CCD

Collaborative Cash Delivery

DESIGN PROCESS REPORT

CCD Data Governance Pilot South Sudan

A DESIGN REPORT FOR THE CCD NETWORK
Prepared by Paul Currion

COLLABORATIVE CASH DELIVERY (CCD) IS A NETWORK OF 14 OF THE LARGEST INTERNATIONAL NGOS WHO COLLECTIVELY DELIVER OVER $1BN IN LAST MILE CASH AND VOUCHER ASSISTANCE EVERY YEAR.

www.collaborativecash.org

AUTHOR: Paul Currion for the Collaborative Cash Delivery (CCD) Network.

ACKNOWLEDGEMENTS

European Union

CONTENTS

INTRODUCTION

Disclaimer

This is a design document and not a research document. The analysis included is not exhaustive, as it is only intended to provide context for the design decisions. Any factual errors in the note reflect the limitations of the design process, rather than the contributions of the participants in that process. The recommendations that it makes are provisional and do not represent commitments by the CCD or any of its members. The solution proposed in the design note is intended as a starting point for the CCD to test new models of governance in the course of addressing a specific problem. In this document we discuss the problem of deduplication in South Sudan.

THE SOLUTION PROPOSED IN THE DESIGN NOTE IS INTENDED AS A STARTING POINT FOR THE CCD TO TEST NEW MODELS OF GOVERNANCE IN THE COURSE OF ADDRESSING A SPECIFIC PROBLEM.

Design Process

Following research carried out by CCD in 2022, a consultant was engaged to develop the design process. The process began with separate webinars delivered to CCD members in South Sudan and Ukraine, followed by Q&A sessions. Following the webinars, a series of design sessions were held remotely with each country, engaging with a group of staff proposed by CCD members. At the same time key informant interviews were carried out to investigate specific issues related to the operating environment, particularly legal and technical. This pilot design note was synthesised from all this information.

We approach the question of data portability by taking a clearly defined concrete challenge faced by aid organisations (deduplication in South Sudan and referrals in Ukraine), and addressing it through an approach which improves accessibility and participation (data stewardship). Once this approach has been proven to work in pilot, it should form a solid foundation for data portability, since there will be:

- ▷ A) the technical basis to move data between organisations,
- ▷ B) participatory mechanisms to ensure that the aid recipients' interests are taken into account, and
- ▷ C) greater understanding among aid recipients (and organisations) about their interests regarding how their data is used.

Definitions

In order to ensure clarity of the different terms used in the design process, we have developed a separate Glossary (Appendix A) which sets out definitions of key terms and sources for those definitions. Here we expand only on the critical terms of Duplication and Deduplication.

Duplication

Duplication can happen for a number of reasons, but is inevitable because A) households are not static entities whose members act in concert, and B) different types of assistance are delivered in different ways and in different timeframes. There are two types of duplication:

1. Duplication of **registration** occurs when two or more organisations hold records on the same individual or household, but the details in those records do not match and so cannot be easily compared. This can occur for different reasons: the individual or household has changed location, and registered with new service providers in that location; or the household is registered multiple times in the same location with different details, such as when different NGOs list different individuals as head of household depending on what the focus of their assistance is.

2. Duplication of **assistance** occurs when different types or tranches of assistance are delivered to a single household at different times, but the records of those deliveries are held by separate organisations (or by separate units within an organisation) and so cannot be easily compared.

It was noticeable that, while participants in the design process could give concrete examples of duplication, there was no shared explicit definition of "duplication". It is important for CCD members to agree such a description in order to ensure that future discussions and developments are based on a shared understanding. Such a description could be:

> Duplication occurs when a Household (HH) appears in more than one database but with different details (e.g. different head of household, different residents, etc) or different schedules (e.g. delivery dates for payments or in-kind assistance, visits from specialist teams, etc). This covers both duplication of registration and assistance.

Deduplication

Deduplication can be described as an algorithm intended to resolve the duplication described above, using the following steps:

1. Compare two or more datasets which may overlap,

2. Establish if and where duplication of records occur, and

3. Deduplicate either by deleting or merging records.

4. Notify the two parties subject to the duplication.

Deduplication is instrumental – it is intended to maximise efficiency of limited resources. This means avoiding duplication of assistance, but also identifying gaps in delivery. There is limited evidence that duplication of assistance is a significant problem in humanitarian aid. From the perspective of aid recipients, gaps in delivery are the greater problem – either inappropriate assistance, delayed assistance, or assistance not arriving at all.

The CCD has already outlined a Deduplication Process (Appendix 2), including a list of mandatory data fields, which can be built on in this pilot project. However some CCD meeting members in South Sudan are not members of CCD globally, and a short scanning exercise will need to be undertaken to ensure both policy compliance and organisational capacity to implement it, regardless of the nature of the pilot.

BACKGROUND

Humanitarian operations in South Sudan are long-running and well-established, but a number of coordination problems remain. During the research phase key CCD member organisations reported a problem with deduplication of data. This problem has been relatively small-scale due to the low density of both population and service providers, but is still an ongoing concern – and potentially a greater concern in the event of a crisis which leads to more population displacement, either internal or external.

Currently CCD members working in South Sudan are not able to deduplicate in a systematic way, nor are larger organisations (specifically UN agencies and government ministries) able to provide this service. While deduplication has been addressed by such large organisations in other country locations (including Ukraine, the other crisis covered by this pilot) there is still a general consensus that this problem has not been solved, and the solutions on offer often create new problems for service providers such as CCD members.

SOUTH SUDAN OFFERS AN OPPORTUNITY TO BUILD ON EXISTING MECHANISMS AT THE COMMUNITY LEVEL TO IMPROVE ACCOUNTABILITY FOR AID RECIPIENTS.

