IMPROVING CASH-BASED INTERVENTIONS
MULTIPURPOSE CASH GRANTS AND PROTECTION
Enhanced Response Capacity Project 2014–2015

Challenges and the State of Play of Interoperability in Cash Transfer Programming

UNHCR
The UN Refugee Agency

World Vision

Challenges and the State of Play of Interoperability in Cash Transfer Programming

April 2015

This document was prepared for the Office of the United Nations High Commissioner for Refugees and World Vision International by ThoughtWorks. It was written by Dan McClure and Brad Menchi.

This document covers humanitarian aid activities implemented with the financial assistance of the European Union. The views expressed herein should not be taken, in any way, to reflect the official opinion of the European Union, and the European Commission is not responsible for any use that may be made of the information it contains.

Contents

CONTENTS

CONTENTS

Executive Overview / Approach to Study

This study explores the factors driving the design of interoperability for Cash Transfer Programming (CTP). The study has been developed under the Enhanced Response Capacity for Multipurpose Cash Grants funded by ECHO and managed by World Vision and UNHCR. The work was conducted by ThoughtWorks, a global technology and innovation consulting firm, from December 2014 through April 2015.

We outline the rapidly changing environment in which digital services are emerging, the types of digital collaborations that could be enabled, and the key design challenges that confront effective interoperability. With this design framework in place, the current state of play is mapped against the trends and challenges.

A companion study examining the legal and policy issues associated with CTP has been done in parallel with this work.

State of Problem and State of Play

There are two potential ways to view a complex design challenge. One is as a problem space, focusing on the elements that drive success and constrain possible solutions. This "State of Problem" is the field onto which any proposed solution must be mapped. The second perspective looks at the solution, a "State of Play" which inventories the current set of solutions and practices in use.

In mature fields attention centers on the State of Play, since the rules of the game are stable and well defined. That is not the situation in CTP. This problem space is young and in flux, where the underlying cash strategies and the supporting stakeholder ecosystem are still evolving. Against this shifting field, the current State of Play is a relatively fragmented group of solutions developed for specific needs.

Given the current maturity of the domain, this report gives significant attention to framing the State of the Problem. This is because how the problem is defined has a significant impact on the resulting interoperability strategy. Later in the report we outline a varied set of practices that make up the current State of Play.

Study Approach – Interviews, Research, Experience

To provide a foundation for building current state of play models, interviews were held with members of NGOs currently engaging in CTP (see appendix for list). Interviewees were selected based on their experience working with multipart or cash-based programs. Interviews lasted from 45 to 90 minutes.

The content of the interviews was based on the current technical environment within the organization and their experience with CTP. The conversations explored each organization's current challenges and their vision for the future of beneficiary data management. Subject areas include:

- Capabilities of existing systems
- Processes surrounding data collection and data sharing
- Lessons learned from the adoption process of existing systems
- Obstacles to data sharing and interoperability
- Data definitions and naming conventions

Interview content was augmented with additional research materials and with insights from ThoughtWorks experience designing and implementing complex multi-party systems.

Insight – An Immature, Complicated, Changing Ecosystem

Cash based aid effectively creates an ability to perform self directed spending on the part of beneficiaries.[1] This is a strategy that is still evolving in the aid sector. During interviews, many of the strategies were discussed in hypothetical terms. In other cases where programs were actively being developed, details of actual strategies were not yet being released.

It was clear that this is a relative immature space with change occurring across multiple dimensions. This was reflected in the interviews by the fact that results were quite diverse, with no two organizations sharing more than one or two of the same technical characteristics.

This however is not just a question of varying levels of adoption for proven best practices. CTP is part of a growing inventory of digital services that are in the early stages of transforming the Humanitarian and Developmental aid sectors. We are seeing the leading edge of a shift from physical in-kind goods and services to digitally delivered aid.

The potential impact of these initiatives is substantial. Compared to pure in-kind programs, Cash Transfer Programs promise decreased supply chain costs, local economy stimulus, and a more dignified aid environment for beneficiaries.

This is not the only digital innovation poised to have an impact on aid efforts. The sudden proliferation of small Internet connected technologies, the Internet of Things (IOT), the coming of age of digital printing, and the ubiquitous access to mobile devices means that valuable services can increasingly be delivered in the moment of need without the overhead of traditional in kind services.

The ecosystem of large international NGO's and formal funding models that support them will be under increasing pressure to reimagine itself. With the digitization of services, the barriers to participating in aid efforts fall. It's no longer necessary to have a robust logistics system and teams on the ground to make a difference. Activists in far flung parts of the world can take part in highly specific and creative ways.

The value of collaboration will also increase. Operationally, sharing digital beneficiary data enables increases accountability to donors and enhances data integrity. However, this collaboration will not only be among established industry participants. New stakeholders will appear and have the opportunity to develop synergistic services that build on each other's contributions to individuals in affected populations.

In short we are moving to an environment with more service options, lower barriers to entry, and increasingly diverse stakeholders. It is a transformation that has just begun.

Framing Models – Making Sense of Complexity

The report seeks to reflect the multi-dimensional aspect of this design challenge. Nimble and secure data sharing in this space will require complex and sophisticated strategies that go far beyond the traditional focus on data mapping that defined traditional data sharing efforts.

Interoperability is a messy multi-part challenge with a level of difficulty that is often underestimated. In mature industries with known participants, stable business practices, and working infrastructure, interoperability has focused on building consensus around data standards and transfer protocols. These detailed negotiations generally presume the stability of the underlying sector ecosystem.

[1] Emergency Economies: The Impact of Cash Assistance in Lebanon – International Rescue Committee

This is not the current or even the future environment of CTP and digital aid services. On every major dimension there are disruptive changes that introduce thorny new design problems. The Health Sector provides a useful analog for complex multi-party interoperability challenge that involve personal information. Dr. W. Ed Hammond outlined ten distinct concerns that lie within complete e-Health interoperability solutions:[2]

- Stakeholder Interoperability
- Semantic Interoperability
- Functional Interoperability
- Technical Interoperability
- User Interface Interoperability
- Privacy Interoperability
- Business Interoperability
- Communications Interoperability
- Legal Interoperability
- Environmental Interoperability

The "simple" challenge of mapping data fields is only a small portion of work needed to create an overall interoperability ecosystem.

For the effort at hand, the solution will be highly dependent on the scope and definition of the interoperability problem. This report is structured to provide models that can help make sense of that complexity. The issues are grouped into four principal areas:

1 Rapidly expanding ecosystem of stakeholders

2 Use cases for interoperability

3 Design challenges for creating an interoperability solution

4 State of play for solutions that are currently in use

5 The conclusion of the report provides thoughts on a path forward for exploring these challenges.

[2] http://www.ehealth-connection.org/files/conf-materials/Perspective%20on%20Interoperability_0.pdf

CONTENTS

PART 1

State of Problem: Expanding Range of Stakeholders

Maturity of the Ecosystem

Overview: The current ecosystem for capturing personal information and using it for programming of aid is still in early days. Major transformations are ahead that will substantially alter the aid delivery model across many dimensions.

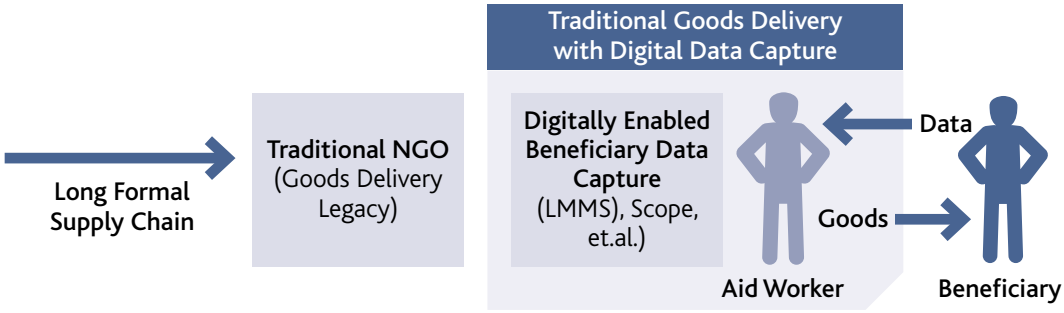
Why This Matters: The current ecosystem is not the target for a long term design. While short term tactical actions will be shaped to current conditions, the overall solution architecture requires a longer term view.

Digital Services Initiate Radical Ecosystem Changes

The shift to cash programming represents a deeper and more pervasive set of changes to humanitarian and developmental programming than the simple shift from a goods to cash benefits strategy.

It is the beginning of a broader shift to a digital services model that enables new benefits strategies, invites in new participants, and expands the range of populations that can be served. The following five contributors the current transformation of the aid ecosystem must ultimately be accounted for in the design of a strategic approach to interoperability.

Stage 1: Digital Beneficiary Capture for Traditional Aid



Tools have been developed to digitally capture beneficiary information in support of traditional goods based aid strategies. Multiple organizations have invested in technology solutions that enable aid workers to capture individual personal information while directly interacting with beneficiaries in the field. In effect the beneficiary data capture tools extend the ability to digitally process information all the way to the point of contact with the beneficiary.

While the marketplace for these tools is still evolving the technologies have become relatively mature and have been deployed in a number of crisis situations. The leading tools have seen robust testing in real life deployments, generating concrete evidence that they can operate at scale in harsh and unpredictable environments.

Legacy Business System Integration

The tools have demonstrated real value in the support of on the ground aid workers. Data can be captured more quickly and consistently, which is a benefit both for the field workers and the beneficiaries. However, much of the strategic operational value occurs when digital information is uploaded to the organization's existing business systems, in support of management functions such as program reporting and audit.

These information flows are inward, requiring the data collection to ultimately align with the types and formats of internal systems. Since many of these legacy operating systems are very large and have their own deeply entrenched technical ecosystems of data repositories, reports and integration points, there is limited ability to change.

As a result, there is pressure on any tool in wide use to support varied demands from different business systems integrations. The tool's ability to enforce a new data standard across the sector may be be limited, even if there is broad adoption.

