - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**The_Artist_(film)**
The Artist is a 2011 French romantic comedy-drama **film** in the style of a black-and-white silent **film**. The **film** was written and directed by Michel Hazanavicius ...
Bérénice Bejo - Jean Dujardin - List of accolades received

The Artist Trailer 2011 HD - YouTube
www.youtube.com/watch?v=O8K9AZcSQJE


Aug 25, 2011 - Uploaded by TrailersApplecom
The Artist - Official Trailer 2011 HD More info: http://trailers.apple.
... George Valentin is a very successful ...

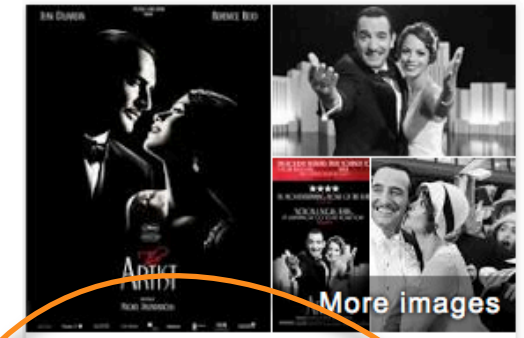
More videos for **the artist movie** »

The Artist - Rotten Tomatoes
www.rottentomatoes.com/m/**the_artist/**
Rating: 98% - 218 reviews
For young extra Peppy Miller (Berenice Bejo), it seems the sky's the limit - major **movie** stardom awaits. **The Artist** tells the story of their interlinked destinies.

The Artist | Film | The Guardian
www.guardian.co.uk › Culture › Film
50+ items – Summary: A silent **movie** star struggles to come to terms with ...

The Artist – review | Film | The Guardian
www.guardian.co.uk › Culture › Film › The Artist
Dec 22, 2011 – This witty and moving homage to the silent **movie** era is one of the most purely enjoyable movies in years, writes Peter Bradshaw.

The Artist | review, synopsis, book tickets, showtimes, movie release ...
www.timeout.com/london/**film/the-artist**



The Artist

2011 Film

8.1/10 - IMDb
98% - Rotten Tomatoes

The Artist is a 2011 French romantic comedy-drama film in the style of a black-and-white silent film. The film was written and directed by Michel Hazanavicius and stars Jean Dujardin and Bérénice Bejo.
Wikipedia

Initial release: October 12, 2011 (Belgium, France)

Director: Michel Hazanavicius

Featured song: Pennies from Heaven

Screenplay: Michel Hazanavicius

Awards: Academy Award for Best Picture, More

Cast

Home | Data.gov

www.data.gov

My Mercurial | LocalData | TR | TR List | 2013 | Private | Social | Mailing lists | SW | Python | RDFa it! | Bookmarklets | Mob. Highlighter

An Official Web Site of the United States Government | Change Text Size | SHARE

DATA.GOV
EMPOWERING PEOPLE

Login

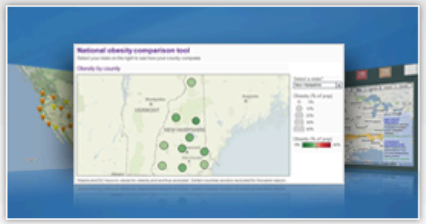
Search Data.gov | SEARCH

HOME | ABOUT | DATA | METRICS | OPEN GOVERNMENT | BLOGS | COMMUNITIES

Latest Datasets

- DOE Green Energy
- DOE Patents
- ScienceCinema
- Gravesite locations of Veterans and...
- Gravesite locations of Veterans and...
- Gravesite locations of Veterans and...
- Gravesite locations of Veterans and...
- Gravesite locations of Veterans and...
- Gravesite locations of Veterans and...
- SciTech Connect

DATA AND TOOLS



- 373,029 raw and geospatial datasets
- 1,209 data tools
- 350 apps
- 137 mobile apps
- 172 agencies and subagencies
- Suggest a dataset

COMMUNITIES

Come explore, discuss, meet others in the same field, and develop the data and apps in the community that you care about. Join in the discussions by going to communities, some of which are listed below, that interest you.

EDUCATION | ENERGY
BUSINESSUSA | AGRICULTURE
OCEAN | SAFETY

OPEN GOVERNMENT

View our government at work and find opportunities to participate.

- Open Government Initiative
- Open Government Platform
- Recovery.gov
- USASpending.gov
- We the People

Join the conversation

LEARN

SEMANTIC WEB

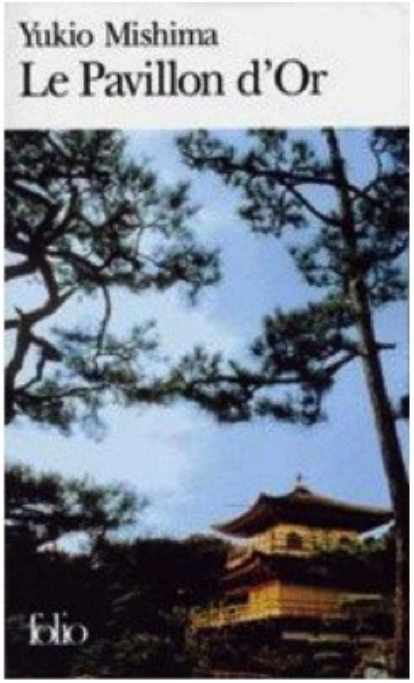
DEVELOPERS CORNER

(5)

Cover of: Le Pavillion D'Or by Yukio Mishima

1 edition of Le Pavillion D'Or by Yukio Mishima • Add edition?

