# Information Management in Emergencies

TIPS TO HELP MANAGE DATA TO MAKE BETTER DECISIONS FASTER

# Contents

Collecting and making sense of data in an emergency is not easy. In some cases, there is an overload of information, in other cases there is a complete lack of it or critical gaps in our understanding. Resource constraints, accessibility, and poor data management practices, among other challenges, hinder effective decision-making about response interventions. This guide is intended to support CRS teams responding to crises to set up effective IM/data management systems to support programming and coordination.

## WHAT IS INFORMATION MANAGEMENT? WHY DOES IT MATTER IN EMERGENCY SETTINGS?

**"Effective IM allows for better decisions to be made faster."[1]** Information management (IM) is the process of controlling how data is gathered, validated, cleaned, analyzed, consolidated, transformed into useful information, and ultimately shared. IM is an important element of effective humanitarian response. Strong IM supports the capacity of different actors to make decisions through timely data collection, processing, interpretation and sharing. It facilitates coordination and complementary action between different humanitarian actors and stakeholders. Additionally, data collected in the early stages of an emergency supports response, recovery, and planning for longer-term development and disaster preparedness.

*IM goes beyond using databases or ICT4D tools. It emphasizes collaboration to capture, share and apply knowledge.*

The purpose of IM is to ensure that "relevant information related to a humanitarian emergency is provided to the right person at the right time in a usable form to facilitate situational understanding and decision-making."[2] IM has been called a key piece of the "engine" that drives humanitarian response. IM helps us understand where we want to go, how to get there, when we have arrived, and what others are doing. When IM is functioning well, it is a seamless part of the response; when it is not functioning well, it can cause delays, bad decisions, and lead to harm to the people we serve.

## COMMON IM CHALLENGES IN EMERGENCY SETTINGS

The rest of this guidance will focus on best practices in information management to support overcoming common challenges faced in emergency contexts. There are also a number of existing complementary resources to help avoid or overcome these challenges.

The rest of this guidance will focus on best practices in information management to support overcoming common challenges faced in emergency contexts. There are also a number of existing complementary resources to help avoid or overcome these challenges.

---

[1] ALNAP Briefing Paper: How to Make the Most of IM in Coordination. 2016. Link.
[2] IASC. Operational Guidance on Information Management. LInk

## Common IM Challenges and Helpful Resources

| Challenge | Existing Resource | Tips in this Guide |
|---|---|---|
| **Influx of New Staff / Staff Turnover** Without effective IM systems, it is a challenge to get new staff up to speed, or to retain or easily find knowledge when old staff leave. | • CRS Proposal Handover Guidance | • Annex: CRS Teams Folder Hierarchy Suggestion |
| **Lack of Reliable, Credible or Up-to-date Secondary Information** In many cases, information available on affected populations is not available or is out of date, only available for certain areas, or difficult to access. This creates pressure to collect new data quickly. | • List of OCHA Common Operational Databases (CODs)<br>• Reliefweb<br>• ALNAP<br>• Humanitarian Data Exchange | • Annex: Key Datasets for Emergencies |
| **Data Quality Issues** Data entry errors can hinder data analysis and interpretation and lead to bad decisions. Non-standardized formats make it time-consuming or even impossible to consolidate information to get the bigger pictures. | • SMILER+ Session 7 Data Collection Forms<br>• SMILER+ Session 8 Reporting Formats<br>• CRS CommCare Application Templates | • Best Practices in Digital Survey Design<br>• Excel Tips: Working with Spreadsheets |
| **Short-time frame to collect and analyze data** Programming teams often need information by yesterday. Time pressure and a lack of agreed procedures and tools make it easier to make mistakes, sacrificing quality for speed. | • Good Enough Guide<br>• CRS MEAL In Emergencies<br>• SMILER+ Session 5 Data Flow Map<br>• CRS ICT4E Checklist | • Database Basics<br>• Data Management Options at CRS |
| **Short-notice requests to share information with donors, clusters, etc.** Requests for information come from all sides during an emergency. It is important to have an agreed process for prioritizing and handling these requests in place so that reporting does not hinder operations**.** | • SMILER+ Stakeholder Communication Plan<br>• SMILER+ Session 8 Reporting Format<br>• Project DIP and MEAL Calendar | • Organizing Files Tips<br>• Sharing Information Tips |
| **Information Sensitivity, Data Protection and Security** Information on needs and locations of affected populations or on planned interventions is often sensitive. Making as making this information public can pose security risks to humanitarian actors or expose affected populations to further harm. | • CRS Responsible Data<br>• Signal Code<br>• CRS Data Privacy Impact Assessment<br>• Data Sharing Agreement Template | • A Note about Data Security |

## DATA MANAGEMENT IN EMERGENCIES

Data management is a foundational aspect of effective IM. Data management refers to the process of collecting, keeping, and using data securely, efficiently, and effectively. There are several principles for data management. It is important to consider the balance between each when designing your data management/IM system.

**Simplicity:** The more complex a solution is, the more places it can fail. One way to reduce the complexity of your information management system is to collect less data in the first place. Consider 'need-to-know' vs. 'nice-to-know' information by ensuring all data collected is linked to a specific purpose. This can be an indicator, a necessary disaggregation, a point that leads to an expected programmatic decision, or another clearly defined use.
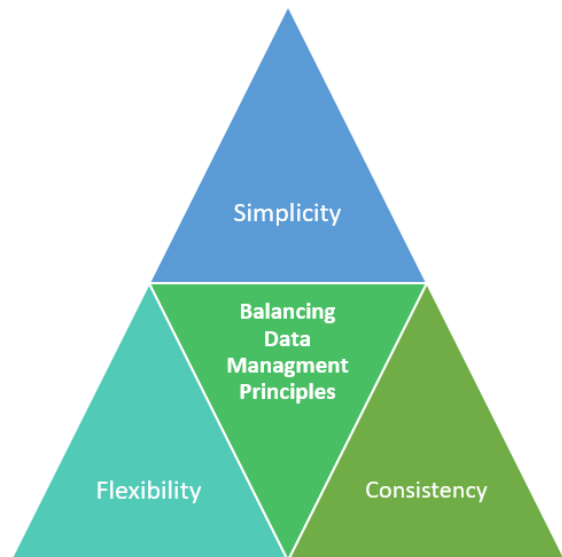
**Flexibility**: Quick solutions may get a system up and running quickly, but as a response grows and evolves the solution may not be flexible enough to incorporate new elements. Simply adding a new column to a registration table to account for distributed items might make counting the distributed items quick to add up in that moment, but if more distributions occur with the same individuals or different kinds of questions need to be asked about who received what when, this system will become rigid and will need to be recreated. For example, creating a new tab in an Excel workbook for each district that receives a distribution is a quick way to organize a small-scale distribution but won't scale well if the distribution increases to many districts.

**Consistency:**: An important "rule" of data management is to be consistent! If you start to organize and store your data, files, etc. in one way, continue to use that system so that you don't have to spend time harmonizing later. Some formatting conventions are arbitrary and only matter to the extent that they are applied consistently, similar to choosing on which side of the street to drive.

**Quick Consistency Tips:**

- Use consistent variable names
- Use consistent codes for categorical values
- Use a consistent code for missing values
- Use consistent file names
- Use consistent report formats and templates
- Use consistent format for dates

In addition to the principles mentioned above, it's a good idea to keep in mind broader IM principles when designing your data management/IM system. OCHA has defined a set of **Principles for Humanitarian Information Management and Exchange**: reliability, verifiability, interoperability, accessibility, sustainability, timeliness,

relevance, inclusiveness, accountability, impartiality, humanity, reciprocity and confidentiality. These are closely aligned with CRS's Responsible Data Values and Principles. More information about OCHA IM principles can be found in the Annex.

