Responsible data management training cycle

cartong

MINISTÈRE DE L'EUROPE ET DES AFFAIRES ÉTRANGÈRES

Session 4
Responsible Data in action - part 2

Introduction to session 4

And now…

1/ The different dimensions of responsible data management

2/ Focus on data protection

3/ The concepts of responsible data management in action- part 1

4/ The concepts of responsible data management in action- part 2

5/ Discover how current stakes apply to you

Session 4 agenda

- Introduction to today's session
- Overview of the « data usage » steps of the cycle
- 1 focus on topic of high interest : De-identification
- Group work on 6 themes & debrief
- Local peer-to-peer exchange
- Conclusion

The development of this training material is supported by the French Ministry of Europe and Foreign Affairs (MEAE-CDCS). Nevertheless, the ideas and opinions presented in this training do not necessarily represent those of MEAE-CDCS.

Feedback on the homework

Feedback on homework

Questions & Answers and Quiz

Quiz

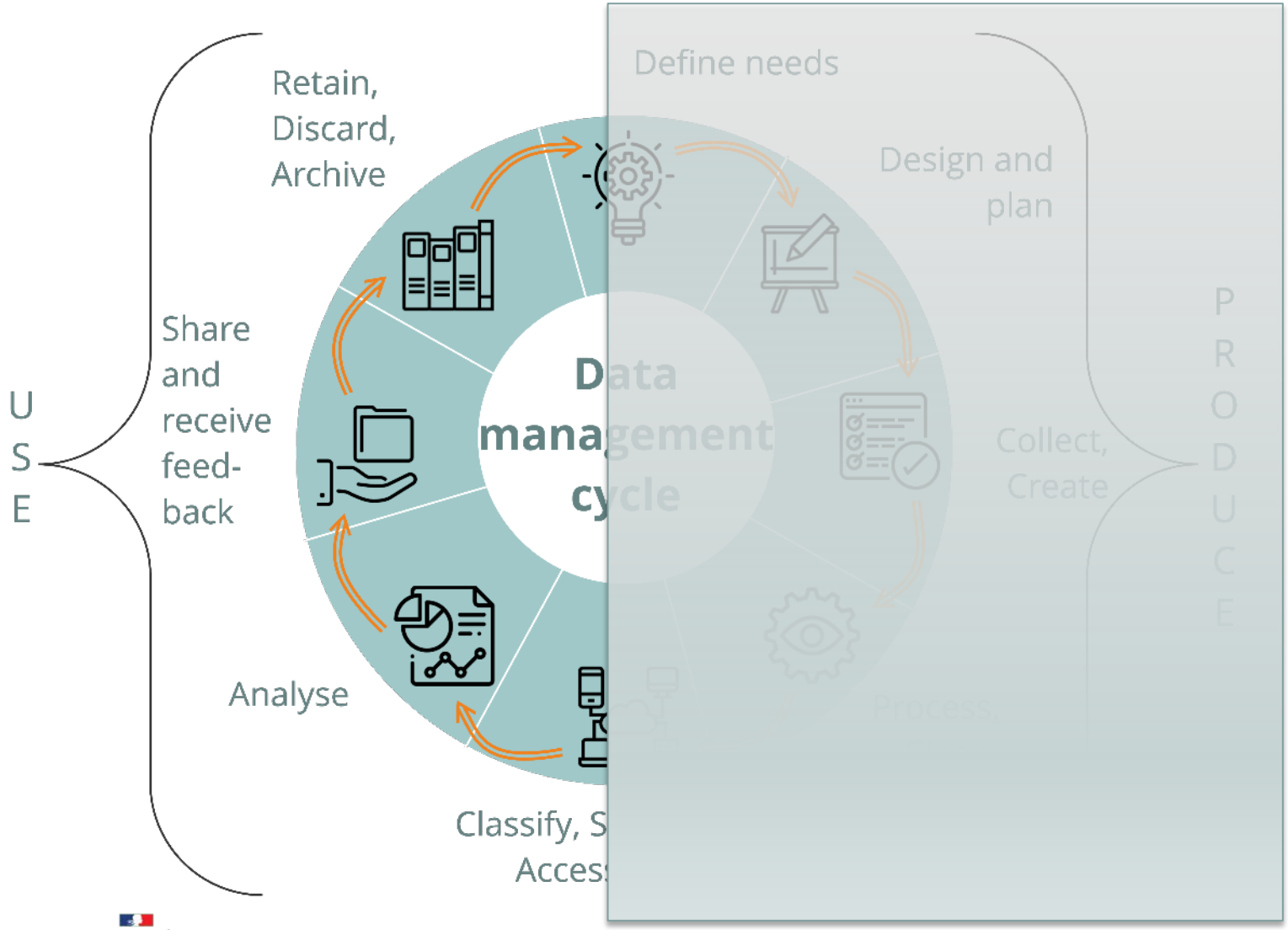
Questions & Answers

You can find all the questions and answers from last session, including those we didn't have time to answer live, in the constant companion.

