The Humanitarian eXchange Language: Coordinating Disaster Response with Semantic Web Technologies

Editor(s): Christophe Guéret, Data Archiving and Networked Services (DANS), KNAW; Stephane Boyera, SBC4D; Mike Powell, IKM Emergent; Martin Murillo, Data Connectivity Initiative – IEEE
Solicited review(s): Christophe Guéret, Data Archiving and Networked Services (DANS), KNAW; Martin Murillo, Data Connectivity Initiative – IEEE; Stefan Boyera, SBC4D; Mike Powell, IKM Emergent; Louiqa Raschid, University of Maryland, U.S.A.

Carsten Keßler [a] and Chad Hendrix [b,*]
[a] *CARSI, Department of Geography, Hunter College – City University of New York, USA*
E-mail: carsten.kessler@hunter.cuny.edu
[b] *United Nations Office for the Coordination of Humanitarian Affairs, Geneva, Switzerland*
E-mail: hendrix@un.org

Abstract. The Humanitarian eXchange Language (HXL) is a project by the United Nations Office for the Coordination of Humanitarian Affairs that aims at refining data management and exchange for disaster response. Data exchange in this field, which often has to deal with chaotic environments heavily affected by an emergency such as a natural disaster or an armed conflict, still happens mostly manually. The goal of HXL is to contribute to the automatization of many of these processes, saving valuable time for staff in the field and improving the information flow for decision makers who have to allocate resources for response activities. This paper gives an overview of this initiative, which is set to significantly improve information exchange in the humanitarian domain. We introduce the HXL vocabulary, which provides a formal definition of the terminology used in this domain, and an initial set of tools and services that produce and consume HXL data. The HXL system infrastructure is introduced, along with its data management principles. The paper concludes with an outlook on the future of HXL and its role in the humanitarian ecosystem.

Keywords: Disaster Management, Humanitarian Aid, Linked Open Data, Vocabulary, Data Management, Tools

1. Introduction

Events such as large-scale natural disasters or armed conflicts often affect major parts of the local population. If these events exceed the capacity of a government to fully respond, support from the international community is required to address the affected population's need for shelter, food, water, sanitation, and medical care. The Office for the Coordination of Humanitarian Affairs[1] (OCHA) is the United Nations' division responsible for the orchestration of response actions to events such as the 2010 Haiti earthquake and the recent armed conflict in Mali.[2] A specific challenge for OCHA—in addition to the often confusing situation in the affected regions—is the large number of organizations that need to be coordinated to address the affected population's needs. These organizations range from large institutions such as the International Red

*Any opinions expressed herein are those of the authors and do not represent official positions of the United Nations Office for the Coordination of Humanitarian Affairs.

[1] See http://unocha.org.
[2] These are only two of the 25 humanitarian operations in which OCHA is currently working.

Cross and Red Crescent Movement to other UN divisions such as the UN Refugee organization (UNHCR) and national governmental bodies down to small, local non-governmental organizations (NGOs) with a handful of employees. The database of OCHA's financial tracking service[3] currently lists over 5500 organizations that participate in consolidated appeals to donors. Each of these organizations use a different system to handle the data about their activities, ranging from full-fledged enterprise information management systems to relational databases to simple spreadsheets. In a large-scale disaster, hundreds of these organizations need to be coordinated; in the aftermath of the Haiti earthquake, response data from an estimated 600 organizations was collected by OCHA, which needs to be compiled into a common operational picture. Although this did not represent the total humanitarian activity in country at the time, it did represent a massive challenge in data reconciliation.

Collecting and integrating data to optimize the response efforts in such a heterogeneous and distributed environment is challenging and still leads to situations where OCHA's information management officers on site integrate data by manually copying data from one spreadsheet into another. Widely varying transliterations of place names and differences in units of measurement for humanitarian interventions add further difficulties to the task of compiling a common operational picture. For the aid recipients, often desperate to fulfill their most basic needs, such manual data handling leads to significant delays in identifying and addressing needs. The improved data handling that HXL strives to enable facilitates faster and more informed decisions to provide water, food, medication, and shelter, where it is most needed. Unified means to exchange data between the different organizations working in the field means that those enacted decisions can also be implemented faster. As an example, an ambulance unit can directly locate a group of people in need for medical help based on the incoming report. Moreover, improved data handling frees up time to invest in other tasks for information management officers, who are always working under extreme time pressure in the onset of an emergency. These aspects all contribute to faster and more targeted response to the needs of populations in need.

While the current situation in terms of data exchange is extremely inhomogeneous, it is also very unlikely that all involved organizations can be convinced to use a common information system. This is due to the differences both in size and topical focus of the different organizations involved. Therefore, the idea of a common exchange format for humanitarian data was brought up within OCHA in 2011. The rationale behind the exchange format was that it would allow each organization to retain their established data management practices, but still facilitate data sharing with OCHA and other collaborators. At this point, this exchange format was supposed to be developed as an XML schema and therefore named the *Humanitarian eXchange Language* (HXL; [ˈhɛksl̩]). HXL was first discussed in public in a breakout session at the International Crisis Mappers Conference in November 2011, where several participants suggested a Semantic Web approach to tackle this large-scale data integration problem.

This paper reports on the current state of the Humanitarian eXchange Language and describes what has been achieved since fall 2011. A schematic overview of the HXL infrastructure and the different components this paper reports on is shown in Figure 1. In the next section, we first refer to relevant related work and point out how HXL differs from these existing solutions that address a similar problem space. The requirements for the development of HXL are laid out in Section 3 to show why we chose an approach based on Semantic Web technologies. Section 4 introduces the current state of the HXL vocabulary and the underlying design principles. We also discuss the development and iteration process we follow in the ongoing extension of the vocabulary. In Sections 5 and 6, we address the different means of producing HXL data and scenarios for its consumption in applications, respectively. Data management principles and work flows are discussed in Section 7, followed by an overview of the surrounding infrastructure in Section 8. Section 9 concludes the paper and gives an outlook on the future of the Humanitarian eXchange Language.

2. Related Work

Data exchange between different actors has been a challenge both for OCHA and the humanitarian field as a whole for a long time. OCHA and other organizations have tried to address different aspects of this problem by developing multiple siloed data collection and management systems, often requiring cut-and-paste operations to feed these systems with data

[3] See http://fts.unocha.org/.

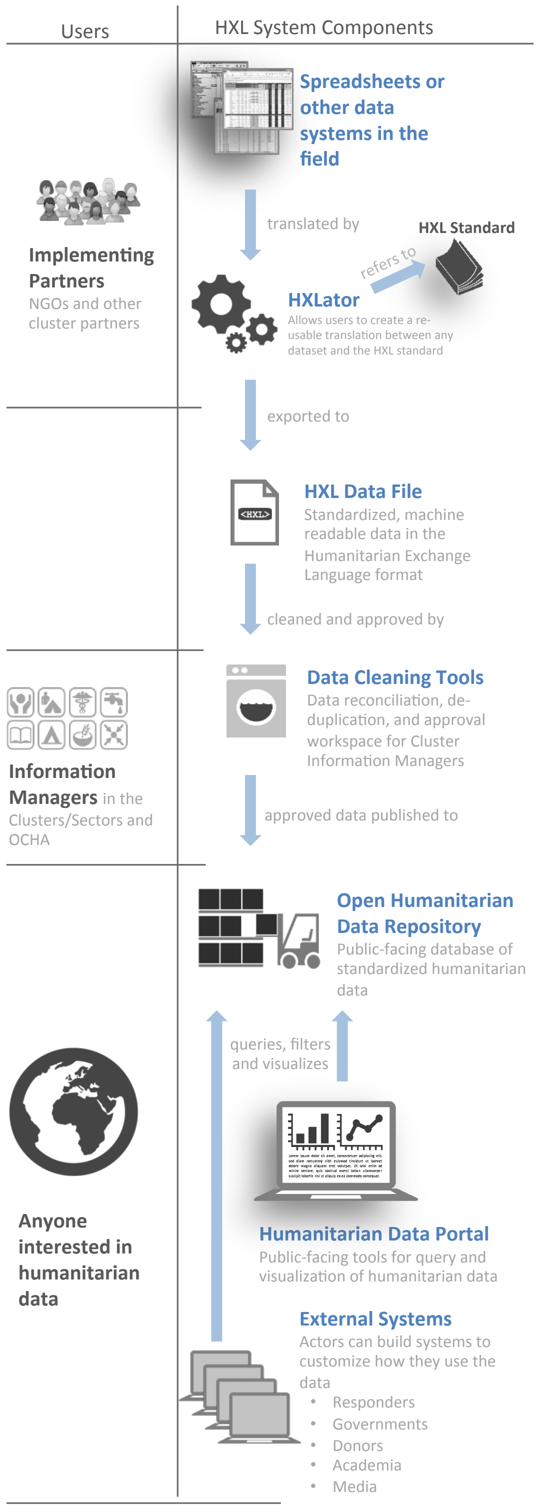

Fig. 1. High-level overview of the HXL system components.