South Sudan is a difficult operating environment in which to introduce new technology initiatives due to the relatively poor technical infrastructure and relatively low levels of digital literacy. However aid organisations need to plan for future developments, including increased penetration of technology (particularly mobile networks) and increased capacity of the national government. In particular South Sudan offers an opportunity to build on existing mechanisms at the community level to improve accountability for aid recipients.

Stakeholders

This section provides an overview of stakeholder groups. Recommendations for stakeholder engagement are provided in the Layer Design section.

Communities

The focus of our design is the group of primary stakeholders in any information system – those who provide the data which makes the system useful, in this case the individuals, households and communities of South Sudan. The pilot should attempt to incorporate their perspective at all points during the project cycle. We are currently commissioning research to better understand community perceptions of how their data is used.
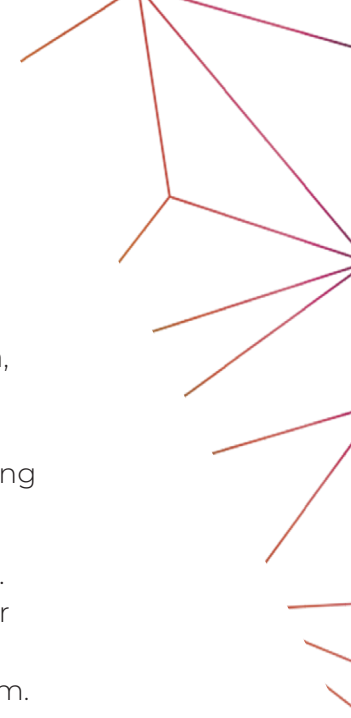
International and National NGOs

It might be appropriate to categorise these two types of NGO separately on the basis of the differing levels of power, capability and reach that they possess. However for the purpose of the pilot design we include them together since the CCD membership in South Sudan includes both international and national NGOs.

United Nations Agencies

While a range of UN agencies are operating in South Sudan, the most relevant to the pilot design is the World Food Programme (WFP). WFP's SCOPE platform holds more, and more granular, data than any other stakeholder (including government institutions), with coverage of most of the country down to household level.

This creates a number of issues for CCD members. While the data in scope was largely collected by NGOs working under contract to WFP, NGO access to SCOPE data is limited. In order to have access NGOs must sign an MoU with WFP, but under such an agreement SCOPE data is only made available on request, and participants reported that responses to these requests were not timely. Some NGOs have refused to sign the MoU on grounds that the terms are unacceptable, specifically:

→ Data ownership lies with WFP, so that NGOs have no access or related rights.

→ Data retention periods in SCOPE are longer than the limits observed by NGOs.

→ Biometric data is included in SCOPE which raises ethical concerns for some NGOs.

In some locations SCOPE is the only source for household-level data, while in other locations data might be held by SCOPE plus multiple NGOs, with identical or overlapping households between datasets. However there is no way for NGOs to deduplicate these datasets using SCOPE because of the barrier to access.

Other UN agencies were also mentioned during the design process. CCD members share data with other UN agencies (FAO in particular was mentioned), and the Cash Working Group in collaboration with IOM is planning to build a humanitarian cash programming platform. However these relationships do not currently appear relevant to the pilot design.

The argument is sometimes made that instead of trying to address issues such as deduplication collectively, a centralised platform should simply take all responsibility. There are both political and technical reasons why this is not necessarily a feasible or desirable solution, with a recent study suggesting that we should move away from "a competition between the best agency systems to be adopted widely in the sector, towards the acceptance of a plurality of systems that currently form the cash IM architecture".[1]

Government institutions

CCD members have working relations with a range of government departments depending on their programme areas. Examples given during the design sessions included: the Ministry of Agriculture and Food Security (food security data); Ministry of Health (health management information systems); and Ministry of Water Resources and Irrigation (water level data, related to climate). Data sharing with these institutions does not appear to be systematic, but in response to specific operational requirements on either side.

Private Sector

All participating organisations were using third-party software and hardware platforms provided by private companies. These platforms provide a variety of functions related to both registration and delivery; these functions overlap but the capabilities of each platform are not identical, and the platforms themselves are not interoperable. This is one reason why the main format used to exchange data between organisations remains MS Excel spreadsheets. It is not likely that these companies will collaborate to create the type of interoperability needed for deduplication, and their corporate structures do not lend themselves to a stewardship approach.

---

1 *Registration, Targeting and Deduplication: Emergency Response inside Ukraine (2022)*, CALP Network.

Constraints

There are three major external constraints on data governance projects in South Sudan, all three of which present significant challenges to increasing the involvement of aid recipients in the governance of their data:

→ **An absence of regulatory framework and legal institutions** to enforce policies or provide recourse relating to data protection or other governance issues.

→ **Significant limitations of technical infrastructure**, in terms of both power and communications, which make some approaches unfeasible.

→ **Low levels of digital access and digital literacy**, especially in rural areas, so that certain solutions which are common in other countries are not viable at present.

These constraints overlap considerably – for example, South Sudan has a mobile penetration rate of less than 30%, which is a combination of all three of these constraints. Any innovation in data governance will need to work with and around these constraints.

Process

The Design Session participants were asked to describe how their organisation collects data from aid recipients, and how they manage that data subsequently. Diagram 1 below is a rough sketch of the processes described.

