Project Acronym: Open-DAI

Grant Agreement number: 297362

Project Title: Opening Data Architectures and Infrastructures of European Public Administrations

Work Package: Data Assessment and Specification

Deliverable Number: D3.2

Revision History

Revision Date	Author	Organisation	Description
21 May 2012	L.Gioppo D. Tatti	CSI-Piemonte DigitPA	Draft ver. 0.1
28 May 2012	F. Cairo M Palanques C. Tosunoglu	Polito BDigital Sampas	Draft ver. 0.2
30 May 2012	L. Gioppo D. Tatti F. Broman F. Morando	CSI-Piemont DigitPA Netport Politecnico di Torino	Draft ver. 0.3

Legal Disclaimer

Statement of originality:
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

1 Overview 3
1.1 Open data approach 3
1.2 Not only open data 3
1.3 Design rules 4
1.3.1 Naming convention 4
1.3.2 Data service design 4
1.4 Project KPI 4
2 Data services 4
2.1 Piedmont Region 4
2.1.1 Air data set 4
2.1.1.1 Legal aspects 5
2.1.1.2 Data service description 5
2.1.2 TWIST data set 10
2.1.2.1 Legal aspects 10
2.1.2.2 Data service description 10
2.1.2.2.1 Legacy data 10
2.1.2.2.2 New citizen data 10
2.1.2.2.3 Web Service description 11
2.1.2.2.4 Geo Service description 12
2.1.3 Transport data set 12
2.1.3.1 Legal aspects 12
2.2 Barcelona Municipality 13
2.2.1 Data sets providing environmental measures 13
2.2.1.1 Data service description 13
2.2.2 Public measurement stations data set 14
2.2.2.1 Data service description 15
2.2.3 City facilities data set 15
2.2.3.1 Data service description 15
2.2.4 Street directory data set 16
2.2.4.1 Data service description 16
2.2.5 Road congestion data set 17
2.2.5.1 Data service description 17
2.2.6 Data services for jogging route planning 18
2.2.6.1 Data service description 18
2.2.7 Data service for tracking the environmental impact of the public works 20
2.2.7.1 Data service description 20
2.2.8 Legal aspects of the data sets 21
2.3 Lleida Municipality 21
2.3.1 Data sets providing points of interest 22
2.3.1.1 Description of the data sets 22
2.3.1.2 Data service description 23
2.3.2 Road incidents data set 24
2.3.2.1 Data service description 24
2.3.3 Municipal bus services data set 25
2.3.3.1 Data service description 25
2.3.4 Legal aspects of the data sets 27
2.4 Ordu Municipality 27
2.4.1 City Dynamics 27
2.4.1.1 Data service description 27
2.4.2 POIs 31
2.4.2.1 Data service description 31
2.4.3 Complains and Demands 33
2.4.3.1 Data service description 33
2.5 Karlshamn Municipality 36
2.5.1 Data sets containing points of interest 36
2.5.1.1 Data sets 36

2.5.1.1.1 Lighting 36
2.5.1.1.2 Ticket machines 36
2.5.1.1.3 Wastebins 36
2.5.1.1.4 Litterbins 37
2.5.1.2 Data service 37
2.5.1.3 Legal aspects of publishing the data sets 37
2.5.2 Data sets containing reports 37
2.5.2.1 Data set 37
2.5.2.2 Data service 38
2.5.2.3 Legal aspects of publishing the data set 39
3 Semantic Data Integration 40
4 General legal consideration 45

1 Overview

Each data set described in D3.1 will have a number of data services that will be designed to query the exposed data. These data services will be deployed in the Open-DAI platform according to architectural rules defined in WP2.

This document describes such data services that will be designed by each partner.
In this chapter we'll explain the importance of publishing data in open-data format and some of the design rules that will be applied in the project

1.1 Open data approach

Accessing, analysing, reusing, combining, and processing governmental data produces important benefits that can be grouped around two dimensions:

- **Transparency**: Open data is perceived as a powerful instrument to increase transparency in public administration, improving visibility on previously not accessible information, informing citizens and business on policies, public spending and outcomes.
- **New services and economic growth**: data combined in innovative and not initially foreseen ways enables new added-value services, boosting job creation, and resulting in substantial improvements in public service provision while contributing to economic growth.

1.2 Not only open data

The project promote the publishing of data sets in open data format, but public administrations will be convinced to proceed to the breakage of silos only if there can be an immediate return of investment and if this opens up new developments for its own internal needs.
Consortium partners have the evidence that within the same PA the usage of "out of silos" data is beneficial from within the same PA since this allows to bypass long internal procedures to get access to information.
Obviously to implement internal services data set can and must contain data that could not fit into the requirements for "open data".
The strength of the project proposal is such that having a virtualization view of the legacy Db is possible to generate different view of the data set: one for the "open data" approach the other for the internal PA usage.
The internal usage will probably be the most fitted for the SOA approach.
Another aspect, not yet considered within the scope of the project, is the Business Intelligence aspect of correlating different data sets.

1.3 Design rules

1.3.1 Naming convention

Should arise the need to use a namespace the following will be used: **eu.opendai.partner**.

1.3.2 Data service design

In the project there will be two types of data services:

- Data service needed by a pilot
- Data service published for external usage

Obviously the data services used by pilots will be more detailed during pilot design phase and could be revisioned accordingly
Data services published for external usage will respect the expected usage of the data set by external consumers and so basic functionality over data will be implemented.
These functionality will follow this guideline:

- Data will be published in the most complete form possible (all reference to decoding tables will be resolved during publication)
- There will be a paginated list request for basic tables with a significant subset of data
- There will be a single record request for complete data

Should a data service decide to implement filtering options these shall be documented and each filter will have to be passed as a single parameter to the data service call.
For example in case of a geographically filtered data service call parameter shall be expressed in the following way:

- Longitude: float (will have to declare the projection system)
- Latitude: float (will have to declare the projection system)
- Radius: float

1.4 Project KPI

The project will have to produce the following number of data sets according to the project proposal B3.4

Indicator	Year 1	Year 2	Year 3
Number of data sets published	0	15	25

To be able to publish the declared amount of data services in year two at the end of development phase the project have to design accordingly so each data provider partner will have to provide ~4 data sets.

2 Data services

In this chapter will be described the data services calls needed by the pilot and in general offered to the public.
The form of the data service will be as one or more WSDL document to attach to the document.
Here will be listed and explained the function calls of each data service

2.1 Piedmont Region

2.1.1 Air data set

The data set contains information about measurement stations and raw data measurement.

This data has to be standardized towards a common format decided by EU initiative. At the moment of the writing the working group that have to decided the service's call at EU level has not yet released any design specification, thus hereafter will be described minimal data service interfaces that will comply with the data structure proposed by EU and a possible implementation of the data service
The expected output is thus corresponding to the following XML.

Transforming the actual data set to the required output will call for a complex mapping operation, that will be done using TEIID IDE.
Once the mapping is managed it will be created a web service that will publish the required XML.

2.1.1.1 Legal aspects

Considering the nature of the data, the Air data set is not expected to raise significant privacy or data protection issues even in case it is published as open data. Moreover, the legal obligations related with the INSPIRE Directive (2007/2/EC) and international agreements, such as the Aarhus Convention already imply significant publicity requirements (which are currently respected by the Piedmont Region). Problems connected with potential misinterpretations of the data will be addressed through the publication of appropriate metadata, for instance in order to help re-users in making appropriate comparisons with other external data.

2.1.1.2 Data service description

GetEnvironmentalMonitoringFacilities
This data service will return a list of XML elements as the one below.
Considering the amount of data the service will not be paginated nor will be filtered.

```
<ef:EnvironmentalMonitoringFacility gml:id="idvalue0" xmlns:base="urn:x-
inspire:specification:gmlas:BaseTypes:3.2"
xmlns:base2="http://inspire.jrc.ec.europa.eu/schemas/base2/0.1"
xmlns:ef="http://inspire.jrc.ec.europa.eu/schemas/ef/2.0"
xmlns:gco="http://www.isotc211.org/2005/gco"
xmlns:gmd="http://www.isotc211.org/2005/gmd"
xmlns:gml="http://www.opengis.net/gml/3.2"
xmlns:gsr="http://www.isotc211.org/2005/gsr"
xmlns:gss="http://www.isotc211.org/2005/gss"
xmlns:gts="http://www.isotc211.org/2005/gts"
xmlns:om="http://www.opengis.net/om/2.0"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://inspire.jrc.ec.europa.eu/schemas/ef/2.0
EnvironmentalMonitoringFacilities.xsd ">
  <ef:inspireId>
    <base:Identifier>
      <base:localId>base:localId</base:localId>
      <base:namespace>base:namespace</base:namespace>
    </base:Identifier>
  </ef:inspireId>
  <ef:beginLifespan>2001-12-31T12:00:00</ef:beginLifespan>
  <ef:measurementRegime>ef:measurementRegime</ef:measurementRegime>
  <ef:mediaMonitored>ef:mediaMonitored</ef:mediaMonitored>
  <ef:mobile>ef:mobile</ef:mobile>
  <ef:operationalActivityPeriod/>
</ef:EnvironmentalMonitoringFacility>
```

getRawData

Due to the potential big amount of data that his service could return the settings of many the filter's parameters will be compulsory.
The XML returned will be done according to the following model

```
<?xml          version="1.0"          encoding="UTF-8"?><gml:FeatureCollection
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:aqd="http://www.exampleURI.com/AQD"                    xmlns:base="urn:x-
inspire:specification:gmlas:BaseTypes:3.2"
xmlns:gmd="http://www.isotc211.org/2005/gmd"
xmlns:gco="http://www.isotc211.org/2005/gco"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:gml="http://www.opengis.net/gml/3.2"
xmlns:gss="http://www.isotc211.org/2005/gss"
xmlns:gts="http://www.isotc211.org/2005/gts"
xmlns:gsr="http://www.isotc211.org/2005/gsr"
xmlns:ef="http://inspire.jrc.ec.europa.eu/schemas/ef/2.0"
xmlns:base2="http://inspire.jrc.ec.europa.eu/schemas/base2/0.1"
xmlns:om="http://www.opengis.net/om/2.0"
xmlns:swe="http://www.opengis.net/swe/2.0"
xmlns:sams="http://www.opengis.net/samplingSpatial/2.0"
xmlns:sam="http://www.opengis.net/sampling/2.0"
xmlns:am="http://inspire.jrc.ec.europa.eu/schemas/am/2.0"     xmlns:gn="urn:x-
inspire:specification:gmlas:GeographicalNames:3.0"                  xmlns:am-
ru="http://inspire.jrc.ec.europa.eu/schemas/am-ru/2.0"
xsi:schemaLocation="http://www.exampleURI.com/AQD                      AQD.xsd
http://www.opengis.net/gml/3.2   http://schemas.opengis.net/gml/3.2.1/gml.xsd"
gml:id="observation">
  <gml:featureMember>
    <om:OM_Observation gml:id="Observation_IT1247A_">
      <om:phenomenonTime>
        <gml:TimePeriod gml:id="ObservationTimePeriodLOCAL_ID_0">
          <gml:beginPosition>2012-01-05T00:00:00</gml:beginPosition>
          <gml:endPosition>2012-01-06T00:00:00</gml:endPosition>
        </gml:TimePeriod>
      </om:phenomenonTime>
      <om:resultTime>
        <gml:TimeInstant gml:id="ObservationResultInstance_8">
          <gml:timePosition>2012-04-23T17:33:34</gml:timePosition>
        </gml:TimeInstant>
      </om:resultTime>
      <om:procedure xlink:href="Process_"/>
      <om:parameter>
        <om:NamedValue>
          <om:name xlink:href="samplingPoint"/>
          <om:value>SamplingPoint_IT1247A_sp1</om:value>
        </om:NamedValue>
      </om:parameter>
      <om:observedProperty xlink:href="#Biossido di azoto (NO2)"/>
      <om:featureOfInterest
xlink:href="SampleFeature_IT1247A_sp1">SampleFeature_IT1247A_sp1</om:featureO
fInterest>
      <om:result>
        <swe:DataArray>
          <swe:elementCount>
            <swe:Count>
              <swe:value>24</swe:value>
            </swe:Count>
```

```
          </swe:elementCount>
          <swe:elementType name="Components">
            <swe:DataRecord>
              <swe:field name="time">
                <swe:Time definition="urn:ogc:property:time:iso8601">
                  <swe:uom code="hour"/>
                </swe:Time>
              </swe:field>
              <swe:field name="Validity">
                <swe:Quantity definition="Validity">
                  <swe:uom code="Codelist"/>
                </swe:Quantity>
              </swe:field>
              <swe:field name="Verification">
                <swe:Quantity definition="Verification">
                  <swe:uom code="Codelist"/>
                </swe:Quantity>
              </swe:field>
              <swe:field name="Biossido di azoto (NO2)">
                <swe:Quantity definition="Biossido di azoto (NO2)">
                  <swe:uom code="Microgrammi al metro cubo"/>
                </swe:Quantity>
              </swe:field>
            </swe:DataRecord>
          </swe:elementType>
          <swe:encoding>
            <swe:TextEncoding     decimalSeparator="."     tokenSeparator="," 
blockSeparator="@@"/>
          </swe:encoding>
          <swe:values>2012-01-05T01:00:00,0,0,28.000@@2012-01-
05T02:00:00,0,0,14.000@@2012-01-05T03:00:00,0,0,19.000@@2012-01-
05T04:00:00,0,0,14.000@@2012-01-05T05:00:00,0,0,10.000@@2012-01-
05T06:00:00,0,0,21.000@@2012-01-05T07:00:00,0,0,59.000@@2012-01-
05T08:00:00,0,0,56.000@@2012-01-05T09:00:00,0,0,63.000@@2012-01-
05T10:00:00,0,0,76.000@@2012-01-05T11:00:00,0,0,69.000@@2012-01-
05T12:00:00,0,0,42.000@@2012-01-05T13:00:00,0,0,41.000@@2012-01-
05T14:00:00,0,0,29.000@@2012-01-05T15:00:00,0,0,36.000@@2012-01-
05T16:00:00,0,0,63.000@@2012-01-05T17:00:00,0,0,76.000@@2012-01-
05T18:00:00,0,0,70.000@@2012-01-05T19:00:00,0,0,77.000@@2012-01-
05T20:00:00,0,0,62.000@@2012-01-05T21:00:00,0,0,66.000@@2012-01-
05T22:00:00,0,0,47.000@@2012-01-05T23:00:00,0,0,18.000@@2012-01-
05T24:00:00,0,0,7.000@@</swe:values>
        </swe:DataArray>
      </om:result>
    </om:OM_Observation>
  </gml:featureMember>
</gml:FeatureCollection>
```

