# CCD DATA INTEROPERABILITY WORKING GROUP

*July 23, 2020*

*Amos Doornbos*

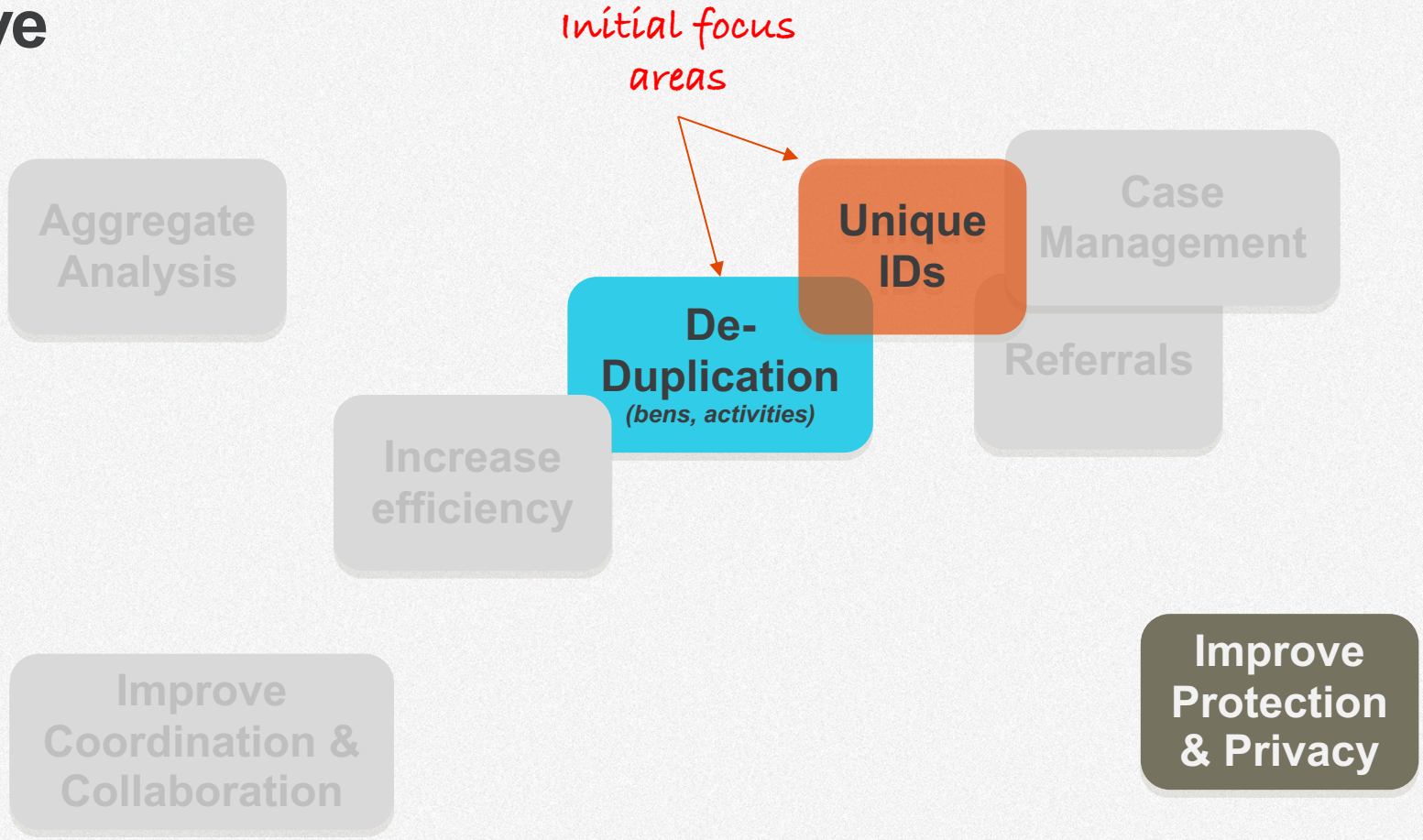
# Objectives

*The objectives of the CCD WG are to agree on :*

- *The purpose of data sharing*
- *The legal basis for collecting and sharing data*
- *What data the CCD members will share and the definitions of each data field*
- *The criteria for accessing the shared data*
- *How the shared data can be used*
- *How CCD members will be collectively transparent with those about whom the data is*
- *Options for the mechanics of sharing (i.e. methods, requirements, pros & cons, etc.) and*
- *Where possible and appropriate, support country consortiums with on going data sharing questions, concerns and problems*