from the actors' own systems. Examples include the Common Request Format (CRF) [35] to streamline information requests and the Multi-Cluster/Sector Initial Rapid Assessment (MIRA) framework [22], which serves as a standardized way for the initial assessment of humanitarian needs at the onset of a disaster. Another example is the Single Reporting Format, which provided a comprehensive system for collecting humanitarian activity data in Pakistan, but met with limited adoption in the field. In some cases, the humanitarian community has been successful at defining standards for describing some humanitarian data. The Inter-Agency Standing Committee (a forum of organizations involved in humanitarian response) established guidelines describing data regarding the attributes and size of populations affected by a crisis [21].

Most of the current information systems hold data that serve very specific needs. Their contents are only combined to generate the periodic reports for a variety of audiences. These reports are largely generated manually by staff members that are highly familiar with the situation on-site and the corresponding data. In addition to the data kept in specialized information systems, some data—such as geographic and demographic information—is required across OCHA and in other cooperating organizations. Such data is organized by country in the Common Operational Datasets (CODs) as downloadable files on a website.[4] OCHA is responsible for identifying and updating these datasets, which ideally form the baseline data for organizations responding to a humanitarian crisis. Updates are agreed in advance by the humanitarian community in-country to make sure that all involved organizations always refer to the same version of the CODs.

Outside of OCHA, several different standards for data exchange in the humanitarian domain have been developed over the past years. The most widely used is the Emergency Data Exchange Language (EDXL),[5] a collection of XML-based messaging standards initiated by the US Department of Homeland Security and developed by the Organization for the Advancement of Structured Information Standards (OASIS). EDXL consists of different components that specify how to exchange data about distributions, resources, hospital availability, situation reporting, and tracking of emergency patients. As the name suggests, EDXL has been designed to speed up the direct communi-

[4]See http://cod.humanitarianresponse.info/.
[5]See http://docs.oasis-open.org/emergency/.

cation between the different actors in an immediate emergency—e.g., when an ambulance quickly needs to find a hospital nearby that can accept a patient. For such purposes, EDXL provides standards to request resources (such as an ambulance), ask for hospital availability, and track emergency patients. HXL, in contrast, focuses on standardizing and streamlining data reporting in long-term humanitarian operations. It provides a method to track information such as affected population counts by source over the course of weeks, months, or even years.

The International Aid Transparency Initiative[6] (IATI) is an effort that does not target the operational side of humanitarian response, but fosters transparency of the spendings for development aid. As the transition from disaster response to development aid is often fluent and many organizations work in both domains, there is an overlap between the resources covered in the IATI registry[7] and those addressed in HXL. Cross-referencing both data sources would hence be a valuable next step, which is already being discussed between both groups. Having the IATI data as Linked Open Data—which has already been proposed [11]—would facilitate this step; so far, the data published in the IATI XML format are available through an implementation of the CKAN API.[8]

An opportunity arises for the development of an ontology—"an explicit specification of a conceptualization" [16, p.199]—that structures the domain and enables semantic interoperability [3] between the different systems in use.While the use of Semantic Web technology for disaster management has been discussed in the literature [7], the proposal for *Crowdsourced Linked Open Data* [34] is the only actual application to date. The authors show how information about the Haiti earthquake collected on the Ushahidi platform [28] can be organized and made available online, using the Management Of A Crisis (MOAC) vocabulary developed for this purpose.[9] Within W3C, synergies between these different efforts are being discussed in the recently established Emergency Information Community Group.[10]

[6]See http://www.aidtransparency.net
[7]See http://www.iatiregistry.org.
[8]See http://docs.ckan.org/en/latest/api.html.
[9]See http://observedchange.com/moac/ns/.
[10]See http://www.w3.org/community/emergency/.

3. Requirements for HXL

This section outlines the peculiarities of data exchange in the humanitarian domain, followed by a discussion of the requirements that were identified for the development of HXL.

3.1. Data Exchange in the Humanitarian Ecosystem

Humanitarian data exists at many scales, from global overviews down to highly granular operational data. An example of the former might be the number of people affected by flooding globally in 2012; an example of the latter might be the number of families provided with sanitation supplies on a given day in a given refugee camp by a given organization. HXL is primarily concerned with this granular, operational-level data. Whereas the more aggregated, global level data benefits from a luxury of time in which to produce and validate it, the operational data has no such luxury. To be useful to humanitarian actors, the data must be compiled, reconciled, validated, analyzed and disseminated within hours or days. Changes, additions, and updates are relentless and take place in a high pressure, high stakes environment. In the onset of an emergency, information about affected populations, destroyed infrastructure, and required response actions changes rapidly and can vary widely, depending on the respective source. These different numbers need to be tracked, verified, and compared against each other into a common operational picture. As an additional complication, in large scale disasters many of the responding organizations will be national actors or small local NGOs who have never had contact with the international humanitarian system. This system consists of different governmental, inter-governmental (such as different UN agencies), and non-governmental (such as the International Red Cross and Red Crescent Movement) organizations. Smaller NGOs will likely have not had exposure to or training on existing information management tools and standards. Instead, they will have their own established, or perhaps ad hoc, information systems. It is this multitude of systems that drives the complexity and duplication of humanitarian information during a crisis response.

The demand for humanitarian data comes from many levels. Small local actors trying to plan their response activities for the coming days may need the number of children to vaccinate, for example. On the other end of the spectrum, donor governments trying to marshal financial or other resources for the response need aggregate data to make informed decisions. A responding organization may find requests for its operational data (or some aggregate of it) coming from its national headquarters, from the national government, from OCHA, from the media, and from one or more major donors. These requests generally do not share a common format. This "reporting burden" is a key reason why the creation of additional reporting systems meets with limited adoption. In conclusion, all organizations participating in response activities both require data to do their work, and produce data that is reported back to coordinating bodies.

The telecommunications environment of large-scale disasters is also a key element to be considered in the development of any information management systems. Until recently, these humanitarian operations often took place in regions that had very poor telecommunications infrastructure even in normal times. During disasters, the reliability of electricity and communications is uncertain. In particular, the lack of reliable Internet connectivity has been a major constraint in the development of humanitarian information systems. However, in the last several years, the increasing coverage of mobile networks as well as the availability of satellite-based Internet connectivity has lessened this constraint and opened a door for the development of Internet-based systems for sharing humanitarian operations data.

The international humanitarian system is organized into "clusters" such as Water and Sanitation, Health, Shelter, and Education, among others. Members of these clusters are humanitarian organizations that are responding in one (or more) of these thematic areas. Data from each organization flows to the "cluster-lead" organization, which has the responsibility for compiling that data into a common operational picture for that cluster. However, because of the myriad of semantic and syntactic differences among the various organizations' data, this compilation task often exceeds the information management capacity available to the clusters.

3.2. Requirements Specification

The primary requirement for HXL is that it address the fundamental information management problem described above: HXL must make the compilation of a common operational picture easier and more timely. The identification of this problem comes from the experiences of information managers who have worked in multiple emergencies over the last several years.

In solving this fundamental problem, any proposed solution must not significantly increase the reporting burden already imposed on humanitarian actors and ideally should reduce it by making it possible for organizations to report data once in a way that serves the diversity of users, from operational partners to analysts at the global level. Furthermore, to be successful, any proposed solution should not require the replacement of existing information management systems, but rather focus on interoperability between existing systems. A standard way of describing and encoding operational humanitarian data can achieve this interoperability, however a standard alone is not adequate to solve the problem. During a crisis, there is not sufficient time for organizations who have not been part of an international humanitarian response to integrate a data standard into their operations; indeed, for many small actors there would not be resources for such integration work. Instead, any solution must include not only a data standard, but also a suite of tools allowing easy translation from the most common information management systems (spreadsheets and relational databases) to the data standard. Additionally, to encourage participation in the data standard, this suite of tools should should also be able to quickly produce some feedback in the form of data visualization and/or maps.

