

Exposing public health surveillance data using existing standards.

Clément Turbelin, Pierre-Yves Boëlle

► **To cite this version:**

Clément Turbelin, Pierre-Yves Boëlle. Exposing public health surveillance data using existing standards.. Studies in Health Technology and Informatics, IOS Press, 2013, 192, pp.802-6. inserm-00869435

HAL Id: inserm-00869435

<https://www.hal.inserm.fr/inserm-00869435>

Submitted on 3 Oct 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exposing Public Health Surveillance Data Using Existing Standards

Clément Turbelin^{a,b}, Pierre-Yves Boëlle^{a,c}

^aUPMC Univ Paris 06, UMR-S U707, F-75012, Paris, France

^bINSERM, U 707, F-75012, Paris, France

^cAP-HP, Hôpital Saint-Antoine, Paris, F-75012, France

Abstract

With the growing use of information technologies, an increased volume of data is produced in Public Health Surveillance, enabling utilization of new data sources and analysis methods. Public health and research will benefit from the use of data standards promoting harmonization and data description through metadata. No data standard has yet been universally accepted for exchanging public health data. In this work, we implemented two existing standards eligible to expose public health data: Statistical Data and Metadata Exchange – Health Domain (SDMX-HD) proposed by the World Health Organization and Open Data Protocol (OData) proposed by Microsoft Corp. SDMX-HD promotes harmonization through controlled vocabulary and predefined data structure suitable for public health but requires important investment, while OData, a generic purpose standard, proposes a simple way to expose data with minimal documentation and end-user integration tools. The two solutions were implemented and are publicly available at <http://sdmx.sentiweb.fr> and <http://odata.sentiweb.fr>. These solutions show that data sharing and interoperability are already possible in Public Health Surveillance.

Keywords:

Public Health; Data Standards.

Introduction

Public Health Surveillance (PHS), as the ongoing, systematic data collection, management, analysis and interpretation of data followed by the dissemination of these data to public health programs to stimulate public health action [1], is now using an increasing number of data sources thanks to information technologies expansion: generalization of database management system and arising of web technologies have made data more available and useable. Electronic health record systems through emergency departments [2] or general practitioners provide such data, as well as Internet search queries [3], or social networks [4], internet robots are used to produce worldwide epidemiological maps [5]. From these data sources, epidemiologists and researchers can compute synthetic indicators to assess the burden of a disease or alert about an outbreak. All these data will be referred to as “public health data” in the following text.

To enable innovative use of data for PHS, for example the integration of different kind of indicators on the same disease, data sharing in public health must be put forward as it was in other domains [6, 7]. Molecular biology (“-omics” domain) set forth a major example by driving data sharing with the Gene Expression Omnibus [8]. Moreover, giving access to data is

increasingly required as a step in scientific articles publication.

From this perspective, data needs first to be organized, represented and documented. Metadata – data about data – should document the source data. Some specialized standards already exist to support this goal, for example, the “minimal datasets” formats designed to standardize data from biological experiments. They are listed and described by the Minimum Information for Biological and Biomedical Investigations project which encourages harmonizing and reutilizing of this type of standards [9]. These kinds of standards should help interoperability, at least at the syntactic level in a given domain.

There is no universal standard to expose public health data, we identified two potential standards and described their implementation to expose public health surveillance data in a standard way over the Internet.

Materials and Methods

Method

The objective of this study was to expose public health data from the French *Sentinelles* network (FSN) using existing standards.

Potential standards were identified from a literature review on Medline including research about data standards and data exchange protocols used in public health or statistical domains, surveillance and public health information systems. It was complemented by an Internet research about data standards and exchange protocols used in science and information technology (Google Scholar, Google). Selected standards were required to have public specifications and licenses, and capabilities to handle aggregated data.

