FROM RAW PUBLISHED DATA TO INTERLINKED SEMANTIC DATA

Authors	Ghislain Atemezing (EURECOM) Charles Nepote (FING) Raphaël Troncy (EURECOM)
Reviewer	Fabien Gandon (INRIA) Thomas Francart (Mondeca)
Date	June, 8th 2012
Reference	
Version	V1.1
Destination	Public

PROJET DATALIFT
De la donnée brute publiée vers la donnée sémantique interconnectée

Appel ANR CONTINT 2010 ANR-10-CORD-009
RAPPORT DE RECHERCHE

June 2012

USAGE SCENARII FOR APPLICATIONS

Authors:
Ghislain A. Atemezing (EURECOM)
Charles Nepote (FING)
Raphaël Troncy (EURECOM)

Reviewer:
Fabien Gandon (INRIA)
Thomas Francart (Mondeca)

Table of Contents

1. Introduction

Nowadays, more than fifteen states and fifty urban communities have begun to open their data among which one finds the largest U.S. cities. This should be added to the cities of the United Kingdom where a hundred of urban communities have lead this movement under the leadership of the British government (1). PublicData.eu lists more than 215 initiatives of public actors in Europe, either at the national level (Belgium, Finland, Sweden, Norway, Spain, Greece, France, UK, ...) or at local levels (region, cities, urban communities...).

The Open data movement in France is evolving and advocates for federating all the actors engaged in projects of sharing public data in France. The visibility of the movement can be seen with the website opendatafrance.net (2) aiming at working together with different actors and initiatives such as LiberTIC (3), Data Publica (4), FING (5), Regards Citoyens (6) and OWNI (7). Many of these initiatives are following the same objective than the data.gouv.fr, the portal of the French Government to open public data earlier this year 2012.

In France, it is Kéolis and Rennes Metropole (8) who launched the first initiative in 2010 with the data warehouse of Rennes (9), joined by other regional and territorial (including platforms of Paris Open Data (10), the Saone-et-Loire (11), Loire Atlantique (12), Loir-et-Cher (13), the Gironde (14), the Bordeaux Urban Community (15), the Grand Toulouse (16), Nantes (17), Montpellier (18), etc.. (the Figure 1 shows the map of open data initiatives published by (3)), then by the creation of EtaLab (19) and the launch of the portal data.gouv.fr (20), the initiative at the state level at the end of 2011.

In two years, the legal landscape has been also simplified. Most of the data sets converge on two licenses: the Open Database License (ODLB)[1] and the Open License[2]. The latter is used by Etalab.

---

[1] http://fr.wikipedia.org/wiki/Open_Database_License
[2] http://ddata.over-blog.com/xxxyyy/4/37/99/26/licence/Licence-Ouverte-Open-Licence.pdf
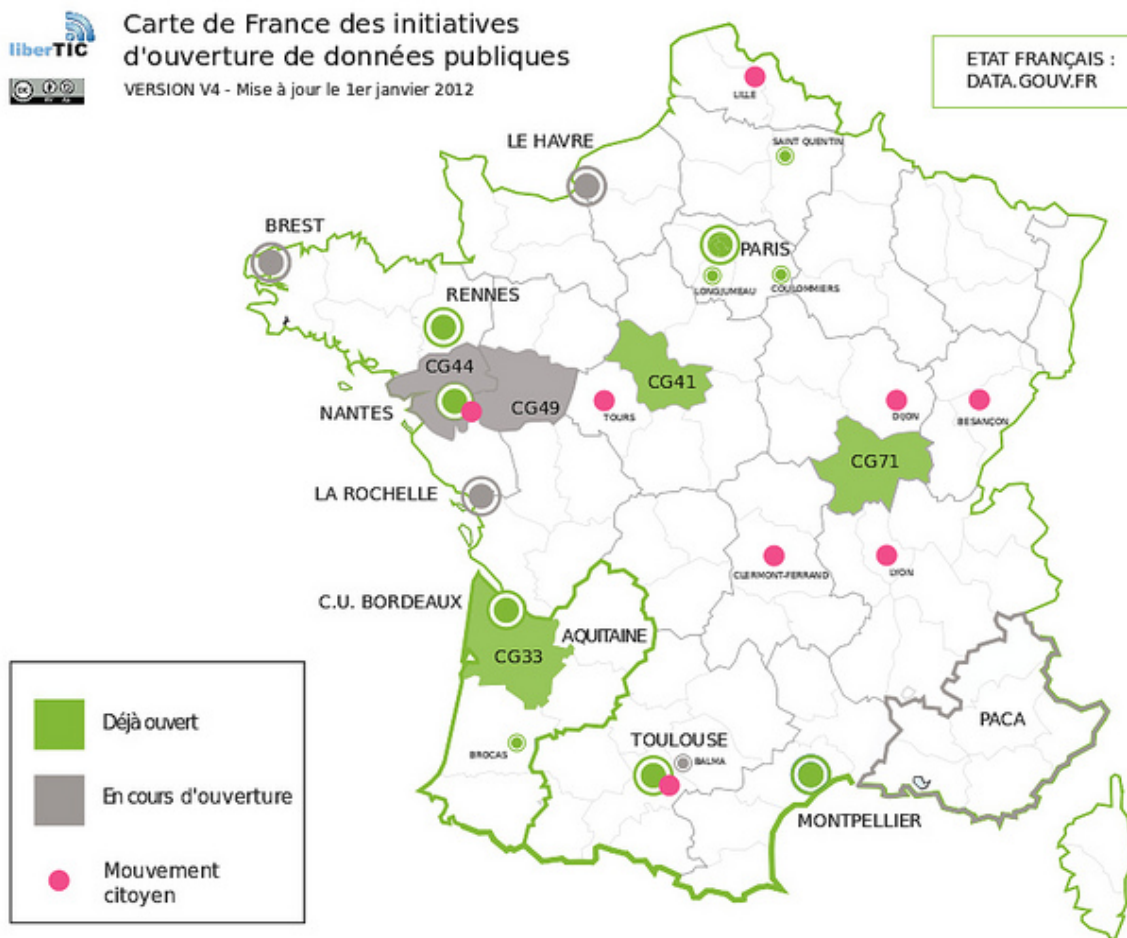
Figure 1 Map of France (V4) with the initiatives of opening public data. Last update as from January, 2012.

The movement is still evolving: in 18 months, 11 big cities from the 15 biggest ones in France will open their data. Regions, General Councils, cities and agglomerations, as well as public agencies or private enterprises (e.g. the SNCF[3]) have joint the movement. It is also to highlight the role of citizens associations that organize the co-production of data such as Open Street Map (21) or "Regards Citoyens". At the moment, apart from the data opened by the State, there are more than 800 open data sets by the territorial actors, leading to produce more than 200 applications (22).

Another movement is gaining interest in the Semantic Web community since two years: the Linked Data movement. Some authors say that "*The Semantic Web has gone mainstream*" (23) and the recent release of Google Knowledge Graph (24) after the Open Graph Protocol from Facebook (25) confirm that idea.

The Semantic Web is not just about putting data on the web. It is about making links, so that a person or machine can explore the web of data[4]. With linked data, one can find other, related data. The goal of Linked Data is to enable people to share structured data on the Web as easily as they can share documents today. The term

---

[3] http://data.sncf.com/
[4] http://linkeddata.org/

6

Linked Data was coined by Tim Berners-Lee in his Linked Data Web architecture note (26). The term refers to a style of publishing and interlinking structured data on the Web. The basic assumption behind Linked Data is that the value and usefulness of data increases the more it is interlinked with other data. In summary, Linked Data is simply about using the Web to create typed links between data from different sources.

As stated in (26), the web of data is an enhancement over the web of documents where links are established between arbitrary things described in RDF[5], and where URIs[6] are used to identify any kind of object or concept. In his famous note, Tim Berners Lee also describes the so-called four linked data principles (26):

1. Use URIs as names for things

2. Use HTTP URIs so that people can look up those names.

3. When someone looks up a URI, provide useful information, using the standards (RDF, HTML[7], SPARQL)

4. Include links to other URIs. so that they can discover more things.

Based on these principles, the web of data has grown since 2007. The Figure 2 shows the actual state of the Linked Open Data Cloud.
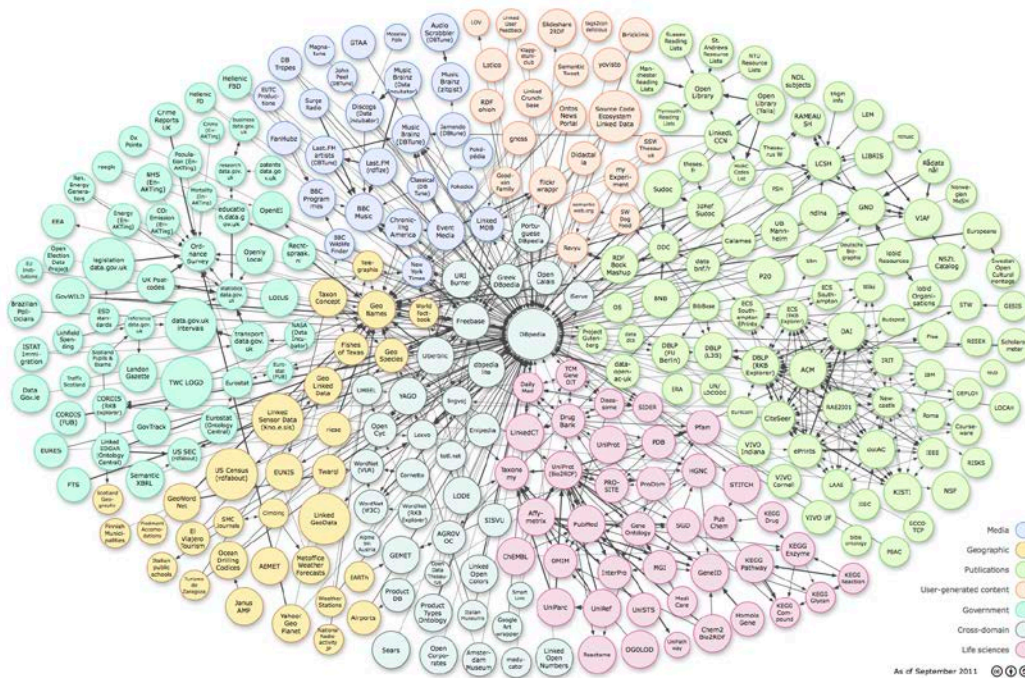
Figure 2 Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

---

[5] http://en.wikipedia.org/wiki/Resource_Description_Framework
[6] http://en.wikipedia.org/wiki/URI
[7] Content negotiation is generally implemented in order to serve either an HTML (for humans) or a RDF (for machines) representation of the resource identified by the linked data URI.

