Linking Enterprise Data

François-Paul Servant

WWW 2008 Workshop: Linked Data on the Web - April 22, 2008 - Beijing



"Business model for the Semantic Web" [Berners-Lee]

- Integrating data sources of a corporation is expensive.
 - ~N² point to point connections between N systems
- Using SW technologies can cut down these costs.
 - N RDF views of the N systems

François-Paul Servant



However, adoption of SW technologies in companies remains slow.

- SW technologies are often simply overlooked,
 - Not advertised by solution providers
- or considered as promising
 - but not ready to be used right now
- Even a negative prejudice
 - yet another technological hype, whose promises have been heard many times before.
- "We already have XML and Web Services. Why do we need SW?"



Linking Open Data / Linking Enterprise Data

The "Linking Open Data" community initiative:

- a major contribution to the concretization of the Web of Data
 - Proof of feasibility
 - Best practices

Why not try the same strategy in an enterprise?

- to publish enterprise data as linked data
- to produce guidelines, sample code,
- to give examples of applications using that data.



Envisioning one company's Information Systems as a space of linked data

- Linked Data principles provide effective solutions for 2 questions regarded by Renault as priorities for its IS architecture:
 - data repositories
 - services
- Our work was intended to demonstrate that

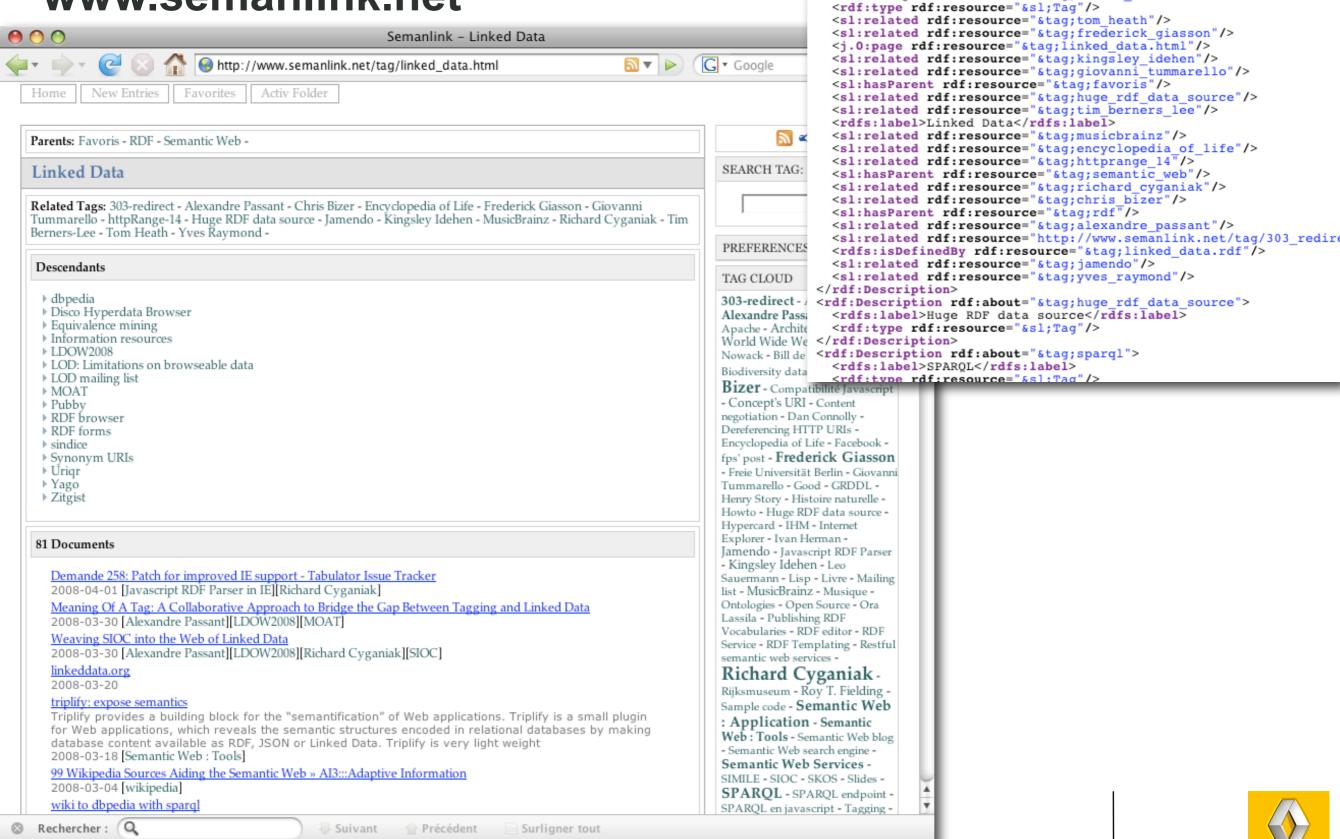


Summary of what we did

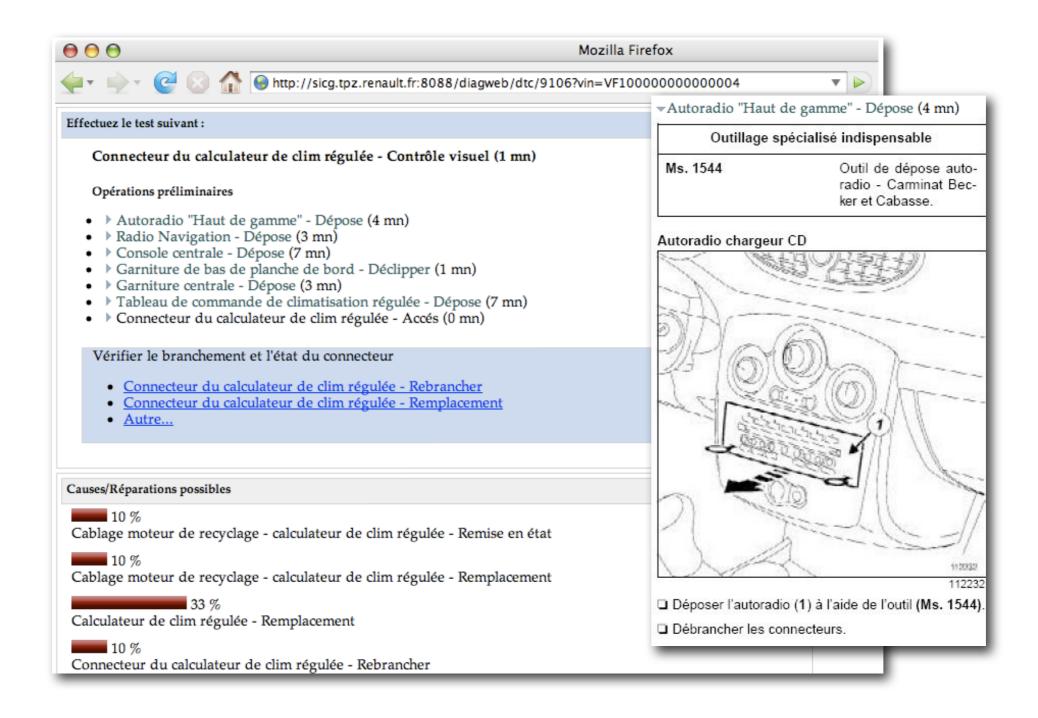
- Publishing of a data repository as Linked Data
 - REST service returning RDF to requests made by programs
