IFRC

USE CASE 1

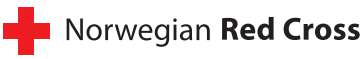
Deduplication of people, families or households

ACKNOWLEDGEMENTS

Funded by
European Union
Humanitarian Aid

This research and publication were made possible thanks to funding from the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Thanks to the authors Robert Worthington and Andrea Duechting, and to all the interviewees and reviewers for their generous contributions.



Thanks also to the members of the Dignified Identities in Cash Assistance consortium (DIGID), the Norwegian Refugee Council, the Norwegian Red Cross and Save the Children Norway for their involvement and support during this research.

Contact us:
Requests for commercial reproduction should be directed to the IFRC Secretariat:

Address: Chemin des Crêts 17, Petit-Saconnex, 1209 Geneva, Switzerland
Postal address: P.O. Box 303, 1211 Geneva 19, Switzerland
T +41 (0)22 730 42 22 | **F** +41 (0)22 730 42 00 | **E** secretariat@ifrc.org | **W** ifrc.org

Cover photo: A woman whose household was affected by an earthquake in Tanahun district, in Nepal receives unconditional cash assistance. (Photo Credit: Danish Red Cross)

PROBLEM SUMMARY

Adoption of cash and voucher assistance (CVA) by the humanitarian community has consistently increased over the last six years. In 2021, 5.4 billion US dollars in transfers reached 1.3 billion people[1]. It is likely that these numbers will continue to grow.

With such large transfer volumes, even a small fraction of duplicate payments may add up to large sums of money with high opportunity costs[2]. In practical terms this means that some people in urgent need of support may not be assisted due to the limited resources available.

There is detailed research on this topic from the social protection sector. As an example, a 2006 RAND Europe study for the UK National Audit Office[3] put the range of fraud, customer error and office error in social protection systems (where data was available) at between two to five per cent of overall government expenditure on social security.

In contrast there is relatively little data available to understand the scale of fraud or error in CVA programmes. Estimates on duplicate cases (which could result from participant error, official error or fraud) from the literature and respondents interviewed ranged from 1 to 15 per cent, with a figure of around 5 per cent being the most commonly cited[4]. These rates are very context specific with a narrow range in some contexts and a wide range in others. This suggests that factors related to the programme design, the type of identity (ID) documents used or the context may influence the level of duplicate registration.

The complex nature of the humanitarian response makes it difficult to arrive at an accurate picture of the scale of the problem. For example, there may be around 100 organizations providing CVA or other complementary types of support in one country. The people they are supporting often lack foundational ID documents. The unit of support may vary from an individual person to a family or a household. Depending on the context, people may also be eligible for different kinds of assistance from different organizations – and at different times. Furthermore, the urgent nature of the humanitarian response may prioritize action over delays which may be needed to verify that the person or household has not been already assisted. These and other factors combine to make this a complex problem.

This use case explores the challenge of ensuring organizations do not duplicate assistance or create gaps in supporting eligible people, families or households.

How does this use case relate to the CVA business process?

Please refer to the CVA business process diagram in the appendix of the main report for an overview of the full CVA business process. The majority of the respondents interviewed incorporated deduplication into the 'Intervention Set-up' stage. This focuses on deduplicating recipient payment lists before a cash transfer or voucher is issued. There is one example of deduplication taking place during the 'Distribution Cycle' stage in Jordan, which is discussed later in this report.

1 Development Initiatives. 2022. Global Humanitarian Assistance Report 2022. https://devinit.org/resources/global-humanitarian-assistance-report-2022

2 This is not always the case. See a later discussion for the nuances of when such opportunity costs may apply

3 National Audit Office, UK. 2006. International benchmark of fraud and error in social security administrations. HC 1387, Session 2005-2006. www.nao.org.uk/reports/international-benchmark-of-fraud-and-error-in-social-security-systems

4 One respondent reported duplicate rates of 70 to 90 in Yemen. However, we were unable to confirm this figure from other sources.

5 UNHCR and CALP. 2017. Review of the Common Cash Facility Approach in Jordan. www.calpnetwork.org/wp-content/uploads/2020/03/ccf-jordan-web-1.pdf

What are the variations of this use case?

There are several variations of this use case:

1. **Deduplication with common unique identifier codes.** This covers the scenario where there is a common identifier code for all the people listed. This could take the form of a foundational ID (for example a national ID number), a functional ID (for example a voter ID card or UN High Commissioner for Refugees (UNHCR) ID). In cases where deduplication takes place at the payment stage, it may be a bank account number, mobile wallet identifier or mobile phone number

2. **Deduplication with biometrics.** This covers the scenario where biometrics is used to verify identify. Concerns related to the appropriateness of biometrics are discussed later in this report. Biometric modalities include fingerprint, face, iris and voice. Raw biometric characteristics are collected and then encoded as a biometric template. Criteria from this template can then be combined using an algorithm to generate a unique identifier for the person.