Linked *Open* Data (LOD) is Linked Data which is released under an open license, which does not impede its reuse for free. Creative Commons CC-BY is an example of open license, as is the UK's Open Government License [OGL]. Linked Data does not have to be open -- there is a lot of important use of linked data internally, and for personal and group-wide data. In 2010, in order to encourage people -- especially government data owners -- along the road to offer good linked data, Tim Berners-Lee developed a five stars rating system. A dataset can have 5-star Linked Data without it being open. However, if it claims to be Linked Open Data, then it must be open to get any star at all.

Under the star scheme, one get one (big!) star if the information has been made public at all, even if it is a photo of a scan of a fax of a table -- if it has an open license. The five stars dataset means the data is published in RDF, entities are identified by HTTP URIs, and are linked to other datasets to enrich information. The table below gives the rates and the meaning of each of the stars for ranking datasets while these 5 stars are nicely illustrated by example at http://lab.linkeddata.deri.ie/2010/star-scheme-by-example/.

★	**Available on the web (whatever format)** *but with an open license, to be Open Data*
★★	As (1) plus available as machine-readable structured data (e.g. excel instead of image scan of a table)
★★★	as (2) plus non-proprietary format (e.g. CSV instead of excel)
★★★★	All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
★★★★★	All the above, plus: Link your data to other people's data to provide context

In this document, we first review the numerous applications that have been developed on top of datasets that have been opened by governments and local authorities. We then derive use cases (UC) that can be developed to consume data from the different providers of the Datalift project: INSEE, DILA, IGN, FING, etc. These UCs, developed and deployed, can be useful to show the benefits of Linked Data in a variety of domains such as Education, Tourism, Cultural Heritage, Civil administrations, Judicial Court, Medicine, etc.

This document is structured as follow: we first describe some innovative applications consuming open data in UK, USA and France. We then present some relevant use cases that could be developed in the DataLift platform, considering the type of data "lifted" from the data providers in the project (INSEE, IGN and DILA). We next capture some requirements from the users' perspectives and conclude in the last section.

2. Innovative Applications consuming Open Data in United Kingdom (UK)

We list here some applications chosen for their diversity, their innovation and the richness re-use of the different data sets. We observe that most of them are map oriented, i.e. the geo-localization of the user is often the starting point of the data exploration journey.

2.1. UK Crime

UKCrimeStats (27) is the UK's only crime ranking platform for Neighbourhoods, Police Forces and Streets with maps, analysis and reports. UKCrimeStats also allocates an individual crime id, map and url page for each registered crime. It is a platform that aims to close the gap between policy and knowledge in new and innovative ways.

For the first time, using the official data of the 43 Police Forces of England and Wales and taking it to a new level of analysis, UKCrimeStats can answer the following questions:

- Which of the 43 Police Forces has the highest or lowest crime rate / total crime / type of crime in which month / over the selected months / and how do they rank against one another?
- Which neighbourhoods in England and Wales have the highest total / violent / vehicle / robbery / other / asbo crime or crime rate in a month or between two months?
- Which streets in England and Wales had the highest total / violent / vehicle / robbery / other / asbo crime in a month or between two months?
- Where does my neighbourhood rank in the Neighbourhood Crime League table? (all neighbourhoods within 5 miles of a given postcode).

Crime Plus ASB Data At A Glance

Updated with crime data from March 2012

Crime and ASB Breakdown - March 2012

- Other Crime 186,283
- Violent Crime 55,268
- Vehicle Crime 35,248
- Burglary 41,247
- Robbery 6,043
- Anti-Social Behaviour 213,677

©Copyright UKCrimeStats.com

Crime and ASB Totals

- ASB
- Robbery
- Burglary
- Vehicle
- Violent
- Other
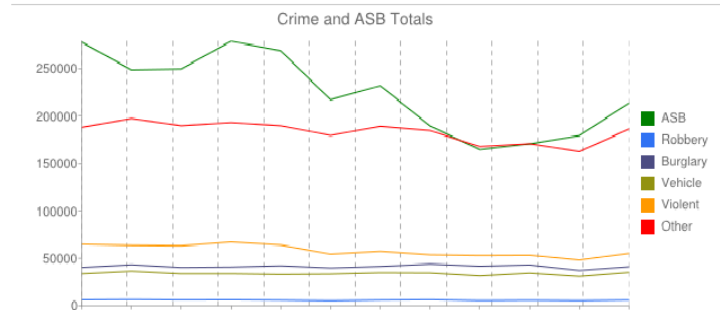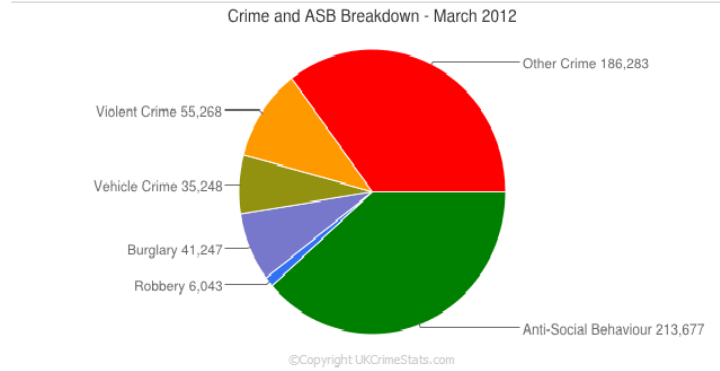
Total all Crimes in England and Wales: last 2 months

Crime	Feb 2012	Mar 2012	Change
Anti-Social Behaviour	179,253	213,677	+34,424
Robbery	5,433	6,043	+610
Burglary	37,247	41,247	+4,000
Vehicle Crime	30,895	35,248	+4,353
Violent Crime	48,100	55,268	+7,168
Other Crime	14,452	15,895	+1,443
Theft - Shoplifting	24,212	27,017	+2,805
Drugs	16,278	18,467	+2,189
Criminal Damage and Arson	44,050	49,763	+5,713
Public Disorder and Weapons	10,676	12,389	+1,713
Theft - Other	52,645	62,752	+10,107
Total Crimes Plus ASB	**463,241**	**537,766**	**+74,525**

Figure 3:A screenshot of the Crime UK application.

Name	UK Crime
Scope/Domain	Crimes in UK
Description	The application helps to find and to visualize information about crimes in a given radius in UK.
Supported Platform	Web
URL Policy	The URI pattern is http://www.ukcrimestats.com//Neighbourhood/COUNTY_Police/CITY#Crime where • COUNTY_Police is the name of the police division where the city belongs, • CITY is the name of the city.

	e.g: http://www.ukcrimestats.com//Neighbourhood/Nottinghamshire_Police/Old_Meadows#Crime gives the crimes statistics for the city of "Old_Meadows". For the streets crime, the URI pattern is http://www.ukcrimestats.com/Street_Crime/ID/ e.g: http://www.ukcrimestats.com/Street_Crime/8453235/
Data Source	UK crime data in England & Wale in Excel (xls)
Type of views	Map, chart, pie chart, tabular views
Visualization tools	Google Maps API, Jquery
Drawback	--
License	Not available
Business Value	Free for basic access. Fees for additional access (points of interest, school area, etc)

2.2. UK Pharmacy

This application helps you find the nearest pharmacy (28) quickly and easily on an Android, iPhone or iPod Touch mobile device. UK Pharmacy is a native application for mobile phone, without a version online for desktops. The application helps to search for a pharmacy/chemist around a current location using the GPS built in the phone, or look for one in another area (whilst out, or for a distant relative) via a simple place name or postcode search. The main dataset provider is the Health and Social Care Information Center which covers 99.37% of all the pharmacies in England, with nearly 10,000 pharmacies.
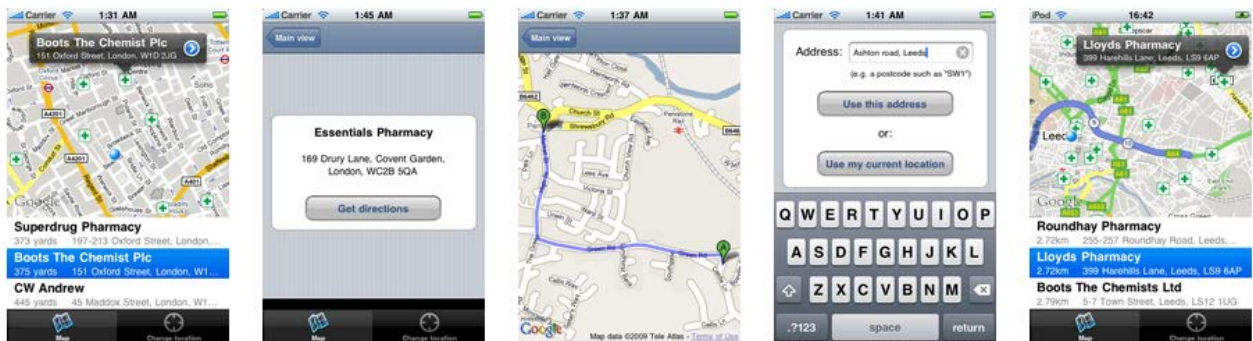
Figure 4: A screenshot of the UK Pharmacy application version for mobile phone.

Name	UK Pharmacy
Scope/Domain	Health, Pharmacy
Description	The application helps searching for a pharmacy/chemist around a current location using the GPS built in the phone, or look for one in another area (whilst out, or for a distant relative) via a simple place name or postcode search.
Supported Platform	Mobile (Android, iPhone, iPad)
URL Policy	---
Data Source	Health and Social Care Information Centre (HSCIC) under the license of the Office of Public Sector Information (OPSI).

Type of views	Map view
Visualization tools	--
Drawback	• Proprietary • Difficult to understand how the data is collected and transformed
License	All rights reserved to Elbatrop Ltd
Business Value	Available for FREE on the Apple App Store and from Google Play with in-app ads. Paid version without ads.

2.3. Numberhood

The Numberhood application (29) for iPhone and iPad shows how your local area fares on the important issues – the economy, unemployment, education, health, crime, housing and community strength. Data is shown for key issues and indicators, with data charts and descriptions to help us understand patterns and trends. Numberhood is currently available for iPhone/ iPad from the App Store, with Android and Blackberry versions planned for the future. The free application covers 11 headline indicators, with an in-app purchase opening up 70 more indicators. The possible users are local communities, councillors, researchers, students and house-movers.