- Implementation of a (very) simple RDF browser
 - Provides the HTML web application
- Access to the data from an outer application



Previous experiences with Linked Data: www.semanlink.net <rdf:Description rdf:about="&tag;linked_data">



Previous experiences with Linked Data: Prototype repository of repair and diagnostics operations





The field of the use case

Technical after-sales documentation

- A typical case for using SW technologies:
 - many objects need to be shared between many different systems,
 - ranging from engineering to repair shops,
 - in a rather complex process.



The repository

A dictionary of the terms that documentation writers may use

- Used to enforce an homogenous naming scheme throughout the whole documentation
- Used to index repair and diagnostics methods, and to define links between documents

Why this repository ?

- Not available as a service
- XML dump available
- Well managed
- Interesting uses

Content:

- for each term:
 - An identifying code
 - Labels (in 20 languages)
 - List of the corresponding "Generic Parts"
 - = link between repair documentation and spare parts catalogs
 - "Generic Parts": codification of the components of a car maintained by the Parts Department.
- A SKOS-like hierarchy of the terms
 - Routine maintenance : Filters : Oil filter



Development environment

- Java
 - Servlet
 - Jena
- Javascript



Implementing Linked Data principles

Minting URIs for the items of the repository

- namespace / item_class / item_code
 - "Slash URI"
 - can be dereferenced one at a time
 - namespace / item_class / item_code ? param=val &...

