

Training on online data collection (Kobo)

Kabul, 24 April 2018



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

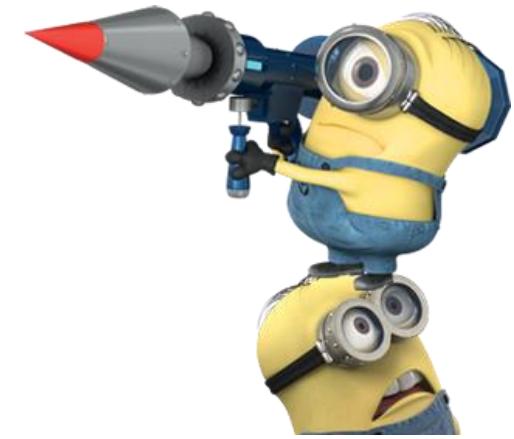
Session 1. INTRODUCTION



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Session 1 Objectives

- Learn skills to **independently develop simple form**, to modify existing one, to extract and analyze the data collected
- To create a **community of practice** with an understanding of data collection & able to develop and relay a **data collection campaign**
- Objective, to get you as **Super User** of Kobo data collection tool



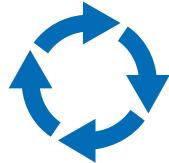
Session 1 Evolution Tree of Users



Session 1 Evaluate yourself



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter



Information Management is a cycle of six stages



Processing



Dissemination



Analysis



Design



Combining



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Why Online/mobile data collection tool?



INTERVIEW



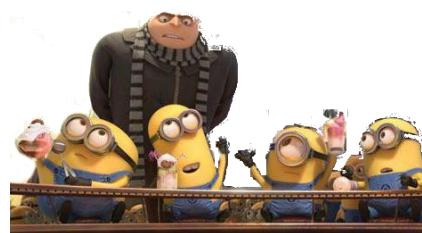
FIELD AGENT

??



DATA ENTRY

??
??



QUALITY CONTROL

??
??



RAW DATA ANALYSIS

??
??
??



REPORT



??



??
??



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Why Kobo ?

Acquee
COMMANDmobile®
CommCare
CommTrack
CSPromo
CyberSync
DevInfo
do For
droidSurvey
Enketo Smart Paper
EpiCollect



FrontlineSMS
Fulcrum
GeoChat
GeoPoll
Humanitarian Data
iSURVEY
KoBo
Last Mile Mobile
Solution

Magpi
Majella Insight
Mabenzi Researcher
Nokia Data
Open Data Kit
openXdata
Pendragon
Poimapper

PSI Mobile –
Fusion
RapidSMS
SoukTel
Telerivet
ViewWorld
Voxiva
Wepi



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Agenda

Morning

08h30-08h45 Session 1 Introduction

08h45-09h30 Session 2 Design

09h30-10h30 Session 3 Collection

10h30-10h45 Coffee/Tea

10h45 -11h30 Session 4 Combining/
Processing

11h30-12h00 Session 5 Analysis/Dissemination 15h30 -15h45 Wrap up

Afternoon

13h00-13h30 Session 6 Programing
My 5 first question on Kobo

13h30-14h30 Session 7 Programing

Adding 10 more: date, pictures...

14h30-14h45 Coffee

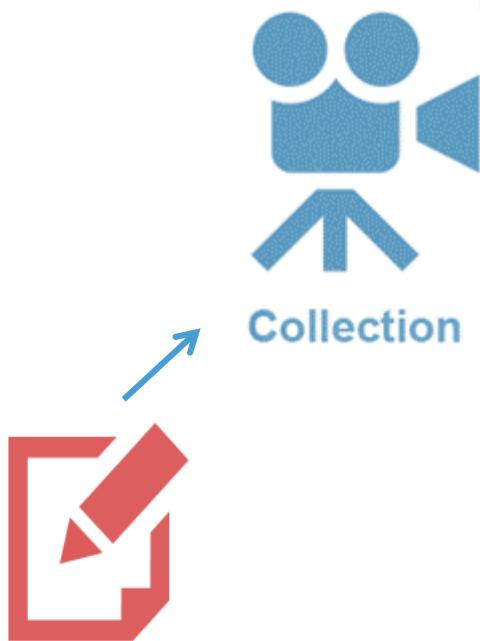
14h45-15h30 Presentation of Participants



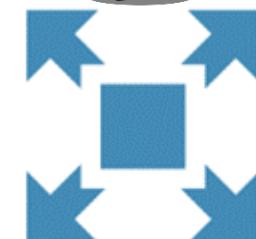
Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter



Information Management is a cycle of six stages



Processing



Dissemination



Analysis



Design



Combining



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Session 2. Making a plan is first step to any data collection!

- Plan before you start data collection;
- Define roles and assign responsibilities;



s;

ment dat



em

- Revisit the plan throughout project



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Session 2. Assess data needs

Strategic

- **WHAT** you want to measure
- **WHY** you need to measure
- **HOW** you will measure



Operational

- What is **level** of data collection?
- Design your **data model**
- **Explain** your data



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter



Session 2. **WHAT** you want to measure



Population data

- # IDPs
- # affected communities



Damage data

- # destroyed houses
- # damaged houses (cat 1,2,3)



Other thematic data

- # communities without water
- # cases of flu per week

Format



Text



Number



Date



GPS Coordinates



Picture

Type



Session 2. WHY you need to measure it?



We do not have
these data! We
need it NOW!



I want to have an overall
picture! What is going
on outside this office???



Bring me
beneficiary
lists NOW!

- Quantify needs;
- Register beneficiaries;
- Monitor and evaluate program
- Etc...



Session 2. HOW you will measure it?

Check for available data



- Government data;
- Assessments;
- Available researches;
- Info from partners...



Going to the field to collect data directly from people



Primarily data collection



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Session 2. LEVELs of data collection



Individual



Note: More detailed data collected requires more time, people and resources!



Household

Levels of data collection are interlinked:
Individual → Household → Community



Community (village, city, district, region)



Institution (school, Collective Center)

etc...

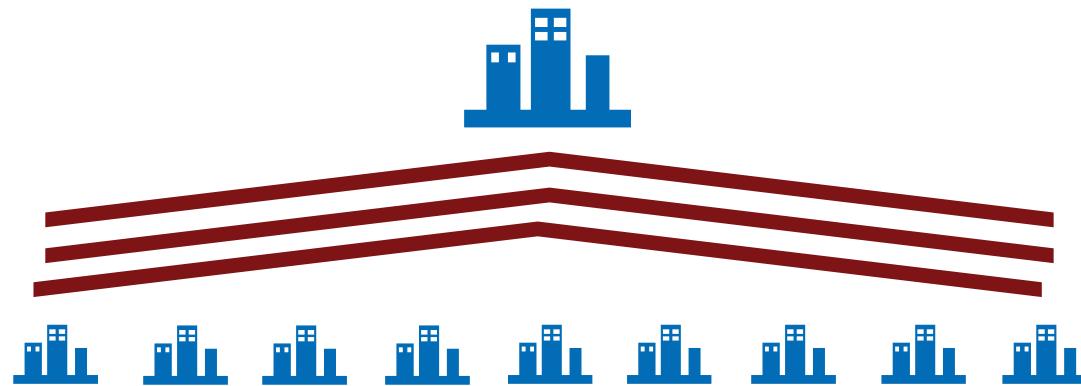


SAMPLING



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

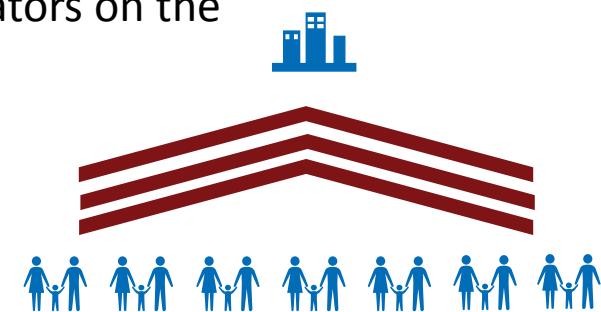
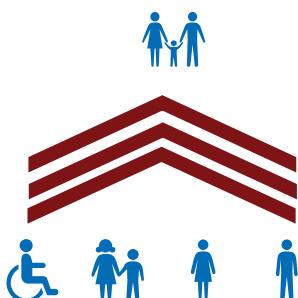
Session 2. Sampling and confidence level - 1



Having enough representative number of assessed units, it is possible to make judgments about the entire group.

Sampling – scientifically defined number of enough interviews to be able to assess entire group.

