Enabling Vertical Search on Railways data

- A 90di Travel Search White Paper
- February 2008 (Updated in October 2009)

For any comments or queries please contact: feedback@90di.com

Overview

This document is a white paper on how Railway organizations like IRCTC and CRIS; the Indian people; and the private travel application developers can all benefit, if Railways opens up some of the basic data such as Train time-table and fare and seat availability as a web service or an API. This has multi-pronged benefits for Railway bodies as well as people who consume the web service. The various parties will stand to benefit from it in the following ways:

Benefits to the Railway organizations:

- An alternate channel of revenue when commercial parties exploit the API to build creative Web 2.0 applications
- Being able to act as an information broker or Infomediary when it comes to the Railways data.
- Having the latest technology in its armory, in terms of having an important technology for Web 2.0 i.e. a Web Service (or an API) to complement its on-line web sites (www.indianrail.gov.in, www.trainenquiry.com)
- To act as an enabler, which fuels innovation in India by allowing information-rich travel applications to be developed

For the travelers in India it will lead to the following benefits:

- Having access to the basic railways data through other channels such as an API, which can be used to apply certain context of their other needs, such as mixing it with Flight data and so on.
- Being able to use the advanced travel applications which address the heterogeneous nature of their travel needs

For private parties who develop travel applications (such as Travel Search portal), can benefit in the following ways:

- They utilize the API to build innovative applications on top of the basic train information data. Such as providing a multi-modal travel routing website for Flights, Trains and even Buses.
- Taking advantage of the other phenomenon happening on the Web 2.0 front such as the social networking movement on the web and applying it on the Trains (and travel in general)

The rest of the document is organized as such: We first explain the emergence of Vertical search for the Internet and why is it important. We go on to explain how a consumer shops on the Internet, whether it is for travel tickets or other things. How she or he would compare goods sold on the Internet and select the best deal as per their requirements. Then we explain as to how vertical search should be applied to travel and what are the benefits of each to all the parties including the Indian Railways.

What is vertical search?

To understand vertical search, we would first need to talk briefly about the search engine technology most commonly *in use* today. What we mean by *in use* is that, the way it is most commonly understood by an average user of the Internet. An average user of the Internet understands very well the concept of search, as its done by Google, Yahoo! and the Microsoft's Bing search engine and many more. All these search engines have been designed for the static content (or the web sites) on the Internet. A typical search engine like Google, caches all the millions of web sites (some billions of pages) on its servers. Hence it is able to get the search results very fast to the user.

This kind of search technology works very well for all the static content on the Internet. Which constitutes about 10% of data on the Internet! Yes, the remaining 90% of the data on the Internet is behind those static web pages – in the data bases of those web sites or in the other systems that those web sites front end. Figure 1 illustrates the information on the Internet as covered by horizontal and the vertical search. Typical example of such data, covered by vertical search is price of some item for sale e.g. a price of a Refrigerator or a price of a travel ticket. So how do we access that *deep* sitting data? The obvious answer is when a user uses the web site forms and pulls out that data, for example by specifying an item type etc. But what if the user wants to compare such deep information across hundreds or thousands of such web sites? The short answer to that is through *vertical search* engines. We'll understand more about it in the paragraphs below.

