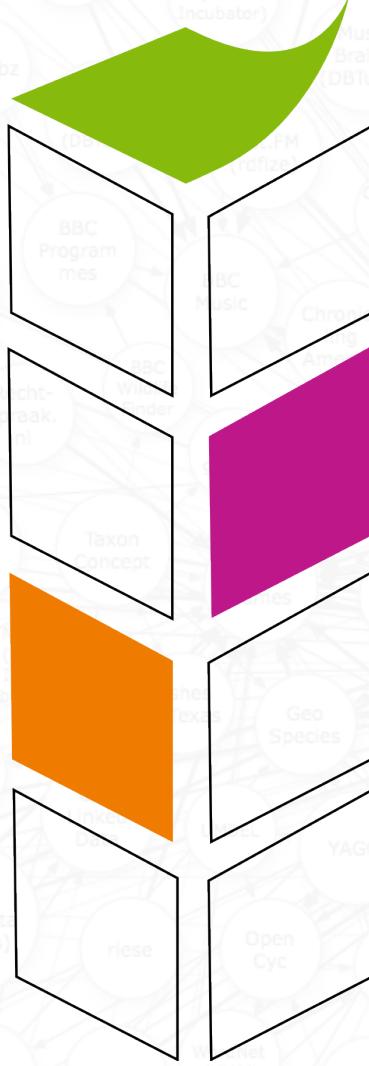


DATA **LIFT**



Conversion et visualisation de données statistiques

Thibaud Colas
Thibaut Marmin



Vendredi 8 mars 2013

Travail d'Étude et de Recherche - GMIN401



Plan général

- ▶ Contexte du projet
- ▶ Modules Datalift
- ▶ Conversion SDMX
- ▶ Visualisation DataCube