In the chaotic environment of an international crisis response, there are often conflicting data on humanitarian needs as well as operational data. These problems are currently handled, albeit with great effort, at the cluster-level with some support from OCHA for certain types of data. A solution to the stated problem must replicate and improve the efficiency of this cluster-level reconciliation process before data are published. Efficiency will be also be greatly improved if differences among standard reference information, such as place names, can also be resolved. A proposed solution should include the ability to serve out standard reference lists that can be ingested into existing information management systems in a variety of ways, from simple spreadsheet-based gazetteers to standards-based web services for geodata.

In conclusion, the problem HXL addresses is to facilitate data exchange between humanitarian actors in a way so that (a) data can be accessed in a standardized way; (b) standardized terminology and identifiers for commonly used entities can be defined; (c) the use of those identifiers is supported at report time; and (d) the semantics, metadata and provenance of the data is unambiguous.

3.3. Proposed Solution

After analyzing these requirements, a solution based on Semantic Web technologies, following the Linked Open Data [2] paradigm, was identified as the most promising solution. An RDF vocabulary for the domain provides a sound definition of the most important domain concepts, and any datasets annotated with the corresponding classes and properties are self-describing, as the respective terms can always be looked up online. The SPARQL query language [18] provides a standardized API, so that there is no need to define a new, proprietary API specific to this standard. The double function of URIs as both identifiers and locators for information about a resource allows for easy sharing and lookup of IDs for commonly required resources, such as organizations, emergencies, or geographic features. For geographic information about administrative boundaries or locations of refugee camps, the Open Geospatial Consortium's Simple Features Model and the corresponding ontology already provide a useful standard [31] and extension to SPARQL for geospatial queries [32].

The first step towards HXL was the definition of the vocabulary, which is described in the following section.

4. Vocabulary

This section introduces the Humanitarian eXchange Language vocabulary.[11] We first outline the current coverage of the vocabulary, followed by a discussion of the underlying design principles and an overview of the development process.

4.1. Coverage of the HXL Vocabulary

The HXL vocabulary—officially entitled *Humanitarian eXchange Language (HXL) Situation and Response Standard*—has been developed to annotate humanitarian data in the absence of properties in established vocabularies that are detailed enough to meet the requirements outlined in Section 3.2. While the development of the MOAC [34] vocabulary was a first step into the right direction and some generic vocabularies provide useful classes and properties that can (and should) be reused, it was apparent that a substantial number of classes and properties have to be introduced to ensure a meaningful annotation of the data at hand. While qualitative, free-text descriptions of the situation are common in the domain and a valuable source of information, the current version of the vocabulary focuses on quantitative information that can be directly used to generate reports, maps, and interactive dashboards. It consists of five sections that organize the vocabulary by topic:

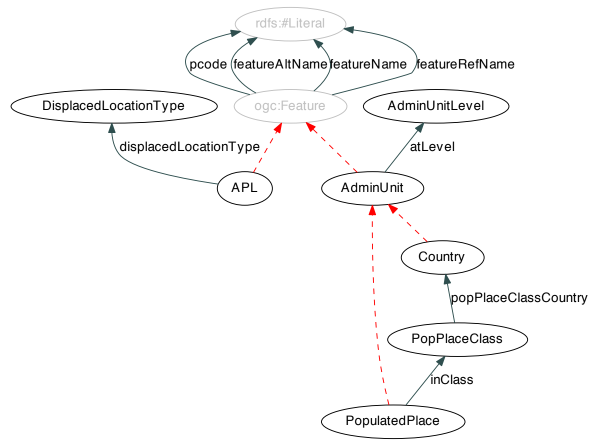

Fig. 2. Overview of the core classes and properties of the geolocation section. Subclass relationships are shown as red dashed arrows.

1. **Geolocation section**. This section provides the classes and properties to annotate geographic information such as the common operational datasets (see Section 2). It builds on the Open Geospatial Consortium's Simple Features model [31] and extends the corresponding ontology.[12] This approach ensures that all HXL data is fully compliant with the GeoSPARQL recommendation [32,1] and hence support complex spatial queries in a standardized way. HXL extends the Simple Features model by the classes and properties required to model the administrative hierarchy in a country, such as `hxl:AdminUnit`. Figure 2 gives an overview of this section of the vocabulary.
2. **Humanitarian profile section**. This section defines the classes and properties required to publish data about the populations affected by an emergency. The classes in this section correspond to the humanitarian profile (see Section 2), breaking down the person counts by the way in

[11] The latest version of the vocabulary is available from `http://hxl.humanitarianresponse.info/ns/`.

[12] See `http://www.opengis.net/ont/geosparql`.

which the corresponding populations are affected (`Casualty`, `Missing`, `Displaced`, etc.). For each of these subclasses of `Population`, the respective properties for `personCount`, `ageGroup`, `sexCategory`, etc. are provided. Figure 3 gives an overview of this section of the vocabulary.

3. **Metadata section**. HXL makes extensive use of named graphs for data management (see Section 7 for details). This section defines the classes and properties to annotate the named graphs, which are declared as instances of the class `DataContainer` in HXL. To track the provenance data relevant in the humanitarian context, each data container has information attached about its `Source`, what person or organization it has been `approvedBy`, as well as timestamps for reporting and validity dates.
4. **Response section**. This section contains the classes and properties to describe the organizations involved in response activities coordinated by OCHA, including name, abbreviation, and internal ID. In the next iteration of the vocabulary, this section will be extended with classes and properties to describe actual response activities such as food distributions, vaccinations, etc.
5. **Situation section**. Similar to the response section, the situation section is a stub that will be extended in the future. Its main purpose is currently to enable the annotation of emergencies in HXL with GLobal IDEntifier (GLIDE) numbers.[13] GLIDE numbers are unique identifiers assigned to all disaster events that meet the criteria of the EM-DAT disaster database[14] and commonly used across the humanitarian domain. The situation section will ultimately contain vocabulary for describing the situation requiring a humanitarian response, including information about needs generated by the crisis (for shelter, water, protection of minors, etc.) and events (security incidents, damage reports, etc.) that shape the response environment.

4.2. Design Principles

Reuse of existing vocabularies is a central principle to ensure interoperability on the Web of Data. However, in the context of HXL as an effort of a United Nations agency, special care had to be taken to make sure that

1. any existing vocabularies used are stable,
2. the definitions of concepts correspond *exactly* to those used in the UN context, and
3. the vocabulary reflects the jargon commonly used in the domain to facilitate adoption.

These points have constrained the number of potential vocabularies to reuse considerably. A class such as `hxl:Country` could have been taken from existing vocabularies such as the DBpedia ontology [5], however, we could not find any definition that includes nations along with dependent territories and other special cases. Those may be considered countries in the humanitarian context—however, this does not imply any endorsement or recognition by the United Nations. For classes and properties such as `hxl:NonDisplaced` or `hxl:householdCount`, there were no existing properties at all. For those reasons, the current version of HXL reuses only the Friend of a Friend (FOAF) vocabulary [8], Dublin Core [14], and the Open Geospatial Consortium's GeoSPARQL ontology [32].

From the technical perspective, the vocabulary has been divided into thematic sections introduced in the previous subsection, using the `rdfs:isDefinedBy` property. This makes the vocabulary more tractable and allows us to automatically generate a well-structured documentation for the vocabulary.[15] All HXL classes are subclasses of an abstract `hxl:BaseClass`. Properties are generally defined with domain and range, and "mandatory" properties are marked using an `owl:minCardinality` of 1. These restrictions are especially useful when developing tools such as the HXLator (see Section 5.1) and to validate whether a HXL dataset is complete with respect to those required properties.

4.3. Development Process

As mentioned in Section 4.2, reuse of terms from the humanitarian field was a mandatory requirement for the development of HXL. The first step was hence to collect as many standards documents, spreadsheet templates, guidelines, and API documentations as possible. From this set of documents, frequently recurring

[13] See `http://www.glidenumber.net/`.
[14] See `http://emdat.be/frequently-asked-questions#FAQ3` for the criteria.
[15] The scripts generating the documentation at `http:http://hxl.humanitarianresponse.info/ns/` are available from `http://github.com/hxl-team/HXL-Vocab/`.