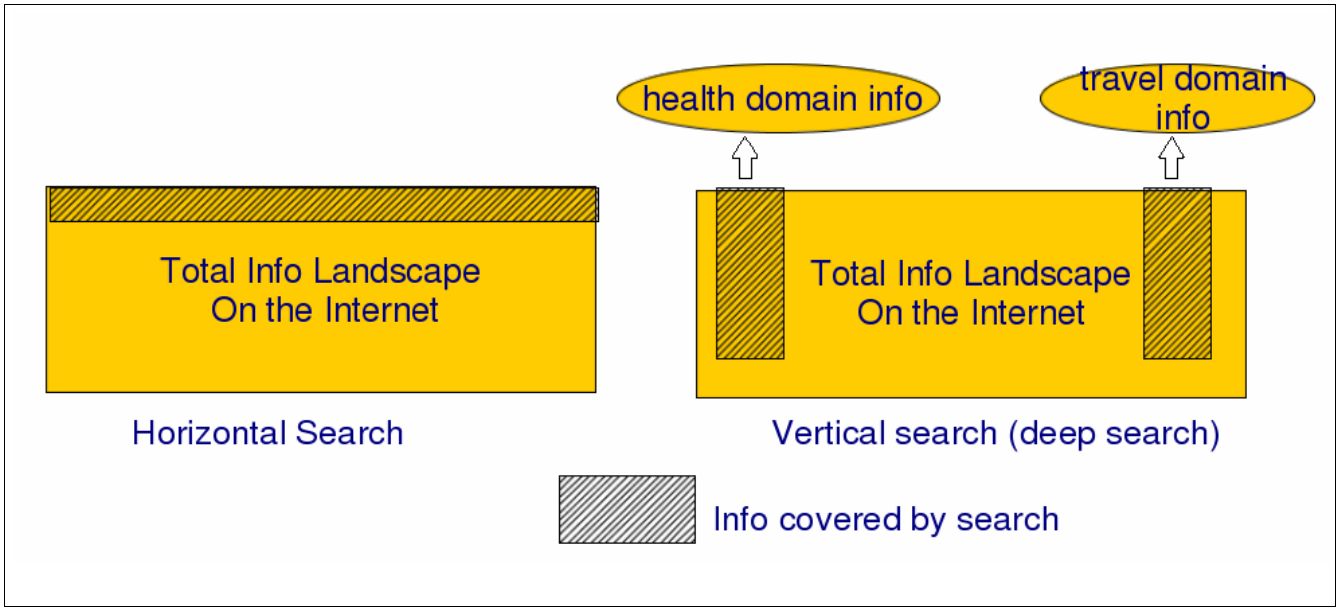

Figure 1: Picture depicting the difference between horizontal and vertical search

Simply put *vertical search* is the natural evolution of the search technology on the Internet. Which can provide answers to the above question. In this scenario, there will be *vertical* (and also *deep)* search engines, which will have a complete list of sellers for a particular domain like electronic goods, machine parts or travel. Apart form maintaining a catalog of seller sites, they will also understand the attributes of the particular domain (e.g. travel time and fare for *travel domain*). And when the user comes to the deep searching web site, it will search the inventory of all the seller web sites in real time and allow the user to choose an item based on her preferred attributes.

Vertical applies to a particular domain like retail, travel, health care, etc. So the term vertical search is often used to mean searching for a particular domain. A good example of a vertical search site is www.kosmix.com , which has vertical search for different categories like Autos, Travel and Health etc. Google blog search and Google book search are also examples of vertical search.

Another point to note is the overlap in terminologies like *vertical* and *deep*. In this document we will use them interchangeably. If it is able to pull out the hidden data which lies behind the web sites pages. A search engine can be a either a vertical search type or a deep search type or a combination of the two.

What is on-line comparison shopping?

The number of web sites in the world is already a few millions and growing very fast. Also almost every body who sells something wants to have a web front end too. In this scenario of e-commerce web site explosion, it is becoming increasingly difficult for a user to find the relevant information based on her needs. Below are some questions for which we will seek the answers:

- How does an Internet user know which site to use to pull out the information?
- Will the user get the best price and the best fit for her needs, just by searching the few sites she knows?
- How does the user know which sellers (sites) have what she needs?
- Also if she wants to compare the price (or other attributes of the product) from multiple vendors (web sites) how does she do?
- Won't she be overwhelmed with data after just accessing a handful of seller web sites?

In the real world before making a purchase, a consumer does a lot of *window shopping* to compare things before buying (Apart from just having a fun time in the mall :-]) . How does she do the same on the Internet? One can argue that just “surfing” the net may be an equivalent of the real world window shopping. I will agree to some extent on this, but the question is: Is the buyer getting the best deal by using just a few seller sites she knows of? How does she come to know about the best deal for her?

Comparison shopping (on the Internet) can be said to be possible by the existence of shopping comparison tools. On-line Comparison shopping tools, is one category of Vertical search application. Needless to say, for it to be effective these tools need to be *vertical* i.e. domain specific, apart from being *deep*. ugenie.com is one of the examples of some of the new players in this field. And it can be said that ebay is another example of a comparison shopping site, but it is is a variant with an emphasis on bids/auctions etc also it has a very broad base of items across categories so it can't be termed as *Vertical*.