3. **Deduplication without common identifier codes.** This covers the scenario where people listed have provided different types of ID documents or proxy ID documents to verify their ID. Alternatively, they may lack any ID documents. This may require community validation from village chiefs or other locals to verify that people are unique. As such there are no common unique identifiers.

4. **Deduplication of family or household data.** This covers the scenario where the data is associated with a family or a household that may be identified by a lead person. The decision to manage cases based on individual, family or household data could be a policy decision made by an organization or a programme decision based on targeting criteria.

The specific variation(s) of the use case that applies depends highly on the context. In the case of biometrics it also depends on the programme design and the organization's ability to implement biometrics.

To what extent is this use case relevant to the practitioners consulted?

This use case was broadly relevant to the practitioners consulted. When speaking of interoperability and data sharing, deduplication is one of the use cases most commonly cited. This is also reflected in the literature.

There are a number of drivers for this use case. Many respondents reported it being a priority expressed by donors. Some felt pressure to focus on deduplication even in cases where the context made this less likely (see box). Others framed it from an accountability perspective, optimizing use of the limited resources available to help the people most in need.

Approaches to tackle deduplication should consider the need to deduplicate within one organization (across different programmes) as well as across a range of organizations working in the same place or with the same caseload.

When might duplication be less likely?

Some respondents reported that errors in registering the same cases twice or more would be less likely in contexts where:

- programmes have a shorter duration (from one-off distributions to three, six or nine-month durations)
- community-based targeting is used for enrolment
- implementing partners coordinate their work across different geographic areas.

To what extent – and in which ways – does this use case prevent people from receiving dignified humanitarian assistance?

There is a paucity of direct research to understand the extent to which this use case affects recipients of humanitarian assistance. Clearly CVA programmes have finite resources (and may often be underfunded in relation to the needs assessed). Logically we can then assume that a failure to avoid duplicate support will result in some people in need of assistance missing out, risking further marginalizing people already vulnerable and at risk.

However, data from the Cash Barometer survey in Somalia[6] reported that 52 per cent of CVA recipients indicated sharing their assistance with others outside their immediate household. Sharing resources is very context specific and in this case ingrained in Somali culture and driven by the urge to help people in need. This consideration, however, makes it complex to generally determine the effect on recipients of humanitarian assistance of not attempting to avoid duplicate support.

What is clearer is the impact of attempting to avoid duplicate support. Two key themes emerge from the literature and the respondents interviewed.

First, in many cases efforts to avoid duplicate support – if not managed well – will have negative consequences for people receiving humanitarian assistance. Several respondents reported that they resulted in longer waits from registration to provision of assistance. The specific challenges that may lead to these delays are discussed later. Other respondents raised that people may be put at risk if their data is shared via insecure means for unregulated purposes of deduplication.

Second, one of the drivers for avoiding duplication is the perception of risk of fraud or error by the people receiving support. However, as discussed earlier, there is limited evidence to substantiate the extent to which fraud or error occurs. We could find no objective evidence to separate the likely rates of duplicates occurring due to deliberate fraud, official error or recipient error. Some respondents provided some anecdotal examples of recipient error that led to duplicate cases:

- A refugee camp in Somalia where people both intentionally and unintentionally registered multiple times. This was understood to largely result from a lack of information and understanding on different programmes and registration approaches. People just wanted to make sure they were on the list of recipients.
- Over 10,000 duplicate registrations on self-registration apps and portals related to the Ukraine response. In some cases, they were understood to be attempts to update the number of people associated with a household, rather than to register a duplicate case.

While skilled people might be in a position to cheat the system, this must be balanced with the risk of excluding people who are less skilled and knowledgeable. Hence, this debate is better framed in relation to inclusion, exclusion and quality programming.

6 Ground Truth Solutions. 2022. Rights, information and predictability. Cash Barometer Somalia. www.groundtruthsolutions.org/projects/cash-barometer

Efforts to avoid duplicate support arising from error should hence consider the following challenges:

- Offering faster, safer and simpler processes which increase trust from potential recipients.
- Improving transparency, information sharing and communication on the full range and types of support that people may be eligible for, from who and for how long, and whether or not they have been accepted for assistance.
- Accommodating the rights of people not comfortable or willing to share their data.
- Ensuring that people can update their eligibility profile if their situation changes.