Confidence level – degree to which indicators on the bigger group are statistically relevant.



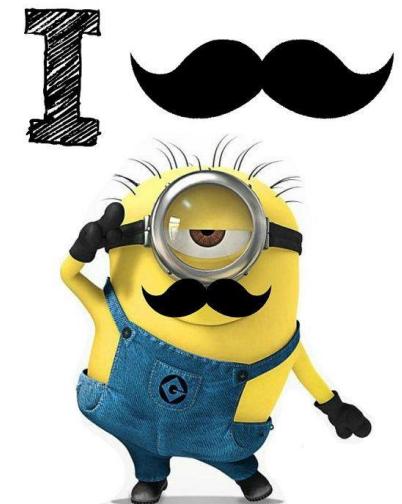
Session 2. Sampling and confidence level - 2

- To calculate sample size, confidence level and confidence interval, see:
<http://www.surveysystem.com/sscalc.htm>
- Let's try calculate sample size for population group **30,000 individuals** and with **confidence interval 5%, confidence level is 95%**...
 - 379...



Session 2. Design a data model

- Design your questions;
- Decide on types of answers;
- Explain your data.



Session 2. Designing your questions

Open questions:

What is your most urgent need ?	Father is ill and cannot work
---------------------------------	-------------------------------

Closed questions:

What is your most urgent need ? (Pick one)	<input type="checkbox"/> Food
	<input checked="" type="checkbox"/> Shelter
	<input type="checkbox"/> Water
	<input checked="" type="checkbox"/> Livelihoods
	<input type="checkbox"/> Health
	<input checked="" type="checkbox"/> Education
	<input checked="" type="checkbox"/> Other (specify)

- ! Requires human reading in order to analyse

- ! What about answers you had not thought of?



Session 2. 1,2,3-choice questions

- ✓ Single choice questions;
- ✓ Multiple choice questions...
- ✓ ...

WHAT IS THE DIFFERENCE and WHEN we use them?



Session 2. SMART Indicators

- ✓ **Specific** – target a specific area; should be clear what is to be measured/improved
- ✓ **Measurable** – quantifiable or clear qualitative measurement; something that can be expressed in numbers or in terms of a meaningful scale of values
- ✓ **Achievable** – has to be possible to measure from an operational standpoint
- ✓ **Relevant** – relevant and useful in measuring the need/activity/objective it's linked to
- ✓ **Timebound** – must be measurable in a specific period of time (more relevant to monitoring)

Session 2. Design a data model

A data model describes how you store your data
Regions

Region	Number of houses	Number of people	GCA/NGCA
TEXT	NUMBER	NUMBER	GCA/NGCA

Agencies

Agency Name	Email address	Phone number	Active
TEXT	TEXT	TEXT	YES/NO

! Too much data in one field makes it difficult to analyze data

Session 2. How to build the model in Excel

- First row headers
- One data type per column
- One piece of data per cell
- A sheet with the definition of your data on

Example

Region	Province	Municipality	Barangay(s) (Preferred option: one barangay per row)	Is location a camp/evacuation centre?
Use drop down menu	Use drop down	Use drop down	Type text	Yes/No

Example

Tarpaulins - 1 per HH	Tarpaulins - 2 per HH	Tents	NFI : Household	NFI : Kitchen Sets	Tech. Assistance (Em. Shelter)
HH	HH	HH	HH	HH	Yes/No
100		100	50	50	Yes
200		30	30	70	

Example

Guidance notes



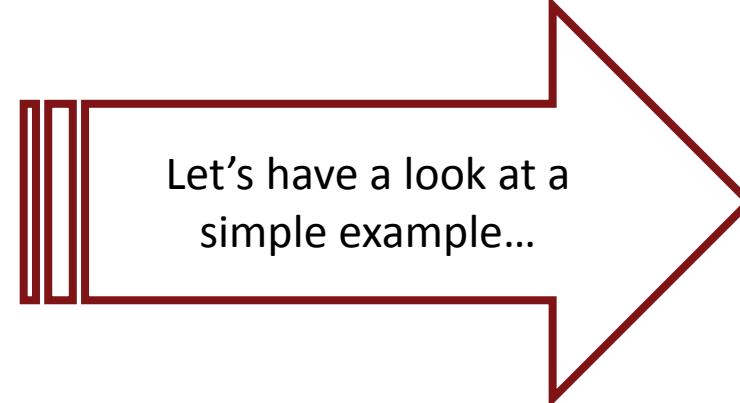
Session 2. Explain your data

- Give explanation of each data

Example

Partially damaged house

A house is partially damaged if it is still repairable.



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Session 2. Data model example

questionnaire	raw_data	guidance_notes
---------------	----------	----------------

2		
3	Questions	
4	What is your name?	First name of the respondent, spelled as in official documents in English.
5	What is your current address?	Current actual address of the respondent (not registration, but actual).
6	How old are you?	Age of the respondent. Only full years.
7	What is your marital status?	Marital status of the respondent. Official, not actual.
8	married	marriage officially registered
9	divorced	currently single, was married before, but divorced
10	single	single
11	widowed	currently single, was married before, but spouse died
12		
13		

RAW DATA

A	B	C	D
1			
2	Question	Type of question	Options
3	What is your name?	open	
4	What is your current address?	open	
5	How old are you?	open, restricted to number	
6	What is your marital status?	closed	married single
7			divorced widowed
8			

QUESTIONNAIRE

A	B	C	D	E
1 name	address	age	marital status	
2 Oleksii	Donetsk region city N, street X, 55	19	single	
3 Maria	Poltava region city N, street X, 56	45	married	
4 Sergii	Kyiv region city N, street X, 57	32	divorced	
5 Igor	Luhansk region city N, street X, 58	67	widowed	
6				



Session 2. P Coding - 1

- Each region, province, district and village have its own pre-defined unique number: the P-code
- This is useful, because many units have the same names
- Sometimes if you have your location question “open”, it is impossible to find the right village...

Example



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Session 2. Common Operational Datasets

- The Common Operational Datasets (CODs) are critical datasets that are used to support the work of humanitarian actors across multiple sectors. They are considered a de facto standard for the humanitarian community and should represent the best-available datasets for each theme.
- They may include:
 - Administrative boundaries;
 - Populated places (settlements);
 - Transportation network (roads, airports, checkpoints);
 - Hydrology (rivers etc)
 - Population statistics (IDPs, resident population etc);



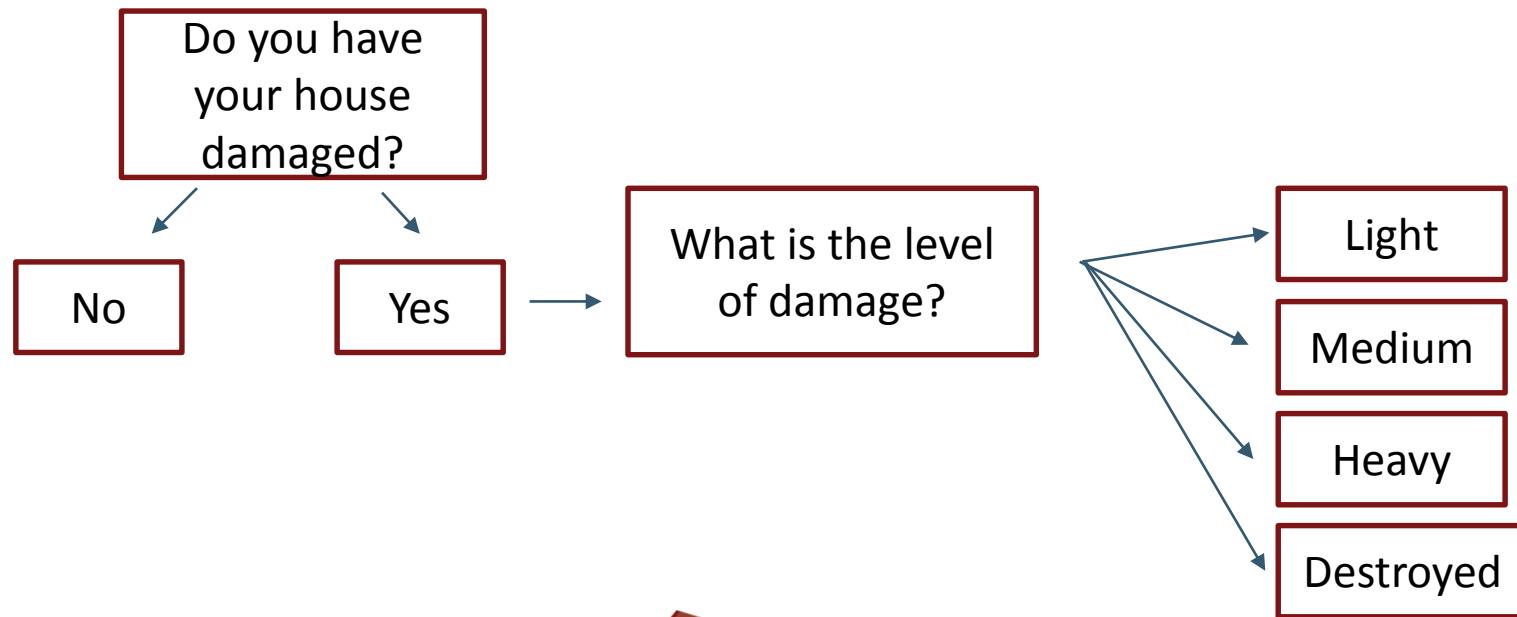
Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Session 2. Where to get CODs?