Comparison shopping offers benefits to both Sellers and the Buyers:

Benefits to buyers

Buyers stand to gain a lot by the existence of Internet comparison shopping tools. The main benefits are:

- Obvious benefit: Buyers able to compare prices of similar products, from different sellers (web sites) and get the best deal
- Buyers able to search and select a product as per their preference (the comparison shopping tool does a parameterized deep search based on the buyers need)
- Encourages and results in healthy competition between sellers, which is always good for a consumer

Benefits to sellers

The benefits to sellers are:

- Automatic information spread - about the best deals offered by them - to the buyers. Since it happens through a parameterized comparison shopping tool
- Easy entry to new players (It can be argued, that it works against the existing ones). This is so because its easier to publish yourself to a few on line comparison tools (may be a few tens or hundreds), than to the numerous consumers (perhaps millions of them).

Examples of vertical search technology used in different verticals

Below listed are some of the real world examples of Vertical search technology in use today:

1) Octopart (http://octopart.com/) is search engine for Electronic parts. The way they describe them selves "Octopart is a search engine which lets you compare the prices and availability of electronic parts across different distributors" on their About Page at http://octopart.com/about
2) Local search engines, catering to a particular geography. There has been a explosion of local search engines, which work like horizontal search engines but target only the web sites having information of a particular region. Some examples are Google local, Yahoo! local and and asklaila.com. Justdial.com, guruji.com (for India).
3) Automobile search engine Edmunds.com, can be used to search for used Cars on sale, when buying a Car in the US.
4) Travel search engines, which cater to information about travel (we will look at this category of search engines in more detail later). Some examples are: kayak.com, mobissimo.com, which are primarily US focussed flight search engines. And 90di.com (yours truly) and ixigo.com, which catering to travel search in India.
5) Jobs search engines: SimplyHired.com, Indeed.com, Bixee.com (India),
6) Search engines looking for Health information: Amniota.com (Currently a Google partner!), CloserLookSearch.com

7) Another recent example is the move by the Government of California to make lots of government data public on its web site : http://www.datasf.org . Which is already being used by creative developers & companies to create innovative applications like providing local directions, to fight drug dealing near schools (http://blog.spatialkey.com/2009/09/visualizing-sfpds-operation-safe-schools/)

Components of the vertical search technology

The components required to realize a vertical search technology implementation are:

1) **Provision of APIs (Application Programming Interfaces)** by the web site owners. In the absence of the APIs a substitute is the 'Screen scraping' technique, in which the program interacts with the web site just like a human would, to pull out the information. APIs are getting popular and getting talked by more and more people but they are still not very common.

2) A **vertical search UI**. Since deep search is typically applied for a vertical i.e. a domain like travel,electronics, health care. We need a user interface specific for that domain. That interface is needed to enter the search parameters for that vertical. For example a travel search site will require from/to places, date, travel mode etc.

3) **Vertical search algorithms**: This aspect is hidden to the users and to the other stake holders. But is the USP of the particular deep (and vertical) search application. This is the main differentiator between vertical search engines. If we look at the present day search engines also, then Google's primary differentiator was the strength of its PageRank algorithm (of course the ability to scale to hundreds of thousands of machines was also a must have)

Vertical search and comparison shopping techniques applied to travel in India

We look at how vertical search technology can be applied to travel in India. To understand this lets begin by asking this question.

How does one travel from Azamgarh (U.P) to Bangalore (Karnataka) ? How does the traveler does her travel planning in this case?

In this case even if we assume that the traveler knows that she could travel by train to Delhi and then take a flight from Delhi to Bangalore. Still she has to go to multiple web sites to do the travel planning. She has to go to the Railways web site to do the train search and then go to the Flight sites to do the Flight search. This may be only be a bit of a problem if there is just one place, say Delhi, as a possible transit point.

On the other hand, If the traveler considers some other transit points also such as Lucknow, then the planning work becomes more tedious. Now she has to do more searches on different web sites. It gets

more complex if for single legs of her journey, she is open to travel by train and flight both and wants to find the best match based on her needs.

Wouldn't it be simple if she were able to ask this question at a single place (a travel search engine which connects with all the back end service providers (Railways and the Airlines) to get all the data?

In this case she can do a single search from Azamgarh to Bangalore and then select the best route based on the attributes important to her example: Total fare, Total duration, selecting a specific mode (Train/Flight) of journey for a particular leg.