Parameters

- temporal filter (the month and the year)
- station ID
- type of pollutant

getZones

This call will return the measurement zones with the information of the pollutants measurements collected.

The method will return the complete collection without any paramters.

```
<gml:featureMember>
      <aqd:AQD_Zone gml:id="IT0116__15">
            <am:inspireId>
                  <base:Identifier>
                        <base:localId>IT0116</base:localId>
                        <base:namespace>aqd.it.zones</base:namespace>
                  </base:Identifier>
            </am:inspireId>
            <am:name>
                  <gn:geographicalName>
                  <gn:nativeness>edonym</gn:nativeness>
                  <gn:nameStatus>standardised</gn:nameStatus>

      <gn:sourceOfName>zones_agglomeration_Piedmont_Region.shp</gn:sourceOfNam
e>
                  <gn:pronunciation nillReason="Unknown" xsi:nil = "true"/>
                  <gn:spelling>
                        <gn:spellingOfName>
                              <gn:text>Verbania 01</gn:text>
                              <gn:script>Latn</gn:script>
                        </gn:spellingOfName>
                  </gn:spelling>
                  </gn:geographicalName>
            </am:name>
                  <am:specialisedZoneType>
                        nonagglomeration
                  </am:specialisedZoneType>
<!-- In spreadsheet of mappings ... this is zoneType - left both here just to
check them both-->
                  <am:zoneType>
                        nonagglomeration
                  </am:zoneType>
                  <am:zoneType                        codeSpace="http://inspire-
registry.jrc.ec.europa.eu/registers/CLR/ZoneCodeType">airQualityManagementZon
e</am:zoneType>

                  <am:validTime nilReason="Unknown" xsi:nil="true"/>
                  <am:competentAuthority nilReason="Unknown" xsi:nil="true"/>
                  <am:legalBasis>
                    <base2:LegislationReference>
                      <base2:legalName>Directive 2008/50/EC of the European
Parliament and of the Council of 21 May 2008 on ambient air quality and
cleaner air for Europe</base2:legalName>
                      <base2:shortName>Air Quality Directive</base2:shortName>

<base2:identificationNumber>2008/50/EC</base2:identificationNumber>

<base2:officialDocumentNumber>32008L0050</base2:officialDocumentNumber>
                      <base2:linkToLegislativeInstrument>http://eur-
lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32008L0050:EN:NOT</base2:lin
kToLegislativeInstrument>
                      <base2:publicationDate>2008-05-
21</base2:publicationDate>
                      <base2:dateEnteredIntoForce>2008-06-
11</base2:dateEnteredIntoForce>
                      <base2:level>international</base2:level>
```

```
                    </base2:LegislationReference>
                  </am:legalBasis>
                  <am:beginLifespanVersion                    nilReason="Unknown"
xsi:nil="true"/>

            <aqd:zoneCode>IT0116</aqd:zoneCode>
            <aqd:LAU>unknown</aqd:LAU>
            <aqd:residentPopulation>116576</aqd:residentPopulation>
            <aqd:residentPopulationYear>2009-01-
01</aqd:residentPopulationYear>
<!-- The GML geometry object ENDS here .. -->
            <am:geometry>
<!-- There are 0 GML geometry objects -->
<!-- this is a gml polygon  - local id supplied and SRS -->
                  <gml:Polygon           srsName="urn:ogc:def:crs:EPSG::32632"
gml:id="GeomID_IT0116__15"><gml:outerBoundaryIs><gml:LinearRing><gml:coordina
tes>nil</gml:coordinates></gml:LinearRing></gml:outerBoundaryIs></gml:Polygon
>
            </am:geometry>
<!-- The GML geometry object ENDS here .. -->
            <!-- Metals and Pollutant listing .. -->
<!-- Codelist not available - used chemical symbols as temporary data-->
<aqd:pollutants>
      <aqd:pollutantCode>SO2</aqd:pollutantCode>
      <aqd:protectionTarget>Health</aqd:protectionTarget>
</aqd:pollutants>
<aqd:pollutants>
      <aqd:pollutantCode>NO2</aqd:pollutantCode>
      <aqd:protectionTarget>Health</aqd:protectionTarget>
</aqd:pollutants>
<aqd:pollutants>
      <aqd:pollutantCode>PM10</aqd:pollutantCode>
      <aqd:protectionTarget>Health</aqd:protectionTarget>
</aqd:pollutants>
<aqd:pollutants>
      <aqd:pollutantCode>PM25</aqd:pollutantCode>
      <aqd:protectionTarget>Health</aqd:protectionTarget>
</aqd:pollutants>
<aqd:pollutants>
      <aqd:pollutantCode>Pb</aqd:pollutantCode>
      <aqd:protectionTarget>Health</aqd:protectionTarget>
</aqd:pollutants>
<aqd:pollutants>
      <aqd:pollutantCode>C6H6</aqd:pollutantCode>
      <aqd:protectionTarget>Health</aqd:protectionTarget>
</aqd:pollutants>
<aqd:pollutants>
      <aqd:pollutantCode>CO</aqd:pollutantCode>
      <aqd:protectionTarget>Health</aqd:protectionTarget>
</aqd:pollutants>
<aqd:pollutants>
      <aqd:pollutantCode>O3</aqd:pollutantCode>
      <aqd:protectionTarget>Health and Vegetation</aqd:protectionTarget>
</aqd:pollutants>
      </aqd:AQD_Zone>
</gml:featureMember>
```

2.1.2 TWIST data set

The TWIST data set will make use of different services for exposing data.
There will be the need to:

- expose legacy data
- publish data collected from citizen
- offer services for the accident application
- publish accident in a map for public administration usage.

2.1.2.1 Legal aspects

Considering the nature of the data, including the geo- and chrono- referenced microdata about individual car accidents (and involved people), the TWIST data set presents major privacy and data protection issues, in particular if the data are to be made available to third parties (or as open data). That applies even if all direct personal identifiers will be anonymized. Moreover, even the use within the Piedmont Region needs to respect a memorandum of understanding with the Italian statistical office (ISTAT).
Within the context of the Open-DAI project, the Piedmont Region identified a set of data which could be re-used internally or with selected partners and another subset which could possibly be published as open data. These data protection issues are currently being explored and a request to the Italian data protection authority is being drafted (at the moment of writing this deliverable, the possibility of drafting a joint request together with ISTAT is under examination).

2.1.2.2 Data service description

Services based on pure data will be distinguished from the origin of the data one set from the legacy DB and a second one from the new collected data from citizen that represent real time data.

2.1.2.2.1 Legacy data

getAllAccidentLegacyData
This data service will return a paginated list of data based on the filter parameters with the following information:

- accident ID
- location data
- temporal data

Parameters

- page
- location filter (center of search and radius); defaulting to all locations
- temporal filter (from date, to date); defaulting to current year

getAccidentLegacyDataDetail
This will return all accident data available.
The parameter accepted will be the accident ID.

2.1.2.2.2 New citizen data

setAccident
This service will be used to add a new accident to the new tableset built for the project to host real time accident, since all these data will have to be anonymized and is not verified by PA authorities it will have to be placed on dedicated tables and exposed with a data quality asserting that it comes from a not verified source.

It will return insert information
Parameters
location information (longitude and latitude)

- temporal information (time of the accident)
- test (boolean information to assess if the information is a real one or a testing data)
- onBehalf (boolean to assert if the user is sending information on behalf of someone else)

getAllAccidentCitizenData
This data service will return a paginated list of data based on the filter parameters with the following information:

- accident ID
- location data
- temporal data

Parameters

- page
- location filter (center of search and radius); defaulting to all locations
- temporal filter (from date, to date); defaulting to current year

getCitizenProcessData
This service will get the customization of the process based on the user saved preferences.
Parameters

- user ID

Data returned will be the customization information for the process; this information will be detailed during the pilot design phase.

SetCitizenProcessData
This service will save the customization of the process and link it to all the phone the user could use.
The detail of this dataset will be depending on the detailed process definition parameters that will be designed during the pilot so is not possible to describe in this document the specifics of the call.
It will be a web service call with an array of parameters that represents the process variables to save in the DB together with the user ID

setCitizenInfo
This will create an account for the citizen and link the account to a list of telephone numbers (a user could have more than one phone).
This also will be dependent on the detailed specifics and design of the pilot.

2.1.2.2.3 Web Service description

This service will be the one used by the mobile application to send the accident data to the process engine.
Since the process engine need to recover all the custom process information of the single user the telephone number will also be passed, but will be used just to customize the process flow and not be saved with the accident information.
Parameters

- user ID (telephone number)
- location information (longitude and latitude)
- temporal information (time of the accident)
- test (boolean information to assess if the information is a real one or a testing data)
- onBehalf (boolean to assert if the user is sending information on behalf of someone else)

There will be two important parameters in the call:
The first represent if the call is a testing one (used both for development and end user testing)
The second represent the possibility to refer an accident on behalf of others to enable social help, but with the opportunity to correctly tag the event with a lower data quality.

2.1.2.2.4 Geo Service description

Accident data will also published as a georeferenced cloud of points for PA usage.
This service will be offered by the GeoServer layer that will access the virtual database and export a KML view of the data.
Filters will be the same as the one used for the alphanumeric data service.

2.1.3 Transport data set

This data set will publish most of the virtual database tables in the logic explained in chapter 1 of this document.
Apart from publishing data as web services the pilot will evaluate also the possibility to publish information as “Transit Feed Specification”.
The specification of the Feed Files is described below and the minimal target is to publish the “Required” Feeds.