Do you have new questions on the last session?

Responsible data applied to data usage

Data Management cycle



Do you remember Akachaland and our Unicorn NGO?

Remember…

In the country **Akachaland,** a major flood during the moon season has devastated around 30 villages, located in the north. You are a member of the **« Unicorn » NGO,** from Finobaka, specialized in the protection of children and women to whom you also provide food and NFI.

Unicorn is working **with a its local partner Akachaland care who is amongst other things collecting the data**. The 2 ONGs have decided to share the database to avoid collecting too much data.

Akachaland

cartong

MINISTÈRE DE L'EUROPE ET DES AFFAIRES ÉTRANGÈRES

Step 6:

Analyse

Confidentiality

•Ensure data for analysis is only **accessible to authorized staff**

Security

- **Respect SOPs** on analysis

Quality

- Follow the analysis plan / check what is produced is inline with what was planned
- **Ensure the analysis is not unduly biased**

Literacy

- Ensure **proper understanding of data quality**, bias, of the teams involved in the cleaning, as well as the limits of the data collected

What would it look like in Akachaland?

Unicorn put in place **internal procedures** alongside all the steps of the data management cycle to harmonise the practises in its missions and provide guidance to its staff, and has worked with Akachaland Care to do the same.

Akachaland Care colleagues collected data amongst the women and children. **This collection aims at identifying the recipients** of Unicorn's assistance **and their specific needs**. The data is sensitive and personal and is stored in a database. 2 members of your Unicorn team need access to the database to analyse it.

How do you check that the collected data matches its designed purpose?
How do manage the access rights to the database?
What measures do you take to follow your own procedures? How do you make sure the analysis process is unbiased ?