Contexte du projet



Contexte du projet

La plateforme Datalift

Inria

Atos



EURECOM
Sophia Antipolis

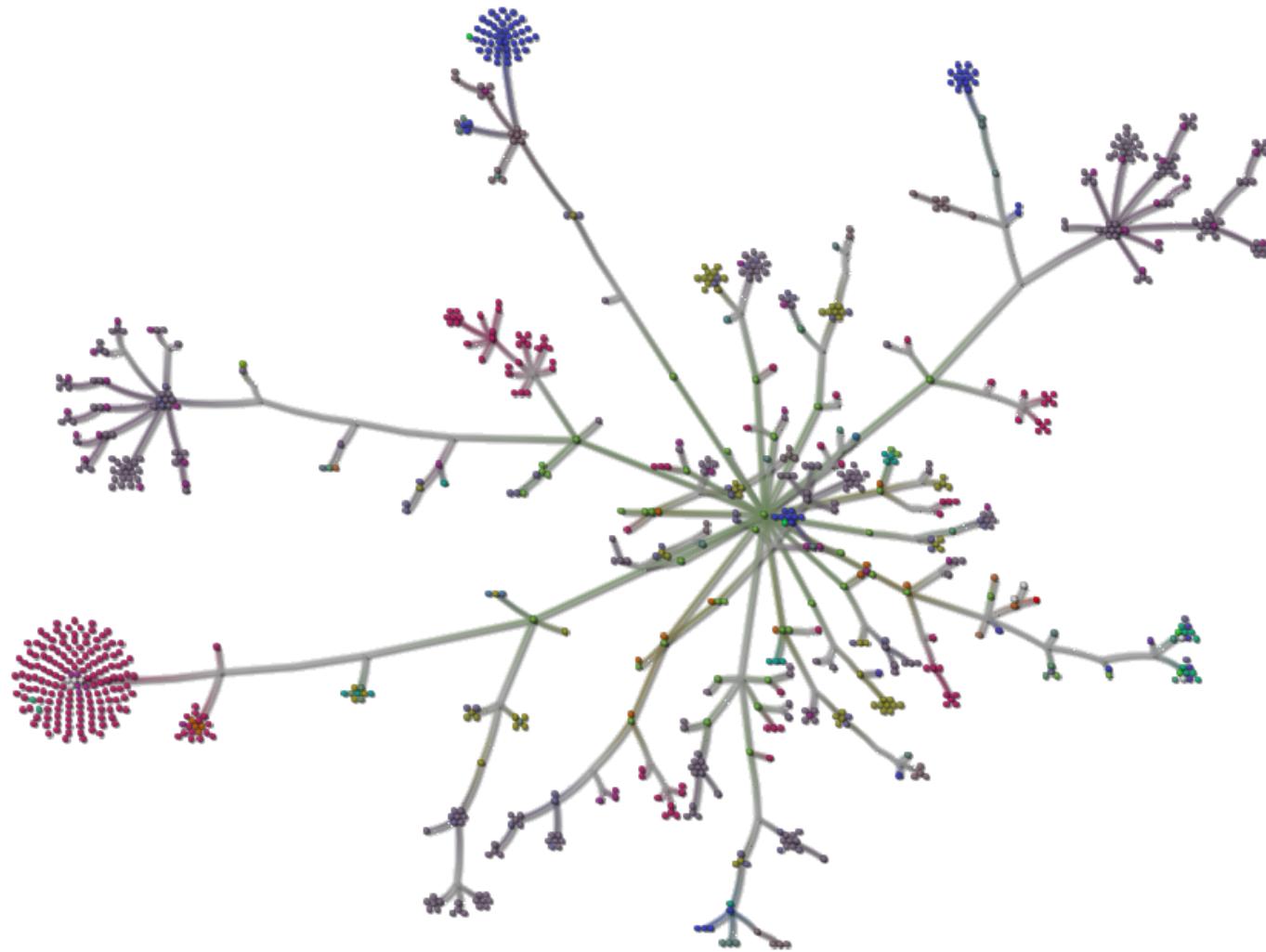


MONDECA



Contexte du projet

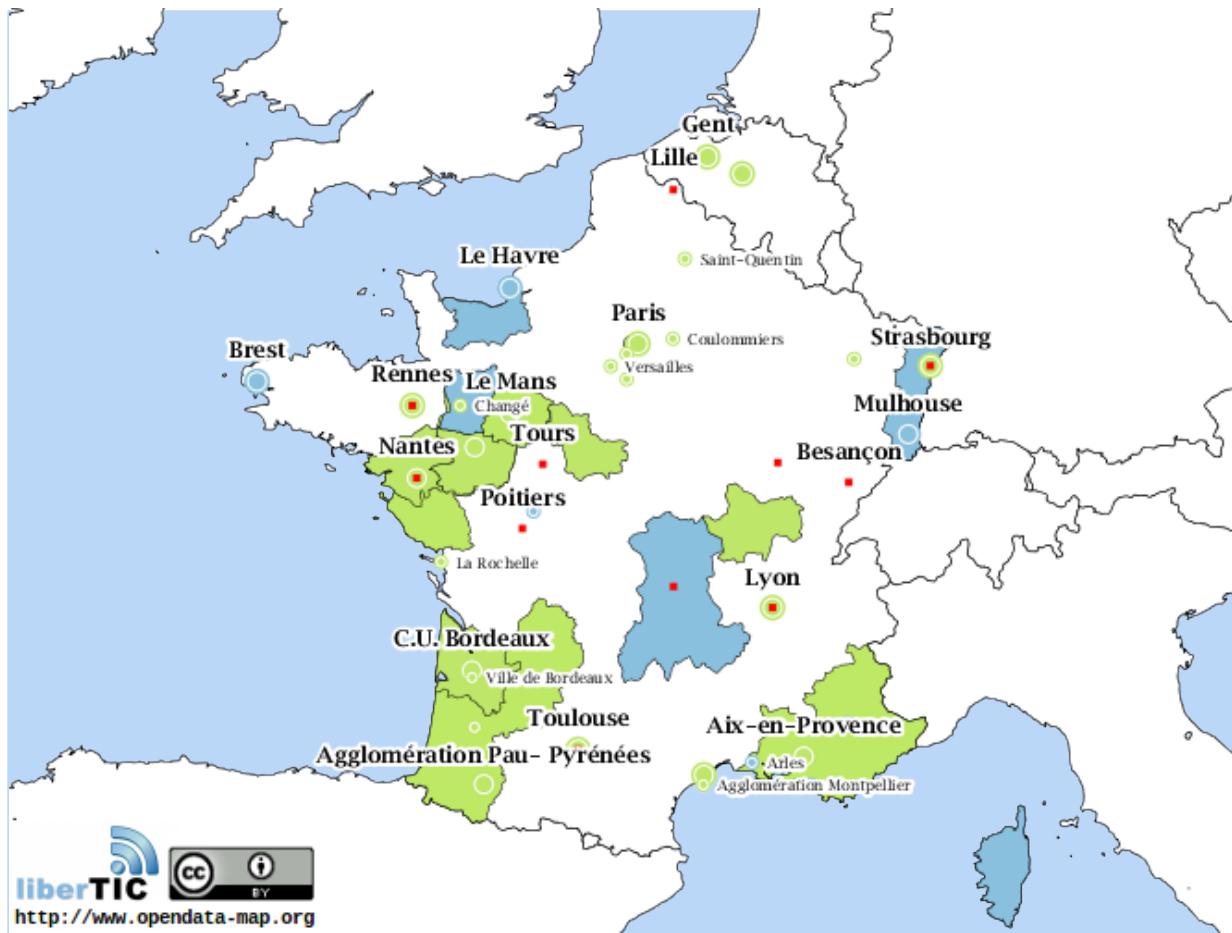
La plateforme Datalift





Contexte du projet

La plateforme Datalift





Contexte du projet

La plateforme Datalift





Contexte du projet

La plateforme Datalift

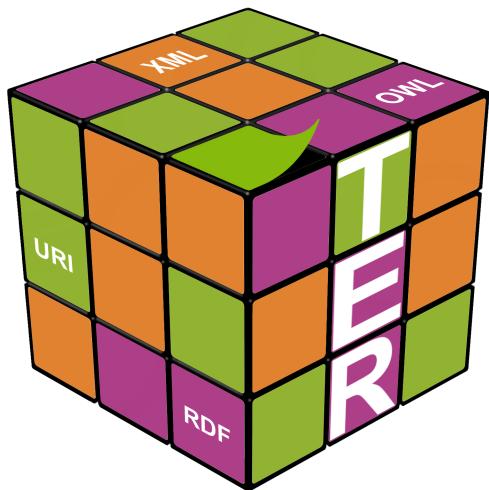
| # | Nom du livrable |
|--------|---|
| D2.1 | Méthodes et indicateurs pour la sélection d'ontologies fiables et réutilisables |
| D3.1 | Politique d'identification des ressources |
| D3.5 | State of the art on data provenance and data licensing |
| D4.1.1 | Methods for automated datasets interlinking |
| D4.2.1 | Dataset Interlinking Module |
| D5.1 | Datalift Architecture |
| D5.2 | Datalift Integration and modularity |
| D5.3.1 | Processus de Publication des Données |
| D5.4 | API Documentation |
| D6.1 | Usage scenarii for applications |

<http://datalift.org/fr/node/19>



Contexte du projet

La plateforme Datalift



D4





Contexte du projet

Nos objectifs et contraintes

```
interlink
├── build.xml
└── lib
    └── silk-2.5.3.jar
src
├── java
│   └── org
│       └── datalift
│           └── interlink
│               └── SilkController.java
└── web
    ├── interlink-form.vm
    ├── public
    │   └── js
    │       └── interlink-form.js
    ├── resources_en.properties
    └── resources_fr.properties
```



