Caribou Digital

Identity at the margins: Identification systems for refugees

A Caribou Digital publication, authored by:
Emrys Schoemaker, Paul Currion, Bryan Pon

Acknowledgements

This report was written by Paul Currion, Bryan Pon, and Emrys Schoemaker,* with research support from Suhail Abualsameed, Dina Baslan, and Pius Gumisiriza.

The research was supported by an advisory committee that included Vision International, Save the Children International, Oxfam GB, Action Against Hunger, Vision Fund, Mercy Corps and Mastercard Inc.; research logistics were supported by Save the Children teams in London, Lebanon, Jordan, and Uganda.

This material has been funded by UK aid from the UK government; however the views expressed do not necessarily reflect the UK government's official policies.

* Corresponding author: emrys@cariboudigital.net

Recommended citation:

Caribou Digital, *Identity at the Margins: refugee identity and data management*,
Farnham, Surrey, United Kingdom: Caribou Digital Publishing, 2018

Reach us at:

info@cariboudigital.net
www.cariboudigital.net

UKaid
from the British people

Executive summary

Interoperable identity systems in the humanitarian sector can enable significant operational benefits for organizations, yet realizing benefits for refugees, including personal control of portable data, will only be achieved when beneficiaries are the focus of system design

Humanitarian organizations are embracing increasingly sophisticated digital systems for managing the identities and personal data of the beneficiaries they serve. While these systems can provide significant operational benefits to the organization, there is growing recognition that the current lack of interoperability between systems has negative impacts for both organizations and beneficiaries.[1] Improving the ability of humanitarian organizations to share identity and related data across systems should, in principle, improve service delivery, simplify reporting, reduce fraud, strengthen data protection, and increase convenience for beneficiaries themselves.[2]

Yet identity and personal information are intrinsically powerful and sensitive topics, and given the vulnerable nature of the beneficiaries being served—including refugees and other stateless individuals—the complexities and risks of these large-scale systems are amplified. Navigating this terrain therefore requires us to understand the experiences of the individual beneficiaries that are subjected to these systems, and to balance the operational benefits with an empathetic assessment of how these systems impact individuals' privacy, dignity, and agency.

Key findings: The refugee perspective

- The current state of humanitarian identity systems presents challenges to refugees, who typically have very limited visibility or agency around the data collected about them by organizations.
- Refugees have little to no knowledge of the institutional processes through which their personal data is managed, including which organizations have access to their personal data.
- Refugees are rarely offered the opportunity to exercise agency with regards to data that is collected on them (e.g., they are rarely given choices about what data is collected or how it is shared), despite having the capacity to do so.
- Systems that record beneficiaries at a household rather than individual level can impact household power dynamics (e.g., when women are registered as heads of households when traditionally the head of the household is the male), amplifying the impact of systems on individual lives.

[1] See, for example, USAID, "Identity in A Digital Age: Infrastructure for Inclusive Development" (USAID, September 2017), https://www.usaid.gov/sites/default/files/documents/15396/IDENTITY_IN_A_DIGITAL_AGE.pdf

[2] WEF, "Digital Identity On the Threshold of a Digital Identity Revolution" (Davos, Switzerland: World Economic Forum, January 2018), http://www3.weforum.org/docs/White_Paper_Digital_Identity_Threshold_Digital_Identity_Revolution_report_2018.pdf; Charlie Ensor, "Biometrics in Aid and Development: Game-Changer or Trouble-Maker?," The Guardian, February 22, 2016, http://www.theguardian.com/global-development-professionals-network/2016/feb/22/biometrics-aid-development-panacea-technology

- At the same time, many refugees make active efforts to negotiate the various identities available to them, consciously weighing the benefits and constraints associated with different statuses (such as registered vs. unregistered) in order to access services, ensure eligibility for employment, and preserve their spatial mobility.

Key findings: Institutional perspective

- The diversity of systems, capacity, and organizational mandates continues to create challenges for sharing identity and related data between humanitarian organizations.
- Beyond the bespoke systems of UN agencies and larger NGOs, many organizations lack technological capacity and instead rely on repurposed and insecure existing technologies (such as spreadsheets) to share the personal information of vulnerable populations.
- The diversity of mandates amongst UN agencies, donors, NGOs, and the private sector further complicates identity system interoperability and the sharing of refugee information.
- The consequence is data sharing practices that are inefficient and insecure, hampering organizational efforts to deliver basic services and uphold protection principles—which should now cover the personal data of refugees.
- Current trends in data management focus on organizational needs for internal efficiency and data security, rather than on the needs of individual refugees, leading to data sharing becoming more opaque and inaccessible to beneficiaries.

Recommendations

It is therefore critical that efforts towards increasing interoperability pay attention to individual refugee needs, and particularly to strengthening their control over the use of their personal data. This report therefore describes a way forward to strengthen identity and data management for refugees in ways that can deliver the organizational benefits of increased interoperability while strengthening individual agency and privacy for some of the most vulnerable populations.

The report concludes with detailed and actionable recommendations for humanitarian stakeholders, specifically DFID. These recommendations are intended to help realize the opportunities presented by digital technologies to strengthen the agency and privacy of refugees, and to enhance the capacity of humanitarian organizations to provide efficient, transparent, and accountable services.

- Donors should align their requirements for policy and practice around data protection, and lobby host governments to establish data protection legislation.
- The humanitarian community should ensure "data accountability" is built into future policy and guidance documents for Accountability to Affected Populations (AAP).
- In the short term, a multi-stakeholder working group on interoperability chaired by UNHCR should be established, to support a longer-term standards body focused on identity data.
- The working group should develop standardized approaches to a "translation layer" that enables interoperability between the data management systems of diverse service providers. The translation layer would enable interoperability of functional identities and reduce the dependence on legal refugee status

as a prerequisite for obtaining access to needed services, removing barriers to access and strengthening inclusion.

- The working group should establish a compliance framework; this could take the form of a “command” mode, but a voluntary, incentive-based strategy will likely be more effective.
- DFID and other donors should incorporate data management into the funding process by: integrating it into proposal criteria and project evaluation, including data management as a separate and protected budget line in every project proposal, funding more pilots, engaging at the senior management level, agreeing a common requirement for the inclusion of data in agency reporting, and incorporating funding for better data management into host government capacity building.
- DFID should particularly support the development of open-source biometric standards and solutions in order to create a more inclusive approach to biometrics overall.
- Crucially, humanitarian service providers, supported by donors, should invest more in participatory design to ensure more ethical identity systems.

Introduction

This document is the final report of a research project commissioned by the Department for International Development (DFID). The starting point for the research was three questions related to data standards for digital identity systems for forcibly displaced people:

- What is the current state of play for data standards within refugee camps?
- What is the design specification required for a data standard for a functional ID for refugees in camps?
- What are the main ethical issues facing functional digital identity systems for forcibly displaced people?

Driving these questions is the belief that the use of data standards in the humanitarian sector can improve the efficiency and efficacy of service delivery, while increasing convenience and utility for refugee beneficiaries. The organizations we spoke with described current systems and processes as highly fragmented and siloed, making it difficult if not impossible to perform many coordination and analytical activities, and there was widespread enthusiasm for the operational benefits of being able to easily share identity and related data across organizations.

A commonly referenced scenario was one where refugees could receive a single identity credential that would enable them to access services across multiple providers without having to register with each, and potentially even access services outside the humanitarian context (e.g., open a bank account).

Realizing these outcomes will require organizations to increase system **interoperability**, the capacity to transfer and process data between digital systems. Achieving interoperability requires organizations to agree on **standards** at multiple levels—not just technical data formats but also operational processes, legal agreements, and governance mechanisms. While technical standards are often the focus of discussions, they are only a small part of the overall framework required to achieve interoperability; it is much easier for organizations to agree on data formats, such as XML or JSON,[3] than to agree on country-specific data management practices or a mutually-agreed definition of "informed consent." So while this research included reviews of the data types and technical systems in use by different organizations, our analysis is focused on organizational and structural dimensions.

It is important to note that this drive to increase the interoperability of identity-based systems is not unique to the humanitarian sector. The health care sector, for example, has seen multiple efforts at developing industry-wide frameworks for sharing patient identity and medical information,[4] and the European Union has enacted as part of its Digital Single Market initiative the eIDAS framework, which specifies how EU member states can rely on digital identity credentials from other member states.[5] There are clearly lessons that can be drawn from these and other efforts, both successful and unsuccessful.

Yet there are many aspects that are unique to the refugee context specifically, and the humanitarian sector more broadly, and it is these that are the focus of this report. For example, the nonprofit nature of most organizations has a large impact on how resources are allocated both internally and between organizations. The UN—and especially

[3] XML and JSON are among the most common data formats used for data exchange; both of which are readable by machines and humans.

[4] For example, see FHIR, http://www.hl7.org/

[5] For an overview, see https://www.cryptomathic.com/news-events/blog/understanding-eidas , Dawn M. Turner, "Understanding eIDAS," Cryptomathic, January 2016, https://www.cryptomathic.com/news-events/blog/understanding-eidas

UNHCR as the intergovernmental organization with the mandate to ensure protection for refugees and seek permanent solutions in refugee crises—is critical for a healthy humanitarian ecosystem, but an ecosystem requires more than one actor, and UNHCR requires effective partners in order to fulfil that mandate.

But the most important consideration by far is that refugees and other displaced persons are by definition extremely vulnerable populations. The power dynamic and information asymmetry in a typical refugee response is tremendous, and the source of many of the challenges refugees face. Therefore while the drive to increase the interoperability of identity systems is a top-down effort driven by organizational imperatives, this research was designed from the beginning to highlight the needs and perspectives of refugees themselves.

As we describe in this report, the current state of humanitarian identification systems already presents challenges to refugees, who typically have very limited visibility or agency around the data collected about them by organizations. It is likely that increasing interoperability between systems will only accelerate the transmission, processing, and storage of personal information in ways that are even more opaque and inaccessible to refugees. It is therefore critical that efforts towards increasing interoperability pay attention to individual needs, and to strengthening individual control over the use of personal data. Increasing interoperability without properly addressing the interests of refugees could amplify the impact of data regimes—such as UNHCR's beneficiary management system, proGres—that have the power to reshape social structures, such as family organization, without recognising that impact.

Humanitarian systems are increasingly digital both indirectly, through their reliance on information technology in all their management systems,[6] and directly, through new types of interventions such as cash programming, which are increasingly delivered through digital channels.[7] This digitisation of aid, if approached from a position that empowers the individual as much as the institution, offers a chance to give refugees back their voices; and the first step is to give them control over their identities.[8]

The report begins by describing the current landscape of identity ecosystems in the humanitarian sector, describing the key organizations, their requirements and an overview of system design for robust data management. This is followed by insight into the experience of refugees, and how their needs and the challenges they face raise ethical questions around identity management and data sharing. The report then focuses on the current status and challenges of interoperability between humanitarian organizations, followed by an overview of pathways for improved data collection, presenting an interoperability framework and models for identity management. The report concludes by describing ways in which the humanitarian community could help achieve this, followed by specific, actionable recommendations that can help make progress towards this ideal state.

[6] OCHA, "Humanitarianism in the Network Age" (New York: UN OCHA, March 6, 2013), https://www.unocha.org/publication/policy-briefs-studies/humanitarianism-network-avge

[7] Dan McClure and Brad Menchi, "Challenges and the State of Play of Interoperability in Cash Transfer Programming" (Geneva: UNHCR & World Vision, 2015), http://www.cashlearning.org/downloads/erc-executive-summary-interoperability-web.pdf

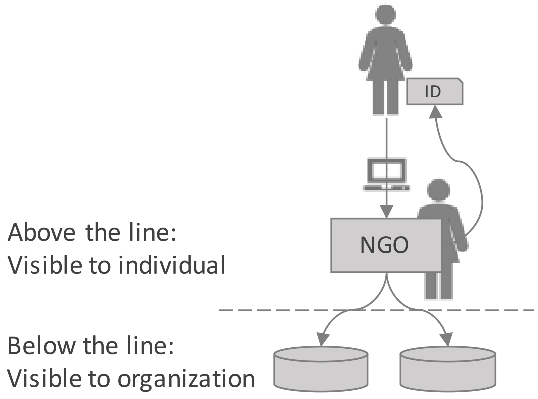
[8] Paul Currion, "The Refugee Identity – Caribou Digital – Medium," Medium (Caribou Digital, March 13, 2018), https://medium.com/caribou-digital/the-refugee-identity-bfc60654229a

Methods

The highly technical nature of identification systems often leads to discussions where technical problems and their solutions are the implicit focus of the analysis. This kind of bias can pre-determine the range of potential responses by leaving out more critical perspectives. In the primary research for this report, we have tried to be more systematic in how we include different stakeholder views in the analysis.