Last edited by Ivan Herman
584 milliseconds ago | History



Manage Covers

Le Pavillion D'Or

Published **October 1, 1975** by French & European Pubns.

The Physical Object

Format	Paperback

ID Numbers

Open Library	OL11039953M
ISBN 10	0785940391
ISBN 13	9780785940395
Goodreads	756054

Read

No readable version available.

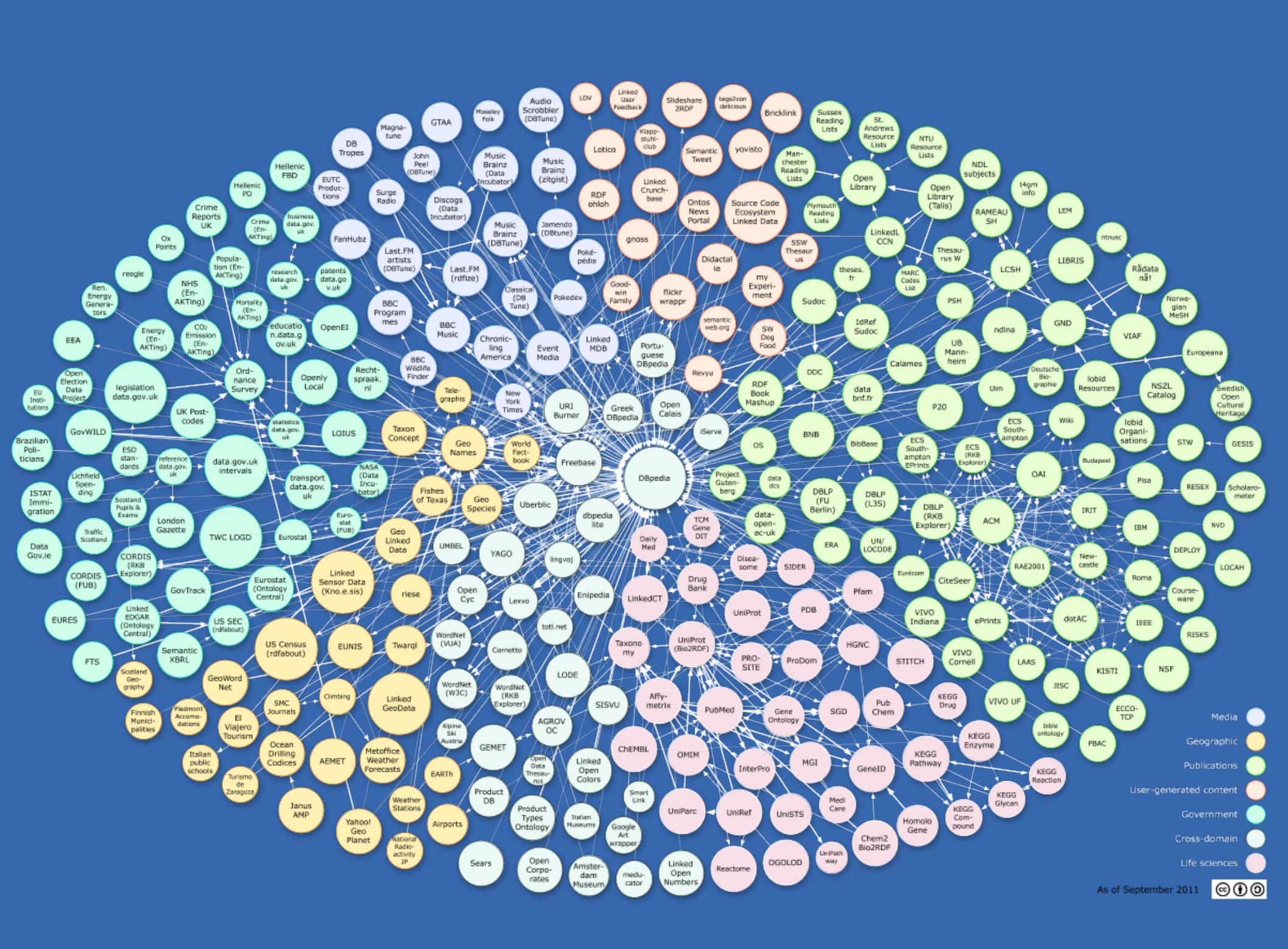
Borrow

Physical copy, local WorldCat

Buy

Alibris
Amazon
AbeBooks
Biblio.com
Book Depository
Powells

Lists

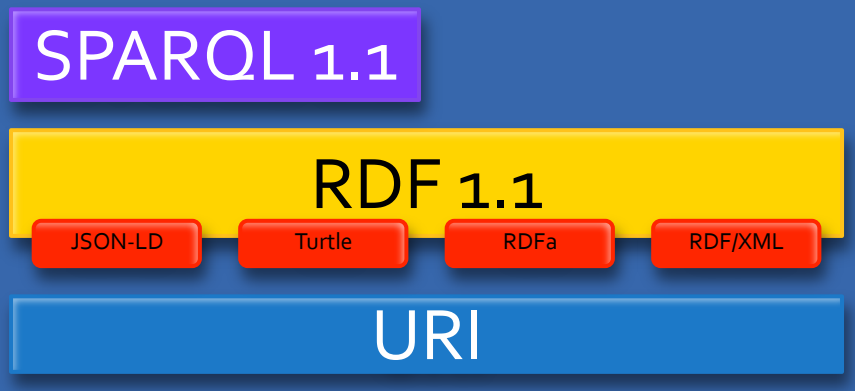

DBpedia
Freebase
Geo Names
Linked Sensor Data (Kno.e.sis)
Linked GeoData
Linked Open Numbers
Linked Open Colors
Linked Open Data Thesaurus
Linked Open Corporates
Linked Open Amsterdam Museum
Linked Open Reactome
Linked Open OGOLOD
Linked Open Sears
Linked Open Yahoo! Geo Planet
Linked Open Weather Stations
Linked Open Airports
Linked Open Product DB
Linked Open Product Types Ontology
Linked Open Italian Museums
Linked Open Google Art wrapper
Linked Open Smart Link
Linked Open UniParc
Linked Open UniRef
Linked Open UniSTS
Linked Open Chem2 Bio2RDF
Linked Open Homolo Gene
Linked Open Medi Care
Linked Open KEGG Compound
Linked Open KEGG Glycan
Linked Open KEGG Reaction
Linked Open KEGG Pathway
Linked Open KEGG Enzyme
Linked Open KEGG Drug
Linked Open GeneID
Linked Open MGI
Linked Open InterPro
Linked Open OMIM
Linked Open ChEMBL
Linked Open PubMed
Linked Open Affy-metrix
Linked Open SISVU
Linked Open LODE
Linked Open AGROVOC
Linked Open GEMET
Linked Open EARTh
Linked Open Alpine Ski Austria
Linked Open WordNet (RKB Explorer)
Linked Open WordNet (W3C)
Linked Open Cornetto
Linked Open totl.net
Linked Open Lexvo
Linked Open Open Cyc
Linked Open riese
Linked Open Twarql
Linked Open WordNet (VUA)
Linked Open EUNIS
Linked Open US Census (rdfabout)
Linked Open Semantic XBRL
Linked Open US SEC (rdfabout)
Linked Open Linked EDGAR (Ontology Central)
Linked Open FTS
Linked Open EURES
Linked Open CORDIS (FUB)
Linked Open CORDIS (RKB Explorer)
Linked Open GovTrack
Linked Open Eurostat (Ontology Central)
Linked Open Eurostat
Linked Open TWC LOGD
Linked Open London Gazette
Linked Open Data Gov.ie
Linked Open ISTAT Immigration
Linked Open Brazilian Politicians
Linked Open GovWILD
Linked Open UK Post-codes
Linked Open legislation data.gov.uk
Linked Open EEA
Linked Open Open Election Data Project
Linked Open EU Institutions
Linked Open Ordnance Survey
Linked Open data.gov.uk intervals
Linked Open Energy (En-AKTing)
Linked Open CO2 Emission (En-AKTing)
Linked Open Ren. Energy Generators
Linked Open NHS (En-AKTing)
Linked Open Mortality (En-AKTing)
Linked Open Population (En-AKTing)
Linked Open Crime Reports UK