Contexte du projet

Nos objectifs et contraintes

- Implémenter la classe

org.datalift.fwk.project.Module

- Définir des routes HTTP

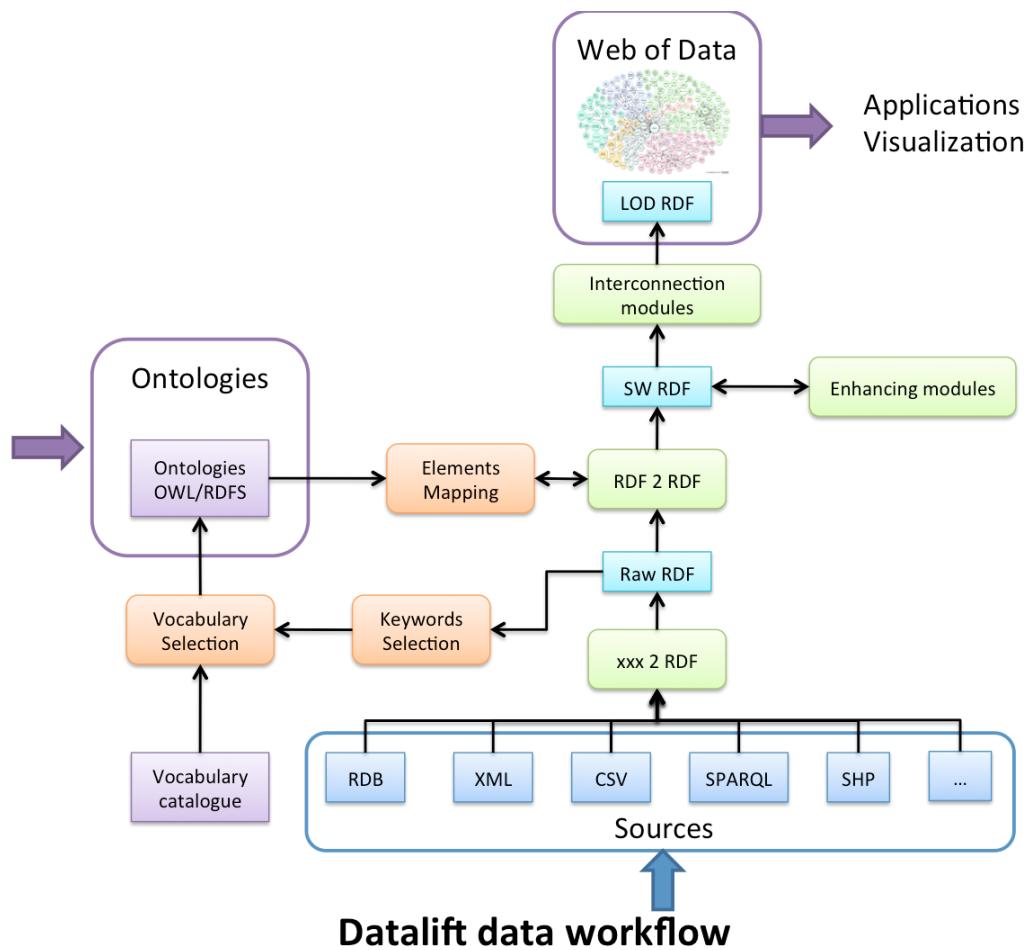
module/launch, module/validate, etc

- Définir un niveau d'élévation



Contexte du projet

Nos objectifs et contraintes





Contexte du projet

Nos objectifs et contraintes





Contexte du projet

Nos objectifs et contraintes

```
231      - // TODO : voir le code de OntologyMapper.java
232      + MessageTransporter transporter = new MessageTransporter();
233      -
234      -     return null;
235      +     transporter.global = projectId + " | " + inputSourceURI + " | "
236      +         + outputSourceURI + " | " + outputSourceName;
237      +
238      +     if (projectId != null) {
239      +         Project project = null;
240      +         try {
241      +             project = this.getProject(new java.net.URI(projectId));
242      +         } catch (Exception e) {
243      +             transporter.projectId = getTranslatedResource("error.projectId.unidentifiable")
```

1



ThibWeb repo collab

just now



Is this a bug ?

[Add a line note](#)

```
254      +             + " (" + projectId + ")";
255      +         }
```



Contexte du projet

Nos objectifs et contraintes



| | | | | | |
|---|--------------------------|----|---|--------------------------|---|
| ● | <input type="checkbox"/> | #7 | Dev Contrôleur | TODO |  |
| | | | by marminthibaut | a month ago | |
| ● | <input type="checkbox"/> | #6 | Dev UI | enhancement |  |
| | | | by marminthibaut | a month ago |  1 comment |
| ● | <input type="checkbox"/> | #5 | Méthode détection xml au format SDMX | question |  |
| | | | by marminthibaut | a month ago | |