Nine standards were identified. Three were not adapted to our type of data (Data Documentation Initiative, mostly used to describe individual-based datasets), HL7 Clinical Document Architecture [10] and related standards like EHR-Public Health (EHR-PH) from the Public Health Data Standards Consortium (PHDSC) [11] were designed to handle health-care data exchange; as the Clinical Data Interoperability Standard Consortium for clinical trial description [12]. Data Access Protocol [13] and netCDF [14] and were more designed to describe physical storage for large datasets with limited metadata. One standard had a limited purpose (Google Data Protocol, used to access Google Services). We ultimately took three remained standards into account: Open Data Protocol (OData) proposed by Microsoft Corp. [15], Google Dataset Publishing Language (GDPL) [16] and Statistical Data and Metadata Exchange – Health Domain (SDMX-HD) [17] proposed by the World Health Organization. GDPL was mostly a standard for describing existing datasets and had similar

purpose than SDMX-HD in a simpler way. OData and SDMX-HD were chosen at the end of our selection process.

Source data

Data were provided by the French *Sentinelles* Network (FSN) [18], a general practitioners-based surveillance network collecting data (number of visits to the GPs for 10 conditions), using a web-based interface and a desktop ad-hoc software [19]. Produced data were a weekly estimation computed from the collected data aggregated at several geographic levels (national and regional level). For each monitored condition, available data were: weekly incidence (estimated value with upper and lower bound of 95% confidence incidence (95%CI)), weekly incidence rate (estimated value and 95%CI lower and upper value). Some conditions could also have a weekly threshold value at the national level, used to determine epidemic periods.

Standards

Statistical Data and Metadata Exchange Health-Domain

SDMX-HD is a specialized implementation of the SDMX technical standard offering solutions to exchange data and metadata in statistical domains. It is based on an information model (SDMX-IM) including entities designed to structure and describe other data. SDMX-IM first specifies first two types of “metadata”: *structural metadata* describing how data are organized and *reference metadata*, describing the context of a data element with another data. As an implementation of ISO 11179 metadata registry, each data element is identified by a conceptual domain (referring to a *Concept*) and a value domain (referring to a data format like text or numeric; or a list of acceptable values called *Codelist* for enumerated data values). Therefore, SDMX-IM identifies entities as *Concept*, *ConceptScheme* (regrouping *Concepts*).

Concept in SDMX-IM, is defined by a unique identifier of the concept (for example “OBS_VALUE”, “CURRENCY”) and a textual description of the concept (“An observation value”, “A currency name”). A *Codelist* is a set of acceptable code (used as the stored value in data document) associated with a textual label (for user presentation).

Data and metadata are exchanged in datasets and metadata sets documents (for example in XML files). The arrangement of these documents is described in *Structure Definition* documents (like a schema describing a database structure). Two types of data structures are taken into account by SDMX-HD: *Compact Dataset* and *Generic Metadata set*.

A *Compact Dataset* includes time series: a set of measures associated with a time step. Each measure is described by a context giving the signification of the measure (what, where, when). Such context is identified by a set of data elements with different roles: a *Measure* (including the actual value for the time step, for example the incidence value), *Dimension* (uniquely identifying an observation within the dataset: for example time, geographic location, type of disease) and *Attribute* (other measure-related contextual data, for example a comment about data completion of missingness). To specify data elements, SDMX-HD associates a *Concept* with a role and a data representation (data format or *Codelist*): for example, a time series has one element, combining the concept “TIME_PERIOD” with the role of Dimension and the concept “OBS_VALUE” with the role of Measure (usually using numeric data type). In the *Compact Dataset*, data elements are organized in a nested structure identifying several levels: *Dataset*, *Group*, *Series*, and *Observation*. A measure is contained

at the *Observation* level. Other levels are useful for grouping observations sharing the same context.

The set of roles and related concepts used in a given instance of *Compact Dataset* is described in a *Data Structure Definition* document. SDMX-HD thus specifies a template of DSD for Compact Dataset, including a minimal set of Concepts used for exchanging public health data (introduced in the results).

Metadata are stored in dedicated structures. Within SDMX, a metadata is the association of a data element value (a given code of a *Codelist*) with a set of attributes. Each attribute is related to a Concept and data format. A given type of data elements (for all codes of a given *Codelist*) are associated with a specific metadata structure (with its own set of metadata attributes). These structures are described in a *Metadata Structure Definition* document.