One way we do this is through a construct for categorizing those activities that are visible to the individual refugee vs. those activities that are visible to the organization in its broadest definition. In any given identification system, a refugee may speak to an NGO field worker, present a credential, be shown documents, and so forth; all of this is visible to the refugee as what we call "above the line" activity.

Figure 1. A construct for considering what is visible to the beneficiary, and what is visible to the organization



But the other side of these engagements are typically invisible to the refugee—the NGO worker may type data into a beneficiary management system, this data may be transferred to a global headquarters for verification, some data may be shared with partner NGOs; all of this is what we call "below the line" activity.[9] We find this simplified construct—illustrated in Figure 1—useful in exploring the organizational and refugee perspectives on identity systems, and to highlight the gaps between those perspectives.

This construct serves as both a research method and a way of presenting our findings. As a form of presentation the construct emphasizes the experience of the individual and the voice of the refugee. This emphasis also establishes the refugee experience as the foundation for the ethical questions that we believe should be central to discussions about the future development of identity systems, and particularly their interoperability. It is apparent that refugees—and other vulnerable populations in humanitarian crises—are not consulted about the data regimes through which they are managed, regardless of the potential impact of those regimes on their lives and livelihoods.

The research was carried out by researchers from Caribou Digital (Paul Currion, Bryan Pon and Emrys Schoemaker) supported by research assistants (Suhail Abualsameed, Dina Baslan, and Pius Gumisiriza). The research consisted of three parts: a literature review,[10] stakeholder consultations, and a series of three country visits (Lebanon, Jordan, and Uganda).

Logistics support for the country visits was provided by Save the Children International, and broader support was provided by an informal consortium of international NGOs (World Vision International, Save the Children International, Oxfam GB, Action Against Hunger, Vision Fund, and Mercy Corps) who have been working with Mastercard Inc. to develop digital identity standards for data interoperability in humanitarian operations.

[9] This construct is based on the model presented in USAID, "Identity in A Digital Age: Infrastructure for Inclusive Development."

[10] The literature review was published by Caribou Digital as a separate essay, Paul Currion, "The Refugee Identity – Caribou Digital – Medium." https://medium.com/caribou-digital/the-refugee-identity-bfc60654229a

The research methods included interviews with stakeholders and refugees. We interviewed over 80 stakeholders working on issues related to identity and data management from headquarters and field offices, using a semi-structured interview guide but allowing respondents to draw on their experience and insights. We talked with over 200 registered and non-registered refugees in Lebanon, Jordan, and Uganda, in and out of camps, between January and May 2018.

We developed a structured interview guide informed by the literature review and stakeholder consultations, and conducted hour-long focus groups and in-depth interviews in which respondents were invited to identify and articulate the issues most important to them. The interview guide was informed by previous Caribou Digital identity research, drawing out the experience over time to explore the "refugee journey" from initial registration, the subsequent use of refugee credentials, and finally the experience of authenticating one's identity in daily interactions.

Below the line: The current landscape

Humanitarian identity systems exist within a "national identity ecosystem" (NIE) and include a diverse set of stakeholders, many of whom share objectives but have distinct roles and interests, which shape the top-down, command-driven identity systems they use

We begin our analysis from the point of view of the organization, since it is NGOs, donors, governments, and other institutional actors that are pushing for increased interoperability of identity and data systems. This section first provides an overview of how humanitarian identity systems fit into broader national identity ecosystems (for a detailed overview of models of identity systems, see Appendix A), and an overview of identity systems in a typical refugee response.

Many of the stakeholders in a humanitarian response share similar objectives and anticipate similar benefits (described in detail in the section "Operational barriers to interoperability") but occupy distinct roles with contrasting agendas. The following section summarizes these differences, taking a top-down perspective in describing the incentives and requirements of the different actors involved in the space, and how these translate into actual technology systems and practices.

Structural dimensions of identity systems

Every identity system is part of what we consider a "national identity ecosystem" (NIE), a "system of systems"[11] which encompasses all the different forms of identification and authentication within the bounds of a nation-state, both formal and informal, including: state-based identity systems such as passport or birth certificate; private-sector systems such as mobile SIM cards or Facebook; and humanitarian identity systems such as UNHCR's or WorldVision's beneficiary management systems.

Even in states which do not issue a national identity credential, an ecosystem of some sort will exist, although that ecosystem may be in poor health, especially in failed states. In any given nation-state, therefore, this constellation of systems will look and operate differently based on countless interdependent factors; like a biological ecosystem, the complexity of these sociotechnical systems evolving over time means it is impossible to alter one system without impacting others; conversely, creating widespread directional change requires complicated, multi-point engagement. While humanitarian identity systems usually (although not always) function within the bounds of one or more NIEs, they are commonly perceived in isolation from those wider ecosystems.

Therefore while the focus of this research is on humanitarian systems, it is critical to also understand how these are situated within the broader ecosystem. For example, in NIEs where the government effectively does not provide formal state recognition of refugees, the humanitarian identity systems take on added importance and value as the most "official" credential refugees can obtain, whereas in other NIEs the state may take a more active role and thus change the power dynamics of the humanitarian response. To take another increasingly relevant example, private-sector actors such as mobile network operators and commercial banks are increasingly involved

[11] Alan Gelb and Anna Diofasi Metz, *Identification Revolution: Can Digital ID Be Harnessed for Development?* (Center for Global Development, 2018); USAID, "Identity in A Digital Age: Infrastructure for Inclusive Development."

with humanitarian programming in the form of cash transfers, and consequently are key actors in conversations about the regulatory environment for extending financial services to refugees.

It is also important to note that "digital identity" systems usually have analog components—e.g. a paper certificate—and almost always have analog precedents, and are thus better thought of as identity systems in an increasingly digital age, rather than as an entirely new phenomenon. There is a rich literature on identity and other administrative systems that is still relevant to our understanding, and we should build on that knowledge with specific research on how these systems are affected by digitization and networked technologies, rather than starting anew.

Identity systems in the refugee response

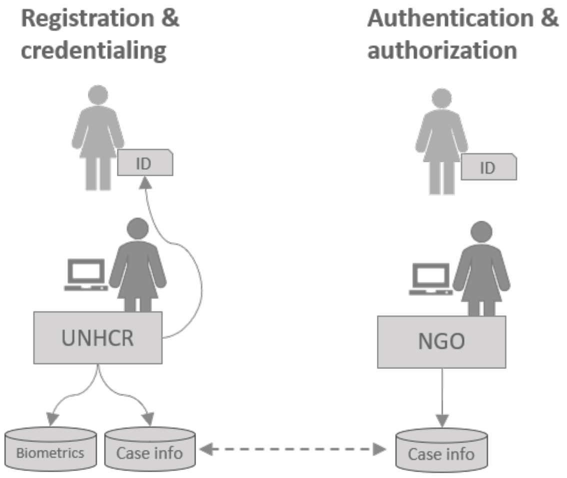
At the system level, a wide range of technology systems are used for managing the identity and related data of refugees. While there are countless use cases and individual experiences, there are key interaction points in which individuals and their data are processed into multiple organizations, systems, and databases. Refugees experience this as a journey through the identity "value chain"[12] in which individuals are **registered** into systems, **issued credentials**, and then **authenticated** and **authorized** each time they access services (see Figure 2 for simplified depiction).

Registration: Most beneficiaries register first with UNHCR. The beneficiary presents any extant identity documents, and provides basic biodata (name, birth date, etc.). UNHCR performs identity proofing to validate the information and check for existing records. UNHCR typically uses biometrics (fingerprint and iris) to help with deduplication to ensure the new record is unique at the global level; if no match is found, a new account is created. This biometric data is typically stored separately from case-level data.

Credentialing: For each new account UNHCR issues a credential (card, paper certificate, etc.) to the individual, incorporating a globally unique alphanumeric identifier—the case number—that subsequently becomes the *de facto* identifier for many service providers (particularly NGOs) who work with refugees.

Authentication: To access services from another provider, beneficiaries may have to re-register, often using their existing UNHCR certificate as a "breeder document" or trusted source of identity information. However many NGOs will simply accept the UNHCR certificate, and will authenticate the beneficiary against the credential, often just by visually comparing the photo.

Figure 2. Simplified view of identity value chain in a refugee context, where the individual first registers with UNHCR and is issued an UNHCR certificate, which is then used to authenticate with other NGO service providers



Authorization: If the beneficiary meets the requirements of the service provider, they will authorize the provision of services. Sometimes this is based on shared case records between an NGO and UNHCR or another organization, or authorization may be completely internal to the NGO.

[12] The identity value chain is described in detail in Gelb and Metz, *Identification Revolution: Can Digital ID Be Harnessed for Development?*; USAID, "Identity in A Digital Age: Infrastructure for Inclusive Development."

The political economy of data management

Although data management is critical, it is something that the humanitarian community has historically struggled with, and the impetus for this research was generated by the challenges that service providers face in sharing data more effectively. Collective data requirements remain relatively ill-defined, although the UN Office for the Coordination of Humanitarian Affairs (UN OCHA) has worked with partners to develop the first data standard for the humanitarian sector, the Humanitarian eXchange Language (HXL). Lessons learned in developing HXL found that "the cost of the lack of data standardization… was (and to a degree still is) not well understood," because "management are more concerned with the end product—a presentable, compiled spreadsheet—than the time and effort required to get it to this point."[13]

It was clear from the interviews with stakeholders at both headquarters and field offices that while there were technical obstacles to data sharing, the larger obstacles related more to specific organizational needs and interests—in other words, the political economy of data sharing in the humanitarian sector. While the specific dynamics between organizations vary considerably from country to country, there are a limited range of roles which they occupy; we outline here those roles and how they fit into the political economy of data management.

Host governments are legally responsible for refugees under the 1951 Refugee Convention, and their policies and practices therefore define the legal and operational context for humanitarian action. While neither Lebanon or Jordan have signed the Convention (although both have Memoranda of Understanding with UNHCR), their governments still play a critical (if controversial) role in refugee management; in contrast Uganda is a signatory and has been hailed as a model for refugee management.

The research identified that government policy was the single most significant determinant of formal refugee identity, both in terms of policy frameworks (e.g., the legal status of displaced people) and political will (e.g., the type and amount of resources committed to their support). The priority assigned to the refugee crisis is reflected in the level of government responsibility—in the Office of the Prime Minister (OPM) in Uganda, the Ministry of Planning and International Cooperation (MoPIC) in Jordan, and the Ministry of Social Affairs (MoSA) in Lebanon—which in turn influences the amount of political attention given to the crisis and the types of resources assigned by the government to address it.

Government policy also informs the way in which, if at all, state-based identity systems interact with the humanitarian response. For example, in Uganda the government has mandated that health centers in refugee camps must follow the same data management practices as government-run health clinics. In principle, humanitarian NGOs delivering health services are supposed to share summary data with the government's Health Management Information Service. In practice, the health department lacks digital capacity and relies on paper records.

Government policy thus shapes the policies and practices—including data management— of other actors, who must respond to these factors whether or not they work directly with government agencies responsible for providing social services. This is clearly the case with identity data, which is seen as a valuable resource both by host governments and humanitarian organizations alike, whether for security management, resource mobilization, or service provision. However none of the three locations in this research project have substantive data protection legislation,[14] which creates additional risks for refugees (including their

[13] Alexandra T. Warner and Alice Obrecht, "Standardising Humanitarian Data for a Better Response: The Humanitarian eXchange Language | ALNAP" (London: ODI / ALNAP, March 10, 2016), https://www.alnap.org/help-library/standardising-humanitarian-data-for-a-better-response-the-humanitarian-exchange

[14] "State of Privacy 2018" (London: Privacy International, January 2018), https://privacyinternational.org/type-resource/state-privacy

livelihoods) and thus an additional burden for the humanitarian community.