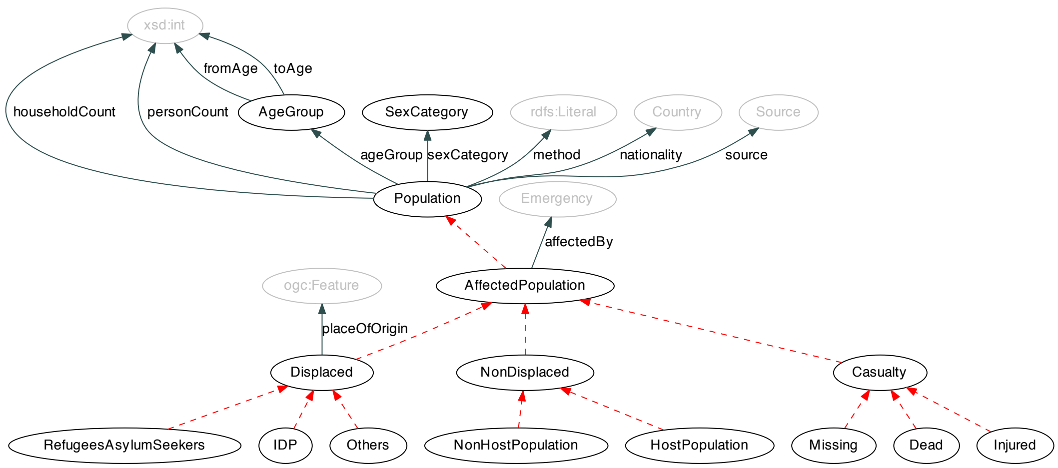

Fig. 3. Overview of the core classes and properties of the humanitarian profile section. Subclass relationships are shown as red dashed arrows.

terms were identified and grouped thematically. These clusters were then arranged as concept maps as a first draft of the class hierarchy, which was discussed in several face-to-face meetings with domain experts for verification. At this point, it turned out that a graph-based data model was a considerable mental leap for the domain experts used to the relational and table-oriented models commonly used in the domain. This resulted in a number of presentations on the Semantic Web, RDF, and Linked Data, which set the stage for further discussions with a better understanding of the technology.

When turning the concept maps into an RDF vocabulary, it quickly became apparent that the scope is too broad for a first version, as the documents that had been taken into account cover tasks as different as needs assessment, financial tracking, and response planning. We therefore took a step back and decided to limit the vocabulary to a specific task for the time being, for which we chose the humanitarian profile. From this first scaled-down version, we have since gone through more than 30 revisions, the latest of which is always available from `http://hxl.humanitarianresponse.info/ns/`. These frequent updates result from a *fail early, fail often* approach to ontology engineering in which we re-evaluate the vocabulary whenever we produce, translate, or consume HXL data and find bugs or missing classes and properties.

5. Data Generation

This section introduces the two current ways of producing HXL data: The HXLator (Section 5.1), an interactive tool that guides the user through the process of translating spreadsheets to HXL; and custom-build system crosswalks (Section 5.2) that publish data from existing information systems as HXL data.

5.1. HXLator

A significant share of the data exchanged in the humanitarian space lives in ad-hoc spreadsheets. In the field, spreadsheets are used to keep track of refugee counts, distribution activities (e.g. of food or shelter kits), needs assessments, etc. Spreadsheet templates have been developed for common use cases that recur in many emergencies. Since these spreadsheets are often used for several reporting purposes, skipping this step and directly entering the data into a tool to generate HXL is not an option at this point. Translating the contents of these spreadsheets to HXL is a straightforward task for someone with a decent background in programming and Semantic Web technologies. Unfortunately, this skill set is usually not covered by the field staff. Existing tools for this problem space [23,17,26], including LODRefine[16] and EasyOpenData,[17] still require a certain amount of techni-

[16]See `http://code.zemanta.com/sparkica/index.html`.
[17]See `http://app.easyopendata.com`.

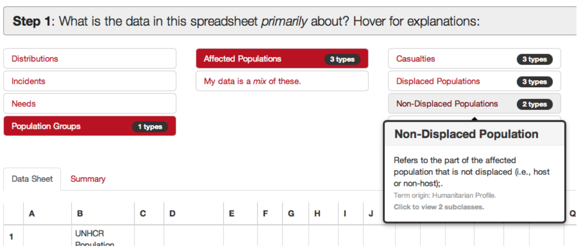

Fig. 4. Class selection in HXLator. When a class is selected, its subclasses expand to the right. Hovering over one of the class buttons brings up the class definition from the HXL vocabulary.

cal knowledge. Moreover, they have been developed as general-purpose tools, whereas our target user group clearly needs a focused tool that guides them through the process to *extract* data from a spreadsheet, *transform* the data to RDF (i.e., HXL), and *load* it into a triple store (ETL). The required level of ease of use for non-technical users and the degree of customization to HXL led to the decision to develop our own solution.

The HXLator is an open source[18] online tool that guides the user through the process of converting the data in a spreadsheet to HXL. It has been developed in PHP, based on the PHPExcel[19] and EasyRDF[20] libraries and the Bootstrap[21] framework for the frontend. The main challenge for the HXL use case was to make the process easy enough so that someone with decent skills in Excel, but no knowledge about Semantic Web technologies, could use it. The HXLator was hence developed with a strong focus on the work flow and hiding most of the complexity of the underlying technology from the user.

After logging in, the user starts by entering some metadata, selecting the emergency the data is about, the report category (currently only supporting *humanitarian profile*), a validity date for the uploaded data, and the file to upload. Optionally, she can reuse an existing translator: once the translation process is complete, HXLator stores the translator the user has created, so that it can be re-applied to a new (or updated) spreadsheet of the same structure. This is especially useful for the commonly used templates mentioned above, as the translators for those can already be provided in HXLator, saving the user the effort to create them.

[18]See `http://github.com/hxl-team/HXLator/`.
[19]See `http://phpexcel.codeplex.com`.
[20]See `http://www.easyrdf.org`.
[21]See `http://twitter.github.com/bootstrap/`.

Once the spreadsheet has been uploaded to the server, it is shown back to the user, who is now guided through a five step process to generate a translator for the file:

- **Step 1**: HXLator shows the classes defined in HXL, organized based on the subclass hierarchy (see Figure 4). Here, the user has to select the class her spreadsheet has data about. This step defines the properties available for mapping, using the domain and range definitions provided in the HXL vocabulary and inferred via subclass reasoning.
- **Step 2**: HXLator asks the user to select the first row in the spreadsheet that contains actual data (not the header row). This row acts as a template and is used for the actual mapping process.
- **Step 3a**: The user selects a cell in this row that identifies an instance of the class selected in Step 1; e.g., a population of refugees and asylum seekers.
- **Step 3b**: The user selects one of the properties available for the selected class. At this point, HXLator distinguishes between data properties and object properties:
  * For data properties, HXLator asks the user to either select the value by clicking the respective cell in the spreadsheet, or by directly typing it in.
  * For object properties, HXLator allows the user to perform lookups for existing resources on the HXL triple store. Again, these can be selected from the spreadsheet or typed in, and the user has to confirm the correct resource to link to (e.g., the URI of a camp that the spreadsheet only contains the name of).