How do practitioners currently address this use case?

Solutions to this use case are highly context dependent. Some of the solutions that are widely used across the sector include those developed by and for UN agencies, notably SCOPE for the World Food Programme (WFP), proGres and PRIMES for UNHCR, BRaVe for the International Organization for Migration (IOM), as well as solutions that are commonly used by non-UN stakeholders, namely RedRose, and Last Mile Mobile Solutions (LMMS) developed by World Vision International.[7] They are also often closely interlinked with the payment and referral use cases.

When discussing the approaches to deduplication below, respondents reported a range of protocols used to share personal identifiable information. These ranged from password protected and non-protected Excel files sent by email to digital storage systems (for example Azure Blob Storage to application programming interface/API). In the majority of cases respondents are looking for a better solution to this use case (for example interoperable and integrated systems between UN Common Cash Statement partners).

Shared registry of recipients

This approach relies on a centralized registry. Different organizations may have access to register a new case or it may be centrally managed. The type and level of access is determined by segregation of duties and the partnerships or contractual relationship (for example consortia, implementing partners).

One example encountered is the Uganda response where UNHCR uses proGres to manage a registry of people needing assistance in refugee camps. Partners able to provide assistance may request a targeting list that matches their assistance profile. In the case of Uganda, data on the assistance provided is shared back with UNHCR, such that the registry tracks who has received what kind of assistance. This is similar to how integrated beneficiary registries work in the social protection sector.

This model relies on a single agency having the mandate and data controller rights to take on this role.

Deduplication with common unique identifier codes

Several respondents shared examples of solutions developed for this use case variation. These are based on several foundations.

First, at the governance level a data-sharing agreement. This describes which organizations are taking part in the deduplication process, what data is shared and for what purpose. Most examples encountered were one-to-one agreements. However, in several cases one organization (UNHCR for example) had similar one-to-one relationships with other organizations.

The agreement must then be operationalized. This means agreeing on which specific data points will be shared and updated, in what format, by what mechanism and how often. Additionally, what criteria will be used to define a potential duplicate and what action will be taken in cases where a potential duplicate is identified.

7 For more information, see the list of tools and systems in the main report.

Respondents report that this approach is time consuming to establish and highly context and relationship specific, perhaps taking up to three months and more to establish the global data-sharing agreement, followed by another month to operationalize it. In cases where the programme only runs for twelve months the results of the deduplication may come too late to be useful.

Due to the nature of different legal frameworks such as the General Data Protection Regulation of the European Union (GDPR)[8] and privileges and immunities, UN-to-UN agreements are faster to negotiate than UN-international non-governmental organization (INGO) agreements, for example. However, once established the process is relatively quick to complete.

In cases where one organization (e.g. UNHCR) is checking for duplicates with a large number of other organizations then presumably the time and complexity of the process will multiply.

Cross-border response to the Ukraine crisis

In Bulgaria, Poland, Romania and Slovakia, the International Federation of Red Cross and Red Crescent Societies (IFRC) collaborates with UNHCR and national societies. IFRC has data-sharing agreements in place with each National Society and a regional two-way data-sharing agreement with UNHCR.

It took around a month to operationalize the process. Each month both organizations generate and export a list of all approved cases. This is consolidated into a standard, pre-agreed set of fields. Separate files are created for each country of enrolment and each assistance programme.

The files are shared using Azure Blob Storage. Both organizations download each other's files and match based on Ukrainian tax number, Ukrainian passport number, or temporary person ID.

Pre-determined business rules set out which organization will retain a matching case based on:

- who provided cash assistance (if only one organization is paid/in process)
- first cash assistance date (if both paid/in process) and first enrolment date (if neither paid/in process)
- rare ties are determined with amount of assistance provided or discussion and agreement.

The outcome file is shared back using the same mechanism. They can then be completed to identify any discrepancies. If necessary a meeting is scheduled to discuss the outcome files and resolve any discrepancies in matched cases or in retaining organization.

Deduplication without common identifier codes

Several respondents described the challenges of solving this use case variant. Take the example of Somalia. There is no foundational ID available. There are 10 different types of functional ID that are commonly used (including for 'know your customer') but an additional 10 to 20 that are sometimes used (e.g. pre-war passports or passports from Southern authorities). Furthermore, if Arabic names are recorded in the Latin alphabet there are different ways they could be spelled.

Therefore, any deduplication process would first require agreement on:

- how to transcribe names from Arabic or any other language to the most commonly used language (usually Latin languages such as English, French or Spanish)
- which data points to collect
- which forms of functional ID to prioritize.