- Humanitarian Data Exchange:
<https://data.humdata.org/dataset/afg-admin-boundaries>
- OCHA Afghanistan

Session 2. Structure and priorities

- There are always several key questions and many additional.
- Use cascading option to prioritize and set a structure:



Session 2. EXERCISE 2

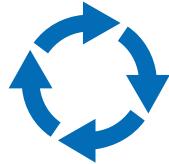
- In groups prepare data model with 10 questions on a specific thematic (15 min):
 - Village assessment;
 - Monitoring and evaluation;
 - Beneficiary registration;
 - Damage assessment.
- Take into account the following:
 - Level of data collection;
 - Types of questions;
 - Are they SMART?
 - Structure.



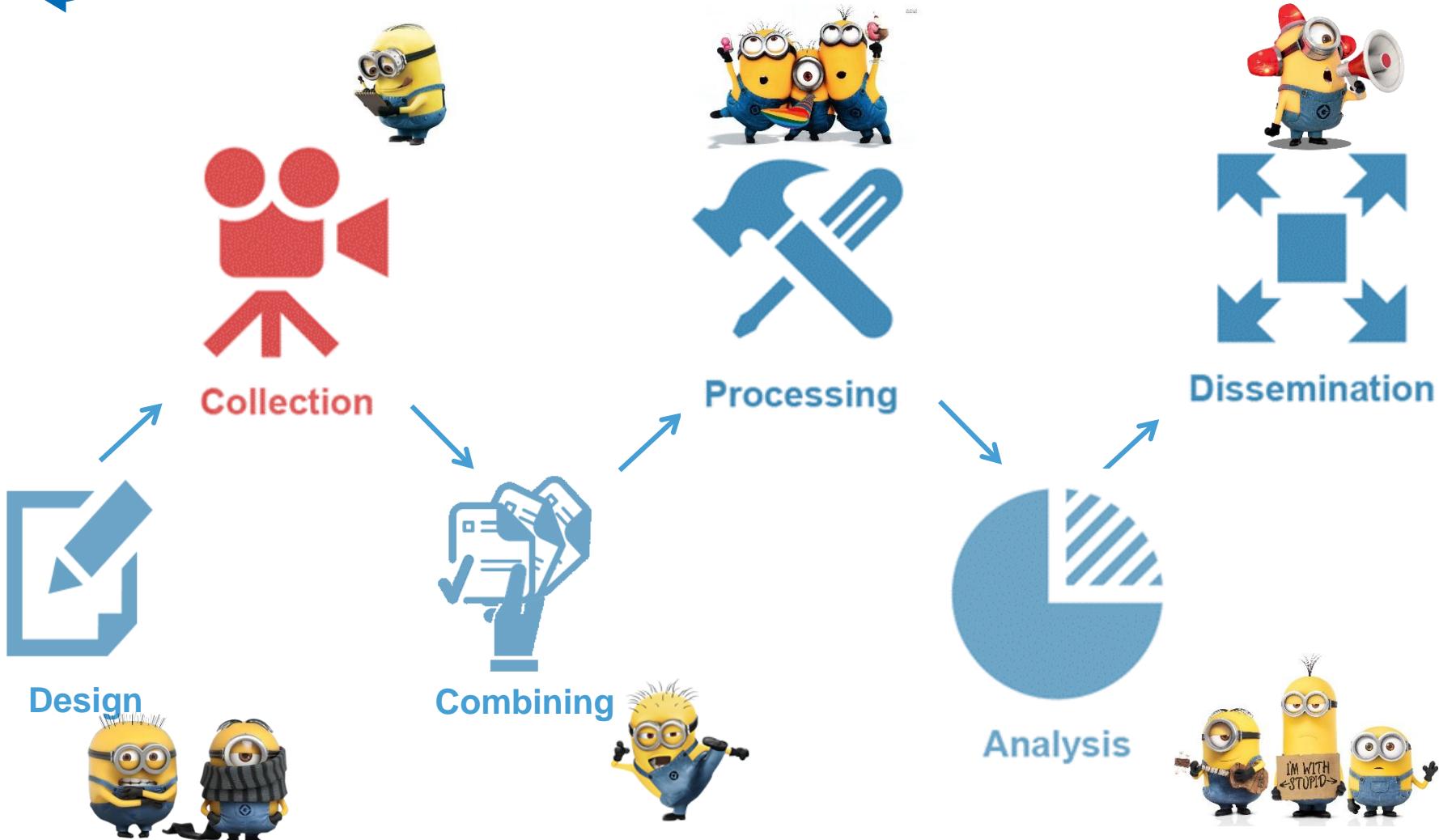
Session 3. COLLECTION



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter



Information Management is a cycle of six stages



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Session 3. Kobo main page



Page address is:
<https://kobo.unhcr.org>

- Account for humanitarian agencies is free of charge;
- As many surveys as possible;
- 5 min to register



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Session 3. KOBO MAIN MENU

The screenshot shows the KoboToolbox web application interface. At the top, there is a navigation bar with tabs: 'Quick Start Overview' (highlighted), 'Multiple Ways To Capture', 'Projects | KoboToolbox', and a download icon. Below the navigation bar is a header with the 'KoBoToolbox' logo and a search bar labeled 'Search Projects'. On the left, there is a sidebar with categories: 'NEW' (selected), 'Deployed' (0), 'Draft' (0), and 'Archived' (1). The main content area displays a table for 'Archived' projects, showing columns for 'Name', 'Shared by', 'Created', and 'Last updated'. One project is listed: 'ESNFI Cluster SAG Meeting' shared by 'haidari' on April 4, 2018. A user profile dropdown menu is open on the right, showing the user 'kadir' from 'mohameke@unhcr.org' with options for 'ACCOUNT SETTINGS', 'Language', and 'Logout'.

- **Projects** – all forms that are deployed and work online
- **New** – To create new and upload forms that user is working on.
- **Deployed**
- **Draft**
- **Archive**
- **Library** – library of questions, if created.
- **New**–To access questions, upload and collection
- **My Library**
- **Public collections**
- **Settings**