Technological solutions are therefore second in importance to national policy frameworks, yet it is also clear that the opportunities offered by new technologies—such as biometric registration—can stimulate debates about national policy.

UN operational agencies are required by their mandates to work with governments; in particular UNHCR has a legal mandate for refugee protection under the 1951 Convention, and is usually agreed to be the lead agency in refugee crises.[15] As a result, UNHCR has a central role in any data regime relating to refugees, which is reflected in a clear commitment to data protection,[16] summed up in one respondent's declaration that "the protection of the data is close to the protection of the individual."

UNHCR data has commonly been stored in the country of deployment, but its new beneficiary management system (proGres v4) is moving towards a global database, hosted in Geneva, in order to strengthen data security as well as improve efficiency; UNHCR has strict limits on data sharing and only provides limited access to refugee data. In all three research locations, UNHCR plays a central role in registering refugees,[17] although the exact nature of this role varies depending on local conditions:

- **Support** - The host government takes primary responsibility for registering refugees, which UNHCR supports depending on government capacity. During the Uganda research, the government was in the middle of a program to verify the identities of all refugees in the country, a program only possible because of the support of UNHCR's biometric software and field staff.
- **Lead** - UNHCR assumes responsibility for registering refugees, usually with a mandate from the host government. This mandate may be subject to policy changes; in Lebanon UNHCR suspended registration of Syrian refugees in May 2015 at the request of the government, causing the exclusion of many *de facto* refugees, which in turn complicated UNHCR's operations and particularly its relationship with NGOs.
- **Parallel** - The host government carries out some form of registration (often for security purposes), but UNHCR also carries out a separate registration process. In Jordan refugees crossing the border were first screened by the General Intelligence Directorate (Jordan's intelligence agency) but then passed to UNHCR for formal registration, creating "parallel" data regimes serving different political purposes.

A UNHCR credential is often the most valuable document that a refugee can obtain in terms of recognition of legal refugee status and access to services, especially for refugees who lack any identity documents from their country of origin. While UNHCR credentials have functional value, refugees often attach equal value to national identity documents not just in legal terms as a claim on a set of rights (including the right to return) but also in symbolic terms, as a social and emotional connection to their place of origin.

In practical terms, a UNHCR credential was the *de facto* identity document for refugees in all three research locations, and served as a breeder document (a document that is used as the basis for other forms of identification) when registering with other service providers, or as a basis for visual authentication by those service providers.

However other service providers also "register" refugees themselves, and sometimes issue their own identity credentials; while this establishes

[15] When UNHCR is not assigned as lead agency by a government, it is considered controversial; see for example the recent decision by the government of Bangladesh to assign IOM as the lead agency in the Rohingya refugee crisis.

[16] UNHCR, "Policy on the Protection of Personal Data of Persons of Concern to UNHCR" (Geneva: United High Commission for Refugees, May 2015), http://www.refworld.org/docid/55643c1d4.html

[17] Other UN agencies also play important roles in service provision, particularly WFP and UNICEF, but the research did not investigate their roles except where it directly related to camp services.

identity management as a common requirement for a range of organizations, usually on a project basis, it also creates a regime of overlapping and often redundant identity systems.

Non-governmental organizations, international and national, usually deliver actual services on the ground. Within refugee camps, UNHCR subcontracts NGOs to distribute food and non-food items, maintain shelters, etc. In addition, many NGOs have their own funds and work both in and out of formal refugee camps, as well as in host communities.

NGOs frequently experience tension between recognition of UNHCR and government mandates, and frustration at the way in which those mandates are expressed; one respondent in Lebanon pointed out that UNHCR's unilateral implementation of its Refugee Assistance Information System (RAIS)[18] without adequate consultation had alienated the NGO community, requiring both UNHCR and its NGO partners to expend resources in rebuilding their relationships.

Data management within the NGO community is weak, primarily due to lack of resources: spreadsheets are the default means of data management, larger organizations working in multiple sectors usually have multiple unrelated datasets, and data sharing between NGOs was extremely weak in all locations. This was recognised by NGOs themselves, but examples did emerge of increased NGO investment in data management.

Some larger NGOs, such as Save the Children and CARE, have invested in bespoke information management systems that include beneficiary and identity management capacity; however the motivation for the development of these systems are internal management requirements rather than external data sharing. There are a small number of platforms designed within the sector which are being used by multiple organizations in multiple locations—one example would be World Vision's Last Mile Mobile Solutions (LMMS)—but data is not necessarily shared between different instances of the same platform. In some cases, consortia have formed to implement joint solutions specifically to facilitate data sharing; in Lebanon, the Danish Refugee Council has developed a Referrals Information Management System (RIMS) to address persistent problems in the existing protection referrals system, which five other NGOs have signed up to.[19]

Community-based organizations (CBOs) are an overlapping subset with NGOs; the term CBO often encompasses a wider variety of organizational forms which may include a wider variety of organizational forms such as faith-based actors, trades unions, women's groups, neighborhood solidarity movements, and so on. The research was not able to explore this group of actors in any depth, but responses confirmed that they form a critical part of the response, particularly with urban refugees and host communities in longer-term programming.

Often there is a tiered relationship, from donor to INGO to CBO; however the variation in CBO technological capabilities varies even more widely than among NGOs, and their priorities in terms of data management are significantly different than more formal humanitarian organizations, due to much of their work being relational rather than transactional in nature. In practice this means that CBOs often do not recognise the value of beneficiary data in their humanitarian action, not just as a basis for planning but also a means of mobilizing resources through advocacy and fundraising.

Private-sector firms support refugees both directly (through philanthropy or corporate social responsibility[20]) and indirectly (through products and services either to refugees, or to other service providers—especially INGOs and UN agencies). In the research locations it was clear that private-

[18] RAIS was developed separately to proGres as part of the Syrian regional response, but UNHCR has expressed its intent to integrate RAIS functionality into the new proGres v.4, as part of the new Population Registration and Identity Management EcoSystem (PRIMES) platform.

[19] InterSos, MedAir, Handicap International, Solidarites, and GVC Italia.

[20] See, for example, Tent Foundation and Refugee Investments

sector involvement in data management was increasing, driven largely by the requirements of cash distribution programs.[21]

Particularly notable was the role of the Jordanian company IrisGuard, contracted by UNHCR to support the expansion of biometric identification in the Syrian refugee response with proprietary hardware, software, and cloud services. Since cash distribution requires individual authentication, and given the centrality of identity to all data regimes, this positions the private sector more centrally in humanitarian response than ever before. It also creates a new market for data in humanitarian response, as private companies—particularly mobile network operators and services provided through their networks—now have the opportunity to gather large amounts of data on a previously inaccessible market segment.

Donor governments, while in some cases frustrated by problems associated with data management (particularly gaps in coverage caused by incomplete data), have not been fully engaged with the issue. Institutional donors often lack the capacity to understand wider issues around data management; for example DFID has been supportive of data management, including funding development of HXL and HDX, but has little of its own capacity to engage with ethical and practical issues around data. As data becomes increasingly critical to humanitarian operations, expertise in, for example, data law and ethics, will also become increasingly critical, requiring new skills within donors.

However these donors play a critical role; a recent report on drivers of change in the humanitarian system pointed out that "those with the greatest power to effect reforms [e.g. donors] are often not those with the strongest interest in their success... reforms are only partially in line with their self-interests or, in the case of accountability to affected populations, even run counter to them."[22]

One vendor commented that "the underlying problem is the beneficiaries are not the customers of the NGOs, the donors are," and this donor-driven model of data collection can create institutional disincentives. More accurate beneficiary data may reveal that service providers have exaggerated their beneficiary numbers, and therefore inflated their project budgets.

Organizational requirements for robust identity and data systems

Humanitarian organizations have varied operational profiles, yet there are common incentives and requirements for operating robust identity and data management systems, and understanding these drivers is instructive for considering how organizations evaluate and prioritize their efforts towards interoperability. In this section we summarize the main themes that arose out of our interviews with UN and NGO staff in Lebanon, Jordan, and Uganda.

Beneficiary information is essential for organizations to assess eligibility for status and services. In addition to the importance of accurate information, e.g., place of origin to determine refugee status, such information is essential for assessing vulnerabilities and prioritising needs in order to provide services. While this type of data, and subsequent data collected during service delivery, is essential for internal business analytics and resource planning, the systems themselves are—in the words of one vendor—"designed to meet the immediate business needs of the organization," i.e., time-limited and organization-specific project requirements, rather than wider sectoral concerns.

[21] Although other types of commercial actors—notably companies working in the supply chain of humanitarian goods—are deeply involved in delivery, they are not involved in response planning and implementation to the same extent, and do not generate data that can be used in the same way by humanitarian organizations.

[22] Julia Steets et al., "Drivers and Inhibitors of Change in the Humanitarian System," *A Political Economy Analysis of Reform Efforts Relating to Cash, Accountability to Affected Populations and Protection. Global Public Policy Institute*, 2016, http://www.gppi.net/fileadmin/user_upload/media/pub/2016/Steets__Binder__Horvath__Krueger__Ruppert__2016__Drivers_and_Inhibitors_of_Change_in_the_Humanitarian_System.pdf

Organizations view proper data management as essential for preventing fraud. One of the most common reasons given for improving data collection—particularly identity data, and especially now biometric data—is to address fraud.[23] One staff member involved in the verification exercise in Uganda said, "one of the main benefits of the system is that it will allow the identification of 'recyclers'—individuals who appear on multiple registration documents." That exercise shows that biometrics can help to reassure donors; however there is limited data on the types and levels of fraud prevalent in the humanitarian sector, and almost no quantification of how much improved data management will reduce this fraud. It was clear that technology alone will not prevent fraud; as one NGO technology lead explained, "a lot of the fraud and waste isn't due to technology, but to processes, to humans."[24]

Donor-funded organizations need beneficiary data to report back on service provision. Service providers need robust management information systems to report back to both institutional donors such as DFID, USAID, and ECHO, as well as individual supporters in the general public. As one representative from World Vision said, "connecting data within [LMMS, our beneficiary management system] to our marketing team helps so they can provide better feedback to supporters, i.e., donors." This is the most obvious example of how data becomes part of the political economy of the aid industry; it is not itself monetized (as it is by commercial service providers) but it is used as leverage to raise money.

Organizations need to share accurate, up-to-date service delivery information for planning and resource coordination. Service delivery data, such as number of beneficiaries reached, services provided and active locations, is critical to support organizational and sectoral planning and reporting. It is particularly important at a response level to enable efficient allocation of resources through coordination processes. Although there are coordination mechanisms in place, such as the UNHCR-convened Information Management Working Group and the "3W" reporting mechanism,[25] in all three research locations these mechanisms were dormant or suffered limited participation.

This is largely because they are not a priority for senior management, since information sharing with other organizations is not deemed mission-critical. Importantly though, these coordination mechanisms largely involve activity-based information, often with a spatial component, but with little if any personally identifiable information (PII) involved. Improvements in this area, whilst important as part of increasing efficiencies in the wider response, are not necessarily closely linked to beneficiary identity and information systems.

[23] This is not the only reason given—nearly all organizations considered in this research assert that improved data collection will lead to better services—but it is the common thread that runs through the rhetoric most prevalent at present. Zara Rahman, Paola Verhaert, and Carly Nyst, "Biometrics in the Humanitarian Sector" (Berlin: The Engine Room and Oxfam, 2018), http://oxfamilibrary.openrepository.com/oxfam/handle/10546/620454

[24] This is not to say that other concerns are not also important drivers of information management systems. For those organizations with relevant policies, data management is viewed through the lens of safeguarding and the protection of individual data, and is one of the incentives for the development of robust identity systems.

[25] 3W is a tool to show Who does What, Where, e.g. operational presence by sector and location within an emergency. It enables organizations to identify potential partners, give a rough understanding of an ongoing response, and identify potential overlaps or gaps.

Above the line: Ethics in identity systems

Refugees often lack understanding and control over the systems and processes in which personal data is collected and managed, despite their interest and capacity to do so

Despite existing to serve the needs of refugees, all too often the ultimate beneficiaries of identity and data management systems are ignored. The following section describes the "above the line" experience of refugees to expose the needs and challenges that they face in their interactions with identity systems. Their exclusion is a common problem in the humanitarian system, where the voices of beneficiaries are commonly marginalised in the design and delivery of humanitarian response,[26] but is by no means unique to the humanitarian sector.[27]

Incorporating the experience and perspective of refugees is critical to the foundation of an ethical approach to the development of future identity and data management systems.