  Steps 3a and 3b are repeated until all cells have been mapped. During this process, all mapped cells and properties are highlighted with green marks (See Figure 5).
- **Step 4**: When the user continues to the next step, she is warned about any properties that have not been mapped. Even though it is unlikely that a spreadsheet will contain data for each of the HXL properties available for the selected class, this step is to make sure the user does not miss any relevant data. Once this is confirmed, the user is asked to select the rows that will be translated in the final step. In case the user has used an existing translator, she will directly jump to this step, i.e.,

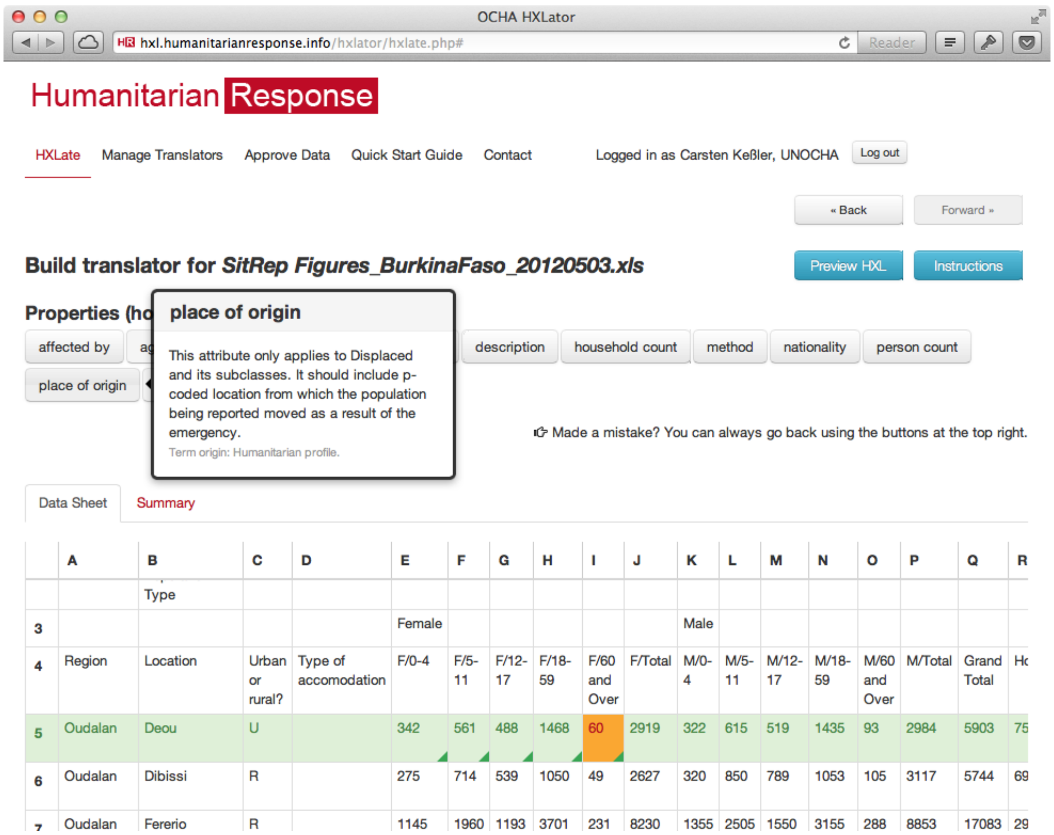

Fig. 5. The HXLator user interface.

skip the actual mapping process and only select the rows to translate.

– **Step 5**: After selecting the rows, the translated data is shown to the user for a final check, before they are submitted to a protected triple store for review (see Section 7).

HXLator contains several functionalities to make this process as straight-forward and efficient as possible, such as the selection of multiple cells that share the same property (e.g., if several cells identify female populations). The preview function mentioned above (see Figure 6) is available for the user throughout the process, so that she can easily check what data has been mapped yet, and whether the mapping is correct so far. If an error is spotted, the user can always go back, as HXLator keeps track of all changes to the translator. The translators are currently implemented as JSON objects, as the whole mapping process happens on the client side, in JavaScript, and JSON is easiest to parse and serialize in this environment. We are considering a switch to R2RML [10] as a more standardized solution that also allows the export of the generated translators to other tools in the future.

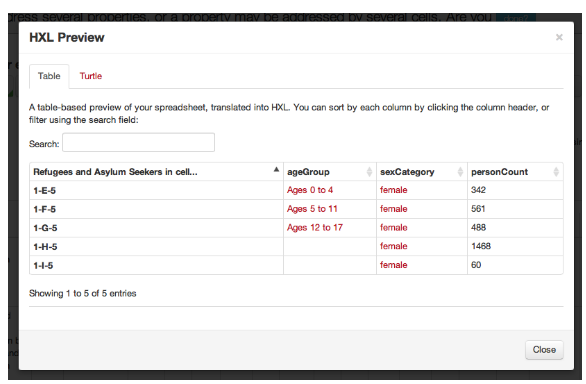

Fig. 6. In HXLator, the user can access a a preview of the generated HXL at all times, both in tabular form and as RDF code in Turtle notation.

The HXLator can be operated either by staff members of the reporting organization, or by OCHA staff. In either case, the biggest effort is creating the initial translator for a certain spreadsheet template. Once this translator has been set up correctly, it can be reused for all spreadsheets of this kind and reduces the translation process to uploading the spreadsheet and selecting the rows to convert. OCHA can provide translators for commonly used spreadsheet templates for reuse.

5.2. System Crosswalks

While HXLator has been developed to map the data from ad-hoc spreadsheets, often filled in on-site, large humanitarian organizations—both within the UN system and international NGOs—already have a wide range of relational databases and information systems in use. As these provide far more predictable and well-structured data, setting up ETL processes for these is more straight-forward. Moreover, the underlying structure (database schema or API) hardly changes, so that it is usually sufficient to have an expert work out a crosswalk to generate HXL from these systems once, and it will work as long as there are no major changes made to the source system.

The first crosswalks developed was the translation of the geographic information contained in the common operational datasets (CODs). The CODs contain GIS shape files for each country OCHA currently has missions in,[22] acting as the geographic reference data for UN agencies in the respective area. These files are central to OCHA's operations, as most other data is directly or indirectly tied to the locations referenced here. Moreover, there was already a system in place that assigned each place a unique ID, a so-called *p-code* that reflects the administrative hierarchy of the feature [20]. Using existing libraries for handling shape files,[23] translating the shape files to RDF using the HXL vocabulary was straight-forward.[24] Since HXL builds on the Simple Features Model [31], HXL data is ready to be queried via GeoSPARQL [32].

The first relational database exposed through HXL is a system for refugee numbers based on their current location, origin, and crisis that affects them. A first version of a custom crosswalk to this database maintained by the United Nations Refugee Organization (UNHCR) was developed from scratch to explore the difficulties that such a translation might bear. It turned out that the translation could only be completed with fairly massive manual intervention, as the database did not make use of p-codes, for example, but only used the place names. In many cases, there are a number of potential spelling alternatives for a place name (and places with the same or very similar names), which cannot be automatically resolved using string distance measures such as the Levenshtein distance [25]. In order to facilitate the translation and streamline the data management practices between UN agencies in general, we aim for a solution that encourages the use of unique identifiers such as the p-codes or GLIDE numbers across different agencies. This will also facilitate the move from translating dumps of the database,[25] as in the current prototype, to "live" exposure of such databases in the future, via tools such as D2RQ [6,4].

With a growing number of tools and scripts producing HXL data, agreed-upon URI patterns [13] gain in importance. A standard pattern per HXL class is especially important to make sure that the same real-world entities are always represented by the same URI, independent of the system that produces the current data at hand. Patterns such as

`http://hxl.humanitarianresponse.info/data/locations/admin/country-code/p-code`

are collected in a shared document[26] and implemented both in the HXLator, as well as the crosswalk tools.

It is worth mentioning that while HXL can help spotting obvious reporting errors, it is still a *garbage in, garbage out* system. Humanitarian data is always messy, potentially biased, and constantly in flux. Numbers from different sources may be conflicting, either because of miscounting, different assessment methods, or intentionally, i.e., for political reasons. These are facts that HXL cannot do away with. Having said that, HXL can help keep track of different data sources and quickly provide emergency management professionals who are aware of these potential pitfalls with the latest assessment numbers.

[22] See `http://cod.humanitarianresponse.info/country-region/afghanistan` for an example.

[23] See `http://www.gdal.org/ogr2ogr.html`.

[24] See `http://hxl.humanitarianresponse.info/data/locations/admin/bfa/BFA050` for an example.

[25] See `http://hxl.humanitarianresponse.info/data/datacontainers/1363644265.2731` for a sample data container generated from the UNCHR database.

[26] See `http://goo.gl/kGnK4`.

6. Data Consumption

The HXL project is driven by the need to be able to re-purpose data into products that support the work of many users: humanitarian actors need a solid basis for planning their activities, donors want to prioritize projects to allocate funds, media need up-to-date information for articles, and academics for scientific studies—hence the need for open, machine readable data. In the following, we introduce two use cases that already leverage the HXL data and discuss how they improve the situation for humanitarian actors.

