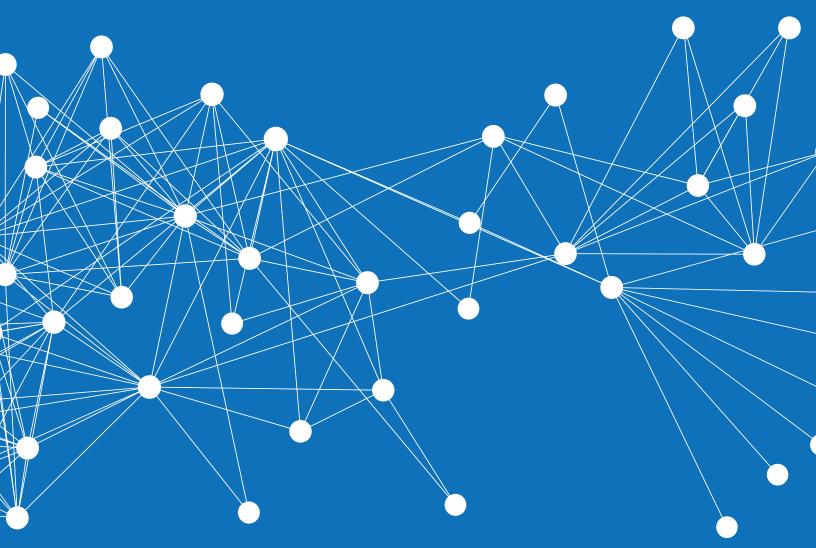
DEVELOPING HUMANITARIAN DATA STANDARDS: an introduction and plan for 2014











THE **DURABLE SOLUTION** FOR OPERATIONAL DATA SHARING IN HUMANITARIAN EMERGENCIES.

CONTENT

- **04** Introduction
- **O5** Starting Inventory
- **O7** One-Year Roadmap
- 10 Major Decisions
- 14 Proposed Working Group Charter

This paper was written by **David Megginson**, an independent consultant, in collaboration with staff from the United Nations Office for the Coordination of Humanitarian Affairs. It is based on the project plan described in the grant agreement that OCHA has with the Humanitarian Innovation Fund.

All network images are credited to Seed Scientific (http://seedscientific.com/).

For ongoing updates on this work go to http://www.humanitarianinnovation.org/projects/large-grants/UNOCHA

INTRODUCTION

eginning in 2014, the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) will lead a community standards effort to define a Humanitarian Exchange Language (HXL). HXL will help the humanitarian community capture and share high-priority response data across a network of disparate actors, such as agencies, NGOs, and governments.

The humanitarian community's existing data-sharing practices usually involve one-to-one exchange of non-standard spreadsheets or individual figures over email at irregular intervals. These ad-hoc methods can cause a significant delay between the collection of data and the formulation of that data into a common operational picture. In the worst case, due to the effort involved, information is simply not shared at all, leaving gaps in the understanding of the field situation.

At OCHA, we have observed several negative consequences of this situation:

- **Delays:** Data typically arrives too late to provide meaningful day-to-day decision-making support during a crisis.
- **Duplication of effort:** Humanitarian stakeholders create numerous point solutions to meet specific needs, but cannot reuse them for other purposes, increasing fragmentation and inefficiency.
- Inaccessibility and uncertain provenance: Humanitarian stakeholders spend unnecessarily large amounts of time searching for data, and they are often uncertain of the source, accuracy and terms of use for what they find.
- Lack of reusability: Humanitarian data is not available as an open data standard for web developers to access and combine with other data domains (development data, financial data, etc).
- Information loss: While the community generally archives reports and other high-level documents after a crisis, the data used to produce them is

usually unrecoverable, and thus, unavailable for future retrospective analysis.