The data used in the Numberhood application has been collected from two main providers: data.gov.uk and data4nr.net. It also collates data from a wide range of sources: *Annual Population Survey (APS), Annual Survey of Hours and Earnings (ASHE), Census 2001, Communities and Local Government (CLG), Department for Business Innovation and Skills (BIS), Department for Business, Enterprise and Regulatory Reform (BERR), Department for Culture Media and Sport (DCMS), Department for Environment, Food and Rural Affairs (DEFRA) - WasteDataFlow system, Department for Education (DfE), Department for Work and Pensions (DWP), Department of Health (DoH), Home Office, Office for National Statistics (ONS), Place Survey, Small Business Service (SBS), Youth Offending Team (YOT).*

Name	Numberhood
Scope/Domain	Local area dynamics
Description	The application shows how your local area fares on the important issues – the economy, unemployment, education, health, crime, housing and community strength.
Supported Platform	iPhone/iPad
URL Policy	----
Data Source	Data.gov.uk, data4nr.net, many other reports (population survey, unemployment, small business service, etc)
Type of views	Chart, histogramm, tabular views
Visualization tools	--
Drawback	• Proprietary • Difficult to understand how the data is collected and transformed.
License	All rights reserved to Oxford Consultants for Social Inclusion (OCSI)
Business Value	Commercial

2.4.   BUSit London

BUSit London (30) uses London Bus data from Transport for London to plan your multi-leg bus journey in the capital. Simply let the web application detect you current location and then click on the map to indicate your destination.

BUSit will then plan your route for you indicating which buses to catch, where to catch them and where to change. Drag the markers to change your start or destination points, or alternatively search by address, postcode or place name. The application is mobile ready using a simplified compact interface when being accessed on mobile browsers.

The main provider of the data is Transport of London[8] [TFL].
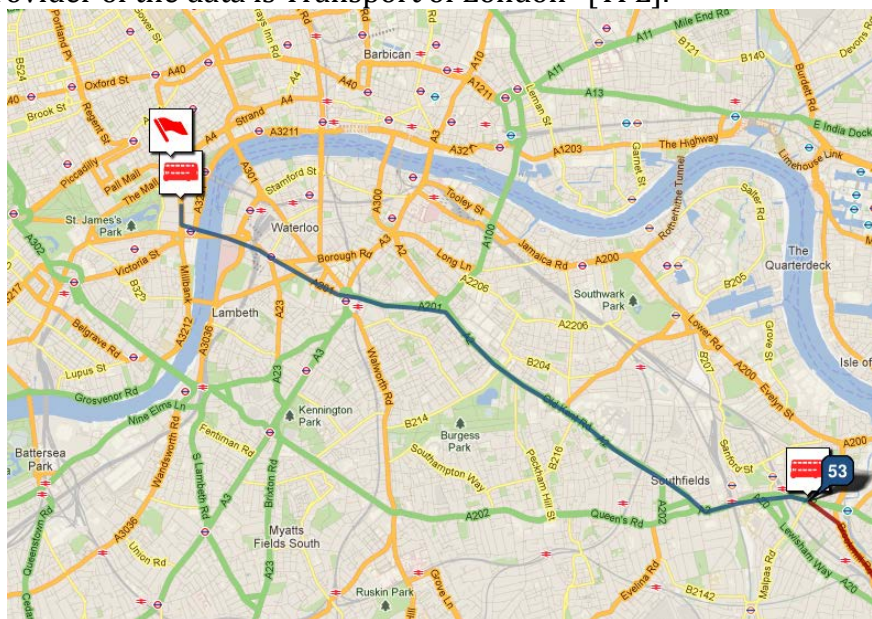
---

[8] http://www.tfl.gov.uk/businessandpartners/syndication/default.aspx

Figure 5: BuSit London screenshot, results from searching a route from London to Lewisham.

Name	BuSit London
Scope/Domain	Public Transportation in London
Description	The application helps users to plan a multi-leg bus journey in London based on his current location. BUSit will then plan the route based on the destination provided with route indicating which buses to catch, where to catch them and where to change. It includes features to change the start or destination points, or alternatively search by address, postcode or place name.
Supported Platform	Web and mobile
URL Policy	No URI policy. All the searching and results functionalities in the main domain: http://www.busitlondon.co.uk/
Data Source	Transport of London (TfL) The data is obtained by syndication from the TfL either in csv or xls via an API.
Type of Views	Map view
Visualization tools	Google Maps API, Jquery
Drawback	• No statistics on estimation time for the journey • Not linked to other data sets.
License	Not available
Business Value	Not commercial

2.5.   UK School Finder

UK School Finder (31) is an HTML5 Web Application that helps you easily find great schools. It provides comprehensive statistics including test results, size, location, age group, absence rate to help evaluating the school quality.
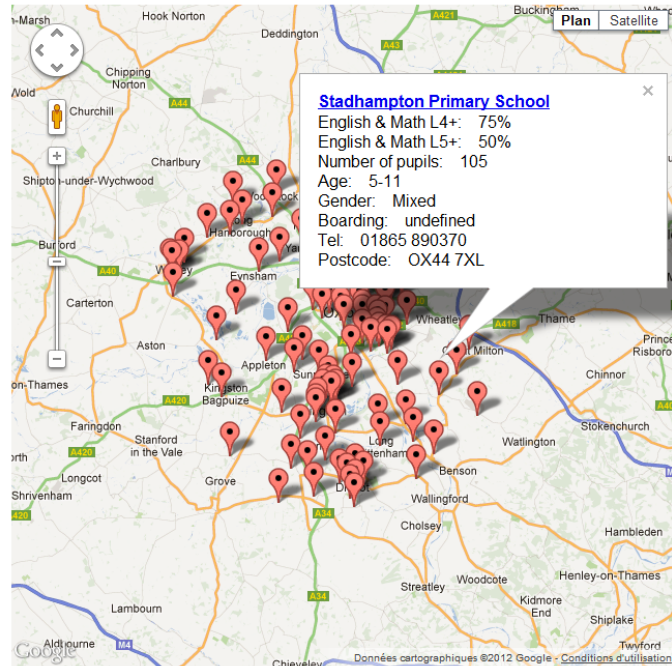
The data is published by the UK department for education, mainly using their data, with more than 30,000 primary school throughout UK.

Figure 6: Partial result of searching schools in Oxford

Name	Grade	English&Math L4+	English&Math L5+	Pupils	Age
St Ebbe's Church of England Aided Primary School	Good	75%	23%	308	5-11
St Barnabas' Church of England Aided Primary School	Good	68%	29%	238	4-11
East Oxford Primary School	Satisfactory	36%	7%	280	4-11
West Oxford Community Primary School		92%	69%	147	3-11

Name	UK School Finder
Scope/Domain	Primary education
Description	The application helps find and visualize information about schools in a given radius in UK, and for each school provides more statistics such as grades, inspection judgment, or comparison metrics.
Supported Platform	Web
URL Policy	http://www.ukschoolfinder.com/ for the main page. http://www.ukschoolfinder.com/#details?id=ID to have the details for a school with a given ID. e.g: http://www.ukschoolfinder.com/#details?id=101127 (URI for the statistics for the "St Clement Daves CofE Primary School" near London.
Data Source	UK inspection Report (pdf)[9]. No information about how the data source format.
Type of Views	Map, tabular and histogram views.
Visualization tools	Google Maps API, Jquery
Drawback	• Links to pdf documents • No best practices for URI policy • Not linked to other data sets (e.g.: transportation, county, etc.)
License	Not available

---

[9] http://www.ofsted.gov.uk/inspection-reports/

Business Value	Not commercial

2.6.   Where-can-I-Live

The application (32) is based on the assumption that people want to live where they can easily get to work. It works on commuting time between stations (Underground and Docklands Light Railway only) and average house prices. By giving which station someone works near and how much he/she can afford, the application shows results of possible options where to live. It is also possible to view property listings around the chosen station and some house price statistics. The Average house prices are updated monthly and the travel time is calculated only between stations.

Primarily, Where-can-I-Live was designed for the city of London. Similarly the same concept is applied for 7 other cities in Europe: Madrid, Barcelona, Berlin, Franckfurt, Munich, Milano and Roma. It could be interested to apply the Linked Data principles to similar application for big cities in France: Paris, Bordeaux, Lyon and Marseille.
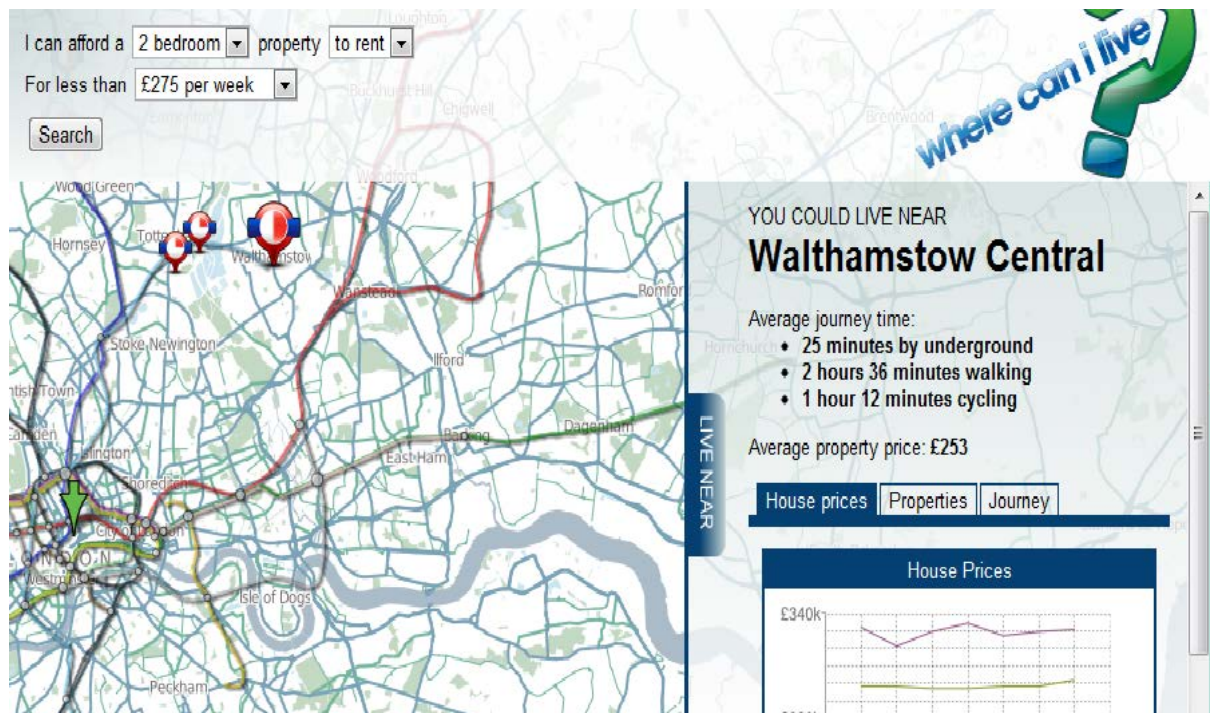
Figure 7: View of the application where-i-can-live

Name	Where-can-I-Live
Scope/Domain	House pricing, properties, public transportations
Description	The application helps find and visualize information about house pricing depending on which underground station specified, calculate the time distance from your work, and recommend houses depending on some criteria (price, etc). It also shows statistics of the house pricing updated each month.
Supported Platform	Web
URL Policy	http://www.where-can-i-live.com/londonproperty/[UNDERGOUND-NAME]/[TIME TO

	STATION]/[BEDROOM NUMBER]/[PROPERTY TYPE]/ [MAX PRICE] e.g: Results of searching in "Covent Garden", a 2 bedrooms to rent, for less than £300 per week, and not far more than 30 min. http://www.where-can-i-live.com/londonproperty/covent-garden/time-30/bedrooms-2/ltype-rent/maxprice-300
Data Source	House pricing, DLR
Type of Views	Map, chart
Visualization tools	YUI Library: http://yuilibrary.com/, php, google maps, javascript
Drawback	• Metadata not shown • No more data to the district where you live (crimes, school, leisure, etc) • Not linked to external data sets (Wikipedia, OS, etc)
License	Not available
Business Value	Not commercial

2.7. Open data communities

This web application (33) provides a selection of statistics on Local Government finance, housing and homelessness and deprivation, as well as supporting geographical data. All of the data is available as fully browsable and queryable Linked Data, and are free to re-use under the Open Government License. A selection of the data from this site is available via a searchable Dashboard.