*It is expected that through the work of this WG issues will arise requiring input from or decisions by other working groups and work streams. When this occurs, these will be documented and shared with the appropriate person(s).*

# Purpose/ Problem to Solve



# Who are we sharing the data with?

Often a fuzzy line, however when CCD has a standard methodology amongst its members, engaging others will be easier

**Phase One**
CCD Country Consortium

**Phase Two**
CCD Country Consortium
+ other NGOs, UN, Gov't?

# Deduplication Working Assumptions

We are seeking to deduplicate at a ***CCD country project*** level

We are deduplicating **individuals** not HHs

We are **not** seeking to share a ***collective*** transaction history for each beneficiary

System to flag duplicates online; resolve offline

Technology **Agnostic**

Biometrics should **not** be a requirement

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

Share the **minimum** data possible

# We’re seeking Exact and Fuzzy Matches

Fuzzy Matches require more data

Exact require less data

# Required Data

# Naming Conventions (Rules)

*To reduce the number of false duplicates, there are a few naming conventions to follow:*

### *Global Decisions*

- System coding language = English (British? American? Canadian? Australian?)
- Singular names are used for tables. Both table names and column names use lowercase letters only with words separated by underscores.
- Standardised list of country names

### *CCD Country Consortium Decisions*

- User language determined by country CCD consortium
- 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.)
- Which gov't issued ID will be captured if present

# Mandatory Data

*For deduplication there 2 types of data required (mandatory) to be shared and 2 types of data required, **not** to be shared, but to be retained by the organisation collecting the data:*

*To be **shared**:*

1. **Shared data about the person**
2. **Shared Metadata**

*To be **retained**:*

1. **Data collected from the person**
2. **Metadata**

# Mandatory Data Details

## *Mandatory for Sharing*

1. first_name
2. family-name
3. Gender (M/F/other)
4. date_of_birth (YYYY
5. place_of_birth
6. community_id
7. mobile_phone_id
8. hh_size
9. Government issued identification document
    A. gov_id_type:
    B. gov_id:

## *Mandatory Metadata*

1. registering_org
2. timestamp_orginal
3. modifying_org
4. timestamp_mod
5. staff_mod
6. reason_mod

## *Mandatory Retained*

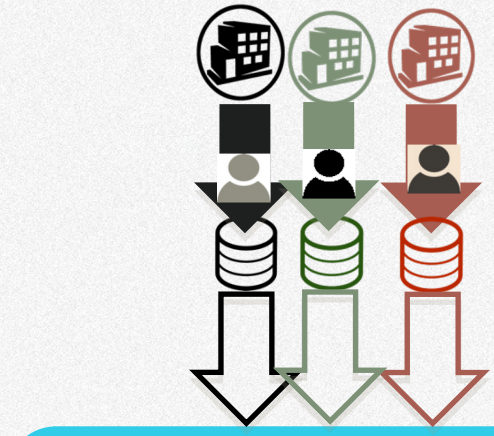
1. consent_to_capture
2. consent_to_share_info