Refugee needs and experiences of identity systems

Refugees experience identity systems very differently than organizations; nearly all the issues described in the previous section exist "below the line" of refugee experience. This difference in perspective can create tensions in the relationship between refugees and the organizations that serve them, particularly because refugee perspectives on identity and data-related issues are rarely heard. This section summarizes key themes we heard in our interviews and focus groups with refugees.

The experience of refugee identity is determined above all by government policy. We heard how access to employment, mobility, and well-being as a recognized refugee was determined by host-government policy. For example, in Lebanon the government has strict policies around refugee employment and mobility, with curfews commonplace—as Amina, a Syrian refugee, told us: "We stay in the village, we don't go far from there. We cannot leave our homes after 9 p.m., no one can." By contrast, in Uganda the government grants legal recognition, freedom of movement, and even issues each refugee with a small plot of land; as Abdul, a South Sudanese refugee in Bidi Bidi camp explained, "they give us land, some poles for a hut, we feel welcomed, like family."

Refugees have partial knowledge of registration processes, which leads them to adopt different strategies—including avoiding it completely in some cases. Many refugees we spoke with in Lebanon, Jordan, and Uganda described concerns about registration interview questions such as place of origin, and anxiety about the consequences of this data being shared. For some, their concerns about sharing personal data to obtain refugee status were so great that they did not register. One such was Yasser, a Yazidi Syrian refugee living with his wife and two children in a one-room apartment in Lebanon, who stated, "Everybody was registering with the UN, but we did not. We were suspicious and scared. We don't know if the UN shares information with anyone, so that is why I did not share many things with them." By contrast, other

[26] There is a growing field of work to strengthen beneficiary engagement—see for example IFRC, "Beneficiary Communication and Accountability. A Responsibility, Not a Choice: Lessons Learned and Recommendations" (International Federation of the Red Cross and Red Crescent Societies, 2011), http://www.ifrc.org/PageFiles/94411/IFRC%20BCA%20Lesson%20Learned%20doc_final.pdf

[27] For example, see previous work by Caribou Digital on state-based identity systems in India, including Aadhaar. Caribou Digital, "Identities: New Practices in a Connected Age" (United Kingdom: Caribou Digital Publishing, 2017), https://www.identitiesproject.com/report/

refugees exercise their agency through subversion or even fraud; in Bidi Bidi we heard how some people "pay the parents of children ten thousand shillings to use them in their registration," inflating their family size and thus the amount of food they are entitled to.

Refugees have limited ability to change identity information, which can have long-term impacts. Rami, a young male Syrian refugee in Lebanon described the difficulty of updating information after marriage: "transferring something from one file to the next it is a hassle. In marriage, to have our own file it is very hard...if you say you want to review or edit my file, it would take you a year." Many respondents described how changing information, such as name, address, or family size, was difficult and time consuming, often because access to staff and systems was limited. In Azraq camp in Jordan, many children's ages were incorrectly registered with UNHCR, yet parents had no mechanism to change their child's record in RAIS. The impact of this might be that their child faces serious educational challenges by being placed in the wrong class at school.

Registration defines family structure and can affect agency and social relationships, particularly for women. In Uganda, because South Sudanese women often fled before male relatives, they were registered in Uganda as heads of households, empowering them as the official recipients of UNHCR rations and Ugandan government land. According to War Child staff in Bidi Bidi, the restructuring of power in formerly patriarchal families was the biggest cause of registered domestic violence cases as male heads of household sought to reclaim their historic positions of power upon arrival at the camp.

Refugees have no control over how service providing organizations share their data. For many, the experience of accessing services is one in which information is shared efficiently, whilst for others, the lack of transparency around data sharing is a cause for concern. Zain, a Syrian refugee in Azraq who had accessed UNHCR services delivered by the NGO CARE, said, "I was not employed through the CARE program, but if I go check my records at CARE I know that this information will be in their records. The issue is very simple, I think it should be up to us to share or not share this information about our employment." For many of the people we spoke to, information sharing between organizations in camps was efficient but opaque.

Despite limited opportunity, refugees have the capability to manage data. Although few systems allow individuals control over their data, there are examples that show people have the capability to do so. For example, in Bidi Bidi, health records are recorded by health services INGOs in a "health book," which is looked after by individual patients. Angela, a South Sudanese refugee, whose health book contained the case history of her sick daughter, described how this works: "The Outpatient Department writes all the information….on a piece of paper for free, or you have to buy your own book. It is better to have a book because a piece of paper can tear or you can lose it." This shows how even the most vulnerable individuals can own, manage, and protect personal data—in contrast to the centralised collection and storage of many large-scale humanitarian data regimes.

Ethical dimensions of identity, data sharing, and privacy

The development of identity and data management systems for refugees is fraught with power asymmetries in which refugees lack the opportunity and means to influence the development and application of systems that affect their lives. An ethical approach to refugee identity and data management must therefore be grounded in the experiences of refugees themselves, not just to ensure that refugee perspectives are incorporated into the design of humanitarian data regimes, but also because addressing the interests of the most vulnerable also contributes to responsible data management in the entire ecosystem.

Data protection and privacy are amongst the biggest ethical considerations related to identity and information sharing. There have been a number of recent reports exposing data vulnerabilities in both NGOs[28] and UN agencies[29]—though to date no research has been conducted into the specific harms caused by data breaches, and our research did not identify any concrete examples of such harms. Despite this, these breaches have raised awareness of the ethical issues around identity provision; this, combined with liability concerns and regulatory compliance, drives the development and adoption of privacy policies and practices.

Larger humanitarian organizations have already begun the process of developing organizational policies and practice guidelines around data protection and privacy. The literature review conducted at the outset of the research found that larger humanitarian organizations—who naturally tend to collect more data than smaller organizations—are already addressing issues of data protection and privacy.[30] However even organizations that have developed these policies and practices struggle to implement them fully across the entire organization; and when the cycle of data management includes partner organizations that are not subject to those policies and have not implemented those practices, those partners are the weak link.

Of specific interest during the research was compliance with the General Data Protection Regulation (GDPR) which came into force under EU law on 25 May 2018.[31] It remains unclear exactly what impact the GDPR will have on the humanitarian sector, but it is clear that there will be an impact.[32] Despite this, awareness of GDPR was almost non-existent among the staff that we interviewed in the research locations, and even with staff specifically working with data the implications of GDPR were poorly understood. This report does not examine the possible impact of the GDPR on the humanitarian community, but it obviously raises compliance issues and suggests humanitarian organizations should follow global developments in data policy more closely.

While some organizations may be developing policies and practices, the humanitarian community in general is behind the curve of the broader ethical debate around data. There is no doubt that humanitarian organizations are concerned about privacy issues, but the increase in the types and amount of data collection in the humanitarian sector has happened rapidly, and organizations have not had time or prioritised the updating of policies and practices in line with such developments. It is important to emphasise that this is not the fault of individual staff, but of entire organizations; however this lag is not an excuse for failing to address these issues, particularly when working with some of the most vulnerable communities in the world.

Refugees are concerned about privacy issues when those issues are framed in concrete ways that relate to their lives. In an informal tented settlement in Lebanon, Mahdi described how people care about privacy in relation to both Facebook and the UN: "Facebook has had many updates since it was launched, and there are more privacies that you can use, not everyone knows. So also with the UN. It isn't about differentiating, if someone is careful with Facebook they are also

[28] Nathaniel A. Raymond, Daniel P. Scarnecchia, and Stuart R. Campo, "Humanitarian Data Breaches: The Real Scandal Is Our Collective Inaction," IRIN, December 8, 2017, https://www.irinnews.org/opinion/2017/12/08/humanitarian-data-breaches-real-scandal-our-collective-inaction

[29] Ben Parker, "Exclusive: Audit Exposes UN Food Agency's Poor Data-Handling," IRIN, January 18, 2018, https://www.irinnews.org/news/2018/01/18/exclusive-audit-exposes-un-food-agency-s-poor-data-handling

[30] A list of policy and practice guidelines is included as Appendix D

[31] The aim of GDPR is "to protect all EU citizens from privacy and data breaches" but "it applies to all companies processing the personal data of data subjects residing in the Union, regardless of the company's location" and therefore has global implications. (https://www.eugdpr.org/), http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

[32] Although this impact is discussed in IRIN, "Aid Agencies Rethink Personal Data as New EU Rules Loom," IRIN, January 8, 2018, http://www.geneve-int.ch/aid-agencies-rethink-personal-data-new-eu-rules-loom, there is a need for further research to advance understanding of how GDPR specifically and advances in privacy more generally will impact the humanitarian community, and how they can respond.

careful with the UN." However it was clear that compared to Facebook, the data regimes provided by the UN and other service providers were less transparent and offered less control to their subjects.

This was most obvious in Jordanian refugee camps, where biometrics are used by refugees to purchase goods at officially-sanctioned supermarkets; camp residents had no choice about whether to participate in this system, and had no understanding of how it worked "below the line." This would seem to go against humanitarian principles, which state that aid should be given on the basis of expressed needs rather than on the basis of data shared.

This tension was recognized by a number of respondents: one NGO respondent encouraged his staff to issue aid even if the beneficiary did not want to be registered. "If you could solve how we get beneficiary consent that would be wonderful," he said, "but I don't believe there is a thing as informed consent in this sector. I think the power dynamics are too great."

Informed consent is therefore the point at which the interests of those communities and the interests of the organizations working with and for them come together, and in some senses conflict. Informed consent forms the foundation of data collection for most of the organizations covered in the research, and features prominently in their policy and practice guidelines. Yet the actual implementation of informed consent is largely aspirational, seldom meaningful, and frequently problematic. While organizations recognize the importance of gaining consent from refugees and other beneficiaries to hold and manage their data, they struggle to translate the principle into meaningful practice.

This is not merely a technical issue, since lack of knowledge of and control over how organizations manage their data robs the refugee of agency; and agency is a critical requirement for ensuring the dignity of disaster-affected communities. Even privacy policies appear to be largely formulated without significant inputs from such communities, and the ethical dimensions of identity and data management are thus characterized by a disconnect between refugee perspective and organizational requirements.

For some this disconnect is a benefit, as one international organization staff member described with pride how beneficiaries had no idea they were participating in a blockchain-based cash transfer program—in other words, that they could not be considered to have given their consent to participate in a program they weren't aware of. This is particularly important in the context of the adoption of new technologies—which even when successful may introduce novel risks for vulnerable communities—and increases wider concerns about experimentation in the humanitarian context.[33]

[33] Katja Lindskov Jacobsen, "Experimentation in Humanitarian Locations: UNHCR and Biometric Registration of Afghan Refugees," *Security Dialogue* 46, no. 2 (April 1, 2015): 144–64.

Operational barriers to interoperability

The humanitarian sector has the foundations for a data standard in HXL, but structural disincentives, misaligned funding practices, and diverse operational mandates are strong barriers to further progress

The current status of interoperability

While the humanitarian sector has not yet embraced a comprehensive push toward interoperability, there have been pockets of activity building on earlier experiences in information management. The UN Office for the Coordination of Humanitarian Affairs (UN OCHA) has incubated development of the Humanitarian Data Exchange (HDX), an open platform for organizations to share aggregate datasets (although they explicitly do not include personally identifiable information, or PII) in a standardized format, including some basic visualization functions. A complementary initiative by OCHA is the Humanitarian eXchange Language (HXL), a set of semantic standards for aggregate data that can be easily applied to tabular data (e.g. csv or Excel files) and a limited subset of JSON data.

The HXL standard, created by the HXL Working Group,[34] was designed to be cooperative, so that rather than being built on a new platform, it works with existing spreadsheets. Users simply add a row of hashtags to an existing spreadsheet or database API[35] output. HXL offers a selection of hashtags that can be mixed and matched to suit reporting needs. There is no new reporting channel and no new skills requirements. In fact, the essential information about HXL hashtags fits on a single 4”×6” postcard, suitable for carrying out into the field; this is all that most staff will need to produce HXL-compatible data.