It becomes very obvious that performing this search at a single place is much more beneficial to her. Figure 2 shows the screen shot of the search results from Azamgarh to Bangalore by a search done on www.90di.com , sorted on Fare, and hence it shows the first route as a pure Indian Railways Train route.

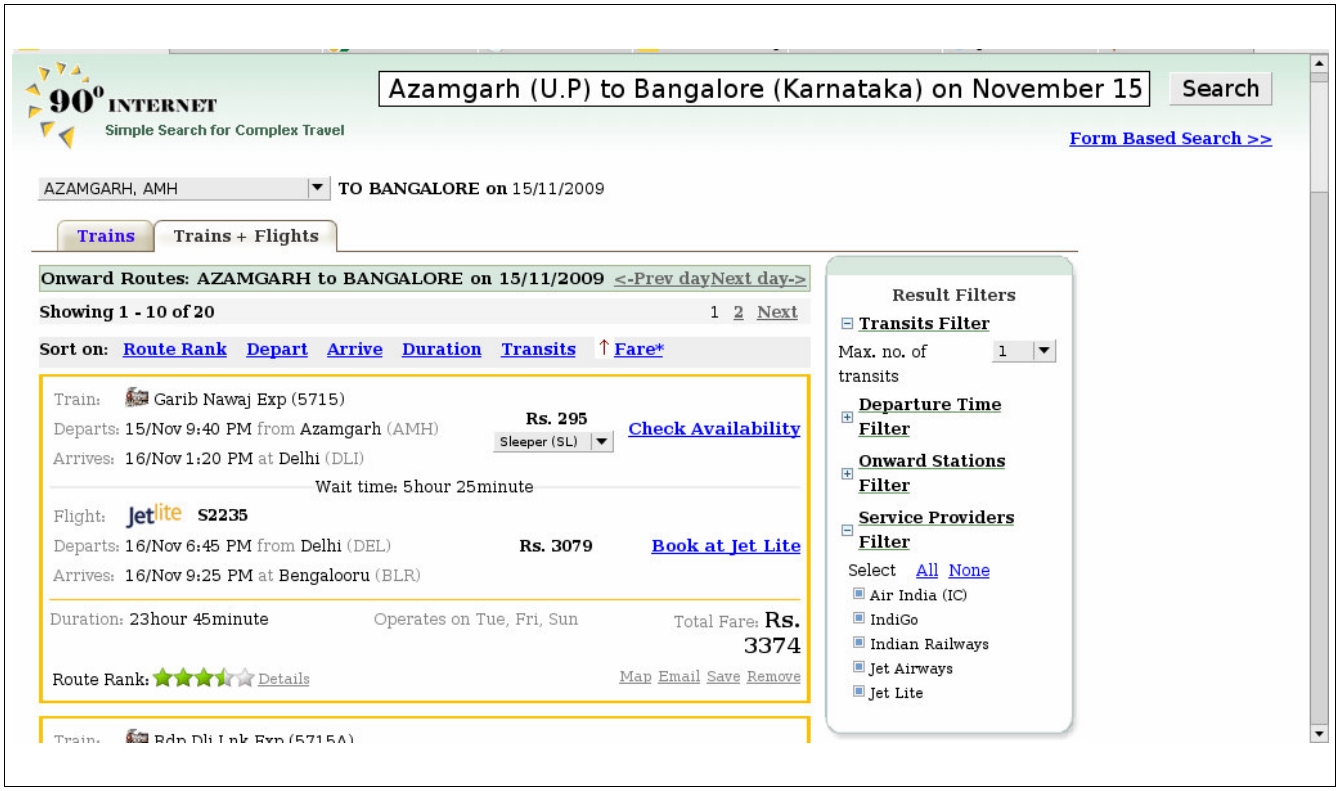

Figure 2: Azamgarh to Bangalore routes sorted on Fare

Figure 3 shows the route graphically on the India map. It is to be noted it shows the complete itinerary of the travel including the Train journey and the Air travel