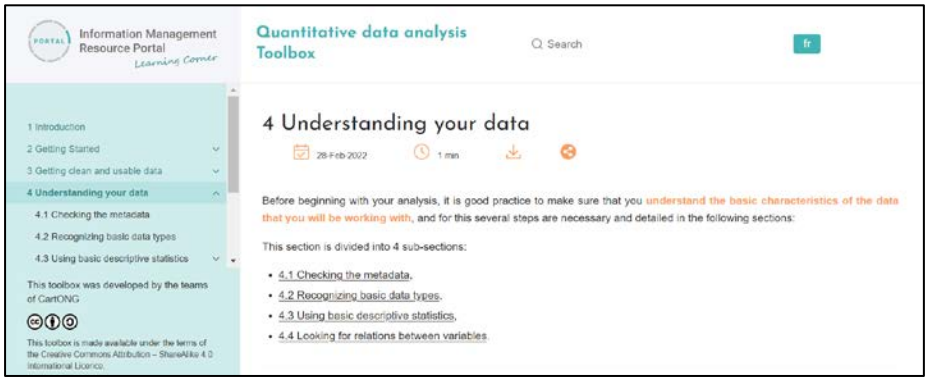
ANALYSIS PLAN

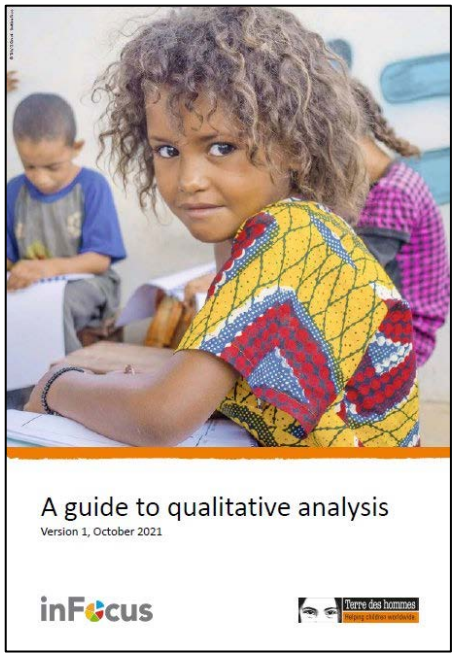
Research question	Indicator / variable	Questionnaire question	Data collection unit	Desired disaggregation	Analysis type
What are the demographics of the population in terms of age, gender and dependency ratio?	Household composition	1. What is the sex of the head of household? 2. What is the age of the head of household? 3. Is the household member currently pregnant? 4. What is the total number of household members? 5. What is the sex of the household member? 6. What is the age of the household member (years)?	Household	Region district	Population pyramids; Frequency distribution on gender; Measures of central tendency of age Measures of variance of age.
	Dependency ratio	What is the age of the household member (years)?	Household	Region	Demographic breakdown between population above and below 15-years old Percentage of 'dependents' (children aged under 15 years and the older population aged 60+) of the total population (total population including 'working age' population, between 15 and 59 years).
What are the most common unmet household basic needs due to lack of affordability and do household's have multiple concurrent unmet needs?	Affordability of basic household needs	Which of your household's basic needs can you not afford?	Household	Region	Frequency distribution.
What proportion of the population has access to improved WASH services and does it vary by region?	Access to improved WASH services	1. What is the principal source of drinking water for members of your household? 2. Where do you, and your household members (including children under 5) usually go to defecate?	Household	Region	Disaggregation per region; Frequency distribution of types of improved water sources includes the response options: 'public tap/standpipe', 'hand/pump/borehole', 'piped connection to house', 'protected spring', and 'bottled water, water sachets'; Frequency distribution of types of improved sanitation includes the response options 'household latrine' and 'communal latrine'.

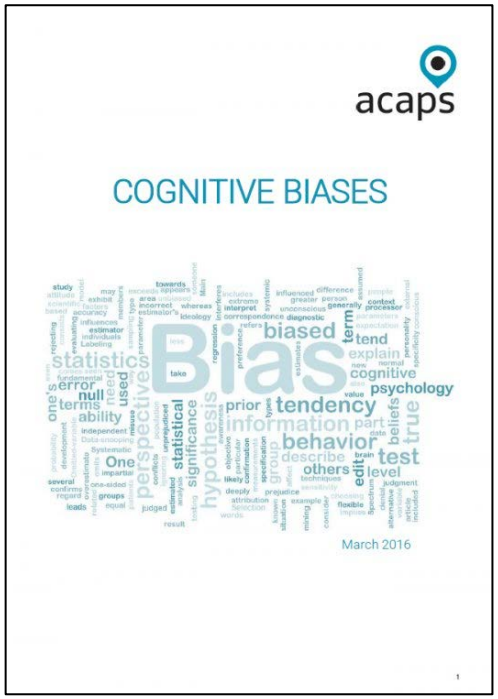
cartong

MINISTÈRE DE L'EUROPE ET DES AFFAIRES ÉTRANGÈRES

Ressources / templates available







Sources: Tdh, ACAPS and CartONG quantitative data analysis Toolbox

cartong

MINISTÈRE DE L'EUROPE ET DES AFFAIRES ÉTRANGÈRES

Step 7 : Share and receive feed-back

Confidentiality

•Ensure data is only **shared to authorized staff and partners**
•De-identify what you can, through aggregation and anonymization (or pseudonymization if no other solution)
•Minimise the personal data sha

Security

Ensure **security** of platform sharing data, with passwords/encryption
avoid use of email if they – or your recipient's- are not encrypted
Respect SOPs on data sharing
Avoid non-professional accounts/emails/hardware used
Ensure passwords are shared in a secure fashion

Quality

- Share analysis with communities if relevant for feedback
- Build spaces for feedback with partners/communities
- **Share aggregated data/ analysis with wider community if relevant / not sensitive**

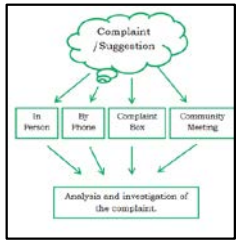
Literacy

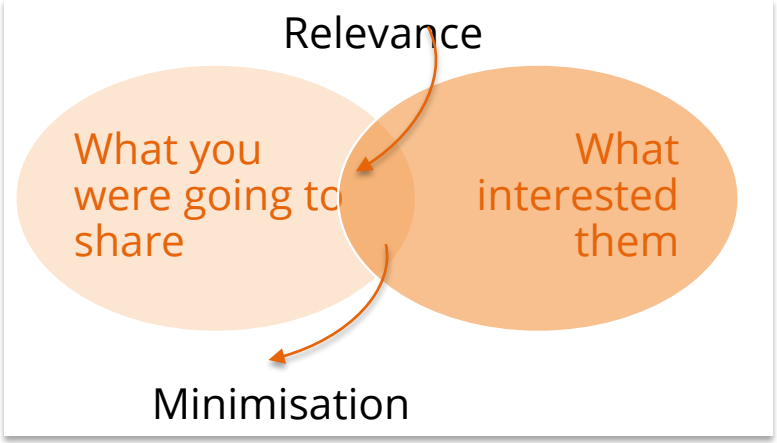
What would it look like in Akachaland?