This document describes the different aspects of the HXL initiative that OCHA, with the kind support of the Humanitarian Innovation Fund, will carry out during 2014. The work will focus on a core HXL working group made up of representatives of different parts of the humanitarian community. That group will draft the standard while consulting the broader community through an open e-mail list and other channels. To frame the HXL initiative at its beginning, this document considers several areas:

- Starting inventory: what OCHA has learned about humanitarian aid data from a proof-of-concept effort during 2012 and 2013, and what feedback we have received.
- One-year roadmap: a tentative schedule for developing the HXL standard, including supporting pilot projects and technical infrastructure that OCHA will provide.
- Major decisions: an overview of some of the major decisions that the working group, and broader HXL initiative, will face during their first year.
- Proposed working group charter: a draft terms of reference for the HXL working group.

¹ http://www.humanitarianinnovation.org/projects/large-grants/UNOCHA

2. STARTING INVENTORY

his section takes inventory of OCHA's past, proof-of-concept work on creating humanitarian data standards, under the "HXL" banner. Much of the prototype work involved refugee data, in cooperation with the United Nations High Commissioner for Refugees (UNHCR), focusing on information entities such as geographical locations, affected populations, and ways of classifying people (see links below for further details).

Another important tool is the prototype humanitarian profile

dashboard,7 using data from UNHCR transformed to the HXL

format and published to a public HXL data store.8

tool,6 a browser-based application for mapping spreadsheet

data to the proof-of-concept HXL format.

2.1. Documentation artifacts

The main starting point for information about the initial HXL proof-of-concept work is the web site http://hxl.humanitari-anresponse.info/. This site hosts two sets of documentation (also stored under the hxl-team organisation at github, described in the next section):

Introductory, tutorial and task-centric documentation.²

Technical, format-centric documentation.³

Additionally, the RDF schemas defining the HXL vocabulary are available in several machine readable formats:

> Turtle: http://hxl.humanitarianresponse.info/ns/hxl.

> RDF/XML: http://hxl.humanitarianresponse.info/ns/

> N-Triples: http://hxl.humanitarianresponse.info/ns/hxl.n3

These initial HXL schemas focussed on the Inter Agency Standing Committee's Humanitarian Profile Guideline,⁴ which defines hierarchical categories of affected people in a humanitarian crisis.

2.2. Technology artifacts

The hxl-team organisation at github⁵ hosts a variety of conversion utilities and browsing tools for the proof-of-concept HXL format. Most prominent among them is the HXLator

2 http://hxl.humanitarianresponse.info/docs/

Most of these projects were last updated during winter and spring 2013, and are written in the PHP programming language (some are in JavaScript). While these artifacts may not be usable with a new HXL format developed by a collaborative working group, they do provide good models for the kind of technical infrastructure support that we will have to provide for the HXL work over 2014.

2.3. Existing interest

The United Nations High Commissioner for Refugees (UN-HCR) is conducting a humanitarian data-exchange pilot with OCHA using the existing HXL prototype standard to publish data about refugees.

AppInventor is a research project at MIT, creating a tool that will allow non-developers to create simple custom mobile applications to display Semantic-Web data. The project overview paper⁹ lists HXL among that standards that the tool will support (assuming that HXL follows Semantic-Web standards such as RDF).

The Ohio Center of Excellence in Knowledge-enabled Computing (Kno.e.sis) at Wright State University has expressed interest in using HXL in their Coordination analysis framework.¹⁰

The European Research Center for Information Systems (ER-CIS) at the University of Münster has expressed interest in

³ http://hxl.humanitarianresponse.info/ns/index.html

⁴ https://assessments.humanitarianresponse.info/files/iasc_guidelines_humanitarian_profile.pdf

⁵ https://github.com/hxl-team/

⁶ https://github.com/hxl-team/HXLator

⁷ http://hxl.humanitarianresponse.info/dashboard/

⁸ http://hxl.humanitarianresponse.info/dashboard/

⁹ http://dig.csail.mit.edu/2013/Papers/IJCAIWorkshop/IJCAI_Workshop SemanticCity.pdf

¹⁰ https://docs.google.com/a/google.com/document/d/1aBJ2egHIC UwaWxR8jOoTIUfEYj10AnUt0q7halKoYGY/ (does not mention HXL explicitly).

developing a logistics extension for HXL (no public information available).