"Non-Information Resources"

- "HTTP-range 14" (303 redirect, with content negotiation)
 - namespace / item_class / item_code
 - namespace / item_class / item_code.rdf
 - namespace / item_class / item_code.html



Answering requests for the RDF about an NIR

- return all statements involving the NIR, either as subject or as object
 - NIR_URI, ? , ?
 - ?,?,NIR_URI
 - and a few more, such as
 - the links between nir, nir.rdf and nir.html
 - the labels of the returned resources that belong to the dataset

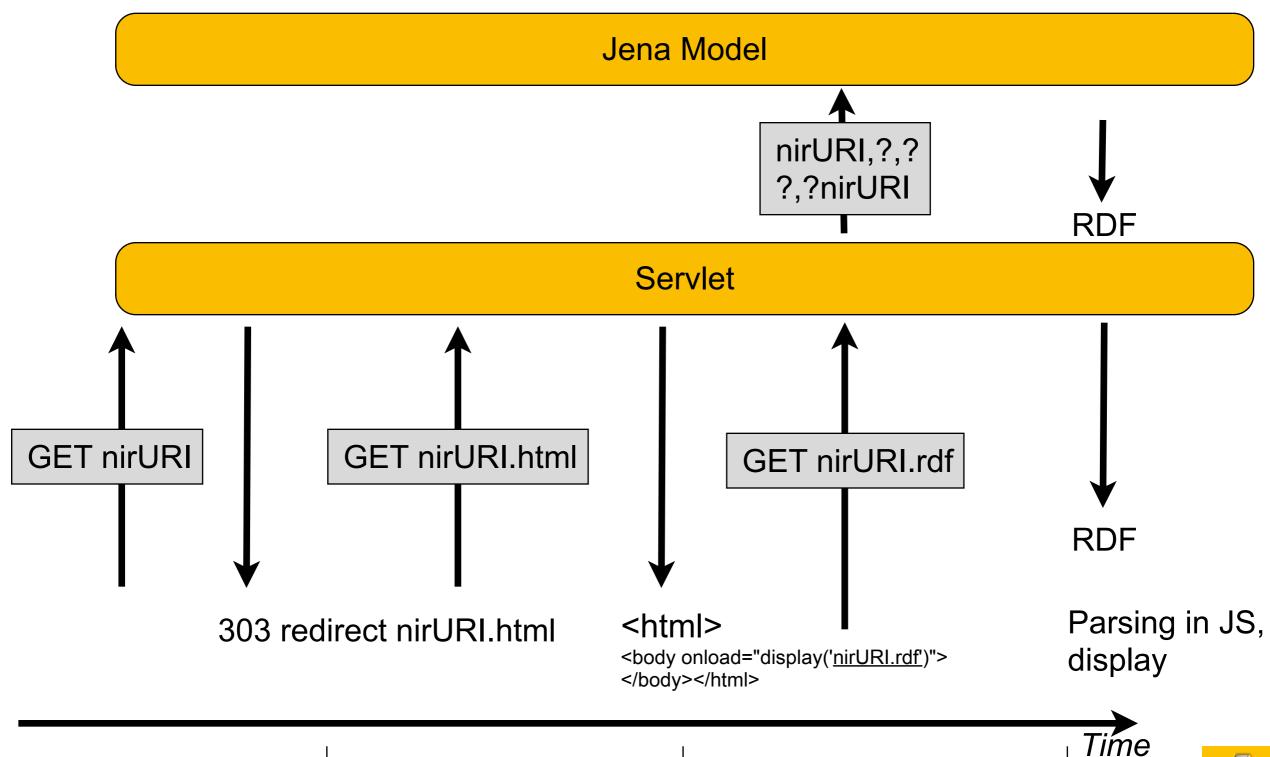


Answering requests for the HTML about an NIR

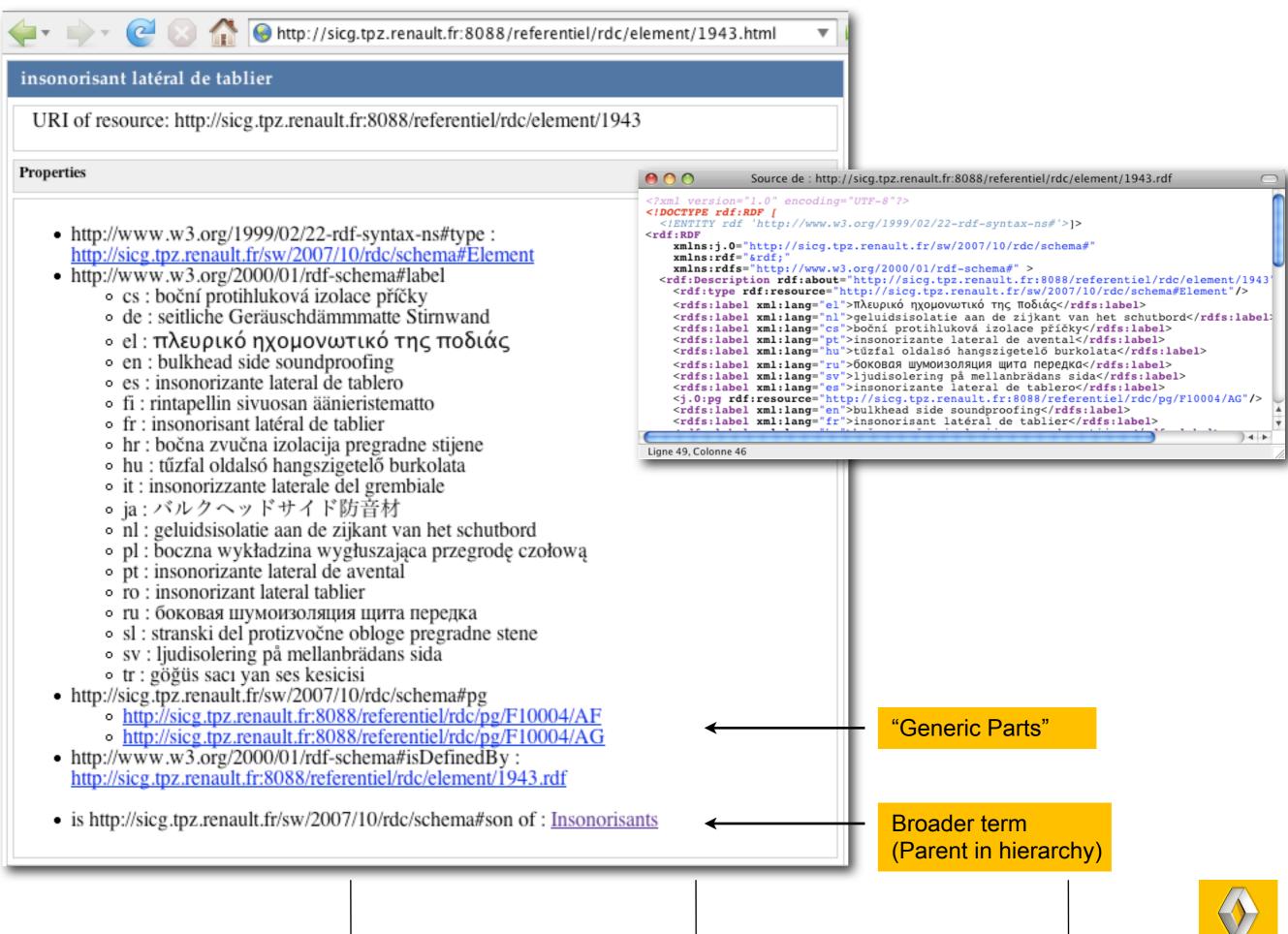
- Several options are possible for HTML generation
 - server side (JSP)
 - client side (Javascript)
- Client side generation of HTML
 - Tabulator's Javascript RDF parser
 - adapted to work with Internet Explorer
 - patch available at http://www.semanlink.net/files/2007/12/js-rdf-parser-ie-modifs2tabulator.zip
 - GET nirURI.html returns a quasi empty HTML page, containing a script that downloads nirURI.rdf and displays it
 - Advantages
 - clean separation of "view" and "model"
 - decreases load on the server
 - possibility to change the display without sending a new request to the server
 - incremental load of RDF data