Still a Fragmented "Tool Market"

This is an early stage marketplace. Products with proven capability have emerged, and development continues to extend their capabilities. However, even as the underlying aid strategies shift, there is no single market leader. Multiple platforms remain in play. Among the leading tools are applications developed by World Vision (LMMS), UNHCR (ProGres), and WFP (SCOpe).

Product: LMMS

Sponsor:

WVI – Developed by World Vision Canada in 2008

Stats:

Canadian International Development Agency Department of Foreign Affairs, Trade and Development (CIDA DFATD) gave a contribution of $900,000 to spread LMMS to other NGOs[3]

By March 2015, LMMS had been deployed to 26 countries spanning Asia, Africa and the Caribbean, and has assisted approximately 2.5 million people

LMMS is a hand-held electronic device and software to support real-time reporting and tracking, remote data collection, beneficiary management, commodity distribution. Beneficiary photos are stored and IDs generated from the software, barcode scanning. Works with no internet, locally. Flexible, can expand. User login to ensure accountability

Key Observations:

Data Capture Tools Available: There are tools available to support the digital capture of beneficiary information. While there has yet to be comprehensive adoption, tools such as LMMS, ProGres, and SCOpe, have been proven in real life deployments

Tool Led Standard Unlikely: There is no dominant tool for capturing beneficiary data, so it is unlikely that a standard led by a single tool will naturally develop.

Integration Drives Variation: Highly varied internal system needs will exert pressure for customization of functionality and data as they are integrated with the data capture tools

Product: ProGres[4]

Sponsor:

Developed by Microsoft and UNHCR, work originally started in 1999, introduced in 2004

Stats:

At end of 2010, operating in more than 250 locations in 82 countries and has provided assistance to nearly 5 million refugees

Product: SCOpe – System for Cash Operation[5]

Sponsor:

WFP

Stats:

2013, expected entire system to be live by end of 2014. By 2013, had been piloted in four countries.

IT solution for cash and voucher project implementation. Has two main parts: Beneficiary and transfer management and an electronic voucher solution. Can register beneficiaries offline with photos and fingerprints. Based on open source technology. Works 'offline'. Can import beneficiary info from other databases.

[3] Project profile: Last Mile Mobile Solutions – World Vision Canada 2013

[4] UNHCR ProGres Solution Overview – Microsoft

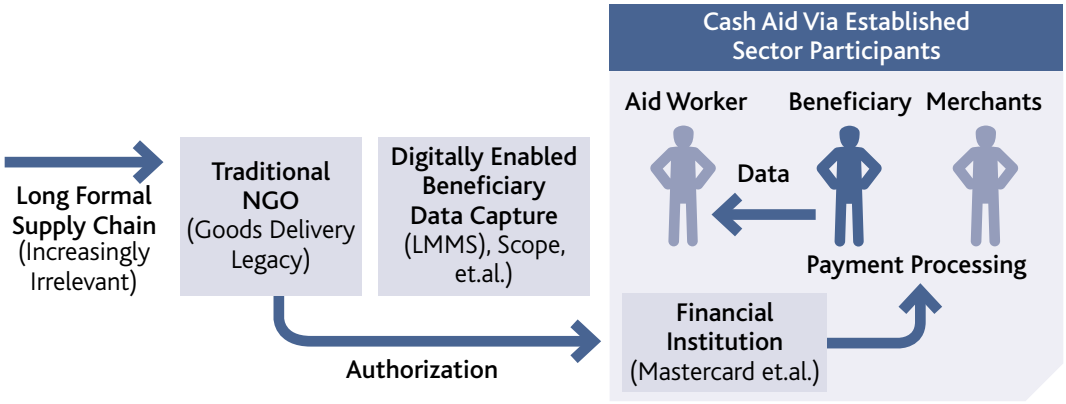
[5] Concept note for WFP technology and innovation demo booth

Key Observations:

- **Data Capture Tools Available:** There are tools available to support the digital capture of beneficiary information. While there has yet to be comprehensive adoption, tools such as LMMS, ProGres, and SCOpe, have been proven in real life deployments
- **Tool Led Standard Unlikely:** There is no dominant tool for capturing beneficiary data, so it is unlikely that a standard led by a single tool will naturally develop.
- **Integration Drives Variation:** Highly varied internal system needs will exert pressure for customization of functionality and data as they are integrated with the data capture tools

CONTENTS

Stage 2: CTP – Swapping Cash for Goods



CTP is a digital version of aid. Digital services enable new players to enter the aid ecosystem, with a wide range of potential distribution options. With so many possibilities, there is no dominant player or approach for these emerging digital services.

Under almost any scenario, the change is significant. At the root of the transformation is a shift from a consuming model to a spending model. In a consuming model, aid is provided to a beneficiary who receives direct benefits. The ecosystem ends with the beneficiary.

A spending model requires the presence of a marketplace. Now the system can easily have four distinctly different players. An aid provider allocates aid. A beneficiary spends their aid. A merchant or vendor provides goods and services in exchange for actual compensation (e.g. cash, credits directly provided to an account) or the promise of payment through some financial provider.

Some of the models discussed include:

- **Individual Account Deposits and Spending:** Beneficiaries receive direct deposits into accounts for use in an existing functional financial system. Tools like M-Pesa have leveraged alternative technologies like mobile phones[6] to create non-traditional local financial networks that while still requiring technology infrastructure (a working mobile phone network and access to mobile phones) is not dependent on traditional banking services.
- **Cash:** In this case, digital tools are used to determine the quantity of the aid, but the actual aid is paid in physical currency. Very little spending infrastructure is required. This model can leverage the local economic ecosystem and naturally provided greater anonymity for recipients. However, it also comes with significant security challenges for both the distributors and recipients.
- **Vouchers / Coupons:** A model similar to the Debit/Credit card, with vendors redeeming the vouchers or coupons. This model in effect creates a shadow economy in vouchers that runs alongside the existing local economy. It requires both a means of distribution and a means of redemption. If the goal is to integrate spending into the local economy, then a significant effort to work with vendors will be needed in addition to the beneficiary allocation. Past experiences by WFP with voucher programs encountered substantial fraud.
- **Captive Marketplace:** Beneficiaries are provided an account but spending is limited to choices within a fixed marketplace of options. Similar to loyalty point programs supported by some commercial credit cards, the market is in effect defined and managed by the aid provider. This increases the ability to setup and control vendor side offerings, albeit with less organic integration into the local economy.

[6] The Economics of M-PESA - William Jack, Tavneet Suri

Variation Today and Tomorrow

At this time there is no a single model for delivering and redeeming cash benefits. In mature commercial markets all of these models co-exist and are applied under varying business and market conditions. Variety exists and continues to grow with the introduction of new financial product innovations. In the aid sector, the following drivers will also make future convergence of service models difficult.

- **Deeply Disrupted/Missing Infrastructure:** Infrastructure is often absent or compromised, significantly limiting which options may be applied. This can be a moving target, since the infrastructure's state typically shifts over time. As the supporting financial and technical infrastructure improve (or degrade) the potential aid solutions also change.
- **Urgency:** The need for a rapid response may force quick solutions that don't provide time for building out supporting market and financial networks. This can also shift over time.
- **Beneficiary's Unique Capabilities:** The education, cultural beliefs, familiarity with technology and other factors demand different types of solutions.
- **Local Context:** The extent to which physical threats are present, whether beneficiaries are centrally located or integrated into a urban area, and the availability of viable market options will drive different solution models.[7]
- **Legal and Political Constraints:** Financial activities can be highly regulated at multiple levels of government.
- **Goals and Purpose of Aid:** There can be multiple dimensions to an aid program's goals. Goals such as empowering women to make financial choices affecting themselves and their families may require different cash delivery strategies.

Who's Legacy Systems Matter Most?

Financial providers have very large deeply integrated processing systems. As a result, financial data interchange standards are in broad use and are well defined for each financial vendor. This will make the existing financial practices of a financial partner a driving force when defining transactional interchange standards.

However, financial service providers are not the only ones using the new cash based data flows. NGO's have deep interest in this information too. Ideally, data interchange standards will also support an NGO's. internal system integration, programming choices, and reporting requirements.

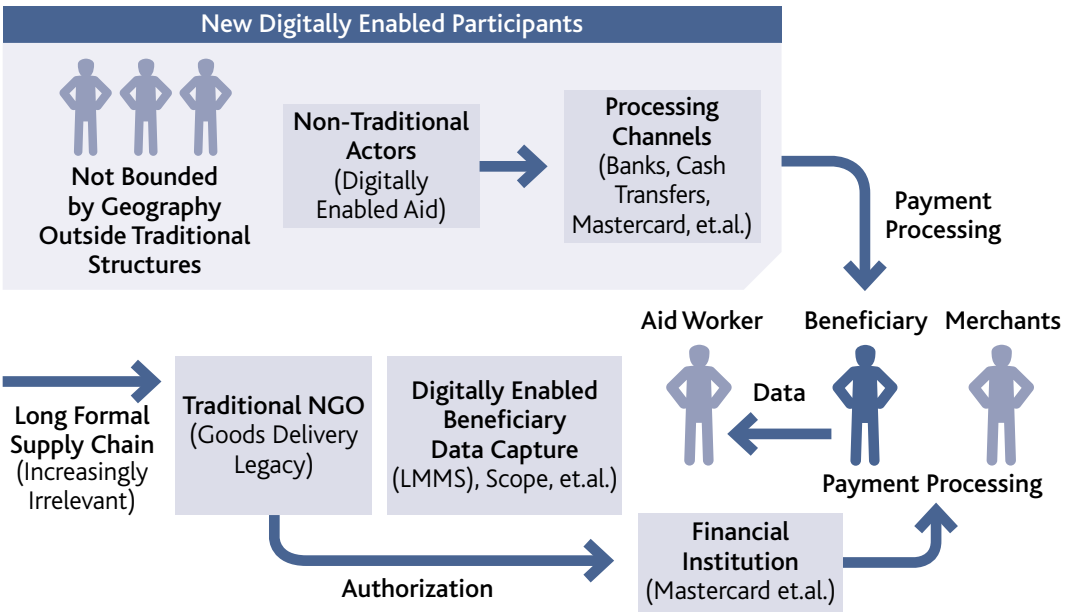
Key Observations:

- **Cash Programs In Early Days:** There are still diverse experiments in the design of cash aid programs. There is no single model that has become a dominant standard for delivering cash aid. The experimentation is likely to continue for some time.
- **Diversity In Approach Likely to Continue:** Multiple drivers including infrastructure, urgency, local context, varying legal constraints, and aid goals make it unlikely that a single model of cash programming will emerge.
- **Multiple Stakeholders Impact Adoption:** While financial providers have a lead role in defining transaction data standards, the ability of other participating organizations to accept and manage data will also impact adoptions.

