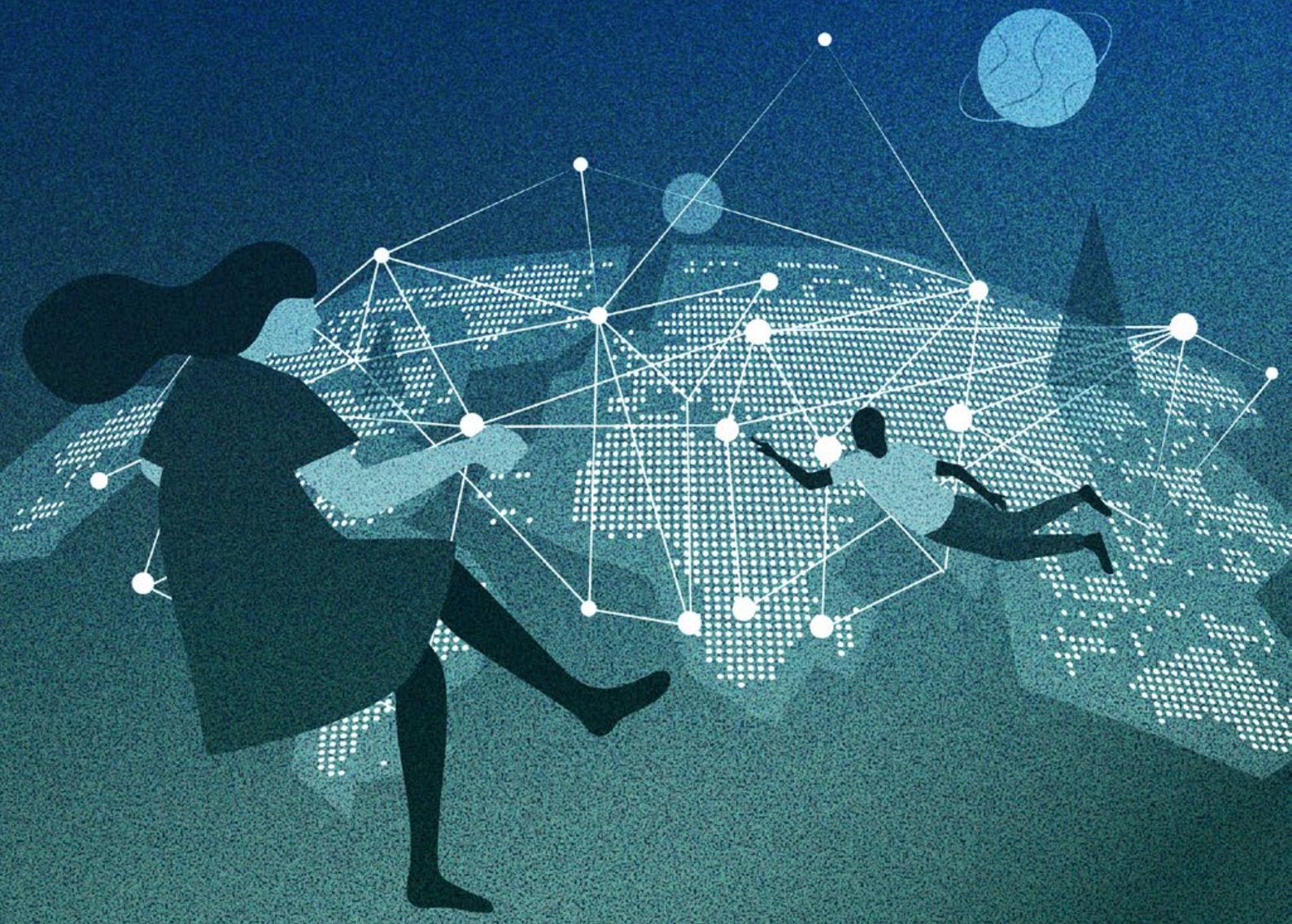
Identifying places:

Improving data connectivity through unique identifiers



October 2022

DH Network
Digital Humanitarian Network

Identifying places:

Improving data connectivity through unique identifiers

By

Elaine Donderer (donderer@homareto.com | @elainedonderer | LinkedIn)
Homa Reto GmbH - The Human Network

Tom Haythornthwaite (haythornthwaite@un.org | @tomhaythornthwaite | LinkedIn)
United Nations Office for the Coordination of Humanitarian Affairs (UN-OCHA)

Andrej Verity (verity@un.org | LinkedIn)
United Nations Office for the Coordination of Humanitarian Affairs (UN-OCHA)

Design

Alexandra Sternin | alexsternin.com

This document was made possible with the support of

OCHA

Table of Contents

Interviewees

Name	Role/Title	Organization/ Affiliation
Alan Mills	Owner of Alan Mills Consulting Ltd. & Volunteer	MapAction
Andrew Alspach	Chief of the Information Management Branch	UNOCHA
Anita Russo	Project Lead for GIS Visualisation Program	ITOS
Catherine Lefebvre	Consultant	WHO
Craig Williams	Freelance Consultant	—
Dan Runfola	Professor	College of William & Mary
David Collins	Government GIS Analyst & Volunteer	MapAction
Elsa Richard	Data Analyst & Visualizer	UNOCHA
Former employee	Information Manager	UNOCHA
Guillaume Le Sourd	Geospatial Officer	UN Cartographic
Ivan Gayton	Senior Humanitarian Adviser	HOT OSM
Jan Burdziej	GIS Consultant	UNICEF
Janet O'Callaghan	Chief of Field Information Services	UNOCHA
Jolynn Schmidt	Project Manager for GRID3	CIESIN
Kristina MacKinnon	Information Management Officer/Geographic Data and Systems Alignment Focal Point Expert	UNHCR

Name	Role/Title	Organization/ Affiliation
Kytt MacManus	Senior Systems Analyst	Columbia University & NASA
Leo Martine	Mission Information Management Officer	UNHCR
Luke Caley	Information Management Lead	IFRC Shelter Cluster
Mansuk Daniel Han	Technical Specialist & Population Data Analyst	UNFPA
Max Malynowsky	Software Developer	REACH Initiative & UNDP
Naomi Morris	Consultant & Information Management Officer	WHO
Neil Bauman	Shelter Adviser & Information Manager	Canadian Red Cross
Nick McWilliam	Preparedness Coordinator	MapAction
Olivier Cottray	Director of Humanitarian Solutions	ESRI
Dr. Prashant Madhukar Hedao	Consultant	WHO GIS Centre for Health
Remi Daniel Galinier	Data Education Analyst	UNHCR
Roberto Colombo	Information Management Officer	UNOCHA
Roxana Nazarie	Database Management Specialist	CartONG
Steve Penson	Data Scientist & Volunteer	ACAPS & MapAction
Tom Haythornthwaite	Geographic Information Officer	UNOCHA