For example, many partners reported transcription problems when wanting to deduplicate or provide payments to Ukrainians affected by crisis. They described cases where, despite national standards to transcribe from Cyrillic to Latin, IDs with Cyrillic characters did not match the system's Latin characters or the information held by the recipient's bank, resulting in rejected assistance.

8 Which sets a higher standard

The Collaborative Cash Delivery Network (CCD) has developed some guidance[9] for its members on this problem. It has been used as the basis for deduplication in Ethiopia, South Sudan and Ukraine.

In some contexts implementers have used proxy identifiers as a partial solution. These might take the form of a QR code or other token that is hard to copy. They can be handed out during registration and returned during distribution. It is not clear whether this approach would avoid a person enrolling twice.

The Dignified Identities in Cash Assistance (DIGID) project has also focused on this challenge. In 2021, the DIGID consortium ran a pilot study on the use of digital IDs in two contexts in Kenya: informal urban settlements and rural areas. The pilot targeted people who had no form of ID and who had been affected by the COVID-19 pandemic, providing them with humanitarian cash assistance. As part of the pilot beneficiaries were issued with digital wallets and credentials. For people without a phone, the credentials took the form of a printed QR code that could be used to withdraw cash from a financial services provider (FSP).

Zero proof approaches

A number of solutions using zero proof approaches have been piloted. Some of the common characteristics of these are:

- A registry holding a list of IDs (or hash codes of IDs or public Pretty Good Privacy (PGP) keys) registered by all implementers.
- The registry can be queried by an implementer to determine if an ID already exists or not.
- This could be done using a hash code generated from the ID instead of transmitting the ID.
- The registry could be centrally managed or it could be decentralized with each implementer running their own node on a blockchain.
- This solution allows implementers to determine if the person exists on the registry without having to share personal identification data.

These approaches provide significant benefits for recipient privacy and security. They focus on sharing a hash code or a public PGP key. This is sufficient to determine if the person to which the hash code or public PGP key relates is held on another database. It can simply respond to confirm if it holds a matching record or not.

If appropriate security precautions are taken, then it is statistically improbable that the person's details could be discovered from the information shared. However, it is important to note that sharing an encrypted version of an ID would still be considered as personal data under GDPR legislation.

One limitation of this approach is that it does not help establish what nature of support the person is entitled too. Hence, it is only a solution to deduplication among organisations providing the same package of support.

What is a hash code?

A hash algorithm (such as SHA-256, SHA-384 or SHA-512) takes an ID number as input and generates what appears to be a random string of digits as the output. The same ID number will always generate the same hash code if the same algorithm is used. The above algorithms are widely used and not known to be cracked. A brute force attack on the SHA-512 hash, for example, would require about 10 to the power of 120 times the age of the universe with all of the Earth's present computing power.

An alternative common attack is to test known ID numbers (stolen somewhere else) or billions of random values formatted like valid ID numbers. To prevent those attacks, IDs must

9 CCD. CCD Deduplication Process. https://docs.google.com/document/d/1OkS2RCqweUlTynnkGffuJ5nt243RNW-Q/edit

be 'salted'. The so-called 'salt' is a long and complex chain of random characters that are appended to the ID number before being hashed. When querying the presence of a number, one needs to know the salt chain and append it before they hash the requested ID number.

Some of the solutions being used include the following.

Building Blocks

Building Blocks started as a pilot in Pakistan in 2017. It has since been adopted in Bangladesh, Jordan, Lebanon and Ukraine, and piloted by different partners in different contexts. The website states that it includes data on 1 million people mainly assisted by WFP[10].

It functions as a collection of blockchain nodes (hosted on its own computer servers which are independently operated by each participating organization). It is therefore an example of a decentralized registry where recipient data is held by each organization independently[11].

While not stated on the website, it presumably stores a salted hash code on the blockchain nodes. This enables participating organizations to query each blockchain node to determine if a recipient is already held on their node.

Genius Tags

Genius Tags[12] has previously been piloted in Syria and is now being piloted in Nigeria[13] with the Cash Working Group. Participating agencies use the GeniusChain app to generate a unique identifier code from a set of personal information. This is stated as being done locally, after which the personal information is deleted. The code is then shared on the blockchain. The technical details of how the code is generated are not provided on the website or literature.

The blockchain is described as being distributed, which would suggest that each agency operates its own node. It also states that "all beneficiaries can have accounts on the GeniusChain, they can interact with the chain directly." This helps get beneficiary feedback on all operations, including opinions, suggestions and complaints.