Geopii is an enterprise-software company in Northern Ireland, spun out from the University of Ulster. They are evaluating the HXL proof-of-concept as an option for communicating among NGOs and other organisations (no public information available).

The Open Humanitarian Initiative (OHI) has expressed interest in HXL as a humanitarian exchange language (no public information available).

3. ONE-YEAR ROADMAP

his section proposes a high-level, one-year roadmap for the entire HXL initiative during 2014, including, but not limited to the HXL working group. The section has three parts, each expanding on the previous:

- **Deliverables:** what we want the HXL initiative to have accomplished at the end of its first year, in terms of high-level deliverables and broader humanitarian outcomes.
- Activities: the different lines of work in which we will have to engage to realise those goals.
- Milestones: the major points where we will be able to verify and evaluate the activities' progress towards the goals.

3.1. Deliverables

The project's goal is to have the following in place by the end of 2014:

- HXL Standard: there is an open, usable, broadlysupported v1.0 standard for exchanging at least a small amount of high-value humanitarian data during a protracted crisis.¹¹
- HXL Participation: at least <u>ten</u> humanitarian organisations have committed to exchanging data in HXL format (including OCHA), and, ideally, at least five of them have begun work.
- HXL Pilots: there are public results available from two field pilots of HXL data exchange, including

specific information on successes, failures, and proposed best practices.

HXL Infrastructure: there is basic free and open technical infrastructure in place to support producers and consumers of HXL data, such as schemas, validation tools, programming libraries/modules/plugins, etc.

3.2. Activities

To reach the deliverable goals listed in the previous section, the HXL project will include the following four major activity streams during 2014:

- A **HXL** working group, led by OCHA and including representatives from at least 5 humanitarian organisations, will develop the HXL standard or standards. The working group will first set the (limited) scope for the v.1 standards work then proceed to create draft standards, and to solicit and respond to community feedback. The group will also propose infrastructure tools, reach out to other organisations, and review the results of pilots.
- A **HXL technical team** within OCHA will create technical infrastructure such as tools and libraries as requested by the HXL working group, will be responsible for the technical implementation of the pilot projects on the OCHA HQ side, and will provide information and assistance to other offices or organisations participating in the pilot projects as required.
- A HXL programme team (including the leader of the working group) will take charge of outreach to other humanitarian organisations and to other parts of OCHA, with the goal of obtaining management-level approval participation from other participants. The programme team will also be responsible for management of the entire activity, including the

¹¹ We propose starting with protracted crises (such as displaced people in Colombia) rather than sudden-onset crises (such as the typhoon in the Philippines), to avoid distracting newly-established emergency response teams with information requests during our development and prototyping period.

HXL pilots, for communications, and for programme-level coordination with the pilot organisations and other HXL participants.

3.3. Milestones

At the beginning of 2014, all three HXL activity teams will be in place and ready to start work. The working group will have been formed already in the end of 2014, with participants selected from OCHA and other humanitarian organisations, and will be ready to begin biweekly meetings immediately in January. The technical team does not need to be fully in place, since its work will be limited until the first draft standard release in March.

The following table proposes quarterly milestones for each activity team:

When	Who	What
March 2014	Programme team	Has identified short list of candidates for the HXL field pilots, and is in advanced negotiation with the candidate countries/organisations.
		Has begun discussions with 15-20 humanitarian partners about future HXL participation.
		Consults with the working group about the proposed pilots.
	Working group	Releases first rough draft of a proposed HXL standard. This draft is unstable, possibly incomplete, and subject to rapid change (alpha-level).
		Circulates draft standard to the wider humanitarian community for feedback.
	Technical team	Begins planning and prototyping technical infrastructure based on the draft standard.
June 2014	Programme team	Has field pilots planned and ready to launch, with all partners committed.
		Begins supervising pilots, based on the second (beta) draft of the HXL standard.
	Working group	Releases second, more-polished draft of the HXL standard (beta-level).
	Technical team	Goes into full working mode for the field pilots, implementing the required processes on the OCHA HQ side, and assisting partners in the field.
		Continues work on infrastructure tools, focusing particularly on those required for the pilots.
September 2014	Programme team	Begins to wrap up field pilots, and reports results to the working group and to the wider humanitarian community.
		Continues outreach. At this point, at least five partner organisations should be on board for ongoing HXL data exchange starting in 2015.
		Begins work on a sustainability plan for HXL past the end of 2014.