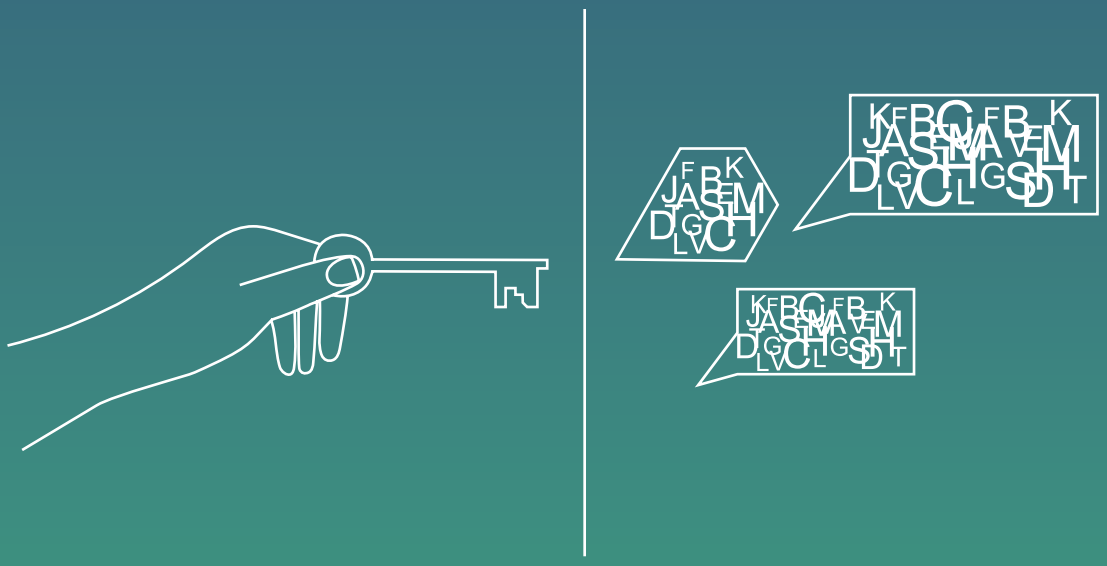
Acknowledgements

We extend our gratitude to UNOCHA for making this paper possible. We are particularly grateful to all interviewees who took their time to meet with us and share their thoughts and expertise. In bringing together all these viewpoints, the authors' overall hope is to help improve information management and exchange through increased leverage of geospatial data in humanitarian relief missions and beyond.

This paper does not represent the views of the interviewees unless explicitly stated. The authors have drafted the paper based on a combination of literature review and interviews.

Key messages

- Coding administrative boundaries eases the coordination of humanitarian agencies and resources during an emergency event.
- Integrated coding processes, with a shared methodology, are more efficient. The governance and maintenance of unique geographical identifiers must be expanded to an inter-agency effort.
- Meaningful collaboration with national Governments and international development organizations can provide coded spatial boundary data for humanitarian relief. Accrediting information sources for data ownership is important to establish these relationships with an anti-colonial approach.
- Unique identifiers can benefit from dynamic elements, for example, documenting the data's temporal lineage.
- Alternative ways of coding geographical boundaries, such as automated processes with machine learning, provide exciting opportunities for the future and should be further explored.



Definitions & Terminology

Geographic Information Systems (GIS): Computer or web-based systems that create, organize, visualize and model spatial positions, patterns and relationships.[1]

Geospatial data: Any data or information that relates to a specific location on the Earth's surface that is or has been generated by, transmitted via, received by, processed by or represented in a GIS resource.[2]

Data interoperability: The process of correctly sharing and interpreting data that crosses from one information source to another (e.g., from one cluster to another, from a cluster to national authority, etc.). "For example: if we assume that the A cluster has information needed by the B cluster, and that data in one cluster system is accessible and understood by the other system, then data can be compared."[3]

Common Operational Dataset (COD): An authoritative data framework that improves interoperability. CODs simplify data exchange and are essential to improve response capacities in an emergency.[4]

Geographic Identifiers (GEOIDs): A variable consisting of alpha-numerical values that are assigned to specific geographic areas. This method is commonly used for a country's national census and facilitates the organization and exchange of information.[5]

P-codes: Unique identifiers that code each administrative unit and populated place. They provide the means to manage and exchange data more easily.[6]

Gazetteer: A geographical index that contains a directory of place names. It may include additional information about a location, including P-codes, administrative feature type, social statistics and physical features of a country, region or continent, alternate spellings or language, and the identity of hierarchically parent features. A gazetteer is typically used in conjunction with a map or an atlas.[7]

1 What is Geographic Information Systems (GIS)? Accessed on 15 September 2022. gisgeography.com/what-gis-geographic-information-systems/

2 What is geospatial data? Accessed on 22 August 2022. www.ibm.com/topics/geospatial-data

3 IASC (2010, p.2). Guidelines Common Operational Datasets (CODs) in Disaster Preparedness and Response.

4 Common Operational Datasets. Accessed on 15 September 2022. cod.unocha.org/

5 Understanding Geographic Identifiers (GEOIDs). Accessed on 15 September 2022. www.census.gov/programs-surveys/geography/guidance/geo-identifiers.html

6 P-codes and gazetteers. Accessed on 3 August 2022. humanitarian.atlassian.net/wiki/spaces/imtoolbox/pages/222265609/P-codes+and+gazetteers

7 Goodchild, M. F., & Hill, L. L. (2008). Introduction to digital gazetteer research. International Journal of Geographical Information Science, 22(10), 1039-1044. Doi: 10.1080/13658810701850497

Methodology

To answer the question around improving data connectivity through unique identifiers, the authors relied on a two-tiered approach using a purposeful literature review and semi-structured interviews. The qualitative methodology serves to understand and interpret context-bound realities and perceptions in the humanitarian field.[8] By interviewing a wide range of experts in the field, the research synthesizes different viewpoints from the global geospatial community. The interviews' semi-structured nature allows for open-ended discussions that can bring in new themes and related concerns or ideas. Hence, applying data collection instruments sensitive to underlying meanings offers a holistic perspective on the barriers for universal identifiers.[9] Thereby, this paper showcases missed opportunities as well as success stories and, if applicable, it proposes new ways to code and join boundaries for information sharing. The constructivist lens enables an in-depth analysis of how different actors construct their data and respective spatial boundaries.

Due to the complexity of international data-exchange processes and administrative landscapes, this methodology is found to be the most appropriate for generating strategies for improved usage of unique identifiers such as P-codes.

The research limits its scope regarding addressing alternate geospatial coding systems in the humanitarian sector, as well as their adoption in the development field. We encourage further review or case studies.

8 Merriam, S. B., & Tisdell, E. J. (2015). Qualitative research: a guide to design and implementation. John Wiley & Sons.