Diagram 1: Sketch of Data Management in South Sudan
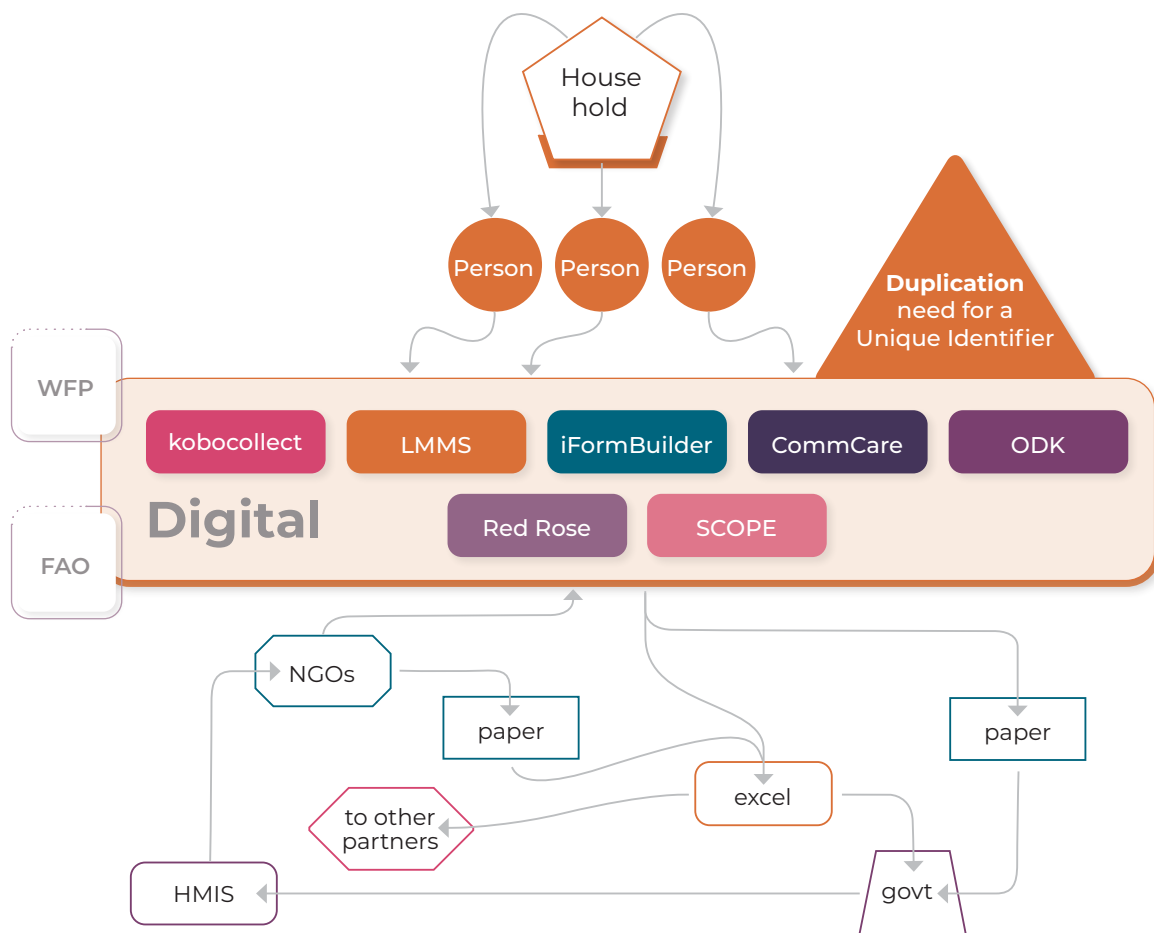
Some salient points emerge from this finding. First is that every participant organisation follows roughly the same steps in their data management, with the main difference between them being which platform(s) they use for different parts of the management cycle (a list of these platforms is given in Table 1, below). Second is that most data exchange takes place in a rudimentary fashion using Microsoft Excel sent via corporate email, and in some cases on paper; this is historically very common in the humanitarian sector. Third is that there is clearly an opportunity for a collective approach to data management as all data passes through a "platform layer" where the main obstacle is interoperability; one can see the similarity between this and the data stewardship model sketched in the design section below.

Table 1: Examples of Aid Platforms in Use

System	Owner	Users	Data Location
SCOPE	WFP	All (except IRC?)	In-country (some locations) Rome? Geneva?
KoboCollect	Kobo	ACF, CMD, WVI	US/Ireland? Vs self-hosted
LMMS/Horizon	World Vision International	WVI	South Sudan?
CommCare	Dimagi	SCI, Plan, IRC	US/Ireland (hosted) Vs self-hosted
iFormBuilder	Zerion	Concern	
KoboCollect	Kobo	Concern, CMD, WVI	
RedRose	RedRose	ACF	
Unnamed cash platform	*IOM (planned)*		

DESIGN PRINCIPLES

The design sessions were partly organised around nine design principles relevant to data governance, grouped into three categories of three principles each in Table 2 below. We have used these principles to structure the design of the pilot, and full definitions of the principles are given for reference in Appendix 1.

Table 2: Key Design Principles for Data Governance

Data Protection	Democratisation	Localisation
Security	Accountability	Localisation
Privacy	Accessibility	Sovereignty
Transparency	Participation	Sustainability

These principles are not binary. They all exist on a spectrum – we can be more or less accountable, for example. There may be tensions between different principles – the most secure system is the least accessible system, for example – and the process of design is partly about managing these tensions.

Data Protection

The design sessions did not spend much time on the topics of security, privacy and transparency of data. The assumption of all participants was that the security and privacy of any deduplication mechanism would be determined by the existing data protection policies of their organisations, most of which aim (at minimum) to apply the provisions of the EU GDPR. Although this is not always possible it remains the benchmark for any pilot project, particularly since South Sudan presently has no data protection regulations in place.

Transparency was discussed more, since participants recognised that at present their systems are not transparent. While CCD members do work on the basis of informed consent, the information in the consent process is limited in scope, and there is little to no transparency for aid recipients once their data has been collected. There is scope to provide this type of transparency via existing community-level structures (such as AAP committees) which are already used for communication and consultation with communities – these are discussed below in the section on Democratisation principles – and through enhancing existing consent procedures to include more detail about data rights.