Linked Open Ox Points
Linked Open reegle
Linked Open Hellenic PD
Linked Open Hellenic FBD
Linked Open Crime (En-AKTing)
Linked Open business data.gov.uk
Linked Open patents data.gov.uk
Linked Open research data.gov.uk
Linked Open education.data.gov.uk
Linked Open Openly Local
Linked Open statistics data.gov.uk
Linked Open reference data.gov.uk
Linked Open transport data.gov.uk
Linked Open LOIUS
Linked Open Recht-spraak.nl
Linked Open OpenEI
Linked Open BBC Wildlife Finder
Linked Open BBC Programmes
Linked Open BBC Music
Linked Open Scotland Pupils & Exams
Linked Open Lichfield Spending
Linked Open ESD standards
Linked Open Traffic Scotland
Linked Open Scotland Geography
Linked Open Finnish Municipalities
Linked Open Piedmont Accommodations
Linked Open El Viajero Tourism
Linked Open Italian public schools
Linked Open Turismo de Zaragoza
Linked Open Janus AMP
Linked Open National Radioactivity JP
Linked Open Ocean Drilling Codices
Linked Open AEMET
Linked Open Metoffice Weather Forecasts
Linked Open SMC Journals
Linked Open Climbing
Linked Open GeoWordNet
Linked Open Scholarometer
Linked Open STW
Linked Open GESIS
Linked Open Swedish Open Cultural Heritage
Linked Open Europeana
Linked Open Norwegian MeSH
Linked Open Rådata nå!
Linked Open LIBRIS
Linked Open ntnusc
Linked Open LEM
Linked Open t4gm info
Linked Open NDL subjects
Linked Open Open Library (Talis)
Linked Open Open Library
Linked Open Sussex Reading Lists
Linked Open St. Andrews Resource Lists
Linked Open NTU Resource Lists
Linked Open Manchester Reading Lists
Linked Open Plymouth Reading Lists
Linked Open RAMEAU SH
Linked Open LCSH
Linked Open PSH
Linked Open GND
Linked Open VIAF
Linked Open NSZL Catalog
Linked Open lobid Resources
Linked Open lobid Organisations
Linked Open Budapest
Linked Open Pisa
Linked Open RESEX
Linked Open IRIT
Linked Open IBM
Linked Open NVD
Linked Open DEPLOY
Linked Open LOCAH
Linked Open Roma
Linked Open Course-ware
Linked Open IEEE
Linked Open RISKS
Linked Open NSF
Linked Open KISTI
Linked Open JISC
Linked Open ECCO-TCP
Linked Open PBAC
Linked Open bible ontology
Linked Open VIVO UF
Linked Open LAAS
Linked Open VIVO Cornell
Linked Open STITCH
Linked Open HGNC
Linked Open Pub Chem
Linked Open SGD
Linked Open Gene Ontology
Linked Open ProDom
Linked Open PRO-SITE
Linked Open PDB
Linked Open Pfam
Linked Open SIDER
Linked Open Disea-some
Linked Open Drug Bank
Linked Open UniProt
Linked Open UniProt (Bio2RDF)
Linked Open Taxono my
Linked Open LinkedCT
Linked Open Enipedia
Linked Open Daily Med
Linked Open TCM Gene DIT
Linked Open dbpedia lite
Linked Open YAGO
Linked Open lingvoj
Linked Open UMBEL
Linked Open Uberblic