Figure 8: Screenshot of the Open Data Communities dashboard, for the case of "Birmingham".

Name	Opendatacommunities
Scope/Domain	Department for Communities and Local Government, datasets access
Description	The application helps find and visualize available datasets (finance, housing, deprivation, geography) by authorities or postcode. On the dashboard, it provides graphs showing the national distribution of a district and how the values for this local authority compare with others in England.
Supported Platform	Web
URL Policy	http://{domain}/**id**/{...} with redirection to the corresponding document at: http://{domain}/**doc**/{...}. For example, in the Local Authorities Dataset, the identifier for the Hampshire County Council is: http://opendatacommunities.org/id/county-council/hampshire. If you put this into your browser you get redirected to an HTML page about Hampshire: http://opendatacommunities.org/doc/county-council/hampshire
Data Source	36 datasets from DCLG, and Administrative Geography and Postcodes from Ordnance Survey.
Type of Views	Graph, Map views.
Visualization tools	HTML5 elements, Intern scripts from Swirrl, maps.googleapis.com, Javascripts (Raphael Library)
Drawback	-
License	Data are open following the open government license [OGL]
Business Value	Not commercial

3. Innovative Applications consuming Open Data in United States (US)

3.1. FlyOnTime

FlyOnTime (34) aims at finding the most on-time flight between two airports or check how late a given flight is on average, in good weather and bad, before someone leaves. The flight and weather information presented on the website is derived from data provided by the United States federal government, while the security line times are submitted by visitors.

This website has four goals:
a. Help American air travelers find the most on-time flights.
b. Help the American public understand the data from government sources by presenting it in interesting and attractive ways.
c. Tap the wisdom of crowds to collect data on airport security line delays.
d. Allow developers to access flight on-time data in a simple, programmatic way[10].
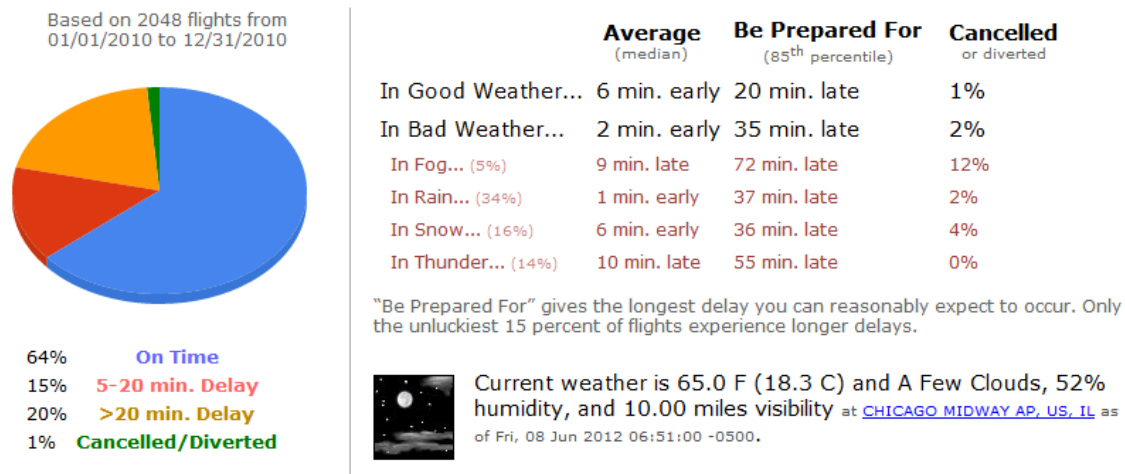
Figure 9: A screenshot of the results when searching the flight from "Chicago, IL" to "Los Angeles, CA"

Name	FlyOnTime
Scope/Domain	Flights, airlines, weather
Description	The application aims at finding the most on-time flight between two airports or check how late a given flight is on average, in good weather and bad, before someone leaves. It also shows statistics on the best days of the week and time of a day to flight from a given

---

[10] http://flyontime.us/developers

	airport.
Supported Platform	Web
URL Policy	The data are exposed as XML and JSON with an API. http://www.flyontime.us/CATEGORY/ where CATEGORY is one of airlines, flights, airports and routes. e.g.: http://flyontime.us/airlines/ (list all the airlines) There are more patterns for /flights and /airports. http://flyontime.us/flights/AIRLINECODE/NUMBER/DEP/ARRI where NUMAIRLINE is the airline code, NUMBER is the flight number, DEP is the departure airport and ARRI is the arrival airport. e.g.: http://flyontime.us/flights/UA/900/LAX/ORD (flight number 900 of the United Airlines (UA) from LAX to ORD.
Data Source	Bureau of Transportation Statistics via data.gov The Federal Aviation Administration The National Oceanic and Atmosphere Administration. *The formats of the data are not explicitly mentioned, but suspect to be compliant to Google Visualization API (csv or DSP)*
Type of views	Pie chart, Histogram, Tabular views.
Visualization tools	Css, Google Visualization (chart, pie chart),
Drawback	• No more information about "things" • Not linked to other external data sets
License	MIT License (http://www.opensource.org/licenses/mit-license.php) for the code source[11]. The website is completely open source.
Business Value	Not commercial

[11] http://code.google.com/p/flyontime/

3.2. White House Visitor Search

The application (35) lets users search visitors and visitees of the White House. The top 100 frequent visited people in the White House are listed.
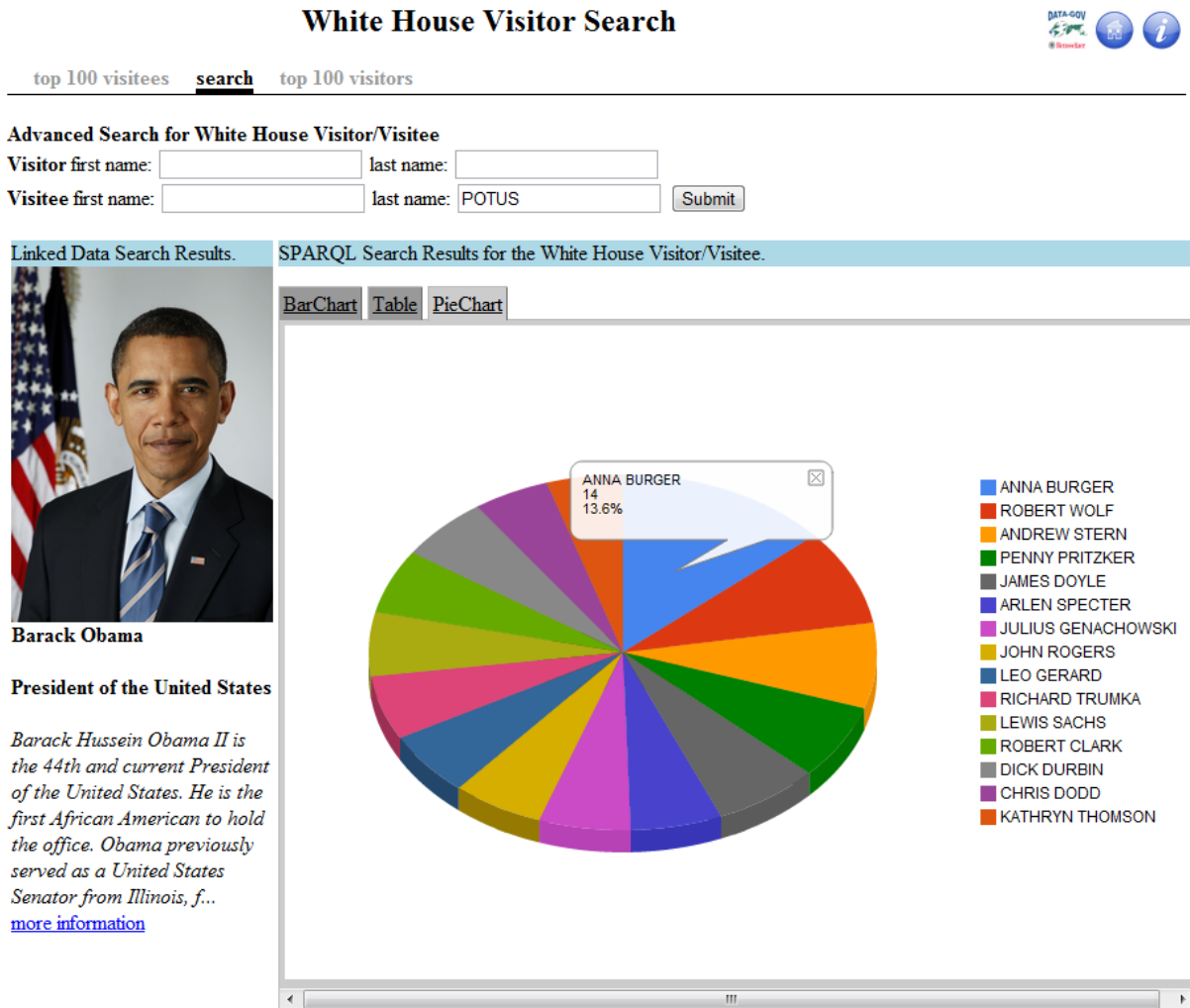
Figure 10: A screenshot of the White House Visitor Application

Name	White House Visitor Search
Scope/Domain	White House (U.S)
Description	The application demo lets users search visitors and visitees of the White House. The top 100 frequent visited people in the White House are listed.
Supported Platform	Web
URL Policy	The application is a proof-of-concept, so there is no stable URL policy
Data Source	The White House Visitor Record[12], The White House Visitee Record and White House staff[13] *The data was transformed in RDF reusing the following ontologies:*

---

[12] http://www.whitehouse.gov/files/disclosures/visitors/WhiteHouse-WAVES-Released-0310.csv

[13] http://www.whitehouse.gov/briefing-room/disclosures/visitor-records

	FOAF, SKOS and SWIVT
Type of Views	Bar chart, pie chart, tabular views.
Visualization tools	Google Visualization Table, Google Visualization API
Drawback	• Links to Wikipedia pages sometimes, not to data sets • Not linked to other external data sets
License	The application is Open Source.
Business Value	Not commercial

3.3. Comparing US-USAID and UK-DFID Global Foreign Aid

The application (36) presents a mashup of foreign aid data (represented in US Dollars) from the United States Agency for International Development (USAID) and UK Department for International Development (DFID) for the 2007 US Fiscal Year. Users may retrieve foreign aid data for specific countries by clicking on a provided world map (shaded based on total combined contributions for USAID and DFID). Upon clicking on a desired country, three kinds of information are presented:

- Aid Figures - These are broken into three categories, each presented through a pie chart: a comparison of total USAID and DFID spending, a breakdown of USAID spending by category, and a breakdown of DFID spending by category.
- News Articles – The data is retrieved through the New York Times Article Search API (14). For an article to be listed, it must have the selected country's name in the title and must contain the word "aid".
- DBpedia - Information from a selected country's DBpedia (dbpdia.org) entry.