## PRACTICAL TIPS FOR SETTING UP A DATA MANAGEMENT SYSTEM

A data management system is a set of solutions for organizing, analyzing, storing, archiving, and sharing data or information. To get starting with your data management system, consider the following:

### Key Steps and Questions To Consider

| | |
|---|---|
| **Planning** | • Does our project have a data flow map from our SMILER+ workshop? <br> • Does our project need a more detailed data management plan? <br> • Does our project require a data privacy impact assessment? Is it relevant even if not required? |
| **Building and documenting the system** | • How can we incorporate ICT4D (like CommCare) appropriately within our project? <br> • How will we use Teams or another system for file sharing and management? <br> • How can we manage partner's access to this system? |
| **Organizing files** | • What is an easy-to-understand way to organize the files we have? <br> • What naming convention do we want to use? |
| **Collecting and Analyzing data** | • Are staff familiar with CommCare, PowerBI, Excel, etc.? <br> • How can we support staff to build data management skills? <br> • How can we create a system that promotes data quality? |
| **Storing data and files** | • Who needs access to this information? When? How often? <br> • How long do we need to store this information? |
| **Sharing information** | • Who do we need to share information with? <br> • Do we have a stakeholders analysis from the SMILER workshop? <br> • How will we share information with different stakeholders? <br> • Would a specific data-sharing agreement/MOU with partners (prime and/or subrecipients) be helpful to clarify ownership, rights and responsibilities? |

Many of the answers to these questions will come from the SMILER+ workshop. See especially: **SMILER+ Session 3:** Stakeholder analysis and communication; **SMILER+ Session 5:** Data flow maps; **SMILER+ Session 7:** Data Collection Forms; and **SMILER+ Session 8:** Reporting Forms.

## PLANNING

IM becomes more efficient when it is known beforehand what information needs to be collected, and how it needs to be analyzed and packaged (visualized, reporting template, etc.). To plan for effective information management, take a look at your stakeholder analysis and data flow map from the SMILER+ workshop, and consider developing a RACI matrix to identify who will be responsible for the following roles:

- **Data collection**: enumerators using specific tools
- **Data storage**: managing databases or repositories
- **Data analysis**: quantitative, qualitative, GIS data, use of PBI
- **Data use**: planning and leading debriefs, reflections, meetings, and learning events, participating in coordination meetings, etc.
- **Reporting and Communication**: Writing reports, developing visualizations, etc.

---

**Planning Tips for IM in Emergencies**

☐ During the earliest phase of an emergency, information may already be available or partially available in different formats and, in most cases, time should not be wasted in re-collecting or repackaging data to fit a certain template or system. (This applies to initial sitreps and early needs assessment).

☐ After the initial phase, the specific formats, templates or information sharing products are helpful to facilitate ongoing coordination.

☐ As the response continues and begins to focus on recovery, it is important to re-evaluated standardized processes and templates to ensure that these products are serving current information needs of different stakeholders. Be sure that all stakeholders in the response (CRS staff and partners) agree to and understand how to use reporting templates.

See CRS MEAL in Emergencies or EFOM for more tips.

---

## BUILDING THE SYSTEM

*Databases are a key component of data management systems, but not the only important piece!*

Have you ever used CommCare? What about Access? Even if you are not familiar with these platforms, you have likely used Excel at some point during your work with CRS to serve as a database or structure for organizing data in multiple tables. It's important to note that a data management system is more than just a database. However, databases are an important part of the system. More information about commonly used databases can be found below.

## DATABASE BASICS

A **database** is a structure that stores organized information. A database contains one of more tables of information. The rows in the table are called **records** and the columns are called **fields** or **attributes**. While in Excel spreadsheet we might talk about rows and columns, when talking about a table in a database it is more common to use the terms fields and records. A database that contains two or more related tables is called a **relational database.** This means that we can identify and access one piece of data in relation to another piece of data in the database**.** The benefit of using a relational database compared to using multiple spreadsheets is that improves efficiency in processing information, improves accuracy by reducing the duplication of data entry, and reduces the time required to maintain the data.

Ideally, each piece of information in a database appears just once, rather than being repeated multiple times. Take a look at the examples below which represents a list of distributions received by beneficiaries

| Table A | | | | | |
|---|---|---|---|---|---|
| **Beneficiary ID Number** | Name | Location | Phone | Distribution Type | Date of Distribution |
| **000000000** | Ana | Luanda | 123-4567 | NFI Kit | 2020-09-02 |
| **000000000** | Ana | Luanda | 123-4567 | WASH Kit | 2020-09-02 |
| **000000000** | Ana | Luanda | 123-4567 | Shelter Kit | 2020-09-09 |

In Table A, you can clearly see that Ana received three types of kits. However, each time we collect the location, phone number, and name, it increases the opportunity for errors during collection or analysis. If one of Ana's phone numbers is different, does that mean she changed numbers, or that it's a mistake? Does that mean there are two people named Ana? Also, storing this data in multiple cells takes up more space, making the database slower to open and results harder to analyze.

### In a CommCare Form Export, Table A would might look more like this:

| Table A (CommCare Form Export) | | | | | | | |
|---|---|---|---|---|---|---|---|
| number | formid | form.case.@case_id | form.case_name | form.location | form.phone | form.distribution_type | form.date |
| 0 | acae8e2d-0b85-43bd-ab06-1af820ba4544 | af7af95e-c778-4a6d-9575-45d1fab0856f | Ana | Luanda | 1234567 | wash_kit | 2020-09-02 |
| 1 | e79f1239-2d81-440d-86de-e2dbdfb973fb | af7af95e-c778-4a6d-9575-45d1fab0856f | Ana | Luanda | 1234567 | nfi_kit | 2020-09-02 |
| 2 | 338e30a6-3223-4e51-bfdf-da7ae5f69d08 | af7af95e-c778-4a6d-9575-45d1fab0856f | Ana | Luanda | 1234567 | shelter_kit | 2020-09-09 |

A better solution is to store the data in two related, tables:

| Table B | | | |
|---|---|---|---|
| **Beneficiary ID Number** | Name | Location | Phone |
| **000000000** | Ana | Luanda | 123-4567 |
| **000000001** | Bob | Luanda | 345-6789 |
| **000000002** | Belinda | Luanda | 567-8912 |

| Table C | | |
|---|---|---|
| **Beneficiary ID Number** | Distribution Type | Date of Distribution |
| **000000000** | NFI Kit | 2020-07-01 |
| **000000000** | WASH Kit | 2020-07-01 |
| **000000000** | Shelter Kit | 2020-08-01 |

In Table B and Table C, we have the same information as in Table A, but presented more efficiently. By looking at Table C, you can see that beneficiary 000000000 (Ana) has received three kits. The other beneficiaries, Bob and Belinda, have received none. Ana's beneficiary ID number links her personal information to the information about the distribution. The Beneficiary ID comes from Table B – and is a unique identifier for each person. We call this the **primary key -** a field whose values are unique in the table, meaning that they appear only once. In Table C, we see that the beneficiary ID number for Ana is the **foreign key**, meaning that it is "imported" from Table B and can be repeated.