Although HXL has been primarily intended for aggregate data sets and not individual beneficiary data, it has the potential to support identity-related interoperability in two ways. First, the HXL standard itself provides the basis for allowing beneficiary data held by individual organizations to be shared between databases and systems, e.g., through the specification of the most commonly used data fields as HXL hashtags. Second, the community around HXL has developed experience and expertise in data sharing in the humanitarian sector that could strengthen the development of new identity-related data sharing standards.

There are also broader initiatives working toward interoperability of identity systems. In 2018 the World Bank’s ID4D team published a report on technical standards for digital identity,[36] and organized multiple workshops to discuss technical standards amongst a wide variety of stakeholders, including technology vendors, UN agencies, government representatives, standards-setting bodies, and NGOs. Similarly a working group of the World Economic Forum (WEF) is leading multi-stakeholder workshops on digital identity standards across both private and public sectors. While these forums are more focused on national identity systems, there is a sense that lessons from the humanitarian context can inform identity provision in the broader context. The ID2020

[34] Comprised of representatives from the Humanitarian Innovation Fund, IOM, OCHA, Save the Children, British Red Cross, INSO, DFID, IFRC, IDMC, UNHCR, UNICEF, USAID, the World Bank, and the World Food Program.

[35] An API (application programming interface) is a protocol or tool for enabling different software components—including those in otherwise separate databases—to communicate with each other.

[36] World Bank, 2017, “Technical Standards for Digital Identity” ID4D, “Technical Standards for Digital Identity Systems for Digital Identity DRAFT FOR DISCUSSION” (Washington D.C.: World Bank, 2017).

Alliance, a public-private partnership originally funded by Microsoft and Accenture, convenes UN agencies, NGOs, and technology vendors to collaborate on developing next-generation identity systems—including but not limited to the humanitarian context.[37]

Specific to the humanitarian sector, the consortium of humanitarian NGOs which supported this research with logistics and advisory inputs is actively working toward the development of an open technology standard for identity management.[38] Outside the humanitarian sector, there are multiple industry associations focused on standardization and best practices of identity systems, including Open Identity Exchange (OIX), Decentralized Identity Foundation, and the Kantara Initiative.

Barriers to increased interoperability

While all humanitarian actors collect data, sharing that data has not historically been a priority for them. For organizations that do wish to share their data, however, other obstacles appear—not because of malicious intent, but as natural outcomes of the way in which the humanitarian sector and the organizations working within it are structured.

Data flow tends to be uni-directional, flowing up the funding chain from implementing organization (e.g., an NGO) to contracting organization (e.g. , a UN agency) to donor. As a result, those lower down the chain often do not see any benefits from their data sharing, even when it relates directly to their activities: one of the most common complaints in Lebanon was that the existing protection referral system made it impossible for NGOs to follow up on specific cases they had referred. Although RAIS does make information available "downstream" to NGOs and service providers, changes to refugee status, eligibility, case management, etc. are usually only unidirectional.

Existing coordination mechanisms are limited in scope, yet most attempts to promote data sharing rely on them. Despite more systematic investment in information management over the last decade or so, the research found very limited coordination activity in the three country locations. Only one active Information Management Working Group was identified, and that group (in Uganda) was meeting for the first time in two years, despite repeated efforts by UNHCR to convene the group. Such coordination mechanisms offer a forum for information management specialists to explore possible solutions to shared problems, but do not address the underlying obstacles to data sharing.

The humanitarian ecosystem is fragmented, creating structural disincentives to interoperability. Adoption of systems that may be useful across an entire sector, such as Child Protection Information Management System (CPIMS), is difficult to ensure, especially when sectoral leads and/or key donors do not mandate the use of those systems. In Bidi Bidi, when UNHCR transfers responsibility for service provision from one implementing partner to another, the incumbent is required to share beneficiary information with the new organization. This is usually done through Excel, and then updated into the CPIMS;[39] however in at least one instance the incumbent refused to send a digital copy of their beneficiary list because they used a proprietary system—and instead sent over paper copies containing over a thousand names. It took the new implementing partner three months to update its new system, physically verifying each and every beneficiary.

[37] https://id2020.org/

[38] Members of this consortium provided logistics support and advisory feedback for this research project.

[39] International Rescue Committee (IRC), Save the Children, UNICEF, UNHCR and Terre des Hommes have developed and promoted CPIMS as a standard interagency system for child protection. In 2014 development began on a "next generation" platform called Primero, which has online and offline capabilities, enhanced features, and a mobile application, and which supports both CPIMS and the Gender Based Violence Information Management System (GBVIMS).

Short-term project-based funding mechanisms are also at odds with developing robust data management. All but the biggest humanitarian organizations lack the core resources to develop adequate systems, and even those with large budgets often continue to make do with Excel spreadsheets (although it should be noted that Excel continues to be used precisely because it meets most of the needs of these organizations). Legacy systems continue to be used because of the costs of introducing new systems, and when new systems are adopted (e.g., biometric registration) it is often difficult to retroactively apply them to existing beneficiary records.

When organizations do build new systems, they are often at a country or at most regional level, e.g., RAIS, which was developed in Jordan and Lebanon, although its features are now being incorporated globally. While this may be inevitable given the differences between different country contexts, and may have some benefits, there are clearly efficiency gains to be had if some kind of global approach can be developed.

All three research locations faced specific challenges to the idea of nominally universal database systems. Critical dependencies such as internet connectivity are often the focus of discussion, but there are simple problems of matching a universal system to the workflows of specific organizations. In Azraq camp, CARE (the focal point for camp services) initially handled referrals using RAIS, but stopped after a few weeks because NGOs working in the camp that were not UNHCR partners did not have access to the system. They moved to communicating with partners via email, but were soon generating thousands of emails per day, making it impossible to track. They then set up 30 Excel spreadsheets, distributed by email on a daily basis, which they reconciled with RAIS at the end of each month. This paved the way for a move to Google Sheets, which saved time but created problems with confidentiality when staff left an NGO but retained their login permission—so they began to export Google Sheets into Excel spreadsheets and then send them via email as before. They plan to move to the newly-developed CARE Database System (CDS) which is being used in the rest of their Jordan operation, but could not be deployed in Azraq previously because of poor internet connectivity.

These obstacles arise systemically from the nature of organizations and the structure of the sector. They are not necessarily problematic in themselves, as segregated data regimes are in some cases desirable and even necessary; it is desirable that protection data is segregated from distribution lists, for example, in order to ensure confidentiality for at-risk beneficiaries. This underlines that many of the barriers to data sharing are not the fault of any single organization or group of organizations, but are systemic in nature and must therefore be addressed systemically.

Frameworks for improved identity and data management

Technical standardization is not enough—full interoperability requires a framework that also addresses governance, legal, organisational and semantic dimensions

"Interoperability is the ability of organizations to interact towards mutually beneficial goals, involving the sharing of information and knowledge between these organizations, through the business processes they support, by means of the exchange of data between their ICT systems."

— *EU Interoperability Framework*[40]

The drive to increase the interoperability of data management systems—including but not limited to identity provision—is not unique to the humanitarian sector. In the healthcare industry, the nonprofit HL7 has successfully developed international interoperability frameworks for sharing patient identity and medical information, reducing costs for providers, insurers, and researchers.[41] At a regional level, as part of its Digital Single Market initiative ,the European Union has enacted the eIDAS framework, which specifies how EU member states can rely on digital identity credentials from other member states.[42] In this section we build on these and other examples to present a view of the different dimensions and structures that should be involved in an interoperability effort, along with considerations specific to the humanitarian sector.

Enabling some level of interoperability requires organizations to agree on standards which define how information systems can communicate and share data. This includes the specific technical formats and protocols being used, e.g. whether the data is stored as XML or JSON, or the type of biometric template used to process iris scans. While the discussion around interoperability often defaults to technical standards, agreement on these alone is insufficient for realizing interoperability, as many standards simply represent best practice or industry defaults; many of the ISO standards fit into this category, for example.[43]

Achieving interoperability instead requires alignment and standardization across multiple dimensions beyond technical formats, including: semantic definitions of different data types, e.g. the possible values and definitions for the data field of "household"; organizational agreements around operational processes and policies, e.g., how informed consent is collected; and legal frameworks, e.g., national data-sovereignty policies.

These different aspects of interoperability can be seen as multiple layers that must be aligned, from the most technical syntax layer at the bottom, through the semantic and organizational layers to the legal layer at the top. Combined, these layers constitute an interoperability framework that structures how organizations reach agreement at each layer of the stack (see Figure 3).

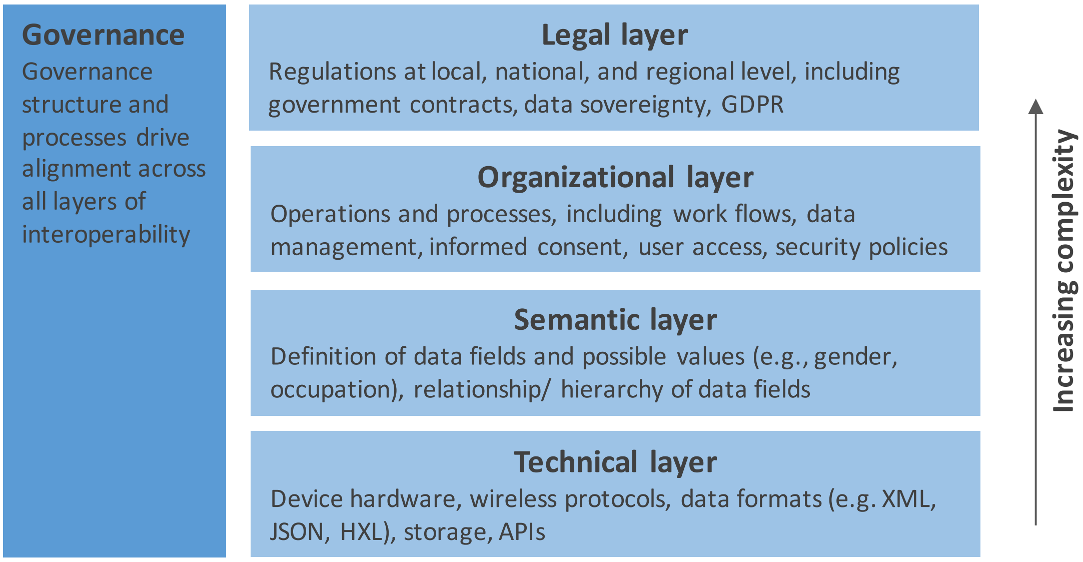
[40] Eliska Kolinkova, "European Interoperability Framework (EIF) - ISA^2 - European Commission," ISA^2 - European Commission, September 15, 2017, https://ec.europa.eu/isa2/publications/european-interoperability-framework-eif_en

[41] For example, see "Personalised Health and Care 2020" (London: Department of Health, November 13, 2014).in which interoperability was a key component, and the FHIR by Health Level 7, http://www.hl7.org/

[42] For an overview, see Turner, "Understanding eIDAS."

[43] For a thorough list of best practice technical standards used in identity systems, see World Bank ID4D, "Technical Standards for Digital Identity Systems"

Figure 3. Interoperability framework for identity systems[44]


Governance
Governance structure and processes drive alignment across all layers of interoperability
Legal layer
Regulations at local, national, and regional level, including government contracts, data sovereignty, GDPR
Organizational layer
Operations and processes, including work flows, data management, informed consent, user access, security policies
Semantic layer
Definition of data fields and possible values (e.g., gender, occupation), relationship/ hierarchy of data fields
Technical layer
Device hardware, wireless protocols, data formats (e.g. XML, JSON, HXL), storage, APIs
Increasing complexity

Governance (cross-cutting) - A governance structure and processes are required to drive alignment across all layers of the interoperability framework. This typically takes the form of a multi-stakeholder industry alliance, with varying levels of formality and membership composition. This body drives the collective discussion and consensus-making process in order to determine not only the standards and interoperability agreements, but also high-level governance mechanisms such as compliance, membership/ affiliation requirements, and use of any associated branding, such as trustmarks. In the case of humanitarian data standards, one of the most critical questions is where such governance sits, and particularly whether it could work within an existing governance structure (such as the Inter Agency Standing Committee), organization (such as the Sphere Project), or process (such as HXL development)—or whether it requires a new and separate body to be established.