Biometrics

A later section of this report discusses the specific interoperability challenges related to biometric data. One example[14] where this has been overcome is between IOM (BRaVE system) and WFP (SCOPE system) in South Sudan. A data-sharing agreement was established in 2018 to govern the exchange of biometric data.

IOM and WFP have exchanged data for more than 700,000 people across the country. The purpose stated for data sharing was to avoid duplication and redundant processes and track population movements in case of further displacement.

9 CCD. CCD Deduplication Process. https://docs.google.com/document/d/1OkS2RCqweUlTynnkGffuJ5nt243RN
10 https://innovation.wfp.org/project/building-blocks
11 While Building Blocks can operate as a collection of nodes, in Jordan the blockchain runs on a single node, operated by the technology providers but controlled by WFP. See https://sovrin.org/wp-content/uploads/14A-Report.pdf
12 www.geniustags.com/geniuschain
13 https://acrobat.adobe.com/link/review?uri=urn:aaid:scds:US:29cb7ab8-d79c-3422-88d4-5d9fd80091a5#pageNum=1
14 IOM. 2019. IOM, WFP Conduct First Beneficiary Data Exchange in South Sudan, IOM news, 28 June 2019. www.iom.int/news/iom-wfp-conduct-first-beneficiary-data-exchange-south-sudan

Current blockages and challenges

This is one of the most complex use cases with a range of challenges and potential blockages.

Governance challenges

A CALP report summarising learning from the Ukraine response[15] emphasizes the need for a high level of harmonization of programming:

Agencies should agree on a common data definition, minimum dataset and purpose of duplication and map out agencies' processes and timeframes from registration to payment. This could be part of mandate of in-country Cash Working Groups and/or bilateral data-sharing agreements.

However, in cases where there is complementary assistance provided by different sectors the governance challenges are more complicated. As CVA is increasing used (on it's own or as a component) to provide protection, food, housing and other types of assistance it is complex to determine what constitutes duplication.

The same person, family or household may be eligible for separate support from different CVA implementers working in different sectors. To fully deduplicate assistance requires a new level of coordination to determine what eligibility means for each sector. This also changes as the humanitarian situation develops over time.

This study was not able to map in detail the extent to which different contexts have made progress tackling these and other governance-level challenges. The picture from the respondents was very varied. In Zimbabwe for example, respondents reported no coordinated data-sharing agreements beyond funding modalities. In other contexts significant progress has been made.

Several respondents highlighted the need for governance frameworks that include mechanisms to proactively identify and correct errors, as well as regularly update the data and respond to requests and complaints from recipients relating to inaccuracies in their data.

A final challenge identified is the tendency for donors to fund innovation pilots. These are often implemented in isolation without a broader strategy to iterate, scale or replicate.

Operational challenges

Establishing a data-sharing agreement (to facilitate deduplication and other collaboration) is an important first step. However, headquarters-level agreements often lack sufficient clarity on the operational aspects of the agreement (e.g. standards, protocols, timing). Several respondents raised the time needed (at least one month) to operationalize a data-sharing agreement in a specific context.

Complying with different data protection legislation and frameworks also presents challenges. This requires careful segregation of duties and data access for data processors (e.g. volunteers, implementing partners) versus (joint) data controllers or data subject rights.

Eligibility criteria is one of the most significant operational challenges. Depending on their vulnerability profile, a household may be eligible for multiple types of assistance or different levels of assistance. One example cited from the Ukraine response covered two agencies deduplicating cases based on common identifiers. However, eligibility criteria were not included in the data shared. As a result cases flagged as duplicates may potentially be eligible for support from another agency.

Several respondents raised the challenge of updating and verifying a household's dataset, particularly in the Ukraine response where people can self-register. The lack of attention to this need was cited as the cause for many duplicate registrations.

15 CALP. 2022. Registration, Targeting and Deduplication: Emergency Response inside Ukraine. Thematic paper. www.calpnetwork.org/wp-content/uploads/2022/09/Registration-Targeting-and-Deduplication-Emergency-Response-inside-Ukraine-Thematic-paper-1.pdf

Others raised cultural dimensions of operationalizing a technology to support deduplication. Iris scans for example require women wearing veils to remove them or face recognition requires taking images which are culturally unacceptable for some population segments. This might have the unintended consequence of encouraging more men to register.