Contexte du projet

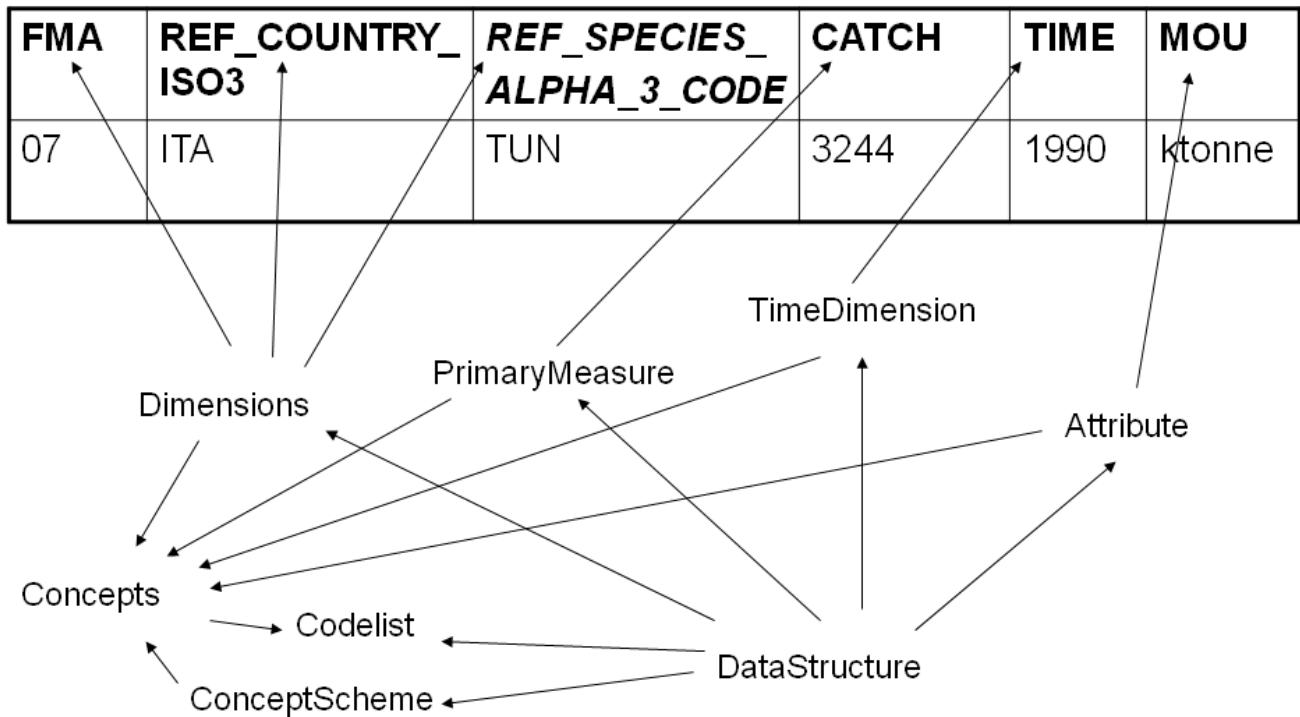
Nos objectifs et contraintes





Contexte du projet

Les données statistiques

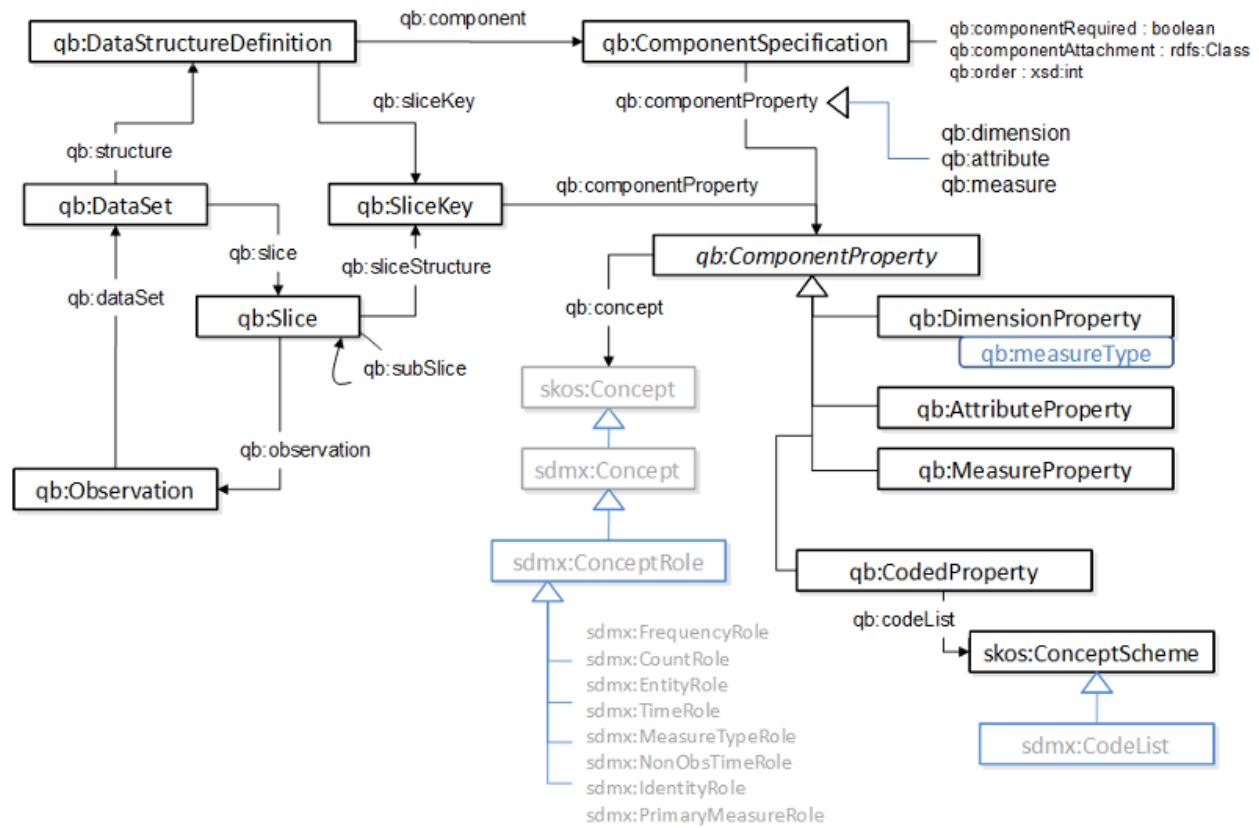




Contexte du projet

Les données statistiques

DataCube



<http://purl.org/linked-data/cube>



Contexte du projet

Les données statistiques



- URNs SDMX -ML

urn:sdmx:org.sdmx.infomo.codelist.Codelist=FR1:RP-GEO-REG-CL
(1.0)

- URIs RDF

<http://data.ign.fr/id/geofla/departement/51>

- Conversion URN -> URI

[http://www.ecb.int/rdf/{0}/{3}\({4}\)/{1}-{5}](http://www.ecb.int/rdf/{0}/{3}({4})/{1}-{5})