Filename	Required	Defines
agency.txt	Required	One or more transit agencies that provide the data in this feed.
stops.txt	Required	Individual locations where vehicles pick up or drop off passengers.
routes.txt	Required	Transit routes. A route is a group of trips that are displayed to riders as a single service.
trips.txt	Required	Trips for each route. A trip is a sequence of two or more stops that occurs at specific time.
stop_times.txt	Required	Times that a vehicle arrives at and departs from individual stops for each trip.
calendar.txt	Required	Dates for service IDs using a weekly schedule. Specify when service starts and ends, as well as days of the week where service is available.
calendar_dates.txt	Optional	Exceptions for the service IDs defined in the calendar.txt file. If calendar_dates.txt includes ALL dates of service, this file may be specified instead of calendar.txt.
fare_attributes.txt	Optional	Fare information for a transit organization's routes.
fare_rules.txt	Optional	Rules for applying fare information for a transit organization's routes.
shapes.txt	Optional	Rules for drawing lines on a map to represent a transit organization's routes.
frequencies.txt	Optional	Headway (time between trips) for routes with variable frequency of service.
transfers.txt	Optional	Rules for making connections at transfer points between routes.
feed_info.txt	Optional	Additional information about the feed itself, including publisher, version, and expiration information.

2.1.3.1 Legal aspects

The Transport dataset does not raise the same data protection issues mentioned for the previous dataset, since it just includes non-personal data; however this dataset includes data provided by several third parties, which are legally independent from the Piedmont Region, which in some cases is just playing the role of collector of these data. The use of these data for internal purposes of the Piedmont Region is already covered by existing agreements, however, in order to maximize the amount of data which could also be published as open data, a rights clearing process is ongoing (and involving other public administrations, such as provinces, and public transport organizations).

2.2 Barcelona Municipality

The data sets available for the Barcelona Municipality store three basic types of contents: environmental measurements, information on public facilities managed by the city council and mobility information in the form of semi-real time road congestion measurements. Since most of the data sets on environmental measures share a very similar model, a single data service will be deployed to expose all these data services. Requests against this service will specify the targeted data set through a URI parameter.

Together with the data services providing raw data, two data services will also offer already processed data required for the Barcelona pilots: the data service on jogging route planning and the one on environmental impact of the public works.

Each of the described data services will be built as a RESTful service and will use JSON as the data-interchange format through their public API .

2.2.1 Data sets providing environmental measures

Pollution data set
This data set contains information on pollution levels, which is collected through 11 sensors around the city and updated every 30 minutes. The 11 sensors are further used to estimate pollution around the city. Such estimation is published after 2 months in a separate data set. However, raw collected information cannot be used to extrapolate pollution levels in any point of the city.

Noise level data set
The city council elaborates a noise level map every 5 years, as required by European regulation, which specifies, mean noise levels. More specifically, each street block is assigned three noise level intervals, depending on the measuring time (from 7 AM to 9 PM, from 9 PM to 11 PM and from 11 PM to 7 AM). The information is currently published by the city council of Barcelona, although not integrated in the Open Data portal.

Weather prediction data set
Weather prediction information is currently published by the regional government including prediction for the morning and the afternoon and expected maximum and minimum temperatures. This information is updated daily. Real data on weather measurements collected through one meteorological station in Barcelona is also daily published, including three daily measurements of temperature, humidity, wind, pressure, presence of clouds and visibility.

Pollen level data set
The pollen level data set contains weekly measurements and predictions on the level of presence in the air of pollen of 19 different families of plants. Current presence level is indicated in a scale from 0 to 4, meaning null and very high respectively, while prediction can be classified as increasing, deceasing, sustained and dangerous. Data is currently published by the Environmental Science and Technology Institute (ICTA) of the Universitat Autònoma de Barcelona (UAB).

2.2.1.1 Data service description

As mentioned before, a single data service is intended to provide the access point to all the similar data sets. Although the return JSON documents will share a common structure, variable-specific attributes will be added within the “entry” attribute when required.

Taking this into account, the general REST API for the data service will provide the following resources:

Resource	Description
GET measurements/type	Returns the complete list of measurements for a given type (pollution, noise…).

GET measurements/type/timeslot	Returns the list of measurements for a given type in the provided time slot.
GET measurements/stationed	Returns the list of measurements retrieved from a given station id.

Together with that, the basic structure of the JSON responses for the service is provided below:

```
{
   "response":{
      "meta":{
         "v":"1.0",
         "status":"OK",
         "code":200,
         "timeRef":1340684854471,
         "msg":""
 },
      "data":{
         "startIndex":0,
         "itemsPerPage":1,
         "totalResults":1,
         "entry":[
            {
               "guid":”21EC2020-3AEA-1069-A2DD-08002B30309D",
               "type":"Pol·lució",
               "description":"Mesura nivell de pol·lució a la Diagonal",
               "begindate":12323123123,
               "enddate":12323123123,
               "measurements":[
               {
                  "00:00":"84",
                  "station":"BN32"

               },
               {
                  "01:00":"82"
                  "station":"BN32"
               }
             ]
            }
         ]
      }
   }
```

The document payloads are divided into two main sections:

- `meta:` includes information on the current API version and on the return codes for the requested resource.
- `data:` collects the actual requested information. Together with the collection of data entries, the object also provides information for pagination.

2.2.2 Public measurement stations data set

This data set contains details (location, managing company, start and end date, and others) of current public works that are equipped with sensors and the measurements collected by the latter. This

information will be used, together with other related parameters such as road congestion nearby the works, to evaluate the impact of those works.

2.2.2.1 Data service description

The REST API for the data service will provide the following resources:

Resource	Description
GET stations/all	Returns the complete list of environmental measurement stations
GET stations/type	Return the list of environmental measurement stations providing information on a given type of environmental measurements (pollution, pollen…)
GET stations/id	Returns full details on a given measurement station.

Together with that, the basic structure of the JSON responses for the service is provided below:

```
{
   "response":{
      "meta":{
         "v":"1.0",
         "status":"OK",
         "code":200,
         "timeRef":1340684854471,
         "msg":""
 },
      "data":{
         "startIndex":0,
         "itemsPerPage":1,
         "totalResults":1,
         "entry":[
            {
               "guid":”BN32",
               "name":"Estació Besós Sud",
               "description":"Estació de mesura nivell de pol·lució a la
Diagonal",
               "type":"pollution",
               "coordinates":"0.2345345345,4.5454545"
            }
         ]
      }
   }
```

2.2.3 City facilities data set

Information on the facilities of the city is currently published in the Open Data portal of the city of Barcelona. The data set contains detailed information of the services provided by every cultural and leisure facility as well as its location and a classification of the facilities.

2.2.3.1 Data service description

The REST API for the data service will provide the following resources:

Resource	Description

GET stations/all	Returns the complete list of environmental measurement stations
GET stations/type	Return the list of environmental measurement stations providing information on a given type of environmental measurements (pollution, pollen…)
GET stations/id	Returns full details on a given measurement station.

Together with that, the basic structure of the JSON responses for the service is provided below:

```
{
   "response":{
      "meta":{
         "v":"1.0",
         "status":"OK",
         "code":200,
         "timeRef":1340684854471,
         "msg":""
 },
      "data":{
         "startIndex":0,
         "itemsPerPage":1,
         "totalResults":1,
         "entry":[
            {
               "guid":"1295081951",
               "name":"Centre Cívic Guinardó",
               "description":"Centre de barri per infants i gent gran",
               "type":[
                   "Centre social",
                   "Espai per nens"
               ],
               "coordinates":"0.2345345345,4.5454545",
               "telephone":"2344343434",
               "url":http://www.bcn.cat/ccg,
               "adapted":"true"
            }
         ]
      }
   }
```

2.2.4 Street directory data set

The street directory data set contains a definition of the street stretches, including a stretch identifier, the street it belongs to, crossing streets and GPS coordinates. This information is currently published through the Open Data portal of Barcelona.

2.2.4.1 Data service description

The REST API for the data service will provide the following resources:

Resource	Description
GET stretches/all	Returns the complete list of street stretches in the city.
GET stretches/id	Returns full details on a specific street stretch.

GET stretches/street	Returns all the stretches in a given street.
GET stretches/coordinates	Returns the list of stretches around the provided coordinates.

Together with that, the basic structure of the JSON responses for the service is provided below:


```
{
   "response":{
      "meta":{
         "v":"1.0",
         "status":"OK",
         "code":200,
         "timeRef":1340684854471,
         "msg":""
 },
      "data":{
         "startIndex":0,
         "itemsPerPage":1,
         "totalResults":1,
         "entry":[
            {
               "guid":"1295081951",
               "name":"Gran Via - Diputació",
               "street":"Pau Claris",
               "originstreet":"Gran Via",
               "destinationstreet":"Diputació",
               "crossstreets":[
               ],
               "coordinates":"0.456464564,0.32342424]
            }
         ]
      }
   }
```


2.2.5 Road congestion data set

The road congestion dataset contains information on current and predicted congestion of each road stretch, as defined in the Street Directory data set, later defined. For each road stretch, the data set contains the current road congestion status (for instance dense or congested) and the estimated status in 15 minutes. The possible statuses include the options of the road being closed and data not being available. This information is currently published through the Open Data portal of Barcelona.

2.2.5.1 Data service description

The REST API for the data service will provide the following resources:

Resource	Description
GET congestion/all	Returns the complete list of congestion information for all the stretches in the city.
GET congestion/coordinates	Returns the list of congestion information for the stretches around the provided coordinates
GET congestion/status	Returns all the stretches in a given status. Useful to retrieve all the dense or congested stretches from a single call to the system.
GET congestion/id	Returns the congestion information on a specific

	stretch.

Together with that, the basic structure of the JSON responses for the service is provided below:

```
{
   "response":{
      "meta":{
         "v":"1.0",
         "status":"OK",
         "code":200,
         "timeRef":1340684854471,
         "msg":""
 },
      "data":{
         "startIndex":0,
         "itemsPerPage":1,
         "totalResults":1,
         "entry":[
            {
               "guid":"1295081951",
               "stretchid":"12223222",
               "status":{
                  "current":"good",
                  "forecast":"dense"
               }
            }
         ]
      }
   }
```

2.2.6 Data services for jogging route planning

2.2.6.1 Data service description

The REST API for the data service will provide the following resources:

Resource	Description
GET routes/all/coordinates	Returns the complete list of routes sorted according to its proximity sorted according to its proximity to the provided coordinates.
GET routes/all/userid/coordinates	Returns the complete list of routes sorted according to its proximity sorted according to its proximity to the provided coordinates. The system takes into account the user id to take into account the previously provided routes in the recommendation process.
GET routes/routeid/	Returns full details on a specific route.
GET routes/prediction/all/coordinates/timeslot	Returns a prediction on best routes for the provided future time slot sorted according to its proximity sorted according to its proximity to the provided coordinates.
GET facilities/facilityid	Returns full details on the provided facility.

The basic structure of the JSON responses for the resources related to routes is provided below:

```
{
   "response":{
      "meta":{
         "v":"1.0",
         "status":"OK",
         "code":200,
         "timeRef":1340684854471,
         "msg":""
 },
      "data":{
         "startIndex":0,
         "itemsPerPage":1,
         "totalResults":1,
         "entry":[
            {
               "guid":”21EC2020-3AEA-1069-A2DD-08002B30309D",
               "name":"Itinerari Eixample esquerra",
               "description":"L’itinerari transcorre durant 4 quilòmetres
pels carrers de l’Eixample esquerra",
               "stretches":[
               {
                  "name":"Aragó – Casanova / Aragó - Urgell"
                  "origincoordinates":"41.656497,0.703125",
                  "endcoordinates":"41.656497,0.703125"
               },
               {
                  "name":"Aragó – Urgell / Aragó - Balmes"
                   "origincoordinates":"41.656497,0.703125",
                   "endcoordinates":"41.656497,0.703125"
               }
              ]
            }
         ]
      }
   }
```


The JSON for the resources related to facilities:


```
{
   "response":{
      "meta":{
         "v":"1.0",
         "status":"OK",
         "code":200,
         "timeRef":1340684854471,
         "msg":""
 },
      "data":{
         "startIndex":0,
         "itemsPerPage":1,
         "totalResults":1,
         "entry":[
            {
               "guid":"21EC2020-3AEA-1069-A2DD-08002B30309D",
               "name":"La Fonteta",
```

```
                "type":"Centre cívic":
                "description":"Edifici històric amb jardí i spa",
                "telephone":"234232344",
                "coordinates":"41.656497,0.703125",
                "adapted":"true"
            }
        ]
    }
}
```

2.2.7 Data service for tracking the environmental impact of the public works

2.2.7.1 Data service description

The REST API for the data service will provide the following resources:

Resource	Description
GET measurements/all/	Returns the complete list of available environmental measurements taken in the proximity of the public works.
GET measurements/all/timeslot	Returns the list of measurements in the provided timeslot
GET measurements/type	Returns the list of measurements for a given type .
GET measurements/type/timeslot	Returns the list of measurements for a given type in the provided timeslot
GET measurements/id	Returns full details on a specific measurement.
GET measurements/id/timeslot	Returns full details on a specific measurement in the provided timeslot.