The lack of formal policy positions on biometrics is another blockage. Relatively few organizations have the expertise or capacity to consider the benefits, costs and risks of biometrics carefully and develop an organizational position, let alone a policy. One particularly complex area is the appropriate legal basis for collecting and processing biometric data. This is particularly challenging for organizations subject to GDPR. While in some contexts consent could be used as a basis, it is harder to justify in emergency situations where fully informed and freely given consent may not be possible.

The lack of organizational and sector-wide policies limit the feasibility of broad adoption of biometrics as a component of a deduplication system.

Voice biometrics is potentially a less contentious option. While it is accurate at confirming ID (i.e. is this person who I think they are?) it is not accurate at using a person's voice to identify them from a database of different voice biometrics. This reduces some of the risks associated with biometrics.

Technical challenges

The sharing of biometric data for deduplication is particularly complex and faces several technical challenges. First, biometric data can be collected using different modalities (e.g. fingerprint, face, iris and voice) which are not interoperable.

However, interoperability can also be a challenge for data collected using the same modality. When enrolling a person, a set of their biometric characteristics are collected and stored. These are then encoded as a biometric template. Criteria from this template can then be combined using an algorithm to generate a unique identifier for the person.

Aside from ISO/IEC 19794-2:2011 (which provides a standard for fingerprint templates) there are no mature interoperability standards for other biometric modalities.

Furthermore, different biometric providers often use different algorithms to generate a unique identifier from a biometric template. Data collected using different providers for the same modality may also not be interoperable.

Hence standardization on the same biometric modality, template and algorithm are important considerations to ensure interoperability. While technically possible, no examples were encountered of different systems using biometrics being interoperable.

Second, the accuracy rates of biometric modalities and systems vary. They include a risk of both false positives (a person is flagged as being a duplicate when in fact they are not) and false negatives (a person is flagged as not being a duplicate when in fact they are).

One respondent reported an example of this type of error in the Ukraine response. A mobile app used to self-register incorporated facial recognition to check for duplicates. However, the matching algorithm was not able to flag duplicates when people used different pictures of themselves or a copy of the same ID uploaded at different locations across the globe.

Technical capacity is a consistent challenge across the sector. The short term nature of appointments amplifies the lack of expertise in areas like data protection, data security, information technology and data governance.

Semantic challenges

Semantic definitions are important and currently need to be negotiated every time from scratch. This is particularly evident in contexts where Cyrillic, Arabic or other scripts are used, given that the global financial system primarily operates on Latin characters. If a name or ID is transcribed into Latin characters then the result can differ depending on the way it is done. As a result agreement on a consistent approach to transcription is needed.

Data protection regime

The purpose of data sharing for this use case is to improve assistance by identifying duplicates. The legal basis is usually consent, but in some cases vital or public interests are instead used.

There are a number of issues related to data protection to consider with this use case. These include:

- Segregation of duties and access rights (for example editing versus viewing rights) for data processors and (joint) data controllers. This impacts how frontline workers (aid workers, volunteers, implementing partners) can follow up on people's requests to access, change or rectify their data.
- Privacy by design and data subject rights: incorporating requests for sharing the purpose, partners (for example humanitarian organizations, third parties and government), and consenting to data sharing without appropriate processes in place that allow follow-up and responses to those requests (particularly in self-registration tools).
- Lack of granular access to data systems for field staff and/or partners that require insecure workarounds (for example sharing data via email or USB drive).
- How to apply the principles of data minimization and proportionality to biometric data given its sensitivity.
- The risk of function creep and the principles of proportionality, data minimization, purpose limitation and data quality.
- Operations usually lack minimum data requirements and tend to collect more data than needed.
- Purpose limitation and data retention: the purpose and legal basis are rarely updated (e.g. emergency versus preparedness purposes) and accordingly people are not informed or involved (e.g. changing/objecting or legal basis); it is more difficult to retain data when data flows are not tracked properly when shared.
- Data quality: lack of processes to keep data up to date and as accurate as required to meet data protection policies.

Cost impact

The cost impact can be understood both in terms of not attempting to avoid duplicate support and in attempting to do so.

Cost of duplicate support

The cost of duplicate support varies depending on the programme. Anecdotal data from respondents suggested that for programmes that use geographical targeting, have a relatively low total transfer value and run for only a few months, the cost would be low.

One example given of a programme that met these criteria is a pilot to use biometrics to check for duplicates. A very small number were identified, which would have results in duplicate payments of less than 1,000 US dollars. In this case the cost of introducing deduplication was not worth the benefit of avoiding duplicate payments.

Other respondents provided details of a large-scale, long-term programme that included self-enrolment options for participants. This programme has implemented deduplication measures and hence was able to estimate the cost of what duplicate payments would be if deduplication were not in place. This came to 945,000 euros.