Modules Datalift



Modules Datalift

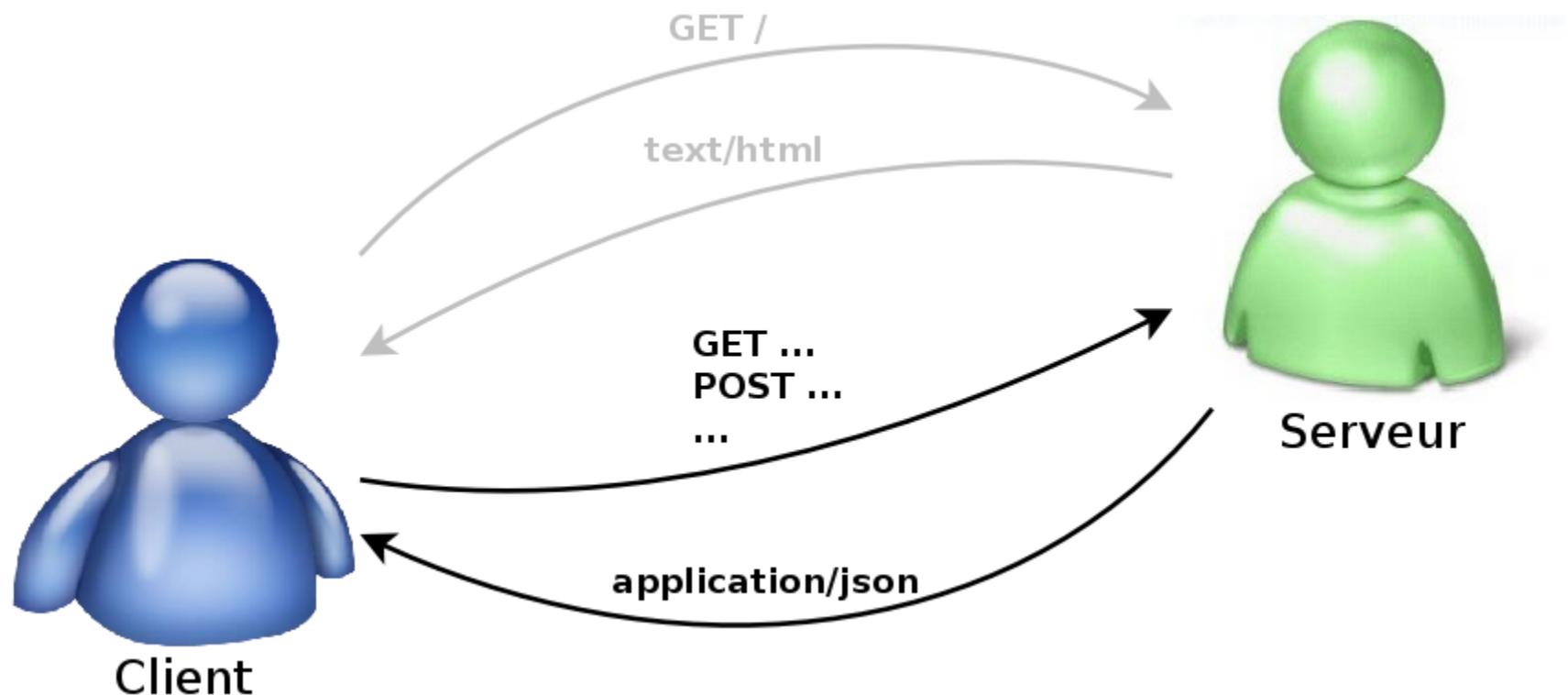
Architecture applicative - WebService





Modules Datalift

Architecture applicative - WebService





Modules Datalift

Architecture applicative - WebService

Méthodes HTTP

Status Code

Content type

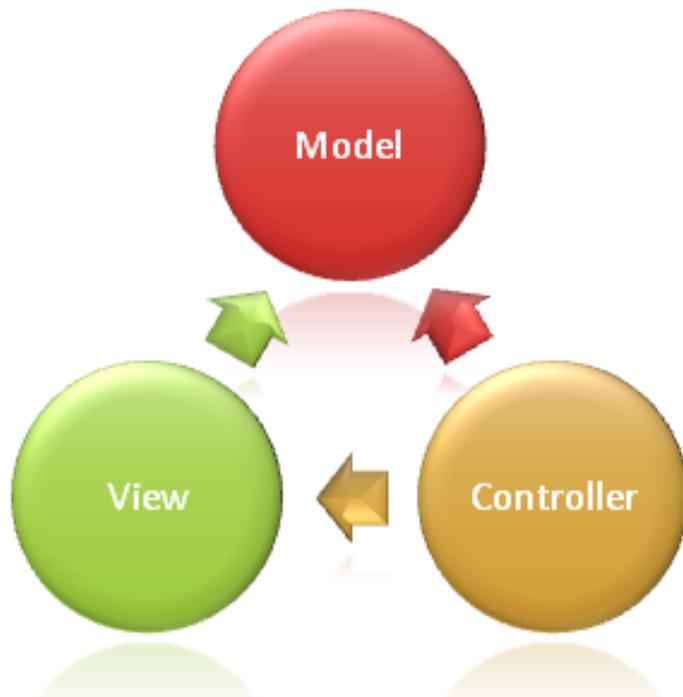


Modules Datalift

Architecture applicative - Pattern

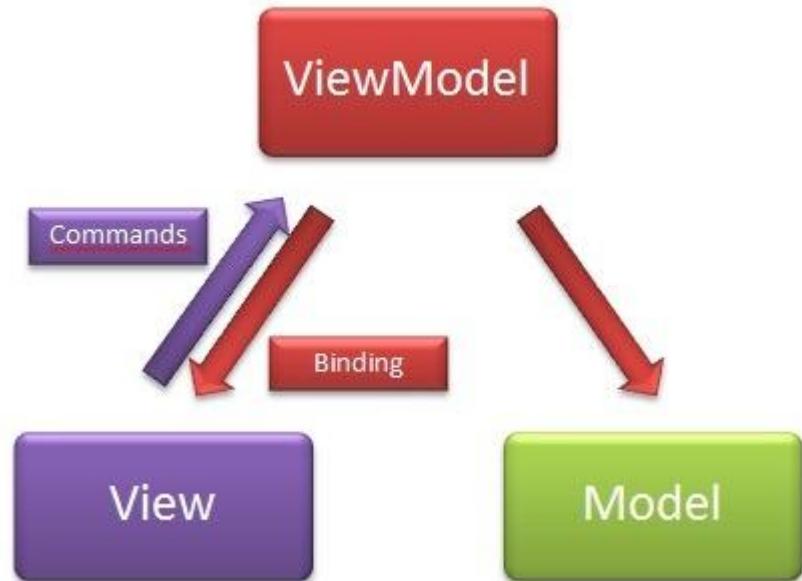
Serveur

MVC



Client léger

MVVM





Modules Datalift

Librairies

Java



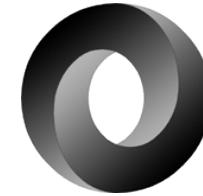
The **Apache Velocity Project**
<http://velocity.apache.org/>



Sesame



jersey



JSON-JAVA

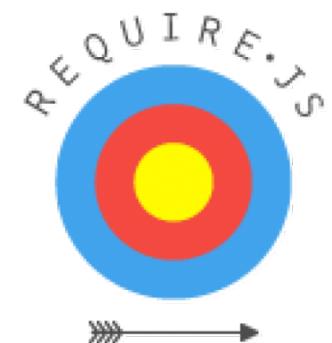


google-gson

JavaScript



Bootstrap, from twitter





Conversion SDMX

SdmxDatcube - Spécifications fonctionnelles



Conversion SDMX

SdmxDatcube - Spécifications fonctionnelles

Conversion source XML / SDMX vers RDF / Datacube

DéTECTER si une source est XML / SDMX

PROPOSER une interface propre et maintenable

Paramétriser la transformation

Paramètres par défaut & aide contextuelle



Conversion SDMX

SdmxDatadcube - Spécifications techniques

Conversion source XML / SDMX vers RDF / Datacube
➤ **SdmxRdfParser**

DéTECTER si une source est XML / SDMX
➤ **Parsing source XML**

Proposer une interface propre et maintenable
➤ **Utilisation de Twitter Bootstrap & knockoutJS**

Paramétriser la transformation

Paramètres par défaut & aide contextuelle



Conversion SDMX

SdmxDatadcube - SdmxRdfParser

openrdf Sesame

```
├── aopalliance
│   └── aopalliance-1.0.jar
├── commons-codec
└── commons-codec-1.4.jar
├── commons-dbcp
└── commons-dbcp-1.3.jar
├── commons-httpclient
└── commons-httpclient-3.1.jar
├── commons-logging
└── commons-logging-1.1.1.jar
├── commons-pool
└── commons-pool-1.5.4.jar
├── dom4j
└── dom4j-1.6.1.jar
├── javax
│   └── servlet
│       ├── servlet-api
│       └── servlet-api-2.5.jar
├── joda-time
└── joda-time-2.1.jar
├── log4j
└── log4j-1.2.14.jar
└── net
    └── sf
        └── opencsv
            └── opencsv-2.3.jar
└── org
    ├── apache
    │   └── xmbeans
    │       ├── xmbeans
    │       └── xmbeans-2.3.0.jar
    │           └── xmbeans-xpath
    │               └── xmbeans-xpath-2.3.0.jar
    ├── aspectj
    │   └── aspectjrt
    │       └── aspectjrt-1.6.11.jar
    ├── openrdf
    │   └── sesame
    │       └── openrdf-sesame-2.6.10-onejar.jar
    └── sdmx
        ├── beans
        │   ├── SDMX1-0
        │   │   └── SDMX1-0-1.0.jar
        │   └── SDMX2-0
        │       └── SDMX2-0-2.0.jar
        └── SDMX2-1
            └── SDMX2-1-3.0.jar
```