	Working group	Reviews results of the field pilots.
		Reviews the technical infrastructure currently in place, and identifies gaps that still need to be filled.
		Publishes a stable release candidate of the HXL standard(s) for broad community feedback, and processes/responds to feedback.
	Technical team	Continues to develop basic HXL infrastructure, as directed by the programme team and the working group, and informed by the pilot results.
		"Productizes" tools by adding documentation, test suites, installers, etc., and making them available through public sources like github.
December 2014	Programme team	Formalizes agreements with at least 9 other humanitarian organisations to begin full production-level HXL data exchange in 2015.
		Finalizes continuity and sustainability plan for HXL in 2015 and beyond.
		Launches large-scale HXL outreach to humanitarian community, including mainstream media outreach, publications, and conference presentations.
	Working group	Releases the final 1.0 version of the HXL standard(s), incorporating feedback from the release candidates.
		Reconstitutes itself, possibly in a different form or as part of a larger group, to provide continuity and stability for HXL in 2015 and beyond.
	Technical team	Releases stable versions of the HXL technical infrastructure tools and assists other organisations in adopting them.

4. MAJOR DECISIONS



his section introduces some of the major decisions that we will need to make during the first year of HXL work.

4.1. Affiliation

Any viable international standard needs reach (wide-ranging familiarity and buy-in from stakeholders) and sustainability (a reasonable guarantee that it will continue to be maintained into the future), both of which require a level of governance that can take years to build. As a result, we need to decide during 2014 whether we will try to build up HXL as a standards body in its own right, or whether we will attempt to affiliate our work with that of another existing, better-established group.

If we choose to affiliate HXL with an existing group, there are several candidates we could consider:

International Aid Transparency Initiative (IATI) — after more than five years of work, IATI¹² has established itself as the leading organisation for open aid data, with visibility and stakeholder buy-in up to the ministerial or head-of-state level in many countries, and a funding and governance plan that should sustain it over at least the next five years. To date, IATI has focussed primarily on financial and organisational data, and has concentrated mainly on official development assistance (ODA); however, IATI is now working with OCHA (particularly the Financial Tracking Service) to extend its financial standards to support humanitarian initiatives, and may be interested in the kind of operational data that HXL will be defining.

Inter-Agency Standing Committee (IASC) — the IASC13 is currently the closest thing the humanitarian world has to a standards group, and already includes many humanitarian stakeholders. However, it tends to work at a fairly high level, and its main body meets only three times/year, so its current structure might not support the kind of detailed standards work required for HXL. The IASC currently oversees the Common Operational Datasets (CODs),14 which include the Humanitarian Profile,15 with mixed success.

12 http://www.aidtransparency.net/

Sphere — Sphere¹⁶ is a voluntary initiative that sets standards for humanitarian work in the field and defines indicators for measuring that work. The Sphere Handbook¹⁷ is a de-facto quide for a wide range of humanitarian activities in areas like WASH and shelter. HXL should cooperate closely with Sphere during its standards-definition process (and the Sphere indicators will be a major input to the HXL standards work), but as an information standard rather than a implementation standard, HXL is likely too different and too labour-intensive for the Sphere board to take on in addition to its current work.