To effectively compare DFID figures with those from USAID, the authors' conditions had to be met. First, as the US and UK use different currencies, conversion from Pound Sterling to US Dollars was required for DFID figures. Second, as DFID data was reported based on the UK fiscal year, conversion of figures to the US Fiscal Year was required. These conditions ensure that figures are represented according to a common fiscal unit and time frame, respectively. However, as manipulations to official DFID figures were required, the authoritative power of this data is possibly diminished. As such, this application presents a compelling case study for the use of provenance in tracking alterations to government data needed for mashup purposes. For the exchange rate (USD to GBP), the authors used the Oanda website[15] with the appropriate settings for date.

---

[14] http://developer.nytimes.com/docs/article_search_api
[15] http://www.oanda.com/currency/average.
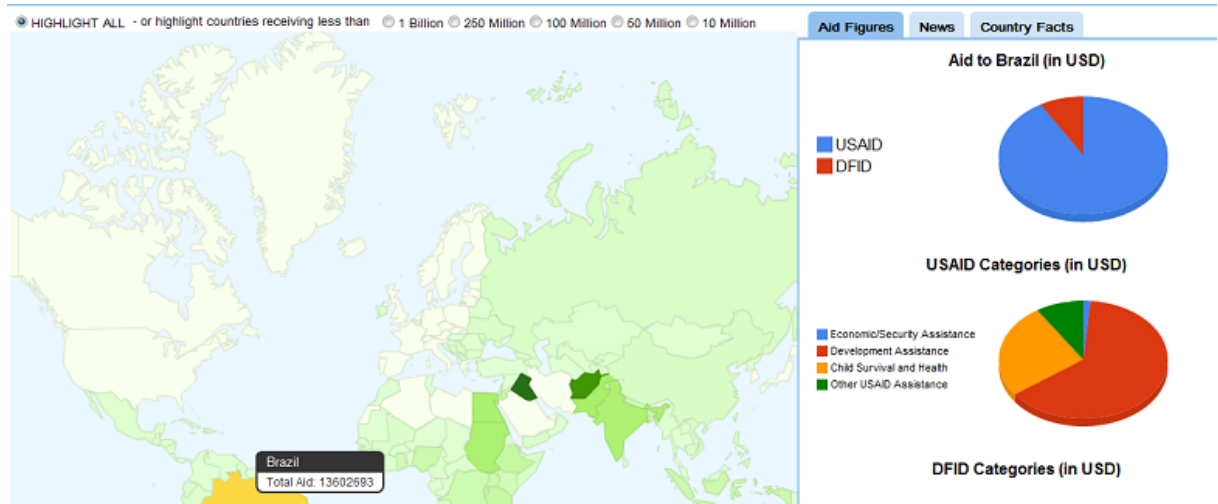
Figure 11: A screenshof of the foreign aid to Brasil

Name	White House Visitor Search
Scope/Domain	Foreign AID in UK and US
Description	The application presents a mashup of foreign aid data (represented in US Dollars) from the United States Agency for International Development (USAID) and UK Department for International Development (DFID) for the 2007 US Fiscal Year.
Supported Platform	Web
URL Policy	The application is a proof-of-concept, so there is no stable URL policy
Data Source	Data.gov/dataset 10030, DBpedia, DFID *The data was transformed in RDF*
Type of Views	Map, Pie chart views.
Visualization tools	Google Maps, Google Visualization API, Google Web Toolkit
Drawback	• Links to Wikipedia pages sometimes, not to data sets • Not linked to other external data sets
License	The application is Open Source.
Business Value	Not commercial

4. Innovative Applications using Open Data in France

Here we list some innovative applications using open Data in France. Some of them have been selected to be in the top of the best applications by Etalab, supported by data.gouv.fr.

4.1.   Fourmisante

The application "Fourmisante" (37) helps to find and visualize information about medical doctors by domain, and gives html pages about a given doctor, where to find the cheapest drugs and how the social security works for a consultation.
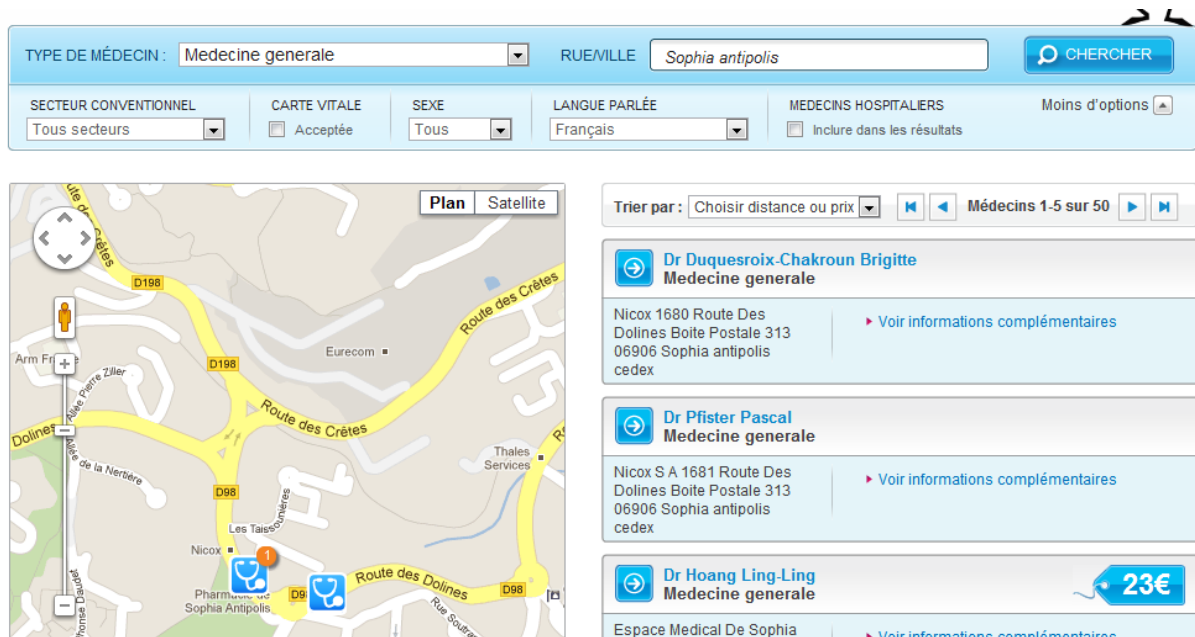
Figure 12: Screenshot of the result of the search in "Sophia Antipolis" for general doctors.

Name	Fourmisante
Scope/Domain	Medicine, health-care, social security,
Description	The application helps find and visualize information about medical doctors by domain; gives html pages about a given doctor, where to find cheapest drugs and how works the social security for a consultation.
Supported Platform	Web
URL Policy	*http://www.fourmisante.com/CATEGORY.html#ProfessionalSearchForm[city]=CITY/page=XX* where CATEGORY ={medecins, dentiste, optique, kine, osteopathe, ambulances, pharmacies, produits, audition} CITY= is the city for searching Page=XX: number of the page e.g: Results of searching for "medecins" in "Paris" http://www.fourmisante.com/medecins.html#ProfessionalSearchForm[city]=Paris/page=1 The information for each practitioner is of the form: *http://www.fourmisante.com/CODE-CATEGORY-SURNAME-*

	NAME.html e.g: http://www.fourmisante.com/11618-medecin-beaumont-marie-louise.html
Data Source	Not specified...They claim to use data from official services, health local services, association of pharmacists; Thesorimed DB, Pages Jaunes.
Type of Views	Map view
Visualization tools	Css, jquery, Sizzle.js, maps.googleapis.com
Drawback	- Metadata not shown - many .html pages - Not linked to external data sets (wikipedia, INSEE, DILA, etc) - No interaction with the data on the map. - Not coherent URI policy
License	All right reserved AMARPI SAS.
Business Value	Not commercial

4.2.   MaVilleVueDuCiel

MaVilleVueDuCiel (38) is an innovative Internet application that aims to highlight the vital forces of local government and commerce to boost the activity. In a Google Maps interface (satellite view), are pinpointed all municipal buildings, practical information as well as all the shops. The added values they also claim to have are the following:

- The concept of "everything under one roof", emphasizing a certain comfort in navigation and rapid access to information (thematic navigation and intuitive);
- Ease of access to information services and technical planning;
- The ability to add value to buildings and equipment by a 360 virtual tour and an elevation in 3D;
- The implementation of promotional tools Commerce: promotional space, additional exposure, 360 virtual tour of the shop;
- A private access that allows municipalities to have a partner internal management tool to help them locate, for example, their land assets, property preempted, assets at risk as well as any relevant or sensitive information;
- A private access which allows businesses to manage their own updates and promotions.
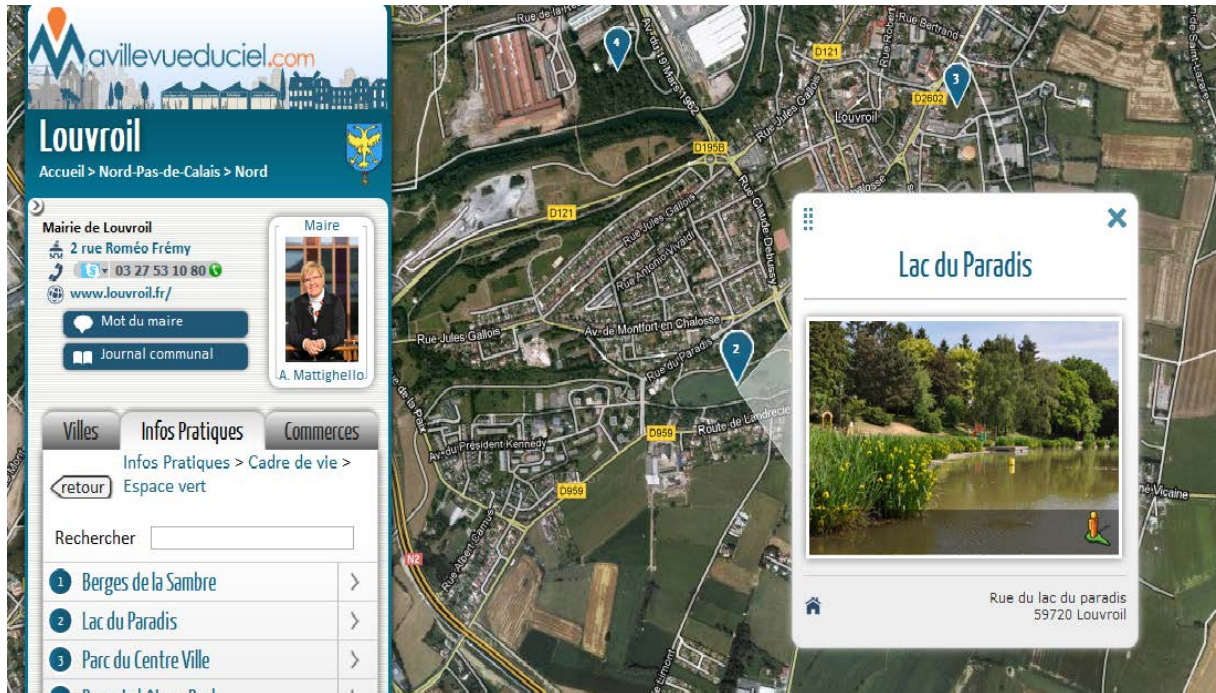
Figure 13: Screenshot for features in the city of "Louvroil" in the region of "Nord-Pas-de- Calais".