Democratisation

There are legitimate questions about the extent to which aid recipients in South Sudan are interested in or able to be more involved in how their data is used. We are in the process of organising survey work with communities to assess this.

Most organisations will need to update their policies to enable some of the steps suggested below, since participants reported that most of those steps were not covered by existing policies. The pilot should guide this on a collective basis.

Accessibility

While the remaining principles in this list are to some extent voluntary, accessibility is not; any CCD member receiving EU funding is technically legally required to make data accessible in their projects. It is however difficult to realise this in practice, particularly in a working environment such as South Sudan – but it is also true that most (if not all) CCD agencies do not have accessibility policies, which means they do not develop the capacity.

Article 15 of the GDPR defines the "Right of access by the data subject". This includes the right to know whether their personal data is being processed, applying only to data that the aid recipient has given to the aid organisation, but not to data related to the aid recipient which the aid organisation has collected themselves. It also includes granting access to that data, but the article does not define what constitutes "access", which leave some flexibility in how we can improve accessibility for aid recipients, which might include them being able:

→ 1   to view their data,

→ 2   to have a copy of their data

→ 3   to correct their data

→ 4   to erase their data

Like the Data Protection principles discussed above, the principle of Accessibility has a strong technical component, and is closely linked with the idea of data portability. We identify two extant models for creating accessibility and portability for APs on any platform, including the deduplication platform: Wallets and Accounts, summarised in Table 3.

Table 3: Two Models of Accessibility

	Wallet Model	Account Model
Who's responsible for the data?	The aid recipient (AR)	The data steward (DS)
Where is the data stored?	On device held by AR	On platform managed by DS
How is the data shared?	AR allows NGO access to their wallet	DS allows NGO access to AR account

A deduplication platform provided by the Data Steward would hold household data in individual "Accounts", but would carry out deduplication on an aggregated and anonymised dataset. Individual NGOs would still hold their own datasets for operational purposes, but APs would be able to access the platform in order to know which NGOs are holding data on them and for what purposes.

There is an option to take this further, giving the AR permission control over their Account. Permission would initially be set to share with NGO partners based on the consent given when the data was collected. However the AR would be able to access their Account, and to change those permissions if they wished. The Data Steward platform would then notify the NGO who had contributed that data of the AR's wishes.

One concern raised during the design process was the potential for increased accessibility to increase the potential for conflict to occur, specifically if (for any reason) account holders share their login with other household members, or other households. In such a scenario there could be friction over the amount of aid that individuals or households are receiving compared to each other. Mechanisms to avoid this should be developed during the Pilot.

Due to the constraints of digital literacy and access mentioned above, we recognise that the scope for accessibility is limited in South Sudan. A small-scale pilot project cannot address those limitations, but there is still value in proof of concept. Access requests and their responses would therefore need to happen offline rather than on-platform.

Regardless of the final design of the pilot, accessibility implies that we promote awareness not just of the process or system being used – including ongoing information about where community data is being stored and how it is being used – but also of their basic rights around accessibility, including which of the four options listed above are available.

Accountability

Accountability means that CCD members can be held to account by aid recipients. This is a matter of power, and it is widely accepted that aid recipients do not have much if any power at present. Accountability should therefore be seen as a process in which our organisations can move towards more or less accountability.

An accountability mechanism already exists in South Sudan, in the form of Accountability to Affected Populations (AAP) Committees. These are used by NGOs to engage with communities at the village level, with information sharing in both directions at different stages in the project cycle, as well as acting as a channel for complaints. Broadly speaking AAP committees lead on identification and selection of aid recipients, while service providers conduct registration and verification following the selection.

While existing accountability initiatives have significant limitations,[2] the AAPs offer a model – and potentially a vehicle – for the pilot's accountability pitch. However there are a limited number of ways in which aid organisations can be held accountable for failing to meet their responsibilities: causing loss of funding or loss of reputation for the organisation; or holding an individual within the organisation to account under that organisation's policies. These options are usually not available to aid recipients.

In the absence of such options, aid organisations themselves must generate meaningful accountability. The best way to do this is collectively through peer accountability mechanisms, with the governance layer including ongoing monitoring of participating agencies' activities with regards to data. While such peer mechanisms are immature within the aid sector, such a mechanism could be built into the governance structure of a data steward.

Feedback by itself is usually insufficient to ensure genuine accountability, and CCD members may wish to consider stronger accountability responses that are conceived during the pilot period. The data steward, as well as taking responsibility for the data under its stewardship, should also hold its member organisations to account for any breaches of their responsibilities. However this might prove to be an obstacle that prevents organisations from joining the project, and it may be better to delay this until after an initial pilot succeeds.

---

2 REACH, Community Perceptions in South Sudan: Findings and Recommendations, June 2022
    https://reliefweb.int/report/south-sudan/community-perceptions-south-sudan-findings-and-
    recommendations-june-2022

Participation

Participation is an important element in many data stewardship mechanisms globally, and would be a huge step forward in the humanitarian sector. Participation takes the involvement of aid recipients further to give them a voice in the decision-making around aid projects. In the case of this pilot, this would mean representation (either elected or appointed) in the governance structure of the data steward. These representatives could (for example) be appointed from existing community structures that CCD members already work with, or be proposed by Community-Based Organisations that work at the local level.

PARTICIPATION TAKES THE INVOLVEMENT OF AID RECIPIENTS FURTHER TO GIVE THEM A VOICE IN THE DECISION-MAKING AROUND AID PROJECTS

WFP has a similar mechanism in the form of Project Management Committees, and Water User Committees formed at village level which are given responsibility for the maintenance of improved water facilities. In South Sudan there are also traditional legal structures in at least some regions which act in parallel to the national court system. Another model could be the Task Force model to combat Sexual And Gender Based Violence, in which a lead agency appoints representatives from communities. Such Task Forces are a focal point for sensitisation around specific issues, and can develop reporting structures which communities can use to raise issues (via a local leader, hotline or suggestion box).