Sub-Working Group on Emergency Telecommunications (WGET) — WGET¹⁸ is a forum within IASC (see above) focussed on telecommunications during disaster relief and mitigation. Since 2008, WGET has also accepted a broader mandate to work on information and communications standards and activities during disaster relief. This mandate is consistent with the goals of the HXL initiative (though WGET may be more interested in the infrastructure level of communications). WGET has two plenary meetings each year, and appears to work at a much-more informal level than standards bodies like the W3C (see below) or IATI; however, it is possible that WGET could expand its scope and formalize its governance to the point that it could manage a standard such as HXL in the future.

Information Management Working Group (IMWG) - the IMWG¹⁹ is an informal (non IASC-affiliated) UN-led group focussed on improving and monitoring information management during a humanitarian crisis. The group includes IASC members, cluster leads, individual humanitarian organisations, academic researchers, and non-traditional aid actors, depending on the topic under discussion. The group is comprised of mostly working-level information management staff at the global (Geneva/Rome) level. Face-to-face meetings are held once a month and focus mostly on information sharing. The group does not govern any standards but given its composition of key humanitarian partners, it would

quidelines humanitarian profile.pdf

^{13 &}lt;a href="http://www.humanitarianinfo.org/iasc/">http://www.humanitarianinfo.org/iasc/

¹⁴ https://cod.humanitarianresponse.info/

¹⁵ https://assessments.humanitarianresponse.info/files/iasc

¹⁶ http://www.sphereproject.org/

¹⁷ http://www.sphereproject.org/handbook/

¹⁸ http://www.humanitarianinfo.org/iasc/pageloader. aspx?page=content-subsidi-common-default&sb=13

¹⁹ http://imwg.humanitarianresponse.info/informationmanagement-working-group

be worth exploring whether the IMWG would be a suitable home for HXL in future years.

Open Government Partnership (OGP) — the OGP²⁰ is newer than IATI, and has a broader scope: its mandate covers all open public data (not just aid), and it has delegated aid-opendata to IATI. Like IATI, the OGP has good international visibility and buy-in, but it does not have a strong governance structure, and operates more as an advocacy group than a standards body. An affiliation with the OGP would give HXL some degree of recognition and credibility, but likely little actual governance support or sustainability.

World Wide Web Consortium (W3C) — the W3C²¹ is de-jure a vendor consortium of companies working with the web, but has become de-facto also an academic research consortium, since the vendors have often been happy to offload the difficult standards-making work onto "invited experts" (independent coninformationsultants and academic volunteers). The W3C was highly effective up until the time of the 2000 dot. com crash, but has gradually lost relevance since then (with the exception of their highly-successful push for "HTML5"); nevertheless, it remains custodian of many important standards. Given that most of the humanitarian stakeholders are not vendors, and that our timelines are relatively tight (we want to have a v.1 standard by the end of 2014), the W3C is likely not a good choice.

Organization for the Advancement of Structured Information Standards (OASIS) — OASIS²² was originally a trade organisation, but changed to a Web and XML standards group in 1998 in reaction to perceived problems with the procedural problems and slow progress of web standards work in the W3C. OASIS has much more transparent processes than the W3C and tends to finish standards faster, but it can also be dominated by vendors, as became clear during the Web Services work in the mid 2000s, where a few large vendors like Microsoft and IBM were essentially able to dictate terms to the organisation. OASIS provides a home and a stamp of approval for standards work, and has had some notable successes, but would has almost no visibility within the humanitarian world,

and would add limited value from a governance/sustainability perspective.

4.2. Coverage

The first important decision of the HXL Working Group will be coverage:

- What type of information should the first version of the HXL standard include?
- How much detail should the standard define for the information (e.g. disaggregated by type of need, gender, age, ethnicity, religion)?
- What time periods should the v.1 data cover (e.g. daily, weekly, monthly)?
- What level of geographical precision should the v.1 data cover (e.g. national, ADM1, ADM2)?

The humanitarian community already provides some guidance on the question of coverage. For example, the IASC's Humanitarian Profile²³ defines a well-understood and widely-used taxonomy for classifying people by their current circumstances and needs, reporting for the Consolidated Appeals Process (CAP)²⁴ includes a set of indicators to measure needs and response for each humanitarian cluster during a crisis, and the Sphere Handbook²⁵ defines a collection of indicators for measuring the effectiveness of implementation.