In CommCare, Table B and C would look more like this:

| Table B (Commcare Case Export) | | | | |
|---|---|---|---|---|
| number | caseid | name | location | phone |
| 0 | af7af95e-c778-4a6d-9575-45d1fab0856f | Ana | Luanda | 1234567 |
| 1 | a3bb726b-5b36-4c71-af31-b62a1cef92a3 | Belinda | Luanda | 3456789 |
| 2 | 7fd1b77a-1647-449e-84ea-f2d8d5edb490 | Bob | Luanda | 2345678 |

| Table C (CommCare Form Export) | | |
|---|---|---|
| number | form.case.@case_id | form.distribution_type | form.date |
| 0 | af7af95e-c778-4a6d-9575-45d1fab0856f | wash_kit | 2020-09-02 |
| 1 | af7af95e-c778-4a6d-9575-45d1fab0856f | nfi_kit | 2020-09-02 |
| 2 | af7af95e-c778-4a6d-9575-45d1fab0856f | shelter_kit | 2020-09-09 |

It's also helpful to understand the relationships between tables in a database in terms of **parent** and **child** tables. In the example above, Table B is the parent table because it is independent of Table C, meaning that it can contain a record without needing a corresponding record in Table C. For example, it is possible to collect beneficiary registration information without distributing the kit to all beneficiaries. Here, we can see that we have Ana, Bob and Belinda registered, but only Ana has received kits. We can create Table A without creating Table B. However, we could not create Table B without Table A because we wouldn't have the beneficiary ID.

The last helpful thing to understand is **cardinality**, or how many records in one table can be related to records in another table in a relational database. There are a few types of relationships possible.

- **One-to-one** relationships:
- **One-to-many** or **many-to-one** relationships (e.g. 1-to-2; 1-to-5; 1-to-∞)
- **Many-to-many** relationships: multiple records in a table are associated with multiple records in another table

*One-to-one relationships are when one record in a table is associated with one and only one record in another table*

An example of **one-to-one** relationships can be seen below. In this example about from beneficiary registration for an education project, Table D contains only the beneficiary (student) ID number and name. In a real situation, Table E might contain other demographic data (sex, age, disability status) and vulnerability information. Table E contains the beneficiary ID number and government-issued identification number. Each beneficiary can only have one beneficiary ID and one government ID.

| Table D | |
|---|---|
| **Beneficiary ID Number** | Government ID |
| **000000000** | 54324885322110 |
| **000000001** | 56432468542210 |
| **000000002** | 98567532589325 |

| Table E | |
|---|---|
| **Beneficiary ID Number** | Name |
| **000000000** | Mariam |
| **000000001** | Ahmed |
| **000000002** | Rita |

In CommCare, Table D and Table E might look something like this:

| Table D (CommCare Form Export) | | |
|---|---|---|
| number | form.gov_id | form.case.@case_id |
| 0 | 5643246854221 | 766d98f2-2943-4e31-81c2-5121213b2c21 |
| 1 | 98567532589325 | 493c650c-f255-4672-8b5e-a3d007c0605a |
| 2 | 54324885322110 | 2c57ebfe-5bb0-449b-bfe0-0cddc83854e0 |

| Table E (CommCare Form Export) | | |
|---|---|---|
| number | form.name | form.case.@case_id |
| 0 | Rita | 493c650c-f255-4672-8b5e-a3d007c0605a |
| 1 | Mariam | 2c57ebfe-5bb0-449b-bfe0-0cddc83854e0 |
| 2 | Ahmed | 766d98f2-2943-4e31-81c2-5121213b2c21 |

*One-to-many or many-to-one relationships are hen one record in a table is associated with one or more records in another table*

An example of **one-to-many or many-to-one** relationships is below. Here, we can see that each school has multiple children (one-to-many) and multiple children are enrolled in exactly one school (many-to-one), and that each record from Table F is related to more than one record in Table G.

| Table F | |
|---|---|
| **Student ID Number** | Name |
| **000000000** | Mariam |
| **000000001** | Ahmed |
| **000000002** | Rita |
| **00000003** | Bassem |

| Table G | | |
|---|---|---|
| **Student ID Number** | School | Date of Enrollment |
| | | |
| **000000000** | Amman Secondary | 2020-09-01 |
| **000000001** | Amman Secondary | 2020-09-01 |
| **000000002** | Amman Secondary | 2020-09-01 |
| **000000003** | Amman Secondary | 2020-09-01 |

In CommCare, these would look something like:

| Table F (CommCare Form Export) | | |
|---|---|---|
| number | form.name | form.case.@case_id |
| 0 | Rita | 493c650c-f255-4672-8b5e-a3d007c0605a |
| 1 | Mariam | 2c57ebfe-5bb0-449b-bfe0-0cddc83854e0 |
| 2 | Ahmed | 766d98f2-2943-4e31-81c2-5121213b2c21 |
| 3 | Bassem | bd9647ea-b506-4484-8b12-cc6330c99189 |

| Table G (CommCare Form Export) | | | |
|---|---|---|---|
| number | form.case.@case_id | form.school | form.date_enroll |
| 0 | 766d98f2-2943-4e31-81c2-5121213b2c21 | amman_secondary_school | 2020-09-01 |
| 1 | 2c57ebfe-5bb0-449b-bfe0-0cddc83854e0 | amman_secondary_school | 2020-09-01 |
| 2 | 493c650c-f255-4672-8b5e-a3d007c0605a | amman_secondary_school | 2020-09-01 |
| 3 | bd9647ea-b506-4484-8b12-cc6330c99189 | amman_secondary_school | 2020-09-01 |

INFORMATION MANAGEMENT IN EMERGENCIES

*Many-to-many relationships are when multiple records in a table are associated with multiple records in another table*

An example of **many-to-many relationships** can also be found below in another education example. In this case, we can see that each child is enrolled in more than one class, and each class contains multiple children.

| Table H | | |
|---|---|---|
| **Student ID Number** | | Name |
| **000000000** | | Mariam |
| **000000001** | | Ahmed |
| **000000002** | | Rita |
| **000000003** | | Bassem |

| Table I | |
|---|---|
| **Student ID Number** | Class |
| **000000000** | Math |
| **000000000** | English |
| **000000000** | Chemistry |
| **000000000** | Art |
| **000000001** | Math |
| **000000001** | English |
| **000000001** | Chemistry |
| **000000001** | Art |
| **000000002** | Math |
| **000000002** | English |
| **000000002** | Chemistry |
| **000000002** | Art |
| **000000003** | Math |
| **000000003** | English |
| **000000003** | Chemistry |
| **000000003** | Art |

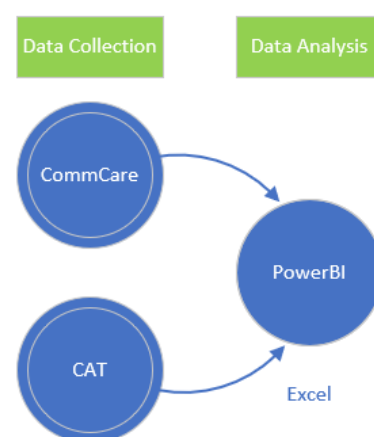**A note on many to many relationships in Power BI**

Implementing many-to-many relationships in CommCare and Power BI is a bit trickier. In Power BI, you need to use a workaround called a **join table**, which combines columns from two or more tables and helps filter the relationship between the different sets of data. More information about how to implement many-to-many relationships in Power BI can be found here.

## DATA MANAGEMENT OPTIONS AT CRS

An ideal dataflow in an emergency may look something like this:: Data is collected using CommCare or CAT (RR Collect), then linked to Power BI for analysis and visualization. Data can also be exported from CommCare or CAT to Excel for simpler analysis and visualizations.