The basic structure of the JSON responses is provided below:

```
{
   "response":{
      "meta":{
         "v":"1.0",
         "status":"OK",
         "code":200,
         "timeRef":1340684854471,
         "msg":""
 },
      "data":{
         "startIndex":0,
         "itemsPerPage":1,
         "totalResults":1,
         "entry":[
            {
               "guid":”21EC2020-3AEA-1069-A2DD-08002B30309D",
               "name":"Pol·len en suspensió",
               "begindate":”123323233322”,
               "enddate":”1233232333232”,
               "values":[
```

```
                {
                   "station":"Collserola"
                   "xvalue":"08:00"
                   "yvalue":"23.5"
                }
             ]
          }
       ]
    }
 }
```

2.2.8 Legal aspects of the data sets

The joint analysis of the licensing policies and terms of use of the original sources from which the Barcelona pilot data sets are collected provides the basis for the specification of this legal framework. On one side, the Barcelona Open Data is published mostly under a Creative Commons CC-by 3.0 license. In practice, this means that it is allowed to reproduce, distribute, to communicate and to transform data in order to generate derivative works. These derivative works can be used for any commercial and non-commercial purposes, with the sole exception of the activities related to public administration. However, for those data sets where third parties beyond the Barcelona Municipality are involved, the applied license is CC BY-ND 3.0.

Together with that, this is also the licensing model for the Open Data portal of the Catalan government, which provides a major part of the data sets to be integrated during the pilot development. In this case, though, some of the provided data sets are published freely with no restriction at all.

Both sites are also compliant with the Spanish regulation on reuse of public information. This regulation states that:

- The purpose of the information should not be modified
- The sources should always be cited
- The date of the last updates should always be provided

Finally, the data sets provided by the UAB University (Universitat Autonoma de Barcelona) belong to the Aerobiological Network of Catalunya (Xarxa Aerobiològica de Catalunya, XAC) and cannot be reproduced without an express authorization. Such authorization is currently under negotiation.

Besides this, no Service Level Agreement on data availability or system performance will be granted by Open-DAI, since it relies on the source services provided by the municipality and these ones do not provide any SLA.

The licensing policies of the Open- DAI data sets for the Barcelona Municipality will be fully aligned with the ones from the original models.

2.3 Lleida Municipality

Since many of the data sets envisioned for the Lleida Municipality provide information on very similar models (geo-located information on points of interest in the city), a single data service can be responsible for the publication of those data sets through the Open-DAI system. Therefore, only three data services are described in the next section: service providing information on the POIs, service providing information on the public bus lines and service providing accessibility information for places.

2.3.1 Data sets providing points of interest

The data sets in the following section provide several views on the different points of interest located in the municipality of Lleida. Each of the data sets is focused in a specific type of points of interest, such as information offices, accommodations or health services.

The data service responsible for the publication of the data services will offer a REST API providing JSON data. This technological approach aims to provide the optimal backend support for the development of mobile services, since the use of JSON for representation reduces the computational cost related with processing XML documents. The transformation from the incoming data stream to the described output will be carried out by the native REST support in TEIID IDE. The process will also carry out the transformation of the textual location information found the incoming data streams into geospatial coordinates.

2.3.1.1 Description of the data sets

Private bus services data set

Information is provided by private companies managing bus services within the city. The data set contains information on the existing bus companies and their location and contact information, and is currently collected by the local police call center.

Information offices data set

The set contains a relation of the municipal information offices and their location. This data is currently collected by the local police call center.

Accommodations data set

The set contains a relation of the accommodations in the city and their location. This data is currently collected by the local police call center.

Educational institutions data set

The set contains a list of the schools, universities, and other educational institutions in the city, and their location, service hours and contact information (phone number, email address and/or web). This information is currently published through the Open Data portal of the city and is updated when changes occur.

Health services data set

Information on hospital, health centers and pharmacies is contained in this data set, together with their location, service hours and contact information (phone number, email address and/or web). This information is currently published through the Open Data portal of the city and is updated when changes occur.

Sports facilities and institutions data set

This set contains data on sport facilities and institutions, together with their location, service hours and contact information (phone number, email address and/or web). This information is currently published through the Open Data portal of the city and is updated when changes occur.

Municipal services data set

The data set lists services offered by the city council, the address of their headquarters, service hours and contact information. This data set is currently published in the Open Data portal of the city council and is updated when changes occur.

Non-municipal public administrations data set

This set lists services offered by other Public Administrations, rather than the city council. Details include their address, service hours and contact information. This data set is currently published in the Open Data portal of the city council and is updated when changes occur.

Means of transport, parking and gas station

Information on private and public services related to means of transport, such as parking places, gas stations, taxis, airports, and so on is collected in this data set. Available data includes address and contact information. This data set is currently published in the Open Data portal of the city council and is updated when changes occur.

Municipal marketplaces, commerce and consumer entities

The data set contains details on municipal marketplaces, and entities related to commerce and consumers, including address, service hours and contact information. This data set is currently published in the Open Data portal of the city council and is updated when changes occur.

List of pharmacies in Lleida

This set contains a list of pharmacies in Lleida, together with their location and contact information (phone number, email address and/or web). This information is currently published through the Open Data portal of the city and is updated when changes occur.

Catering businesses data set

The set contains a list of restaurants, bars and other catering businesses in the city of Lleida, and their location, service hours and contact information (phone number, email address and/or web). This information is collected by the Federation of Hospitality and Catering of Lleida, which updates it every three months, and stored by the city council of Lleida.

Hospitality businesses data set

The set contains a list of hotels, hostels and similar in the city of Lleida, and their location, service hours and contact information. This information is collected by the Federation of Hospitality and Catering of Lleida, which updates it every three months, and stored by the city council of Lleida.

Accessibility data set

The data set, provided by the Association of Paraplegic and Physically Disabled People of Lleida and stored and maintained by the city council of Lleida, contains a specification of whether the Hospitality and Catering businesses listed in the previous datasets are accessible or not.

2.3.1.2 Data service description

The REST API for the data service will provide the following resources:

Resource	Description
GET pois/all	Returns the complete list of points of interest
GET pois/coordinates	Returns a list of points of interest sorted according to its proximity to the provided coordinates
GET pois/type	Returns a list of points of interest in the provided category
GET pois/adapted	Return a list of adapted points of interest
GET pois/id	Return the complete information on the provided POI

The basic structure of the JSON responses provided through the API is shown below:

```
{
   "response":{
      "meta":{
         "v":"1.0",
         "status":"OK",
         "code":200,
         "timeRef":1340684854471,
         "msg":""
 },
      "data":{
         "startIndex":0,
         "itemsPerPage":1,
         "totalResults":1,
         "entry":[
            {
               "guid":”21EC2020-3AEA-1069-A2DD-08002B30309D",
               "type":"hospitality",
               "name":"Lleida Hotel",
               "address":"Major 12",
               "postcode":"25194",
               "coordinates":"41.656497,0.703125",
               "telephone":"973555555",
               "city":"Lleida",
               "url":"http://www.lleidahotel.com",
               "email":"info@lleidahotel.com",
               "schedule":"Monday to Sunday from 8:00 to 22:00",
               "adapted":"true"
            }
         ]
      }
   }
}
```


The document payload is divided into two main sections:

- `meta:` includes information on the current API version and on the return codes for the requested resource.
- `data:` collects the actual requested information. Together with the collection of data entries, the object also provides information for pagination.

2.3.2 Road incidents data set

This set contains detailed information of mobility incidents taking place in the city of Lleida, which is used by the local police to define and coordinate response actions. Information is collected in a call center which receives information from citizens, governmental entities and public and private organizations. The original data set contains information of any kind of incident and will be filtered for purposes of the project, in order to include only road incidents, road works and events information, which are relevant to vehicle and walking transit.

2.3.2.1 Data service description

The REST API for the data service will provide the following resources:

Resource	Description
GET roadincidents/all	Returns the complete list of active road incidents-

GET roadincidents/id	Returns the complete information on the provided road incident.
GET roadincidents/type	Returns the list of road incidents in the provided type.
GET roadincidents/coordinates	Returns a list of incidents sorted according to its proximity to the provided coordinates.
GET roadincidents/status	Returns a list of incidents in the provided status. Can be used for planning upcoming trips.

The JSON payload for these resources is described below:

```
   "response":{
      "meta":{
         "v":"1.0",
         "status":"OK",
         "code":200,
         "timeRef":1340684854471,
         "msg":""
 },
      "data":{
         "startIndex":0,
         "itemsPerPage":1,
         "totalResults":1,
         "entry":[
            {
               "guid":”21EC2020-3AEA-1069-A2ED-08022B30309D",
               "name":"Tall tram central c/Major",
               "type":"Obra"
               "description":"Per obres a la vorera, el tram central del
c/Major es troba tallat tant per automòbils com de les persones",
               "status":"actiu",
               "coordinates":"41.656497,0.703125",
               "begindate":123123123123,
               "enddate":123123343322                ]
            }
         ]
      }
   }
}
```

2.3.3 Municipal bus services data set

Information is provided by the city council and encompasses the existing bus lines and their routes, stops and geographical covered areas on Google Earth. Municipal bus information is currently published in the Open Data portal of the city of Lleida and updated when changes occur.

2.3.3.1 Data service description

The REST API for the data service will provide the following resources:

Resource	Description
GET publicbusline/all	Returns the complete list of public bus lines
GET publicbusline/id	Returns the complete information on the provided

	public bus line
GET publicbusline/stops/id	Returns the list of bus stops for the provided public bus line
GET publicbusline/stops/coordinates	Returns a list of bus stops sorted according to its proximity to the provided coordinates

The JSON payload for the resources returning information on the lines is as follows:

```
{
   "response":{
      "meta":{
         "v":"1.0",
         "status":"OK",
         "code":200,
         "timeRef":1340684854471,
         "msg":""
 },
      "data":{
         "startIndex":0,
         "itemsPerPage":1,
         "totalResults":1,
         "entry":[
            {
               "guid":”21EC2020-3AEA-1069-A2DD-08022B30309D",
               "name":"L12 Pl. Ajuntament – Av. Catalunya",
               "area":"Centre urbà",
               "route":"Plaça de l’Ajuntament – Rambla – Av. Catalunya”
            }
         ]
      }
   }
}
```

The payload for the resources providing information on the stops is as follows:

```
{
   "response":{
      "meta":{
         "v":"1.0",
         "status":"OK",
         "code":200,
         "timeRef":1340684854471,
         "msg":""
 },
      "data":{
         "startIndex":0,
         "itemsPerPage":1,
         "totalResults":1,
         "entry":[
            {
               "guid":”21EC2020-3AEA-1069-A2DD-08022B30309D",
               "name":"Pl. Ajuntament",
               "coordinates":"41.656497,0.703125",
               "lines": [
                   "L12",
                   "L24",
```

```
                    "L26"
                ]
            }
        ]
    }
}
```

2.3.4 Legal aspects of the data sets

The legal framework for publishing the data sets required for the Lleida Municipality pilots must be encompassed with the licensing approach implemented through the existing Lleida Open Data portal. Although these data sets provide information on road incidents coming from the emergency management system, any sensitive information will be removed from the data sources before its integration in the Open-DAI platform. In addition, the site is also compliant with the Spanish regulation on reuse of public information.

Again, no Service Level Agreement on data availability or system performance will be granted by Open-DAI, since it relies on the source services provided by the municipality and these ones do not provide any SLA.

According to those criteria, the data sets described in the previous sections will be provided both under Creative Commons CC by-3.0 license and Creative Commons CC BY-ND 3.0 license accordingly to the licenses in the source datasets.