Linked Open Geo Species
Linked Open Fishes of Texas
Linked Open Geo Linked Data
Linked Open NASA (Data Incu-bator)
Linked Open Eurostat (FUB)
Linked Open data-open-ac-uk
Linked Open ERA
Linked Open UN/ LOCODE
Linked Open DBLP (RKB Explorer)
Linked Open DBLP (L3S)
Linked Open DBLP (FU Berlin)
Linked Open data dcs
Linked Open Project Guten-berg
Linked Open Eurécom
Linked Open CiteSeer
Linked Open RAE2001
Linked Open ACM
Linked Open OAI
Linked Open ECS (RKB Explorer)
Linked Open ECS South-ampton
Linked Open ECS South-ampton EPrints
Linked Open VIVO Indiana
Linked Open ePrints
Linked Open dotAC
Linked Open New-castle
Linked Open Wiki
Linked Open Deutsche Bio-graphie
Linked Open Ulm
Linked Open UB Mann-heim
Linked Open ndlna
Linked Open BibBase
Linked Open BNB
Linked Open data bnf.fr
Linked Open DDC
Linked Open Calames
Linked Open P20
Linked Open Sudoc
Linked Open IdRef Sudoc
Linked Open Thesau-rus W
Linked Open MARC Codes List
Linked Open theses. fr
Linked Open LinkedL CCN
Linked Open SSW Thesau-rus
Linked Open my Experi-ment
Linked Open SW Dog Food
Linked Open semantic web.org
Linked Open Revyu
Linked Open RDF Book Mashup
Linked Open flickr wrappr
Linked Open Good-win Family
Linked Open gnoss
Linked Open Ontos News Portal
Linked Open Source Code Ecosystem Linked Data
Linked Open Didactal ia
Linked Open Portu-guese DBpedia
Linked Open Linked MDB
Linked Open Event Media
Linked Open Chronic-ling America
Linked Open Classical (DB Tune)
Linked Open Pokedex
Linked Open Poké-pédia
Linked Open Jamendo (DBtune)
Linked Open Music Brainz (DBTune)
Linked Open Last.FM (rdfize)
Linked Open Last.FM artists (DBTune)
Linked Open Discogs (Data Incubator)
Linked Open Music Brainz (Data Incubator)
Linked Open Music Brainz (zitgist)
Linked Open Surge Radio
Linked Open FanHubz
Linked Open EUTC Produc-tions
Linked Open DB Tropes
Linked Open John Peel (DBTune)
Linked Open Magna-tune
Linked Open GTAA
Linked Open Moseley Folk
Linked Open Audio Scrobbler (DBTune)
Linked Open Lotico
Linked Open RDF ohloh
Linked Open Linked Crunch-base
Linked Open Klapp-stuhl-club
Linked Open Semantic Tweet
Linked Open yovisto
Linked Open Slideshare 2RDF
Linked Open tags2con delicious
Linked Open Bricklink
Linked Open Linked User Feedback
Linked Open LOV
Linked Open New York Times
Linked Open BBC Music
Linked Open iServe
Linked Open Open Calais
Linked Open URI Burner
Linked Open Greek DBpedia
Linked Open Taxon Concept
Linked Open Geo Names
Linked Open World Fact-book
Linked Open Freebase
Linked Open Tele-graphis
Linked Open OS
Media
Geographic
Publications
User-generated content
Government
Cross-domain
Life sciences
As of September 2011