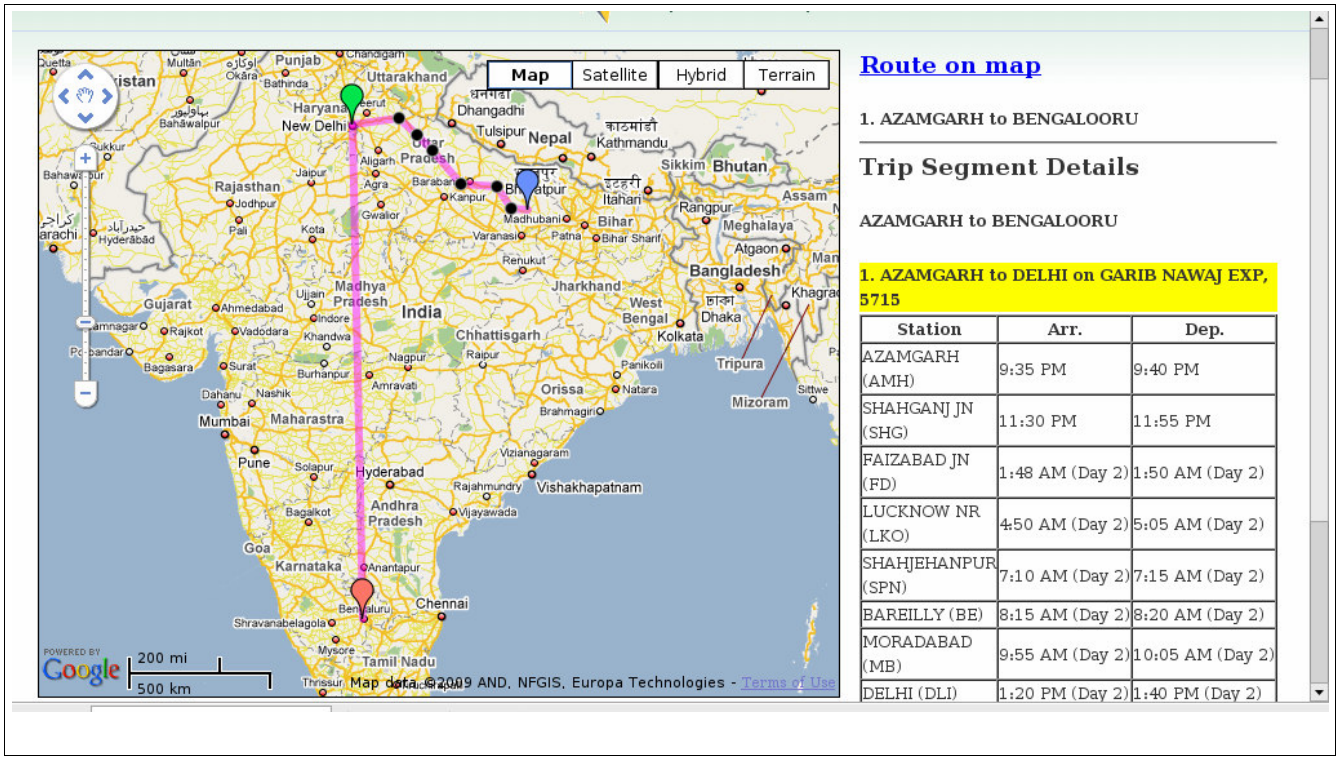

Figure 3: Azamgarh to Bangalore route shown graphically on the India map

So how does it become possible and how does it benefit all the stake holders including the service providers like CRIS/IRCTC (Indian Railways) and the Air lines. Lets explore this in the next section.

How does service providers like the Indian Railways (CRIS/IRCTC) can enable and at the same time benefit from vertical search

Enabling travel search through APIs

Lets look at the enabling thing first. To enable vertical search for travel, the Indian Railways (like all the other Travel service providers) has to expose some kind of APIs (or Web Services). Through which clients can programmatically get the schedule, fare and seat availability information.

Some examples of kinds of Web Services (APIs) which CRIS/IRCTC/Railways could expose are:

1) Getting of all Train numbers. The query would be to get all the trains which are currently in operation by the Indian Railways.

API Name: GetALLTrainNumbers

INPUT: NONE

OUTPUT: List of all the Train Numbers

2) Getting of a Train Schedule

API Name: GetTrainSchedule

INPUT: Train number (e.g. 5016)

OUTPUT: Schedule of the train as XML (or in some other format)

3) Get Availability for a Train between stations for a given travel date

API Name: GetStationTrainAvailability

INPUT: Train number (e.g. 5016) ,

From Station code (e.g. YPR)

To Station code (e.g. LJN)

Date (e.g. 19/Nov/2009)

Travel Class (e.g SL or 3AC etc.)