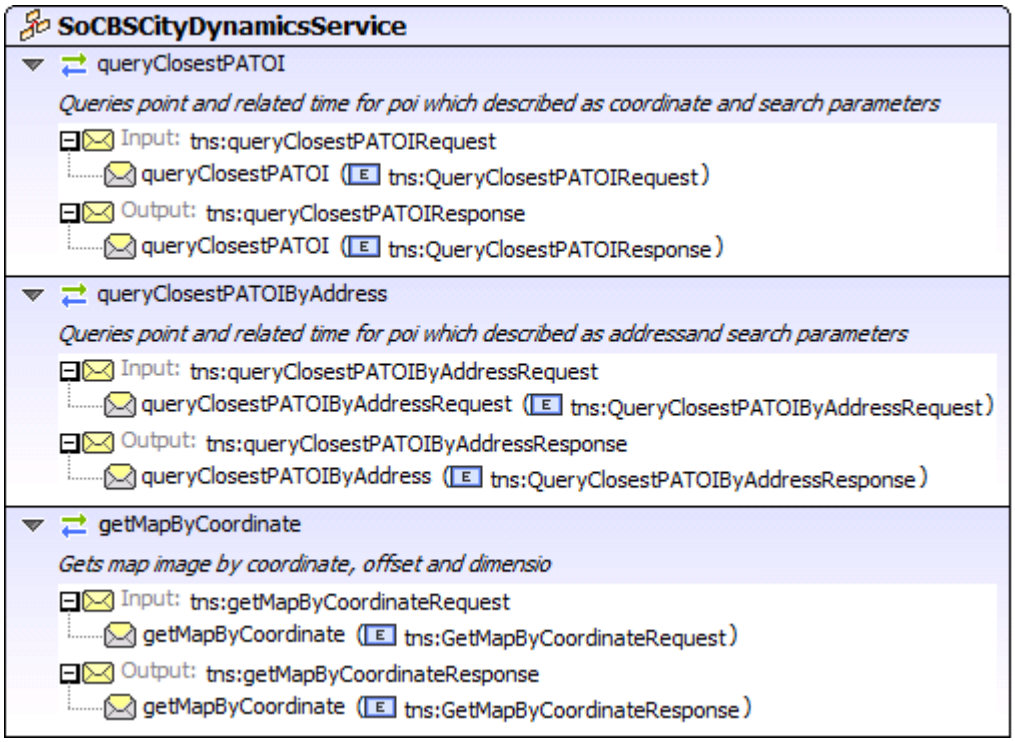
2.4 Ordu Municipality

2.4.1 City Dynamics

City dynamics data set includes kinds of POI like garbage cans, bus stops etc. and data services includes services for query this data.

2.4.1.1 Data service description

Web services definition for reach the dataset data.



type	**tns:SoCBSCityDynamicsService**
extensibility	<soap:binding style="document" transport="http://schemas.xmlsoap.org/soap/http"/>
operations	**queryClosestPATOI** extensibility <soap:operation soapAction="http://sampas.com/SoCBSCityDynamics/queryClosestPATOI"/> input <soap:body use="literal"/> output <soap:body use="literal"/> **queryClosestPATOIByAddress** extensibility <soap:operation soapAction="http://sampas.com/SoCBSCityDynamics/queryClosestPATOIByAddress"/> input <soap:body use="literal"/> output <soap:body use="literal"/> **getMapByCoordinate** extensibility <soap:operation soapAction="http://sampas.com/SoCBSCityDynamics/getMapByCoordinate"/> input <soap:body use="literal"/> output <soap:body use="literal"/>
source	<wsdl:binding name="SoCBSCityDynamicsServiceSOAP" type="tns:SoCBSCityDynamicsService"> <soap:binding style="document" transport="http://schemas.xmlsoap.org/soap/http"/> <wsdl:operation name="queryClosestPATOI"> <soap:operation soapAction="http://sampas.com/SoCBSCityDynamics/queryClosestPATOI"/> <wsdl:input> <soap:body use="literal"/> </wsdl:input> <wsdl:output>

```
      <soap:body use="literal"/>
    </wsdl:output>
  </wsdl:operation>
  <wsdl:operation name="queryClosestPATOIByAddress">
    <soap:operation
soapAction="http://sampas.com/SoCBSCityDynamics/queryClosestPATOIByAddress"/>
    <wsdl:input>
      <soap:body use="literal"/>
    </wsdl:input>
    <wsdl:output>
      <soap:body use="literal"/>
    </wsdl:output>
  </wsdl:operation>
  <wsdl:operation name="getMapByCoordinate">
    <soap:operation soapAction="http://sampas.com/SoCBSCityDynamics/getMapByCoordinate"/>
    <wsdl:input>
      <soap:body use="literal"/>
    </wsdl:input>
    <wsdl:output>
      <soap:body use="literal"/>
    </wsdl:output>
  </wsdl:operation>
```

operations

queryClosestPATOI

Queries point and related time for poi which described as coordinate and search parameters

input **tns:queryClosestPATOIRequest**

output **tns:queryClosestPATOIResponse**

queryClosestPATOIByAddress

Queries point and related time for poi which described as addressand search parameters

input **tns:queryClosestPATOIByAddressRequest**

output **tns:queryClosestPATOIByAddressResponse**

getMapByCoordinate

Gets map image by coordinate, offset and dimensio

input **tns:getMapByCoordinateRequest**

output **tns:getMapByCoordinateResponse**

source

```
<wsdl:portType name="SoCBSCityDynamicsService">
  <wsdl:operation name="queryClosestPATOI">
    <wsdl:documentation>Queries point and related time for poi which described as coordinate
and search parameters</wsdl:documentation>
    <wsdl:input message="tns:queryClosestPATOIRequest"/>
    <wsdl:output message="tns:queryClosestPATOIResponse"/>
  </wsdl:operation>
  <wsdl:operation name="queryClosestPATOIByAddress">
    <wsdl:documentation>Queries point and related time for poi which described as addressand
search parameters</wsdl:documentation>
```

	<wsdl:input message="tns:queryClosestPATOIByAddressRequest"/> <wsdl:output message="tns:queryClosestPATOIByAddressResponse"/> </wsdl:operation> <wsdl:operation name="getMapByCoordinate"> <wsdl:documentation>Gets map image by coordinate, offset and dimensio</wsdl:documentation> <wsdl:input message="tns:getMapByCoordinateRequest"/> <wsdl:output message="tns:getMapByCoordinateResponse"/> </wsdl:operation>

message **getMapByCoordinateRequest**

parts	**getMapByCoordinate** element **tns:GetMapByCoordinateRequest**
source	<wsdl:message name="getMapByCoordinateRequest"> <wsdl:part name="getMapByCoordinate" element="tns:GetMapByCoordinateRequest"/> </wsdl:message>

message **getMapByCoordinateResponse**

parts	**getMapByCoordinate** element **tns:GetMapByCoordinateResponse**
source	<wsdl:message name="getMapByCoordinateResponse"> <wsdl:part name="getMapByCoordinate" element="tns:GetMapByCoordinateResponse"/> </wsdl:message>

message **queryClosestPATOIByAddressRequest**

parts	**queryClosestPATOIByAddressRequest** element **tns:QueryClosestPATOIByAddressRequest**
source	<wsdl:message name="queryClosestPATOIByAddressRequest"> <wsdl:part name="queryClosestPATOIByAddressRequest" element="tns:QueryClosestPATOIByAddressRequest"> </wsdl:part> </wsdl:message>

message **queryClosestPATOIByAddressResponse**

parts	**queryClosestPATOIByAddress** element **tns:QueryClosestPATOIByAddressResponse**
used by	Operation **queryClosestPATOIByAddress** in PortType **SoCBSCityDynamicsService**
source	<wsdl:message name="queryClosestPATOIByAddressResponse"> <wsdl:part name="queryClosestPATOIByAddress" element="tns:QueryClosestPATOIByAddressResponse"> </wsdl:part> </wsdl:message>

message **queryClosestPATOIRequest**

parts	**queryClosestPATOI** element **tns:QueryClosestPATOIRequest**
source	<wsdl:message name="queryClosestPATOIRequest"> <wsdl:part name="queryClosestPATOI" element="tns:QueryClosestPATOIRequest"> </wsdl:part>

	</wsdl:message>

message **queryClosestPATOIResponse**

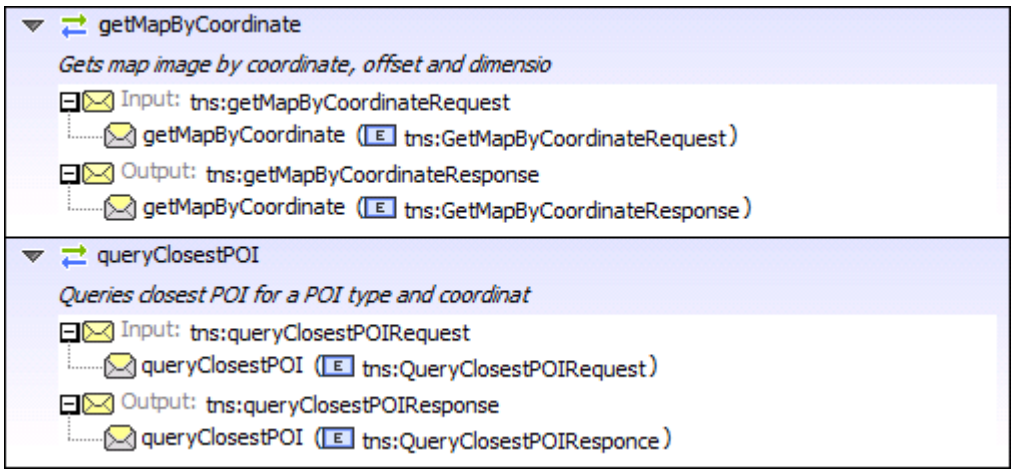
parts	**queryClosestPATOI** element **tns:QueryClosestPATOIResponse**
source	<wsdl:message name="queryClosestPATOIResponse"> <wsdl:part name="queryClosestPATOI" element="tns:QueryClosestPATOIResponse"> </wsdl:part> </wsdl:message>

2.4.2 POIs

POIs data set includes approximately 180 POI Types and POI data for related public administration.

2.4.2.1 Data service description

Web services definition for reach the dataset data.



operations	**getMapByCoordinate** extensibility <soap:operation soapAction="http://sampas.com/SoCBSCityDynamics/getMapByCoordinate"/> input <soap:body use="literal"/> output <soap:body use="literal"/> **queryClosestPOI** extensibility <soap:operation soapAction="http://sampas.com/SoCBSCityDynamics/queryClosestPOI"/> input <soap:body use="literal"/> output <soap:body use="literal"/>
source	<wsdl:binding name="SoCBSCityDynamicsServiceSOAP" type="tns:SoCBSCityDynamicsService">

```
  <soap:binding style="document" transport="http://schemas.xmlsoap.org/soap/http"/>
  <wsdl:operation name="queryClosestPATOI">
    <wsdl:operation name="getMapByCoordinate">
   <soap:operation
soapAction="http://sampas.com/SoCBSCityDynamics/getMapByCoordinate"/>
    <wsdl:input>
     <soap:body use="literal"/>
    </wsdl:input>
    <wsdl:output>
     <soap:body use="literal"/>
    </wsdl:output>
  </wsdl:operation>
  <wsdl:operation name="queryClosestPOI">
   <soap:operation soapAction="http://sampas.com/SoCBSCityDynamics/queryClosestPOI"/>
   <wsdl:input>
    <soap:body use="literal"/>
   </wsdl:input>
   <wsdl:output>
    <soap:body use="literal"/>
   </wsdl:output>
  </wsdl:operation>
```

operations

getMapByCoordinate

Gets map image by coordinate, offset and dimension

input **tns:getMapByCoordinateRequest**

output **tns:getMapByCoordinateResponse**

queryClosestPOI

Queries closest POI for a POI type and coordinat

input **tns:queryClosestPOIRequest**

output **tns:queryClosestPOIResponse**

source

```
<wsdl:portType name="SoCBSCityDynamicsService">
  <wsdl:operation name="getMapByCoordinate">
   <wsdl:documentation>Gets map image by coordinate, offset and
dimensio</wsdl:documentation>
   <wsdl:input message="tns:getMapByCoordinateRequest"/>
   <wsdl:output message="tns:getMapByCoordinateResponse"/>
  </wsdl:operation>
  <wsdl:operation name="queryClosestPOI">
   <wsdl:documentation>Queries closest POI for a POI type and
coordinat</wsdl:documentation>
   <wsdl:input message="tns:queryClosestPOIRequest"/>
   <wsdl:output message="tns:queryClosestPOIResponse"/>
  </wsdl:operation>
```