(7)

W3C's standardization focus was, traditionally, on Web scale integration of data

- Some basic principles:
  - use of URIs everywhere (to uniquely identify things)
  - relate resources among one another (to connect things on the Web)
  - discover new relationships through inferences
- This is what the Semantic Web technologies are all about

We have a number of standards

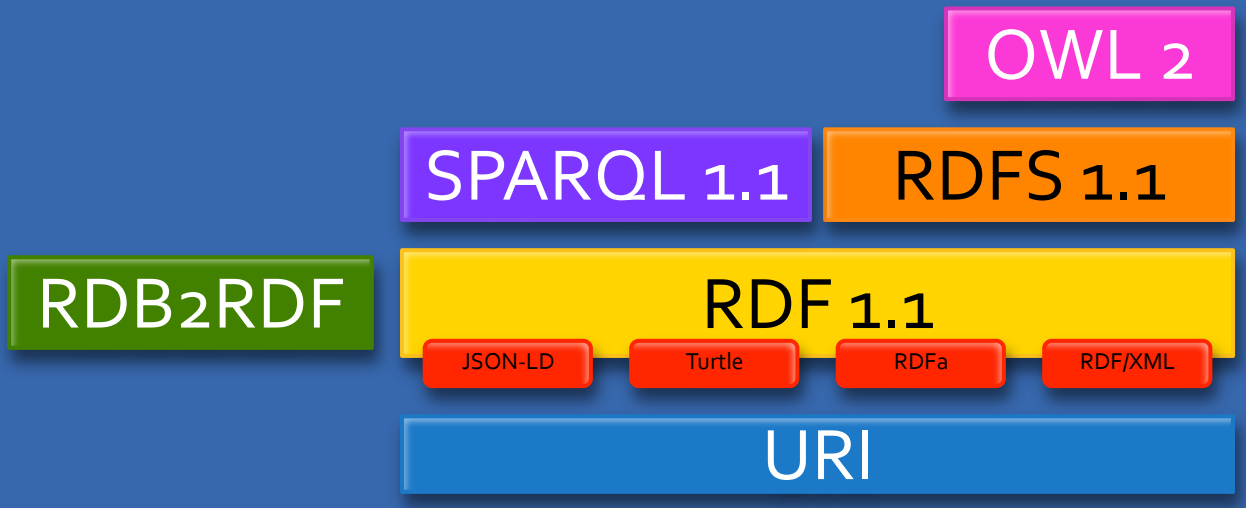


SPARQL: querying data

RDF: data model, links, basic assertions; different serializations

A fairly stable set of technologies by now!

We have a number of standards



RDB2RDF: databases to RDF

SPARQL: querying data

OWL: complex vocabularies, ontologies

RDFS: simple vocabularies

RDF: data model, links, basic assertions; different serializations

A fairly stable set of technologies by now!

We have Linked Data principles

Integration is done in different ways

- *Very roughly*:
  - data is accessed directly as RDF and turned into something useful
    - relies on data being "preprocessed" and published as RDF
  - data is collected from different sources, *integrated internally*
    - using, say, a triple store

BBC – Music – Eric Clapton

http://www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294

BBC Music

Netvibes ▾ Feedly Social ▾ Private ▾ Mailing lists ▾ SW ▾ Python ▾ RDFa it! ▾ Bookmarklets ▾ Add Zemanta bit.ly To Mendeley TinyURL To Faviki Dokuwiki

MUSIC BETA GENRES ARTISTS REVIEWS NEWS BLOG QUICK FIND Enter an artist name ...

BBC Music > Artists > Eric Clapton

Eric Clapton

Born 30 March 1945.

MOST PLAYED ON BBC RADIO 2

Biography

Eric Patrick Clapton, CBE (born 30 March 1945) is an English blues-rock guitarist, singer, songwriter and composer. Clapton has been inducted into the Rock and Roll Hall of Fame as a solo performer, as a member of rock bands; the Yardbirds and Cream. Clapton is the only person ever to be inducted three times. Often viewed by critics and fans alike as one of the most important and influential guitarists of all time, Clapton was ranked fourth in Rolling Stone magazine's list of the "100 Greatest Guitarists of All Time" and #53 on their list of the Immortals: 100 Greatest Artists of All Time

Latest Tracks Played On The BBC

Promises
BBC Radio 2 | Ken Bruce 22/02/2010

Bad Love
BBC Radio 2 | Alex Lester 22/02/2010

Lay Down Sally
BBC Radio 2 | Chris Evans Breakfast 18/02/2010

I Ain't Gonna Stand For It
BBC Radio 2 | Alex Lester 15/02/2010

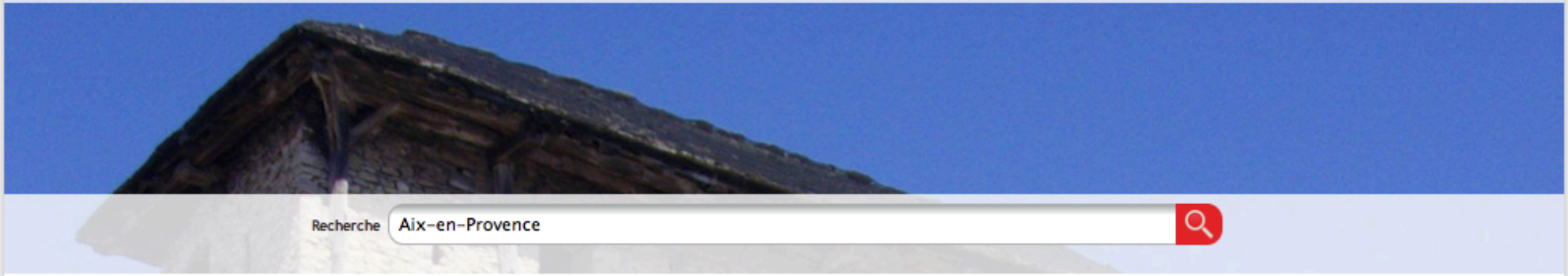
Wonderful Tonight
BBC Radio 2 | Ken Bruce 10/02/2010

Audio Previews From Latest Album Review

Me And Mr Johnson

8 Milkcow's Calf Blues

10 Come on in My Kitchen



Recherche: Aix-en-Provence

Affiner

Photos du monument ?
- Non 4
- Oui 2

Région / Département / Ville
- Provence-Alpes-Côte D'Azur 6

Type de monument
- Eglise 3
- Monument 1
- Hôtel de ville 1
- Monastère 1

Période historique
- Moyen Âge 6
- Époque moderne 5
- Renaissance 2
- Époque contemporaine 1