[7] New technologies in cash transfer programming and humanitarian assistance - Smith, G; MacAuslan, I; Butters, S; and Trommé, M; for the CaLP

CONTENTS

Stage 3: Inviting In New Participants



New creative responses to crisis needs can be developed by established aid organizations. NGO's, which currently structure services around long supply chains for goods and big teams on the ground, will increasingly be able to reframe their services in digital forms.

At the same time, the move to cash dramatically lowers the barriers to entry for other non-traditional participants that want to engage in aid efforts. Organizations like Give Directly provide direct cash transfer to the rural poor of Kenya via the mPesa network.[8] The need for long costly supply chains and logistics capabilities is largely removed from the picture.

Even individuals can actively participate. The ability to send money can be largely democratized. Diaspora populations now transfer funds directly to individuals in areas of crisis, as well as to those who live in areas of persistent poverty.

New Players Arrive / Old Positions Threatened

While the aid sector has a multitude of small participants who move in and out of the field, there has been a striking stability among the major players. Many of the largest aid organizations (Oxfam, World Vision) were founded along with the UN over 60 years ago in the wake of World War II. Some such as the Red Cross have over a century of work behind them.

Stable, long term operating models like these are common in industries where large investments in infrastructure and operating capabilities are needed to support scaling and operational efficiency. Where long supply chains combined and an army of aid providers are necessary, it difficult for newcomers to alter the existing order. You need to be a big organization to participate in this kind of operation.

The shift to digital products alters this equilibrium. New aid providers with their own approach to services and business model, can enter the field. Offerings can be redefined and operations can be developed without the baggage of a goods delivery model.

[8] https://www.givedirectly.org

In commercial fields, digitally delivered services have a history of being highly disruptive to established scale driven players. The aid sector should expect similar challenges to big incumbents. From an interoperability stand point, many more organizations and individuals will need to be involved in the effort to define approaches to interoperability. Big players will have fewer opportunities to define a de facto standard.

Inviting In New Voices – Whose Priorities Will Lead?

Interoperability standards and strategies necessarily reflect tradeoffs between different priorities. In the previous section the tension between different legacy system implementations was called out. As the sector continues there will be more diversity in how data is viewed and used.

In this increasingly heterogeneous environment the designers of an interoperability strategy will need to determine if a central standard for attributes will be developed and enforced or will multiple standards be supported? Who will be certified as trusted partners and how? Who will arbitrate on changes to standards and uses of data?

As digitization of aid progresses, there is less and less reason to assume the continued dominance of leading funding agencies, the UN, and major NGO's. While there are certainly advantages to having established players to help drive efforts forward, "all the usual players" may not reflect the priorities and the fast emerging needs in the space.

Reporting – A Major Use Case With Troubles

The digitization of aid creates both an opportunity and a challenge for reporting. On one hand, the collection and reporting of activity can be done far more quickly and accurately. It finally becomes possible to integrate the data of many different actors to create a holistic view of aid and needs in a community.

At the same time, the previously mentioned fragmenting of the sector and the entry of non-traditional providers, reduces the footprint of the traditional service providers. For major players the likely result, is a better picture of a smaller part of the problem.

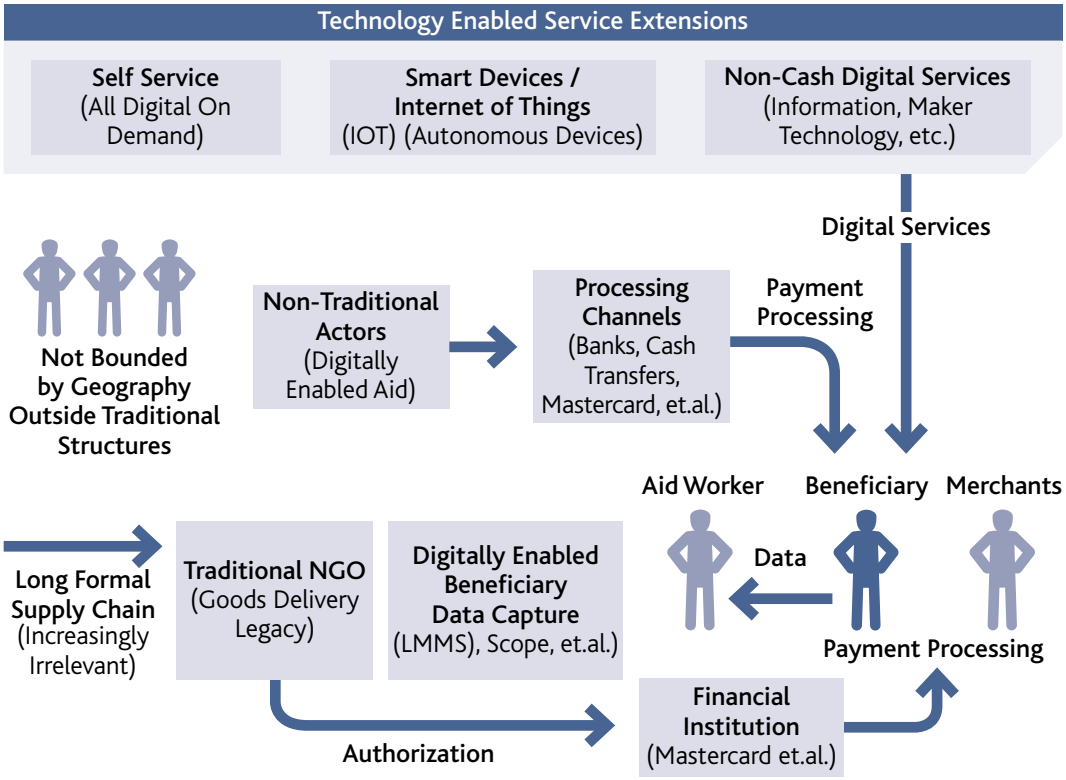
This particular issue matters because reporting is one of the key use cases driving the need for interoperability. The substantial effort needed to standardize data for reporting needs to be balanced against the shifting value such consolidated reports may provide.

Key Observations:

- **Diversifying Sector Participation:** Digitization of aid opens the door for non-traditional aid providers to enter the field. It also undermines the position of dominant incumbents.
- **Questioning Big Player Strategies:** The priority of issues addressed by interoperability strategies will likely shift with the entry of new players. Strategies suited to a small number of large participants will be less viable.
- **Mixed News for the Reporting Use Case:** Digitization will improve the ability to create consolidated reporting from traditional industry participants, while at the same time reducing the footprint subject to their action.

CONTENTS

Stage 4: Expanding Possibilities for Digital Aid



Aid remains a labor-intensive industry, often with low levels of technology use in the field. A number of factors drove this reality. Supporting infrastructure was often absent in crisis areas, technology was relatively expensive and fragile, and technology primarily served business operations needs.

The development of digital data collection and cash based aid is a clear step toward incorporating technology into the actual delivery of aid. However, the change need not be limited to an empowerment of on the ground aid workers.

Broader Definition of Digital Aid

A series of major technology developments are occurring which will make it possible for smart technologies to be deployed directly in the hands of beneficiaries. Digital aid can potentially encompass many more services.

The Internet of Things (IOT) is a series of technologies that provides very small sensors that can actively report on real time conditions. Information that previously had to be manually captured from individuals reporting their data can potentially be obtained in real time without human intervention. The temperature of a home, the health of a baby, the integrity of a building structure are all easily within reach with these technologies.

The ability to act remotely and in real time is increased too. Cash is one form of digitization of aid, but other services can be digitized too, such as the delivery of information. Digital actuators make it possible to control systems so that digital triggered aid becomes possible.

The increasing sophistication of digital printing (maker technologies) extends this capacity even further. Prosthetics tailored to individual needs can be digitally designed anywhere in the world and then printed locally, as can more mundane elements of shelter.

Broader Reach of Privacy and Data Interchange Concerns

The emergence of other forms of digital aid in addition to cash raises the question of where the line should be drawn when defining interoperability standards. At the most basic level a cash standard could cover the basics of financial transaction processing. This however is largely a solved problem in the financial industry.

Extending the view of interoperability to include all the different digital services that might affect an individual's well being is a much bigger challenge. Yet, if different digital services can be traded off or have dependencies, just such an interchange may be needed.

This extension of concern to include independent technologies will impact a wide variety of subjects including privacy and data sharing standards.

Reduced Focus on Manual Data Collection

As more information is collected in context and in real time, data collection tools like LMMS, SCOpe, and ProGres take on a different, and perhaps smaller role. The labor intensive acts of data collection and aid management shift away from field work.