message **getMapByCoordinateRequest**

parts

getMapByCoordinate

element **tns:GetMapByCoordinateRequest**

source

```
<wsdl:message name="getMapByCoordinateRequest">
  <wsdl:part name="getMapByCoordinate" element="tns:GetMapByCoordinateRequest"/>
</wsdl:message>
```

message **getMapByCoordinateResponse**

parts	**getMapByCoordinate** element **tns:GetMapByCoordinateResponse**
source	<wsdl:message name="getMapByCoordinateResponse"> <wsdl:part name="getMapByCoordinate" element="tns:GetMapByCoordinateResponse"/> </wsdl:message>

message **queryClosestPOIRequest**

parts	**queryClosestPOI** element **tns:QueryClosestPOIRequest**
source	<wsdl:message name="queryClosestPOIRequest"> <wsdl:part name="queryClosestPOI" element="tns:QueryClosestPOIRequest"/> </wsdl:message>

message **queryClosestPOIResponse**

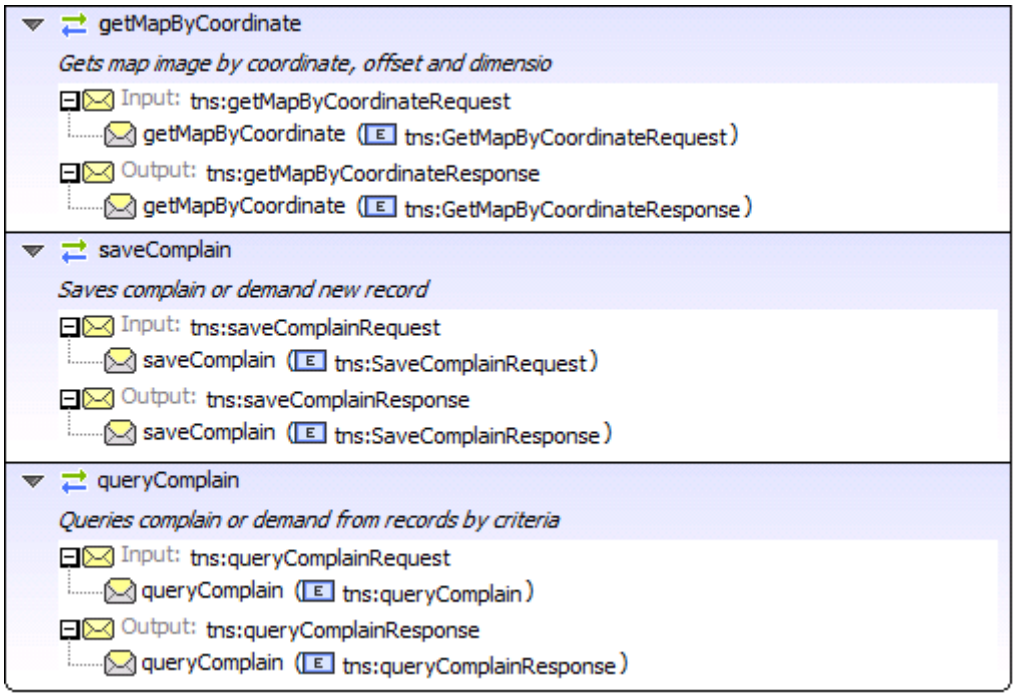
parts	**queryClosestPOI** element **tns:QueryClosestPOIResponce**
source	<wsdl:message name="queryClosestPOIResponse"> <wsdl:part name="queryClosestPOI" element="tns:QueryClosestPOIResponce"/> </wsdl:message>

2.4.3 Complains and Demands

Complains and Demands data set includes requests from citizen to public administration. And also data can be related by a geographic location.

2.4.3.1 Data service description

Web services definition for reach the dataset data.



operations

getMapByCoordinate

extensibility `<soap:operation soapAction="http://sampas.com/SoCBSCityDynamics/getMapByCoordinate"/>`

input `<soap:body use="literal"/>`

output `<soap:body use="literal"/>`

saveComplain

extensibility `<soap:operation soapAction="http://sampas.com/SoCBSCityDynamics/saveComplain"/>`

input `<soap:body use="literal"/>`

output `<soap:body use="literal"/>`

queryComplain

extensibility `<soap:operation soapAction="http://sampas.com/SoCBSCityDynamics/queryComplain"/>`

input `<soap:body use="literal"/>`

output `<soap:body use="literal"/>`

source

```
<wsdl:operation name="getMapByCoordinate">
  <soap:operation soapAction="http://sampas.com/SoCBSCityDynamics/getMapByCoordinate"/>
  <wsdl:input>
    <soap:body use="literal"/>
  </wsdl:input>
  <wsdl:output>
    <soap:body use="literal"/>
  </wsdl:output>
</wsdl:operation>
<wsdl:operation name="saveComplain">
  <soap:operation soapAction="http://sampas.com/SoCBSCityDynamics/saveComplain"/>
```

```
    <wsdl:input>
      <soap:body use="literal"/>
    </wsdl:input>
    <wsdl:output>
      <soap:body use="literal"/>
    </wsdl:output>
  </wsdl:operation>
  <wsdl:operation name="queryComplain">
    <soap:operation soapAction="http://sampas.com/SoCBSCityDynamics/queryComplain"/>
    <wsdl:input>
      <soap:body use="literal"/>
    </wsdl:input>
    <wsdl:output>
      <soap:body use="literal"/>
    </wsdl:output>
  </wsdl:operation>
</wsdl:binding>
```

message **getMapByCoordinateRequest**

parts	**getMapByCoordinate** element **tns:GetMapByCoordinateRequest**

source	`<wsdl:message name="getMapByCoordinateRequest">`
	`<wsdl:part name="getMapByCoordinate" element="tns:GetMapByCoordinateRequest"/>`
	`</wsdl:message>`

message **getMapByCoordinateResponse**

parts	**getMapByCoordinate** element **tns:GetMapByCoordinateResponse**

source	`<wsdl:message name="getMapByCoordinateResponse">`
	`<wsdl:part name="getMapByCoordinate" element="tns:GetMapByCoordinateResponse"/>`
	`</wsdl:message>`

message **queryComplainRequest**

parts	**queryComplain** element **tns:queryComplain**

source	`<wsdl:message name="queryComplainRequest">`
	`<wsdl:part name="queryComplain" element="tns:queryComplain"/>`
	`</wsdl:message>`

message **queryComplainResponse**

parts	**queryComplain** element **tns:queryComplainResponse**

source	`<wsdl:message name="queryComplainResponse">`
	`<wsdl:part name="queryComplain" element="tns:queryComplainResponse"/>`
	`</wsdl:message>`

message **saveComplainRequest**

parts	**saveComplain** element **tns:SaveComplainRequest**
source	<wsdl:message name="saveComplainRequest"> <wsdl:part name="saveComplain" element="tns:SaveComplainRequest"/> </wsdl:message>

message **saveComplainResponse**

parts	**saveComplain** element **tns:SaveComplainResponse**
source	<wsdl:message name="saveComplainResponse"> <wsdl:part name="saveComplain" element="tns:SaveComplainResponse"/> </wsdl:message>

2.5 Karlshamn Municipality

The Karlshamn pilot allows citizens and staff to connect reports and suggestions to equipment in the Municipality. Therefore two datasets will be published as services in the Open-DAI system: reports and points of interest. User credentials and rights management will be handled via third party systems.

2.5.1 Data sets containing points of interest

The following data sets contain information about different points of interest (POI) located in the Municipality of Karlshamn. Different categories of POIs are located in separate data sets, and are directly connected to the municipal GIS system that is used by city planners and maintenance workers. This ensures that the published data sets are always updated with the latest changes to equipment that is added, removed, replaced or moved in the Municipality.

The service for publishing POIs will provide JSON data, which is a modern protocol for sending information between applications on the web. This fits the choice of technologies for developing the Karlshamn pilot, and will also provide a well-known interface for other developers wishing to leverage the information. TEIID will handle the transformation of the data stream into the desired format, e.g. location information into geospatial coordinates.

2.5.1.1 Data sets

2.5.1.1.1 Lighting

Information about fixtures, such as model and manufacturer, light direction and type of lamp.

2.5.1.1.2 Ticket machines

Information about ticket machines (for parking), such as model, placement, parking zone and pricing.

2.5.1.1.3 Waste bins

Information about Municipality-owned waste bins, such as date of last maintenance, schedule for emptying and placement (e.g. indoors/outdoors).

2.5.1.1.4 Litterbins

Information about bins for dog litter, such as date of last maintenance, schedule for emptying and placement (e.g. on a light post).

2.5.1.2 Data service

The REST API for the data service will provide the following resources:

Method	Url	Comment
GET	/poi/#id	Fetch POI from unique id
GET	/poi/report/#report_id	Fetch POIs connected to report id

The JSON response will follow this basic structure:

```
{
  "response" : {
    "meta" : {
      "version" : "1.0",
      "status" : "OK",
      "code" : 200
    }
  },
  "data" : {
    "object_id" : 1,
    "title" : "Example Title",
    "LatLng" : {
      "Lat" : "String",
      "Lng" : "String"
    }
  }
}
```

2.5.1.3 Legal aspects of publishing the data sets

The data sets are fully owned and maintained by the municipality and contains no sensitive information. Therefore we foresee no legal aspects to apply to the publishing of the data. A licensing model for the published information has yet to be determined.

2.5.2 Data sets containing reports

The following data sets contain reports and suggestions created by citizens and staff. Reports are connected to POIs and assigned with statuses such as “new”, “reviewing”, “in progress”, “complete” and “cancelled”. Newly created reports will be assigned with the “new” status, and only municipality staff is given the authority to modify the status.

The service for publishing POIs will provide JSON data, which has been transformed from its original data stream with TEIID.

2.5.2.1 Data set

A report object contains information about classification, times of status changes (including creation), associated users and POIs and URL to a photo.

2.5.2.2 Data service

The REST API for the data service will provide the following resources:

Method	Url	Comment
GET	/report/#report_id	Fetch report from unique id
GET	/reports	Fetch all reports with limit
GET	/reports/status/#status_id	Fetch reports depending on status

The JSON response for a single report will follow this basic structure:

```
{
  "response" : {
    "meta" : {
      "version" : "1.0",
      "status" : "OK",
      "code" : 200
    }
  },
  "data" : {
    "id" : 1,
    "object" : {
      "object_id" : 1,
      "title" : "Example Title",
      "LatLng" : {
        "Lat" : "56.17030300",
        "Long" : "14.86307300"
      }
    },
    "title" : "Example Report",
    "description" : "Example Description",
    "img_url" : "Test",
    "types_id" : 1,
    "user" : {
      "id" : 1,
      "fb_firstname" : "Example Firstname",
      "fb_lastname" : "Example Lastname"
    },
    "added" : "2012-01-01 00:00:00",
    "updated" : "2012-01-01 00:00:00"
  }
}
```

The JSON response for a list of reports will follow this basic structure:

```
{
  "response" : {
    "meta" : {
      "version" : "1.0",
      "status" : "OK",
      "code" : 200
    }
  },
  "data" : [
    {
      "added" : "2012-01-01 00:00:00",
      "id" : 1,
      "object" : {
        "LatLng" : {
          "Lat" : "56.17030300",
          "Long" : "14.86307300"
        },
        "object_id" : 1,
        "title" : "Example Title"
      },
      "title" : "Example Report",
      "types_id" : 1,
      "updated" : "2012-01-01 00:00:00"
    },
    {
      "added" : "2012-01-01 00:00:00",
      "id" : 2,
      "object" : {
        "LatLng" : {
          "Lat" : "56.17030300",
          "Long" : "14.86307300"
        },
        "object_id" : 1,
        "title" : "Example Title"
      },
      "title" : "Example Report",
      "types_id" : 2,
      "updated" : "2012-01-01 00:00:00"
    }
  ]
}
```


2.5.2.3 Legal aspects of publishing the data set

The data set is fully owned and maintained by the municipality and contains no sensitive information. No personal information is stored within the data set, because user credentials are handled via third party systems. A licensing model for the published information has yet to be determined.