Open Data standard

OData is a technical protocol, proposed by Microsoft Corp. (Redmond, WA), based on well-known web existing standards (HTTP, AtomPub) suitable for building a standard API to expose a database. Three types of documents are accessible through this protocol: a *Metadata Document* (describing the available data structures), a *Service Document* (listing available datasets) and *EntitySet* documents providing the datasets. This protocol is available through a REST-based web service, the URL itself specifying how a document is accessed. For example, adding “\$metadata” to the service URL is the standard way for accessing the metadata document. Adding the name of the dataset to the service URL is how the dataset is retrieved.

The *Metadata document* describes the structure of the database using the Conceptual Schema Description Language (CSDL, based on XML) as an object-oriented data model. It defines each data structure as an *EntityType* (acting as a class) containing properties (strongly typed using primitive types – integer, string,..., complex types (set of primitive typed properties) or referring to another *EntityType* –acting as a foreign key).

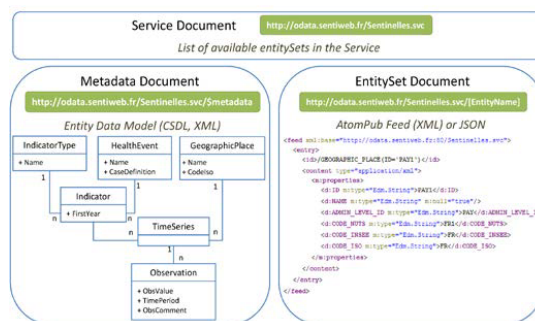


Figure 1 - OData service overview and OData Entity Data Model used for the FSN's data

EntityType could use inheritance to extend from another *EntityType* (within the current schema or from an external metadata document). EntitySets that contain data are then defined as an instance of *EntityType* (thus available datasets in an OData system are the EntitySets).

The *Service Document*, accessed at the base service URL, provides the list of available *EntitySets* and their URL. The *EntitySet* documents are a AtomPub feed embedding representation of each instance of *EntitySet*, i.e., the actual data of

the service. Figure 1 shows an overview of available documents within an OData-based service.

Results

SDMX-HD-based solution

Following SDMX-HD standard, French Sentinelles network's data were described by a set of *Dimensions* which uniquely identifies each measure in relation to a Concept:

- GEOGRAPHIC_PLACE_NAME (location),
- BOUNDARIES (estimation, lower or upper bound),
- AGROUP (Age group)
- INDICATOR (described below)
- TIME_PERIOD (time step of the measure).

Two others dimensions were defined by SDMX-HD but hold the same information for all data: VALUE_TYPE (to distinguish observation from target value), ORGANIZATION (organization providing data). The data element for the measure was related to the concept 'OBS_VALUE'.

These data elements were also related to a data format, and to a Codelist where needed. An example of Compact DataSet document is illustrated in Figure 2. In this example, Influenza-like Illness incidences and incidence rates (using the INDICATOR coded values respectively '3I' and '3TI' and the BOUNDARIES code '2' – estimation point) are given for the national level (GEOGRAPHIC_PLACE_NAME coded value "FR") for the weeks 2012-W09, 2012-W11.

```
<?xml version="1.0" encoding="UTF-8" ?>
<CompactData>
  <DataSet CALENDAR_TYPE="ISO8601" ORGANIZATION="SENTIWEB"
    FREQ="W" AGROUP="ALL" VALUE_TYPE="1">
    <Series INDICATOR="3I" GEOGRAPHIC_PLACE_NAME="FR" BOUNDARIES="2">
      <Obs TIME_PERIOD="2012-W11" OBS_VALUE="128945"/>
      <Obs TIME_PERIOD="2012-W10" OBS_VALUE="174704"/>
      <Obs TIME_PERIOD="2012-W09" OBS_VALUE="225648"/>
    </Series>
    <Series INDICATOR="3TI" GEOGRAPHIC_PLACE_NAME="FR" BOUNDARIES="2">
      <Obs TIME_PERIOD="2012-W11" OBS_VALUE="203"/>
      <Obs TIME_PERIOD="2012-W10" OBS_VALUE="275"/>
      <Obs TIME_PERIOD="2012-W09" OBS_VALUE="355"/>
    </Series>
  </DataSet>
</CompactData>
```