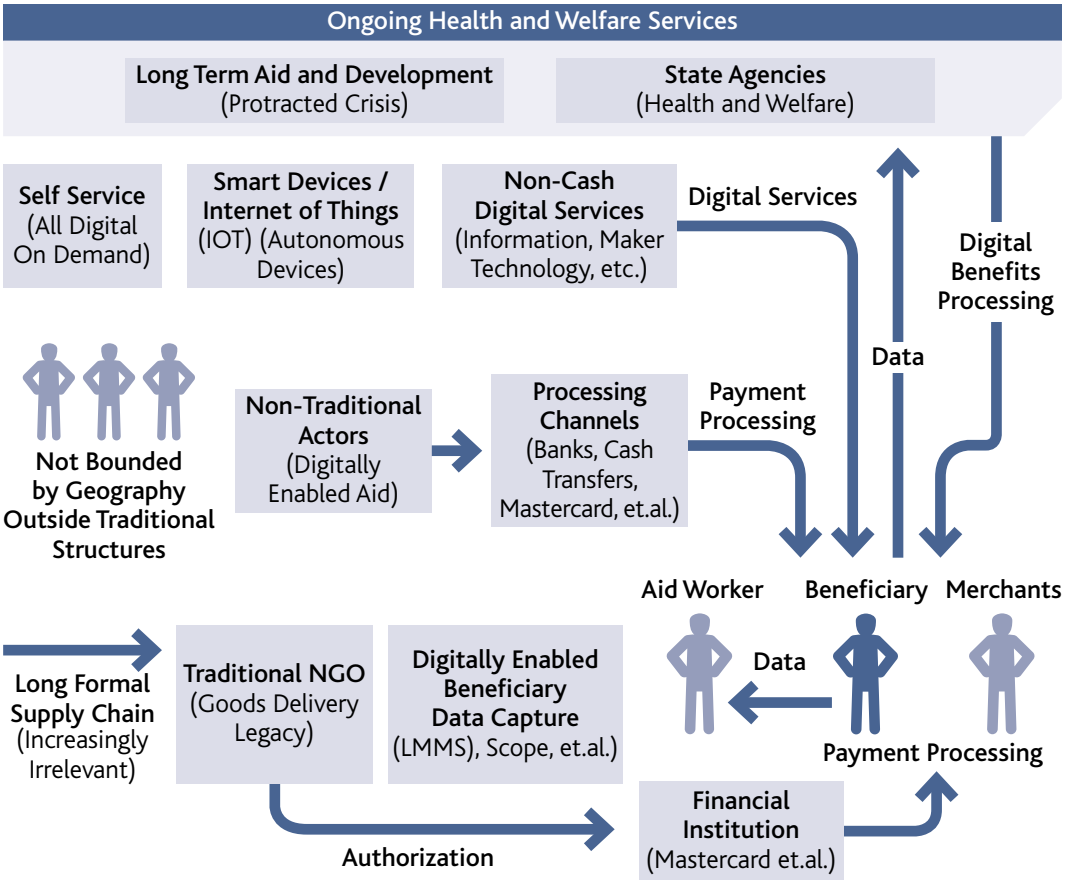
As interoperability strategies and standards are developed, the relative importance of such tools may decline or shift in focus.

Key Observations:

- **More Digital Services:** Cash is not the only digital aid. A growing number of digital services are possible that leverage advances in real time, in context data gathering and aid delivery.
- **Interoperability Beyond Cash:** New digital services may stretch the definition of interoperability and add to challenges defining privacy and data interchange standards.
- **Manual Data Collection Declines:** The development of real time in context alternatives to labor intensive manual data collection may diminish the importance of aid worker tools and processes.

CONTENTS

Stage 5: Expanding the Role from Aid to Human Welfare



The line between Humanitarian Aid and Developmental assistance is already blurry. The extended multi-year displacements are becoming a dominant aid challenge, with populations living in a persistent state of distress. As cash and other digital services become the medium for aid, the rational for this divide becomes less and less relevant. The digitization of services should allow a far more seamless transition from Humanitarian to Development support.

Additional Stakeholders – Government Health and Welfare

A digital services toolset potentially invites in another major group of stakeholders, government agencies that are focus on ongoing services to their populations. Government and other state agencies have similar data collection, analysis, and aid distribution challenges when providing health and welfare benefits to poor and disadvantaged populations. The same digital solutions that enable better aid management and digital delivery of services in crisis could easily be of value to state actors.

This is significant, because these large governmental users may bring their own priorities and focus to the use cases for interoperability. They are also unique in being local in their perspective, not part of the international community.

Key Observations:

- **Transitioning Services Over Time:** Digital services can be more easily transitioned over time. A key interoperability need is therefore between successive aid providers.
- **Government Stakeholders:** Digital services for aid could be used by government agencies for ongoing health and welfare services. This adds local governments to the key stakeholder list.

CONTENTS

PART 2

State of Problem: Major Use Cases

What is the Problem Being Solved?

During the interview process, one of the frequent questions was "What do you mean by interoperability?" As respondents describe challenges and experiences in the space, a wide range of possible answers to this question emerged.

Ultimately, the question of which problem is being solved dramatically changes the types of issues that must be resolved and which organizations are the principal stakeholders in the effort. These functions divide into three major use case, each with substantially different data interoperability needs:

Three Interoperability Use Cases		
Use Case	**Data Needs**	**Level of Data**
Use Case 1 – Transaction Processing	Transaction Values – stable well defined data structures that can be explicitly defined and exchanged.	Individual + Transaction
Use Case 2 – Consolidated Reporting	Attributes – Descriptive data about additional elements that help group, sort, and assign values to elements in a data set	Anonymized Individual Data and / or Consolidated Aggregate Data
Use Case 3 – Collaborative Programming	Transaction + Attributes + Context – Additional enrichment with time and context sensitive information	Individual. Not just transactional information.

CONTENTS

Use Case 1 – Processing Financial Transactions

Goal

As an Aid Provider I want to deliver cash based aid to a beneficiary via a financial intermediary.

Overview

This is traditional processing of a business transactions through a third party. The Aid organization determines the amount of aid and the beneficiary. A financial organization or other third party is then responsible for executing the delivery of funds through their systems.

Participants

This is typically a two party exchange between the aid provider and the financial services processor. The relationship is stable and tailored to the system need of the financial services processor.

Data Privacy

Individual financial information is required to link a beneficiary to an account. Additional demographic or other personally identifiable data are not required as part of the interchange (although the aid organization may maintain these internally).

Key Functionality

The key challenges here are tied to robust transaction processing and account management. Data interchange is generally highly structured and stable.

Data Transfer

- Establish data sharing and business agreements
- Define a data transfer channel
- Define necessary data formats and structures

Account Setup

- Identifying new beneficiaries
- Setting up accounts
- Managing changes to accounts

Transaction Processing

- Request a payment
- Verify request success
- Respond to error conditions

Reporting and Reconciliation

- Reporting of transactions processed
- Reporting on use of funds (account balance etc.)
- Dispute and correction of transactions

Use Case 2 – Consolidated Reporting

Goal

As a Reporting Agency I want to consolidate data about beneficiary needs and aid deliveries to provide a holistic picture of a crisis response.

Overview

Many separate efforts combine together to form a crisis response. Creating a holistic view of the needs and the resulting aid response is a strategically important capability. It drives the ability to plan response efforts, focus resources on areas of high need, and measure the impact of efforts. Information is collected from multiple sources, integrated into a common form, and then processed for analysis and reporting.

Participants

There may be multiple practices involved in both data generation and data reporting. Relationships may shift over time and according to context. Data contributors may leverage varied tool sets, data collection strategies, and local data definitions. For a given data collection program, multiple organizations may use the data to perform their own reporting and analysis.

Data Privacy

To the extent that aggregate information is the output of the reporting, there should be no need for individual name and account information. However, even if the individual records are anonymized, there is still a risk of determining a specific individual's identity from associated attributes (father with three girls and one son arriving from a particular village on a given date). Aggregating individual records can mitigate this danger, but reduces the flexibility in performing later analyses.

Key Functionality

The key challenge here is obtaining information whose meaning and format are sufficiently aligned to allow combined reporting and analysis. Data interchange is likely to be more difficult to standardize.

Data Transfer

- Establish data sharing agreements
- Define available data exchange options
- Define available data formats and structures

Comprehensible Data

- Define available lexicon of terms and values
- Define available information syntax
- Define semantic meaning of terms and values
- Define organization and relationship of data elements
- Define meaning of missing fields (no value, not collected, hidden, etc.)
- Define model for revising data terms, values, and structures
- Define approach for translating from one data model to another

Data Consolidation and Storage

- Define where aggregate data is stored
- Define who can access aggregate data
- Define sunset rules and handing of history
- Define handling of data conflicts
- Data Use and Privacy
- Define what can be shared
- Define who can share
- Define how rules for security, privacy and use will be enforced.

Use Case 3 – Support Personalized Aid Programming

Goal

As an Aid Provider I want to personalize my aid services to the current needs of a specific individuals.

Overview

This use case represents the truly transformational opportunity in aid programming. Here detailed insight in a specific individual's needs and resources provides an opportunity to tailor services in ways to maximize impact. This could be through a single aid provider examining several service options that might be provided to a family, or it could involve multiple agencies coordinating their efforts to avoid overlap and promote synergies between services. In depth information that is current and context specific is the key to this kind of programming.

Participants

There may be multiple practices involved in both data generation and program execution. As with the reporting use case, relationships may shift over time and according to context, however in this case instead of just reporting on data in aggregate, the individual information is used to craft service responses at a personal level.

Data Privacy

This use of personal data provides the greatest opportunity for delivering added value to the individual. However it also comes with the greatest potential risk. Personal identity and details are important, including up to date changes in situation. The very specificity of the data increases its potential use, but also escalates the privacy and security risk.

Key Functionality

The key challenge here is obtaining rich real time information that can be used to form a current picture of an individual's needs and resources. Data interchange is likely to be very rich but also difficult to standardize.

Data Transfer

- Establish data sharing agreements
- Define available data exchange options
- Define available data formats and structures

Comprehensible Data

- Define available lexicon of terms and values
- Define available information syntax
- Define semantic meaning of terms and values
- Define organization and relationship of data elements
- Define meaning of missing fields (no value, not collected, hidden, etc.)
- Define model for revising data terms, values, and structures
- Define approach for translating from one data model to another

Individual Current State Data

- Define what level "beneficiaries" are defined (individual, family, etc)
- Define how individuals are identified
- Define where current state data can be obtained
- Define who can access the current state data
- Define sunset rules and handling of data history
- Define handling of data conflicts

Data Use and Privacy

- Define what can be shared
- Define who can share
- Define how security of use will be assessed
- Define how rules for security, privacy and use will be enforced

CONTENTS

PART 3

State of Problem: Design Challenges

Overview – Designing an Complex Interoperability Solution

Interoperability is a tool that enables collaborations. This is not a simple design space. There is a great deal that needs to be evaluated. While interoperability is often associated with the last century's program of defining data dictionaries and file layouts, this is only a part of the broader challenge that exists with complex collaborations.