9 Gray, D.E. (2014). Doing Research in the Real World. London: SAGE.

Introduction

Within the UN system, the United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA) is mandated to develop and maintain baseline data on areas affected by humanitarian crises.[10]

At the onset of any disaster, information management (IM) is a critical task that helps to provide an overview of needs and resources, and to coordinate humanitarian and development organizations. Thus, a pre-existing information infrastructure is vital for delivering reliable and timely spatial and thematic information to responders.[11]

A baseline component of datasets for humanitarian coordination is the use and exchange of geographic data and relationships. The principal element for geospatial data interoperability is an agreed methodology or approach in coding geospatial boundaries.[12] This enables the joining and exchanging of data across agencies, thereby circumventing a waste of resources, duplicated efforts and missed opportunities in saving lives.[13]

The P-code, or "place code," is among the most well-known attributes for linking the administrative boundaries, population statistics and other thematic datasets. P-codes are unique identifiers that code each administrative unit and populated place, allowing the creation of geocoded gazetteers of administrative units and related facilities.[14] They provide the means to manage and exchange data conveniently.

10 World Health Organization (2007). "Country Level Minimum Common Operational Datasets" (PDF). Archived from the original (PDF) on 27 October 2013. Retrieved from: web.archive.org/web/20131027063940/http:/www.who.int/hac/network/global_health_cluster/country_level_minimum_common_operational_datasets.pdf

11 Ulgen, S. & Williams, C. (na.). Standardization of Geographic Names in Humanitarian Information Management (Towards a Humanitarian Spatial Data Infrastructure).

12 Woldeghebriel, T. T. (2021). Common Operational Datasets (CODs) in Humanitarian Emergency Preparedness and Response: Framework for Sound Information Management. Master Thesis.

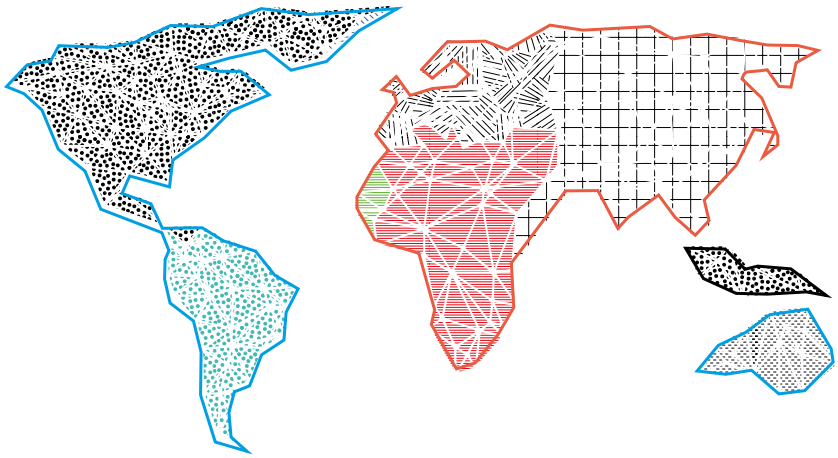
13 D. Guha-Sapir, R. Below, P. Hoyois, Data on disasters: Easier said than done, Joint Session of Executive Board - UNICEF, UNDP, WFP, UNFPA, New York, 19 January 2006.

14 P-codes and gazetteers. Accessed on 3 August 2022. humanitarian.atlassian.net/wiki/spaces/imtoolbox/pages/222265609/P-codes+and+gazetteers

> *No amount of money spent on complex GIS systems would be worth its value if there is no unique key to tie it all together – that is, the P-code."*
>
> **— Mansuk Daniel Han, Technical Specialist & Population Data Analyst, UNFPA**

Nevertheless, P-codes are only one solution for overcoming the difficulties of data exchange across a given response. With different organizations using their own unique identifiers for their datasets, hurdles between global, local and national geospatial datasets have arisen in recent years.[15] Another challenge is documenting the lineage of unique identifiers and creating inclusive processes and ownership for Governments and development organizations beyond the humanitarian field.[16]

This paper hopes to aid information managers and humanitarians in developing a joint system for identifying places. Due to the paper's scope, the authors focus on the evolution of coding systems for geospatial boundaries and associated P-code attributes in providing a framework for linking relevant datasets. The paper also discusses the barriers and experiences of professionals in the field when applying these systems. Therein, this study may support an information environment in which many organizations and relevant actors can contribute to and benefit from geographic information and improved data connectivity



15 Interview with Olivier Cottray.

16 Interview with Naomi Morris.

The History of P-codes

In understanding the present landscape and discussing the future of GEOIDs, the following section looks at the origin of P-codes to showcase the often-operational nature of unique identifiers.

In the early 2000s, IM was moving towards the digital format. Although not as new as the digital concept, humanitarian information managers were increasingly adopting and expanding on the possibilities of modern GIS systems. The ability of computer systems to support IM was promising, and the search for digital tools to facilitate humanitarian response greatly expanded. Notably, the process of developing a digital toolkit for humanitarians was often led by on-the-ground decision makers rather than GIS specialists and data scientists.[17] As part of this toolkit, P-codes showed considerable outcomes in improving the quality of information for relief missions,[18] especially those in places that have many languages, spellings, translations and anglicized versions of local languages, all of which can lead to confusion, duplication of efforts and a waste of desperately needed resources.[19]

The history of P-codes goes back to the Kosovo War in the Balkans, from 1998–1999, when unique identifiers became necessary to coordinate relief operations and conduct rapid needs assessments.[20] Humanitarian staff had to leave the country, together with the refugees, and they conducted their work for Kosovo from outside the country.

During a five-month period, humanitarians came together in Macedonia and scoped out what the ideal situation would be upon their return to Kosovo. Geospatial data collected beforehand became the basis for a data system to coordinate all the different agencies.[21] In these datasets for multi-cluster rapid assessments, field names were titled M-code for municipalities, and P-code for populated places. The concepts were used to describe the idea of a common set of data standards to consolidate inter-agency data.[22] Around 140 organizations gathered data through this rapid assessment form, using centralized GIS and P-codes to compare, share and cross-analyse their information.[23] This success story illustrates that P-codes are simply a brand for unique identifiers that originated as an inter-agency effort for improved data connectivity.

17 Interview with Neil Bauman.

18 Williams C., & Marinos J. (2011). GIS Enables the Humanitarian Response: A Perspective from the United Nations. Retrieved from: www.directionsmag.com/article/1932

19 Interview with Dr. Prashant Madhukar Hedao.

20 Interview with Andrew Alspach.

21 Williams C. (2001). Data Coordination at Work Coordinating Housing Damage Surveys in Kosovo.

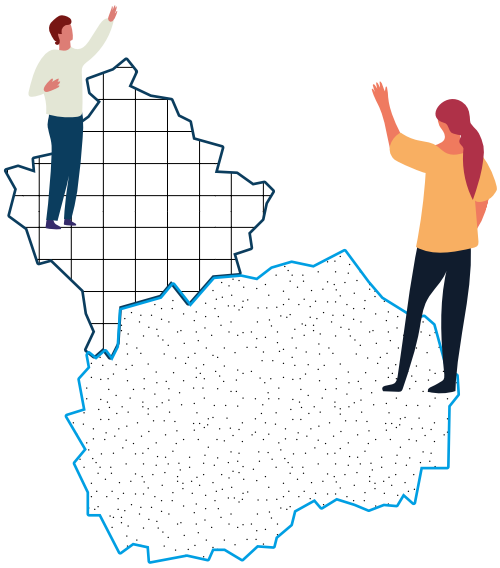
22 Interview with Craig Williams.