Standards for detail, time, and geographical precision are more difficult to find, so the working group may need to engage in significantly more community outreach to discover

²⁰ http://www.opengovpartnership.org/

²¹ http://www.w3.org/

²² https://www.oasis-open.org/

²³ https://assessments.humanitarianresponse.info/files/iasc_guidelines_humanitarian_profile.pdf

²⁴ http://www.unocha.org/cap/

²⁵ http://www.sphereproject.org/handbook/

the best balance between the effort to gather information and the value of the information to its users.

provide the higher-level semantic layer (e.g. indicators and entities). To be broadly useful as part of an open standard, HXL's base model and format must meet the following criteria:

4.3. Pilot countries

The programme team, in consultation with the working group and the technical team, will need to choose two countries for field pilots. The roadmap above suggests that the programme team have a short list ready by the end of March 2014, and that the pilots be finalized and ready to begin by June 2014.

The following are some of the considerations for choosing the HXL pilot countries:

- Geographic diversity: the two pilots should be in different regions.
- **IM/data capacity:** each of the pilot countries must have an OCHA office and humanitarian partners with the technical capacity to participate in the pilot.
- Potential for impact: the pilots should take place in countries where we can demonstrate a real impact through data sharing.
- Ease of travel access: if the pilots will require onsite visits, the countries must be safe for team members on the ground, and reasonably accessible within the project's limited travel budget.²⁶

4.4. HXL base model and format

The HXL Working Group will, fairly early on, need to choose a base model and format for HXL. The base model/format is the underlying framework for HXL; it specifies where characters and bytes go and provides a very low-level information model (e.g. columns or resources), while HXL will

- It is free and system-/vendor-independent.
- It is well-understood (at least in the IT world).
- It is well-supported by software tools.
- It has sufficient expressive power to represent humanitarian data.

There are four base model/formats that are reasonable candidates for consideration:

Comma-separated values (CSV)

pro: very well understood; good developer-level and consumer-level tool support (e.g. any spreadsheet program or relational database); very commonly used for public open data.

con: not as strictly-defined as other formats; can describe only tabular data; lacks formal schema/ validation support.

model: implies that data will be represented as multiple tables (similar to spreadsheets), related by shared keys.

Extensible Markup Language (XML)

pro: excellent developer-level tool support; moderately well-understood within the IT community; very commonly used for public open data; rich expressive power; properly-defined and -maintained international standard.

con: limited consumer-level tool support, and

²⁶ Note, however, that we may choose to carry out a pilot in a country that we cannot visit, specifically to test the feasibility of that approach.

virtually no consumer-level understanding; too complex for some IT teams.

model: implies that data will be represented as hierarchical trees (similar to a filesystem on a hard drive), optionally linked by shared keys.

hard drive), optionally linked by shared

) JavaScript Object Notation (JSON)

pro: very popular among web developers; moderately well-understood within the broader IT community; rich expressive power.

con: no consumer tool support at all; almost never used for public open data; virtually unknown outside the web-development community; lacks formal schema/validation support; limited tools for transformation and querying; not a properly-maintained international standard (though strictly defined through an adhoc consensus specification).

model: implies that data will be represented as nested arrays (similar to trees).

Resource Description Framework (RDF)

pro: occasionally used for public open data; good developer tool support for validation, querying, and transformation; some (very limited) tool support for consumer browsing; rich expressive power; properly-defined and -maintained international standards.

con: much less-well understood than the other formats in the broader IT community; virtually no familiarity outside of the IT community; supports multiple representation formats, increasing complexity and costs for information consumers; very limited software support compared to CSV or XML (though better than JSON); uses an advanced semantic model that

is too complex for most users to understand; used much less often than CSV or XML for open public data.

model: implies that data will be represented as directed graphs, similar to linked web pages.

It is, of course, possible to allow multiple base formats (e.g. CSV and XML), but doing so is generally a misguided choice: while a selection of base formats makes life easier for information producers (who can support any format), it makes life harder for information consumers (who must support all formats).