Whether these models are a good fit for data governance is a question which only CCD members who already work with them can answer – for example, data governance is different from project governance, and in particular is not a village-level concern – however it seems likely that elements from them could be adapted.

Localisation

Localisation

Although some of the requirements for successful localisation may be met if the principle of participation is successfully addressed, the two principles should be treated separately. It is unlikely that significant steps towards localisation will be taken during the pilot itself.

Sovereignty

Sovereignty in this case refers to where the data that is collected is primarily stored, i.e. where the data servers are located, both in terms of geographic location and legal jurisdiction. Most aid organisations now store data in cloud-based servers rather than in their country of operation. Participants did not feel that this was a significant issue, and did not believe that aid recipients would recognise it as an issue; however it is also true that aid recipients are not informed where their data will be stored.

We did hear that the Government of South Sudan may recognise this as an issue, shown in the case of a CCD member organisation collecting data related to water levels. The data was originally planned to be transmitted to and stored on a server in Nairobi managed by the organisation, but the Ministry of Water Resources and Irrigation expressed a preference that the data remain in country; it is now held on a server in the Ministry offices, which is shared by multiple organisations. Whether this is an official policy or not is unclear.

Sustainability

Due to lack of time the design sessions did not address sustainability. In the pilot context the main sustainability question is how it will be funded following the pilot phase, if it is successful and the CCD members wish to continue. Since data governance mechanisms cannot be self-sustaining financially in the absence of a commercial model, the project will require donor support until such time as e.g. deduplication is no longer the concern of aid organisations, and responsibility can be handed over to another institution. Whether the mechanism is sustainable will therefore depend on whether it is eventually supported by an institutional actor – preferably the government, possibly a UN agency.

THE DATA STEWARD

Role

Multilateral deduplication requires a single register against which all parties' data can be compared, and which is agreed to be the basis for deduplication. The primary model in the humanitarian sector is that of a central authority to which all other stakeholders are required to defer. While this model may be able to resolve some amount of duplication - it is not clear how much, since their "successes" are usually self-reported - the model itself is flawed in a number of ways, most obviously in terms of creating a monopoly with no accountability.

Diagram 2 below shows a simplified model of how this has developed in South Sudan. NGOs collect data from individuals (or individual households), some of which is then shared with WFP on the SCOPE platform, and some of which is also shared bilaterally with each other. The model lacks transparency, accessibility and accountability, and is broadly unsatisfactory to most if not all stakeholders.

Diagram 2: The "Command and Control" model
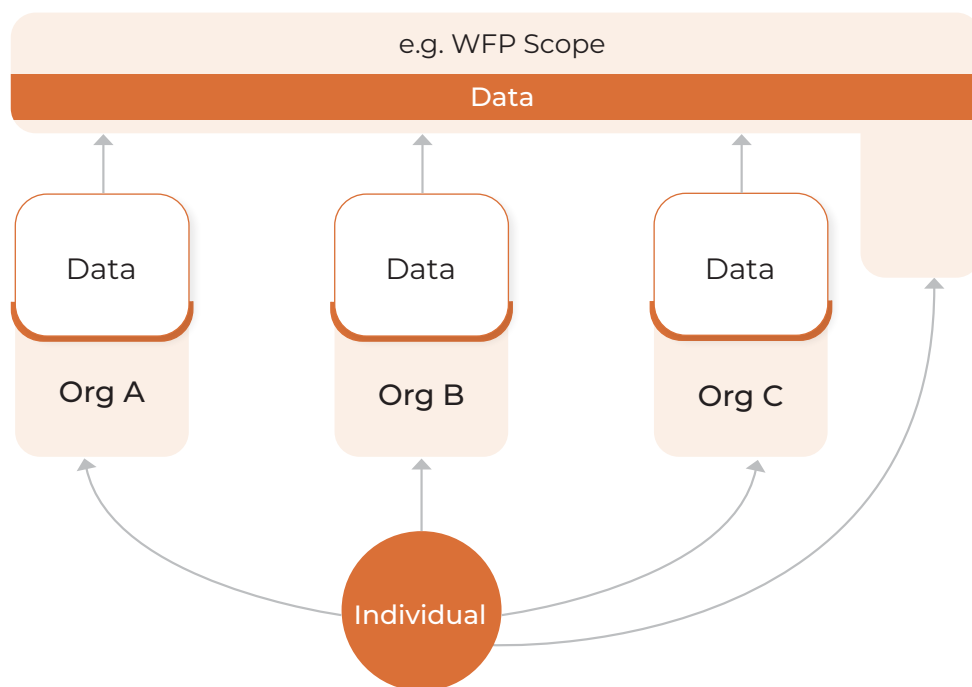
We propose an alternative model based on CCD members' existing collaboration, in which actors voluntarily pool their data and develop a collective mechanism for deduplication.

Diagram 3 below shows – again in the broadest possible terms – what this would look like in comparison to the existing structure.