As mentioned in the introduction, Dr. W. Ed Hammond identified 10 different design dimensions that contribute to interoperability in complex working domains (in this case e-health). Solving for all possible collaboration scenarios presents an extremely difficult task.

This section of the report provides a basic framework for describing these design challenges. Before breaking the problem down into component parts, it will first be necessary to define how the problem is bounded.

A Key Task Ahead – Placing Bounds on the Interoperability Problem

Aspiring to create an all powerful interoperability solution is a dangerous goal. The number of parties involved, the almost certain conflict between needs, and the variety of edge cases all conspire to reduce the likelihood that a collaborative solution can actually be put in place. Ironically, trying to do the best job possible results in expands the development time and slows adoption.

Therefore a good start to the design effort, is to bound the problem for:

- **Stakeholders:** Which parties are involved? The prior sections of the report have outlined an increasingly complex ecosystem of participants that is rapidly evolving. It would be possible to select the "usual suspects" from the traditional aid sector, but this may not be the best target audience for digital aid strategies that are focused on future innovations. For example Segovia[9] is working on cash distribution platforms for the Humanitarian Sector, but are not deeply embedded in the traditional NGO world.
- **Context:** Which aid challenges are the focus of the interoperability effort? As the scope of possible aid situations increases, the complexity of the interoperability solution escalates.
- **Use Case / Goal:** What is the ultimate goal of the collaboration? Is it primarily to support processing, reporting or programming?

[9] http://staging.thesegovia.com/

Even if being everything to everyone is the ultimate goal, the design strategy is much more likely to succeed if a clearly defined target group of stakeholders and a desired set of use cases is selected.

Focus will help craft an immediately valuable solution. For example a pair of large organizations sending the same operating data year after year can have very tightly defined standards, while a loose consortium of groups working together in new combinations every few months with shifting challenges, will necessarily have a much different approach.

… But Also Designing with Change and Growth in Mind

Cash is likely to be only the first in a number of new and innovative services that can be delivered or enabled in digital form. Given the rapidly evolving nature of this technology and the disruptions that digitization of operations and aid are bringing to the sector, it makes little sense to solve for just the current moment in time.

Solutions not only have to work in the present context, they need to support evolution and adaptation to reasonably expected changes. So, even as the problem space is bounded, it is helpful to outline the expected types of change that may occur.

Not every change needs to be supported right away, but building resilience into the architecture up front will help enable later flexibility, extensibility and scalability.

Challenge 1: Designing for Semantic Interoperability

The first interoperability design challenge is linked to the meaning of the data elements themselves and their logical relationship with one another. Three types of semantic meaning need to be aligned when exchanging data:

- **Meaning of Terms:** What does this variable mean? For example if there is a variable "Family", what types of conditions constitute being part of a family? Is this just parents and children? Extended blood relationships? Anyone cohabiting? Is there a contextual element to the term? Is family the same in every cultural setting?
- **Meaning of Values:** If there is a term Education Level, what are the values that the term can hold? If the allowable values are illiterate, primary, secondary, what defines the scope of each value?
- **Hierarchy and Relationships:** Data gains value by being linked to other terms. For example being able to identify the children of single mothers requires a relationship to be built between the mother and the child. The types of relationships and their meaning give a higher order meaning to the data.

One of the most insidious elements of semantic inconsistency is that it is often invisible in practice. The failure of a Mars mission due to the substitution of metric for English units in a variable is an example of the kind of error that can slip through without notice. Even that error (which would show up as a different range of values) would be easier to catch than a difference in meaning on the definition of family.

There are several common solutions for the problem of aligning semantic content. None are fool-proof and most come with high levels of effort and overhead. Controlled vocabularies or technical systems will be more successful in areas where there are a few strong voices who can define a limited set of fairly stable terms. Schema mapping provides a more flexible and resilient approach, but is less robust and precise.

Strategies for Semantic Interoperability – Controlled Vocabularies

Here the allowable terms, including their definitions and the values they may take are codified and shared as a standard. This is an extremely common strategy, but one that has a shaky history of success. Three challenges exist with this strategy:

- **Consistent Use:** Defining and enforcing consistent semantic meaning with diverse communities is enormously difficult. Forcing a poorly aligned semantic meaning onto a local user community results in misuse or simple repurposing of the variable in ways that are consistent with the local communities sense of the term. Referring back to Dr. Hammond and his experience in the healthcare field. "The issue is further complicated by subtle differences in vocabulary, terminology, nomenclature, taxonomy, classification, and now ontology. In frustration, most sites – large or small – use a local vocabulary and frequently that local vocabulary is not used consistently throughout the institution."

- **Difficult Creation:** A substantial effort is needed to create the standard in the first place. Which stakeholders are invited to define the terms? How are genuine conflicts in meaning handled? Who is the official arbiter? This final question is particularly troublesome in the current space for digitization of aid since the traditional sector leaders are not necessarily the dominant players in the future. Long delays in developing the standard can force participants to move ahead with their own vocabulary definitions.
- **Rigidity in the Face of Change:** The meaning of terms can change either because of shifting usage or because new stakeholders enter the field with different perspectives and needs. Changing standards is particularly difficult at a semantic level because the assumed meanings are typically deeply embedded in practices and systems.

Strategies for Semantic Interoperability – Mapping

An alternate strategy maps meaning between one vocabulary and another. This is a much more flexible approach. It allows terms and values to be developed in ways that respond to local needs and use of stakeholders. New stakeholders can enter the system simply by mapping their terms to one of the other vocabularies or a central shared reference.

This flexibility is well suited to a volatile and evolving domain, however it comes with a great deal of overhead. Achieving consistent meaning (not format) across many data sets is particularly difficult. There will be differences in level of detail or scope of a term that are hard to map and as new participants enter the fray, the overall complexity only gets deeper. We'll refer to Dr. Hammond's experience one last time: "Mapping between terminologies is expensive and will contain errors. Further, the task is never finished, and synchronization among terminology sets is impossible. "

Strategies for Semantic Interoperability – Technology Enforced Rules

One option for relatively simple vocabularies is to enforce compliance and consistency through by embedding the terms and their values into a technical system. Here the terms and their values can be locked down in technology or something as simple as a form.

Complex terms can be broken down into simpler less ambiguous questions and then aggregated up into a consistent meaning by the technology. No one computes their own credit score. The value is constructed from a set of more specific data with less ambiguous meanings.

This strategy requires a technical hegemony, broad based use of a solution that is common across all participants. It also is naturally inflexible and limited in the scope of terms it can effectively manage.

The Impact of Use Cases on the Solution

This is a design challenge that is significantly impacted by the choice of use case. Interoperability scenarios that limit the number of participants and establish a strong controlling voice over interchange content will be much easier to implement.

- **Simplifying the Problem (Use Case 1):** Use Case 1, which focuses on interoperability between transaction processing partners fits this model. Data is tightly constrained based on the requirements of the financial services partner and remains stable, even as other aspects of the aid strategy shift. This narrowing of the problem space reduces the risks associated with semantic consistency.

- **Embracing the Complexity of the Domain (use Case 2 and 3):** If Use Case 2 (consolidated reporting) or Use Case 3 (collaborative programming) are to be supported, there will be many challenges associated with semantic interoperability. These use cases are not simply concerned with performing financial transactions. The most interesting reporting analysis and personalized programming recommendations will be linked to data with more complex meanings. Instead of worrying just about account balance, there will be a need to work with data based on concepts like level of need and risk.

Limited Standardization Today

During the interviews it appeared that selected parts of standard beneficiary information had generally consistent use. Unambiguous items like date of birth and mobile number have similar form and meaning across all systems.

However, in some cases terms were used across several organizations, but still had significantly different meanings. They were semantically inconsistent. For example, vulnerability criteria is broadly used to determine the amount of relief a household or individual should receive, but it is calculated in very different ways. While one score may incorporate sexual violence statistics, another may instead focus on HIV victims in its calculation.

Other data fields have the same meaning and use, but have been collected or recorded in different ways. For example, a beneficiary name could come in many forms based on program location. Given name and surname fields may not be appropriate in many cases. However, just providing one freeform field where this data could be placed doesn't provide the same level of detail for analysis or beneficiary identification.

Not surprisingly, standard global definitions are in use for a number of common concepts.

- **Goods and services:** The UN Standard Products and Services Codes (UNSPSC) has been used to identify goods and services for in-kind relief.
- **Locations:** UN/LOCODE, ISO and GPS have be used to identify locations of distribution points and domiciles.
- **Vulnerability criteria:** The UN provides a standard for vulnerability criteria calculation. Some organizations directly use this standard, while some consider other factors.

Barriers to Single Standards

During the interviews we found that there was often limited standardization even within a single organization. A single mandatory standard is unlikely to be achieved easily. Persistent broad based diversity will make development of a centrally recognized standard both difficult and time consuming.

New data standards also don't need to happen all at once. It may be more manageable to transition certain concepts one at a time. Any shift in language towards a globally recognized standard will contribute to interoperability efforts.

Selectively focusing on areas with established standards, such as standardizing based on location, program, or type of effort (i.e. distribution planning, case studies, short-term disaster relief) may provide a starting point.

- **Which Data Matters:** Financial service providers will drive financial transaction data definitions. The difficult part of the challenge will be determining what data is needed to support the broader reporting and collaborative programming strategies.
- **Create or Reuse:** There are a multitude of data standards available. Once the needed data types are defined, there should be significant advantages from leveraging existing models. This could speed of adoption and provide the opportunity to integrate with an existing community. This would be particularly true if digital aid strategies extend beyond cash and start to more directly integrate with fields like e-health.
- **Standards Outside the Aid Sector:** As the sector restructures itself and non-traditional participants play a bigger role, there may be a need to explore standards that do not originate in the international aid community.
- **Enforce or Map:** What will be the strategy for adoption? Will the standard be presented as a fixed model that must be adopted as is? Or will a capability be developed to map between different versions of the data?