Clearly the cost of duplicate payments are very significant in large-scale programmes.

Cost of deduplication

There are several factors that contribute to the cost of deduplication and risk of missing out on potential duplicates. These include the time and capacities needed to negotiate a data-sharing agreement, develop or customize data systems, and operationalize a deduplication process which is particularly challenging for smaller organizations with limited resources. Interviewees stated that UN data systems are cost and resource intense, and thus particularly used in large-scale emergencies. The most common time frame reported was 3 to 4 months. This could take longer if there is no agreement on the basis for functional identification. However, these figures relate to a deduplication process between two organizations. Presumably the time needed will increase further as more organizations are involved.

Once an agreement has been operationalized, most respondents reported that a monthly deduplication process then takes about one or two person days a month per organization.

Use of biometrics is also an additional cost driver. This cost can be significant for smaller programmes with limited budgets.

Overall the cost-benefit analysis strongly supports deduplication in larger programmes where there are multiple agencies providing assistance. In programmes with a shorter duration and lower total transfer values, the result may differ.

Potential new approaches to solve the use case

UNIRIS

UNIRIS generates a PGP pair from biometric measurements. However, unlike other solutions the private key is not stored. Instead it is regenerated each time the person touches the biometric device. If the key generated matches the public key then the person is verified.

No examples were found of UNIRIS being piloted in the humanitarian sector.

See www.uniris.io

What is a PGP key

A Pretty Good Privacy (PGP) key is a public encryption key. A PGP key can be used to sign and encrypt emails and files. When a PGP key is created, a keypair having a public key and a private key is generated.

It is an example of asymmetric cryptography, where something encrypted by one of the keys can only be decrypted by the other key of the pair. Usually, one of the keys goes public (which could be on a blockchain) while the other is kept secret.

Payment provider-led approaches

We are not aware of this approach being piloted yet. It focuses on deduplication of cash payments at the point of transfer. This might work if multiple implementers are using the same bank or mobile money provider. In theory the FSP could determine if the same recipient account or mobile wallet has already received a payment from another implementer. Hence, business rules could be defined to determine payment ceilings for recipients. Once reached further payments could be blocked.

This would in theory allow multiple implementers to maintain their own recipient databases, but handle deduplication of payments at this stage. This has the added benefit of potential cost savings through bulk procurement with FSPs.

This relies on a group of implementers agreeing to use and trust the same FSP, which may not be feasible or desirable. Solutions like Mojaloop may offer a way to broaden this, acting as a payment aggregator that includes multiple FSPs.

There are lessons to explore here from the Jordan Common Cash Facility as well as CARE's pilot using voice biometrics in Somalia . Further work would also be needed to consider which data protection regulations apply and on which basis they should collect people's consent.

Data stewardship approaches

The Collaborative Cash Delivery Network (CCD) is exploring data stewardship as a potential solution to some of the data governance aspects of interoperability in relation to deduplication and referrals. CCD defines data stewardship as "a model of data governance in which an intermediary facilitates or holds consent and decision-making on behalf of users, sometimes with a fiduciary responsibility under law".

Data stewardship provides an alternative to commercial models of data management, which may be exploitative and disempowering when applied in the humanitarian context. It may provide more agency to affected populations in relation to how their data is processed and used.

CCD is piloting this approach in South Sudan and with the Ukraine response with a focus on aspects related to accountability, accessibility and participation. Figure 1 shows CCD's 'stack approach' to data stewardship, which goes beyond the technical layers required for data portability to include all layers needed for implementation. These components could be adapted to respond to different use cases and contexts.