5. PROPOSED WORKING **GROUP CHARTER**

his section contains a draft charter for a HXL working group, describing the scope of its activities and deliverables, and how it will make decisions. This is simply a sample; OCHA and other stakeholders will need to approve the actual charter as the first step of standards work.

The Humanitarian Exchange Language (HXL) Working Group is a team of humanitarian stakeholders, led by OCHA in its aid-coordination role.

5.1. Mission

The mission of the HXL Working Group is to create consensus standards to support the exchange of up-to-date operational data among stakeholders during a humanitarian crisis (initially focussing on protracted crises).

5.2. Scope

The HXL Working Group is responsible for creating and approving the standards and for consulting with the broader humanitarian community to collect requirements and feedback. The Working Group will coordinate with other teams inside UN-OCHA as they work on field pilots for the HXL standards.

5.3. Deliverables

- One or more normative HXL standards documents.
- Supporting technical artifacts (such as schemas and controlled-vocabulary lists).
- Non-normative user documentation, such as examples and tutorials.

5.4. Expected Milestones

- March 2014: first (alpha) version of the standard(s), possibly incomplete.
- June 2014: second (beta) version of the standard(s), complete.
- September 2014: third (stable release candidate) version of the standard(s).
- **December 2014:** final (v.1.0) release of the standard(s).

5.5. Duration

The HXL Working Group shall begin work in January 2014, and continue in the same form until the end of the year. By the end of 2014, it will have developed a plan for its renewal in 2015 and continuing sustainability.

5.6. Participation

The HXL Working Group will consist of 5-7 members, including a chair and a technical lead. The membership will represent at least five different humanitarian organisations (including UN-OCHA), with at least one representative from an NGO.

5.7. Meetings

The HXL Working Group will have a weekly one-hour teleconference, and two face-to-face meetings (timed to coincide with other humanitarian gatherings) over 2014.

5.8. Communication

In addition to its weekly teleconference, the HXL Working Group will have a mailing list for internal work and broader community outreach. All interested parties will be welcome to join the mailing list.

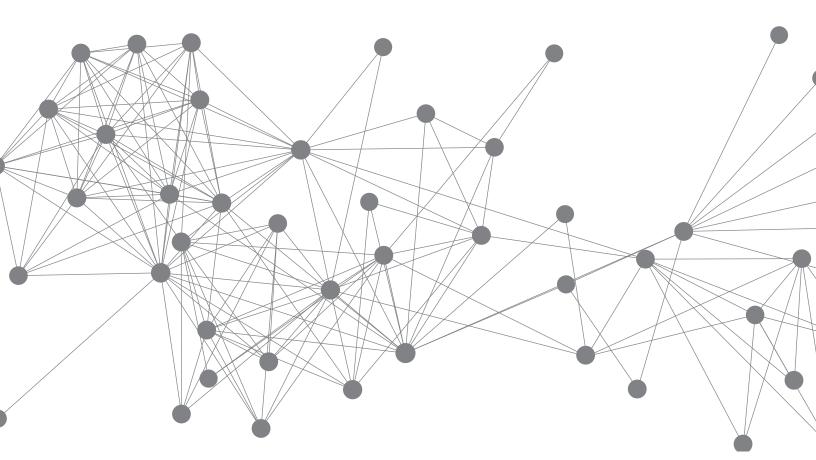
The Working Group will also have web space for draft standards, etc.

5.9. Confidentiality

All activities of the Working Group, including its mailing list and minutes of its weekly meetings, will be publicly available in the interest of transparency.

5.10. Intellectual Property

All work of the HXL Working Group, including standards documents, schemas, and other artifacts, will be released as open content (possibly into the public domain). All working group members will signal their agreement to this policy prior to beginning work.





STAY IN TOUCH

Email: ochadataproject@un.org

GitHub: @OCHA-DAP **Twitter:** @humdata

Site: hxl.humanitarianresponse.info