Gares les plus proches
- Gardanne 6
- Aix-en-Provence 6
- Simiane 6

Type de propriété

Résultats de la recherche

6 résultats (105ms) Trier par Pertinence Afficher 10 résultats par page

Mis à disposition sur Wikimedia Commons

Hôtel de Ville

Type : Hôtel de ville
Adresse : place de l' Hôtel-de-Ville
Localité : Aix-en-Provence (Bouches-du-Rhône)
Période historique concernée : Moyen Âge - Renaissance - Époque moderne -
Liens externes : Base Mérimée - Wikimedia Commons

Description de l'ensemble inscrit ou classé : Hôtel de Ville (cad. AS 63) : classement par arrêté du 12 octobre 1995

Mis à disposition sur Wikimedia Commons

Eglise Saint-Jean-de-Malte

Type : Eglise
Localité : Aix-en-Provence (Bouches-du-Rhône)
Période historique concernée : Moyen Âge - Époque moderne -
Liens externes : Base Mérimée - Wikipedia francophone - Wikimedia Commons

Description de l'ensemble inscrit ou classé : Eglise Saint-Jean-de-Malte : classement par liste de 1840

However…

- There is a price to pay: a relatively heavy ecosystem
  - many developers shy away from using RDF and related tools
- *Not all applications need this!*
  - data may be used directly, no need for integration concerns
  - the emphasis may be on easy production and manipulation of data with simple tools

Typical situation on the Web

- Data published in CSV, JSON, XML
- An application uses only 1-2 datasets, integration done by direct programming is straightforward
  - e.g., in a Web Application
- Data is often very large, direct manipulation is more efficient

Prescribing Analytics

www.prescribinganalytics.com

My Mercurial | LocalData | TR | TR List | 2013 | Private | Social | Mailing lists | SW | Python | RDFa it! | Bookmarklets | Mob. Highlighter

prescribing analytics

- Home
- Detailed statin analysis
- Why prescriptions matter
- FAQ
- Methodology
- Contact us
- About this site

About Prescribing Analytics

Prescribing Analytics is a joint venture of a group of UK technologists and NHS doctors, who believe in the power of data and technology to help the NHS.

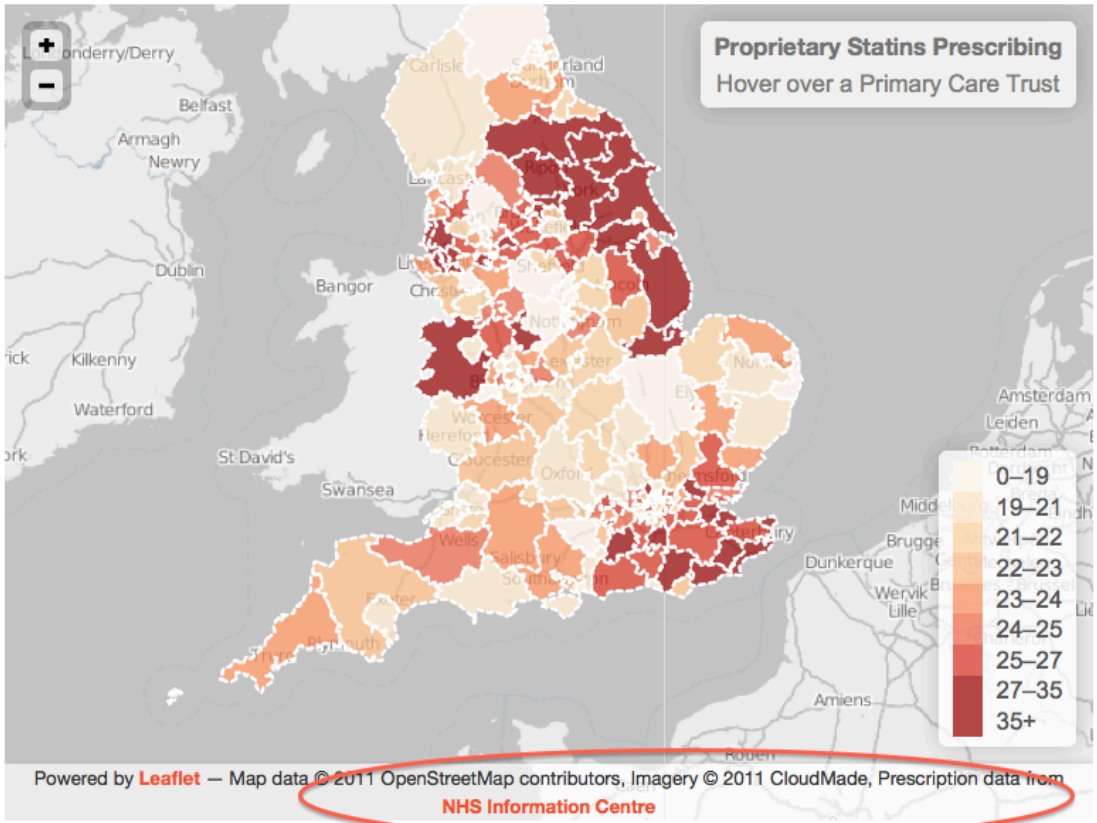
Drop us an email at
info@prescribinganalytics.com