Name	MaVilleVueDuCiel:
Scope/Domain	Local government, commerce
Description	The application aims to highlight the vital forces of local government and commerce to boost the activity. In a Google Maps interface (satellite view), are pinpointed all municipal buildings, practical information as well as all the shops.
Supported Platform	Web
URL Policy	http://www.mavillevueduciel.com/ID/CITY/NUMBER where ID is an identifier, CITY is the name of the current, NUMBER a number. e.g: Information about the city of Louvril. http://www.mavillevueduciel.com/-23/Louvroil/1 But from the URI, it is not possible to know from with department or region it belongs.
Data Source	Not specified...
Type of views	Map, facets views
Visualization tools	Css, jquery, maps.googleapis.com
Drawback	• The actual data provided cover only Nord-Pas-de-Calais • Not linked to external data sets (wikipedia, INSEE, DILA, etc) • Depict only image and address • No coherent URI policy
License	All right reserved to the owner of the website.
Business Value	Not commercial

4.3. Home'n'Go

Home'n'go (39) is a service intended to simplify finding apartment by centralizing all the offers, and also points to the nearest offers next to your location. There exists three ways to use the service: (a) via a URL address of the offer, (b) via an extension on a navigator and (c) via a Bookmarklet. Home'n'Go gives a private space for users to manage their search, plot them on a map, manage the agenda, and collaboration with other users.

The application uses open data from data.gouv.fr and INSEE data of nine categories: education, finance, housing, parks, politics, population, catering, health and transportation. All the list of the data source is provided in their website.[16]
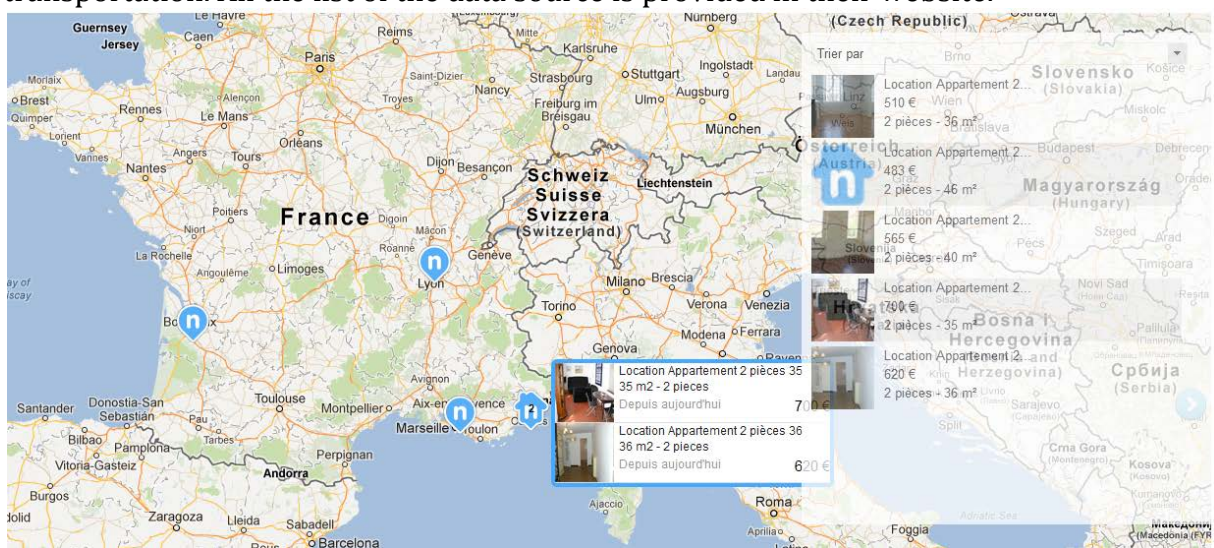
Figure 14: Screenshot of the results of some apartment offers in Antibes, Bordeaux and Lyon.

Name	Home'n'Go
Scope/Domain	Housing
Description	The application aims to help centralize housing offers and show more statistical data regarding the location such as variation of prices, population, education, etc..
URL Policy	Two types of policy is applied: 1. For user's private space: http://www.homengo.com/ma-recherche/CATEGORIE/ where CATEGORIE is carte, agenda, recherché, annonces, aperçu. e.g: http://homengo.com/ma-recherche/recherche/ for the search space. 2. For housing statistics and related information: http://www.homengo.com/immobilier/REGION/DEPARTMENT/COMMUNE e.g: http://homengo.com/immobilier/bretagne/cotes-d-armor/
Data Source	High school statistics (2011), taxes by department (2011), data on housing (2011), Parks in Montpellier, Nantes (2011), Regional

---

[16] http://homengo.com/immobilier/sources/

	elections (2010), presidential election (2012), data on population (2011), catering from opendata71 (2011), statistics of hospitals (2011), metros and velibs of Paris (2011), Transportation data of Toulouse and Nantes (2012). The formats are mainly csv and xls.
Type of Views	Map and tabular views.
Visualization tools	Css, Jquery, maps.googleapis.com
Drawback	• No more information about "things" • Not linked to other external data sets • Need to have an account to play with
License	All right reserved to the owner of the website.
Business Value	Not commercial. But they plan to add ads for companies

5. Report on Publication of Government Linked Data Worldwide

Recently, the W3C GLD WG (40) has made a progress report on publication of Government Linked Data Worldwide (41). Here we select some of them as they are datasets in 4-5 stars. In this report, Data.gov.uk remains the leader with the most published 5 star government LOD[17], while US is making some solid progress with 4-5 star LOD in 2012. The list below summarizes some of these big initiatives:

- The UK Environment Agency: Water Quality. Five star publication of assessments of the water quality at Bathing ("Swimming" for a US audience) sites in England and Wales. The data is a linking of three data streams - annual quality assessments, weekly in-season assessments and profiles (describes environmentally significant features and plans for each site). The data is live - updated at least weekly and sometimes daily as new assessments arrive (42). External links include geographic links (to Ordnance Survey LD], administrations and legislation (http://www.legislation.gov.uk/)
- The University of Southampton Data Service (43) : list 5 data sets for the university including buildings, places, educational resources, events, facilities & equipment, calendar info, student statistics & transportation information.
- The US Library of Congress Linked Data Service (44) includes vocabularies and authorities owned and maintained by the Library of Congress made available as Linked Data.
- The Clean Energy Data (45) for high quality information on renewables, energy efficiency and climate change to help accelerate the clean energy data, manuals and endpoint as 5 star LOD for re-use.
- The Spanish Geographic Data (46) includes information sources from the National Geographic Institute of Spain, and the National Statistical Institute in Spain.
- The XBRL dataset for Edgar, US Securities and Exchange Commission required filings (47) : The project is called the Edgar Linked Data Wrapper, and leverages the RDF Data Cube vocabulary. Companies are linked to Freebase.

---

[17] http://data.gov.uk.

6. Use Cases for DataLift

In this section we describe some scenario that could be innovative for our platform and taking into account the different providers present in the DataLift project.

6.1.   Find the "perfect" Judicial Court.

ID:	**UC-1**
Title:	**Find the "Perfect" Judicial Court**
Description:	This application displays the result of the research of a judicial court depending on three parameters: the location (e.g.: Paris, Valbonne), the type of the jurisdiction[18] and a given radius.
Primary Actor:	A user looking for a Judicial Court.
Preconditions:	A map centered in the location where the user is geo-localized by default or Paris as fallback, with a level of zoom at 13.
Postconditions:	The maps with tiles (of different colors by type of Courts) with the services location, information (aperture, summary, access, etc..), with links to more information (DBpedia, web page, etc..)
Main Success Scenario:	In a form, the user enters the location and the radius (in meters) for the research. The form has two buttons: "search" for the searching purpose and the "clear" button for changing the values entered or modified them.
Extensions:	Optionally it can be added services of distance to the service (e.g.: how to arrive) or some pictures of the service (e.g.: using Google Street view). In case of error, the system should respond with an friendly alert such as "*Try another address*" or "*No Court available at the address proposed*".
	It could also be possible to take the Geolocation of the user and with that information, gives the Court based on that location, before any research query.
Frequency of Use:	As many times as permit by the different other APIs used for the application (e.g.: Google Maps, etc..)
Datasets	
Interest of the UC	
Difficulty	

6.2.   MediVeCinty

ID:	**UC-2**
Title:	**MediVeCinty**

---

[18] See http://www.data.gouv.fr/donnees/view/Donn%C3%A9es-g%C3%A9ocod%C3%A9es-des-structures-de-la-Justice-30378257?xtmc=tribunaux&xtcr=1

Description:	The application gives on a map the list of medical doctors of your department, and find the cheapest one which is the closest to your current location.
Primary Actor:	A user looking for a medical doctor
Preconditions:	A map centered in the center of Paris by default, with a level of zoom at 13. Use data from Ameli.fr by scrapping techniques. We could also use geolocation data from the browser to center the user.
Postconditions:	The maps with tiles (of different colors by prices) with the name of doctors, specialty, phone number, hospital of consultation, the ratio with the price fixed by CNAM (*Assurance Maladie*) with links to more useful information (DBpedia, web page, etc..)
Main Success Scenario:	The user explores the results on a map of France by cities, to found out the information about the medical doctors. The application can provide a search form by cities or department or even region.
Extensions:	Optionally it can be added services of distance to hospital (e.g.: how to arrive) or some pictures of the hospital street (e.g.: using Google Street view). In case of error, the system should respond with an friendly alert such as "*Try another address*" or "*No doctor available at the address proposed*". Optionally, using crowd sourcing techniques, we could also aggregate the average waiting time for a patient to get an appointment and offer this data as an additional criteria for selecting his doctor.
Frequency of Use:	As many times as permit by the different other APIs used for the application (e.g.: Google Maps, etc..)
Datasets	
Interest of the UC	
Difficulty	