A helpful place to start is with the CRS ICT4E Checklist, which can guide your decision-making on: Go/No-go for ICT4E, Human Resources and Capacity, Hardware and Software choices, and related MEAL processes.

The tips in this guidance focus on the data management platforms most commonly used in emergency settings.[3] Many of these platforms combine the functions of a database, data analysis, and visualization and storage, but some are stronger at some functions than others.

| PLATFORM PRIMARY USE | DATA ANALYSIS | VISUALIZATION | STORAGE AND ARCHIVING |
|---|---|---|---|
| **CommCare** Data Collection | **Yes-Limited** Export to Excel or connect to PBI for deeper analysis. | **Yes-Limited** Connect to PBI to visualize data. | **Yes** Create backup of de-identified data and store in SharePoint or Gateway.[4] |
| **Power BI (PBI)** Data Analysis and Visualization | **Yes** | **Yes** PBI is a great option for creating dashboards and interactive reports. | **No** |
| **Excel** Data Analysis | **Yes** | **Yes** Pivot Charts, other charts are easy to make in Excel. | **Yes** |
| **CAT** RR Collect for markets-based/cash programming; monitoring on CAT | **Yes-limited** Reports are predefined in CAT. For further | **Yes-limited** Visualizations are predefined in CAT. For further visualization, connect CAT to PBI. | **Yes** |

---

[3] Many development projects, or more complex and longer projects have custom database systems or use Access.

[4] See Compass Standard 16 (Accountable and Timely Close Out) Key Action 4 (Prepare and store all required close-out documentation and other key project documentation, records and data) for more guidance.

## ORGANIZING FILES

Some tips on organizing files during an emergency response are below. Take a look to help your team find what they need, when they need it!

### FILE STORAGE CONVENTIONS

It is helpful to create shared file storage space so that different staff can access key information and collaborate on key documents. CRS teams are increasingly adopting Microsoft Teams for this purpose.

You might consider using a standard file organization structure for Teams across your response/project. In the absence of an established convention, default to the following tiered structure:

| TIER | TYPE |
|------|------|
| 1 | Department (e.g., Programs, HR, Finance, etc.) |
| 2 | Discrete Projects (Project A, B, C) and Cross Project Elements (e.g., Assessments that are not project specific, databases that contain data from multiple projects [e.g., beneficiaries registered in overlapping projects], warehouse protocols that cover multiple projects) |
| 3 | Project specific activities (proposals, MEAL plans, financial projections for a specific project, |

A skeleton structure for file storage on Teams can be found in the annex. This includes empty folders already labeled and organized according to the tiers above.

### FILE NAMING CONVENTIONS

If you are going to use information, you have to be able to find it! This starts with creating a system to name and organize your files. Naming conventions are just that – conventions. The most important thing isn't whether you drive on the right side of the road or the left, but rather that if one person drives on the right, that every else follows suit. A good practice is to pick one naming convention and storage (cataloging) system and stick to it!

### ✓ DO: be descriptive and consistent

A good practice is to include a descriptive title and the date in the file name. You might consider the following descriptors:

- Project name or acronym
- Donor name or acronym
- Location/spatial coordinates
- Date or date range
- Type of data
- Version number of file
- Date of creation/last modification

For example, **OFDA_WASH_Workshop_Attendance_2020.07.31.** Here you can see clearly the donor (OFDA), sector (WASH), activity (Workshop), what the file is (Attendance List) and Date.

There are a few standard ways to list dates that will facilitate easy sorting and filtering:

- YYYYMMDD
- YYYY.MM.DD
- YYYY-MM-DD

While this style (YYYYMMDD) may be counterintuitive for non-Americans, the rationale behind this method is that if files are sorted by name, they will also sort chronologically.

**More Naming Convention Do's and Don'ts**

✓ **DO** use a sequential numbering system, including leading zeros ("_001, _002") so that files are ordered correctly (i.e. so that 1 is not followed by 11).

✕ **DON'T** use file names that are too long, since long names sometimes can't be opened by other software, uploaded to SharePoint, or sent by email.

✕ **DON'T** use spaces (i.e. "OFDA WASH Workshop Attendance 2020 08"). Instead use underscores (i.e. "Workshop_Attendance_2020.08).

✕ **DON'T** use special characters (! @ # $ % ^ () ; < > etc.

✕ **DON'T** use "_FINAL" as part of file names. Often, files are saved as "_FINAL" and then updated/edited afterwards, making it difficult to know which is really the most recent version. If an element of finality is needed, save a file with an absolute reference (e.g. File_20190530_as_submitted)

### FILE VERSIONING

Versioning refers to saving new copies of files when you make changes or updates so that you can go back and see the previous version if needed.

You may be tempted to save versions using a version number, e.g. "_v1" at the end of the file name. This is most appropriate if only one person is working on the file. In this case, it is important to be diligent about saving and updating the file name in this case. In general, **this is not recommended**, as it is easy to lose track of which version is the most recent (e.g. "is v4 the last version? Is there a v5?") or if "v1" was last modified more recently than "v2". The same goes for including "_FINAL" at the end of file names (see Naming Conventions Do's and Don'ts above).

If multiple people will be updating or using the file, it is recommended to add the date of most recent update to the file name, or to use SharePoint or another file storage system to manage the file library. You might also consider adding a table at the top to keep track of who has edited and what changes have been made. For example:

| Date | Reviewer / Editor | Changes Made |
|------|-------------------|--------------|
| **01/01/2020** | Sandra W. | Revised "Result Section" according to feedback from team. |
| **01/03/2020** | Paul Q. | Added "Responsible Data" section to report. |
| **01/04/2020** | Ella M. | Updated graphics and charts with new data from June |

It's also good to note that **SharePoint** libraries can be configured to automatically store previous versions. More information about how to enable the versioning feature on SharePoint can be found here. Version history can be displayed as a metadata column in SharePoint. **Teams** is an interface for a SharePoint library. **Google Drive[5]**, which is often used with or by CRS partners, automatically creates new versions when you edit files, and you can see the version history. More information here.

## COLLECTING AND ANALYZING DATA

The tips below are intended to help you create digital tools that will generate easy to use exports and facilitate analysis.

## ICT4D/COMMCARE TIPS: BEST PRACTICES IN DIGITAL SURVEY DESIGN

✓ **DO: use meaningful -- but concise -- codes for questions**

For variables (which will become spreadsheet column titles), it is a good idea to be descriptive, yet concise (and also, consistent!)

Many data collection platforms (like CommCare) will assign a default code to a question. For example, a question like "If you had a question or comment about the project, who would you contact?" might be given the default code if_you_had_a_question_or_comment_about_the_project_who_would_you_contact . And a follow up question like "Please specify (other)" with a code like please_specify_other

The first code is too long and is hard to read in a table of many questions. A better, more meaningful code would be "feedback_source". The second code is too vague, since it's unclear what question is being specified. A better code would refer to the initial question with something like "feedback_source_other".

✓ **DO: assign unique codes to beneficiaries**

f you are using CommCare, use a case list to register beneficiaries (and not a survey). This will automatically assign a unique value to each registered beneficiary or household (depending on the nature of the registration).

If you are not using CommCare, consider using a Universally Unique ID Generator (https://www.uuidgenerator.net/) to create unique IDs for your beneficiaries.