Maven

sdmxsource

```
└── sdmxsource
    ├── SdmxApi
    │   └── 0.9.13
    │       └── SdmxApi-0.9.13.jar
    ├── SdmxBeans
    │   └── 0.9.13
    │       └── SdmxBeans-0.9.13.jar
    ├── SdmxDataParser
    │   └── 0.9.13
    │       └── SdmxDataParser-0.9.13.jar
    ├── SdmxEdiParser
    │   └── 0.9.13
    │       └── SdmxEdiParser-0.9.13.jar
    ├── SdmxRdfParser
    │   └── 0.9.13
    │       └── SdmxRdfParser-0.9.13.jar
    ├── SdmxSourceBase
    │   └── 0.9.13
    │       └── SdmxSourceUtil-0.9.13.jar
    ├── SdmxSourceUtil
    │   └── 0.9.13
    │       └── SdmxSourceUtil-0.9.13.jar
    ├── SdmxStructureParser
    │   └── 0.9.13
    │       └── SdmxStructureParser-0.9.13.jar
    ├── SdmxStructureRetrieval
    │   └── 0.9.13
    │       └── SdmxStructureRetrieval-0.9.13.jar
    └── slf4j
        └── slf4j-api
            └── slf4j-api-1.6.1.jar
```

spring fwk

```
└── springframework
    ├── spring-aop
    │   └── spring-aop-3.0.7.RELEASE.jar
    ├── spring-asn
    │   └── spring-asn-3.0.7.RELEASE.jar
    ├── spring-aspects
    │   └── spring-aspects-3.0.7.RELEASE.jar
    ├── spring-beans
    │   └── spring-beans-3.0.7.RELEASE.jar
    ├── spring-context
    │   └── spring-context-3.0.7.RELEASE.jar
    ├── spring-context-support
    │   └── spring-context-support-3.0.7.RELEASE.jar
    ├── spring-core
    │   └── spring-core-3.0.7.RELEASE.jar
    ├── spring-expression
    │   └── spring-expression-3.0.7.RELEASE.jar
    ├── spring-test
    │   └── spring-test-3.0.7.RELEASE.jar
    ├── spring-tx
    │   └── spring-tx-3.0.7.RELEASE.jar
    ├── spring-web
    │   └── spring-web-3.0.7.RELEASE.jar
    └── stax
        ├── stax
        │   └── stax-1.2.0.jar
        └── xml-apis
            └── xml-apis-1.0.b2.jar
```



Conversion SDMX

SdmxDatadcube - SdmxRdfParser

Spring

Dépendance de sdmxsource

Difficultés d'intégration :(

```
// Initialize the sdmxDatadcubeTransformer, which is a Spring bean.  
// It is also referenced in spring-beans.xml.
```

```
Thread.currentThread().setContextClassLoader (  
    this.getClass().getClassLoader());  
ClassPathXmlApplicationContext ctx = new  
ClassPathXmlApplicationContext();  
ctx.setConfigLocation ("spring/spring-beans.xml");  
ctx.refresh();  
  
sdmxDataCubeTransformer = ctx.getBean(SDMXDataCubeTransformer.class);
```



Conversion SDMX

SdmxDatadecube - SdmxRdfParser

```
// TODO Find a way to define URI templates elsewhere.  
private InputStream convert(XmlSource source, RDFFormat rdfFormat) {  
    ByteArrayOutputStream convertedStream = null;  
    if(!isValidSource(source)){  
        LOG.fatal("Source {} is not valid" , source.getUri());  
    } else {  
        try {  
            convertedStream =  
                sdmxDatadecubeTransformer.convertSDMXToDataCube(  
                    source.getInputStream() , rdfFormat);  
        } catch (IOException e) {  
            LOG.fatal("Failed to load stream of source {}: {}" , e,  
                source.getUri() , e.getMessage());  
        }  
    }  
  
    return new ByteArrayInputStream(convertedStream.toByteArray());  
}
```



Conversion SDMX

SdmxDatcube - isValidSource

Critères d'une source valide

XMLSource

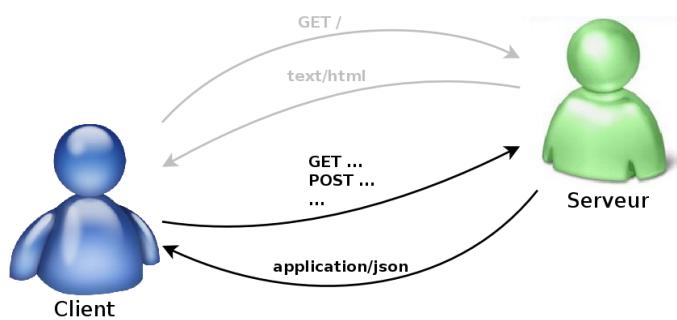
Namespace SDMX en début de fichier

"http://www.sdmx.org/resources/sdmxml/schemas/v2_1/"



Conversion SDMX

SdmxDatcube - Séquence



GET /

200 OK text/html

POST /validate

200 OK application/json

400 BAD REQUEST application/json

POST /

201 CREATED text/plain

400 BAD REQUEST text/plain



Conversion SDMX

SdmxDatadcube - Validation



```
{  
    "dest_graph_uri": "Output source URI already exists (http:  
//localhost:8080/datalift/project/test1/source/agequinquennal6809-full-  
21-qb)" ,  
    "dest_title": "Output source name already used" ,  
    "global": "dest_title: Output source name already used" ,  
    "valid": false  
}
```



Conversion SDMX

SdmxDatadcube - Demo



Visualisation QB



Visualisation QB

Spécifications du module

- Choisir un jeu de données à visualiser
- Explorer les données sous forme tabulaire
- Choisir les dimensions affichées
- Manipuler leur ordonnancement à l'écran
- Visualiser différents graphiques dessinés en temps réel



Visualisation QB

Spécifications du module

The screenshot shows the 'Explorateur de données' (Data Explorer) interface. The left sidebar displays a tree structure of data categories. The 'Statistiques générales et régionales' category is expanded, revealing various sub-sections such as 'Indicateurs européens et nationaux pour l'analyse à court terme (euroind)', 'Enquêtes de conjoncture et de consommation (source: DG ECFIN)', 'Balance des paiements (ei_bp)', and 'Régions maritimes (mare)'. Each category has a small icon and a blue link to its detailed description.