Challenge 2: Aligning Syntax and Communication

By comparison, the alignment of syntax is easier. Here the concern is that the specific format and structure of information is consistent. Because these are explicitly specifiable using a number of technologies, it is far easier to know what the syntax of a data set is and validate compliance with it.

Data Format and Verifying Syntax

As software design models have evolved, the approaches for structuring shared data have sought to balance the need for identifying specific values in a data stream and the desire to have flexibility to adjust the content as needs change.

- **Fixed Format Files:** In this model, a fixed file structure is specified and fields are given designated locations. The specification tells where the data is located, but there may be nothing within the file itself to designate the contents of each column or field in the record. This is in many ways a worst of both worlds solution. Data is not labeled, so only the position in a file identifies the field. It is also difficult to change, since any variation in layout needs to be incorporated by the systems generating and reading the information. Commonly used in EDI schemes, this is largely obsolete as a system design strategy, however it is still a common output format for spreadsheet exports and manual tabulations.
- **XML / Structured Schemas:** In this model data is tagged with an identifier that conforms to a defined schema layout. In effect there is a specified hierarchical structure that can be use to validate that all the data is included and that the data values are actually matched up with the intended variable. This makes the data syntax highly verifiable and easily readable, but it can be difficult to change over time.
- **Serialized Data:** Most recently, lighter weight models such as JSON leverage simpler name-value pairs as the primary way of identifying data. Data elements are still specifically identified, but there is less overhead than XML schema based models and greater flexibility for change.

Data Access Strategies

A number of data exchange models are in wide use. Multiparty technical ecosystems typically involve several of these strategies in order to accommodate existing legacy system requirements and particular user needs.

- **Export / Import – File transfer models:** Data is exported to a file by one application. The file is then either transmitted to a receiving application or the receiving application "picks up" the file at a predefined location. The content is then imported into the receiving application. These jobs are often run on a predefined schedule. Additional communication may occur to identify the success of the load and identify errors.
- **Publish / Subscribe:** When a potential user is interested in obtaining an ongoing stream of data from a publishing organization, they subscribe to the service. When the publisher has information ready, they push the information out to all the current subscribers.
- **Central Repository:** Data is collected in a central repository. Authorized users are given rights to access the data store and perform specific functions.
- **Services:** A specialized service provides users with a specific set of functionality. Data input and output is specifically constrained to support this operation. Many recent technology architectures have migrated to the service oriented architecture.

Added Complexity of Providing an Individual's Real Time State

Historically, organizations in both the commercial and non-profit space have leverage their large data stores to provide reporting and overall operational support. Large quantities of data are processed after the fact for inward facing business needs.

The shift to personal digital services for beneficiaries provides an opportunity to flip this focus around. Data can be leveraged to craft individualized services that are tailored to the specific beneficiary's need at a particular moment in time.

This creates a new type of data need. Customized services require an ability to know the "state of the beneficiary" at a given time. Now instead of simply aggregating up data on a set of services provided, there is a need to know what things have been done, what needs still exist, and what circumstances are present for an individual. Has a mother given birth? Have children gone to school. Has a man's family arrived?

Tracking the current state of an individual requires the ability to see a broad range of information about an individual (not just the aid being provided by a particular person), and see the cascading effects of new events or facts.

Providing this consolidated view, which can be updated in real time, extends the interoperability challenges beyond the simple exchange of data.

Varied Current Technical Maturity

The most apparent conclusion from the interviews was the vast difference in experiences of the organizations interviewed. While some had adopted a digital data management system, many still relied completely on manual methods or digital spreadsheets. Most organizations used a mix of methods depending on the location of work being done. In more remote locations, especially those

lacking consistent technical infrastructure, manual methods were much more prevalent. Pilots of new software solutions and collaborations with 3rd-parties were also common. The varying missions and values for each organization contributed to varying approaches to to data management. For this reason, some organizations have begun work on their own in-house systems, from fully-realized software solutions to detailed manual processes.

Currently, organizations resort to manual methods of data collection for most program situations. Forms or spreadsheets can be customized to fill the needs of a specific program and updated with little effort to account for any program changes. This flexibility comes with increased effort and data insecurity when compared with currently available software solutions. Data is temporarily stored in a file on a remote server and is periodically uploaded to a central location. Any mapping from the spreadsheet to the central data store needs to be done manually to fit the organization's central beneficiary data schema, or separate programs store their in different locations, each with their own schema. In some cases a multi-purpose spreadsheet is used that is responsible for data collection and analysis, with pivot tables being used for any required calculations. These files might be uploaded as-is. After the upload process is complete, users must manually delete files from the remote servers to keep beneficiary data in the one secure location.

Mature solutions like LMMS, SCOpe, and ProGres have been adopted or tried by multiple organizations, especially when collaborating with other users of the software. These solutions offer an extended feature set, which allows a little more flexibility of use between regions and programs. There is also a larger community around such systems, and users access support more readily. For some systems, security concerns are reduced due to built-in protection, and data is shared and uploaded safely. Some systems also provide for the ability to work in remote locations with little technical infrastructure. In such cases, data would be kept encrypted on the remote server until it came in range of a secure network.

Technical Foundations for Data Trust

Knowing the state of the data, when it was collected, when it was updated, and how it has been validated is an important element of building trust and promoting data use.

Interviewees indicated that data received from external organizations was not necessarily relied upon when performing crucial operations. They cited relatively inconsistent data collection processes that organizations undertake over the course of a crisis response. In the first few hours of a disaster, when teams are first on-site, the main object of the relief effort is to get help to as many people as possible. Data collected tends to be much more sparse: a beneficiary might provide only a name and location in order to receive aid. As programs move forward, data collection is much more rigorous and the data becomes more trustworthy.

Trust can be enhanced by providing insight into a number of factors.

- **Reliability of Capture:** To what extent is the data likely to be accurately captured in the field
- **Data Obsolescence:** Is the data fresh? To what extent is the data out of date? Note that what makes data obsolete may be very contextual. In a stable environment data could be months old and still be considered reasonably accurate, while in a rapidly changing situation data that is days or even hours old may be stale.
- **Perspective / Bias:** Were data captured subject to any particular perspective or bias.
- **Provenance:** Who provided the data and under what circumstances
- **Validation:** Has the data been "cleaned" or validated, or is it a raw information feed

Unfortunately, the current technical infrastructure of the sector often does little to help build this trust. It was implicitly recognized in the development of the HXL standard (see Part 4 of the report) that much of the Humanitarian effort is managed thorugh spreadsheets. Without formal models to assure consistent syntax and meaning, it is difficult to determine which data sets are structurally sound. During the interviews a number of examples of spreadsheets traded by email were mentioned.

Data Sharing Policies – Legal Foundation for Trust

Data sharing is generally founded on overarching agreements regarding the bounds of how sharing will occur. Current data sharing agreements between organizations are typically worked out bilaterally or on a case-by-case basis. For example, WVI and WFP have recently agreed on data sharing for a number programs; WVI will provide data collected from these programs for WFP to use in distribution planning. These agreements work in the short-term while the number of players in the space is low, but are a difficult model to expand and diversify.

There can be a chicken and the egg problem here. The development of broad robust data sharing policies may restrict experimentation at the early stages of the development of cash programming. On the other hand the absence of any such policies leaves open important ownership and privacy concerns.

An iterative process, in which data sharing policies are defined, tested and refined may be necessary. Some interview respondents advocated having the technical details of data sharing follow organizational data sharing decisions, not lead them. As it is difficult to forecast technical standards in this space, technical data sharing strategies can converge to certain commonly used standards as they appear. Once data sharing policy details are decided, technical details can 'work themselves out'.

Challenge 3: Designing for Data Privacy and Security

CTP and the broader digitization of aid has the potential to drive an increased level of personal data collection and sharing. This raises a number of personal privacy issues and increases the importance of data security.

The need to deal with these issues is made all the more urgent by two factors that combine to elevate the risk of working with personal information. First, the beneficiaries are often uniquely vulnerable and can be subject to physical danger as a result of data they provide. Second, managing this escalated risk is particularly difficult in a technology ecosystem that has many relatively unsophisticated players amongst a wide shifting network of participants.

To unpack the design challenges, three types of concerns will be defined.

Defining Requirements for Privacy

There are multiple levels of privacy requirements. Each is defined by a different party and is subject to a specific time, place, and circumstance. This makes the determination of constraints on a privacy policy highly contextual.

- **Legal:** These rules are set by outside agencies. This is the minimum bar for privacy requirements, legal restrictions that apply to the collection, storage, and use of personal information. Legal requirements vary from one jurisdiction to another and are continuing to evolve as governments wrestle with digital data rights. These laws are covered in some detail in the companion study to this report.
- **Ethical:** These rules should be set by the organizations gathering and using data. These reflect the ethical questions collectors and users of data need to consider in an attempt to avoid doing harm to their beneficiaries. If a beneficiary is likely to be put at risk as a result of an organization's data collection effort, there is an ethical burden to take that into consideration before acquiring the data.
- **Authorization:** These are constraints that an individual may put on the collection and use of their data. This is based on the growing belief that individuals should be empowered to control the use of their personal information.

Defining the Elements of Privacy

Privacy involves multiple dimensions of data collection, storage and use. In the case of an interoperable environment, these activities may be performed by a variety of different parties.

Presence of Personal Identifiers: The extent to which personal identifiers can be present in the data. Put in a different way, must the data be anonymized? If so, what level of detail must be removed (id, name, personal information, etc.)?

- **Allowable Use:** What types of use are permitted.
- **Retention:** How long can the information be retained.
- **Sharing:** Who can share the data. What restrictions on the type of data shared and the use of the data exist when the data is shared.
- **Permission/Revocation of Permission:** What level of personal control does the beneficiary have over the use of their data. How is permission given and how can it be revoked.