6.3. OpenedDrugstore

ID:	**UC-3**
Title:	**OpenedDrugstore**
Description:	The application gives on a map the list of drugstores "opened 24 hours" (*all-night drugstore and doctor on call*)[19] during a period of time (per day, or per week at most), in a given city.
Primary Actor:	A user looking for a drugstore
Preconditions:	A map centered in the center of Paris by default, with a level of zoom at 13. We could also use geolocation data from the browser

---

[19] The so-called in french « pharmacie de garde » et « médecin de garde »

	to center the user. Use data from Ameli.fr by scrapping techniques.
Postconditions:	The map with tiles (differentiating by those which are open 24 hours) with the name of drugstore, phone number, contact.
Main Success Scenario:	The user explores the results on a map of the region, city or commune. The application can geo-locate the current position of the user to give results in a 2km radius.
Extensions:	Optionally it can be added services of distance to the drugstore. Optionally it can be added data of epidemic ("gastro", flu, etc), doctors, hospitals, anonym declaration of patients.
Frequency of Use:	As many times as permit by the different other APIs used for the application (e.g.: Google Maps, etc..)
Data sets	IGN Maps with street view, Pharmacies information (location, opening days).
Interest of the UC	Simple UC to show-case semantic concepts. UC concrete and useful to the citizens Many interfaces are possible: web, mobile, etc. Data from the health domain (not too much exploited in the open data)
Difficulty	No standard for the data, and heterogeneity of the data sources. Data are not even exposed as 1-star in some departments. Some existing applications for mobile, such as "MaPharmacie" available on Appstore[20].

6.4.   Public Service Next2U

ID:	**UC-4**
Title:	**Public Service Next2U**
Description:	The application gives on a map public services or services of interest next to your home (or elsewhere). It gives information (opening hours, etc) of local services, public national administrations
Primary Actor:	A user looking for a local or national service.
Preconditions:	A map centered in the center of Paris by default, with a level of zoom at 13. We could also use geolocation data from the browser to center the user.
Post conditions:	The maps with tiles (of different colors by prices) with the all the relevant information needed by the users to help him knowing where he/she can find what it is helpful for his needs. For more details, it provides URLs to homepages or external pages (e.g.:

---

[20] http://itunes.apple.com/fr/app/ma-pharmacie/id352231905?mt=8

	Wikipedia).
Main Success Scenario:	The user explores the results on a map of a region, city or commune.
Extensions:	Optionally it can be added services of distance to the service (e.g.: how to arrive) or some pictures service street (e.g.: using Google Street view). Also add if it has hotspots wifi or not around (e.g. data from Paris town hall[21]).

Optionally it could be integrated the possibility to locate someone's voting circumscription for local or national level elections. |
| Frequency of Use: | As many times as permit by the different other APIs used for the application (e.g.: Google Maps, etc..). |
| Datasets | DILA, Local Data, National Data, BDAdresse (IGN), INSEE, COG, BPE, Ministry of Justice, Aquitaine Data, Public Pigma data |
| Interest of the UC | Concrete case and useful for the citizens
Service could be stable since the actors are professionals
Complex and many different datasets |
| Difficulty | Interconnection and transforming some proprietary formats to RDF, such as the Pigma format.
Many actors to coordinate.
Interest of the UC according to the actual concurrence (existing applications, crowd sourcing data). |

6.5. A Pilgrimage To Religious Heritage in France

ID:	**UC-5**
Title:	**A Pilgrimage To Religious Heritage in France**
Description:	The application gives on a map the route from one location to another by placing the prominent religious along the way.
Primary Actor:	A user looking for a religious monument, church or
Preconditions:	A map centered in the center of Avignon by default, with a level of zoom at 13. We could also use geolocation data from the browser to center the user. The user is asked to enter three data: two cities (initial, end); and the Religion (Catholic, Jewish, etc..)
Post conditions:	The map with tiles (of different colors by types of features) with the name of sites, gallery of photos, date of construction, fist used and actual one, and the distance (km) of the two cities. The data could be linked to Commune and airports.
Main Success Scenario:	The user explores the results on a map with line depicting the two cities and distance.
Extensions:	--

---

[21] http://www.data-publica.com/opendata/13489--liste-des-sites-hotspots-wifi-a-paris

Frequency of Use:	As many times as permit by the different other APIs used for the application (e.g.: Google Maps, etc..)
Data Sets	
Interest of the UC	
Difficulty	

6.6. Lost And Found

ID:	**UC-6**
Title:	**Lost And Found (objets trouvés)**
Description:	The application gives on a map the adequate service where objects lost are found.
Primary Actor:	A user looking for a document lost.
Preconditions:	A map centered in the center of Paris by default, with a level of zoom at 13. We could also use geolocation data from the browser to center the user. The data from data.gouv.fr[22] can also be used, alongside with services of DILA.
Post conditions:	The maps with tiles for services "Points d'accueil" by region, department or city.
Main Success Scenario:	The user explores the results on a map of France by cities, to found out the information about the *points d'accueil*, with address, contact. The application can provide a search form by cities or department or even region.
Extensions:	Optionally it can be added information about police station with contacts in the case you want to make a reclamation when loosing something.
Frequency of Use:	As many times as permit by the different other APIs used for the application (e.g.: Google Maps, etc..)
Datasets	
Interest of the UC	
Difficulty	

6.7. Find the "Perfect" School.

ID:	**UC-7**
Title:	**Find the "Perfect" School**
Description:	The application gives on a map the top list (ranking) of schools of you department, pupils, rate at national level, private/public, fees,

---

[22] http://www.data.gouv.fr/content/search?SearchText=points+d+acceuil+police+pour+paris

	etc. If it is a university, number of students, rate at national level, private/public, price, research activities, etc. are also provided
Primary Actor:	A user looking for a school or university.
Preconditions:	A map centered in the center of Paris by default, with a level of zoom at 13. We could also use geolocation data from the browser to center the user. The user enters on a form the location, the radius for searching at the right-side of the map.
Post conditions:	The map with the desired information, with pop-up about relevant information (name, web page, type, address, fees, admission conditions, period to send candidate, etc.)
Main Success Scenario:	The user explores the results of the schools on a map. And a table at the bottom given more information about the school (name, pupil, type of education, ranking level, etc.). For universities, access conditions, ranking level (national/international), research fields covered, etc. are also provided
Extensions:	--
Frequency of Use:	As many times as permit by the different other APIs used for the application (e.g.: Google Maps, etc..)
Datasets	
Interest of the UC	
Difficulty	

6.8.Collaborative Genealogy

ID:	**UC-9**
Title:	**Collaborative genealogy**
Description:	Today, hundred thousands of people search and collect genealogical data. The Gedcom database format – well adopted and documented – allows data exchange. In fact, most Genealogical software produce their own dialect of gedcom: thus, data exchange is sometimes very difficult. Data comparison is even more complex: softwares don't do it well and not in real time. Some data are not easy to compare: e.g. syntax of places' names. Online services such as geneanet.org haven't yet addressed this issue. These services produce also data (wedding listing, etc.) and may not have interest to let people share data to easily. Public actors – such as "archives départementales" – digitalise their documents on the web. While it's a good effort, it is mainly driven by a documentalist perspective and not on a web or data perspective. For example, a page of a document is not accessible by a URI. This situation has many issues:

<table>
<tr><td rowspan="1"></td><td>

- it's bad for the research: thousands of people are wasting their time by retyping things, digitalize documents or reproduce searches that have been already done by others
- it's bad for ancient archives: while the data are not well shared, many researchers will reproduce the same surveys and thus, make more manipulations of the documents
- it's bad for collaboration
- it's bad for data quality: most of the time, data are not referenced by amateurs
- it's bad for data comparison with others historical data
- it's bad for longer term research: when a genealogist dies, his past work passes with him
- last, it's bad for innovation: private enterprises waste their time for data offering (not very innovative) instead of service offering (comparison, sharing, archiving, etc.).

So we propose to let everybody convert his genealogical data, to publish it and to perform actions and complex surveys. Here are some examples:

- I want to copy in my database this new branch while keeping the link to its original source
- I want to share my annotations/comments on a ancestor identified by one of my contacts
- Who did already worked on this ancestor and what does he says about him?
- When this surname had been common?
- Am I a cousin of this people? of this personality?
- Demographic questions such as: what is the expected life duration during the wars of the XVIIth century?
- historical questions such as: what were the local facts between 1750 and 1780 in this place?

</td></tr>
</table>

Primary Actor:	Genealogists
Preconditions:	Each user has already imported genealogical data. The user has a form to search for people/places/dates/facts...
Postconditions:	The user navigates in the results for his search.
Main Success Scenario:	The user explores the results and is allowed to import them in its own database.
Extensions:	The user can import its own database. He can enter data by hand.
Frequency of Use:	As many times as permit by the different other APIs used for the application (e.g.: OpenStreetMaps, etc.)
Datasets	• Datasets from dozen of genealogists • Some big genealogical databases (Roglo (INRIA), etc.). • DBpedia. • INSEE (places), DILA (administrative history of places (deletions, etc.) ?), IGN (hamlet)

Interest of the UC	- Real disruptive innovation in a domain involving hundred thousands of people - Many type of- and many data producers: Institutions (archives, etc.), private producers, non-profit organizations, individuals. - Non data-geeks producers - Functional: how to interconnect data? - Data curation and data access (for privacy concerns) - Frequent updates, feedbacks, between data sources
Difficulty	- Private actors involvement? Payment issues? - Finding critical amount of data

7. Requirements

The most interesting use cases for DataLift are the ones which show the added value of having interconnected datasets. The need for interlinking data is almost never spontaneously expressed by the actors. During our interviews, only a few of them did express it, namely CG 33, Etalab and Montpellier. However, what is expressed, is the desire for interoperability and ease of cross-linking. Interoperability allows developers to not duplicate efforts, thus for example not duplicating a connector each time one is changing his geographical region. In Bordeaux, for example, Sebastian Dugue, author of iRennes, iBordeaux confessed in 2011: "*In Rennes data transport are in GTFS format, and I started with that. In Bordeaux, when I saw that it was in the Trident format, I did not even look at this further*".

We list below some key functionalities expected from the users and actors/providers of Open Data in France. It is clear that there exits some business needs that can be summarized as:

- Access by territory: case of DILA
- Access by time period: History, genealogy, anecdotic history (it happens someday in.....), temporal search, temporal inquiry, games, etc.