6.1. Dashboards

The UN is a large organization with many operational and administrative layers. The information collected by information management officers in the field needs to support decision-making at many of these levels. It is often repackaged into a variety of infographics or other types of reports for different audiences. Many of these data visualization products are put together manually in a process that could easily last several days. Obviously, the situation on-site often already differs from the numbers that the decision makers are looking at because of this lengthy process.

The goal was hence to develop a dashboard that is completely driven by HXL data and that can be easily configured and set up for a new emergency. Once up and running, the dashboard should require no manual work and always display the latest data from the HXL triple store. Figure 7 shows a prototype for such a dashboard that was developed during a hackathon at OCHA in Geneva in November 2012. The dashboard was conceptualized as an empty frame that is "put to life" with data dynamically loaded via AJAX from the HXL triple store and an instance of the Humanitarian Response platform. Besides the fact that it is always up to date with the latest data from the triple store, it also allows the decision makers to explore the data in detail, for example through the interactive charts and the mapping module.

6.2. HXL Geo Web Services

The HXL triple store also contains (for those countries used in the prototype HXL development) geographic reference data provided by OCHA (see Section 5.2). While we are sure that having the different features available as Linked Data will be beneficial in the long run, there are not many software solutions available yet that can take advantage of this offering. Instead, most GIS systems and Web mapping frameworks build on the geo web service specifications developed by the Open Geospatial Consortium, most importantly the Web Map Service (WMS) for pre-rendered map images, and the Web Feature Service (WFS) for vector data [33,30].

In order to make the data available in an easily digestible format for GIS analysts and Web mapping applications, a service chain was set up that publishes new geographic reference data through a suite of standardized web services for geospatial data provided by an ArcGIS server instance hosted by an OCHA partner.[27] Since the reference data only change very infrequently, and changes are communicated to the partners beforehand to make sure everyone is always using the same reference data, it was not necessary to develop a "live" mapping that wraps the triple store as a WFS/WMS. Instead, we have implemented a pull-based solution that checks the triple store every night for changes to the geographic information.[28] If any of the data should have changed, a process is triggered that generates an `INSERT` request to the transactional WFS. The ArcGIS server instance hosting the WFS then automatically creates a WMS based on the data.

Both services are available to the whole community and already in use on the dashboard shown in Figure 7, for which the WMS delivers the base map. The beauty in this solution is that the geo web service infrastructure is automatically synced with the triple store on a daily basis, where the changed shape files would have required manual updates to the ArcGIS server instance before. Moreover, this approach easily supports setting up additional server instances that can be run e.g. onsite for the local staff in an emergency with poor Internet connectivity, acting as a "Spatial Information Infrastructure [29] on a USB stick".

7. Data Management

A data management approach that is designed around existing structures within OCHA and the humanitarian sector as a whole was a fundamental requirement during the development of HXL. This section discusses the governance and workflow requirements and how they were implemented in the standard itself and the surrounding tools.

[27] The GIST, hosted by the University of Georgia; see `https://gistdata.itos.uga.edu/node/5`.

[28] See `http://github.com/hxl-team/HXL2WFS`.

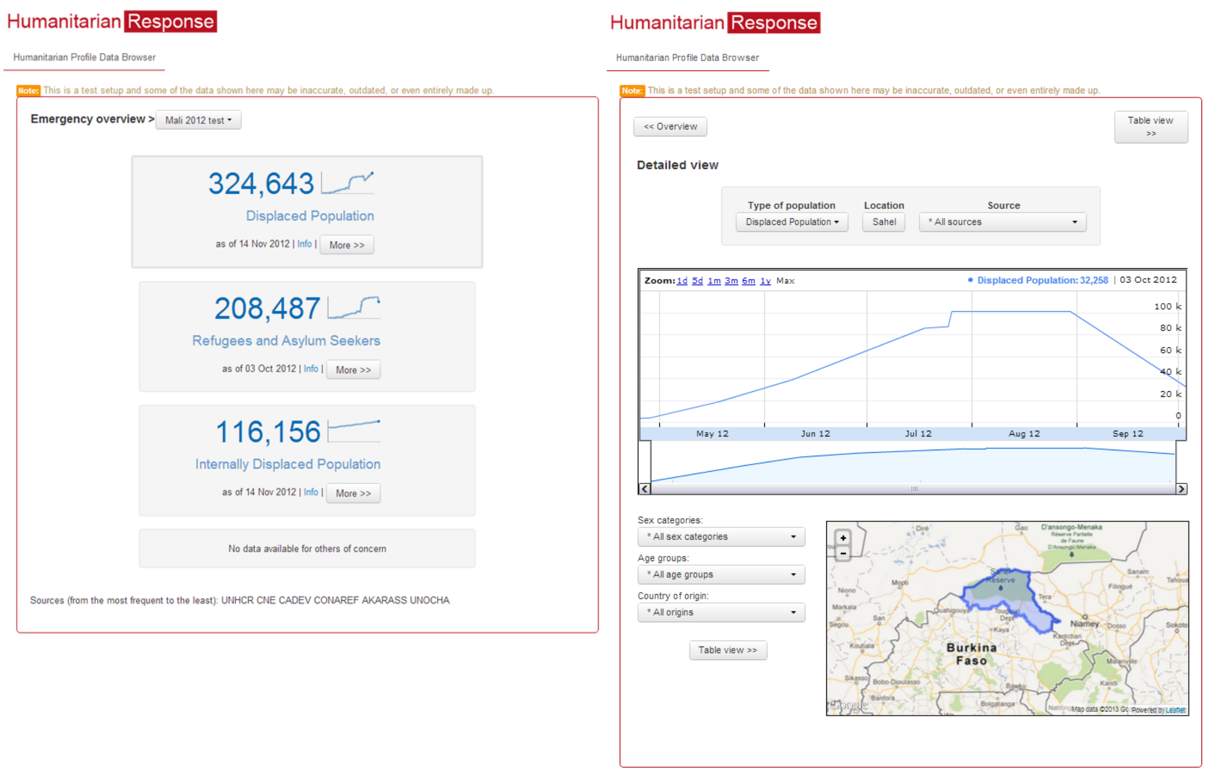

Fig. 7. A prototype dashboard based on HXL data, available from `http://hxl.humanitarianresponse.info/dashboard`. The left side shows the initial page that allows the user to select the emergency and provides the key numbers. The right side shows the detailed view for an emergency, where numbers can be broken down by age or sex group, for example.

7.1. Governance and Workflows

Data collected by OCHA and its partner organizations are the basis of decisionmaking for the international community, especially concerning the allocation of resources in the aftermath of a disaster. Wrong or incomplete data can easily lead to insufficient political actions, putting lives in the affected areas at risk. Within the humanitarian domain, erroneous data can lead to a wrong focus in planning the response activities, or even put field staff in dangerous situations. It is therefore of utmost importance that all data published through HXL go through a review by staff members who are familiar with the overall situation in a specific crisis. Most data compilation already happens at the cluster-lead level, which has a broader scope than the information management officers on a specific site. At the same time, the cluster-lead staff is still highly familiar with the situation, so that any obviously wrong data will immediately catch their eye.

The overall workflow for HXL is hence adopted from existing practice within the humanitarian domain (see Figure 8): Any data collected in the field is compiled and translated to HXL by the information management officers before it propagates to the headquarters in Geneva and New York through the respective cluster lead. This approach also applies the *many eyes principle* to reduce the number of potential errors.

7.2. Implementation

The workflow introduced in the previous section has been implemented at two different levels: In the server setup, and in the HXL vocabulary.