**Diagram 3: the Data Stewardship model**
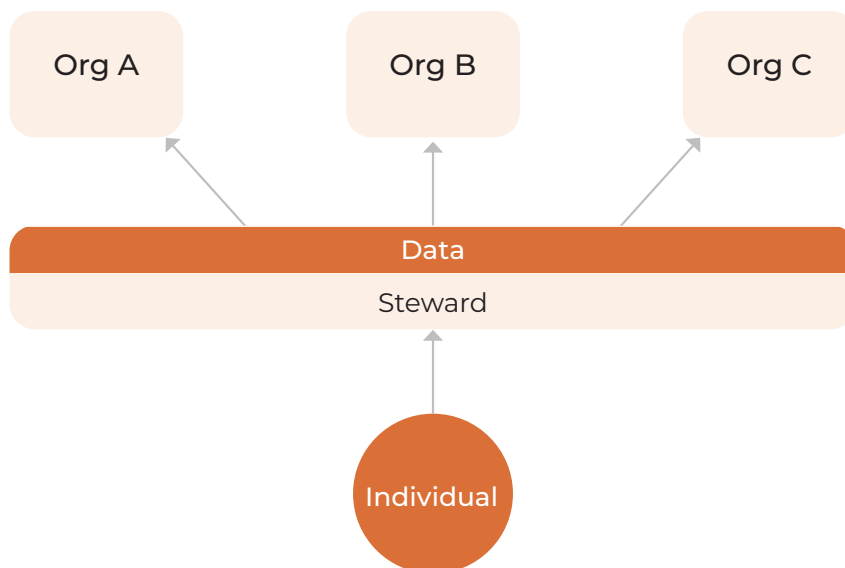
The Steward is mandated in two ways:

1 To represent the interests of aid recipients. As well as incorporating the democratisation principles described above, the steward can also provide a platform for data portability, since it holds a collective register which can respond to portability requests, either from a participating organisation or an aid recipient.

2 To provide a deduplication service to aid organisations. The burden of responsibility for deduplication is taken off participating organisations, becoming a shared responsibility with no single agency holding a monopoly and accountability generated through peer mechanisms.

This model should be able to deliver at least the same level of deduplication as the centralised model as long as the underlying platform for deduplication is effective. It also forms an equally solid basis for data portability, since portability queries can be directed to a single entity which in principle can track where the data it holds has been sent and received; and the data itself provided either to the aid recipient themselves or to the service provider of their choice (as long as that service provider is participating). The design question is what sort of intermediary is required to perform this function, and how it should be set up.

Form

We have developed a Stack Model as the basis for design (Diagram 4 below), where the stack consists of four Layers: Data, Legal, Technical and Governance. All four Layers must be addressed during design and implementation, particularly to ensure that the different stakeholder groups are able to engage with each layer.

Diagram 4: Stack Model of the Deduplication Pilot
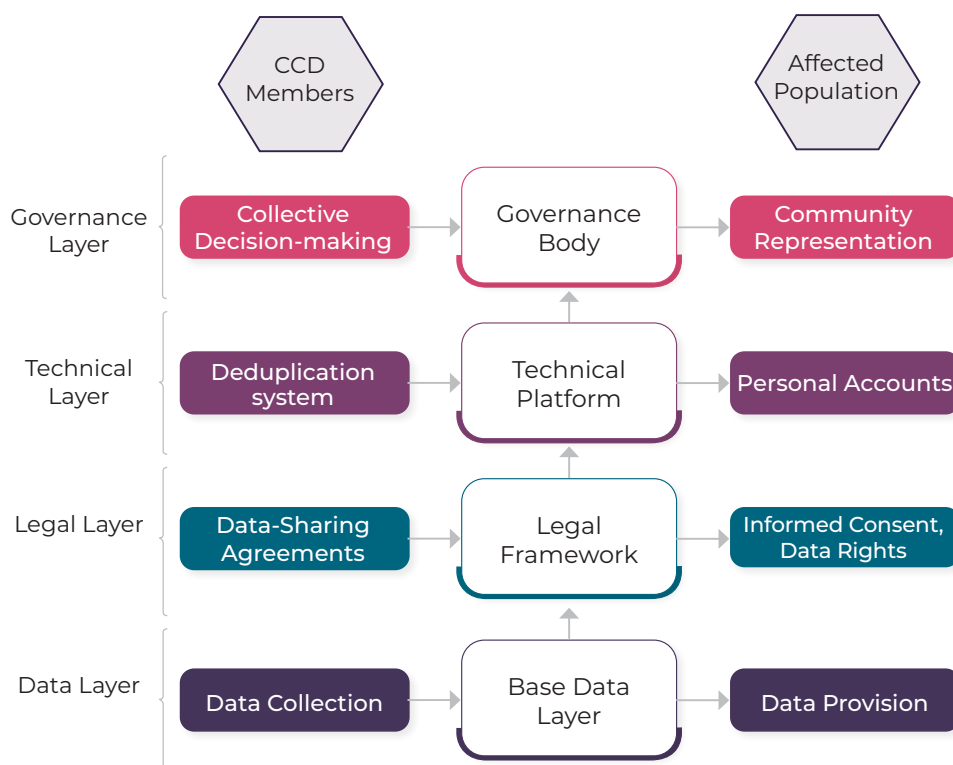
The key design question is what sort of intermediary is required to perform this function, and how it should be set up. Rather than creating a new entity, we propose a "virtual" data steward (as opposed to a separate legal entity) which builds on existing structures to deliver additional value. The structure should be federated rather than centralised, i.e. any lead agency which hosts the data steward mechanism does not act as an authority but as a steward on behalf of the participating organisations – and on behalf of the aid recipients.

It is worth noting that the Technical Layer is just one part of the design, and not usually the most important part. The Technical Layer does not necessarily require that everybody must use the same platform to achieve the interoperability necessary for a design to function; while it may involve developing a new platform, organisations could continue to use their existing platforms while the exchange of data is handled through e.g. API calls to create an aggregated dataset for deduplication.

Layer Design

Stakeholder Recommendations

Stakeholder Group	Recommendation
Local communities	The pilot must engage local communities, but the best approach is unclear. We are currently commissioning a survey to better understand community perceptions of how their data is used, and will continue to discuss with CCD members.
International and national NGOs	The pilot should include both national and international NGO for participation and governance.
UN agencies	The pilot should not include UN agencies as part of the governance structure, but it may be worth ensuring that WFP are aware of the pilot in order to manage expectations.
Governance agencies	The pilot should not include government institutions in the pilot phase, but they should be considered important stakeholders in any future data governance mechanism.
Private companies	The pilot should not include private companies unless there is a specific technical requirement which cannot be met by CCD members themselves.