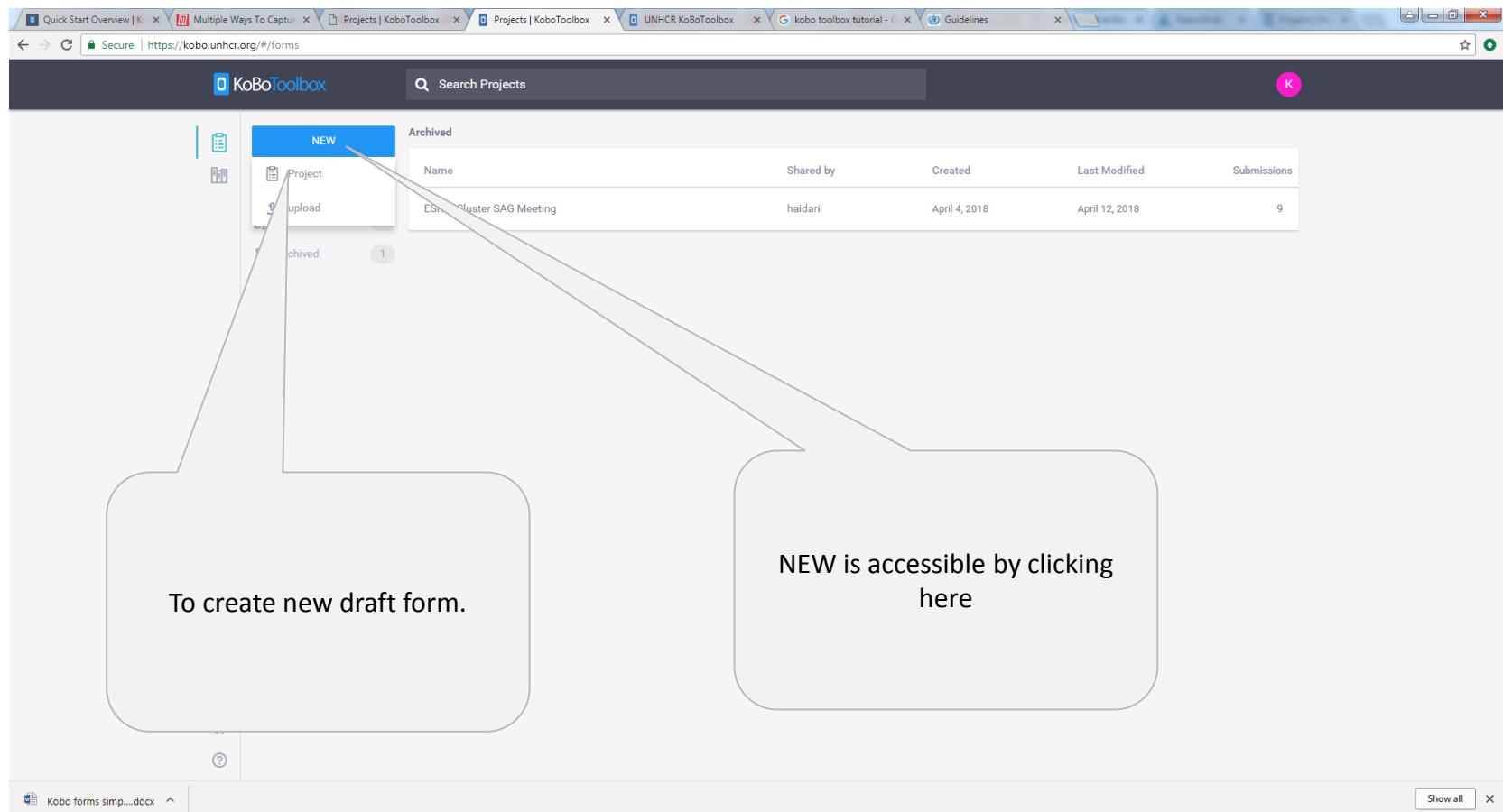


Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

MENU → NEW

Session 3. New

Click on Projects, then on “new”, and then two choice “ project” and “upload”



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

NEW → PROJECT

The screenshot shows the KoboToolbox interface for creating a new project. The main form is titled "Create New Project from Scratch". It includes fields for "Project Name" (with placeholder "Enter title of project here") and "Description" (with placeholder "Enter short description here"). Below these are dropdown menus for "Sector" (labeled "Select...") and "Country" (labeled "Select..."). A checkbox option "Help KoboToolbox improve this product by sharing the sector and country where this project will be deployed. All the information is submitted anonymously, and will not include the project name or description listed above." is present. A large blue "CREATE PROJECT" button is at the bottom right. The interface has a sidebar with categories: NEW (selected), Deployed, Draft, and Archived. A search bar at the top says "Search Projects". The browser's address bar shows the URL <https://kobo.unhcr.org/#/forms>. Callout boxes provide instructions: "Write here the name of the form" points to the Project Name field; "Write here the short description of the form" points to the Description field; "Finally click here to start creating the form" points to the CREATE PROJECT button; "Select here your sector" points to the Sector dropdown; and "Select here country" points to the Country dropdown.

Write here the name of the form

Write here the short description of the form

Select here your sector

Select here country

Finally click here to start creating the form



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Show all

PROJECTS → Deployed → SETTINGS

The screenshot shows the 'SETTINGS' tab of a project titled 'Beneficiary selection assessment tool'. The interface includes sections for 'Form Settings' (Accept Submissions: YES), 'Existing Form Files' (+ Add Document), and 'Delete project with all data and its form' (Delete button). A large callout box highlights the 'Accept Submissions' setting with the text: 'Form can be deployed, but disabled for submissions'. Another callout box points to the '+ Add Document' button with the text: 'Additional documents may be uploaded, in case form uses media embedded in questions and/or notes media (pictures, sounds)'. A prominent warning message 'DO NOT DELETE WITHOUT THINKING TWICE!!!' is displayed in a large box. The right side of the screen shows sharing options: 'Share Project Publicly' (NO selected), 'Share Data Publicly' (NO selected), and a 'Turn on' button for 'Sharing by link is off', which is currently off. It also shows user permissions for 'kedir': Can Edit, Can Submit to, Can Validate, and 'For user:' dropdown set to '2dhis2'. A 'Save Permissions' button is present. A large callout box on the right explains the sharing options: 'Link to the form can enable users without login and password access and view project.' and 'Link to the form can enable users without login and password access and view all the data.' A final callout box at the bottom right explains the sharing feature: 'Projects within Kobo can be shared with other Kobo users. Three level of rights: to edit, to view and to submit data.'

Form can be deployed, but disabled for submissions

Accept Submissions
YES

+ Add Document

Share Project Publicly
NO

Share Data Publicly
NO

Turn on

Sharing by link is off

kedir: Can Edit | Can Submit to | Can Validate

For user:
2dhis2

Perm type: Can view

Save Permissions

DO NOT DELETE WITHOUT THINKING TWICE!!!

Link to the form can enable users without login and password access and view project.

Link to the form can enable users without login and password access and view all the data.

Projects within Kobo can be shared with other Kobo users. Three level of rights: to edit, to view and to submit data.



MENU → PROJECTS → NEW → Deployed → FORM

The screenshot shows the KoboToolbox interface for a project titled 'Beneficiary selection assessment tool'. The 'FORM' tab is selected. On the left, there's a sidebar with 'NEW' (highlighted), 'Deployed' (with 1 item), 'Draft' (with 0 items), and 'Archived' (with 1 item). The main area displays deployment options:

- Online-Offline (multiple submission)**: Allows online and offline submissions.
- Online-Only (multiple submissions)**: Best for entering many records at once.
- Online-Only (single submission)**: Allows a single submission and can be paired with a return URL.
- Embeddable web form code**: Use this HTML5 code snippet to integrate your form into your own website.
- View only**: Use this version for testing and previewing without submitting data.
- Android**: Use this to collect data in the field with your Android device.

On the right, there are several buttons: a pencil icon, an eye icon, a copy icon, an open icon, and a 'DEPLOY' button. A pink callout bubble points to the 'DEPLOY' button with the text: "It shows how it look like the online form. It is similar with preview." Another callout bubble points to the 'OPEN' button with the text: "It allows to copy and share the link of the form". A third callout bubble points to the 'COPY' button with the text: "Replace with XLS". A fourth callout bubble points to the 'OPEN' button with the text: "Download data in XLS and XML. Share and Clone the project".



MENU → PROJECTS → Deployed → SUMMARY

The screenshot shows the KoboToolbox interface for a deployed form titled "Partner's CWC Focal Point". The main area displays the "SUMMARY" tab, which includes sections for "Description" (containing the text "Partner's CWC Focal Point"), "Submissions" (showing 0 submissions for the past 7 days and 0 for the past 31 days, with a total of 15), and "Quick Links" (including "Collect data", "Share form", "Edit form", and "Preview form"). To the right, there are sections for "Form details" (last modified March 7, 2018, latest submission March 19, 2018, 6 questions), "Data" (Reports, Table, Gallery, Downloads, Map), and a "Gallery of pictures" (represented by a purple speech bubble). A large callout bubble on the left side points to the "Submissions" section, stating "It shows the summary of submissions of form in table". Another callout bubble at the bottom points to the "Data" section, stating "Export data submitted to date in XLS, CSV, ZIP, KML or Excel Analyzer". A third callout bubble on the right points to the "Gallery" section, stating "Gallery of pictures".

It shows the summary of submissions of form in table

Export data submitted to date in XLS, CSV, ZIP, KML or Excel Analyzer

It shows the summary report of submission of form

Gallery of pictures



ONLINE FORM

The screenshot shows the KoboToolbox platform interface for a 'Beneficiary selection assessment tool'. At the top, there are two tabs: 'Draft ENG_Dari_Pashto_B...' and 'Beneficiary selection ass...'. The main title 'Beneficiary selection assessment tool' is displayed above the form area. On the left, there's a sidebar with icons for 'Form Preview', 'Form Settings', 'Data Entry', and 'Help'. The main content area is titled 'Beneficiary selection assessment tool' and contains the following fields:

- * Enumerator name: An input field.
- * Province: A dropdown menu showing 'none selected'.
- * District: A dropdown menu showing 'none selected'.
- * Village: An input field.
- * Please record GPS: A note at the bottom.