Figure 2 - Example of a Compact DataSet using XML syntax

SDMX-HD provided the Concept of 'INDICATOR' related to the Health Indicator Conceptual Framework (ISO DIS 21667) and defined as "A single summary measure, [...], that represents a key dimension of health status, the health care system, or related factors.". Since each surveillance produces several kinds of estimations (incidence, incidence rate ...), an 'INDICATOR' was consequently the combination of one surveillance ('Chicken pox surveillance' for example) and a type of outcome ('Incidence rate').

Description of each indicator was provided by a *meta dataset* document: each 'INDICATOR' code was related to a set of metadata attributes (each identified by a Concept): 'FDATA' specifying the first year of surveillance, 'HEALTH_EVENT' indicating the code of the surveillance, and 'INDICATOR_TYPE' indicating the type of estimation. In the same way, each 'HEALTH_EVENT' code was related to a set of metadata attributes, including for example the case definition used in a given surveillance.

OData-based solution

OData does not differentiate data and metadata structures. Nevertheless, the final model included two entities dedicated to access to FSN's data: the 'Observation' Entity embedded each data value (in an 'ObsValue' property) for a given time

step ('TimePeriod' property). Observations were then associated to a 'TimeSeries' property, holding other context than time (Geographic location, type of measure). The data model is illustrated in Figure 1.

Other entities related to the 'TimeSeries' entity aimed at describing data dimensions and provided useful information to help understand our data: an 'Indicator' entity describing the type of measure (related to a 'HealthEvent' entity, with the case definition, and an 'IndicatorType' entity, corresponding to the type of outcome, for example 'Incidence' or 'Incidence rate').

Technical implementation

Each solution was implemented as a public website providing direct access to data and metadata. Specific features were also added to help understand the model or the technical solution: <http://sdmx.sentiweb.fr> and <http://odata.sentiweb.fr>.

The SDMX-HD-based website provides access to all SDMX-HD XML documents (datasets, metadatasets, structure definitions) using an "SDMX explorer" in which XML files are converted to HTML pages through XSL transformation stylesheet. JavaScript is then used to enable user interaction. An example of the output for the DSD document is shown in Figure 3. As defined by SDMX-HD standard, the set of XML files is available in a zip container.

The OData website provides an endpoint to the OData service exposing FSN's data. Data are available in a XML document and can be viewed as HTML pages with an explorer using XSL stylesheet conversion of XML documents.

ConceptRef	ConceptSchemeRef	CodeList	TextFormat
VALUE_TYPE [1.0]	CS_DISAGGREGATION [SDMX-HD]	CL_VALUE_TYPE [SDMX-HD]	[1.0]
FREQ [1.0]	CS_COMMON [SDMX-HD]	CL_FREQ [SDMX-HD]	[1.0]
BOUNDARIES [1.0]	CS_DISAGGREGATION [SDMX-HD]	CL_BOUNDS [SDMX-HD]	[1.0]
AGROUP [1.0]	CS_DISAGGREGATION [SDMX-HD]	CL_AGROUP [SENTIWEB]	[1.0]
GEOGRAPHIC_PLACE_NAME [1.0]	CS_COMMON [SDMX-HD]	CL_GEOGRAPHIC_PLACE_NAME [SDMX-HD]	[1.0]
INDICATOR [1.0]	CS_COMMON [SDMX-HD]	CL_INDICATOR [SENTIWEB]	[1.0]
ORGANIZATION [1.0]	CS_COMMON [SDMX-HD]	CL_ORGANIZATION [SENTIWEB]	[1.0]
Time Dimensions			
TIME_PERIOD [1.0]	CS_COMMON [SDMX-HD]		
PrimaryMeasure			
OBS_VALUE [1.0]	CS_COMMON [SDMX-HD]		Decimal

Figure 3 - Compact Dataset Dimensions showed by SDMX-HD explorer website

To ensure usability of this solution, we also provided an example usage page: data are downloaded via the OData web service and then displayed with a Google visualization widget. This page was created with Javascript (using OData.js client library) and could be hosted by third party: <http://websenti.u707.jussieu.fr/cturbelin/odata-example>. A screenshot of this usage example is shown in Figure 4.