The **data analysis is completed**. Your donor from Brajoki is requesting the database as planned for an audit. **Some affected communities are concerned about the data being shared too widely**, as the armed group AKaidnippers is active and specialized in stealing the localisation data of children.

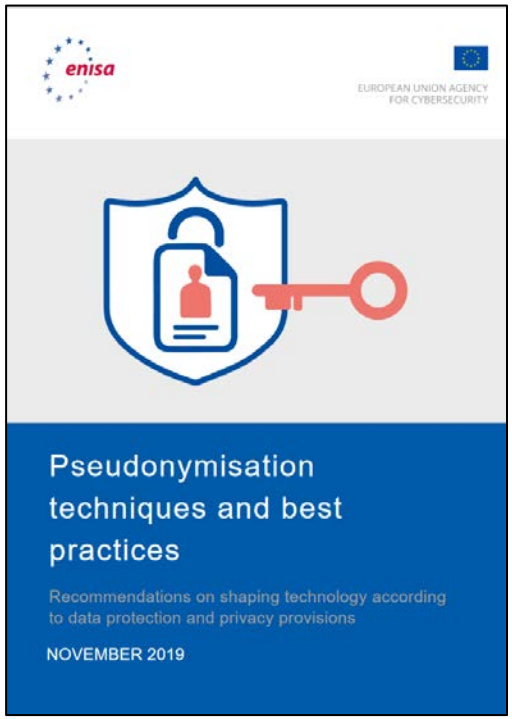
How do you prepare the data sharing to make sure it's secure? Which platform are you going to use to share data? How do you make sure the shared data will be used properly with confidentiality by the donor?

Which mechanism was planned to insure the communities can share feedback? How will you respond to their worries about being identified?



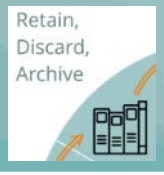


Ressources / templates available



Sources: Enisa (EU) & OHCHR

Step 8 :



Confidentiality

- Follow your RAD plan: make sure you anonymise / archive (/ intermediary archive) or delete as soon as it can be
- **Check there are no duplicates elsewhere**
- Continue applying **data subject rights**
- Check that your partners also respect the RAD plan!

Security

- Ensure **physical & IT security** of data (paper & digital)
- **Respect SOPs** on access

Quality

- Document any issues encountered for accountability

Literacy

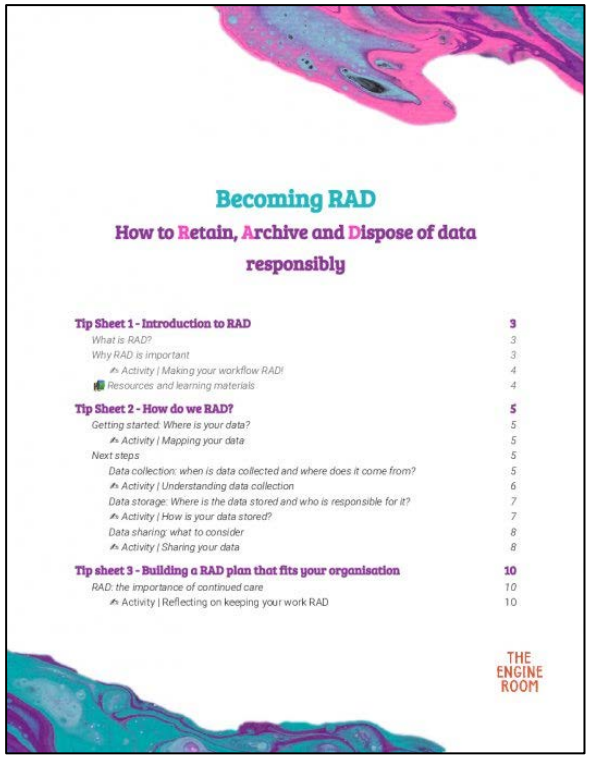
- Accompany partners on the importance of becoming RAD
- Share learnings from this project

What would it look like in Akachaland?