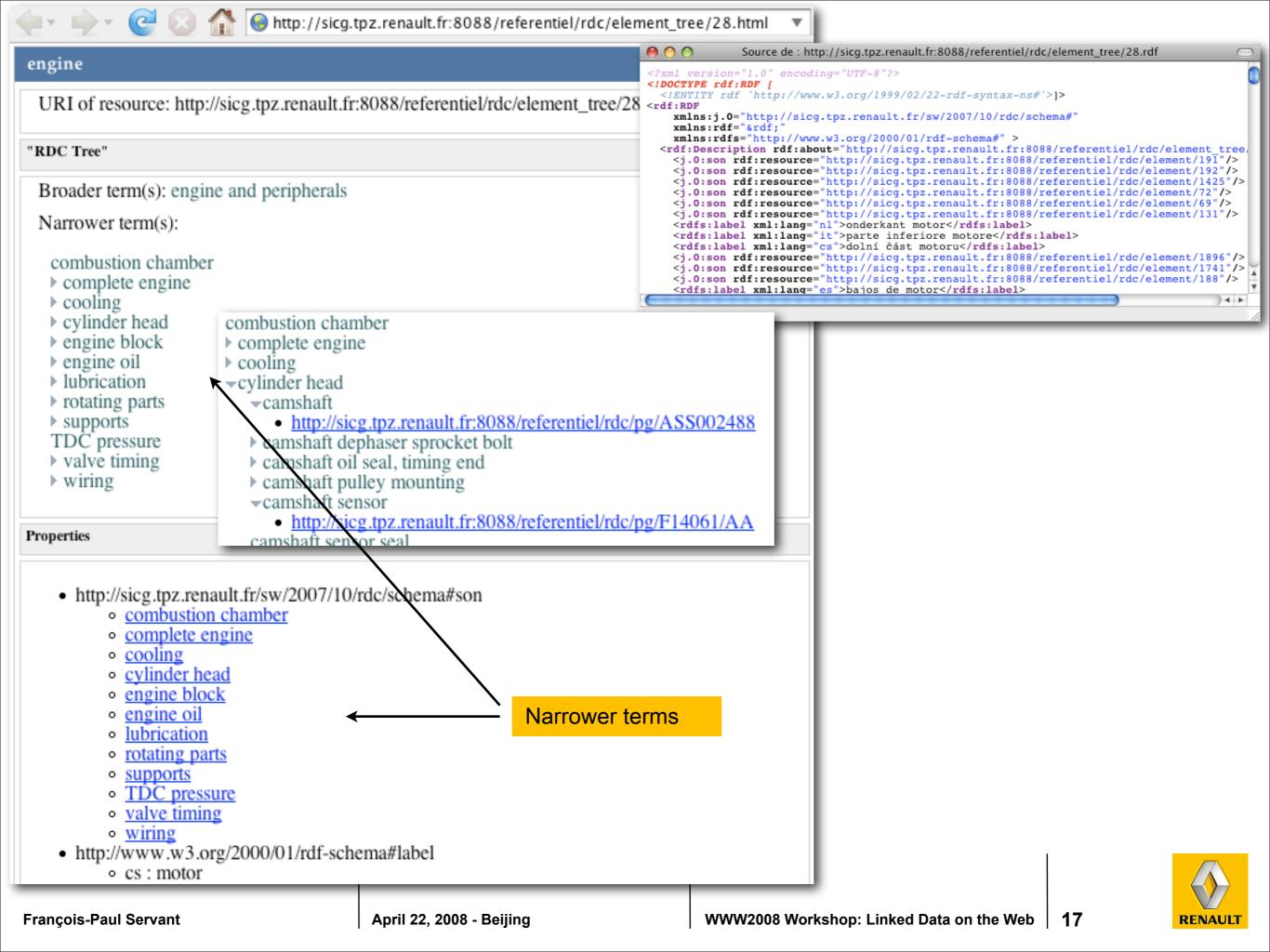
Generating HTML on the client with Javascript