Creating End to End Data Security

Data security is the actual physical implementation of privacy policies, including the protection of data from theft and improper use. It is a "weakest link in the chain" problem, where vulnerabilities at any point along the path from collection to use can result in compromised information.

- **Local Storage – Collection:** This is of particular concern in for the aid sector. Since connectivity is often not available in crisis areas, the tools for data capture often have a local storage capability. This means that a large number of remote devices must be made secure in environments with low degrees of oversight and physical control.

- **Transmission:** Information transmitted over the open Internet is subject to multiple points of attack. The vulnerability is not limited to illicit hackers, but also includes government agencies. Data passing through a jurisdiction may be subject to compromise even if the original collection and ultimate use are elsewhere.
- **Centralized Storage:** The storage locations of data are also points of vulnerability. If data is distributed in many locations, each becomes a possible point for attack.
- **Local Storage – Use:** When data is exported or downloaded to a local device for analysis or other use, there is a risk of poorly managed security.
- **Access and Use:** Authorized users of data are verified through authentication and authorization schemes. Stolen credentials or hacks that bypass an authorization can put data at risk.

Security Risk Mitigation Strategies

One of the most difficult elements of deploying a secure data solution is that frequently the risks are not visible. Leaks and vulnerabilities can easily go undetected. As the sophistication of bad actors increases or new participants extend the system, new points of compromise can emerge in a previously secure ecosystem.

- **Design Secure Systems:** Develop secure system architectures across the full ecosystem.
- **Encrypt Information:** Encrypt information for both storage and transmission.
- **Remove Personal Identifiers:** Only retain personal identifying information when it is actually needed for a critical functional operation.
- **Sunset Data:** Remove data from the system once it is no longer current, or only store fully anonymous versions of the data.
- **Manage Access:** Architectural strategies can be used to place sensitive data in managed secure locations which are then accessed through verified points of use.
- **Require Participant Security:** Validate the security strategies and practices of all participants in the collection, storage and use of the data.
- **Share Only What is Needed:** Limit the data shared and transmitted to that which is actually needed by the user. Don't ship data needlessly.

Significant Risks in a Leaky Ecosystem

Data security and privacy is a key concern with significant current risks. Organizations engaging CTP are collecting increasing amounts of personal data, which can be used to identify the individual from which it was collected. Information from separate datasets can also be cross-referenced to de-anonymize data that previously did not directly identify individuals. This information can put individuals at risk or be used for criminal or political gain. Collecting more information from beneficiaries puts both the organization and the individual at higher risk.

Many data collection methods currently in use, both digital and manual, require manual upload from a remote server (a laptop or other source) to a secure central location. This might be an automated process or require a manual mapping from an initial schema to the current data management standard. Beneficiary data is required for more than solely distribution purposes. Many actors need

access to information for supply chain work and data analysis. Currently, organizations rely on export utilities to create sharable copies of data.

Data management systems only provide security to data stored within them. Files created in the data collection, exporting, and sharing processes may not only remain out of control of a central system for longer than necessary, they may multiply and spread to other locations within the organization, threatening the security of the data within. If care is not taken to ensure files are deleted after being uploaded or used, they will remain possibly forgotten where they were originally placed. If data needs to be shared within an organization, exported files are often sent as email attachments, saving the copies of the file within the email archives of both the sending and receiving parties. Files will remain in remote locations (for example, a personal computer or email archive) until manually deleted.

Each of these remote locations has potential to be compromised, opening up sensitive beneficiary information to outside parties.

Still Evolving Definition of the Problem

Data privacy and security is increasingly recognized as a major data rights issue. For example a conference for Responsible Data for Humanitarian Response was held in The Hague to explore the subject.[10]

It is a complicated space with evolving technical and social dimensions. Perhaps indicative of the evolving nature of the discussion, the conference proceedings from the Hague meeting have not yet been publically released.

[10] http://www.responsible-data.org/programme.html

CONTENTS

PART 4

State of Play: Existing Tools and Standards

Centrally Managed Systems

Key Factors Enabling Centrally Managed Systems

If the problem space of an integration challenge can be clearly defined and controlled, it is possible to develop very robust and comprehensive data strategies. In this case "centrally managed systems" does not mean that there is no exchange of interchange outside the bounds of a single technology platform, but rather that key elements of the problem are under the control of the system owners.

- **Consistent Business Purpose:** The reason for gathering data and the desired business outcomes are stable and well defined.
- **Consistent Business Process:** The processes used for collecting, processing, and using data are defined and in the control of the system owner.
- **Control Over Data Standards:** Since both purpose and process are known and stable, it is possible to define and enforce consistent semantic and syntactical definitions of the data sets.
- **Control Over Change:** Changes to data standards can be evaluated against consistent measures and implemented according to a plan that involves relatively few players.

In effect both the business drivers (purpose and process) and the outputs (standards and change) can be defined and controlled.

proGres – Example of a Mature Registration System

Where conditions exist for a Centrally Managed System, it is possible to create robust and stable platforms with large and complex data sets. An excellent example of this is UNHCR's proGres platform.

proGres captures extensive personal data sets in support of refugee populations and has a clear role in defining an individual's legal status as a refugee. The data collection process is formally defined and is often conducted by professionals. Because the data is tied to a legal qualification and aid processes, there is long-term engagement with individuals who are naturally incentivized to provide updates and ongoing verification of information.

While the system does not currently rely on one master instance, UNHCR has the ability to enforce common standards and system versions amongst copies of the database toolset. Data can be shared outside of proGres, but UNHCR can enforce a single common data schema and base technology platform. Both the semantic meaning of data sets and their syntax can be well defined.

Today 6.6 million people are registered within proGres. It is widely adopted, with an estimated 70% of the global refugee population being managed through the system. The remaining 30% are either managed through governmental system in industrialized countries or are part of small refugee populations for whom a fairly heavy solution is not appropriate.

proGres is not the only large system within this data ecosystem. WFP is currently moving onto SCOPE as its central platform. Integration does occur between these major systems. For a program in Bangladesh proGres handled initial registration, then provided data to SCOPE.

proGres data extracts have also been done with large government based system, such as the US Department of Immigration. Integration capabilities will expand with the introduction of proGres 4.0. The major upgrade will offer a central services model with web client access and expanded API's for data sharing.

Extensibility – Challenges Growing Centrally Managed Systems

The technical ability to pass information to another system is not the only concern when integrating with Centrally Managed Systems. Their robust integrity is driven by an underlying tight alignment of the business and data processing lifecycle. This strength can become a handicap when new needs and priorities enter the design process, either because of new business needs (e.g. the introduction of cash programming) or because new stakeholders bring different priorities.

For example, serving Internally Displaced Persons has different needs and on the ground realities than refugee registration. IDP data tends to be more fragmented and limited in scope. Registration may only be done at the household level with less individual detail.

A system like proGres can (and is) incorporate these variations into its model, but each additional need and change complicates the underlying process and data models. This vulnerability to change can cascade throughout the network of data users. Changes in assumptions about the data (e.g. the core entity may now be either an individual or a household) can require rethinking the handling of information in other systems that have integrated with a presumably stable data platform.

There is also an issue of governance and responsiveness to change. During the interviews this was not cited as an issue in proGres. However, the very nature of centralized control and the need to negotiate and deploy change among a growing number of stakeholders inevitably causes backlogs in requests. This drives the creation of spin off initiatives that reflect the unique needs of a group of stakeholders.

Point to Point Integrations

Point to point integrations tend to grow up around the established islands formed by closed solutions. While smaller in scale, integrations established between, a pair of partners, or a small set of similarly engaged players, benefit from alignments of process and data in the same way that closed systems do.

Point to Point integrations are generally built to support a specific business collaboration. They have the luxury of having stakeholders that can negotiate the details of meaning and syntax. They also are implicitly aligned to some business purpose or process, so the decisions made about data management strategies are anchored in a clear use case.

During our interviews a number of early forays into two party data interchange were mentioned. For example, World Vision International (WVI) has worked with a number of organizations to supplant manual registrations processes with LMMS, a digital tool for last mile data capture. Oxfam has

integrated with LMMS, using it at scale for a number of relief efforts. UNHCR worked closely with LMMS in the Central African Republic, leveraging the digital solution that would avoid the high levels of fraud present in a ration card system.

The Power and Danger of Customization – Biometrics Example

Biometrics is a powerful technology that exposes deep policy questions. These are explored in some detail in the companion report on data policies.

As a strictly technique interchange challenge, biometrics exposes the temptations and difficulties of numerous point to point integrations. Biometric data is highly confidential personal information which is generally encoded to prevent unauthorized use even if the security of a data transmission or data repository are breached. The encoding model is a form of syntax, far more complex than the usual field length and type, but nonetheless providing the same challenges in alignment.

Less obviously, there is a semantic element to determining just how closely biometric readings have to align to be considered a match. The meaning of "match" can differ based on the intended use. Financial transactions might be set to a different standard than general aid qualification.

Point to point integrations have the virtue of enabling customized definitions of these items. This creates flexibility but feeds a growing diversity of approaches. Note that the driver for this diversity is an external change to the available technology toolkit. Changes to programming strategies can have a similar impact, encouraging customization in response to new opportunities.

Open Data and Other Tools Facilitating Data Interchange

A great deal of progress is being made in providing access to data collected by organizations. Two primary developments in this space include the Open Data movement and the HXL lightweight data tagging model.

The Open Data Movement

The open data movement has sought to make large data sets available to users outside the organizations that originally collected the information. Much of the focus of these efforts has been centered on government and other public institutions, where large data sets are gathered as part of their operational mission.