## *Mandatory Metadata*

1. staff_reg
2. staff_id

# Sharing Model

# Chosen Model
# Phase 1



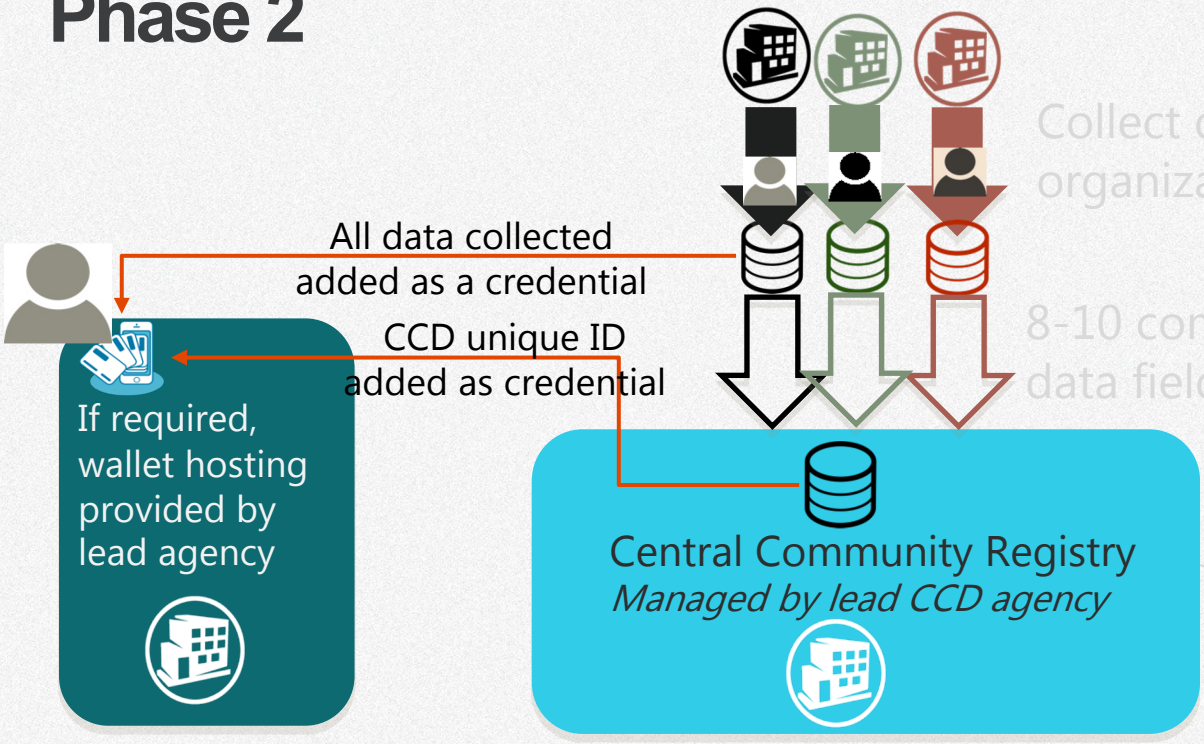
*Each organisation to collect data as required by organisational policies*

Only the 8-10 common data fields shared

Central Community Registry
*Managed by lead CCD agency*

*Checks for potential duplicates when new records added. Creates CCD unique record for each person*

# Chosen Model
# Phase 2




Collect data as required by organizational policies

8-10 common data fields shared

*Checks for potential duplicates when new records added. Creates CCD unique record for each person*

# Sharing Process & Community Registry Notes

InfoSec teams will need to define the requirements better for us

*Security*

- *All Sharing of data must be done through encrypted means*
- *Community registry database must always be encrypted and meet the highest IT security global standards and GDPR standards*
- ***Default** will be to meet GDPR standards due to some CCD members needing to be compliant*