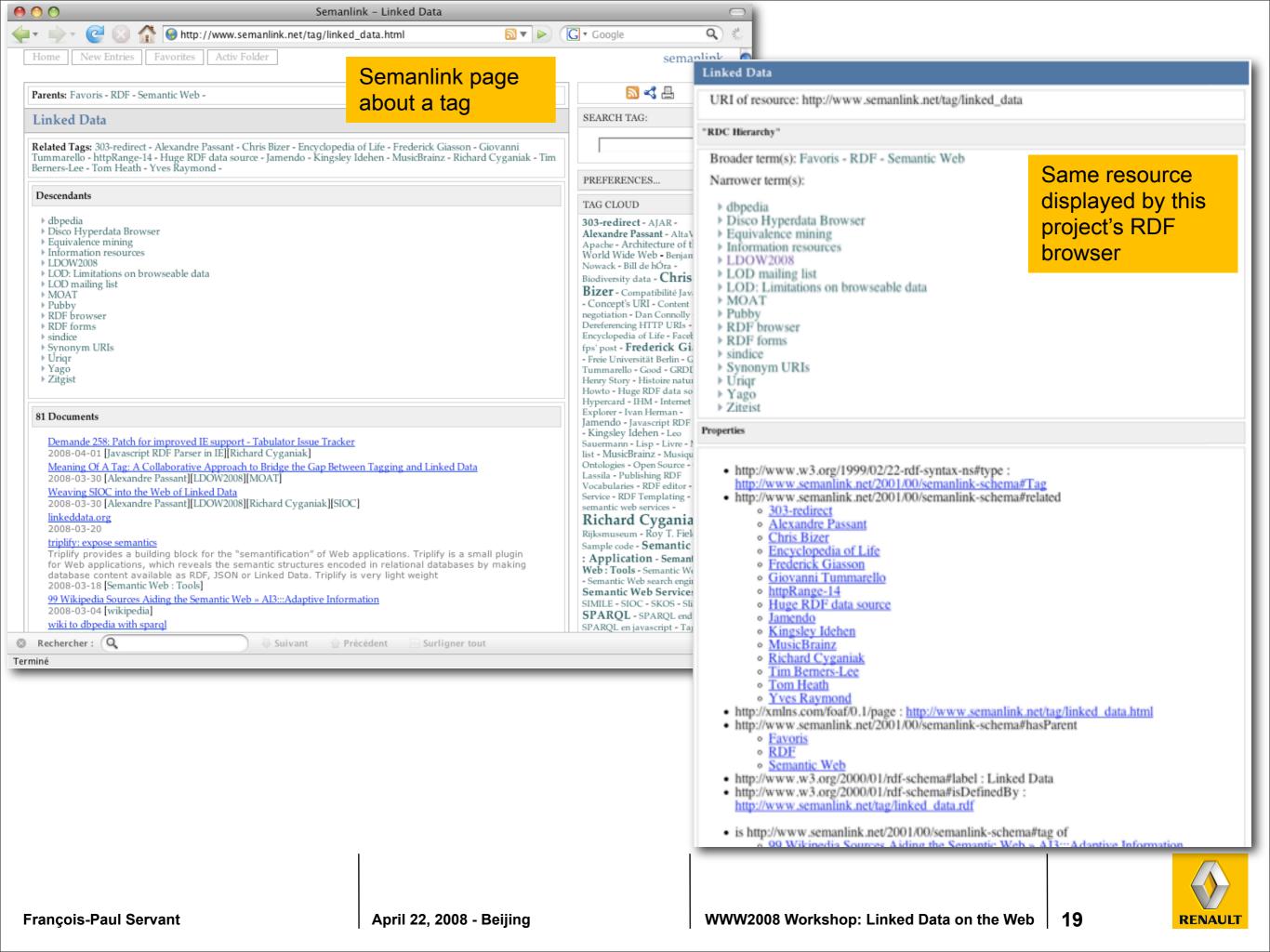




Turning the solution into a simple, yet generic RDF browser

- With what we described until now, we can only dereference NIR from our own dataset
 - because it is Javascript that gets the RDF
 - Standard Javascript security restriction
 - Changing browser's default settings (as Tabulator does) was not an option
- We implemented the usual trick of the "HTTP proxy"
 - Requests to dereference externals URIs are sent to the servlet,
 - the servlet forwards them to the actual server,
 - and then returns the result.