On the right side of the form, there are buttons for 'Choose Language' (set to 'default') and a printer icon. A large callout bubble points to the printer icon with the text 'Print entire form for offline use'. Another callout bubble points to the language selection dropdown with the text 'Language selection (for Multilanguage forms only)'. The status bar at the bottom right shows '0 submissions' and a profile icon.



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Session 3. Steps to create and edit question

Steps to follow

- Step 1: write the question
- Step 2: Click Add question
- Step 3: Define type of variable
- Step 4: Go to settings and set options of question
- Step 5: Set hints if necessary



Session 3. Creating and editing forms

Manual

- Coding in Excel, transforming in XML;
- Needs advanced skills and knowledge;
- Allows to use more functionality (ex.: cascading is not yet available in visual editor);
- Allows design form better.



Visual editor

- Easy for beginners;
- Reduced functionality;
- Quick to start;
- Allows immediate preview.



Session 3. Adding questions to the form

The screenshot shows a mobile application interface for creating a form. At the top, there are four categories with callout bubbles: "One choice question", "Multiple choice questions", "Open questions", and "Only numbers are accepted as an answer". Below these are buttons for "form name", "CREATE", "CLOSE", "PREVIEW", and "FILE". A "METADATA" section includes "START DATE, END TIME" and a gear icon. A large central area displays a grid of question types:

<input checked="" type="radio"/> Select One	<input type="checkbox"/> Select Many	abc Text	123 Number
1.0 Decimal	Date	Time	Date & time
GPS	Photo	Audio	Video
Note	Barcode	<input checked="" type="checkbox"/> Acknowledge	1+1 Calculate
Matrix / Rating	Ranking		

At the bottom left, a callout bubble says: "Additional explanation and text. May be introduction message."



Session 3. Labels and values - 1

- Label: what is seen in user mode: question text and options as answers etc.
- Value: the way data recorded in the online database. The same format would be seen when export data for analysis.

!!! Proper values at the design stage are critical to simplify analysis process later. !!!

Question text

Did you receive any assistance

Answers' labels: what user will see when using form

yes

Value:

01

no

Value:

00

No answer

Value:

no_answer

- Click to add another response...

Value:

AUTOMATIC

Answers' values: what YOU will see when exporting submissions



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Session 3. Labels and values - 2

- Labels and values are critical for analysis:
 - All data is exported in the format of values, not labels.

Column header: is made out of value, not label (available only in manual coding)

Column Headers		Answer Values				
	Show XML Values		Show XML Values			
rt	adaptation_gaps	used_information_internally	help_advocacy	useful_advocacy_donors	useful_advocacy_government	Involvement_nationalAuthorities
some_adaptation	yes		no	yes	do_not_know	somewhat_effective
no_adaptation	no		no	no	no	somewhat_effective
some_adaptation	yes		no	no	do_not_know	i_know
yes_completely	yes		yes	yes	no	mostly_effective
do_no_know	yes		yes	yes	yes	somewhat_effective
some_adaptation	no		no	no	no	mostly_effective
yes_completely	yes		yes	yes	yes	very_effective
some_adaptation	yes		no	yes	yes	mostly_effective

All answers are values, not labels. Easiness of their understanding is upon form designer!



Session 3. Additional features of Kobo

- Hints;
- Restrictions and restrictions messages;
- Validation of data;
- Adding picture;
- Adding geodata;
- Introduction to designing form manually;
- Multilanguage support
- Using mobile devices for data collection...



Session 3. Before going for data collection

- Have a plan! Who goes where and when, how often, logistics planned etc.
- Have idea what you will do with data after it is collected!
- Brief and train your team! Communicate priorities!



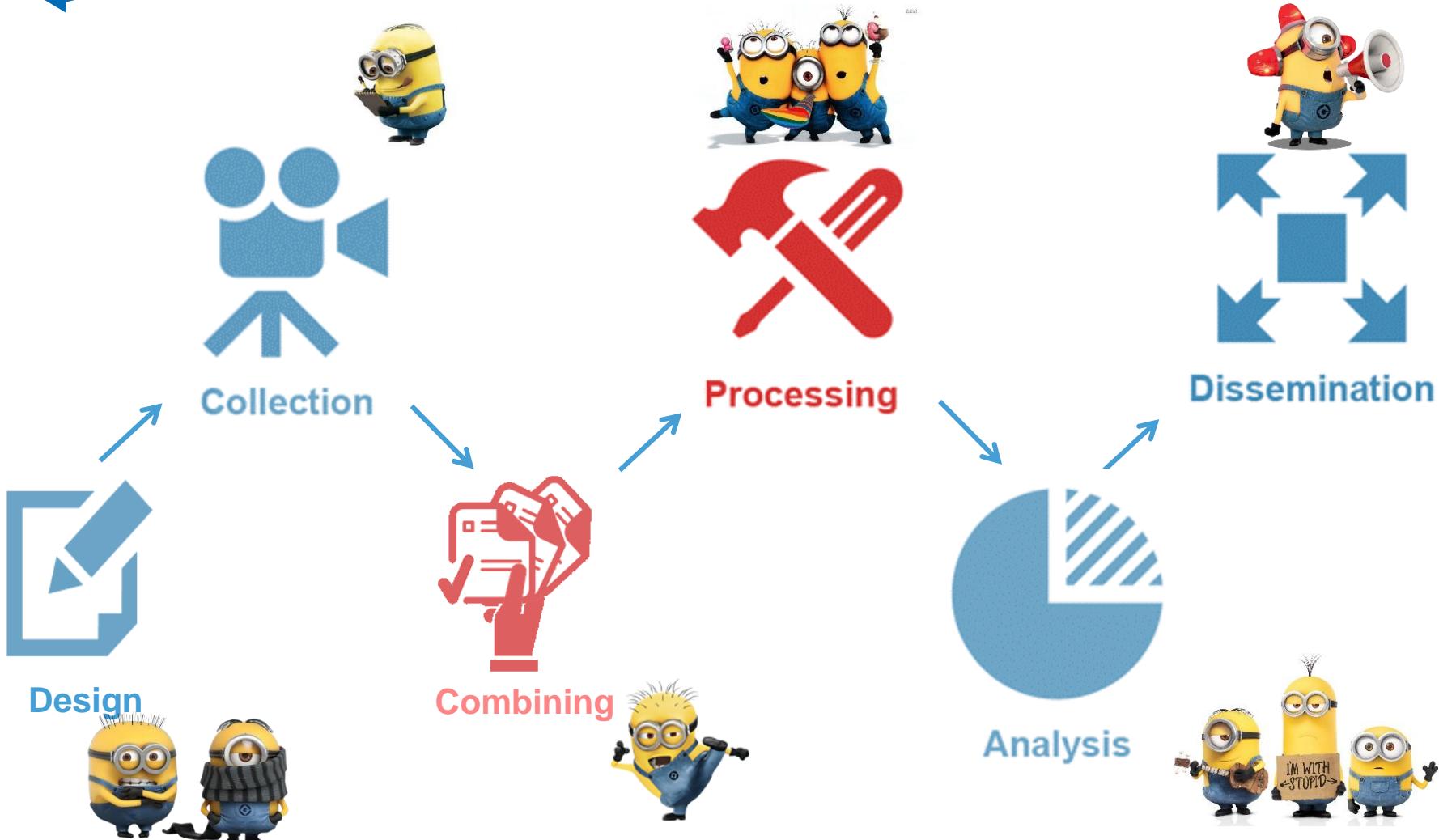
Session 4. Combining and processing data



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter



Information Management is a cycle of six stages



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Session 4. Combining data

- Always all the data is pulled together into one document for further processing and analysis.
- Depending on the way data was collected combining data may happen in different ways and steps...



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Session 4. Paper based form

- **Digitalizing data** – entering data from paper form into Excel or any online form;
- **Merging all data** from different localities into master one;
- **Cleaning data**, validating data and beginning of analysis.



Session 4. Excel based form

- ~~Digitalizing data—entering data from paper form into Excel or any online form;~~
- **Merging all data** from different localities into master one;
- **Cleaning data**, validating data and beginning of analysis.



Session 4. Kobo form

- ~~Digitalizing data—entering data from paper form into Excel or any online form;~~
- ~~Merging all data from different localities into master one;~~
- **Cleaning data**, validating data and beginning of analysis.