The same development feature is available at the the SDMX website.

Solution Comparisons

SDMX-HD proposed a predefined set of concepts and structures to hold health related data that was fully reused to represent our data using the Compact Dataset data structure. Two

concepts were added for utilization in metadata to document the INDICATOR Concept usage and Geographical levels (combining geographical codes into French and European coding systems).

OData did not provide any predefined model to represent such data; the model was therefore built from scratch inspired by our database. We try to provide a model as generic as possible, especially for structures handling our data (keeping in mind extensibility and reusability). The final model included 6 *EntityTypes* and is shown in Figure 1. Entity Data model may be shared through an inheritance mechanism (an external model could be used and extended) providing a decentralized way for sharing structures.

OData protocol was implemented in common languages (php, java, .Net, ruby, objective-c, javascript) as a client or server library enabling easy development of producer or consumer applications. An Excel© plugin (PowerPivot) was successfully used to access data via our OData endpoint.

SDMX (based on SDMX-HD) defined technical specifications for a registry service that enables sharing Concepts, Codelists and structure definitions. SDMX-HD tools were available in java and .Net, and most of the tools are authoring tools or registry services. Some tools are available for end-user utilization (for example to access data in an office or statistical software).

Discussion

Data opening is an emerging challenge in all areas of science [6, 20-23], including Public Health [7, 24] [25]. We described

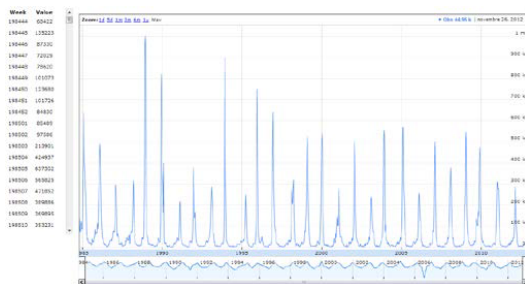


Figure 4 - Usage example page showing visualization of FSN's data from OData service

here how public health surveillance data can be opened via existing standards. This solution was preferred to the development of a dedicated API (Application Programming Interface) to prepare for interoperability of PHS systems.

This work presented a proof-of-concept by considering two different standards: SDMX-HD from statistical domain expertise proposed by WHO for exchange of public health data, and OData a general purpose data exchange protocol from information technology, now provided as an OASIS standard, international open standards consortium. Interestingly, a Public Health Data Consortium was also established with the goal of making recommendations for standards in public health data exchange. However, its recommendations are so far focused on systems already implementing HL7 and IHE (Integrating HealthCare Enterprise), so that health care data can be used for public health purposes. It was therefore not an option in our surveillance system, where EHR systems may not be compliant with HL7.

In the first part of this work, we represented our data using each solution, which implicitly led to assess their expressivity. SDMX-HD, by using controlled vocabulary (Concept) and allowing complex data and metadata structure (possibly hierarchical) has good expressivity. Nevertheless, expressivity of such a system could probably not be compared to a knowledge model since relationships between concepts are not supported. OData is quite simpler, and is limited to a simple relational model, like a database schema. Both standards included a way to add custom annotations to the data model and the data, which could enhance expressivity (and documentation) at the expense of interoperability.

Interoperability was not the primary objective of this work but it is clear that SDMX-HD solution, by providing a set of predefined concepts and data structure results in better interoperability, at least at the syntactic level. SDMX Registry, that enables sharing of Concepts, Codelists and data structures across organizations, could really improve public health data standardization. As a proof-of-concept, WHO is building a worldwide "Indicator and Measurement", based on SDMX-HD, designed to be a centralized resource for metadata and public health data [26].