Legal layer - This layer is about ensuring compatible approaches between organizations and across different legal jurisdictions, which may include local, national, and regional regulations relevant to digital identity and data management. At the organization level, different contractual agreements with donors may have to be resolved before organizations can legally adhere to new interoperability agreements around data sharing, for example. And for international organizations which operate in dozens of countries, each of which has its own unique legal and regulatory frameworks, ensuring compliance with the different national-level policies can complicate standardization efforts, including, for example, KYC/AML/ATF[45] regulation. At the regional level, many organizations will be subject to the EU's GDPR, yet there remains a lack of clarity around whether UN agencies, such as UNHCR and WFP, will have to comply; this and similar regulations may create yet another set of additional regulations across which to find common ground.

[44] Adapted from European Commission, "New European Interoperability Framework" (Luxembourg: Publications Office of the European Union, 2017, March 23, 2017).

[45] Know your customer, anti-money laundering, and anti-terrorist financing are related regulations governing financial services.

Organizational layer - Interoperability at the organizational layer requires agreement on the policies and processes for managing beneficiary identities and related data, which is complicated by the diversity in programming and operational activities across different organizations. For example, it requires organizations to agree on informed consent practices and privacy policies for beneficiaries—if one organization is going to process data from another, the first has to have confidence that the latter's policies and practices are as reliable as its own. This includes, for example, the processes and information requirements (collectively, the "level of assurance") for conducting identification and/or authentication,[46] as well as general security policies, including encryption practices, authorized users, backups, and disclosure of breaches.

Semantic layer -The semantic layer requires agreement on how to define different data fields and the possible values that these may take in order to facilitate data sharing. For example, is the data field called "gender" or "sex"; and are possible values only "Male" and "Female," or include other options? This layer also refers to data relationships or hierarchies, e.g. ,are children's records connected to the records of the mother, father, or both? Can an individual exist in more than one household? How is a household defined? It is clear that these questions do not just reflect different organizational practices, but also cultural norms and expectations about the individual and their relation to their community. Standards must also consider the level of aggregation of the data, e.g., if data is sampled weekly vs. monthly, or country vs. regional level.

Technical layer - Agreement on the technical formats and protocols for the storage and transmission of data is the most foundational set of standards to enable interoperability. This includes the general data format (e.g., XML or JSON), as well as identity-specific technical standards such as those used for biometrics (e.g. iris templates) or authentication protocols (e.g., SAML, OAuth). The codified nature of technical standards allows them to be more easily defined and documented by standards-setting bodies such as ISO, NIST, IEEE, W3C, and ITU, and the World Bank ID4D team has recently published a review of identity-specific technical standards.[47]

[46] For example, both eIDAS and NIST have published explicit specifications for meeting different levels of assurance

[47] ID4D, "Technical Standards for Digital Identity Systems"

Conclusion

Overcoming the barriers to interoperability in the humanitarian sector will require a sector-wide working group that can agree to a lightweight framework for interoperability amongst existing systems; importantly, for increased interoperability to improve, and not worsen, the agency and dignity of beneficiaries, they must be involved in the design process

This research found that the institutional and financial costs of lack of interoperability are becoming clearer. CARE's experience in Azraq camp, related earlier in this report, led one staff member to note that "sometimes you will go to a location and see 20 staff doing nothing but opening their laptops," but such stories can have a happy ending: each of those staff were estimated to save 2-3 hours per day once they move to the CARE Database System, a total time saving of 40-60 person-hours per day.

By contrast, the costs and benefits to refugees—of both the status quo and proposed changes—have not been accounted for in any depth. While there is well-documented[48] frustration with duplicative research assessments from which they see no benefit, there was no indication from refugees that multiple registrations were wasting their time, in part because registration for services provides tangible value. As one respondent in Bidi Bidi said, "time is the one thing we do have"—anecdotes such as this merely underline the fact that almost no aid organizations have properly measured the costs incurred by refugees due to poor data management.

Thus what emerges from the research is a humanitarian system that talks the language of accountability, but which is actively building systems which makes accountability difficult, if not impossible. This is not because humanitarian organisations are working against accountability, but because the main drivers of information system development are internal management requirements rather than transparency or refugee needs. Institutional interests set a direction, and path dependency then makes it difficult to adapt at a later date.

There is positive movement; policies have been developed, systems are being secured, and approaches such as "privacy by design" are increasingly being discussed. However, critical weaknesses in the humanitarian sector's approach to data requirements remain, and until the refugee perspective is more fully understood and incorporated into design processes, humanitarian data regimes will continue to fall short of the promises made to the refugees themselves.

Based on our review of the landscape, our interviews with staff at headquarters and field level, and our discussions with refugees themselves, our conclusion is that it is neither possible or desirable for any one single identity system to solve all problems and address all the needs of every humanitarian organization. Institutional requirements and capabilities vary widely, and organizations will always need their own internal systems dedicated to their specific operations.

To illustrate this: In each of the three country locations, a significant segment of the refugee population was not included in UNHCR databases. This was not a deliberate policy of exclusion, but a result of the assumptions behind the database design and the specifics of the

[48] Parham et. al., "Lessons from assessing the humanitarian situation in Syria and countries hosting refugees," (November 2013), https://odihpn.org/magazine/lessons-from-assessing-the-humanitarian-situation-in-syria-and-countries-hosting-refugees/

operating environment, particularly the policies of host governments.

For example *de facto* refugees may not have registered with UNHCR, yet still be included in community-based projects in which a CBO addresses the needs of both refugees and hosts, or they may benefit from legal support services for which they do not need to prove their refugee status. UNHCR (or any other single actor) can never provide a universal database simply because they can never have universal coverage.

Therefore instead of focusing on the quality of a single identity system, we propose an approach that focuses on enabling the development of a healthy ecosystem constituted by diverse data regimes. These diverse data regimes, designed to meet the needs of all stakeholders in the humanitarian system, supported by the right set of incentives to build compliance, will contribute to good identity ecosystem outcomes—including the privacy and security of refugees. In this final section we describe what we consider to be the optimal approach to improving interoperability through data standards, and outline what steps key stakeholders can take to achieve this.

This approach will be complicated by the political economy of the humanitarian sector, where competing interests have previously hindered the development of similar standards in areas such as supply chain management. The fragmented nature of the community creates a collective action problem; even once a discussion about aligning policies and practices begins, it is likely to take a long time to mature; there may be ongoing disagreements about exactly how it should be implemented; and there are likely to be NGOs that refuse to participate completely, such as Médecins Sans Frontières (MSF), which frequently refuses to participate in initiatives which might compromise its vision of humanitarian principles. However it is not an option to avoid these discussions; NGOs deliver services outside of the purview of UNHCR databases, to individuals who are not registered with UNHCR, yet who continue to be part of the response.

Furthermore, sunk cost investment in existing systems means organizations are very unlikely to wholly substitute existing systems in favor of a new solution, regardless of its benefits. The focus should instead be on developing a lightweight layer of common standards that each organization can map to its own data when importing and exporting to other entities. This layer should build on the core data fields of the proGres database and the technical foundation of the Humanitarian eXchange Language (HXL) to incorporate beneficiary data and related attributes into data exchange standards.

The requirement for interoperability has become too pressing to ignore, and the need for progress towards standards—including not just technical, but organizational and legal—has become urgent. The benefits are potentially enormous: Investments in data management, particularly in interoperability, will release more staff time from collecting and sharing data to focus on accountability, including data accountability; will help address inefficiency and reduce fraud; and will strengthen data protection across the entire sector by minimizing the number of weak links.

If this is combined with investment into participatory design to ensure that refugee perspectives are included, and that they are given more control over their data—including determining which organizations can access that data, and delegating authority to those organizations as desired—the humanitarian community will finally be able to capitalize on the full potential of digital technology in identity and data management.

Recommendations for DFID

As an immediate short-term measure, DFID should convene a group of donors to align their requirements for policy and practice around data protection, to fund further research into good practice, and to ensure a common approach (for example using the GDPR as a baseline for discussion). They should then use this common approach as the basis for lobbying host governments to ensure data protection legislation extends to displaced peoples; and as a way of engaging implementing agencies (including the Red Cross/Red Crescent movement, UN agencies, and NGOs) in a wider dialogue on data protection and related issues.

The humanitarian community should ensure that "data accountability" is built into future policy and guidance documents for Accountability to Affected Populations (AAP) and other accountability initiatives, such as the Humanitarian Ombudsman.

DFID should support the formation of a short-term multi-stakeholder working group on interoperability, the goal of which will be to agree on an initial baseline standard for interoperability and establish the governance requirements for a longer-term standards body with a specific focus on identity data. The working group should comprise key humanitarian stakeholders, and could be formed under an existing structure (such as the IASC), existing organization (such as the Sphere Project), or existing process (such as HXL development). The NGO consortium that supported this project has a history of engaging on these issues and could serve as an initial forum for said working group.

DFID should support UNHCR to co-chair the working group under its mandate for refugee protection; and also as part of its commitment to opening up proGres to make it both more accessible to service users and more interoperable with other service providers through APIs (since these will only be useful if there are participating organizations on the other side of those APIs). However implementing organizations will need to be fully engaged in order to ensure the success of any standards body, and an NGO representative should fill the other co-chair position.

DFID should support the working group to develop a "translation layer" to enable interoperability between the data management systems of diverse service providers. This would initially be based on a core set of fields that incorporate beneficiary data and related attributes into data exchange, to be maintained alongside (if not integrated into) HXL.

To be as inclusive as possible, the solution should be based on a lightweight set of data and semantic standards that any organization can read/write using APIs or manual translation/tagging; we recommend this be built on HXL. Taking this forward would later become the responsibility of the standards body, which would maintain it as an open (rather than proprietary) standard.

The working group should therefore focus on operational requirements rather than legal mandates, establishing a translation layer that accommodates legal credentials, such as those issued by UNHCR, as well as functional credentials recognised by both governmental and non-governmental service providers. Although legal registration would continue to be important, this translation layer would ensure more consistent access to services, and enable portability between services even in the absence of legal identification.

The provision of a translation layer would enable interoperability of functional identities and reduce the dependence on legal refugee status as a prerequisite for obtaining access to needed services, removing barriers to access and strengthening inclusion.

The standards body should develop explicitly as a successor to the working group, but be open to a wider range of stakeholders. Since these will potentially include commercial vendors and government bodies, it will need to be a new technical body rather than formed under the auspices of any existing structure, in order to maintain its independence from any single stakeholder or set of stakeholders.

This should draw on experience in establishing technical standards in the humanitarian sector (e.g. SPHERE, HXL) as well as experience from other industries (e.g. OIX, W3C) to create a more rapid and flexible development process than earlier efforts. As well as developing and maintaining technical standards, the standards body should also promote core design principles for implementation of the standards—specifically privacy, portability, and shareability by design.

The standards body should establish a compliance framework. The success of these standards will be dependent on the degree of adoption; the standards body could adopt a "command" mode, but a voluntary, incentive-based strategy would be more effective, as the fragmented nature of the sector means a mandatory approach would likely fail.

A possible incentive could be a kitemark, which could be displayed by service providers who meet the standard. Donors (institutional and individual) could use the kitemark as a criterion for supporting service providers, and which beneficiaries could refer to it as part of a broader accountability framework. Commercial vendors who comply with the standards would be able to display the kitemark, which humanitarian actors could then use as a criterion for selecting responsible vendors.

The standards body should also be the hub for a network of trusted organisations—not necessarily limited to UNHCR implementing partners—whose procedures for onboarding refugees have been verified (through external audit, peer review, or some other process) and who can subsequently issue a functional identity on behalf of the entire network.

This trust network would mean that once a refugee has been through the onboarding process with one organization, they would not be required to go through it again with any other organization within the trust network; and larger service providers could offer their identity data as an authentication service to smaller organizations. It may be possible in some countries to extend the trust network to a wider set of service providers (for example, government departments and commercial vendors), which could support refugee integration into services beyond the initial humanitarian response.

DFID and other donors should incorporate data management into the funding process, in order to unlock resources for organizations that are struggling with a lack of capacity. While there is a need for investment at a systemic level—a paradigm shift that recognises that data management must be taken more seriously as a core element of humanitarian operations—medium-term steps by donors could involve:

- Incentivizing better data management through integrating it into proposal criteria and project evaluation more explicitly;
- Including data management as a separate and protected budget line in every project proposal;
- Funding more pilots that test new approaches to data management, not limited to developing new software but also new processes;
- Engaging at the senior management level in order to ensure that there is increased internal support for better data management;
- Agreeing a common requirement for the inclusion of data in agency reporting, easing the burden on agencies' data collection;
- Incorporating funding for better data management into host government capacity building, including service ministries, cartographic and statistical offices.