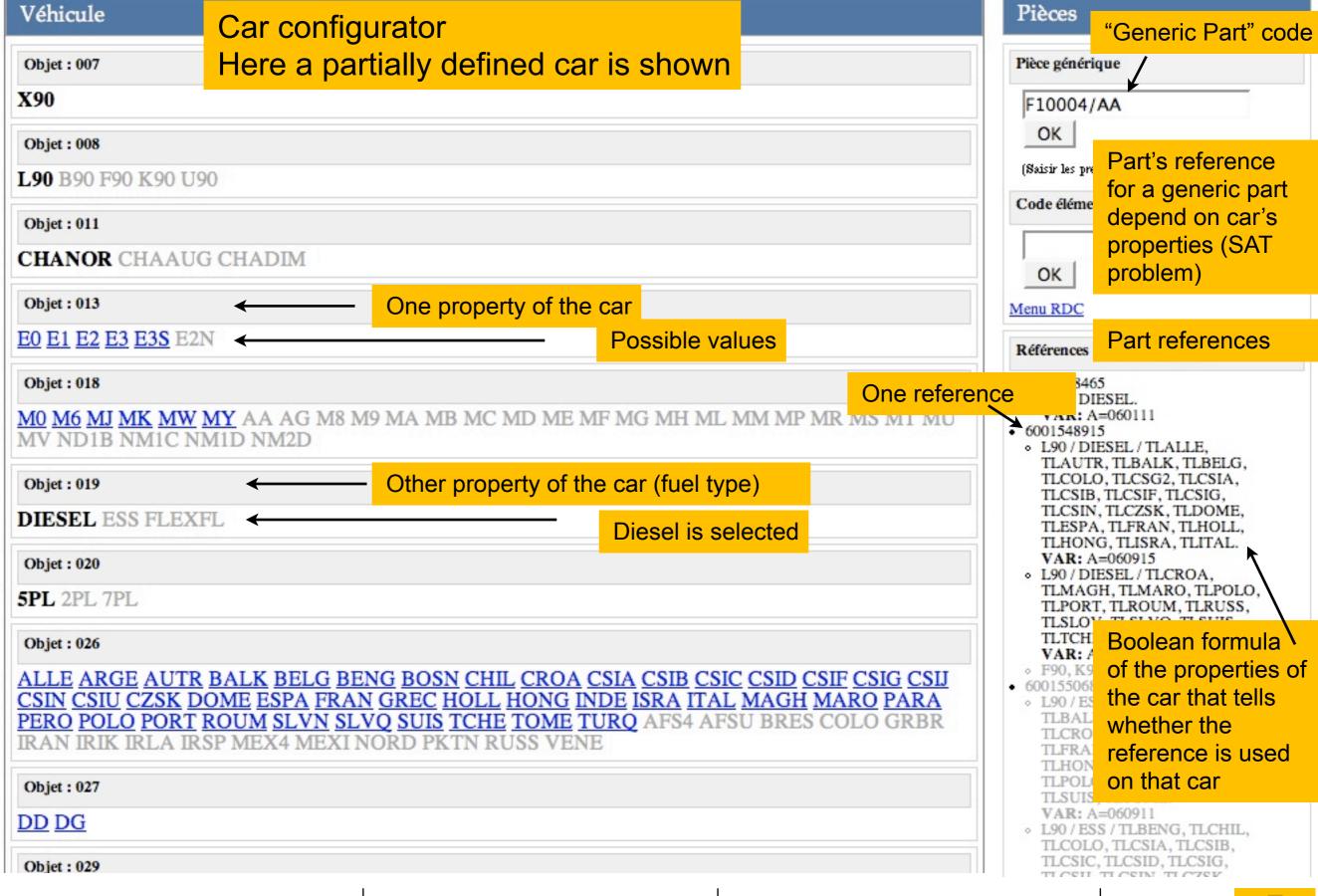




Access to repository data from outer application

- The repository, published as Linked Data, is a service
 - easy to get connected to (HTTP GET)
 - no dedicated code is needed to understand its RDF data
 - a RDF parser is enough
- Lets' use it!







Véhicule Objet: 007 accessories - breakdown - puncture Choose in the X90 bodywork hierarchy defined ▶ chassis in the repository of ▶ clutch - gearbox - transmission Objet: 008 electrical/electronic management repair terms L90 B90 F90 K90 U90 engine and peripherals general information Objet: 011 heating system - climate control instruments CHANOR CHAAUG CHADII opening element controls ▼passenger compartment - tailgate Objet: 013 fittings console E0 E1 E2 E3 E3S E2N dashboard interior trim Objet: 018 luggage compartment liner opening element trims MM MP MR MS MT MU MO M6 MJ MK MW MY AA A MV ND1B NM1C NM1D NM2 rear parcel shelf √soundproofing acoustic insert Objet: 019 bulkhead central soundproofing DIESEL ESS FLEXFL bulkhead side soundproofing bulkhead soundproofing Objet: 020 gearbox control unit soundproofing interior accessory 5PL 2PL 7PL lighting seats Objet: 026 routine vehicle maintenance ALLE ARGE AUTR BALK BI safety CSID CSIF CSIG CSIJ visibility - signalling - exterior lighting CSIN CSIU CZSK DOME ESP MAGH MARO PARA wiring - connectors PERO POLO PORT ROUM SI J BRES COLO GRBR Objet: 027 DD DG Objet: 029

Pièces

Pièce géné	rique	
OK		
(Saisir les p	remiers caractères suffit)	
Code élém	ent RDC	
1941		
OK	or enter term's	
Menu RDC	code	

1941 : INSONORISANT DE TABLIER

Références pour F16585/AC

Pas de référence correspondante

Références pour F10004/AD

· Pas de référence correspondante

Références pour F10004/AB

Pas de référence correspondante

Références pour F10004/AA

- 6001548465
 - L90 / DIESEL.
 - VAR: A=060111
- 6001548915
- L90 / DIESEL / TLALLE, TLAUTR, TLBALK, TLBELG, TLCOLO, TLCSG2, TLCSIA, TLCSIB, TLCSIF, TLCSIG, TLCSIN, TLCZSK, TLDOME, TLESPA, TLFRAN, TLHOLL, TLHONG, TLISRA, TLITAL.
 VAR: A=060915
- L90 / DIESEL / TLCROA, TLMAGH, TLMARO, TLPOLO, TLPORT, TLROUM, TLRUSS, TLSLOV, TLSLVQ, TLSUIS, TLTCHE, TLTOME, TLTURQ.
 VAR: A=060915

VAR: A=060915

- F90, K90 / DIESEL.
- 6001550681



Use of the repository for technical documentation

clutch - gearbox - transmission

- driveshaft
- engaging the clutch engine and gearbox assembly
- b final drive wheel shafts
- gearbox
 - ▶ complete gearbox
 - converter converter lock-up
 - differential
 - fork
 - gearbox gearbox lubrication
 - housings

1. étape de préparation à la dépose

Mettre le véhicule sur un pont élévateur à deux colonnes (voir Véhicule : Remorquage et levage) (MR 388, 02A, Moyen de levage).

IMPORTANT

Pour éviter une chute du véhicule, arrimer le véhicule au pont élévateur à l'aide d'une sangle.

la batterie (voir Batterie : Dépose - Repose) (MR 388, 80A, Batterie),

le bac à batterie,

l'écrou de la tresse de masse sur le support du calculateur d'injection,

le boîtier de filtre à air (voir Boîtier de filtre à air : Dépose - Repose) (MR 388, 12A, Mélange

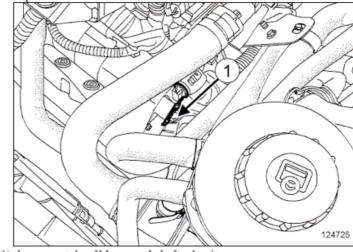
le calculateur d'injection (voir Calculateur d'injection d'essence : Dépose - Repose) (MR 388, 17B, Injection essence),

la fixation du câblage sur le support du calculateur d'injection.