Home

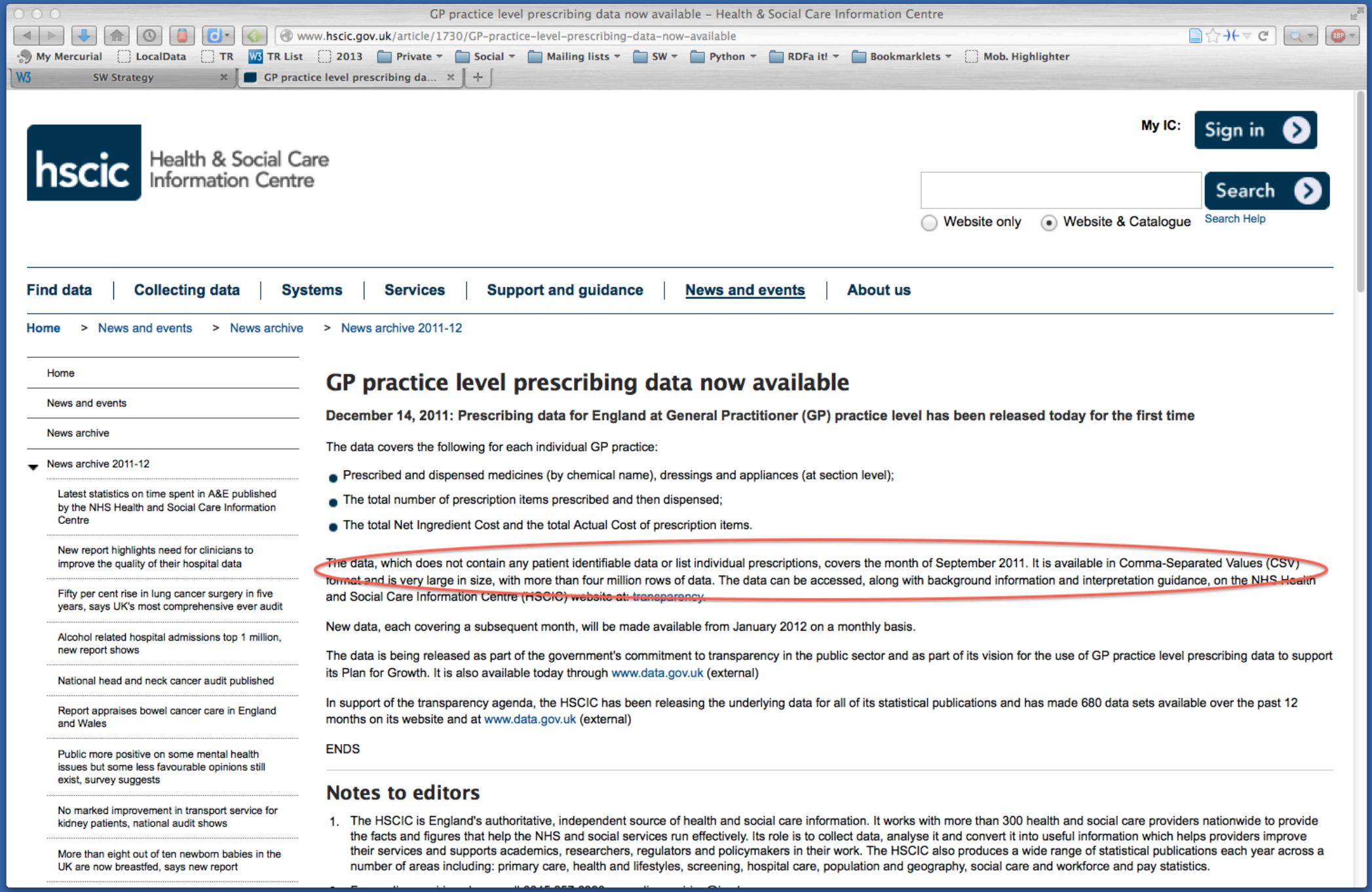
NHS efficiency savings: the role of prescribing analytics

The NHS has been challenged to make £20 billion in "efficiency savings" by 2015 (1). £10 billion a year are spent by the NHS on essential drugs. Often, there's a choice between a cheap "generic" medication, or an expensive "branded" one. Branded drugs can cost over ten times as much, for the same therapeutic benefit. "Prescribing Advisors" in the NHS, with the support of NICE, encourage doctors to use the most cost effective treatments. We have analysed exactly how much is spent on expensive "branded" medicines, for one class of drugs, namely statins, in England.