If you need to give the beneficiary an ID and need them to reference it at a future point (for example, if you are creating an ID card for them) and so you want a shorter (but less guaranteed to be unique) ID, consider using the following formula in Excel to generate a 5 digit alphanumeric ID that is likely to be unique for projects with fewer than 100,000 unique individuals =DEC2HEX(RANDBETWEEN(0,1048575))

✕ **DON'T: use multiple select (checkbox) questions)**

When using a mobile data collection platform, multiple select questions have little practical value. These types of questions serve mostly to save space on paper forms. They hide the fact that they make surveys longer than they appear, since instead of reflecting a single question, they represent a series of yes or no questions (one for each multiple choice option). Consider asking for a superlative (e.g. which of the following is your *most important* source of income?) or replacing the multiple choice question

---

with a series of yes/no questions for each option. Keep nice-to-know vs. need-to-know in mind when making this decision.

## EXCEL TIPS: WORKING WITH SPREADSHEETS

**✓   DO: use Excel Tables formatting when using Excel as a database**

Excel tables force column headers to be unique and make working with datasets more efficient (faster). It is much easier to filter data and identify errors when using tables. To use Excel tables, select a range of data and press ctrl+t. Alternatively, you can select Insert → Table

**×   DON'T: store data in columns that can be determined by other columns**

This can lead to problems organizing and analyzing data and ensuring accurate results. For example, do not create a table like the following

| Total Number of Males | Total Number of Females | Total Number of Participants |
|---|---|---|
| 20 | 30 | 40 |

Since the total number of participants can be determined by the previous two columns, data should not be entered into this column since it can produce a logical inconsistency. Data should only be stored in such a column by using a formula

| Total Number of Males | Total Number of Females | Total Number of Participants |
|---|---|---|
| 20 | 30 | =[Total Number of Males] + [Total Number of Females] |

**×   DON'T: repeat columns of a similar type**

For example, if you have a table of distributions, do not create the table to look like the table below:

| BeneficiaryID | Is_female | Distribution #1 | Distribution #2 | Distribution #3 |
|---|---|---|---|---|
| FF22A | 1 | 1 | 1 | 0 |
| E223D | 0 | 0 | 1 | 1 |
| 23FD2 | 1 | 1 | 0 | 0 |
| 23DD4 | 0 | 1 | 1 | 1 |

It may seem intuitive, but it will not make a database flexible. With a table like this it would be difficult to count how many women received distribution 1 and 2 but have not yet received distribution 3. Furthermore, it would be even more difficult to include any additional information about what was distributed and when without making the table even more unmanageable.

It is better to separate this into two tables. Refer back to the [database basics](#) section for more information.

| BeneficiaryID | Is_female |
|---|---|
| FF22A | 1 |
| E223D | 0 |
| 23FD2 | 1 |
| 23DD4 | 0 |

| BeneficiaryID | Distribution | Date | Item |
|---|---|---|---|
| FF22A | 1 | 1/5/20 | Tarp |
| FF22A | 2 | 8/5/20 | Hygiene Kit |
| E223D | 2 | 8/5/20 | Hygiene Kit |
| E223D | 3 | 15/5/20 | Cash |
| 23FD2 | 1 | 1/5/20 | Tarp |
| 23DD4 | 1 | 1/5/20 | Tarp |
| 23DD4 | 2 | 8/5/20 | Hygiene Kit |
| 23DD4 | 3 | 15/5/20 | Cash |

This makes it easier to quickly query the table even if the project grows over time. In terms of form design, this could mean either submitting one form for each item received (one form per one record) or it could mean manipulating the data set in analysis. In PowerBI, this would mean unpivoting the columns so that each record is repeated for the number of columns. The choice depends on where you have the capacity, in field staff/enumerators or in MEAL staff/office analysts.

### ✕  DON'T: use multiple sheets/tabs for multiple districts

It is not recommended to separate identical tables into multiple tables based on an aspect (e.g. a district) of that table. If the intention is to make it easier to only look at the data from one district, this is achieved more efficiently with a filter than with multiple tables.

### ✕  DON'T: merge cells

When possible avoid using multiple rows for column headers. This makes creating and organizing a table more difficult. Also, NEVER MERGE CELLS. This is not compatible with table formatting in excel and makes analysis more difficult. An easy work around, if you must show text spanning 1+ columns, is to use "center over selected" cell format ([see here for how to do this](#) in excel): Select the cells you want to center → Right click → Format Cells → Alignment → Horizontal → Center Across Selection.

### ✓  DO: avoid empty cells

When entering data manually from paper forms to excel, fill in all cells so that it is clear what data was collected or intentionally not collected, rather than simply missing. Use a consistent code to mark blanks.

### ✓  DO: keep raw data and analyzed data separate

It is always a good idea to keep analysis separate from raw data. If you are using PowerBI, this happens automatically as the analysis you do there does not affect the underlying data. If you are working in Excel, save a copy of the raw data to a new file, then do the analysis there.

When entering data manually from paper forms in Excel, use data validation rules to minimize data entry errors. There are three main types of data entry errors[6]:

| DATA ENTRY ERROR TYPE | WHY IT'S A PROBLEM | HOW TO AVOID IT | HOW TO FIX IT IF IT HAS ALREADY HAPPENED |
|---|---|---|---|
| **Impossible Value**<br><br>e.g.: Age of Head of HH= 180 | Data is not usable | Use data validation in the tool | Recollect and update with correct response, if possible |
| **Improbable value**<br><br>e.g.: Age of Head of HH = 6 | Technically possible, but unlikely | Use data validation in the tool | Use summary statistics and charts to detect outliers.<br><br>Confirm values with M&E staff and enumerators |
| **Inaccurate:**<br><br>e.g.: Age of Head of HH =26, but it's really 62 | This is impossible to detect post-hoc. | Train enumerators, conduct sport checks, perform inter-rater reliability tests (two enumerators collect the same data) | Cross-check physical proof (scan of ID card, ex), |

Other problems you might come across, especially when receiving a spreadsheet from another actor (when CRS was not responsible for designing tools and collecting data) include:

| OTHER COMMON ERRORS | WHY IT'S A PROBLEM | HOW TO AVOID AND ADDRESS IT |
|---|---|---|
| **Values are missing in the dataset** | If the dataset has missing or null values, it is hard to know if the data was never collected, if a respondent refused to answer, or if some other problem occurred. | In surveys, include an option for the participant to respond "I don't know", "I refuse to answer" or "not applicable"<br><br>Note: In CommCare, by default, questions that were skipped have the value '---'. |

---

[6] For more information on data entry mistakes and data cleaning, see ACAPS Spotting Dubious Data guidance.

| | | Empty cells have to be intentionally created. |
|---|---|---|
| **Zeros replacing missing values** | Sometimes datasets might have 0's instead of blanks, null values, "not applicable" etc. | Code "does not wish to respond", "not applicable" etc. using numbers or letters like "998; 999; xx" so that it is easy to tabulate these responses and separate them from true null values or "nothings". |
| **Spelling mistakes** | Misspelled names of people and places can make it hard to count the true number of project participants.<br><br>This is a common problem when transliterating text (e.g. from Arabic script to Roman script). | For people's names, use the spelling on an official ID document.<br><br>When transliterating, identify a common standard.<br><br>When designing digital tools like CommCare, use lookup tables to pre-load place names, making it easy for enumerators. |
| **Inconsistent date formats** | Which date is in August? 9/8/2020 or 8/9/2020? If the first one was written by a European and the second by an American, they both are! | This can be avoided by using digital tools like CommCare for data collection.<br><br>For data that is manually entered into Excel, consider using an expanded date format, like "August 9, 2020". |

## STORING/ARCHIVING DATA

Once you've collected and organized your data properly, you want to keep it safe. Consider following the 3-2-1 methodology for storing files to prevent loss of data/information:

- 3: Store **three** copies of data
- 2: Using **two** types of storage methods/locations
- 1: With **one** offsite

This can be easily accomplished by syncing Teams folders to your desktop and, in the file properties, enabling "always keep a copy" on the computer. This creates two copies (one in the cloud, and one on the device). Then, for the third copy, once per quarter (or more often as appropriate) do a manual backup to separate hard drive or flash drive which is kept in a locked cabinet.