The metadata section of the HXL vocabulary (see Section 4.1) is designed around the principle of using named graphs for data management. In HXL, a named graph is an instance of the class `DataContainer`. A data container is self-describing in that it contains all corresponding metatada. The corresponding triples are automatically generated by the data generation and approval tools (see Section 5). In the data management context, the following properties are particularly

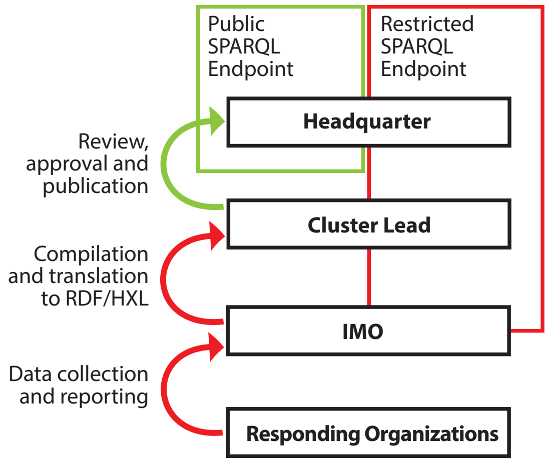

Fig. 8. Overview of the data propagation process.

important, as they contain the reporting and approval provenance data:

- `aboutEmergency`: Links the data container to the corresponding emergency it contains data about.
- `reportCategory`: Links the data container to a specific report category such as *humanitarian profile*, *shelter* or *security*. The report category is used to determine who is responsible for reviewing the data before publication.
- `reportedBy`: Metadata property that links the data container to the person or organization that has reported the contained data.
- `approvedBy`: Metadata property that links the data container to the person or organization that has approved the contained data for publication.

As indicated in Figure 8, the HXL server setup hosts two separate triple stores (both running Fuseki[29]): A public endpoint[30] that can be queried without restrictions, and a second endpoint for unapproved data with restricted access. We refer to this protected triple store as our *incubator store*. Any data coming in from the field, such as from the HXLator, are staged for review in the incubator store. The corresponding cluster lead is notified that new data is waiting to be reviewed, and can look at the data in a web frontend. In case of suspicious data, such as near-duplicates (which may point to double reporting) or extreme outliers, the reviewer needs to find out whether there has been a reporting error, or whether the situation in the field has really changed dramatically. This is usually achieved by calling the colleagues in the field. In the data publication step, i.e., when a data container gets approved at the cluster-lead level, the corresponding triples are moved from the restricted triple store to the public one.

8. Infrastructure

This section introduces the ideas behind the infrastructure in which HXL is set to operate. It discusses how the design decisions behind the infrastructure addresses situations with poor or no connectivity, and lays out the next steps towards a more decentralized setup.

8.1. Low-bandwidth and Offline Situations

HXL is set to operate in a diverse and distributed environment. The technical preconditions in which HXL data is produced, managed, and consumed, vary widely, ranging from high-end, fiber-connected computing facilities in the UN headquarters down to situations in the field with very basic equipment and potentially poor connectivity. Most of the bandwidth and electricity related issues are anticipated to be resolved in the foreseeable future with improved mobile setups for OCHA's missions and the steady development of faster and more reliable satellite-based communication. Nonetheless, HXL and the surrounding tools have been designed to work in low bandwidth and even offline situations.

The key element in the organization of HXL data is the concept of the data container. As mentioned in the previous section, each instance of the class `hxl:DataContainer` is a named graph. The rationale behind this naming choice is that it should also be possible to transfer a data container offline as a file, e.g. on a USB stick, when OCHA staff works in situations without online access. The minimum requirement to have HXL work in the field is therefore a local setup to host and deliver the data at the base camp. Along with a local setup delivering spatial information (see Section 6.2), this configuration is designed to work at a base camp that has only occasional online access to sync the collected data with the headquarters, but can still maintain an overview of the situation in the field, including relevant dashboards which can be ran off a

[29] See `http://jena.apache.org/documentation/serving_data`.
[30] See `http://hxl.humanitarianresponse.info/sparql`.

local server. The same applies for the HXLator; however, its reconciliation functionality will be limited to resources that are available on the local triple store.

8.2. Decentralization

Having separate HXL setups in the field that only hold the locally relevant data, and are synchronized on a regular basis with the headquarters, points to a decentralized infrastructure. While the current setup is fairly centralized, the need for a more decentralized approach is also underlined by the HXL data that will be hosted by partner organizations. We are currently working on exposing the first relational databases maintained by partner organizations through HXL. This decentralized approach also leaves the data with the experts on the respective data, such as refugee numbers or demining activities, which greatly facilitates quality control. In order to still be able to query those distributed datasets without the need to copy everything into a single store, we are currently evaluating different options for query federation [19], with query performance being a critical criterion [27].

9. Conclusions and Future Work

The Humanitarian eXchange Language is an emergent standard for operational data in the humanitarian domain. It is the foundation for an infrastructure that makes strong use of Semantic Web technology in producing, maintaining, and using the data which form the basis for the planning of humanitarian activities in the international community. In this paper, we have reported on the first steps to making HXL a central reference point for the whole domain. The peculiarities of this field are reflected in a set of requirements that HXL needs to address concerning the structure and existing practices in the humanitarian ecosystem. The HXL vocabulary formalizes established terminology from the domain. It currently focuses on humanitarian profile data as well as core reference data, such as geographic information. Data according to the HXL vocabulary can currently be produced either using the HXLator, an interactive tool to translate spreadsheets to HXL, or via crosswalks that produce HXL from existing information systems. HXL data already drive the first applications, including emergency dashboards and a Web service infrastructure for geographic information. An initial data governance system has been set up to ensure that any data coming in from the field are approved at the cluster lead level before publication.

The current version of the vocabulary and the corresponding tools are at a prototyping stage to demo the capabilities of HXL within the UN system and to outside partners to foster adoption. The next steps include the extension of the approval tool chain with consistency checks and automatic highlighting for data that drastically diverge from existing data. This could either point to significant changes in the situation in a camp, for example, or it could point to a reporting error; resolving such issues will require the expertise of an information manager familiar with the situation on the ground. The main task will hence be to build a system that reliably identifies potential problems in the data, and provides an easy-to-use interface to resolve them. The HXL vocabulary will be gradually extended as required, depending on the next reference datasets to be included.

Defining the HXL vocabulary for the humanitarian system as a whole clearly goes beyond the capabilities and expertise of OCHA. In order to achieve this goal, the involvement of the global clusters in developing their respective components, such as vocabulary extensions and cluster-specific tools, is required. As emergency management has substantial overlap with development aid, mappings to initiatives in this domain have to be created, including AGROVOC maintained by the Food and Agriculture Organization of the United Nations [24], the International Aid Transparency Initiative (IATI),[31] and others [12]. Links to several of these initiatives already exist within OCHA, and different models for collaboration are being discussed. In order to increase interoperability with systems and communities outside of the humanitarian domain, HXL should be aligned with existing vocabularies such as DataCube [9]. This would make the statistical nature of most HXL data explicit.

So far, the HXL vocabulary and the tools built around it are only available in English. This is due to the focus of our first development phase, which targets the staff working at camps and the headquarters. In both cases, English is the main language used for communication. Having said that, it is clear that for a broader adoption beyond these circles, a translation into more languages is required. For classes and properties that have equivalents in other languages, we

[31] See http://support.iatistandard.org/categories/20001338-The-IATI-Standards.

can add the corresponding labels and comments with the respective language tag. More research will be required, though, for labels that do not have direct equivalents, and to make sure that the language used also reflects the language of the affected communities [36].

Moreover, the volunteer and technical community needs to be included. Collaboration with this community has already been established in a *random hacks of kindness*[32] event at the international crisis mappers conference in Washington, DC in fall 2012. Future involvement should also address the development of models for crowd-sourced data, which is not yet covered in HXL. The underlying technology, however, has the potential to vastly improve the integration of official agency data information and data collected by the volunteer community.

An approach based on Semantic Web technologies already covers many of the key pieces of a distributed data management system, such as unique and reusable identifiers, clearly defined semantics, a standardized API, and scalable storage. In addition, this approach does not force any of the affected organizations to change the workflows and systems they have in place. In contrast to a "super system" that imposes radical change on all involved organizations, HXL can be rolled out gradually. Many of the existing inter-organizational standardization efforts are either small patches for very specific problems, or lead to bigger information silos. The amount of feedback and interest we receive from the humanitarian domain shows that a more comprehensive solution is required. The biggest obstacle we currently see for the adoption of HXL is the lack of developers familiar with Semantic Web technologies. A graph-based model adds a steep learning curve for programmers who mostly deal with relational data models. We are therefore working on an easy-to use RESTful [15] API layer on top of the SPARQL endpoint that enables straightforward access to frequently requested data.

References

[1] Robert Battle and Dave Kolas. Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL. *Semantic Web*, 3(4):355–370, 2012.

[2] Tim Berners-Lee. Linked Data – Design Issues. Available from `http://www.w3.org/DesignIssues/LinkedData.html`, 2009.