DFID should particularly support the development of open-source biometric standards and solutions in order to create a more inclusive approach to biometrics overall. Biometrics are a controversial yet central part of identity management in humanitarian response, either in a full-stack identity provision model or through a dedicated biometric service provider (BSP) model: both approaches free humanitarian organizations from having to stay up-to-date with a highly specialized, quickly evolving technology sector.

Depending on the system design, these approaches reduce organizational liability for handling sensitive personal data by absolving them from having to process or store biometric data, which would be completely separate and outside the organization's control; and the ability to pay for complete identity management and/or biometrics as a service instead of a capital investment in hardware and system would be potentially valuable.

Crucially, humanitarian service providers, supported by donors, should invest more in participatory design to ensure ethical identity systems. The research found that individuals have limited understanding of what their data is used for, no visibility into the way their data is shared, and no ability to exercise control over their data. Investment in user experience and data literacy will help refugees make informed choices about their data, potentially increasing registration as common concerns about data sharing are addressed. These measures could also be incorporated into wider communications and accountability mechanisms, particularly regarding policies and practices around informed consent.

Appendix A: Models and Dimensions of identity systems

Models for identity management

The humanitarian sector isn't unique in its pursuit of more interoperable identity systems to replace the traditional siloed model. We describe in this section the different models for identity management in order to give context to how interoperable models may evolve in the humanitarian sector.

Siloed identity systems

The traditional model of identity management is siloed, with every organization or service provider maintaining its own database of user credentials and any associated data. If an individual wants to open a bank account at two different banks, each bank will have to do its own due diligence (KYC check) to enroll the individual, and each bank will maintain a separate record of that user. If she needs to update her information— for example, her address— she has to call up each bank (and the dozens or hundreds of other services that keep a file on her) separately to update her info.

The advantage of siloed systems is that organizations don't need to coordinate, as there is no formal interoperability. But organizations are liable for the personal information they hold, which is increasingly seen as a liability given the probability of a hack or leak. And there are many disadvantages for the end user, who has little control over her accounts and must manage a multitude of different credentials, one for each organization she has a relationship with, which results in high friction and less secure credential management.

Federated identity systems

Commercial online services long ago recognized the value of shared identity systems, and industry standards around OpenID, OAuth, and SAML have enable a generation of internet users to access other services using their Google, Facebook, or Amazon credentials. In this federated model, one or more organizations provide an identity credential that other organizations can use for that individual. Essentially, the identity provider (often called an "IDP") does the initial registration or proofing of the individual, and provides her with credentials, such as a username and password. The user can then use those credentials with any other organization ("relying party") that accepts the IDP as a trustworthy source. The user would enter the username/password with the relying party, the relying party would then send those over to the IDP via its API, the IDP would return back an answer that says "Yes those credentials are valid, and this is the information we can provide you about her profile."

Importantly, the federated model isn't limited to private-sector services. The UK government's Gov.UK Verify program is a highly visible example of a national government using federated identity scheme for access to state services. In that model, the UK government has approved a short list of IDPs, including Barclay's, Experian, and the national Post Office, who provide competing IDP services to individuals.

The advantage of the federated model is that by relying on IDPs, most organizations can avoid having to themselves manage identity credentials and personal data, which are increasingly seen as liabilities that increase security risks. For end users, the federated model can offer significant convenience by being able to use one set of credentials across many service providers. Of course, the federated model only works if the IDPs are trusted entities, and the centralization of identity credentials in a relatively small number of firms can increase the impact of a hack, leak, or downtime.

Decentralized identity systems

Advances in encryption and distributed database technology have enabled a new model for identity typically referred to as decentralized, or sometimes "self-sovereign," identity. In this model the individual controls a digital "wallet" where all of her identity credentials and related personal data are stored. The individual can then grant access to those credentials— or even just some parts of those credentials, such as proof of age—to requesting entities. The decentralized identity model relies on distributed ledgers (e.g., blockchains) and encryption to enable peer-to-peer transactions and digital signatures for verifying claims.

One advantage of the decentralized model is that there are no centralized data stores of PII ("honey pots") or centralized power in a few IDPs, as all credentials and data are distributed across individual wallets. The model embodies "privacy by design"[49] in that all data sharing is by consent and controlled by the user, making it inherently compliant with privacy regulations such as GDPR.[50] Other benefits include the ability to selectively share information, granular control, zero-knowledge proofs, and portability of personal data.

The disadvantages of a decentralised system include a reliance on network infrastructure that is still in its infancy, with very few implementations beyond pilot scale.[51] Importantly, while the technical architecture enables very high levels of individual agency or control, that control requires a level of digital literacy, and possibly technology (e.g. smartphone), that is unrealistic for many populations being served in the humanitarian context. Individuals that cannot self-manage their digital wallets, including not only low-literacy individuals but also children, will have to rely on "guardians" (e.g.. NGOs) to act on their behalf, which negates that benefit for many beneficiaries. While different models have different approaches to governance, even fully open-source, decentralized solutions tend to rely on a core group of developers or stakeholders who have disproportionate influence over how the network evolves.

Potential models for the humanitarian sector

Identity in the humanitarian sector is still highly siloed, though UNHCR does often serve as a *de facto* IDP with its implementing partners, i.e., when partner NGOs check a beneficiary's paper UNHCR credential when providing services, UNHCR is essentially acting as a (passive) IDP. As discussed previously, both UNHCR and WFP have expressed interest in taking on a more formal role in identity provisioning via their respective beneficiary management systems. For example, UNHCR could offer BIMS-certified iris scanners and related hardware, and open up a BIMS authentication service via APIs such that other organizations could scan beneficiaries and immediately get their case file (or whatever part of it UNHCR authorizes). Alternatively, in a fully federated model, multiple NGOs could provide such identity services. This would require consensus around some technical standards and registration/ proofing processes, but would allow more diversity and competition (e.g. around pricing), which should improve services for the relying parties.

Another possible model is for specialized IDPs to carve out a niche in a federated model; the idea is that firms with expertise in authentication technologies such as biometrics could provide dedicated biometric authentication services to NGOs—though some NGOs are wary of using

[49] Privacy by design, as referenced in the EU's GDPR article 25, describes a principle whereby the privacy and protection of individual data--such as pseudo-anonymization and data minimization-- is prioritized in the actual system design and architecture, such that these benefits are structurally integrated into the system and its operations. What constitutes privacy by design will likely continue to evolve with new technological advances and policy changes.

[50] The EU Blockchain Observatory, part of the European Commission, has suggested that many blockchain-based identity platforms may not actually be GDPR-compliant. https://www.eublockchainforum.eu

[51] One example in the refugee context is the WFP "Building Blocks" pilot in Jordan, where the agency has connected parts of its cash-transfer program to a private fork of the Ethereum blockchain. https://innovation.wfp.org/project/building-blocks

biometrics.[52] Instead of being full-fledged identity providers with detailed profiles on beneficiaries, these "biometric service providers" would only hold an iris and/or fingerprint scan and a unique identity number—no PII. They would lease their technology to NGOs, and when the organization scans an individual, the BSP would simply return a "yes" or "no" that the biometrics match the number.

Dimensions of identity systems

Although the metaphor of a value chain and its multiple stages is generalizable across identity systems, there are several critical factors that shape how identity systems are designed and implemented. We describe three of those dimensions here to provide additional insight into the organizational perspective, and to highlight the challenge of reaching interoperability across systems with fundamentally different objectives and structures.

Level of assurance: The operational activities of the organization determine, at least in principle, the type of identity management employed. UNHCR and other organizations that work with individuals in a specific legal framework need to have assurances that they know who they are dealing with, and that that person stays the same throughout the service provisioning process. They are therefore more likely to employ advanced biometrics, whereas organizations distributing, for example, shelter materials to households may only record addresses with names attached, as there is limited utility in ensuring unique identity at the individual level.

The robustness of the identity proofing and subsequent authentication processes is referred to as the "level of assurance." While both the EU (through eIDAS) and the US (through the National Institute of Standards and Technology) have defined clear standards for multiple levels of assurance, most organizations implement variable processes for identity registration and authentication based on local requirements and conditions, with an emphasis on pragmatism.

Scale of uniqueness: Another dimension of these systems is the scale at which they operate. While every identity system strives to establish uniqueness across a population, the population in question can vary greatly.

UNHCR's Biometric Identity Management System (BIMS) is a global database that can determine with a high degree of confidence that every individual UNHCR registers receives a unique record. This provides UNHCR with the capability to know that if a woman registers at a camp in Jordan, and then two years later registers at a camp in Greece, she will be correctly matched to her previous records. Most NGOs do not have such global systems simply because they do not need them, especially because the benefits of such a global system do not outweigh its security risks and budget implications; an NGO providing medical services probably only needs to ensure uniqueness within the immediate community in which it provides those services.

Regulatory compliance: Increasingly the most important use case driving identity system adoption is cash transfer programming, with the Report of the High-Level Panel on Humanitarian Cash Transfers recommending that the humanitarian community should "[w]here possible, deliver cash digitally."[53] This exposes humanitarian organizations to international know-your-customer (KYC), anti-money laundering (AML) and anti-terrorist financing (ATF) regulations;[54] while humanitarian organizations themselves are not subject to KYC regulations since they are not financial entities, their private sector partners are.

A key challenge is that many refugee beneficiaries simply do not have sufficient documentation

[52] Rahman, Verhaert, and Nyst, "Biometrics in the Humanitarian Sector."

[53] The High Level Panel on Humanitarian Cash Transfers (2015). Doing cash differently: How cash transfers can transform humanitarian aid. London: Overseas Development Institute.

[54] UNHCR / World Vision (2012) 'Know Your Customer Standards and Privacy Recommendations for Cash Transfers'.

to satisfy standard KYC identity proofing requirements.[55] In some cases humanitarian actors avoid the issue by ensuring that any fund disbursement remains below the threshold which triggers KYC requirements. Another potential solution is for the NGOs to establish themselves as the recipient of the funds, and therefore become the subject of the KYC process instead of the beneficiaries. In general this remains a very actively negotiated space between government regulators, humanitarian actors, and private sector firms such as banks and mobile operators.[56]

[55] FATF is the most widely used international standard for KYC regulations: http://www.fatf-gafi.org/publications/fatfrecommendations/documents/fatf-recommendations.html

[56] See also: GSMA, 2017, "Mobile Money, Humanitarian Cash Transfers and Displaced Populations"

Appendix B: Interviews and focus groups

We spoke to 36 stakeholders, including donors, headquarters and frontline humanitarian service delivery staff, private sector representatives and academics.

We spoke to at least 200 refugees in focus groups and in-depth interviews, with the following composition:

Refugees	In-depth interviews	Focus group discussions*	Total
Lebanon	9 IDIs with a total of 16 individuals (including with 3 couples & 2 widows)	9 FGDs with a total of at least* 50 individuals (including youth (2), women (2), men (2), mixed adults (3)	At least 66 individuals
Jordan	4 IDIs with a total of 6 individuals (including youth and relatives of people with disability)	8 FGDs with a total of at least 60 individuals	At least 60 individuals
Uganda	26 IDIs with a total of 27 individuals (including youth, 8 Ugandans, 2 unregistered refugees, 8 men, 7 women)	6 FGDs with a total of at least 46 individuals	At least 74 individuals
	39 IDIs	23 FGDs	200

* Most FGDs were conducted in informal contexts or people's homes, so there were multiple people entering and exiting the conversation.

Appendix C: Key identity systems

This appendix includes a summary of the key beneficiary management systems in use in the humanitarian sector, based on the literature review, stakeholder interviews, and field research.

UNHCR: proGres/BIMS

The current version of UNHCR's beneficiary management system is proGres v.3, which is in use in over 70 countries, with over 500 distinct databases globally in use (all data stored in country). ProGres only stores biodata and a photo; any biometrics are stored separately. In the MENA region, UNHCR contracted with technology vendor IrisGuard to provide a full-stack iris-scanning solution, including hardware, proprietary software/templates, and cloud storage and authentication service. Elsewhere UNHCR has deployed its BIMS solution, which captures fingerprint and iris scans into a single database in Geneva? For global deduplication. The next version of proGres, v.4, is currently being rolled out as part of a new internal architecture dubbed PRIMES (Population Registration and Identity Management EcoSystem). The new architecture will combine proGres and BIMS into a core platform that is designed to connect with external partner systems and other UNHCR services. UNHCR is the primary holder of refugee biometric data and granting refugee status.