- Base de données par thèmes
- Statistiques générales et régionales
 - Indicateurs européens et nationaux pour l'analyse à court terme (euroind) (Euroind - Old codes-New codes)
 - Enquêtes de conjoncture et de consommation (source: DG ECFIN) (ei_bcs)
 - Enquêtes de consommation (source: DG ECFIN) (ei_bcs_cs)
 - Consommateurs - données mensuelles (ei_bsco_m)
 - Consommateurs - données trimestrielles (ei_bsco_q)
 - Enquêtes de la conjoncture (source: DG ECFIN) (ei_bcs_bs)
 - Balance des paiements (ei_bp)
 - Prix à la consommation (ei_cp)
 - Commerce international (ei_et)
 - Industrie, commerce et services (ei_is) (Collecte commandes industrielles supprimée)
 - Marché du travail (ei_im)
 - Indicateurs monétaires et financiers (ei_mf)
 - Comptes nationaux (ei_qna)
 - Statistiques régionales par classification NUTS (reg)
 - Autres statistiques infranationales (reg_nat)
 - Régions Métropolitaines (met)
 - Régions maritimes (mare)
 - Statistiques démographiques par régions côtières (mare_demo)
 - Superficie des régions (en km²) (mare_d3area)
 - Densité de la population (mare_d3dens)
 - Population par sexe et groupes d'âge au 1er janvier (mare_pjanaggr3)
 - Bilan démographique et taux bruts (mare_gind3)