Percentage of proprietary statin prescribing by CCG Sep 2011 - May 2012



Show PCT data


GP practice level prescribing data now available – Health & Social Care Information Centre
www.hscic.gov.uk/article/1730/GP-practice-level-prescribing-data-now-available
My Mercurial
LocalData
TR
TR List
2013
Private
Social
Mailing lists
SW
Python
RDFa it!
Bookmarklets
Mob. Highlighter
SW Strategy
GP practice level prescribing da...
hscic
Health & Social Care Information Centre
My IC:
Sign in
Search
Website only
Website & Catalogue
Search Help
Find data
Collecting data
Systems
Services
Support and guidance
News and events
About us
Home > News and events > News archive > News archive 2011-12
Home
News and events
News archive
News archive 2011-12
Latest statistics on time spent in A&E published by the NHS Health and Social Care Information Centre
New report highlights need for clinicians to improve the quality of their hospital data
Fifty per cent rise in lung cancer surgery in five years, says UK's most comprehensive ever audit
Alcohol related hospital admissions top 1 million, new report shows
National head and neck cancer audit published
Report appraises bowel cancer care in England and Wales
Public more positive on some mental health issues but some less favourable opinions still exist, survey suggests
No marked improvement in transport service for kidney patients, national audit shows
More than eight out of ten newborn babies in the UK are now breastfed, says new report
GP practice level prescribing data now available
December 14, 2011: Prescribing data for England at General Practitioner (GP) practice level has been released today for the first time
The data covers the following for each individual GP practice:
Prescribed and dispensed medicines (by chemical name), dressings and appliances (at section level);
The total number of prescription items prescribed and then dispensed;
The total Net Ingredient Cost and the total Actual Cost of prescription items.
The data, which does not contain any patient identifiable data or list individual prescriptions, covers the month of September 2011. It is available in Comma-Separated Values (CSV) format and is very large in size, with more than four million rows of data. The data can be accessed, along with background information and interpretation guidance, on the NHS Health and Social Care Information Centre (HSCIC) website at: transparency.
New data, each covering a subsequent month, will be made available from January 2012 on a monthly basis.
The data is being released as part of the government's commitment to transparency in the public sector and as part of its vision for the use of GP practice level prescribing data to support its Plan for Growth. It is also available today through www.data.gov.uk (external)
In support of the transparency agenda, the HSCIC has been releasing the underlying data for all of its statistical publications and has made 680 data sets available over the past 12 months on its website and at www.data.gov.uk (external)
ENDS
Notes to editors
1. The HSCIC is England's authoritative, independent source of health and social care information. It works with more than 300 health and social care providers nationwide to provide the facts and figures that help the NHS and social services run effectively. Its role is to collect data, analyse it and convert it into useful information which helps providers improve their services and supports academics, researchers, regulators and policymakers in their work. The HSCIC also produces a wide range of statistical publications each year across a number of areas including: primary care, health and lifestyles, screening, hospital care, population and geography, social care and workforce and pay statistics.

What that application does…

- Gets the data published by NHS
- Processes the data (e.g., through Hadoop)
- Integrates *the result* of the analysis with geographical data

Ie: the raw data is used without integration

The reality of data on the Web…

- It is still a fairly messy space out there ☹
  - many different formats are used
  - data is difficult to find
  - published data are messy, erroneous,
  - tools are complex, unfinished…

How do developers perceive this?

'**When transportation agencies consider data integration**, one pervasive notion is that the analysis of existing information needs and infrastructure, much less the organization of data into viable channels for integration, **requires a monumental initial commitment of resources** and staff. Resource-scarce agencies identify this perceived major upfront overhaul as "unachievable" and "disruptive."'

-- Data Integration Primer: Challenges to Data Integration, US Dept. of Transportation

One may look at the problem through different goggles

- Two alternatives come to the fore:
  1. provide tools, environments, etc., to help outsiders to publish Linked Data (in RDF) easily
     - a typical example is the Datalift project
  2. forget about RDF, Linked Data, etc, and concentrate on the raw data instead

DANGER!
RELIGIOUS
WARS!
www.txt2pic.com

But religions and cultures can coexist… ☺

Open Data on the Web Workshop

- Had a successful workshop in London, in April:
  - around 100 participants
  - coming from different horizons: publishers and users of Linked Data, CSV, PDF, …

We also talked to our "stakeholders"

- Member organizations and companies
- Open Data Institute, Open Knowledge Foundation, Schema.org
- …

Some takeaway

- The Semantic Web community needs stability of the technology
  - do not add yet another technology block ☺
  - existing technologies should be maintained

Some takeaway

- Look at the more general space, too
  - importance of metadata
  - deal with non-RDF data formats
  - best practices are necessary to raise the quality of published data

We need to meet app developers where they are!

Metadata is of a major importance

- Metadata describes the characteristics of the dataset
  - structure, datatypes used
  - access rights, licenses
  - provenance, authorship
  - etc.
- Vocabularies are also key for Linked Data

Vocabulary Management Action

- Standard vocabularies are necessary to describe data
  - there are already some initiatives: W3C's data cube, data catalog, PROV, schema.org, DCMI, …
- At the moment, it is a fairly chaotic world…
  - many, possibly overlapping vocabularies
  - difficult to locate the one that is needed
  - vocabularies may not be properly managed, maintained, versioned, provided persistence…

W3C's plan:

- Provide a space whereby
  - communities can develop
  - *host vocabularies at W3C* if requested
  - annotate vocabularies with a proper set of metadata terms
  - establish a vocabulary directory
- The exact structure is still being discussed:

  http://www.w3.org/2013/04/vocabs/

-1
3
FIRE EXIT ONLY

CSV on the Web

- Planned work areas:
  - metadata vocabulary to describe CSV data
    - structure, reference to access rights, annotations, etc.
  - methods to *find* the metadata
    - part of an HTTP header, special rows and columns, packaging formats…
  - mapping content to RDF, JSON, XML
- Possibly at a later phase:
  - API standards to access CSV data

Open Data Best Practices

- Document best practices for data publishers
  - management of persistence, versioning, URI design
  - use of core vocabularies (provenance, access control, ownership, annotations,…)
  - business models
- Specialized Metadata vocabularies
  - quality description (quality of the data, update frequencies, correction policies, etc.)
  - description of data access API-s
  - …

Summary

- Data on the Web has many different facets
- We have concentrated on the integration aspects in the past years
- We have to take a more general view, look at other types of data published on the Web

In future…

- We should look at other formats, not only CSV
  - MARC, GIS, ABIF,…
- Better outreach to data publishing communities and organizations
  - WF, RDA, ODI, OKFN, …

Enjoy the event!