Session 4. Well-designed Kobo form

- ~~Digitalizing data – entering data from paper form into Excel or any online form;~~
- ~~Merging all data from different localities into master one;~~
- ~~Cleaning data, validating data and beginning of analysis.~~



Session 4. Extracting data from Kobo

- Simply click “**Download data**” at the project management form and select format: XLS, CSV, ZIP, KML...

The screenshot shows a user interface for managing Kobo projects. At the top, there are three tabs: 'SUMMARY', 'FORM', and 'DATA', with 'DATA' being the active tab. On the left, there's a sidebar with icons for 'Reports', 'Table', 'Gallery', 'Downloads' (which is currently selected), and 'Map'. The main area is titled 'Download Data' and contains a dropdown menu labeled 'Select export type'. The dropdown menu lists several options: 'XLS' (which is highlighted with a blue background), 'XLS (legacy)', 'CSV', 'CSV (legacy)', 'Media Attachments (ZIP)', 'GPS coordinates (KML)', 'Excel Analyser', and 'SPSS Labels'. To the right of the dropdown, there's a section labeled 'Value and header format' with a dropdown set to 'English'.



Session 4. Data cleaning

- **Data cleaning** is the process of detecting and correcting corrupt or inaccurate records from a database.



Session 4. Data cleaning

■ Be creative!

- Lookup functions
 - Easy to find non-existing codes (typos)
- Formulas
 - Check for mathematical and logic consistency
- Compare with other sources (**Triangulation**)
 - Validation of values/expected ranges (do we have approximately the same)
- Compare with previous years
 - Validation of values/expected ranges (do we have approximately the same)



Session 4. Useful Excel Tools

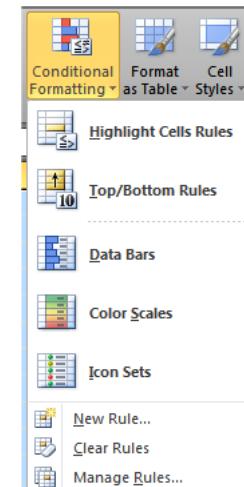
- Data -> Validation (allows only certain values)



- Data -> Sort & Filter



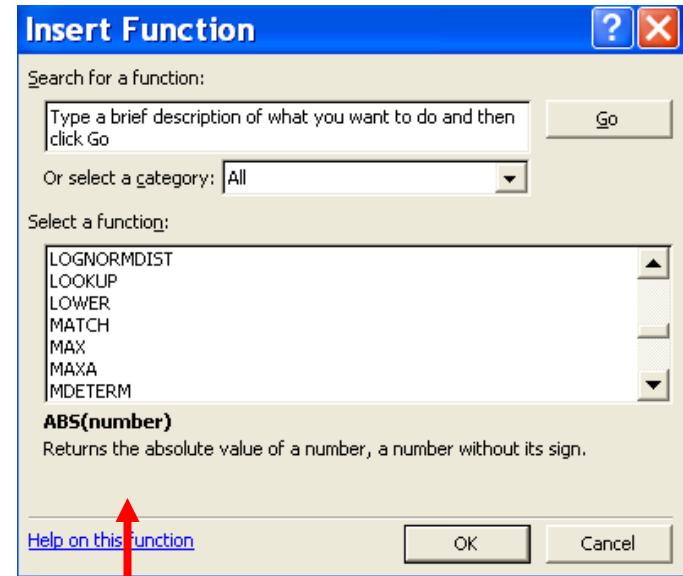
- Home-> Conditional Formatting



- Pivot Tables
- Formulas

Session 4. Some useful Excel functions

- Logic
 - AND
 - OR
 - IF (THEN)
 - NOT
- Mathematical/Statistical
 - AVERAGE
 - COUNT
 - COUNTA
 - COUNTBLANK
 - COUNTIF
 - DSUM
 - SUMIF
 - RANK
- Information
 - TRIM
 - CLEAN
 - VLOOKUP
 - CONCATENATE
 - LEFT
 - RIGHT
 - MID
 - LEN
 - FIND
 - PROPER
 - LOWER
 - UPPER
 - ISBLANK
 - ISTEXT
 - YEARFRAC
 - TODAY



Use the help in Excel which gives guidance on the use of each formula



Session 5: **ANALYSIS/DISSEMINATION**



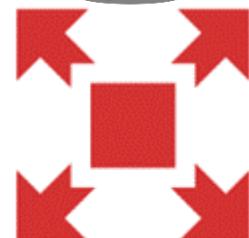
Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter



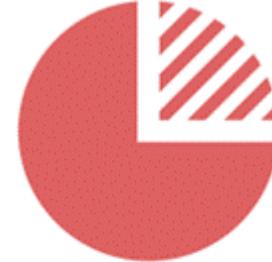
Information Management is a cycle of six stages



Processing



Dissemination



Analysis



Design



Combining



Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

Session 5: ANALYSIS

- Analysis aims to **proceed clean raw data** into indicators to prove or revoke a hypothesis.
- Analysis involves a level of expertise for the interpretation of the data.
- Decide the final **confidence level** (usually the same as sampling)
- Keep in mind what you are looking for or want to demonstrate.
- Decide to use , or not, **proxy indicator**.

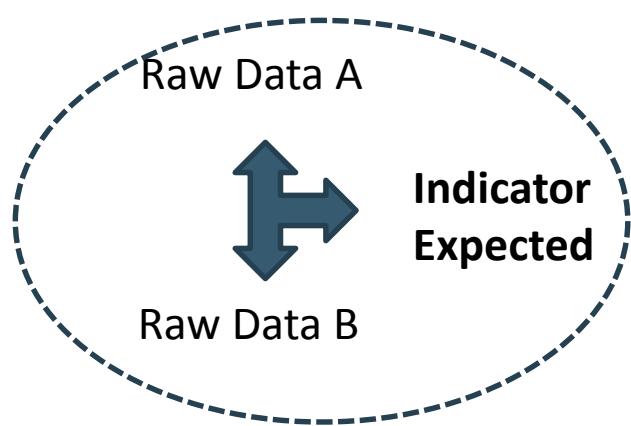
Even better plan a list of what you expect and don't worry we will always find interesting other pieces of information along road.



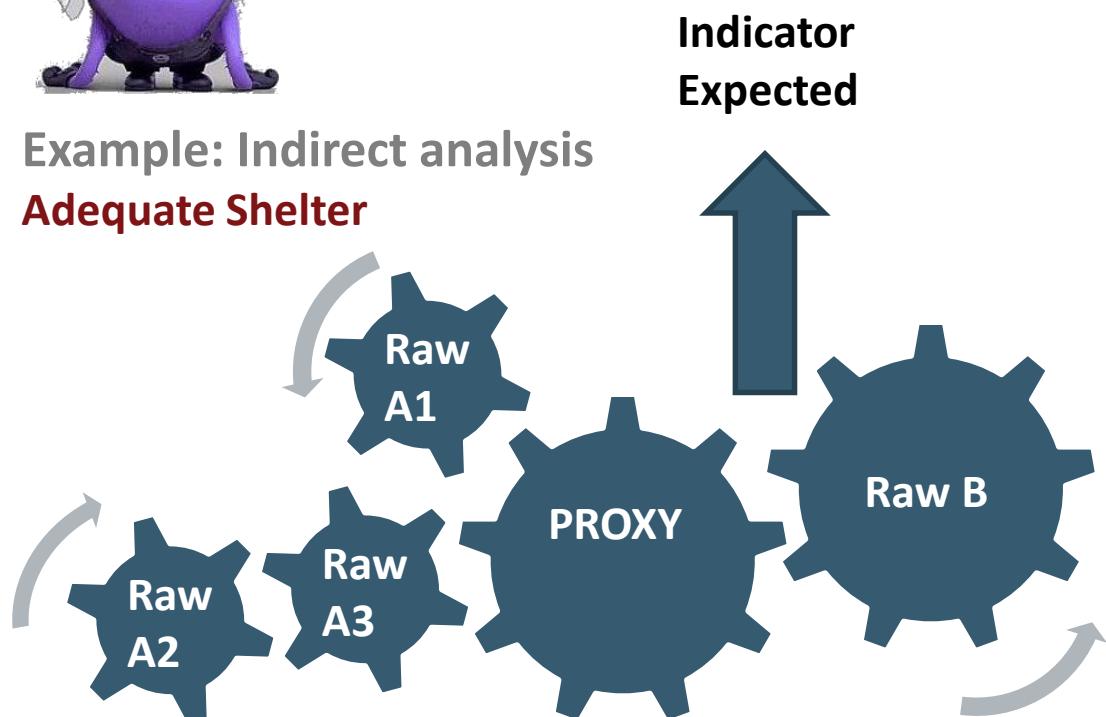
Session 5: Proxy Indicator



Example: Direct analysis
of Individual per Household



Example: Indirect analysis
Adequate Shelter



- What is a proxy indicator? (intermediate calculation)
- Be careful to not use too early and too many proxy indicators, it will jeopardize the confidence level.