OData solution is not developed with interoperability in mind but with a uniform data access solution to expose database through a generic API. This standard is more focused on data access protocol and physical representation than data semantics. It could be possible, using annotation and/or by extending produced XML document to add semantics information, but this will not be a part of the actual standard.

From an end-user perspective (here, public health professionals, epidemiologists or researchers), we assessed usability of each solution by searching for available tools facilitating the use of data from each standard (integration in every day office tools, statistical software, data visualization). In this respect, OData was a suitable solution for an easy end-user usage thanks to the number of available libraries and end-user tools. SDMX-HD does not provide a ready-to-use solution to explore, visualize and easily import data into common tools (like spreadsheets or statistical softwares).

Sustainability is another issue to consider for the adoption of a standard. OData, is developed by a major actor in information systems, and has been used by a large community. SDMX-HD originated from a major actor in public health (WHO) but has not been widely adopted. Therefore, uncertainties remain today regarding the right choice to make.

Study Limitations

The data used for this work was described by a minimal set of metadata ensuring self-comprehensiveness. This choice was led by information already available on our public website and within our database. As we did not attempt to add more metadata, our assessment of what could be expressed with the standards was limited.

The two solutions were built in parallel and it is clear that the OData model was widely inspired from the SDMX-HD solution. The choice of organizing data in 'TimeSeries' entities was indeed lifted from the Compact Dataset definition. This structure was finally a way to streamline the service by offering a single way to access all data, regardless of the indicator and geographic level. To help understand and harmonize the implemented solutions, we used names in the OData model close-related to the concepts identifiers used within SDMX-HD. This work was done with interoperability in mind, but it was not a primary goal for this step. In this respect, SDMX-

HD solution seems to have greater capabilities to achieve interoperability, with the use of controlled vocabulary and pre-defined data structure. Furthermore, SDMX is currently moving forward to web semantics technology (<http://publishing-statistical-data.googlecode.com>) enabling for example, the use of knowledge model (SKOS) to improve expressivity.

Conclusion

This work is a first step for exposing public health surveillance data. Implementation of SDMX-HD and Open Data solutions emphasizes the need for a well-defined data model to support interoperability as well as the need for an end-user perspective to promote utilization of such standards.

Acknowledgments

We would like to thank all general practitioners from the French Sentinelles Network. Authors state having no conflict of interest.