*Database location*

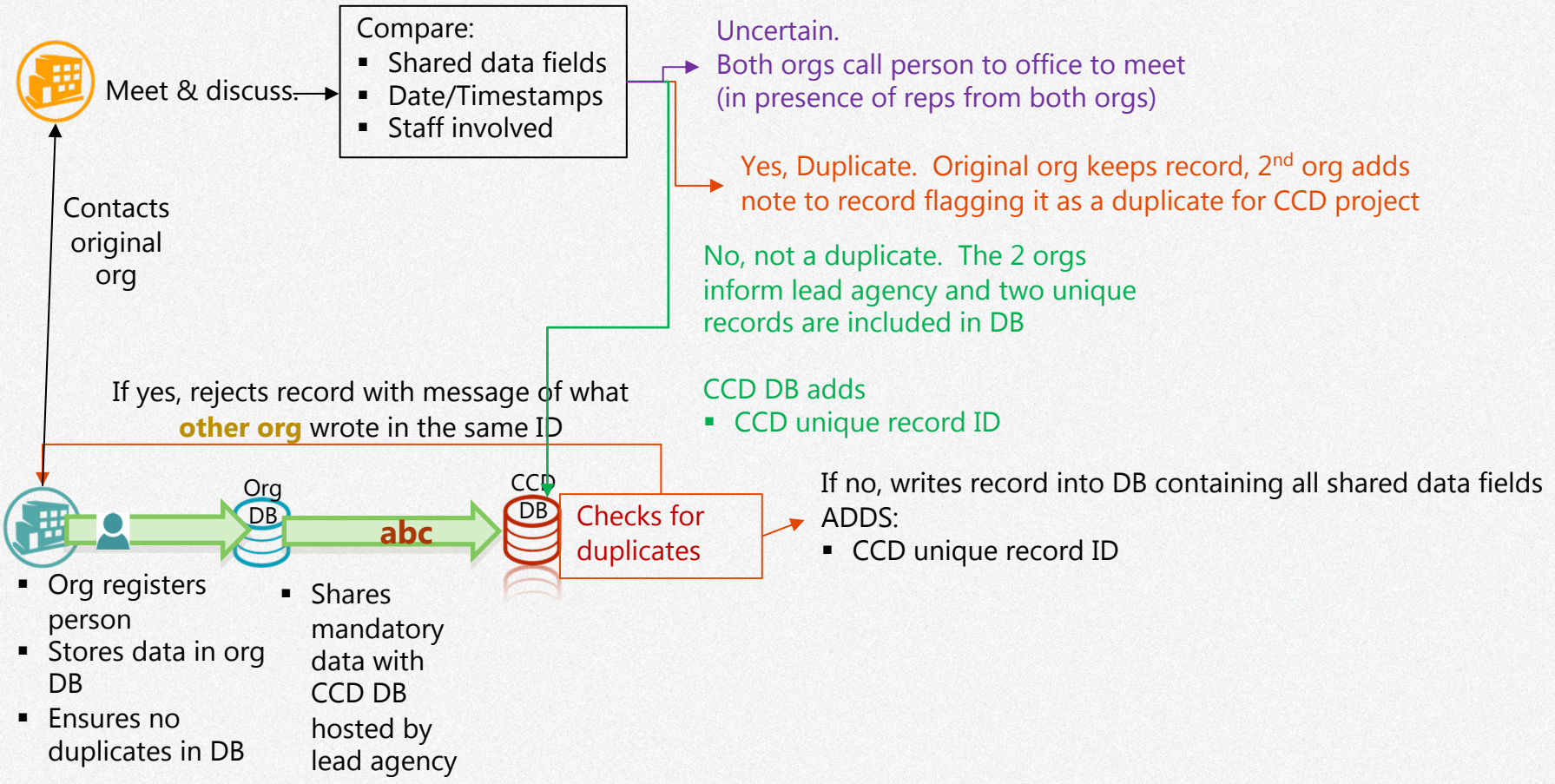
- ***Default** will be for the community registry database to be hosted within the country of operations.*
- *However, if the lead agency would like to host the community registry outside of the country of operations, express, written approval must be obtained from all participating agencies*

*Access*

- *Access to the community registry should be strictly limited to the appointed database manager in the lead agency*
- *Roles with access should be stated as part of the data sharing agreement*

# Duplicate Resolution Process

# Duplicate Resolution Process



# Where Gov't issued IDs exist

*100% coverage of common gov't ID*

*- simply use gov't issued ID as deduplication check*

*- check still done in common registry*

*Coverage less than 100%*

*1. run deduplicate check on common gov't issued IDs*

*2. run deduplicate check using all data fields*