23 Smith. D. (2001). Kosovo: Applying GIS in an International Humanitarian Crisis. UN GIS Consultant in Bosnia, Kosovo, Macedonia and Albania between July 1997 and December 1999.

Despite the collaborative environment of P-codes' origin, thereafter the codes were often implemented in relatively closed systems, stayed in organizational silos and too often developed only at the immediate onset of a disaster. Through the advocacy and training by UNOCHA and MapAction, and the Inter-Agency Standing Committee's *Guidelines Common Operational Datasets (CODs) in Disaster Preparedness and Response*[24], the IM community has gradually increased its awareness and use of P-codes in recent decades.[25] With increasing preparedness measures in place, as showcased by the 2006 earthquake in Yogyakarta, Indonesia, P-codes have found their way into the mainstream discourse on resilience strategies.[26]

Further improvements have been made since the establishment of the Humanitarian Data Exchange [platform] and the incorporation of P-codes into selected datasets.[27] P-codes are now consistently found in Administrative Boundary CODs (COD-ABs) and Population Statistics CODs (COD-PSs). Their use is encouraged for alternate geographic data sources and any potentially associated humanitarian dataset.

Nevertheless, P-codes remain predominantly used by humanitarians rather than being owned and developed by Governments and development organizations in advance of a crisis. Numerous practitioners criticize a lack of inter-agency inclusion in the process of establishing P-codes, and they point out that communities may be further marginalized if not accounted for correctly.[28] Thus, it is paramount to always consider the political dimension of defining spaces and geographic boundaries. Analysing success stories and missed opportunities of unique identifiers can help navigate this process. Although P-codes are not a one-size-fits-all approach, there seems to be a consensus that lessons learned from their application can draw a map for a joint approach to inclusive and efficient geospatial data exchange.



24 IASC (2010). Guidelines Common Operational Datasets (CODs) in Disaster Preparedness and Response.
25 Interview with Nick McWilliam.
26 Interview with Neil Bauman.
27 Interview with Leo Martine.
28 Interview with Alan Mills.

P-codes Today: Lessons Learned

What is a P-code?

P-codes are generated only if a national coding scheme is not already available, or if no informal coding scheme has been adopted by national or international organizations.

The schematic below shows a basic structure of a P-code system for a country with four administrative boundary levels (see figure 1).[29] For further information, see the official P-code guidance.

Country code - This is usually taken from the ISO 2-letter country code standard. This makes the P-code recognizable as text rather than a number, which ensures that leading zeroes are not dropped, and it reduces the risk of database key incompatibility. The country code also ensures P-code uniqueness when merging databases.

Administrative level 1 to Administrative level X - Each administrative level receives a numeric code string with enough character width to ensure that any adding of administrative units will not exceed the number of codes available. In the figure below, two digits are adequate for administrative levels 1 - 3.

Incremental Settlement Numbers (if required) - For maintenance, it is best to use a single set of incremental numbers across the whole country. This allows places to be referred to by settlement number without the admin unit codes, making them shorter. Enough digits should be used to allow for the possibility of adding new datasets.

29 P-codes and gazetteers. Accessed on 3 August 2022. humanitarian.atlassian.net/wiki/spaces/imtoolbox/pages/222265609/P-codes+and+gazetteers

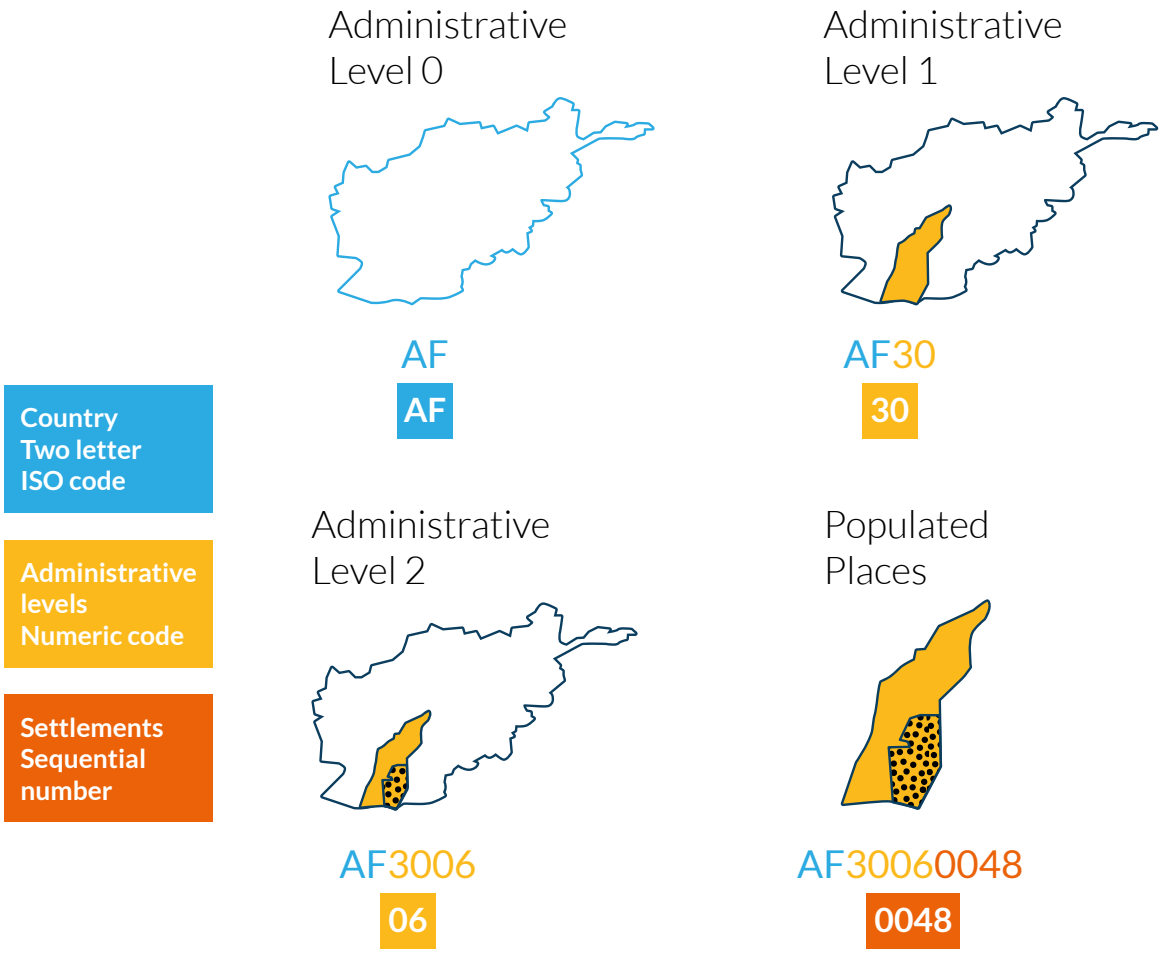


Figure 1: Structure of the P-code[30]