WFP: SCOPE

The SCOPE beneficiary management system may be the largest currently in use in the humanitarian sector, with more than 20 million beneficiaries registered and a stated aim to enrol all 80+ million WFP beneficiaries. SCOPE was built internally, due to lack of commercially available off-the-shelf solutions that met its needs, and concerns of proprietary IP/lock-in with a vendor-built solution. It is based on a centralized database housed in Geneva, accessed via a web tool, with numerous modules for program management, tracking distributions (in-kind and cash), and business intelligence. While SCOPE does include biometric data, only about 20%-25% of profiles have biometrics attached, as different WFP country offices have different policies on collecting them. WFP is actively promoting SCOPE to be used by other organizations.

IOM: PIRS

The IOM identity management system is the Personal Identification and Registration System (PIRS), which according to IOM is designed to collect, process and store travellers' information including biodata at entry and exit border posts for the purpose of identification, authentication and analysis. It is the only trans-national identity system in the humanitarian sector, though it is not a unified single system. Although there is data sharing between databases, this is on a country-by-country basis. Transfer happens via XML extract and hard disk. WFP and IOM also share data bidirectionally, though IOM systems are fragmented.

WorldVision: LMMS

World Vision contracted out development of its Last Mile Mobile Solution starting in 2006; the system is now used in 29 countries with 8 million beneficiaries registered. Importantly, LMMS has been adopted by more than a dozen other NGOs in the sector, including Oxfam, ICRC, MercyCorps, CARE, NRC, and more. The system provide four core modules: identity management, distribution, cash programming, and business analytics. LMMS wasn't designed specifically to support biometric identity registration, and as World Vision explores the use of biometrics it expect those data to be maintained separately.

Appendix D: Inventory of policy guidance and practice guidelines for data management and cash transfers in humanitarian response

Title	Institution	Year
Responsible Data at Oxfam: translating policy into practice	Oxfam	2017
Model agreement on the sharing of personal data with Governments in the context of registration	UNHCR	
Policy on the Protection of Personal Data of Persons of Concern to UNHCR	UNHCR	2015
Camp Management Toolkit	Norwegian Refugee Council	2008
Responsible Program Data Policy	Oxfam	
Protecting Beneficiary Privacy: Principles and Operational Standards for the Secure Use of Personal Data in Cash and E-Transfer Programmes	Cash Learning Partnership	2013
Privacy Impact Assessment of UNHCR Cash Based Interventions	UNHCR	2015
Responsible Data Management training pack	Oxfam	
The Signal Code	Harvard Humanitarian Initiative	
Data Protection Manual	IOM	2010
Open Information Policy	WorldVision International	
Handbook on Data Protection in Humanitarian Action	ICRC	2017
Data Sharing Policy	MSF	2013
Resolution on Privacy and International Humanitarian Action	International Conference of Data Protection and Privacy Commissioners	
Guidelines on the protection of privacy in the use of mobile phone data for responding to the Ebola outbreak	GSMA	2014
Draft Guidelines for Third - Party Use of Big Data Generated by Mobile Network Operators	LIRNEasia	2014
Privacy and Data Protection Principles	UN Global Pulse	
Principles on Public-Private Cooperation in Humanitarian Payments	WEF	
Principles on Identification for Sustainable Development	Various	
Principles for Digital Payments in Humanitarian Response		
Rules on Personal Data Protection	ICRC	2017
Guide to Personal Data Protection and Privacy	WFP	2016
Building Data Responsibility into Humanitarian Action	OCHA	2016
Principles of Protection Information Management	PIM Working Group	2015
Data Management and Protection Starter Kit	The Electronic Cash Transfer Learning Action Network	????
Policy on Biometrics in Refugee Registration and Verification	UNHCR	2010
The Hand-Book of the Modern Development Specialist	The Engine Room	2016
Protection Information Management Principles	Protection Information Management Initiative	2015
Data Protection and Document Retention Policy	Save the Children	2014
510 Data Responsibility Policy	510 (Netherlands Red Cross)	2017

Appendix E. Humanitarian Exchange Language (HXL) overview

Humanitarian Exchange Language (HXL) tags are easily added to any tabular data source (e.g., csv) by simply adding a new row after the last header row in the spreadsheet.

A basic HXL hashtag looks like "#adm1" (for a subnational administrative level one) or "#affected" (for the number of people affected by a crisis). You can refine the meaning of hashtags by adding attributes after the hashtag, like "+f" (applies to females) or "+children" (applies to children). For example, "#affected +f" means the number of female people affected, while "#affected +f +children" further refines the meaning to the number of female children (in other words, girls) affected. The order of attributes does not matter.

Some hashtags have required datatypes. For example, the #date hashtag must always tag a column of dates or the humanitarian caseload hashtags #affected, #inneed, #targeted, #reached, which must always tag a column of numbers.

Some hashtags have default vocabularies: those are taxonomies of the values that are expected to appear in the column below the hashtag when you add the +code attribute: for example, the default vocabulary for #adm1 +code is +v_pcode (UN place codes from IASC Common Operational Datasets), but you can override any defaults by supplying a different +v_ attribute.

Examples of common HXL tags and their accompanying attributes:

HXL tag	Description	Sample attributes
#country	Country (often left implied in a dataset). Also sometimes known as admin level 0. Defaults to +v_iso3 with the +code attribute if you do not specify a vocabulary.	+code (Is a code or ID) +dest (Is a destination/place) +name (Is a name or title) +origin (Is a place of origin)
#affected	Number of people or households affected by an emergency. Subset of #population; superset of #inneed. Every value must be a number.	+adults (Are adults (people) +children (Are children (people) +displaced (Are displaced (people) +noncamp (Are not in camps (people) +refugees (Are refugees (people)
#beneficiary	General (non-numeric) information about a person or group meant to benefit from aid activities, e.g. "lactating women".	+code (Is a code or ID) +name (Is a name or title) +type (Classifies something by type)
#respondee	Descriptive information, such as name, identifier, or traits, for a single respondee (person, household, etc.).	+name (Is a name) +id (Is a unique identifier)

Additional resources:

- Complete HXL dictionary
- Machine-readable HXL schema
- Printable postcard reference

Appendix F: Bibliography

Caribou Digital. "Identities: New Practices in a Connected Age." United Kingdom: Caribou Digital Publishing, 2017. https://www.identitiesproject.com/report/

Currion, Paul. "The Refugee Identity – Caribou Digital – Medium." Medium. Caribou Digital, March 13, 2018. https://medium.com/caribou-digital/the-refugee-identity-bfc60654229a

Ensor, Charlie. "Biometrics in Aid and Development: Game-Changer or Trouble-Maker?" *The Guardian*, February 22, 2016. http://www.theguardian.com/global-development-professionals-network/2016/feb/22/biometrics-aid-development-panacea-technology

European Commission. "New European Interoperability Framework." Luxembourg: Publications Office of the European Union, 2017, March 23, 2017.

Gelb, Alan, and Anna Diofasi Metz. *Identification Revolution: Can Digital ID Be Harnessed for Development?* Center for Global Development, 2018.

ID4D. "Technical Standards for Digital Identity Systems for Digital Identity DRAFT FOR DISCUSSION." Washington D.C.: World Bank, 2017.

IFRC. "Beneficiary Communication and Accountability. A Responsibility, Not a Choice: Lessons Learned and Recommendations." International Federation of the Red Cross and Red Crescent Societies, 2011. http://www.ifrc.org/PageFiles/94411/IFRC%20BCA%20Lesson%20Learned%20doc_final.pdf

IRIN. "Aid Agencies Rethink Personal Data as New EU Rules Loom." IRIN, January 8, 2018. http://www.geneve-int.ch/aid-agencies-rethink-personal-data-new-eu-rules-loom

Jacobsen, Katja Lindskov. "Experimentation in Humanitarian Locations: UNHCR and Biometric Registration of Afghan Refugees." *Security Dialogue* 46, no. 2 (April 1, 2015): 144–64.

Kolinkova, Eliska. "European Interoperability Framework (EIF) - ISA^2 - European Commission." ISA^2 - European Commission, September 15, 2017. https://ec.europa.eu/isa2/publications/european-interoperability-framework-eif_en

McClure, Dan, and Brad Menchi. "Challenges and the State of Play of Interoperability in Cash Transfer Programming." Geneva: UNHCR & World Vision, 2015. http://www.cashlearning.org/downloads/erc-executive-summary-interoperability-web.pdf

OCHA. "Humanitarianism in the Network Age." New York: UN OCHA, March 6, 2013. https://www.unocha.org/publication/policy-briefs-studies/humanitarianism-network-age

Parker, Ben. "Exclusive: Audit Exposes UN Food Agency's Poor Data-Handling." IRIN, January 18, 2018. https://www.irinnews.org/news/2018/01/18/exclusive-audit-exposes-un-food-agency-s-poor-data-handling

"Personalised Health and Care 2020." London: Department of Health, November 13, 2014.

Rahman, Zara, Paola Verhaert, and Carly Nyst. "Biometrics in the Humanitarian Sector." Berlin: The Engine Room and Oxfam, 2018. http://oxfamilibrary.openrepository.com/oxfam/handle/10546/620454

Raymond, Nathaniel A., Daniel P. Scarnecchia, and Stuart R. Campo. "Humanitarian Data Breaches: The Real Scandal Is Our Collective Inaction." IRIN, December 8, 2017. https://www.irinnews.org/opinion/2017/12/08/humanitarian-data-breaches-real-scandal-our-collective-inaction

"Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)}." Vol. L119. Brussels: European Parliament, May 4, 2016. http://eur-lex.europa.eu/legal-content/EN/

TXT/?uri=OJ:L:2016:119:TOC

"State of Privacy 2018." London: Privacy International, January 2018. https://privacyinternational.org/type-resource/state-privacy

Steets, Julia, Andrea Binder, Andras Derzsi-Horvath, S. Krüger, and L. Ruppert. "Drivers and Inhibitors of Change in the Humanitarian System." *A Political Economy Analysis of Reform Efforts Relating to Cash, Accountability to Affected Populations and Protection. Global Public Policy Institute*, 2016. http://www.gppi.net/fileadmin/user_upload/media/pub/2016/Steets__Binder__Horvath__Krueger__Ruppert__2016__Drivers_and_Inhibitors_of_Change_in_the_Humanitarian_System.pdf

Turner, Dawn M. "Understanding eIDAS." Cryptomathic, January 2016. https://www.cryptomathic.com/news-events/blog/understanding-eidas

UNHCR. "Policy on the Protection of Personal Data of Persons of Concern to UNHCR." Geneva: United High Commission for Refugees, May 2015. http://www.refworld.org/docid/55643c1d4.html

USAID. "Identity in A Digital Age: Infrastructure for Inclusive Development." USAID, September 2017. https://www.usaid.gov/sites/default/files/documents/15396/IDENTITY_IN_A_DIGITAL_AGE.pdf

Warner, Alexandra T., and Alice Obrecht. "Standardising Humanitarian Data for a Better Response: The Humanitarian eXchange Language | ALNAP." London: ODI / ALNAP, March 10, 2016. https://www.alnap.org/help-library/standardising-humanitarian-data-for-a-better-response-the-humanitarian-exchange

WEF. "Digital Identity On the Threshold of a Digital Identity Revolution." Davos, Switzerland: World Economic Forum, January 2018. http://www3.weforum.org/docs/White_Paper_Digital_Identity_Threshold_Digital_Identity_Revolution_report_2018.pdf v

Caribou Digital is the leading independent think tank on digital economies in emerging markets.

Readers are encouraged to reproduce material from Caribou Digital for their own publications, as long as they are not being sold commercially. As copyright holder, Caribou Digital requests due acknowledgment and a copy of the publication.

For online use, we ask readers to link to the original resource on the cariboudigital.net website. The views presented in this paper are those of the author(s) and do not necessarily represent the views of Caribou Digital.

Caribou
Digital