There are multiple data sets and providers. Indicative of the energy behind this activity, datacatalogs.org provides some 400+ local, regional, and national open data catalogs from around the world. Groups such as the Open Knowledge Initiative[11] are actively promoting and extending open data practices. In the Open Knowledge Initiative's terms[12] the drivers for Open Data are:

- Transparency
- Releasing social and commercial value
- Participation and engagement

[11] https://okfn.org/

[12] https://okfn.org/opendata/

As with other "open" movements there is a goal of broad participation and effective use. This is not just about making data visible, but also making it useful. To that end the recommended approach to Open Data includes:

- **Availability:** Data must be available as a whole and at no more than a reasonable reproduction cost
- **Accessibility:** Data must be machine readable, preferably by downloading over the internet and permit intermixing with other datasets.
- **Universal participation:** Everyone must be able to use and redistribute the data, including for commercial or semi-commercial purposes.

Open Data: How Open? How Useful?

International aid agencies are moving to open data standards. The World Bank has an open data policy as do a number of other large organizations in the sector.

What these policies mean in practice can vary depending on what kind of data is published, what level of aggregation is applied, and how often new content is made available. Truly open data, data which is available to anyone, will generally need to be scrubbed of private personal information either by redacting key fields or by presenting only aggregate field levels.

If data is to be used primarily for aggregate analysis and reporting (Use Case 2) then these limitations may not be troubling. However, if the full potential of a personalized digital aid service is to be delivered to individuals based on their private information, then the anonymized aggregated data may be of less use.

One possible alternative to these challenges is to create tools that allow controlled access to data through formal API's (application programming interfaces) that can be easily connected to, but still have selective authorization and authentication. The move to create new API's that facilitate partner access to proGres data is an example of this model.

HXL – Light Weight Humanitarian Exchange Language

HXL, the Humanitarian Exchange Language (HXL)[13], is a joint project by ICT4Peace Foundation, the Humanitarian Innovation Fund, OCHA, Save the Children, UNHCR, UNICEF, USAID, World Bank, and WFP. HXL is a lightweight data standard[14] for tagging information using "hashtags" style labels for a field. A selection of tags are available and each field is labeled as appropriate. It is designed to allow very flexible selection of tags and even to allow different tags to be used in different circumstances. It is compatible with low tech solutions, including spreadsheets.

This model highlights the tradeoffs between flexibility and robust consistency and verifiability. It is very easy to apply and therefore allows quick and easy adoption without the painful process of reconciling data syntax and semantics with other players. The explicit goal of the effort is to "minimise the work and maximize the value of sharing information."

However, these same virtues mean that there is no assurance of repeatable consistency in data structure or meaning. Nor is it easy to codify hierarchical structures and verify that a data conforms to those structures. In Use Case 2 these tradeoffs may be quite acceptable, but they become much

[13] http://hxlstandard.org/

[14] http://docs.hdx.rwlabs.org/standards/

more significant with the details of syntax, semantics and structure are key to processing transactions (Use Case 1) or personalized digital aid decisions (Use Case 3).

Nor does the flexibility of the technique prevent the disruption that comes with change. As of this writing, there is a proposal to delete 70 tags from the data dictionary. This is in no way a critique of HXL's approach, it simply illustrates that the core issues of consistency and change exist even with these alternative toolsets.

Formal Data Standards

There was no clear standard of choice for cash programming during the interviews. This would seem to be expected in the case of the emerging use of financial data. However, the concerns about data standards often extended further, into the broader data surrounding the individual. Four explanations for this were hinted at during the conversations.

- **Scope of Cash Interoperability:** The very definition of what is meant by interoperability in cash programming is still to be defined. Each of the three use cases outlined earlier has very different data needs and would drive significantly different standards.
- **Legacy Integration:** For organizations with existing internal system integrations, the de facto standard is defined by their existing technology. This is a common data interchange problem, which can be solved by data mapping between the internal and external standard. This works, but can create a drag on the ability to change in response to new business needs.
- **Multiple Standards:** As with other sophisticated data rich domains, there are already many formal data standards in the Humanitarian Sector. These may exist strictly as a data standard, or be living data sets. For example, structured XML data sets are available, such as those provided by OCHA through the online COD/FOD database.[15] With multiple options available, simply choosing the right standard can be a challenge to adoption.
- **Shifting Needs:** Even with all these standards, there was a consistent message during the interviews that the standards didn't work for the problems that interviewee faced or that the variability in needs made it difficult to leverage standards consistently. Here the concept of a data standard runs into conflict with the desire to create flexible responsive aid services. The growing diversity of needs and stakeholders might well be expected to drive the development of ever more competing standards.

These challenges to standards adoption, raises the question of how important creating a new data standard would be to meeting the shifting needs of the sector as it moves into digital aid services.

[15] www.humanitarianresponse.info/applications/data

CONTENTS

CONCLUSION

Moving Forward

Defining the Scope of Interoperability Efforts

The "state of problem" in this new rapidly evolving field is potentially large, complex and changing. The scope of the coming work will depend on how much of this challenge is undertaken now and how much is deferred into the future.

Shifting Future: Traditional Aid with Cash/Personal Digital Services

Currently many of the applications for cash in aid are acting as substitutes for historical goods in kind programming. Interoperability in this case is linked to making traditional business processes work with the new technology. Use Cases 1 and 2 provide for the processing of financial transactions and aggregate reporting, functions that are central to current aid practices and will be necessary short term capabilities

The coming disruption will ultimately be far greater. Use Case 3 looks forward to the broader potential of highly individualized real time services delivered through digital channels. These will reimagine current sector practices in multiple ways, substantially changing the types of analysis and data that are used to the types of participants that are involved.

Shifts from Traditional Aid Models to Personalized Digital Services		
From: Traditional Aid	**To: Digital Services**	
Anonymous Crowd	Detailed Individuals	The level of data collected and used will become finer grained and linked to specific individuals.
Historical Reporting	Person's Real Time State	The use of data primarily as an after the fact evaluation of service will shift to an in the moment assessment of need.
Traditional Actors in Silos	Collaborative Ecosystems	The number of aid strategies will grow and diversify. The innovations will shift from optimizing current practices to disruptive new approaches.
Optimized Practices	Disruptive Innovation	The number of aid strategies will grow and diversify. The innovations will shift from optimizing current practices to disruptive new approaches.

Designing for a Robust Present or Resilient Future

It is increasingly common to distinguish between Robust and Resilient system designs[16]. When considering how much scope to undertake when developing a system interoperability, this can be a useful model for contrasting the choices.

Robust designs provide reliable performance of a known set of activities. It is common to lock down variability by centralizing activities in a few official systems or players, establishing common standards that are slow to change, and emphasizing conformance to approach. The practices work passably well in stable business operations where activities are relatively slow to change and are handled by a defined group of stakeholders.

Robust systems are brittle. They work well ... until they don't. Change is resisted and contained until the system finally breaks under the unanticipated demands. A heavy XML based data standard is an example of a Robust design strategy.

In contrast Resilient systems typically have less rigid constraints, providing less structural efficiency, but allowing for growth and change. Resilient systems can evolve to serve new contexts and to accommodate creative new innovation. The lightweight HXL standard aspires to be an example of a resilient approach.

The Danger of Dead Ends

There is always a tension between designing for current and future needs, and it is particularly strong here. In this rapidly evolving environment with ongoing innovation and disruption, three key design traits are needed to create a resilient solution.

- **Extensibility:** New innovations can be incorporated. New players, practices, and services can be invented and deployed without paying a substantial "tax" for interoperability.
- **Flexibility:** Existing practices can be adapted to different situations quickly and easily, and still maintain interoperability.
- **Scalability:** Solutions can be used by multiple players. Adoption and use at scale is easy.

Robust practices do not necessarily lead to future systems with this resilience. Some of the most effective short term solutions, such as the use of centralized systems and single standards, are also the ones most likely to lead to a dead end with few options to add in extensibility, flexibility and scalability.

[16] Zolli, Andrew, Resiliance, Simon and Schuster 2013

Recommendations – Next Steps

Based on our review or this rapidly evolving space, there are a wide range of opportunities to foster interoperability capabilities for CTP and other digital aid services. Potential activities range from near term actions (1–3) to long term strategic enablement (4–5)

1 Answer Questions of Scope

Determine the scope of the interoperability effort (who participates, what services, what use case) that is to be undertaken. These choices will significantly impact the problems that need to be tackled and the actions that should follow.

2 Foster Ongoing Collaborations / Intentionally Gather Learning

This is a domain where both the problem space and the practices are actively evolving. There are new initiatives underway that expand the use of tools and collaborations in the service of new aid strategies. These are an excellent lab for identifying needs and opportunities.

A program of active engagement and learning would help build a rich, ongoing view of the development of the sector.

3 Opportunistically Create Resources

Traditional resources such as data sharing agreements and core data sets can be developed with the recognition that they will not be the whole of a long term resilient solution.

4 Support the Architecture of a Resilient Ecosystem

Provide a focus for thinking and design of the overall ecosystem in support of personalized digital aid. This is not something that can be engineered from the top down, but it will benefit from a choreographer's big picture view of the problem space. Aid in the architecture of the overall ecosystem as it evolves.

5 Strategically Address the Wicked Problems

Some problems, like assuring verifiable data security across a loosely configured ecosystem of collaborators need deep study and solution architecture. These should be prioritized for cross industry efforts, since it is unlikely that single players will have the perspective, resources, or influence to address them.

CONTENTS

Appendix

Interview List

Interviewee	Organization
Keith Chibafa and Gerald Marzano	WVI
Louisa Seferis	DRC
Lili Mohiddin	CaLP (consultant)
Claire Durham	IFRC
Laura Eldon	Oxfam
Andrew Cusack	UNHCR
Edgardo Yu and John McHarris	WFP
Shelley Gornall, Erica Pilcher, Joanna Friedman	UNHCR

CONTENTS

This material was developed as part of the European Commission Humanitarian Aid and Civil Protection Department's Enhanced Response Capacity funding (2014–15).

This inter-agency project was led by the Office of the United Nations High Commissioner for Refugees on behalf of its partners: the Cash Learning Partnership, Danish Refugee Council, International Rescue Committee, Norwegian Refugee Council, Save the Children, Oxfam, United Nations Office for the Coordination of Humanitarian Affairs, Women's Refugee Commission, World Food Programme, and World Vision International.

Humanitarian Aid
and Civil Protection