## A NOTE ABOUT DATA SECURITY:

It's always recommended to use administrative, technical, and physical safeguards to protect data.[7] Consider the following:

| SAFEGUARD | NOTES |
|---|---|
| **Access Groups and Permissions**<br><br>Users are grouped by specific privileges according to their role and what they "need to know" or access (read/edit) | Refer to your RACI matrix.<br><br>Default to minimal access for more users.<br><br>Consider creating private Teams channels and manage access to SharePoint folders as appropriate.<br><br>Manage user permissions and create user groups in CommCare. |
| **Physical Safeguards**<br><br>Restrict physical access to data and devices used for data collection. | Always store databases securely (e.g. on CommCare, RR, TeamSite/SharePoint Site). If data must be collected on paper, store it in locked cabinets**.**<br><br>Do not use personal cloud storage services (e.g. Google Drive, Dropbox, etc.) to store data. |
| **Two factor/multi-factor authentication**<br><br>A security system that verifies a user's identity using 2 or more credentials.<br><br>Two factor uses something you know (password) plus something you have (e.g. phone to receive SMS code).<br><br>Multifactor may include more than one thing you know (e.g. password and response to security questions), something you have (e.g. phone, computer with email) or biometric data. | It's a good idea to set strong passwords for your CRS account, online services like CommCare, including for mobile users. Use a different password for each service.<br><br>CRS already uses a secure single-sign in system for the organization to manage access Microsoft Services (e.g. SharePoint, OneDrive, Power BI). |
| **Encryption**<br><br>Translation of data into another code that requires a private digital key to translate the data back into a readable form. | No action by user necessary.<br><br>Microsoft and other cloud service providers already encrypt data in transit to and upon storage to their servers. |

## SHARING INFORMATION

OCHA has outlined **3 C's** of IM to support effective responses: coordination, communication and collaboration. Some tips are:

---

[7] More information about securing data can be found here, from the Principles for Digital Development resource page, or here, from *Shooting our HardDrive into Space, and Other Ways to Practice Responsible Development Data.*

- **Internal Meetings** are a simple and effective way to share information within CRS. It is helpful to prepare for these meetings with a standard agenda and regular schedule to ensure they are efficient. Opportunities include: regular debriefs, learning events, etc. Meetings can also happen virtually over Teams, Skype, or WhatsApp. Consider sharing out notes by email or logging in a shared Teams channel.
- **Information Working Group and other WG Meetings.** IMWG bring together IM and MEAL focal points from different actors involved in a response (UN, NGOs, government, etc.). The role of the IMWG is to identify and improve common operational datasets (CODs), identify available data sources, create a list of key terms and definitions, and identify data gaps.[8]
    o Key questions for CRS staff to consider during working group meetings are:
    o What information are other groups sharing? Is this consistent with CRS information? Why or why not?
    o What information needs do other actors have? How can CRS support (e.g. by sharing information to fill gaps or influence decisions)?
- **Log Notes from coordination meetings** It's a good practice to log which coordination meetings (with clusters, government officials, etc.) are attended, by whom, and to document key notes. Some teams have found it helpful to create a shared log using the Microsoft Notes app within a Teams channel. It's up to you how to store these logs in a way that is easy to track and access and refer back to.
-

---

[8] OCHA. IM Guidance: Sudden Onset Emergencies. p14. Link.

INFORMATION MANAGEMENT IN EMERGENCIES

## ANNEX 1. FILE STORAGE - TEMPLATE STRUCTURE FOR TEAMS

When creating a new file system for an emergency response, consider using the template below to get started. This is particularly helpful for Teams sites or other collaborative spaces, where many people need to be able to find, edit and share documents quickly.

A sample folder structure can be found below. *A copy-able template with this folder structure can be found here, for use on Teams.*

| Tier | Type |
|------|------|
| 1 | Department (e.g., Programs, HR, Finance, etc.) |
| 2 | Discrete Projects (Project A, B, C[1]) and Cross Project Elements (e.g., Assessments that are not project specific, Databases that contain data from multiple projects [e.g., beneficiaries registered in overlapping projects], Warehouse protocols that cover multiple projects) |
| 3 | Project specific activities (proposals, MEAL plans, financial projections for a specific project |

1. Programs
    1.1. Project A
        1.1.1. Planning
        1.1.2. Funding and Needs
        1.1.3. Proposal and Project Documents (Award documents)
        1.1.4. Orientation Documentation
        1.1.5. General Information and Emergency Background
        1.1.6. Handover Notes
    1.2. Project B
    1.3. Project C

2. MEAL
    2.1. Databases
        2.1.1. Registration List
            2.1.1.1. Site A
            2.1.1.2. Site B
        2.1.2. Distribution Lists
            2.1.2.1. Site A
            2.1.2.2. Site B
        2.1.3. Digital Data Collection Tool Exports
    2.2. Project A
        2.2.1. MEAL Design (SMILER+ Documents, RF/Proframe, IPTT, MEAL Calendar, etc)
        2.2.2. Monitoring
            2.2.2.1. Tools
        2.2.3. Evaluation
            2.2.3.1. Needs Assessments
            2.2.3.2. Baseline
            2.2.3.3. RTE
            2.2.3.4. Final Evaluation
        2.2.4. Accountability
            2.2.4.1. Feedback forms
            2.2.4.2. Hotline tools
            2.2.4.3. Feedback database
            2.2.4.4. Protection Procedures
        2.2.5. Learning
            2.2.5.1. After Action Review
            2.2.5.2. Reflection Events

  2.2.6. Reports
    2.2.6.1. Reporting Templates
      2.2.6.1.1. 4W
      2.2.6.1.2. Weekly Reports
      2.2.6.1.3. Monthly reports
      2.2.6.1.4. Final reports
    2.2.6.2. Completed/submitted reports
      2.2.6.2.1. 4W
      2.2.6.2.2. Weekly Reports
      2.2.6.2.3. Monthly reports
      2.2.6.2.4. Final reports

3. Human Resources
 3.1. Employee Details (Employee Roster, Org Chart)
 3.2. Recruitment (Job Descriptions, Volunteers)
 3.3. TDYs (SOWs)
 3.4. Policies (Per Diem, R&R, etc.)