When working with unique identifiers today, organizations may use different coding systems to serve their missions and mandates. For example, UNHCR works with layers that contain only administrative boundary levels 0, 1 and 2 in its GIS system. These boundary codes are predominantly sourced from Governments and other official sources, such as national cartographic agencies, other UN bodies or NGOs. In compiling its codes, UNHCR considers the COD P-codes, but it is more focused on populated places as to understand the movement of people. All the data is linked back to a source code and is publicly available.[31] As with UNOCHA, UNHCR sometimes adopts operational boundaries that are mission-driven and therefore do not imply official endorsement or acceptance by the United Nations.

The branches and local units of the International Federation of Red Cross and Red Crescent Societies (IFRC) within a country can span multiple administrative boundaries. Therefore, IFRC requires a parallel data infrastructure that can fit national administrative structures. When conducting assessments and

30 P-codes and gazetteers. Accessed on 3 August 2022. humanitarian.atlassian.net/wiki/spaces/imtoolbox/pages/222265609/P-codes+and+gazetteers

31 UNHCR Core GIS Data. Accessed on 15 September 2022. data.unhcr.org/en/geoservices/

building surveys with KoBoToolbox, IFRC's information managers work with a P-code look-up table to conveniently associate data. Frequently, the P-codes are sourced from MapAction, but they need to match the P-codes of national societies and Governments. Hence, although IFRC has its own unique identifiers for administrative boundaries, P-codes can be referenced to ensure interoperability.[32]

The unique identifiers of the World Health Organization (WHO) are similarly required to locate information in any humanitarian mission. Although WHO works with the P-code, administrative units for health facilities do not always align with the administrative boundaries referenced by the P-code. Health districts can change frequently, particularly in areas with a strong population growth, and they do not remain as static as the administrative boundaries. The WHO is working on releasing geodatabases with global health boundary information, with a special focus on projects related to polio eradication and African health boundaries. The WHO AFRO GIS Centre is a driving force in creating, documenting and updating these changes of geographical units.[33] These efforts helped, inter alia, with the 2018 Cyclone Idai response in Mozambique.

The related databases are designed to handle temporal changes, but implementation has been limited, thereby demonstrating the challenge of collecting and maintaining such granular information. In most real-life cases, there is a lack of time stamps. The WHO is also pursuing work with authorities on generating master lists for health facilities, which will all require unique identifiers.[34]

Notably, it is not only international organizations that use unique identifiers to ensure data connectivity; national Governments and agencies, local NGOs and even the private sector use them. There can be conflicting identifiers and duplicated efforts, but there is a consensus that the value added by GEOIDs can facilitate data linkages and help structure spatial frameworks. As stressed by Dr. Hedao, not only will we need to bring data together but also people from different disciplines. Hence, **finding an inter-agency adapter to link these different coding schemes in a collaborative manner will be paramount in maximizing their benefits.**[35]

> *"GIS is not just a tool to bring data together but to bring people together and break down the silos. It builds data relationships between disciplines."*
>
> **— Dr. Prashant Madhukar Hedao, Consultant, WHO GIS Centre for Health**

32 Interview with Luke Caley.

33 AFRO GIS: Reaching All Populations. Accessed on 18 August 2022. storymaps.arcgis.com/stories/3a26d81380234693a8ed2c0c38342ac5

34 Interview with Naomi Morris.

35 Interview with Dr. Prashant Madhukar Hedao.

Success Stories

A wide array of success stories illustrates the benefits of P-codes. As discussed above, gazetteers, or lists of P-codes and place names at all administrative levels, can structure relief activities such as assessments, distributions and activity monitoring, even without anyone drawing a map. Hence, they become key to creating innovative, multidimensional analyses for solving real-world problems. For example, if one dataset holds the location of health clinics and the services they offer while another holds information on health-service utilization, those P-coded datasets could be superimposed onto COD-AB, COD-PS and local infrastructure mapping data to allow a geospatial assessment of people's access to health services and what can be improved in terms of services.[36]

Another health-care application is the patient-origin problem. Frequently, registries of health-care facilities contain labels with the patient's origin entered as a free text. These labels tend to be phonetically transcribed and do not indicate a category or administrative division. In this case, P-codes can solve the patient-origin problem by differentiating village names. Cultures usually have some way of describing geography more or less unambiguously through language; a universal translation of these languages and human descriptions can be the hierarchical division reflected in the P-code.[37]

The 2006 Yogyakarta earthquake is a textbook example of P-coded CODs as a preparedness measure. Shortly before the earthquake, UNOCHA, in collaboration with the national Government, published up-to-date CODs on the region, ensuring efficient rapid assessments and mapping of housing damage. An atlas was published showing the situation before, immediately after and month by month for reconstruction. Visual representation of the disaster and how the response operation addressed it eased information access for a variety of relief and reconstruction operations.[38]

IFRC also records positive experiences with P-codes. In collaboration with REACH, MapAction, Médecins Sans Frontières and some UN partners, IFRC carried out needs assessments to inform the Instituto de Meteorologia e Geofísica, in Cape Verde. The datasets used P-codes as spatial data infrastructure, which helped to make more refined estimates of the needs but also of the affected population in need. Even in such successful cases, people in informal settlements remain a vulnerable demographic who can suffer from undercounting. Nevertheless, the system enabled an association with the most recent census data, aggregated at different administrative levels using WillPop, and it supported relief missions across organizations.[39]

36 Interview with Mansuk Daniel Han.
37 Interview with Ivan Gayton.
38 Interview with Craig Williams.
39 Interview with Luke Caley.

Success stories between sectors and organizations are particularly interesting for the evolution of inter-agency coordination. Learning from these cases can improve the information landscape.[40] During the 2019 flood response in Djibouti, existing mapping was flagged as incorrect. Consequently, professionals in the field connected with local people on the ground and established national-regional districts to improve the P-coding. Characterized by the emergency's urban environment, maps were coded down to administrative level 5 or 6. Coding with granularity down to the neighbourhood level is rare, but it was constructive in this setting.