Layer	Brief Description	Notes
Data	The existing data that is already collected by participating organisations.	There should be no additional data collection requirement for the Pilot. Analysis will need to be carried out regarding what data is specifically needed for deduplication, and how that data can be shared systematically by participating agencies. If possible, the pilot should ensure that the data remains hosted in country.
Legal	The agreed policies and processes required to deduplicate pooled data.	Participating organisations should apply their existing data protection policies to the pilot project, including application of GPDR. The layer should build on existing data sharing agreements between CCD members. There will also need to be an additional agreement that assures members that the deduplication process itself meets their data protection requirements.
Technical	The hardware and software required to carry out the deduplication function, as well as any protocols necessary for data sharing.	The layer will implement the existing CCD Deduplication Process. A narrative description exists, but it will need to be updated after consultation with CCD members, including which data fields will be used as the basis for de-duplication.

There are outstanding technical requirements which the design process was not able to answer. The primary requirement is an encrypted data processing platform to handle deduplication. The precise technical requirements of the platform will be developed in the next phase of pilot planning, most likely using existing CCD member capabilities. |

Layer	Brief Description	Notes
Technical (continued)		This platform should complement rather than replace the existing systems used by humanitarian actors, carrying out only the single function of multilateral deduplication which individual organisations cannot carry out themselves. There may therefore be an interoperability requirement in order to connect the Stack to organisational internal platforms. This should be addressed in the wider context of the project, engaging with DIGID to establish interoperability solutions.
Governance	The agreements and structures that determine how the stakeholders relate to each other and make decisions collectively – the stewardship mechanism itself.	This layer is what distinguishes the pilot from other initiatives. We do not recommend setting up a new organisation separate to CCD members in South Sudan, since the lack of regulatory framework would not give a separate organisation a stronger legal basis for the responsibilities of a data steward. The data steward will therefore be "virtual", based on a joint and democratic decision-making process regarding use of pooled data.

Layer	Brief Description	Notes
Governance (continued)		Although the steward is "virtual", one of the participating organisations will need to take responsibility for hosting any platform in a lead agency model, with participating agencies making an institutional commitment to pool relevant data. However this structure should be federated rather than centralised, i.e. the lead agency does not act as an authority but as a steward on behalf of the participating organisations – and of course on behalf of the aid recipients.
		The "virtual" data steward will also have terms of reference which require it to establish channels to communicate with communities about why deduplication is a problem, how they intend to address it, and communities' rights. If possible this layer should also incorporate participation in the decision-making process by representatives from those communities, either appointed by CCD members who work with them, or represented by one or more Community-Based Organisation.
		The governance documents will be developed in the next phase of pilot planning. During this period the participants should seek to identify a pathway towards hosting by a local organisation rather than an international organisation, which will require seeking multi-year funding from additional sources.
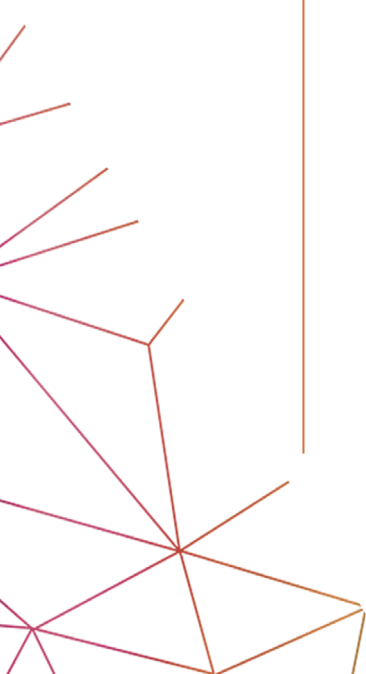
APPENDIX 1:

Design Principle Definitions

The 9 Data Governance Design Principles

Principle	Definition	Source
Data Protection group		
Security	Data security is the practice of protecting digital information from unauthorized access, corruption, or theft throughout its entire lifecycle.	https://www.ibm.com/topics/data-security
Privacy	Data privacy... is an area of data protection that concerns the proper handling of sensitive data... to meet regulatory requirements as well as protecting the confidentiality and immutability of the data.	https://www.snia.org/education/what-is-data-privacy
Transparency	Data transparency provides customers with an inside look into how their data is collected and used... customers should know why you need their data, how their information is gathered, where it is stored, and how it is protected.	https://blog.hubspot.com
Democratisation group		
Accountability	Accountability to affected people is a commitment by humanitarians to use power responsibly: to take account of, give account to, and be held to account by the people we seek to assist.	https://www.unocha.org/themes/accountability-affected-people

Principle	Definition	Source
Accessibility	The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and [associated] information.	https://gdpr-info.eu/art-15-gdpr/
Participation	effective "participation" of people affected by humanitarian crises puts the needs and interests of those people at the core of humanitarian decision making, by actively engaging them throughout decision-making processes.	https://interagencystandingcommittee.org
Localisation group		
Localisation	the process of having a humanitarian response owned in part or whole by a national and/or local constituency well-versed in the needs and socio-cultural context of the area in crisis.	https://resourcecenter.undac.org
Sovereignty	the understanding that data which are stored outside of an organization's host country and still subject to the laws in the country where the data are stored.	https://purdue.edu/critical-data-studies
Sustainability	the ability of an organization to continue its mission or program far into the future. All projects have to end eventually, but the project impact should continue.	https://proposalsforngos.com/what-is-project-sustainability/

APPENDIX 2:

CCD Deduplication Process – A Narrative Description

Overall Purpose

The Collaborative Cash Delivery (CCD) Platform is a coalition of humanitarian organisations aiming to harmonise cash delivery for increased scale, efficiency, effectiveness and collective impact. The CCD benefits from the members' wide range of practical experience and expertise, while all share a common objective of increasing the use of cash in meeting multiple needs of crisis-affected populations globally. Our collective cash programming will be adapted to the local context and ensure a community-focused approach. The CCD acknowledges that the unique strengths and strategic priorities of each organisation are mutually beneficial for cooperation in humanitarian actions.