Any photos taken by a user are subject to Swedish copyright laws (1960:729, 1993:1212 and 1994:193). Therefore we will require users to accept a licencing agreement before publishing their photos.

3 Semantic Data Integration

From an architectural point of view, Open-DAI allows semantic data integration by means of D2RQ tool. D2RQ is able to expose data in RDF / SPARQL according to a common conceptual model and from diverse VDBs, owned whether by the same or by different project partners. Given the complexity and variety of the legacy databases, we decided to start experimenting semantic integration among the three VDBs of Regione Piemonte, then extending the scope, if the opportunity arises, to other partners' VDBs. The reason for this choice is that Piedmont datasets are very detailed, have more homogeneous domain compared to all others, and can be linked to other datasets released as Linked Open Data (LOD) by Regione Piemonte through the dati.piemonte.it portal.
Legacy datasets of Regione Piemonte, transformed into VDBs by JBoss Teiid, cover a quite coherent domain:

1. "Aria" dataset contains data on the measurement stations placed around the region that collect raw measures on air quality; these are information on the model of the station, which sensors are aboard and other management information.
2. "Incidenti" dataset includes information about the car accident and all the related data as the condition of the road at the moment of the accident, in which hospital the wounded people are brought, the condition of people involved in the accident, type of car, etc.
3. "Trasporti" dataset holds the timetables of public bus transport.

Therefore, domain shared by this datasets is geo-spatial and especially focused on urban environment. All datasets involve concepts such as points of interest (POIs), roads, populated places, transports and road events (both accidental and not).
Some standard vocabularies can be taken into account to represent classes and relationships in this domain. Conforming to the best practices on Linked Data publication, it is worth reusing existing vocabularies wherever possible, if they contain suitable concepts in your domain, rather than reinventing one. Reuse of existing terms is highly recommended because it increases the probability that data can be consumed by applications tuned to well-known vocabularies, without requiring further processing of the data. An important criterion for choosing a vocabulary is its widespread dissemination on the Web, especially in the LOD community. If the vocabulary is used by other authoritative LOD sets, most likely it will be helpful, because in re-use of vocabularies we achieve the benefits of Linked Data. The following table is a result of the study carried out by Free University of Berlin and DERI Institute, that provides statistics about the structure and content of the LOD cloud1. It lists the most widely used vocabularies and provides links to the data sources that use a specific vocabulary.

Vocabulary prefix	Vocabulary link	Number of usages in data sets	Data sets that use the vocabulary
dc	http://purl.org/dc/elements/1.1/	92 (31.19 %)	Data sets that use dc
foaf	http://xmlns.com/foaf/0.1/	81 (27.46 %)	Data sets that use foaf
skos	http://www.w3.org/2004/02/skos/core#	58 (19.66 %)	Data sets that use skos
geo	http://www.w3.org/2003/01/geo/wgs84_pos#	25 (8.47 %)	Data sets that use geo
xhtml	http://www.w3.org/1999/xhtml/vocab#	19 (6.44 %)	Data sets that use xhtml
akt	http://www.aktors.org/ontology/portal#	17 (5.76 %)	Data sets that use akt
bibo	http://purl.org/ontology/bibo/	14 (4.75 %)	Data sets that use bibo
mo	http://purl.org/ontology/mo/	13 (4.41 %)	Data sets that use mo

[1] URL: http://www4.wiwiss.fu-berlin.de/lodcloud/state/#terms. Una lista più completa più essere trovata all'indirizzo: http://stats.lod2.eu/vocabularies

vcard	http://www.w3.org/2006/vcard/ns#	10 (3.39 %)	Data sets that use vcard
sioc	http://rdfs.org/sioc/ns#	10 (3.39 %)	Data sets that use sioc
cc	http://creativecommons.org/ns#	8 (2.71 %)	Data sets that use cc
GeoNames	http://www.GeoNames.org/ontology#	6 (2.03 %)	Data sets that use GeoNames
frbr	http://purl.org/vocab/frbr/core#	6 (2.03 %)	Data sets that use frbr
xsd	http://www.w3.org/2001/XMLSchema#	6 (2.03 %)	Data sets that use xsd
time	http://www.w3.org/2006/time#	5 (1.69 %)	Data sets that use time
event	http://purl.org/NET/c4dm/event.owl#	5 (1.69 %)	Data sets that use event
dbpedia	http://dbpedia.org/resource/	5 (1.69 %)	Data sets that use dbpedia
gr	http://purl.org/goodrelations/v1#	4 (1.36 %)	Data sets that use gr
dbo	http://dbpedia.org/ontology/	4 (1.36 %)	Data sets that use dbo
ore	http://www.openarchives.org/ore/terms/	4 (1.36 %)	Data sets that use ore
bio	http://purl.org/vocab/bio/0.1/	4 (1.36 %)	Data sets that use bio
dbp	http://dbpedia.org/property/	4 (1.36 %)	Data sets that use dbp
tag	http://www.holygoat.co.uk/owl/redwood/0.1/tags/	3 (1.02 %)	Data sets that use tag
void	http://rdfs.org/ns/void#	3 (1.02 %)	Data sets that use void
scovo	http://purl.org/NET/scovo#	3 (1.02 %)	Data sets that use scovo
http	http://www.w3.org/2006/http#	3 (1.02 %)	Data sets that use http
uniprot	http://purl.uniprot.org/core/	3 (1.02 %)	Data sets that use uniprot
umbel	http://umbel.org/umbel#	3 (1.02 %)	Data sets that use umbel
rev	http://purl.org/stuff/rev#	3 (1.02 %)	Data sets that use rev
qb	http://purl.org/linked-data/cube#	2 (0.68 %)	Data sets that use qb
geospecies	http://rdf.geospecies.org/ont/geospecies#	2 (0.68 %)	Data sets that use geospecies
sdmx	http://purl.org/linked-data/sdmx#	2 (0.68 %)	Data sets that use sdmx
sawsdl	http://www.w3.org/ns/sawsdl#	2 (0.68 %)	Data sets that use sawsdl
org	http://www.w3.org/ns/org#	2 (0.68 %)	Data sets that use org
vann	http://purl.org/vocab/vann/	2 (0.68 %)	Data sets that use vann
admingeo	http://data.ordnancesurvey.co.uk/ontology/admingeo/	2 (0.68 %)	Data sets that use admingeo
wdrs	http://www.w3.org/2007/05/powder-s#	2 (0.68 %)	Data sets that use wdrs

Table 1 - LOD Cloud Vocabularies Stats

According to the “Best Practices for Publishing Linked Data”2 drafted by the W3C Government Linked Data Working Group, other selection criteria in addiction to diffusion are the following:

1) Vocabularies must be documented. This includes the liberal use of labels and comments; tags to language used. Human-readable pages must be provided by the publisher describe the classes and properties, preferably with use cases defined.
2) Vocabularies should be self-descriptive. Each property or term in a vocabulary should have a Label, Definition and Comment defined. Self-describing data suggests that information about the encodings used for each representation is provided explicitly within the representation. The ability for Linked Data to describe itself, to place itself in context, contributes to the usefulness of the underlying data.
3) Vocabularies should be described in more than one language. Multilingualism should be supported by the vocabulary, i.e., all the elements of the vocabulary should have labels, definitions and comments available in the government's official language, e.g., Spanish, and at least in English. That is also very important as the documentation should be clear enough with appropriate tag for the language used for the comments or labels.
4) Vocabularies should be accessible for a long period. The vocabulary selected should provide some guarantee of maintenance over a specified period, ideally indefinitely.
5) Vocabularies should be published by a trusted group or organization. Although anyone can create a vocabulary, it is always better to check if it is one person, group or authoritative organization that is responsible for publishing and maintaining the vocabulary.
6) Vocabularies should have persistent URLs. Persistent access to the server hosting the vocabulary, facilitating reusability is necessary.
7) Vocabularies should provide a versioning policy. The publisher ideally will address compatibility of versions over time. Major changes to the vocabularies should be reflected on the documentation.

Taken into account the popularity of a schema and all criteria listed above, the following metadata vocabularies can be used to represent the domain of the Piedmont datasets:

1. **DC (Dublin Core).** It is suitable to describe metadata about the vocabulary itself as a document (Title, Creator, Subject, Description, Date, Contributor, Format, Language, etc). Dublin Core allows the description of a wide variety of resources in different formats, being general enough to include any indication of semantic contents. Given its simplicity, Dublin Core is currently widely used and several standards and projects refer to it or specify a mapping on it.
2. **Geo (Geo Ontology)**. Geo is an RDF vocabulary that uses the concepts of latitude and longitude to represent spatial information according to WGS84 standard (World Geodetic System 1984). The vocabulary defines a class 'Point', which can be described using the 'lat', 'long' and 'alt' properties. The 'lat' and 'long' properties take literal (ie. textual values), each in decimal degrees. The 'alt' property is decimal meters about local reference ellipsoid.
3. **CC (Creative Commons).** CC can be used to describe the license under which Open-DAI Conceptual Model (ODCM) is released. The Creative Commons Rights Expression Language (CC REL) lets you describe copyright licenses in RDF. Main classes of the schema are: Work, License, Jurisdiction, Permission, Requirement, Prohibition, Reproduction, Distribution, Derivative Works, Sharing, etc. Main properties are : permits, requires, prohibits, jurisdiction, legalcode, deprecatedOn, license, morePermissions, attributionName, attributionURL, useGuidelines.
4. **Time (Time Ontology).** This is essential to represent anything that happens over time and has a duration. Time is an ontology of temporal concepts, which provides terms for expressing facts about topological relations among instants and intervals, together with information about durations and dates. Main classes are: TemporalEntity, Instant, Interval, ProperInterval. Main object properties are: before, after, hasBeginning, hasEnd, inside. Datatype properties for

[2] URL: https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html

describing duration includes concepts such as: years, months, weeks, days, hours, minutes and seconds.

5. **Event (Event Ontology).** In ODCM Event can be used to represent, for instance, road traffic accidents. This ontology is centered around the notion of event, as the way by which cognitive agents classify arbitrary time/space regions. Event reuses in turn Time and Geo to contextualize the event in space-time. Classes are: Event, Factor and Product. Main properties are: agent, factor, hasProduct, hasSubEvent, place, time.
6. **DBO (DBpedia Ontology)** and **DBP (DBpedia Properties).** DBO is the DBpedia conceptual model, in which all types of resources present in the ontology are defined. DBP, instead, is a list of all object properties and datatype properties within DBpedia, along with their sub-properties. DBpedia is obviously not specific for the urban space domain, as it contains concepts in almost every domain. Rather than directly use classes and properties of DBO and DBP, may be useful to link them to ODCM's own classes and properties, in order to further clarify their semantics, relying on encyclopedic definitions. This can be done, for example, for classes such as City, Lake Road, Bridge, Car, Traffic Light, etc.

Some notes on using **GeoNames Ontology**. GeoNames Ontology presents peculiar Knowledge Representation (KR) choices that make it difficult to adapt to the description of a particular domain. Its classes are:

- Class: A class of features.
- Code: A feature code.
- Feature: A geographical object uniquely defined by its GeoNames id.
- Map: A Web page displaying a map.
- RDFData: A Document containing RDF description of one or several features.
- WikipediaArticle: A Wikipedia article.

This is clearly a not intuitive classification of the geographical/spatial domain. Entities that common sense would recognize as classes (Populated Place, Park, Port, Road, etc.) in GeoNames are individuals of the class "Code". Unfortunately this makes the wealth of GeoNames Codes quite impossible to reuse as a vocabulary for ODCM. Also object properties must be reused with caution, because they frequently set as domain or range the class "Feature", which obviously makes sense only in GeoNames context. These features discourage the reuse of GeoNames vocabulary in describing our own resources.
And now some positive notes on using LinkedGeoData Ontology. Unlike GeoNames Ontology, LinkedGeoData Ontology has an easier and more reusable representation model. Most popular places of the urban landscape (squares, gymnasiums, swimming pools, parks, cinemas, etc.), including Points Of Interest (POIs), are represented as classes with labels in all major languages. LinkedGeoData uses the comprehensive OpenStreetMap data collection to create a large spatial KB consisting of more than 1 billion nodes and 100 million ways. These data are interlinked with DBpedia and GeoNames. Since data are the same contained in OpenStreetMap, they can be easily viewed through "slippy maps" implemented by OpenLayers. A slippy map represents in a layered and dynamic fashion users' favourite POIs. For these reasons, it seems clear that classifying our POIs and other resources with LinkedGeoData Ontology concepts, and sharing its vocabulary, is a more suitable choice than using GeoNames Ontology.
Although these remain the basic tools for knowledge representation in the Government Linked Data field, recently many community's efforts were aimed at a more accurate definition of best practices for the PSI. For these purposes the Government Linked Data Working Group was born in 2011. First the GLDWG has drafted a set of best practices in this area — guidelines designed to make it easier for the public sector to make more of its data available as 5 star linked data. Secondly, it was chartered to standardize or, where necessary, create vocabularies that help to increase interoperability of data sets. Alongside this work, three 'Core Vocabularies' is now close to completion:

- **People** — a vocabulary for describing a natural person (things like name, family name, date and place of birth etc.);
- **Business** — a vocabulary for describing legal entities as found in company registers;

- **Location** — geographical locations and addresses, noting the importance of interoperability with the EU's INSPIRE Directive on spatial information.