Towards the end of MapAction's mission in Djibouti, the World Bank came in for post-disaster needs assessments (PDNAs) and adopted the P-code system with all the relevant maps.[41] This case exemplifies continuity from response to early recovery. Organizing field teams around recognized ward boundaries and data from different sectors simplified logistics and cross-analysis.[42]

The military conflict in Ukraine, escalating in 2022, is a recent case showing successful inter-agency coordination and application of P-codes as unique identifiers. Ukraine is a relatively big country with multiple locations using the same name. For instance, there are two places named Mariupol. In some cases, newly deployed international actors have been looking at the wrong location when planning humanitarian aid. Moreover, the crisis response in Ukraine is characterized by the involvement of many international, national and local actors, and with a geography that can be difficult to understand and communicate. By adding a condition to baseline datasets that contain P-code attributes, data exchange was simplified, and coordination improved.[43]

Another helpful factor was the initial global support and significant funding for IM.[44] As a result, information managers had the capacity to update and create codes for administrative levels 1 and 2 for Ukraine and neighbouring countries, and to make the data accessible on the Humanitarian Data Exchange (HDX) platform to address displacement within one day.[45] Nevertheless, improvements can still be made, especially for very low-level administrative boundaries (e.g., point data). In the case of Ukraine, a long string of characters could have been beneficial, associating individual humanitarian service points, such as psychosocial support from IFRC.[46]

Comprehensive, complete and consistent global P-code adoption has supported both national and cross-border humanitarian coordination. One can safely assume that it will continue to do so in the future.[47]

40 Interview with Roxana Nazarie.

41 From challenges to opportunities: Rapid Needs Assessments as entry points for building resilience. Accessed on 6 September 2022. blogs.worldbank.org/arabvoices/challenges-opportunities-rapid-needs-assessments-entry-points-building-resilience

42 Interview with Alan Mills.

43 Interview with Elsa Richard.

44 Interview with Janet O'Callaghan.

45 Interview with Leo Martine.

46 Interview with Luke Caley.

47 Interview with Steve Penson.

Missed Opportunities

Despite the above-mentioned benefits of P-codes, the remaining difficulties in data exchange across organizations' missions and mandates must be addressed. These barriers can be conceptual, technical, political and organizational.

Problems Identified for Geospatial Data Connectivity

Difficulties in data exchange across organizations' missions and mandates, as identified in the interviews, are perpetuated by the following barriers:

a. Lack of documentation of the temporal lineage of codes.
b. Different coding systems and granularity of boundaries and contained data.
c. Joining data when there are different polygons for boundaries/geospatial units (e.g., health boundaries vs. administrative boundaries) is time consuming.
d. No clear indication if datasets contain unique identifiers on HDX.
e. Difficulties in inter-agency coordination and in-house data systems lead multiple organizations to use different coding schemas.
f. The added value of unique identifiers is not clear to many professionals outside the data management community.
g. Limited adoption of P-codes beyond UNOCHA's CODs.
h. Challenge of generating P-codes in an efficient and inclusive manner.
i. Imposing top-down unique identifiers can further marginalize communities if geographies are not defined at the right level.
j. Lack of accreditation of information sources.
k. Lack of national ownership and maintenance of administrative boundary codes.
l. Discrepancies between operational and official data.

First, barriers for geospatial data connectivity can be conceptual. One major limitation with P-codes is that they do not contain any indication of their temporal lineage.[48] There are arguments for the benefits of static P-codes, but in an environment of changing boundaries, as outlined in the context of health and population dynamics, temporal information can have an advantage. For instance, an archive of P-codes could indicate a timeline with an end date for the code to show when new P-codes were established.[49] Another option is to insert a four-digit code of the year of origin into the code itself (e.g., AA0012022) and archive the codes annually.

48 Goodchild, Michael & Yuan, May & Cova, Thomas. (2007). Towards A General Theory of Geographic Representation in GIS. International Journal of Geographical Information Science. 21. 239-260. 10.1080/13658810600965271.

49 Interview with Dr. Prashant Madhukar Hedao.

In some countries and regions, such as the Sahel, which has a high birth rate, the health districts are modified every year to fit the new population distribution. In Chad, the dataset for health facilities and districts are updated according to these annual changes.[50] Not all district boundaries change, but one or two districts might be split every year.

Places with unique identifiers must become resilient to territorial changes over time.[51]
Inserting the year into the code helps to ensure use of the most recent reference and links other data to this shapefile for mapping purposes.[52] Any complementary GPS data can then avoid further confusion with nearby health facilities.

Second, technical challenges arise in applying P-codes. In achieving data connectivity, the conditions of good spatial data and inter-agency efforts must be fulfilled. Agencies often come up with their own unique identifier coding systems, as it may be the fastest option. For instance, for agencies such as UNHCR, the P-codes available on HDX lack granularity, as administrative levels beyond administrative level 3 are rare. Therefore, UNHCR uses its own coding scheme. Hence, including populated places down to a village level is a missed opportunity in the CODs on HDX for the context of migration and displacement.[53]

In other cases, coding systems by the private and humanitarian sectors compete.[54] As seen during the 2020 Beirut port explosions, such competition can lead to confusion regarding which coding system to use, thereby affecting overall relief efforts.[55] Furthermore, as P-codes constitute an artificial layer of geographic reference, disagreements over location can waste time and even exacerbate security issues.[56]

Third, P-code-related issues have remained limited to the field of technical workers and IM officers, which constitutes an organizational challenge. Rather than solving problems at the source, symptoms are addressed without reorganizing data sourcing and establishing inter-agency communication. Increasing the advocacy and effort to align and change systems will eventually reduce data-cleaning processes and improve information responses. Many agencies use look-up tables as a workaround method, but they are no cure for the lack of inter-agency coordination. Until today, P-codes are rarely understood or used outside the IM community.[57]

50 Chad - List of health facilities and health districts. Accessed on 15 September 2022. data.humdata.org/m/dataset/chad-list-of-health-facilities-and-health-districts

51 Interview with Jolynn Schmidt.

52 Interview with Remi Daniel Galinier.

53 Interview with Kristina MacKinnon.

54 Interview with Roberto Colombo.

55 Interview with Steve Penson.

56 Interview with Olivier Cottray.

57 Interview with Janet O'Callaghan.

> *When your computer works you do not even notice it, but when it breaks down it can create big problems. Meaning, we do not always take stock of the fact that the P-coding system has worked, because we take it for granted."*
>
> **— Janet O'Callaghan, Chief of Field Information Services, UNOCHA**

Perhaps the greatest organizational difficulty is that humanitarians aspire to act fast, which can sometimes be to the detriment of more long-term and sustainable approaches. The issue of operational versus official data extends into establishing the responsibility to create and maintain unique identifiers. Tools such as a resource directory and contact list for generating P-codes would be helpful,[58] but the core challenges may be cured only through a global P-coding community that facilitates fast coordination channels.[59]

Another potential solution is capacity development, as illustrated by the 2015 earthquake in Nepal, when knowledge transfer was ensured beyond UNOCHA's mandate through preparedness initiatives.[60] Development organizations play a vital role in filling these gaps and establishing partnerships for disaster preparedness and embedded IM.[61] Further work is required on how to implement these strategies with a joint initiative.