Members of the CCD will share Personal Information (outlined in Annex A) at a country consortium level of programming participants of CCD jointly implemented projects the purposes of:

→ generating a unique ID for recipients across the CCD agencies at a country project level, **not** globally

→ avoiding duplicative registrations and/or enrolment in cash programming;

Overall Assumptions:

For the avoidance of doubt, the following purpose statements have been agreed:

→ We seek to share the **minimum** data required to achieve the purpose. Therefore, we will **not** share a *collective* transaction history for each beneficiary, but will only share the data outlined in Schedule B

→ We are deduplicating **individuals** not households

→ Biometrics should **not** be a requirement for collection or sharing

→ This agreement is technology **agnostic**, organisations can use whichever system they choose as long as it meets the security, privacy and protection requirements

→ Digital data collection and registration **is** a CCD requirement

The Purpose of the Global Agreement

The purpose of the global framework agreement is for the CCD organisations to agree on the process for deduplication and the roles of each actor. It is to set out a legal agreement which enables a country agreement to be quickly signed off after any minor country specific adjustments are made. The global framework agreement is to be signed by representatives of each organisation (or family of organisations) that comprise CCD.

The Purpose of the Country Level Agreement Annex

A template for a country level agreement will be included as an annex in the global framework agreement. It is to be signed by the appropriate legal entity representing each CCD member agency who are operating in the specific country.

The country level agreement shall specify:

1. Any country specific data protection, privacy, security requirements not covered by the global framework agreement requirements

2. Agreement between country level CCD actors on if there is an appropriate government issued ID for staff to be collecting as part of the mandatory data

3. Agreement on which CCD member is the Lead agency responsible for establishing, hosting, and maintaining the CCD community database

4. Standardised list of place names with 3 levels of geo-hierarchy or administrative divisions in the country (i.e. country, province/state, district/municipal, etc.)

The Deduplication Process

1. Each organisation digitally registers their specific recipients for the CCD project collecting the mandatory data outlined in Annex A. Each organisation also collects additional data according to their own organisations policies and guidelines.

2. Each organisation ensures there are not duplicate records in their database of CCD project recipients

3. Before sharing the new record with the CCD community database (hosted by the lead agency), each organisation queries the community database asking 'does a duplicate of new beneficiary record with mandatory data XYZ exist?'

4 CCD Community database returns an answer either NO or POSSIBLY

  a. If NO, the new record is added to CCD Community database

  b. If POSSIBLY, the CCD community database's response includes the name(s) of the orgs that have registered a beneficiary that looks like a duplicate

5 Verification and resolution of duplicates would happen by a meeting (in person or by telephone) among organisations with potential duplicate.

6 After verification and resolution is finalised,

  a. If the record is a duplicate, one organisation removes record from CCD project file

   i. Add comment field explaining deletion...

  b. If the record is not a duplicate, both organisations keep their records and are included in the CCD community database

ANNEX A: Definitions of Mandatory Data Fields

September 2020

Mandatory Data Fields for Collection and Sharing

1. **first_name**
   a. String of characters

2. **family_name**
   a. String of characters

3. **gender**
   a. Male/Female/Other
   b. Captured as multiple choice

4. **date_of_birth**
   a. Stored in the International date format YYYY-MM-DD; however can be captured as a number (i.e. 34)
   b. (if unknown, 1 Jan 1900 for adult, 1 Jan 2010 for child)
   c. Numerical string

5. **place_of_birth**
   a. If in country of operations, captured by choosing from standardised drop-down list as determined by country consortium
   b. If born outside of country of operations, chosen from drop-down list of countries

6. **community_id**
   a. Current location(community) person(s) resides. These can be permanent settlements like villages and hamlets or temporary housing establishments like refugee camps and temporary shelters for survivors of disasters
   b. Captured by choosing from standardised drop-down list as determined by country consortium

7. **mobile_phone_id**
   a. Mobile Phone number of person
   b. Captured as a numerical string with no country code

8. **hh_size**
   a. number of people living in same location
   b. captured as a numerical string

9 Government issued identification document (if applicable)

   a. Need to capture both type and ID number

      i. gov_id_type: Select from **drop-down** lists the types of government-issued ID that are applicable for the country (national identity card, passport, etc.)

      ii. gov_id: identification number of ID (i.e. passport number) captured as alphanumeric string

Meta Data Fields for Sharing

1. registering_org
→ name of organisation creating original record of person

2 timestamp_orginal
→ Time/Date record set created

3 modifying_org
→ name of organisation modifying record

4 timestamp_mod
→ Time/Date record set modified

5 staff_mod
→ name of staff member modifying record

6 reason_mod
→ reason for modification

Retained Mandatory Data Fields (not shared)

1 consent_to_capture
→ Indicates if this person has given consent for their personal details to be captured and stored by organisation

2 consent_to_share_info
→ Indicates if this person has given consent for their personal details to be shared with other agencies.

Retained Meta Data Fields (not shared)

- staff_reg
→ name of staff member who is original the data collector

- staff_id
→ ID number of staff member

DESIGN PROCESS REPORT
CCD Data Governance Pilot South Sudan

Collaborative Cash Delivery