Session 5 Tool for the ANALYSIS

Processing an indicator could be an average (cost for rent in a city), a range (minimum/maximum), a sum (number of IDP's per oblast).

To get them, we mainly use excel with filters and subtotal formula (useful for monitoring master lists on regular manner) or create a pivot table.

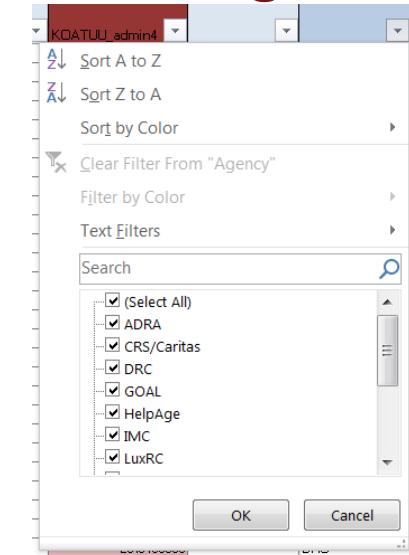


Session 5 Working on master list 1

- REMINDER Be sure that data is organized on sequential and consecutive manner, one data per cell, avoiding cell merging
- Set some automatic filters

Screenshot of Microsoft Excel showing the 'Winterisation Planning Matrix_2015-12-12_publish - Excel' spreadsheet. The table is titled 'Location' and includes columns for GCA/NGCA, REGION, KOATUU_admin1, DISTRICT/CITY, KOATUU_admin2, CITY/VILLAGE, KOATUU_admin3, SETTLEMENT, Settlement (if not in list), Agency, Multi-functional Cash, and Planned regi. Row 8 shows data for GCA Donetskia, while rows 9-16 show data for NGCA Donetskia. A red circle highlights the 'Filter' button in the Data tab ribbon, and another red circle highlights the value '1413870300 (Garnyak)' in the Settlement column of row 8.

GCA/NGCA	REGION	KOATUU_admin1	DISTRICT/CITY	KOATUU_admin2	CITY/VILLAGE	KOATUU_admin3	SETTLEMENT	Settlement (if not in list)	Agency	Multi-functional Cash	Planned regi
8 GCA	Donetskia	1400000000	Selvidyska	1413800000	Hirnytska	1413870300	Hirnyk	1413870300 (Garnyak)	ADRA		5,610
9 NGCA	Donetskia	1400000000	Donetska	1410100000		#N/A	#N/A	#N/A	DRC		
10 NGCA	Donetskia	1400000000	Khartyszka	1415000000		#N/A	#N/A	#N/A	DRC		
11 GCA	Donetskia	1400000000	Marinskyi	1423300000	Krasnohorivska	1423310400	Krasnohorivka	1423310400	DRC		
12 GCA	Donetskia	1400000000	Marinskyi	1423300000	Marinka	1423310100	Marinka	1423310100	DRC		
13 GCA	Donetskia	1400000000	Mariupolska	1412300000	Mariupolska	1412300000	Mariupol	1412300000	DRC		
14 GCA	Donetskia	1400000000	Pershotravnevyi	1423900000	Manhushka	1423955100	Manhush	1423955100	DRC		
15 GCA	Donetskia	1400000000	Pershotravnevyi	1423900000	Urzufska	1423985500	Urzuf	1423985500	DRC		
16 GCA	Donetskia	1400000000	Pershotravnevyi	1423900000	Yalta	1423985500	Yalta	1423985500	DAR		



Session 5 Working on master list 2

- Insert on title the formula =subtotal(109,\$A\$1:\$A\$100) for summing visible cell
- Use formula =subtotal(9,\$A\$1:\$A\$100) for summing complete column range

Please note that \$ argument is used to fix a variable in excel facilitating the copy paste of a determined fixed range.

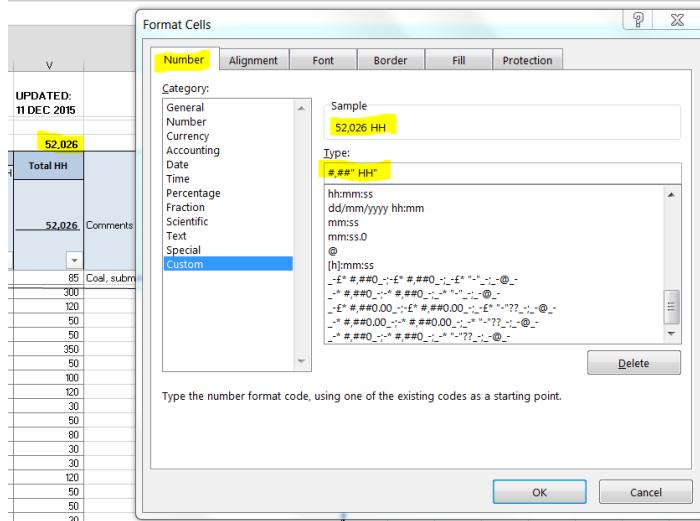
		k shelter activities, Donetsk & Luhansk GCA													UPDATED: 11 DEC 2015
		Planned regular activities			Planned winterisation activities									Total HH	
Location		Agency	HH	HH	HH	HH	HH	HH	HH	HH	HH	HH	HH	HH	
GCA / NGCA	REGION		Multi-functional Cash	NFI	Shelter repair	Winter cash	Winter NFI	Clothing	Coal/wood	Heater	Shelter insulation	Other			
GCA	Donetska	ADRA							85				85		
NGCA	Donetska	DRC		300									300		
NGCA	Donetska	DRC			120								120		
GCA	Donetska	DRC				50							50		
GCA	Donetska	DRC				50							50		
GCA	Donetska	DRC				350							350		
GCA	Donetska	DRC				50							50		



Session 5 Working on master list 3

- For better presentation use group function ...

Winterisation Planning Matrix_2015-12-12_publish - Excel												
Shelter Cluster Ukraine ShelterCluster.org Coordinating Humanitarian Shelter												
Winter shelter activities, Donetsk & Luhansk GCA												
			5,610	1,792	1,335	3,450	14,169	9,941	17,501	3,702	2,750	3,525
			Planned regular activities	Planned winterisation activities								
			HH	HH	HH	HH	HH	HH	HH	HH	HH	Total HH
GCA / NGCA	REGION	Agency	Multi-functional Cash	NFI	Shelter repair	Winter cash	Winter NFI	Clothing	Coal/wood	Heater	Shelter insulation	Other
85	85	85										52,026
85	85	85										Comments
85	85	85										85
85	85	85										Coal, submitted - expecting results - not secured funding yet



... & cell formatting

#,## meaning a figures formatted with separator like 1,000

" " is to included a text label (visible but not counted as cell with text)



Session 5 Working with Pivot Table 1

- REMINDER Be sure that data is organized on sequential and consecutive manner, one data per cell, avoiding cell merging, each column of the future pivot table shall get an unique title.
- Pivot table is a two steps creation



2) Click on pivot table

1) SELECT ONLY the part with column header

The screenshot shows an Excel spreadsheet titled "Winterisation Planning Matrix_2015-12-12_publish - Excel". A yellow box highlights the "PivotTable Recommended" icon in the "INSERT" tab of the ribbon. A red circle highlights the "PivotTable Recommended" icon. A yellow box highlights the "Planned regular activities" section of the PivotTable report. The PivotTable dialog box is open, showing the "Select a table or range" field set to "mastertable!\$A\$7:\$V\$35" and the "New Worksheet" option selected under "Choose where you want the PivotTable report to be placed".