Fourth, as P-codes de facto define spatial boundaries, the politics around disputed areas and electoral systems must not be overlooked. Professionals point out a lack of political will or expertise, as unique identifiers are not understood as a valuable contribution or top priority in a national political agenda. Compounding that, data sensitivity remains a barrier to collaboration, as public data in the context of military conflicts can constitute an additional danger for vulnerable populations.[62] Therefore, ethical data governance must be at the core of P-coding practices.

On the other hand, the people and national agencies on the receiving end of humanitarian aid can be overlooked and further marginalized by exclusionary IM practices. In some cases, Member States complain about representation and require diplomatic intervention. Harvesting datasets from countries without crediting, consulting or establishing ownership seems inconsistent with the values of the UN system and humanitarian principles. National cartographic agencies have complained about the

58 Interview with Jolynn Schmidt.
59 Interview with Janet O'Callaghan.
60 Interview with Kristina MacKinnon.
61 Interview with Roberto Colombo.
62 Interview with Naomi Morris.

lack of collaboration and feedback on the data. Such behaviour is not a sustainable practice for good governance.[63] Most importantly, establishing and maintaining long-lasting relationships and meaningful inter-agency collaboration can support reconciliation between the global geospatial community and national initiatives.[64]

In summary, P-codes, so far, are not a one-size-fits-all solution. Questions remain on granularity, temporality, responsibility and the politics of unique identifiers for spatial boundaries. Complementary approaches with systems such as what3words may bridge some of these concerns in the future.[65] But regardless of potential future technologies, an inter-agency framework will be indispensable in achieving an anti-colonial, innovative and effective strategy for data connectivity.

P-codes in the Future

what3words

what3words assigns a unique three-word address to every 3x3 m^2 worldwide. The words were randomly assigned to each square, and they never change over time. The service is currently available in 50 languages and can help to find and share any place with just three words.[66]

In the past, practitioners used what3words complementary to P-codes, for instance, to refer to the location of infrastructure, such as hospitals.[67] Although a joint approach shows promise, reservations about the lack of connection between what3words' locations and administrative units must be considered, as well as potential language barriers in joining attribute tables with what3words.

63 Interview with Guillaume Le Sourd.
64 Interview with Nick McWilliams.
65 Interview with Olivier Cottray.
66 what3words. Accessed on 15 September 2022. what3words.com
67 Interview with Olivier Cottray.

Raster Coding

Similar to what3words, grid cells can divide the Earth's surface into uniform and closely adjacent grid arrays. Rasters are the simplest and most direct spatial data structure and can remain static despite changes in administrative boundaries. Each grid contains a code representing the attribute type or magnitude of the pixel. It therefore represents the distribution of spatial objects or phenomena by regular arrays, and it can contain the non-geometric attributes of objects or phenomena.[68] Grid cells can be generated at different levels of granularity and can contain not only surface-level data but 3D objects. This is particularly interesting when applying innovative technologies such as remote sensing and computational Earth-surface modelling.[69]

geoBoundaries Open Administrative Boundaries Dataset

The geoBoundaries Global Database of Political Administrative Boundaries Database is an online, open-licence resource of boundaries (e.g. State, county) for every country. The platform tracks and maintains around 1 million boundaries within over 200 entities, including all 195 UN Member States. The geoLab at the College of William & Mary applies artificial intelligence (AI) machine learning tools to identify, compare, correct and maintain these administrative boundaries, while flagging boundaries where discrepancies in datasets are most pronounced.[70] All boundaries are available to view or download in common file formats, including shapefiles. In this open-source community, crediting sources is the only requirement for use.[71]

The approach of the geoLab at the College of William & Mary seeks to diversify decision makers in the geoboundary community and enhance a collaborative process. Free access to data paired with research at the intersection between AI and big data shows great promise for future development.

68 Raster data structure and its coding. Accessed on 22 August 2022. drr.ikcest.org/tutorial/k1072

69 Topic 18: Understanding Grid-based Data. Accessed on 22 August 2022. www.innovativegis.com/basis/mapanalysis/topic18/topic18.htm

70 Visualise & Compare Boundaries. Accessed on 15 September 2022. www.geoboundaries.org/geoContrast.html

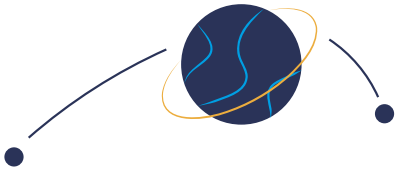
71 geoBoundaries. Accessed on 15 September 2022. www.geoboundaries.org/

Conclusion and Future Outlook

Unique identifiers have undergone a great evolution in recent decades. Experiences with P-codes show that they do not work for all mandates and missions, but coding geographical boundaries and providing interoperable data can bring great benefits to humanitarian and development interventions and must therefore be pursued with concentrated efforts.

For the successful trajectory of increasing data connectivity, there are three determining factors: educational training, adopting new technologies and engaging with national Governments and local authorities. Adding P-codes into spatial datasets from a variety of actors onto HDX, while ensuring correct ownership and accreditation, will advance humanitarian and development operations.[72]

Nationally led data creation and maintenance of unique identifiers is clearly the most sustainable strategy for data preparedness.[73] Nevertheless, where Governments lack the capacity to provide information in the early onset of a disaster, UNOCHA can take advantage of open-source coded boundaries and pre-established relationships, e.g., collaborations with NGOs such as MapAction, to ensure the national adoption of such datasets post-disaster.[74] Given UNOCHA's role in such creation and standardization, we recommend that it joins the UN Geospatial Network to promote a dialogue around strategies for coding geoboundaries.



72 Interview with Elsa Richard.
73 Interview with Luke Caley.
74 Interview with Alan Mills.

> *We do not have a standardized electric plug across the world. You need an adapter to make sure you do not get into trouble. We are now at a point when it is too late to change all the plugs – the only solution is travelling with a universal adapter. The same goes for unique identifiers. This is why we should be careful in expanding on different unique identifiers. We need to make sure this is set up in an interoperable manner with inter-agency standards to ease information management and create an adapter for all kinds of data needs."*
>
> **— Dr. Prashant Madhukar Hedao, Consultant,WHO GIS Centre for Health**

The authors hope this research will help humanitarian and non-humanitarian organizations to improve daily operations by fostering a collaborative information environment based on unique identifiers. Integrating this culture into a comprehensive inter-agency strategy is paramount to building a robust, efficient and inclusive coding system for areas all over the world.

The authors encourage further research to uncover more complementary approaches and to expand the lessons learned.