7.1. Capturing datasets from multiple internal and external sources.

This aims to facilitate the capture of heterogeneous datasets. This process can help centralized datasets that will thus be readily available to re-users: hence the data is easy to search, to query if there is an effort of standardization of the metadata by the publishers. Some protocols are more or less dedicated to this: OAI-PMH (used by GC33), RSS, GEoRSS. Several platforms are now using open data.

7.2. Managing updates of the data.

Managing data updates is crucial. Data from INSEE and IGN, for example, are continuously updated and should benefit from having a model of time integrated, in order to monitor data changes (the inverse of RSS).

7.3. Managing feedback for data quality.

The feedback for data quality is also a big issue, shared by both data producers and consumers. The first ones, for example, are often arguing that they cannot publish a dataset because of its quality. Consumers answer that data quality is not the real issue. Data quality depends on the usage of the data: some usages may deal with relatively poor precision of the data to be relevant. The real issue, instead, is the possibility to give feedback. Consumers can understand that data quality depends on the producer activities. They can also understand that everyone can make mistake. But they want to give feedback to help for data quality. This need is very important for them. When a consumer sees an error in a dataset he is facing two options:

- fix the data in its own database, thus creating an exception in the normal workflow; he will have to manage this exception when the data will be updated by the producers; this option is often costly.
- tell the producer for this issue and wait for the fix.

Of course the second option is less costly and improves data quality for everyone. But the feedback should be well managed to be efficient. It is not possible to report by email every problem to give feedback . The feedback process should be held by the same tools used for data exchange. The SPARQL query language seems to be well suited for that, bringing the possibility of two ways communications, although sometimes the fix should happen in the source data. This case could happen when the RDF-ized process start from a database, thus one wants the fix to happen in the original database, but not in its RDF version.

Datalift should act as a go-between to help data feedback. The consumer could suggest data fix to Datalift by SPARQL queries, thus, making data fixed for the consumer. And Datalift could inform the producer of a data fix request.

7.4. Convert datasets in "ready-to-use" formats.

In general, RDF is never asked spontaneously by the actors. The issue is to not provide dozens of different data formats. Datalift should be able to operate from very basic structured formats such as CSV and then lift and re-expose this data.

7.5. Explicit reconciliation of data from different sources.

This avoids repeating data that are present in different databases. Data duplication is a huge issue nowadays. Firstly, duplicating data is a very common way to avoid the technical connection between heterogeneous databases: bringing common standards for all, the semantic web is a good solution to solve this problem. Secondly, duplicating data is necessary when datasets producers are not able to deal easily with other producers to share data. Thirdly, data duplication is mostly not known by data collectors. Collecting the same data for different databases and usage is very common. The semantic web, and especially linked data technologies, is also a good solution to discover already published datasets.

7.6. Ease the interconnection of data to users.

It implies exposing data on the Web with their license to be consumed and linked to other data already openly available. The Web of Data is rarely mentioned as a solution, and most users are going for proprietary APIs or standard API of their profession or domain of interest. Thus, there is a strong need for a single protocol (not repeating a connector for each dataset nor different API), and a single web service.

This is a key issue and re-using the standard HTTP protocol together with REST principles are probably one of the most valuable solution. Dozen of data re-users are complaining that they are not able to make money with their application. A developer creating a small application will not be able to sell it more than 1 or 2 euros. But its market is generally restricted to one city or dataset. He is not capable to adapt his application to as many APIs as there are databases. Using a unique protocol (SPARQL)

with a unique data format is a real opportunity to make possible a real market for open data applications.

7.7. Control the access to data.

It concerns here privacy issue: formal identification, identification by interconnection. It is also about license and users authentication.

Who does access my data?
It's a common question from public open data producers. They are ok to open their data but they request to know who uses it. It's not a question of control, but a way to improve "customer" relationships.
Control data access for privacy reasons
In France, the privacy laws are strongly developed. Any organization using personal data has to declare it. Everyone must be informed about privacy related data and especially about their usage. Interconnecting privacy related data is a big issue:
- is it compliant with the law?
- how to inform the people?
Even if a person is not clearly identified, data interconnection can give the power to do it.

7.8. Find / Search data

The aim here is to search, link, navigate, see and understand the proximities. A way to do so is to use tags, search form, thesaurii, etc. In general, this issue is badly addressed by data portal, making data mashups more difficult to create and imagine.

7.9. Data validation prior to publication

Data publishers willing to open their data want to make sure the data they publish is of the appropriate quality, not for technical but rather for communication or reputation reason. They don't want to be known as "the service which publishes garbage".
Data publishing usually implies a series of transformation steps, so even if the quality of the original data has been assessed, the quality of the final lifted data must also be assessed.

7.10.   Dataset on-the-fly creation

The aim is to create a tailored dataset specific to an application, containing data from previously separated datasets. Using the links in the datasets, one must be able to extract precisely the data needed for an application, and use that for an application.

8. Conclusion and Outlook

In this survey, we have provided user needs for some challenging and innovative applications to be developed with the data gathered and linked through the DataLift platform. We have also described some Use Cases that make use and enrich the data of the main data providers (INSEE, IGN, DILA, etc.) of the DataLift consortium, showcasing the use of linked data technologies.

We observed that many successful applications that have been developed visualize structured data around the geographical, time and concepts dimensions. We plan to develop a small vocabulary that could be used to describe OpenData publications.

References

1. **Europe, Public Data.** Publicdata.eu. [Online] [Cited: June 08, 2012.] http://publicdata.eu.

2. **OpendaFrance.** opendatafrance.net. [Online] [Cited: June 8, 2012.] http://opendatafrance.net/.

3. **LiberTIC.** [Online] http://libertic.wordpress.com.

4. **Publica, Data.** data-publica.com. [Online] http://data-publica.com.

5. **FING.** fing.org. [Online] http://www.fing.org/.

6. **Citoyens, Regards.** regardscitoyens.org. [Online] http://regardscitoyens.org.

7. **OWNI.** owni.fr. [Online] http://owni.fr.

8. **Open Data Perennes**. Open Data Perennes. [Online] [Cited: June 06, 2012.] http://owni.fr/2012/02/21/open-data-perennes/.

9. **Rennes, Data.** Data Rennes Metropole. [Online] http://data.rennes-metropole.fr.

10. **Data, Paris.** Open Data Paris. [Online] http://opendata.paris.fr.

11. **Data, 71.** Open Data 71. [Online] http://www.opendata71.fr.

12. **Loire-Atlantique, Data.** Open Data Loire-Atlantique. [Online] http://data.loire-atlantique.fr.

13. **41, Data.** Open Data 41. [Online] http://www.pilote41.fr.

14. **Data, CG33.** Open Data CG33. [Online] http://www.datalocale.fr.

15. **Bordeaux, Open Data.** data.lacub.fr. [Online] http://data.lacub.fr.

16. **Toulouse, Open Data.** Data Grand Toulouse. [Online] http://data.grandtoulouse.fr.

17. **Nantes, Open Data.** Open Data Nantes. [Online] http://data.nantes.fr.

18. **Montpellier, Open Data.** Open Data Montpellier. [Online] http://opendata.montpelliernumerique.fr.

19. **EtaLab.** etalab.gouv.fr. [Online] http://www.etalab.gouv.fr.

20. **France, Open Data.** data.gouv.fr. [Online] http://data.gouv.fr.

21. **France, Open Street Map.** openstreetmap.fr. [Online] http://openstreetmap.fr.

22. **Guillard, Hubert.** Open Data, un premier bilan Français. [Online] May 31, 2012. [Cited: June 08, 2012.] http://owni.fr/2012/05/31/open-data-france-premier-bilan-francais-etalab/.

23. **Sequeda, Juan.** Semanticweb.com. [Online] [Cited: June 6, 2012.] http://semanticweb.com/the-semantic-web-has-gone-mainstream-wanna-bet_b27329.

24. **Google.** Google Knowledge Graph blog. [Online] [Cited: June 6, 2012.] http://googleblog.blogspot.fr/2012/05/introducing-knowledge-graph-things-not.html.

25. **Fracebook.** Open Graph Protocol. [Online] http://ogp.me.

26. **Berners-Lee, Tim.** Linked Data. [Online] http://www.W3.org/DesignIssues/LinkedData.html.

27. **UKCrimeStats.com.** [Online] http://uskcrimestats.com.

28. **Ltd, Elbatrop.** ukpharmacy. [Online] [Cited: June 6, 2012.] http://www.ekbatrop.com/ukpharmacy.

29. **Numberhood.** Numberhood.net. [Online] http://www.numberhood.net.

30. **London, BUSit.** busitlondon.co.uk. [Online] http://busitlondon.co.uk/.

31. **Finder, UK School.** UK School Finder. [Online] http://ukschoolfinder.com/.

32. **Christopher Osbone, Lokku Labs.** where-can-i-live. [Online] http://where-can-i-live.com/.

33. **Limited, SwirrlIT.** opendatacommunities.org. [Online] http://opendatacommunities.org/.

34. **FlyOnTime.us.** FlyOnTime. [Online] http://flyontime.us.

35. **Constellation, Tetherless World.** data-gov.tw.rpi.edu. [Online] http://data-gov.tw.rpi.edu/demo/stable/white-house-visitor/search.php.

36. **Aid Visualization**. Demo Aid Visualization. [Online] http://data-gov.tw.rpi.edu/demo/linked/aidviz-1554-10030.html.

37. **SAS, AMARPI.** Fourmisante. [Online] http://www.fourmisante.com/.

38. **MaVilleDuCiel.** mavillevueduciel v1.0. [Online] [Cited: June 06, 2012.] http://www.mavillevueduciel.com.

39. **homengo.com.** Home'n'Go. [Online] http://www.homengo.com.

40. **GLD, W3C WG.** gld. [Online] http://www.w3.org/2011/gld/wiki/Main_Page.

41. **Hyland, Bernadette.** [Online] June 05, 2012. [Cited: June 07, 2012.] http://www.slideshare.net/3roundstones/progress-update-on-government-linked-data-world.

42. **Agency, UK Environment.** [Online] http://environment.data.gov.uk/.

43. **Southampton, University of.** University of Southampton Data Service. [Online] http://data.southampton.ac.uk/datasets.html.

44. **Congress, US Library of.** Authorities and Vocabularies. [Online] http://id.loc.gov/.

45. **Reegle.info.** Clean Energy Datasets. [Online] http://data.reegle.info/.

46. **OEG-UPM.** Spanish GeoLinkedData. [Online] http://geo.linkeddata.es/.

47. **OntologyCentral.** SEC Edgar Linked Data Wrapper. [Online] http://edgarwrap.ontologycentral.com/.

48. **UKCrimeStats.com.** *http://www.ukcrimestats.com.* UK : s.n., 2011.