The release of SEMIC vocabularies[3] in May 2012, especially the Location Core Vocabulary, is certainly noteworthy for our purposes. They are explicitly dedicated to semantic interoperability within the Public Administration domain. Semantic interoperability in PSI is the ability of information systems to exchange public data and to enable the sharing of public information and knowledge. It aims at the mental representations that citizens have of the meaning of any given data. These vocabularies are defined as "core" for the PSI domain, because they aim to provide the most basic and less ambiguous concepts to facilitate information representation and exchange. So mapping to or extending such Core Vocabularies are the minimum required to guarantee a level of cross-domain and cross-border interoperability that can be attained by Public Administrations.

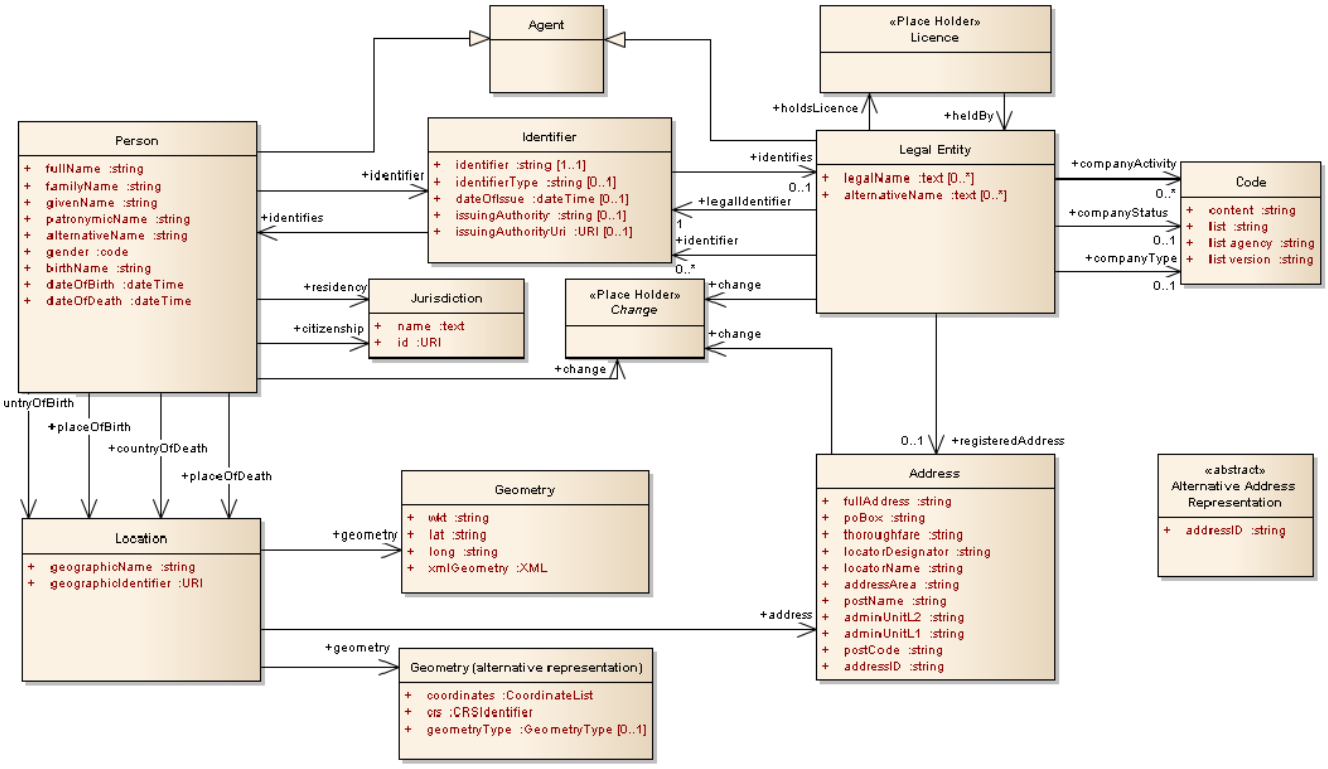


Figure 1 - ISA Core Vocabularies Diagram

The SEMIC Core Vocabularies import several well-known metadata schemata, such as Dublin Core, FOAF, Schema.org, Creative Commons, SKOS, Geo, SIOC, etc. The conceptual model we're most interested in is Location Core, which focuses on the concepts of Location, Geometry and Address.

1) **Location**. Location Core allows you to describe a place in three different ways: by using a place name, an address or a geometry. These three forms of representation may be more or less appropriate in different contexts and for different needs. Place name is the most popular and easy way to indicate a place, but can be ambiguous since there are often several places that share the same name. So Location Core defines a property that allows a Location to be defined by a URI, such as a GeoNames or DBpedia URI. The INSPIRE Data Specification on

[3] SEMIC Core Vocabularies may be found on Joinup (joinup.ec.europa.eu), the platform hosted by the European Commission to support and promote metadata management. SEMIC is a project funded by the ISA Programme (ec.europa.eu/isa).

Geographical Names4 provides a detailed model for describing a named place: the concept of a geographic name used in Location Core is consistent with INPIRE specifications.

2) **Address**. The representation of addresses can be a very prickly problem, not only because rules vary from country to country, but also because in a single country these rules are not always observed, and the end result is usually just a mess. The relevant properties of Location Core for an address are:
   - PO Box (a specialization of locator designator)
   - Thoroughfare (a road, a waterway etc.)
   - Locator designator (a building number, entrance number etc.)
   - Locator name (a proper name for a building or room within a building)
   - Address area (usually a city area or village)
   - Locality (usually a town)
   - Admin unit level 2 (usually a county or state)
   - Admin unit level 1 (almost always a country)
   - Post Code

3) Following this field structure, an address description can be INSPIRE-compliant. However, if you want to put an address into a single string, you can use the following properties:
   - full address (the complete address as a formatted string)
   - addressID (a unique identifier for the address)

   The addressID is part of the INSPIRE guidelines and provides a hook that can be used to link the address to an alternative representation, such as vCard or OASIS xAL.

4) Geometry. The Geometry Class represents a place on the basis of its coordinates, and can be encoded in different formats including WKT, GML, KML, RDF+WKT/GML (GeoSPARQL), RDF (WGS84 lat/long, schema.org) and GeoHash URI references. It has the following properties:
   - coordinates (which gives the coordinate list);
   - crs (an identifier for the coordinate reference system) ;
   - the geometry type (point, line or polygon).

According to the features described above, arguably SEMIC Core Location Vocabulary is the most suitable choice for representing the spatial domain in ODCM. Together with Time Ontology, Event Ontology and DBO, it can definitely provide the main structure for metadata representation in our domain of interest. With regard to the mapping between our resources and other Linked Open Data, we saw that LinkedGeoData Ontology is the most attractive option and is preferable over GeoNames.

4 General legal consideration

General European data protection and data security issues related to cloud computing have been singled out already in:

- recent European Commission Communications;
- documents adopted by the Article 29 Data Protection Working Party;
- ENISA report: Cloud Computing Risk Assessment

Since the project is processing data within a cloud environment we will try to summarize the most relevant for the present analysis.

Restriction on applicability of the Directive 95/46/EC (Article 13(1))
Pursuant to Article 13(1) of Directive 95/46/EC, Member States may restrict the application of certain provisions of Directive 95/46/EC for matters of national and public security or the prosecution and

[4] URL: http://inspire.ec.europa.eu/index.cfm/pageid/2

prevention of crime.11 Thus, depending on local law in a Member State, in certain circumstances some data that municipalities handle may not be subject to all of the regulations under Directive 95/46/EC.

Data Controller – Data Processor (Directive 95/46/EC, Articles 2(d) and (e))
It is necessary to identify the controller, the processor, and their interactions in order to determine 'who is responsible for compliance with data protection rules, how data subjects can exercise their rights, which is the applicable national law and how effective Data Protection Authorities can operate. Directive 95/46/EC clearly distinguishes between controller and processor.
The controller is the individual or entity that determines the purposes of and means for processing of personal data.
The processor is the individual or entity that processes personal data on behalf of the controller. However, applying such a definition to the cloud computing environment is quite challenging. At first glance, one might conclude that the PA is the controller and the Cloud Service Provider the processor.13 Nevertheless, Cloud Service Providers often determine the means and sometimes also the purposes of the processing – thus falling within the definition of controller. To address this issue and provide some guidance on Article 29, the Data Protection Working Party issued an opinion on 16 February 2010 in which it adopted a viewpoint on interpreting such definitions in complex environments. However, the opinion did not shed much light on the specifics of the cloud computing environment, for which the roles of controller and processor still need to be determined on a case-by-case basis and in relation to the nature of the cloud services.
In the case of the Open-DAI project we designed a role for the processor that is totally passive in the sense that it merely hosts the computing resources that publishes the data and data is not really hosted in the cloud.

Prior checking (Directive 95/46/EC, Article 20)
Pursuant to Article 20 and depending on national law, prior checking may be necessary for the processing. This depends on the type of service and types of data being processed.

Appropriate technical and organizational measures (Article 17): data integrity, identity management, and access control
Data integrity and availability are essential elements in the provision of cloud computing services. According to Directive 95/46/EC, the controller and its processors must implement technical and organizational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorized disclosure or access; having regard to the state of the art and the cost of their implementation, such measures must ensure a level of security appropriate to the risks represented by the processing and the nature of the data to be protected (article 17). The problem is that the concept of appropriate has been interpreted in different ways throughout EU Member States. Thus, although Cloud Service Providers quite often implement widely recognized technical standards (e.g., ISO 27001) to secure customer data, these may not match perfectly to national requirements for appropriate measures. Further consistency and harmonization across the EU is required. In addition, the high level of data security requested of a Cloud Service Provider in an e-health scenario is worth noting with special regard to identity management and access control.
In the case of the Open-DAI project the security and access control apply to the access to the platform deployments and general security, since there is no direct access to the data there is no data integrity problem within the model

Data transfer to countries outside the EEA (Articles 25-26)
Cloud models entail that customer information and data may involve the transfer of data by the Cloud Service Provider from one data-center in the EEA to another that can be located anywhere in the world. However, Directive 95/46/EC prohibits transfers of personal data from the EEA to countries which do not ensure an adequate level of protection within the meaning of article 25 – unless the data subject has previously given unambiguous consent to the proposed transfer or other procedures are in place in accordance with article 26 (eg, 'Model Contracts for the transfer of personal data to third countries', 'Safe Harbor Principles' (where the data is being transferred to the United States), or 'Binding Corporate Rules')15. There are challenges with each of these ways to legitimize a transfer, however: basing it on

the consent of the data subject exposes the transfer to the uncertainties of possible withdrawals of that consent; the Safe Harbor Principles, which apply to data transferred to the United States. may fall short in a cloud environment, where data flows may concern non-EEA countries other than the United States; and Binding Corporate Rules have yet to be fully endorsed by large Cloud Service Provider.
Even in this case the project does not transfer data but it can be considered as an extension of the datacenter of the PA subject.

Service Level Agreements
At the stage of defining the present data assessment and specification document, no uniform decision has been taken about the provision of any Service Level Agreement on data or services availability or system performance. As a rule, no SLA will be provided and services will formally be available on a “best effort basis”. However, this issue will be re-examined considering the user-driven feedbacks that will be collected withing WP6 and the exploitation goals defined within WP8.