Annex

List of Case Studies for Further Research

Country and Regional Cases	Type of Incident	Description
Ukraine	Russian invasion 2022	Advanced use of cross-border place codes and relatively successful inter-agency coordination.[75]
Lebanon	Beirut port explosions 2020	No agreed standards and conflicting P-codes, led to confusion in analysis.[76]
Djibouti	Floods 2019	Establishment of national-regional-district-ward-level P-codes after the floods. The World Bank came in for PDNA and could organize the field teams around recognized ward boundaries and data from different assessments to be shared and cross-analysed more easily.[77]
Horn of Africa and the Great Lakes region	Refugee crisis 2017 – today	Unified P-coding systems across the region with refugees in at least two countries. Successful data management across borders with Governments as the custodian of those codes.[78]
Iraq	Earthquake 2017	Lack of capacity to maintain governmental P-codes was addressed through focused training and capacity- and expertise-building for geographic information systems in collaboration with UNDP.[79]

75 Interview with Elsa Richard.
76 Interview with Steve Penson.
77 Interview with Alan Mills
78 Interview with Andrew Alspach.
79 Ibid.

Country and Regional Cases	Type of Incident	Description
Fiji	Cyclone Winston 2016	MapAction designed a P-code system and shared it with HDX, UNOCHA and the Government to establish ownership.[80]
Nepal	Earthquake 2015	Long preparations by national authorities and UN agencies facilitated pre-earthquake P-codes. Needs assessment unit mastered the use of P-codes. Reconciliation tables to make differing data compliant.[81]
Yemen	Civil war since 2014	All data analysis carried out at ACAPS[82] utilizes P-codes to ensure interoperability between datasets.[83] Yemen Core Dataset adopts P-codes throughout.[84] English translations of Arabic place names are inconsistent, so consistency in P-codes is vital. P-code adoption in Yemen has eased interoperability between datasets.
Togo	Preparedness measures West Africa Ebola virus epidemic 2014-2016	In 2015, UNOCHA data showed the sourcing and data are current but were not adopting some of the available data.[85]

80 Interview with Alan Mills.

81 Benini, A., Chataigner, P., Noumri, N., Tax, L., & Wilkins, M. (2016). Information gaps in multiple needs assessments in disaster and conflict areas. Retrieved from: www.acaps.org/sites/acaps/files/resources/files/info_gaps.pdf

82 ACAPS was initially formulated as the acronym for 'Assessment Capacities Project' in 2009. Accessed on 14 August https://www.acaps.org/who-we-are/faqs.

83 Yemen Economic Tracking Initiative. Accessed on 14 August 2022. yemen.yeti.acaps.org/

84 Yemen CrisisInSight Core Dataset. Accessed on 14 August 2022. data.humdata.org/dataset/yemen-crisisinsight-core-dataset-2021

85 Interview with Roberto Colombo.

Country and Regional Cases	Type of Incident	Description
Democratic Republic of the Congo	West Africa Ebola virus epidemic 2014-2016	Effort to synchronize the P-codes in DRC with the health information system boundaries. Congo already has health-related administrative divisions. The Ministry of Health implemented the DHIS2 system from a Scandinavian university, but it is all connected to its internal system of health divisions, which did not work with UNOCHA's P-codes for the humanitarian cluster. The top-level lesson is that you have to know what individuals are using for geography and what the operational implementing actors are using, e.g. in DRC the health boundaries.[86]
South Sudan	Civil war 2013-2020	Issues regarding disputed areas and changes in administrative boundaries. Operational boundaries were sourced from local authorities and mapping with GPS because of a lack of reliable data.[87]
Philippines	Typhoon Haiyan 2013	Established, agreed-to-use framework but very few actors applied it. Only establishment with the endorsement of local groups. But lessons learned for standardized products such as CODs in structuring data to support and keep track of response operations.[88]
Northern Syria	Civil war since 2011	CODs with P-codes were developed fairly quickly and aided joint rapid needs assessment. In conjunction with what3words, health facilities could use P-codes as reporting units.[89]

86 Interview with Ivan Gayton.

87 Interview with Leo Martine.

88 Comes, T. & Vybornova, Olga & Walle, Bartel. (2015). Bringing Structure to the Disaster Data Typhoon: an Analysis of Decision-Makers' Information Needs in the Response to Haiyan. Retrieved from: www.researchgate.net/publication/273886809_Bringing_Structure_to_the_Disaster_Data_Typhoon_an_Analysis_of_Decision-Makers%27_Information_Needs_in_the_Response_to_Haiyan

89 Interview with Leo Martine.

Country and Regional Cases	Type of Incident	Description
Myanmar	Cyclone Nargis 2008	The Myanmar Information Management Unit (MIMU) maintains a repository of common datasets and countrywide geospatial information to the lowest administrative unit. [90] A single and unified system of referring locations, using the P-code system, brought compatibility and combination of multiple datasets, enabling actors to continue generating data and information on different areas.[91]
Indonesia	Yogyakarta earthquake 2006	Building CODs as a preparedness measure ensured the availability of P-code data standards right before the earthquake. This preparedness by the UN and the Government, and transparency in sharing this data, enabled the use of historical data, the swift integration of survey data and community assessments of damage for shelter assistance.[92]
Pakistan	Earthquake 2005	Absence of solid lists of communities and admin units, equivalent baseline population data and community-level infrastructures. Assessments to fill the gaps led to duplicated surveys. Data was only complete months after the incident.[93]

90 MIMU. (2021). Place codes (P-codes). Myanmar Information Management Unit. Retrieved from: themimu.info/place-codes

91 MIMU. (2020). Summary of the Myanmar P-Code version- 9.2 Myanmar Information Management Unit, March 2020.

92 Ulgen, S. & Williams, C. (na.). Standardization of Geographic Names in Humanitarian Information Management (Towards a Humanitarian Spatial Data Infrastructure).

93 Wilder A. (2008). Perceptions of the Pakistan Earthquake Response: Humanitarian Agenda 2015 Pakistan Country Study. The Feinstein International Center. Retrieved from: reliefweb.int/report/world/perceptions-pakistan-earthquake-response

Country and Regional Cases	Type of Incident	Description
Afghanistan	Humanitarian crisis and war since 2001	Up to at least 2010, there were two competing sets of P-codes. Two Afghan ministries disagreed on the recognition of specific districts. One of the most common questions became: "Which set of P-codes are you using?"[94]
Balkans	Refugee crisis 1999	Development of the earliest P-codes to conduct rapid needs assessment and analysis during the return of refugees from the former Yugoslav Republic of Macedonia and Albania.[95]

94 Interview with Tom Haythornthwaite.

95 Smith. D. (2001). Kosovo: Applying GIS in an International Humanitarian Crisis. UN GIS Consultant in Bosnia, Kosovo, Macedonia and Albania between July 1997 and December 1999.


DH Network
Digital Humanitarian Network