[3] Yaser Bishr. Overcoming the Semantic and Other Barriers to GIS Interoperability. *International Journal of Geographical Information Science*, 12(4):299–314, 1998.

[4] Christian Bizer and Richard Cyganiak. D2R Server—Publishing Relational Databases on the Semantic Web. Poster at the the 5th International Semantic Web Conference (ISWC2006), Athens, GA, USA, 5–9 November 2006. Available from `http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/Bizer-Cyganiak-D2R-Server-ISWC2006.pdf`, 2006.

[5] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. DBpedia—A crystallization point for the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165, 2009.

[6] Christian Bizer and Andy Seaborne. D2RQ—treating non-RDF databases as virtual RDF graphs. Poster at the 3rd International Semantic Web Conference (ISWC2004), Hiroshima, Japan, 7–11 November 2004. Available from `http://iswc2004.semanticweb.org/posters/PID-SMCVRKBT-1089637165.pdf`, 2004.

[7] Eva Blomqvist. The Use of Semantic Web Technologies for Decision Support – A Survey. *Semantic Web Journal*, accepted.

[8] Dan Brickley and Libby Miller. FOAF Vocabulary Specification 0.98. Available from `http://xmlns.com/foaf/spec/20100809.html`, 2010.

[9] Richard Cyganiak, Dave Reynolds, and Jeni Tennison. The RDF Data Cube Vocabulary. W3C candidate recommendation available from `http://www.w3.org/TR/2013/CR-vocab-data-cube-20130625/`, 2013.

[10] Souripriya Das, Seema Sundara, and Richard Cyganiak. R2RML: RDB to RDF Mapping Language. W3C recommendation available from `http://www.w3.org/TR/r2rml/`, 2012.

[11] Tim Davies. IATI Linked Data. Discussion paper available from `http://goo.gl/Xc8s4`, 2012.

[12] Tim Davies and Duncan Edwards. Emerging Implications of Open and Linked Data for Knowledge Sharing in Development. *IDS Bulletin*, 43(5):117–127, 2012.

[13] Leigh Dodds and Ian Davis. Linked data patterns – a pattern catalogue for modelling, publishing, and consuming linked data. Available from `http://patterns.dataincubator.org/book/`, 2012.

[14] Dublin Core Metadata Initiative. DCMI Metadata Terms. Avaliable from `http://dublincore.org/documents/dcmi-terms/`, 2012.

[15] Roy Thomas Fielding. *Architectural Styles and the Design of Network-based Software Architectures*. PhD thesis, University of California, Irvine, USA, 2000.

[16] Thomas R. Gruber. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.

[17] Lushan Han, Tim Finin, Cynthia Parr, Joel Sachs, and Anupam Joshi. RDF123: a mechanism to transform spreadsheets to RDF. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI 2006). AAAI Press, Menlo Park*, 2006.

[18] Steve Harris and Andy Seaborne. SPARQL 1.1 Query Language. W3C recommendation available from `http://www.w3.org/TR/sparql11-query/`, 2012.

[32] See `http://www.rhok.org`.

[19] Olaf Hartig, Christian Bizer, and Johann-Christoph Freytag. Executing SPARQL Queries over the Web of Linked Data. In Abraham Bernstein, DavidR. Karger, Tom Heath, Lee Feigenbaum, Diana Maynard, Enrico Motta, and Krishnaprasad Thirunarayan, editors, *The Semantic Web – ISWC 2009*, volume 5823 of *Lecture Notes in Computer Science*, pages 293–309. Springer Berlin Heidelberg, 2009.

[20] Chad Hendrix. Field Guide for the Use of Geo-Codes. Available from http://goo.gl/Ya7fh, 2011.

[21] Inter Agency Standing Committee. IASC Guidelines Common Operational Datasets (CODs) in Disaster Preparedness and Response. Available from http://cod.humanitarianresponse.info/node/53, 2010.

[22] Inter-Agency Standing Committee. Multi-Cluster/Sector Initial Rapid Assessment (MIRA). Available from http://ochanet.unocha.org/p/Documents/mira_final_version2012.pdf, 2012.

[23] Andreas Langegger and Wolfram Wöß. XLWrap – Querying and Integrating Arbitrary Spreadsheets with SPARQL. In Abraham Bernstein, David R. Karger, Tom Heath, Lee Feigenbaum, Diana Maynard, Enrico Motta, and Krishnaprasad Thirunarayan, editors, *The Semantic Web – ISWC 2009*, volume 5823 of *Lecture Notes in Computer Science*, pages 359–374. Springer Berlin Heidelberg, 2009.

[24] Boris Lauser and Margherita Sini. From AGROVOC to the Agricultural Ontology Service/Concept Server: An OWL Model for Creating Ontologies in the Agricultural Domain. In *Proceedings of the 2006 International Conference on Dublin Core and Metadata Applications: Metadata for Knowledge and Learning*, DCMI '06, pages 76–88. Dublin Core Metadata Initiative, 2006.

[25] V.I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8):707–10, 1966.

[26] Fadi Maali, Richard Cyganiak, and Vassilios Peristeras. Reusing Cool URIs: Entity Reconciliation Against LOD Hubs. In Christian Bizer, Tom Heath, Tim Berners-Lee, and Michael Hausenblas, editors, *WWW2011 Workshop on Linked Data on the Web, Hyderabad, India, March 29, 2011*, volume 813 of *CEUR Workshop Proceedings*, 2011.

[27] Gabriela Montoya, Maria-Esther Vidal, Oscar Corcho, Edna Ruckhaus, and Carlos Buil-Aranda. Benchmarking Federated SPARQL Query Engines: Are Existing Testbeds Enough? In Philippe Cudré-Mauroux, Jeff Heflin, Evren Sirin, Tania Tudorache, Jérôme Euzenat, Manfred Hauswirth, Josiane Xavier Parreira, Jim Hendler, Guus Schreiber, Abraham Bernstein, and Eva Blomqvist, editors, *The Semantic Web – ISWC 2012*, volume 7650 of *Lecture Notes in Computer Science*, pages 313–324. Springer Berlin Heidelberg, 2012.

[28] Ory Okolloh. Ushahidi, or'testimony': Web 2.0 tools for crowdsourcing crisis information. *Participatory Learning and Action*, 59(1):65–70, 2009.

[29] Harlan Onsrud, Barbara Poore, Robert Rugg, Richard Taupier, and Lyna Wiggins. The Future of the Spatial Information Infrastructure. In Robert B. McMaster and E. Lynn Usery, editors, *A Research Agenda for Geographic Information Science*, pages 225–255. CRC Press, 2005.

[30] Open Geospatial Consortium. OpenGIS Web Feature Service 2.0 Interface Standard (also ISO 19142). Available from http://opengis.org/standards/wfs, 2010.

[31] Open Geospatial Consortium. OpenGIS Implementation Specification for Geographic information – Simple feature access – Part 1: Common architecture. Available from http://opengis.org/standards/sfa, 2011.

[32] Open Geospatial Consortium. OGC GeoSPARQL – A Geographic Query Language for RDF Data. Available from http://opengis.org/standards/geosparql, 2012.

[33] Open Geospatial Consortium. OpenGIS Web Map Service (WMS) Implementation Specification. Available from http://opengis.org/standards/wms, 2012.

[34] Jens Ortmann, Minu Limbu, Dong Wang, and Tomi Kauppinen. Crowdsourcing Linked Open Data for Disaster Management. In Rolf Grütter, Dave Kolas, Manolis Koubarakis, and Dieter Pfoser, editors, *Proceedings of Terra Cognita 2011, Workshop at the 10th International Semantic Web Conference (ISWC2011)*, volume 798 of *CEUR Workshop Proceedings*, Bonn, Germany, October 2011.

[35] United Nations Office for the Coordination of Humanitarian Affairs. OCHA Strategic Framework, Objective 2.4. Available from http://www.unocha.org/ocha2012-13/strategic-plan/objective-2_4, 2012.

[36] Jessica Seddon Wallack and Ramesh Srinivasan. Local-global: Reconciling mismatched ontologies in development information systems. In *42st Hawaii International International Conference on Systems Science (HICSS-42 2009), Proceedings, 5–8 January 2009, Waikoloa, Big Island, HI, USA*, pages 1–10, 2009.