4. Supply Chain
 4.1. Warehouse Forms
  4.1.1. Waybills
  4.1.2. Receiving Guide
  4.1.3. Warehouse Form
  4.1.4. Lot Form
  4.1.5. Tally Sheet
  4.1.6. Stock Request
  4.1.7. Distribution Request
  4.1.8. Dispatch Plan
  4.1.9. Distribution Report
  4.1.10. Loss Report
  4.1.11. Pipeline
  4.1.12. Procedures Manual

5. Finance
 5.1. Finance Templates
  5.1.1. Reimbursement Form
  5.1.2. Liquidation Form
  5.1.3. Advance Form
  5.1.4. Waiver Templates
 5.2. Policy
  5.2.1. Agreements Guidance
  5.2.2. Per Diem Policy
 5.3. Project A
  5.3.1. Financial Reports
  5.3.2. Reimbursements
  5.3.3. Liquidations
  5.3.4. Cash Flow Forecasts
  5.3.5. Cash Requests
  5.3.6. Budget Forecasts
 5.4. Project B
 5.5. Project C

6. Admin
 6.1. Vehicle Log
 6.2. Procurement
  6.2.1. Template for Procurement Request
  6.2.2. Procurement Procedures
 6.3. Inventory
  6.3.1. Receipt of materials
  6.3.2. Database of inventory
 6.4. Contracts

7. Security
 7.1. Security Briefings
 7.2. Staff Movement Plan
 7.3. Security Plan

8. Photos
    8.1. Donor visits
    8.2. Monitoring visits
    8.3. Distributions
    8.4. Workshops
    8.5. Photo consent forms

## ANNEX 2: KEY AVAILABLE DATA SETS FOR EMERGENCY RESPONSE

OCHA maintains Common Operational Datasets and Fundamental Operational Datasets. OCHA is usually responsible for updating and distributing these during emergencies. Sometimes HCTs or other actors fill this role.

**Common Operational Datasets (**COD**)** are the standard for the humanitarian community. They are an authoritative reference that represent the best-available data on a range of topics:
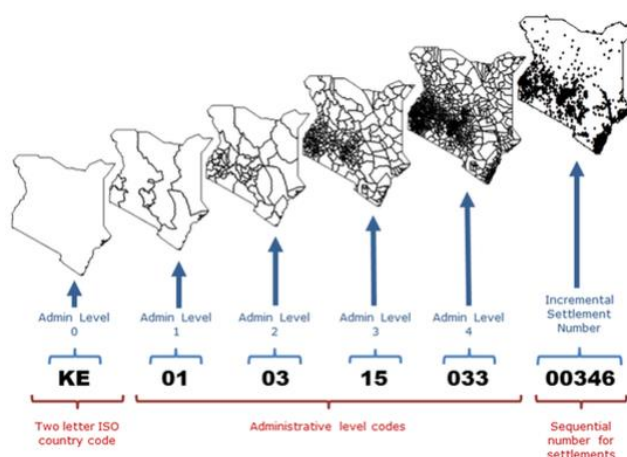
- **Core CODs** contain:
    - o Administrative boundaries (**COD-AB)**
    - o Population statistics (**COD-PS)**
    - o Caseload figures (e.g. number of affected people)
- **Country-specific CODs** consist of other information, such as:
    - o Populated areas (settlements, towns, cities)
    - o Transportation Networks (roads, ports, etc.)
    - o Hydrology (streams, lakes, other bodies of water)
    - o Hypsography (elevation, contours)
    - o Education: schools and other facilities
    - o Health system: hospitals, clinics, etc.
    - o Affected persons locations: refugee and/or IDP camps, etc.
    - o Logistics: warehouse locations
    - o Security: police stations, hazard areas, etc.
    - o Other examples: wells and water points, fuel stations, etc.
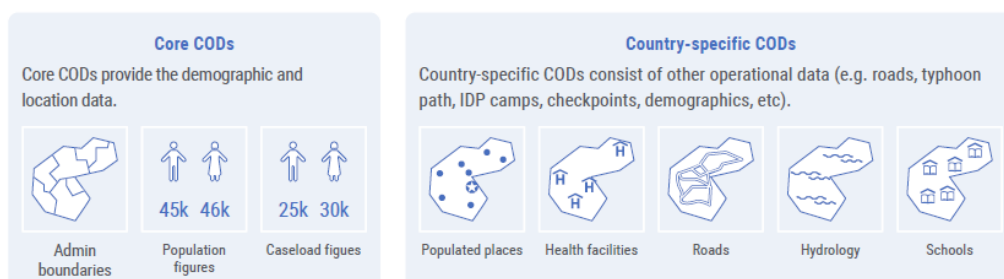


The purpose of CODs is to support coordinated response among different actors. CODs provide a common operational picture of the crisis, and a geographic framework for data collection, analysis and consolidation through place codes (**P-codes)**.

P-codes are unique geographic codes, represented by letters and/or numbers to location or feature on a map (geo) identification combinations of identify a specific or within a database.

The data for CODs and FODs sources, including: national agencies, clusters, NGOs, etc. can be found here. comes from external authorities, UN More information

# ANNEX 3: OCHA IM PRINCIPLES

A description of OCHA's **Principles for Humanitarian Information Management and Exchange** can be found below.[9]

Table 1. Principles for Humanitarian Information Management

| Function | Principle | Explanation |
|---|---|---|
| **Check** | Reliability | Users must be able to **evaluate** the reliability and credibility of data and information by knowing its source and method of collection. Collection methods should adhere to global standards where they exist to support and reinforce credibility. Reliability is a prerequisite for ensuring validity and verifiability. |
| | Verifiability | Information should be accurate, consistent and based on sound methodologies, **validated** by external sources, and analysed within the proper contextual framework. |
| **Share** | Inter-operability | All sharable data and information should be made available in **formats** that can be easily retrieved and shared by humanitarian organizations. |
| | Accessibility | Humanitarian information and data should be **made accessible** to all humanitarian actors by applying easy-to-use formats and by translating information into common or local languages. Information should be widely distributed through a variety of online and offline distribution channels including the media. |
| | Sustainability | Humanitarian information and data should be preserved, catalogued and archived, so that it can be retrieved for preparedness, analysis, and evaluation. When possible, post emergency data should be transitioned to relevant recovery actors and host governments and training provided on its use. |
| **Use** | Timeliness | Humanitarian information should be collected, analysed and disseminated efficiently, and must be kept **current**. |
| | Relevance | Information should be practical, flexible, responsive, and driven by **operational** needs in support of decision-making throughout all phases of a crisis. Data that is not relevant should not be collected. |
| **Humani-tarian Princi-ples** | Inclusiveness | Information management should be based on collaboration, partnership and sharing with a high degree of **participation** and ownership by multiple stakeholders including national and local governments, and affected communities. |
| | Accountability | Information providers should be **responsible** to their partners and stakeholders for the content they publish and disseminate. |
| | Impartiality | Information managers should consult a variety of sources when collecting and analysing information so as to provide varied and **balanced perspectives** for addressing problems and recommending solutions. |
| | Humanity | Information should never be used to distort, to mislead or to **cause harm** to affected or at risk populations and should respect the dignity of victims. |
| | Reciprocity | Information exchange should be a **beneficial** two-way process between the affected communities and the humanitarian community, including affected governments. |
| | Confidential-ity | The processing of any personal data shall not be done without the prior explicit description of its purpose and will only be done for that purpose, and after prior informed consent of the individual concerned. Sufficient safeguards must be put in place to protect personal data against loss, unauthorized processing and other misuse. If sensitive information is publicly disclosed, the sources of such information will not be released when there is a reasonable risk that doing so will affect the security or integrity of these sources. |

---

[9] Bartel Van de Walle and Tina Comes. On the Nature of Information Management in Complex and Natural Disasters. Humanitarian Technology: Science, Systems and Global Impact, HumTech 2015. Procedia Engineering. 107.2015. Link

# ANNEX 4: IM FAQS

## GENERAL INFORMATION MANAGEMENT FAQS

☐ **What is the difference between Information Management, Knowledge Management, and Data Management?**

These terms are often used synonymously; however, they have slightly different connotations. ***Information management*** refers to the process controlling how data is gathered, validated, cleaned, analyzed, consolidated, transformed into useful information and ultimately shared, emphasizing data use for communication, collaboration and coordination. ***Knowledge management*** tends to refer to a similar process but emphasizes longer-term institutional memory and cross-team learning. ***Data management*** is an aspect of IM/KM focused on collecting, keeping, and using data securely, efficiently, and effectively.