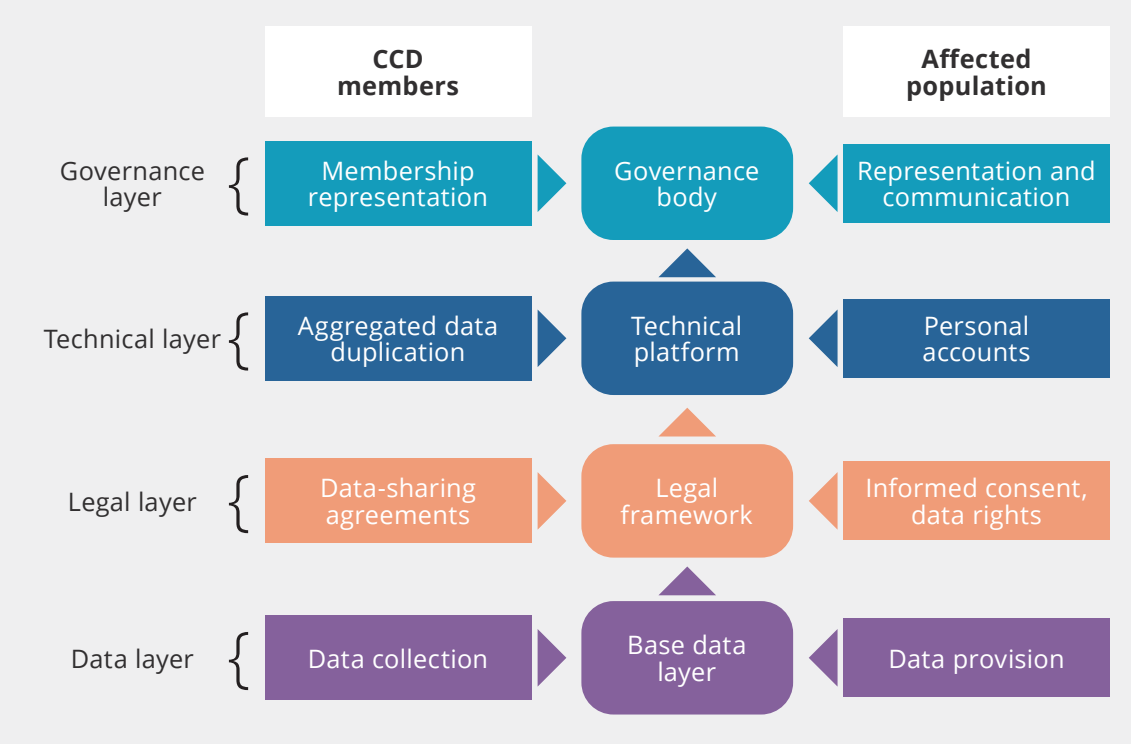


Figure 1. The CCD 'stack approach' to data stewardship

16 https://mojaloop.io
17 UNHCR and CALP. 2017. Review of the Common Cash Facility Approach in Jordan. www.calpnetwork.org/wp-content/uploads/2020/03/ccf-jordan-web-1.pdf
18 GSMA and CARE. 2020. Verifying recipients of cash assistance through Voice ID: Pilot project lessons and outcomes. www.gsma.com/mobilefordevelopment/wp-content/uploads/2021/08/M4H_VoiceID_Evaluation.pdf
19 CCD. 2022. Safe Passage: Options for Data Portability in the Humanitarian Sector. www.collaborativecash.org/safepassageoptionsforsector

THE FUNDAMENTAL PRINCIPLES
OF THE INTERNATIONAL RED CROSS AND RED CRESCENT MOVEMENT

Humanity

The International Red Cross and Red Crescent Movement, born of a desire to bring assistance without discrimination to the wounded on the battlefield, endeavours, in its international and national capacity, to prevent and alleviate human suffering wherever it may be found. Its purpose is to protect life and health and to ensure respect for the human being. It promotes mutual understanding, friendship, cooperation and lasting peace amongst all peoples.

Impartiality

It makes no discrimination as to nationality, race, religious beliefs, class or political opinions. It endeavours to relieve the suffering of individuals, being guided solely by their needs, and to give priority to the most urgent cases of distress.

Neutrality

In order to enjoy the confidence of all, the Movement may not take sides in hostilities or engage at any time in controversies of a political, racial, religious or ideological nature.

Independence

The Movement is independent. The National Societies, while auxiliaries in the humanitarian services of their governments and subject to the laws of their respective countries, must always maintain their autonomy so that they may be able at all times to act in accordance with the principles of the Movement.

Voluntary service

It is a voluntary relief movement not prompted in any manner by desire for gain.

Unity

There can be only one Red Cross or Red Crescent Society in any one country. It must be open to all. It must carry on its humanitarian work throughout its territory.

Universality

The International Red Cross and Red Crescent Movement, in which all societies have equal status and share equal responsibilities and duties in helping each other, is worldwide.

The International Federation of Red Cross and Red Crescent Societies (IFRC) is the world's largest humanitarian network, with 192 National Red Cross and Red Crescent Societies and around 14 million volunteers. Our volunteers are present in communities before, during and after a crisis or disaster. We work in the most hard to reach and complex settings in the world, saving lives and promoting human dignity. We support communities to become stronger and more resilient places where people can live safe and healthy lives, and have opportunities to thrive.

Follow us:

www.ifrc.org | twitter.com/ifrc | facebook.com/ifrc | instagram.com/ifrc | youtube.com/user/ifrc | tiktok.com/@ifrc