OUTPUT: Availability and fare for the journey as XML (or in some other format)

4) Batch requests to get the availability for multiple trains between two station pairs.

API Name: GetAvailabilityInBatch

INPUT: Batch data of several records of INPUT specified in API 3 (GetStationTrainAvailability) above.

OUTPUT: Batch data of several records of OUTPUT specified in API 3 (GetStationTrainAvailability) above.

NOTE: The above are a few examples and not an exhaustive list of the APIs

Benefiting through those APIs

We saw some of the benefits right at the top of this document. Lets understand them in more detail:

1) An alternate channel of revenue for the CRIS/IRCTC (Indian Railways)

The Railways could charge for the use of that API, for commercial use. The detailed model needs to be explored by CRIS/IRCTC, but at a high level – it may have the following usage model:

- Free for non-commercial use by individual users Or up to a certain number of APIs call per day
- Charging the usage of the API for commercial parties. It could be in any of the following ways:

- Share of the revenue made by the private party as a result of using the Railways API. For example it could be a share of the revenue made by the private web site by displaying Ads etc. on its web page (which had data from the Indian Railways)
- Pure usage based. For example flat rate per no. of API calls. The rate could vary for different APIs.

2) Fueling and encouraging innovative travel applications in India

The travel scenario in India is quite complex and different than the western countries like US. The primary difference being that it is much more heterogeneous in India with people using Trains, Flights and Buses, for long distance travel. Whereas in US it is quite homogeneous with Flights being the most preferred mode of long distance travel.

As seen earlier in this document, consumers would do comparison shopping across the different modes of travel to select the mode (or modes) of their choice. In this scenario CRIS/IRCTC and the Indian Railways being the sole provider of the Train mode of service, it becomes obligatory on them to expose their data in a way which benefits the consumers and also fuels innovative travel applications.

One such example of a creative application is the existing 90di web site (www.90di.com) , which adds the following important features to the existing CRIS/IRCTC web sites

- Multi-hop routing: How to go from any station A to another station B in India. It shows the connecting trains through all the possible junctions between A & B, if there is no direct train.
- Often people have to just go from one place to another and are okay to mix the mode of their travel (Air & Train route). 90di.com has built a multi-modal routing solution connecting the Air and Train routes.

3) Readiness to meet the next generation Web

The vertical Internet is happening. The common jargons of Web 2.0 (some people have even started to use Web 3.0! Whatever that means), API and AJAX are heard every now and then when people talk of Internet.

In simple terms it means that the amount if information, that exists on the Internet is growing at a very rapid pace and the search engine technology of today is not able to satisfy the consumer need for the information. *Vertical search* is the answer for these questions and the way of the future is through Semantic web, existence of APIs and such.

By providing Web services and APIs, CRIS and IRCTC shall be naturally moving in the direction of Web 2.0

Summary

Vertical search is the way in which the search is evolving on the Internet. Consumer shopping is one key category of application, which can be built on Vertical search engines. The way people plan

their travel and buy travel tickets is also going to evolve to be a type of a consumer shopping application. In this scenario, the heterogeneous and complex nature of travel has to be simplified when presenting to the user (traveler) .

In the Indian scenario Trains, Flights and Buses are very common modes of long distance transport. Users weigh each of these options when doing their travel plan. With the explosion of other sectors such as aviation and the expansion of already existing existing sectors like Railways, travel planning is increasingly becoming complex for the traveler.

Travel search engines are best suited to address this issue. They would need willing support from all the participants such as the Air Lines and the Indian Railways (CRIS/IRCTC). It is very clear, that it is the future of travel information access and of travel planning. Also all the parties: The service providers, the people (travelers) and the private players stand to gain from innovative travel search applications.

Bibliography

1) Book: The Search, By John Battelle
2) Search Engine watch & Search Engine Land: A good resource for evolution of search technology

   http://searchenginewatch.com/ ; http://www.searchengineland.com/
3) Programmable Web : A good resource tracking APIs/Web services on the Internet

   http://www.programmableweb.com/
4) Wikipedia entry on vertical search

   http://en.wikipedia.org/wiki/Vertical_search
5) San Francisco, the city that's open for data

   http://www.guardian.co.uk/technology/2009/oct/14/san-francisco-open-city-data