Category	Planned regular activities			Planned winterisation activities		
	HH	HH	HH	HH	HH	HH
Emergency	Multi-functional Cash	NFI	Shelter repair	Winter cash	Winter NFI	Clothing
				300	120	50
				50	50	350
				100	120	30
				50	80	50
				30	30	30
				85		

Session 5 Working with Pivot Table 2

- Pivot table is a way to organized data

Winterisation Planning Matrix_2015-12-12_publish - Excel

PIVOTTABLE TOOLS ANALYZE DESIGN

Name of selected column to be 'drag & drop'

Place for main filters ex: working only on certain oblast

Row &/or Column to organized categories

Value of what will be represented example counting in #of HH or summing # of beneficiaries

Clicking inside pivot table range will activate it

PivotTable Fields

Choose fields to add to report:

- GCA / NGCA
- REGION
- KOATUU_admin1
- DISTRICT/CITY
- KOATUU_admin2
- CITY/VILLAGE
- KOATUU_admin3
- SETTLEMENT
- KOATUU_admin4

Drag fields between areas below:

FILTERS	COLUMNS
ROWS	Σ VALUES

Defer Layout Update UPDATE

Sheet1 mastertable +

READY 133%



Session 5 Working with Pivot Table 3

Note:

- Small arrows allow specific selection
- Sum precise the function of the value



Screenshot of Microsoft Excel showing a PivotTable setup. The PivotTable Fields dialog box is open, with 'REGION' selected under 'Row Labels' and 'Sum of Total HH' selected under 'VALUES'. Two Value Field Settings dialog boxes are overlaid on the main window, both showing 'Source Name: Total HH' and 'Custom Name: Sum of Total HH'. The left dialog shows 'Summarize Values By' set to 'Sum' and the right dialog shows it set to '% of Grand Total'. Both dialogs have 'Show Values As' highlighted.

By clicking on the Value, we can change the function, sum, average or the format (figures, %, % per column)



Session 5 Working with Pivot Table 4

- You can change until you reach the output expected

PivotTable Fields

Choose fields to add to report:

- GCA / NGCA
- REGION
- KOATUU_admin1
- DISTRICT/CITY
- CITY/VILLAGE
- KOATUU_admin2
- KOATUU_admin3
- SETTLEMENT
- KOATUU_admin4
- Settlement (if not in list)

Drag fields between areas below:

FILTERS

REGION	Agency
GCA / NGCA	

COLUMNS

ROWS

VALUES

Sum of Total HH

PivotTable Fields

Choose fields to add to report:

- GCA / NGCA
- REGION

Tip If you need to refresh your data set (source)

PivotTable Fields

Choose fields to add to report:

- GCA / NGCA
- REGION
- KOATUU_admin1
- DISTRICT/CITY

Drag fields between areas below:

FILTERS

REGION	Agency
GCA / NGCA	

COLUMNS

ROWS

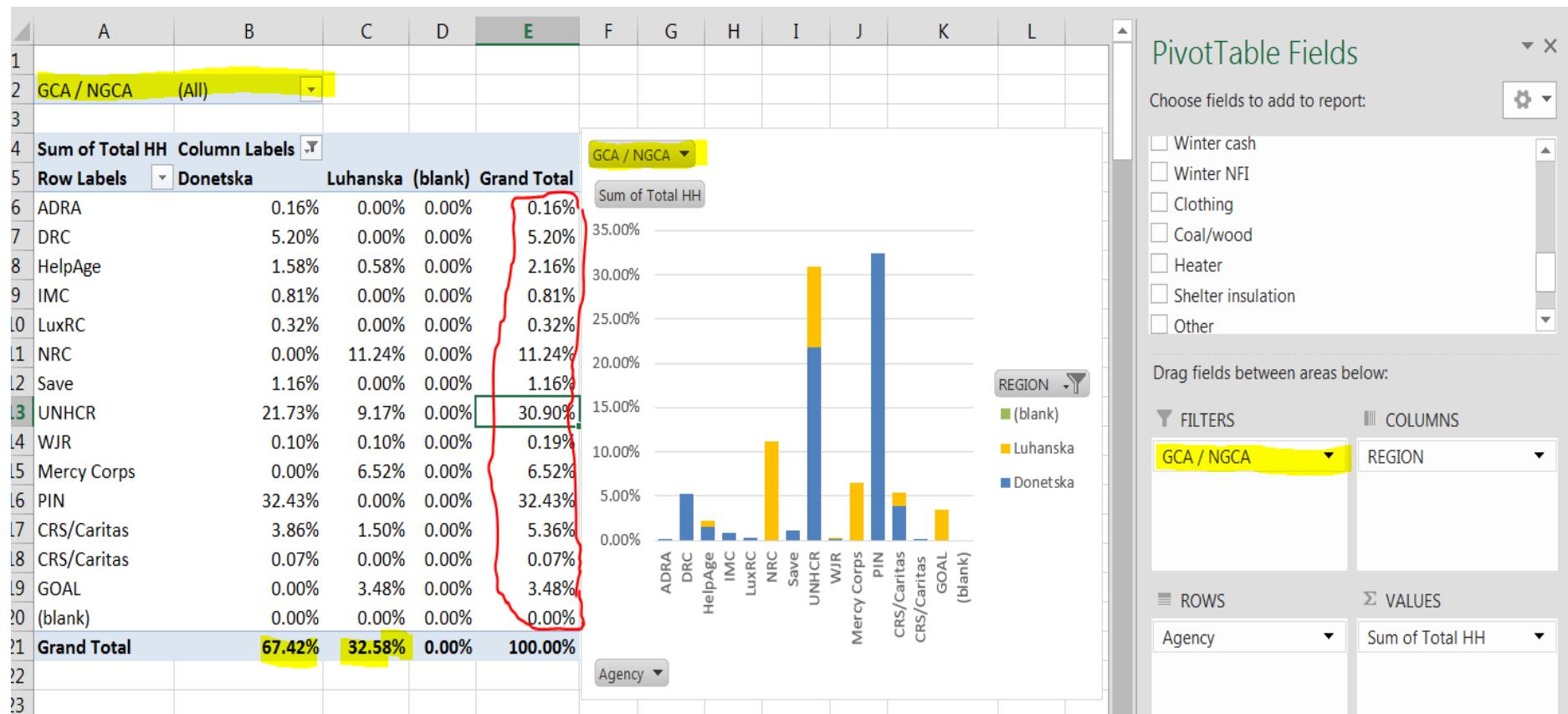
VALUES

Sum of Total HH



Session 5 Working with Pivot Table 5

- You can link a graph to visualize your main findings

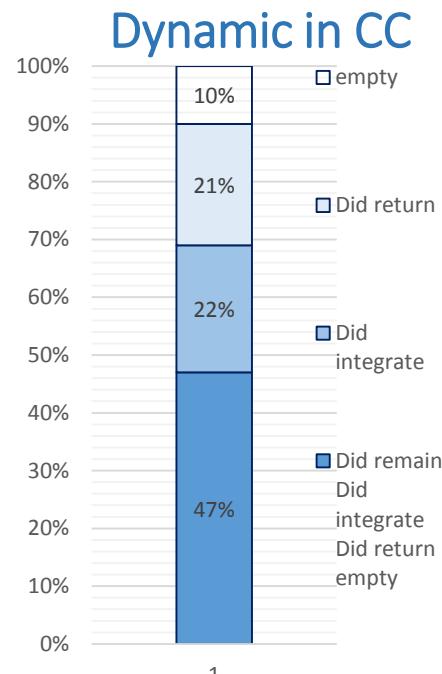


Session 5 Tool for Dissemination

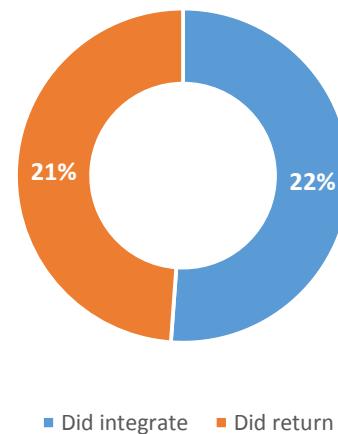
How to visualize and represent your indicators might significantly emphasize one aspect or the others



A9	B
A	B
1	
2	Did remain
3	47%
4	Did integrate
5	22%
6	Did return
7	21%
8	empty
9	10%



47% Occupancy Ratio
Intentions





Afghanistan Shelter Cluster
ShelterCluster.org
Coordinating Humanitarian Shelter

session 6- PROGRAMMING

My 5 first question on Kobo

In 30 minutes set up the 5 first question from the exercise of the morning.

The exercise is on individual computer but you can still keep your group in order to help each other.

Take the first and most simple one

Your objective is to get a first run



session 7 - PROGRAMMING

Another 10 please....

On the 5 first questions, please select 10 others but covering field as date, photo, restrictions.

This step will last for 45 minutes...



Presentation from participants

1 example will be shown per group

- State which difficulty you met
- Give your feed back and specific questions



session – Wrap up

- Identify what did you learn today (referring to the self evaluation form and use another color for ticking boxes)
- Fill up the feed back form
- Have a good rest