References

- [1] Thacker SB, Qualters JR, Lee LM. Public health surveillance in the United States: evolution and challenges. *MMWR Surveill Summ.* 2012; Jul 27;61 Suppl: 3-9.
- [2] Tsui FC, Espino JU, Dato VM, Gesteland PH, Hutman J, Wagner MM. Technical description of RODS: a real-time public health surveillance system. *J Am Med Inform Assoc.* 2003; Sep-Oct;10(5): 399-408.
- [3] Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. *Clin Infect Dis.* 2008; Dec 1;47(11): 1443-8.
- [4] St Louis C, Zorlu G. Can Twitter predict disease outbreaks? *BMJ.* 2012; 344:e2353.
- [5] Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. *N Engl J Med.* 2009; May 21;360(21): 2153-5, 7.
- [6] Hanson B, Sugden A, Alberts B. Making data maximally available. *Science.* 2011; Feb 11;331(6018): 649.
- [7] Walport M, Brest P. Sharing research data to improve public health. *Lancet.* 2011; Feb 12;377(9765): 537-9.
- [8] Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muerterer RN, Holko M, Ayanbule O, Yefanov A, Soboleva A. NCBI GEO: archive for functional genomics data sets--10 years on. *Nucleic Acids Res.* 2011; Jan;39(Database issue): D1005-10.
- [9] Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, Ashburner M, Ball CA, Binz PA, Bogue M, Booth T, Brazma A, Brinkman RR, Michael Clark A, Deutsch EW, Fiehn O, Fostel J, Ghazal P, Gibson F, Gray T, Grimes G, Hancock JM, Hardy NW, Hermjakob H, Julian RK, Jr., Kane M, Kettner C, Kinsinger C, Kolker E, Kuiper M, Le Novere N, Leebens-Mack J, Lewis SE, Lord P, Mallon AM, Marthandan N, Masuya H, McNally R, Mehrle A, Morrison N, Orchard S, Quackenbush J, Reecy JM, Robertson DG, Rocca-Serra P, Rodriguez H, Rosenfelder H, Santoyo-Lopez J, Scheuermann RH, Schober D, Smith B, Snape J, Stoekert CJ, Jr., Tipton K, Sterk P, Untergasser A, Vandesompele J, Wiemann S. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat Biotechnol.* 2008; Aug;26(8): 889-96.
- [10] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, Shabo Shvo A. HL7 Clinical Document Architecture, Release 2. *J Am Med Inform Assoc.* 2006; Jan-Feb;13(1): 30-9.
- [11] Orlova AO, Dunnagan M, Finitzo T, Higgins M, Watkins T, Tien A, Beales S. Electronic health record - public health (EHR-PH) system prototype for interoperability in 21st century healthcare systems. *AMIA Annu Symp Proc.* 2005: 575-9.
- [12] CDISC. Clinical Data Interchange Standards Consortium. [cited 2012 09/08/2012]; Available from: <http://www.cdisc.org>.
- [13] Gallagher J, Potter N, Sgouros T, Hankin S, Flierl G. The Data Access Protocol — DAP 2.0. NASA; 2007; Available from: <http://www.opendap.org/pdf/ESE-RFC-004v1.2.pdf>.
- [14] Unidata. Network Common Data Form (netCDF). [11/09/2011]; Available from: <http://www.unidata.ucar.edu/software/netcdf/>.
- [15] Open Data Protocol. Microsoft; 2012; Available from: <http://www.odata.org/>.
- [16] Google. DSPL: Dataset Publishing Language. 2012 [29 nov 2012]; Available from: <https://developers.google.com/public-data/>.
- [17] SDMX-HD (Health Domain). Available from: <http://www.sdmx-hd.org/>.
- [18] Flahault A, Blanchon T, Dorleans Y, Toubiana L, Vibert JF, Valleron AJ. Virtual surveillance of communicable diseases: a 20-year experience in France. *Stat Methods Med Res.* 2006; Oct;15(5): 413-21.
- [19] Turbelin C, Boelle PY. Improving general practice based epidemiologic surveillance using desktop clients: the French Sentinel Network experience. *Stud Health Technol Inform.* 2010; 160(Pt 1): 442-6.
- [20] Reichman OJ, Jones MB, Schildhauer MP. Challenges and opportunities of open data in ecology. *Science.* 2011; Feb 11;331(6018): 703-5.
- [21] Dealing with data. Challenges and opportunities. Introduction. *Science.* 2011; Feb 11;331(6018): 692-3.
- [22] Overpeck JT, Meehl GA, Bony S, Easterling DR. Climate data challenges in the 21st century. *Science.* 2011; Feb 11;331(6018): 700-2.
- [23] Lang T. Advancing global health research through digital technology and sharing data. *Science.* 2011; Feb 11;331(6018): 714-7.
- [24] Sankoh O, Ijsselmuiden C. Sharing research data to improve public health: a perspective from the global south. *Lancet.* 2011; Jul 30;378(9789): 401-2.
- [25] Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. *Bull World Health Organ.* 2010; Jun;88(6): 462-6.
- [26] Whitaker P. WHO Indicator and Metadata Registry (IMR), SDMX-HD Aggregate Data Exchange 2009. Available from: http://sdmx.org/wp-content/uploads/2009/10/sdmx_who_imr.pdf.

Address for correspondence

Clément Turbelin, turbelin@u707.jussieu.fr
Réseau Sentinelles, UMR S 707, Faculté de médecine St Antoine
27 rue Chaligny, 75571 PARIS CEDEX 12