Débrancher :

le connecteur de la sonde à oxygène,

le connecteur du capteur de point mort haut,



Déposer le vis (1) du support du câblage sur la boîte de vitesses.

gearbox

URI of resource: http://127.0.0.1:8080/referentiel/rdc/element/101 le connecteur du capteur de vitesse vehicule

Properties

- http://www.w3.org/2000/01/rdf-schema#label
 - cs : převodovka
 - de : Getriebe

Repair methods about gearbox

- Boîte de vitesses mécanique : Dépose Repose
 - X90/D4D/JH3
 - (v:model==v:X90) AND (v:engine==v:Engine_D4D) AND (v:gearbox==v:Gearbox JH3)
- Boîte de vitesses mécanique : Dépose Repose
 - X90/K4M/JR5JH3
 - (v:model==v:X90) AND (v:engine==v:Engine K4M) AND ((v:gearbox==v:Gearbox_JR5) OR (v:gearbox==v:Gearbox_JH3))
- Boîte de vitesses mécanique : Dépose Repose
 - X90/K9K 792/JR5 JH3
 - (v:model==v:X90) AND (v:engine6==v:Engine6_K9K_792) AND ((v:gearbox==v:Gearbox_JR5) OR (v:gearbox==v:Gearbox_JH3))
- Boîte de vitesses mécanique : Dépose Repose
 - X90/K9K_790/JR5,JH3
 - (v:model==v:X90) AND (v:engine6==v:Engine6_K9K_790) AND ((v:gearbox==v:Gearbox JR5) OR (v:gearbox==v:Gearbox JH3))
- Boîte de vitesses mécanique : Dépose Repose

Boolean formula of the characteristics of a car. Document applies for a given car iff this formula is true.



Veb

A (current) limitation of linked data principles

- Attached to each document, there is an "Applicability"
 - a Boolean formula of the characteristics of a car that defines the set of cars this
 document is relevant to.
- When searching information for a repair, you're only concerned with the documents relevant to the car being repaired.
- It doesn't make sense to return the documents about all the cars:
 - the client wouldn't be able to understand the Boolean formulae
 - Only the server can deal with them.

A form is needed

- Let's suppose that those formulae depend of 3 variables: v:model, v:engine and v:gearbox
- In a web 1.0 application, a form would be displayed to the user, asking to enter values for those variables.
- How can we handle that in the context of the web of data?
 - With Linked Data, there is no standardized equivalent: no way for the server to request some input from the client



RDF Form



Forms in the web of data

- "The Semantic Web [...] is about making links, so that a person or machine can explore the web of data." [Berners-Lee]
- Beside hypertext links, forms are an important feature of the web.
- They are needed also in the web of data.



Conclusion

- Several other such points need to be discussed among the LOD community
 - e.g., the question of the amount of data to return when fetching the URI of a NIR
- Anyway, Linked Data principles proved to be effective in a corporate context:
 - we published a repository as Linked Data
 - and that turned it into a service
 - easy to get connected to and easy to use
 - that fully respects the web architecture
 - » with immediate and free benefits (caching)
 - (this has to be compared with WS-* services)
 - the repository was simple, but the solution is largely reusable and expandable
- We will now
 - improve the prototype of technical documentation based on RDF metadata
 - and take all opportunities to continue working on Linked Data!



Thank you!



Dealing with applicability

Use the VIN ("Vehicle Identification Number" - kind of URI for cars)

- Constructors manage a database of cars, indexed by VIN, and that stores their technical characteristics
- We could define the URI of the "gearbox of a given car", adding the VIN as parameter:
 - http://.../element/1018?vin=VF123...
 - http://.../element/1018 gives the list of all documents about gearbox
 - The VIN database gives the characteristics of car VF123..
 - The server evaluates the applicability of each document
 - returns the filtered list
- But
 - this URI doesn't fit well within Linked Data
 - we need a way for the server to answer to a GET http://.../element/1018
 "information for a given car available by passing the VIN as parameter".
- Also, cache is of little use for URIs with a parameter such as the VIN
 - the server can know that the applicabilities of the documents about gearbox involve only (for instance), the model of the car, its engine type and its gearbox type
 - So, it would be nice if the server were able to ask the client who dereferenced the URI of the gearbox: "please enter model, engine type and gearbox to get the document you're concerned with."