Visualisation QB

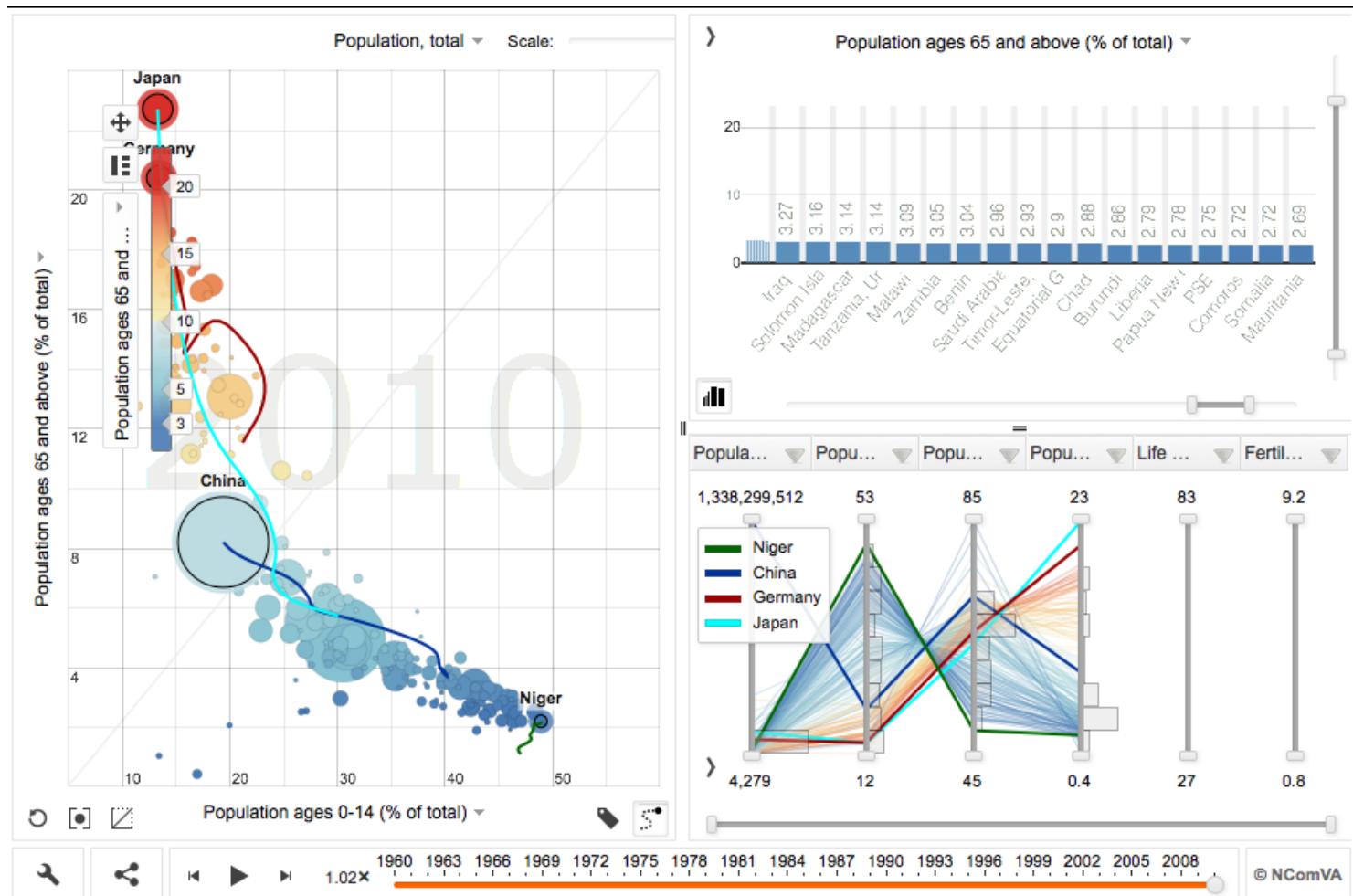
Spécifications du module

| TIME | GEO | INDIC | | | | | | |
|------------------------------|---------------|---------------------------------------|---------|---------|---------|---------|---------|---------|
| S_ADJ | UNIT | Consommateur, indicateur de confiance | | | | | | |
| Données non désaisonnalisées | Bilan (solde) | | | | | | | |
| TIME ▾ | 2012M03 | 2012M04 | 2012M05 | 2012M06 | 2012M07 | 2012M08 | 2012M09 | 2012M10 |
| GEO ▾ | | | | | | | | |
| Union européenne (28) | -19,2 | -19,8 | -18,2 | -18,7 | -19,5 | -21,8 | -22,9 | -24,3 |
| Belgique | -12,3 | -11,1 | -8,5 | -8,0 | -11,8 | -13,4 | -13,4 | -18,6 |
| Bulgarie | -45,4 | -44,4 | -43,2 | -39,7 | -43,0 | -42,0 | -44,9 | -45,5 |
| République tchèque | -28,9 | -25,6 | -28,2 | -25,9 | -24,6 | -24,4 | -27,0 | -29,2 |
| Danemark | 11,9 | 11,9 | 9,5 | 7,3 | 8,5 | 8,3 | 5,3 | 2,5 |
| Allemagne (incluant l' | -1,3 | -1,9 | 0,0 | -1,3 | -3,7 | -8,1 | -9,0 | -8,6 |
| Estonie | -12,5 | -10,6 | -6,1 | -7,9 | -8,5 | -8,0 | -10,1 | -14,6 |
| Irlande | -20,7 | -22,5 | -19,9 | -19,7 | -17,5 | -21,9 | -28,4 | -26,3 |
| Grèce | -79,6 | -78,5 | -76,0 | -72,0 | -64,3 | -62,1 | -74,4 | -76,7 |
| Espagne | -31,6 | -30,0 | -33,1 | -25,2 | -28,4 | -38,2 | -35,6 | -35,6 |
| France | -20,9 | -18,7 | -14,3 | -16,6 | -24,5 | -26,0 | -28,8 | -31,0 |
| Italie | -29,6 | -37,5 | -38,2 | -40,5 | -34,9 | -37,4 | -37,2 | -38,0 |
| Chypre | -42,2 | -37,4 | -41,5 | -46,8 | -51,3 | -55,6 | -53,8 | -58,5 |
| Lettonie | -18,8 | -20,7 | -17,6 | -11,3 | -14,2 | -11,2 | -12,2 | -11,7 |
| Lituanie | -20,6 | -21,0 | -18,5 | -15,5 | -17,6 | -16,8 | -19,6 | -19,5 |



Visualisation QB

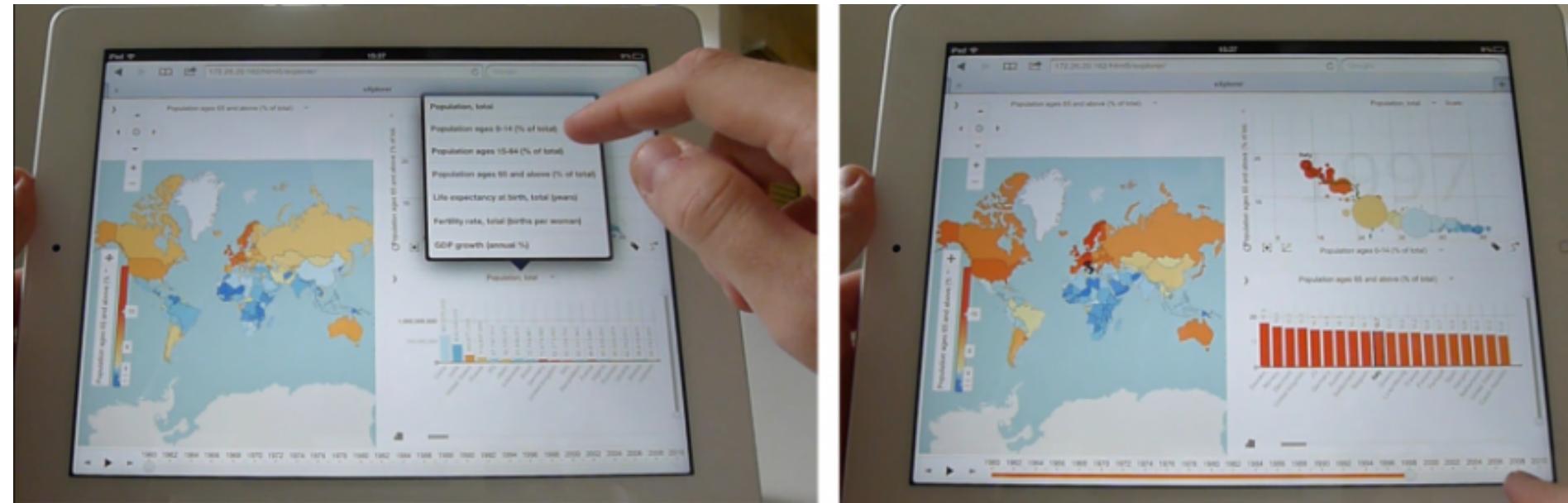
Spécifications du module





Visualisation QB

Spécifications du module





Visualisation QB

Nos choix d'implémentation

| Critère | Poids |
|---------------|-------|
| Usabilité | 20% |
| Visualisation | 35% |
| Accessibilité | 25% |
| Données | 15% |
| Extensibilité | 5% |



Visualisation QB

Déroulement de l'exécution

```
SELECT ?obs ?compoProp ?value
WHERE {
  ?obs a qb:Observation .
  {?obs qb:dataset <http://thib.me/ratp/trafic-2011>}
  UNION
  {?obs qb:dataSet <http://thib.me/ratp/trafic-2011>} .

  ?obs ?compoProp ?value .
  FILTER (?compoProp != rdf:type && ?compoProp != qb:dataset && ?compoProp != qb:dataSet)
}
```



```
?ds rdfs:comment ?descriptionEN} UNION {?ds dct:description ?descriptionEN} UNION {?ds dc:description ?descriptionEN} .  
FILTER langMatches(lang(?descriptionEN), 'EN') .  
}.  
OPTIONAL {  
{?ds rdfs:comment ?descriptionOther} UNION {?ds dct:description ?descriptionOther} UNION  
{?ds dc:description ?descriptionOther} .  
FILTER (!langMatches(lang(?descriptionOther), 'EN') && !langMatches(lang(?  
descriptionOther), 'FR')) .  
} .
```

Visualisation OB

Déroulement de l'exécution

```
OPTIONAL {?ds dct:license ?license} .  
OPTIONAL {{?ds dct:source ?source} UNION {?ds dc:source ?source}} .  
OPTIONAL {?ds rdfs:seeAlso ?seeAlso} .  
  
OPTIONAL {{?ds dct:date ?date} UNION {?ds dc:date ?date}} .  
OPTIONAL {?ds dct:created ?created} .  
OPTIONAL {?ds dct:issued ?issued} .  
OPTIONAL {?ds dct:modified ?modified} .  
  
OPTIONAL {  
{?ds dct:subject ?subject} UNION {?ds dc:subject ?subject}} .  
  
OPTIONAL {  
FILTER (!isLiteral(?subject))  
OPTIONAL {  
?subject a ?subjectType  
FILTER (?subjectType = skos:Concept || ?subjectType = sdmx:Concept || ?subjectType =  
sdmx-subject)  
OPTIONAL {  
{?subject skos:prefLabel ?subjectLabel} .  
}.  
}  
} .
```



Visualisation QB

Démonstration



Conclusion



Conclusion

- Un excellent projet
- De nombreux concepts
- De nombreuses technologies
- Il reste du chemin



Remerciements

- François Scharffe (LIRMM)
- Guillaume Duffes (INSEE)
- Laurent Bihanic (Atos)