☐ **How does IM relate to MEAL?**

Information management practices complement M&E activities and serves as the bridge between M&E activities and decision-making, adaptive management and coordination.
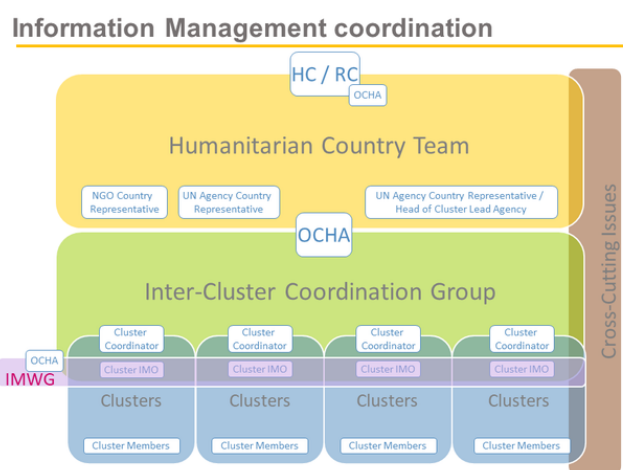
☐ **How to IM relate to coordination?**

IM enables effective coordination by providing timely, accurate information to relevant stakeholders, helping humanitarian actors decide how, where and when to respond to crises.

Typically, OCHA is ultimately responsible for IM between/across clusters, including: coordinating the MIRA, supporting cluster-level assessments, coordinating the production of the HNO and supporting the HCT to develop a response strategy, maintaining and updating CODs, and identifying gaps in the response based on the 4Ws.

Within clusters, the cluster responsible for IM, appointing an IM focal point, inter-cluster IM ensuring alignment with standards, generating maintaining 4Ws, common cluster datasets, consolidated SitReps, and TORs, technical etc, and ensuring data standards are met. lead agency is including: contributing to coordination, global IM contact lists, inventorying producing meeting minutes, guidance, datasets, protection



Information Management coordination

At the cluster member level, CRS and our partners are organizations like responsible for providing information, and for taking into account the information of other actors in decision making.

## DATABASE AND DATA MANAGEMENT FAQS

☐ **I usually store my project's data in Excel spreadsheets. Is that ok?**

<u>Quick answer</u>: Yes, but it might not be the best solution.

<u>Long Answer</u>: Excel is a good place to store data that one person is actively analyzing. However, if you have a lot of different fields and/or multiple spreadsheets, it is better to use a database or platform like PBI. Additionally, as Excel is a software that is continually updated, the file format that works now may not work in the future, so for longer term storage/archiving, consider saving the data in a .csv format, rather than .xls. Other concerns about Excel are that it is limited to 1,048,576 rows by 16,384 columns, meaning if you have a lot of records (such as cash transfers), you will need multiple sheets. Finally, Excel is heavier than other

programs, so if expect it to slow down the more data you have – always remember to hit save frequently when analyzing to avoid crashes.

- **Do I really need to use Power BI?**

<u>Quick answer</u>: No.

<u>Long answer</u>: It's not a requirement to use Power BI, however, it is recommended in many cases. PowerBI is a powerful data analysis and visualization platform that pulls data from other sources, including CommCare, CAT, and Excel, making the analysis more efficient. Power BI can automatically refresh, meaning once you have set it up, results will be updated as data comes in, saving time in the long run. Power BI also helps facilitate *data use* by making results easy to understand through visualizations, helping program managers, senior management, partners and donors identify trends quickly.

- **Our partner only uses Excel, and their reports are hard to analyze. How can we support them to improve?**

Consider developing sub-recipient reporting templates in a participatory workshop (such as a SMILER+). Be sure to include data validation rules in Excel spreadsheet templates. Support partners to access and use CommCare and ICT4D tools and PBI, and, depending on context and resources, prioritize capacity building on MEAL and ICT4D.

- **No one in the field is familiar with CommCare (or CAT or Power BI). What support is available?**

HRD can support field teams to set up CommCare, CAT and Power BI and train staff. Additionally, GKIM and Regional or CP-level MEAL and ICT4 staff can support.

# ANNEX 5: GLOSSARY[10]

| Term | Definition |
|---|---|
| Common Operational Dataset (COD) | An authoritative reference dataset to support operations and decision-making for all actors in a humanitarian response. CODs are the "best available" dataset on a range of topics. |
| Data | A collection of related facts usually organized in a particular format, such as a table or database and gathered for a particular purpose. |
| Data cleaning | The process of correcting and/or standardizing data from a record set, table, or database. |
| Data Management | Data management is an aspect of IM/KM focused on collecting, keeping, and using data securely, efficiently, and effectively. |
| Data responsibility | Data responsibility entails a set of principles, processes and tools that support the safe, ethical and effective management of data in humanitarian response |
| Data security | A set of physical, technological and procedural measures that safeguard the confidentiality, integrity and availability of data and prevent its accidental or intentional, unlawful or otherwise unauthorized loss, destruction, alteration, acquisition, or disclosure |
| Data standard | A published specification, e.g. the structure of a particular file format, recommended nomenclature to use in a particular domain, a common set of metadata fields, etc. Conforming to relevant standards greatly increases the value of published data by improving machine readability and easing data integration. Source Open Data Handbook. The Humanitarian Exchange Language (HXL) is a lightweight data standard for exchanging humanitarian data in an interoperable way based on spreadsheet formats such as CSV or Excel. |
| Geodata | Information about geographic locations that is stored in a format that can be used with a geographic information system (GIS). Geodata can be stored in a database, geodatabase, shapefile, coverage, raster image, or even a dbf table or Microsoft Excel spreadsheet. Source ArcGIS. |
| Information | Data that has been interpreted, verbalized, translated, or transformed to reveal the underlying meaning and context. |
| Information Management | Information Management refers to the process controlling how data is gathered, validated, cleaned, analyzed, consolidated, transformed into useful information and |

---

[10] Terms adapted from Center for Humanitarian Data - https://centre.humdata.org/glossary/

| | |
|---|---|
| | ultimately shared, emphasizing data use for communication, collaboration and coordination. |
| **Knowledge** | The internalization of information, data, and experience. **Tacit Knowledge** is the personal knowledge resident within the mind, behavior and perceptions of individual members of the organization. **Explicit Knowledge** is the formal, recorded, or systematic knowledge that can easily be accessed, transmitted, or stored in computer files or hard copy. |
| **Knowledge management** | The systematic strategy to collect, store and retrieve knowledge and then help distribute the information and knowledge to those who need it in a timely manner. |
| **Metadata** | Data about data, or data that defines or describes other data. Metadata is additional information or documentation about your dataset that makes it easier for others to understand and put your data into context. (Source HDX.) |
| **Open data** | data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike. (Source Open Data Handbook) |
| **Sensitive data** | Data that, if disclosed or accessed without proper authorization, is likely to cause harm to any person, including the source of the data or other identifiable persons or groups, or a negative impact on an organization's capacity to carry out its activities or on public perceptions of that organization. (Source OCHA Data Responsibility Guidelines) |