After 5 years, the project of Unicorn **will end soon** in the northern region of Akachaland. As part of the project the Unicorn team is still using some data for their daily activities with some women and children, but many have been displaced to another country, and much of the initial data is no longer operationally useful.

When do you archive the data? Which data will be anonymised and how? What about the data shared with Akachaland Care: how do you make sure they have planned a retention period and make sure to delete the data then?

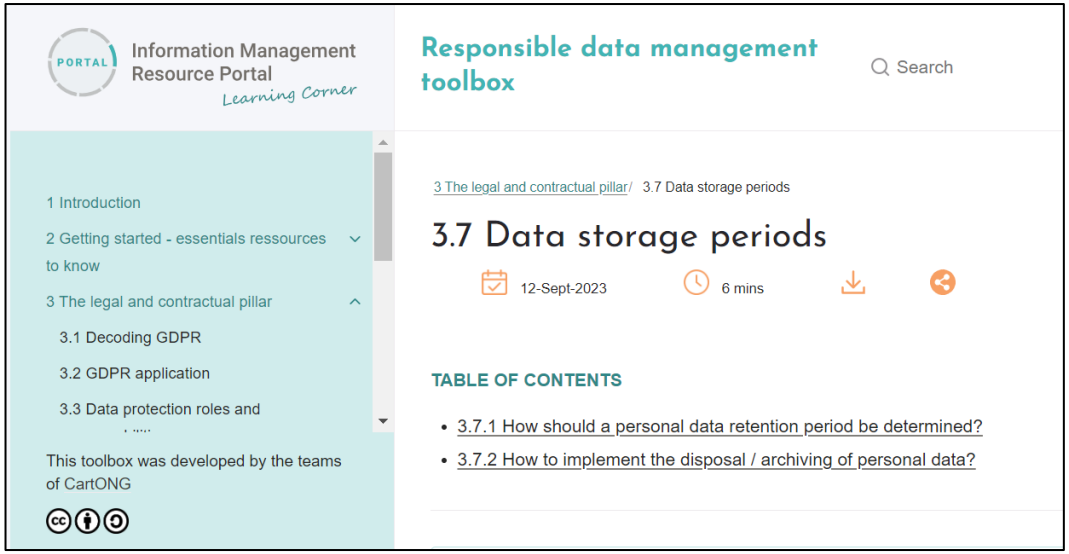
Ressources / templates available



Becoming RAD

How to Retain, Archive and Dispose of data responsibly

THE ENGINE ROOM



PORTAL Information Management Resource Portal
Learning Corner

Responsible data management toolbox

Search

This toolbox was developed by the teams of CartONG

3 The legal and contractual pillar / 3.7 Data storage periods

3.7 Data storage periods

12-Sept-2023 6 mins

TABLE OF CONTENTS

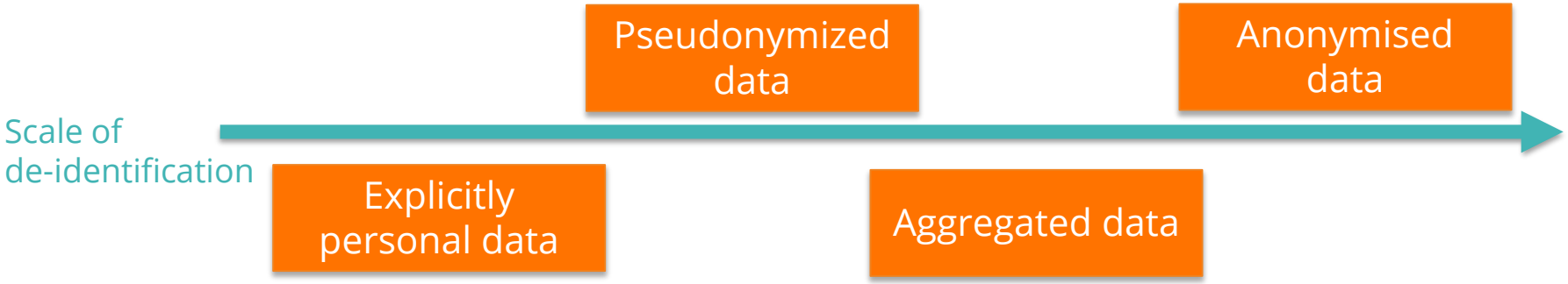
Sources: The Engine Room/CartONG & the Responsible data management toolbox

Key focus on deidentification

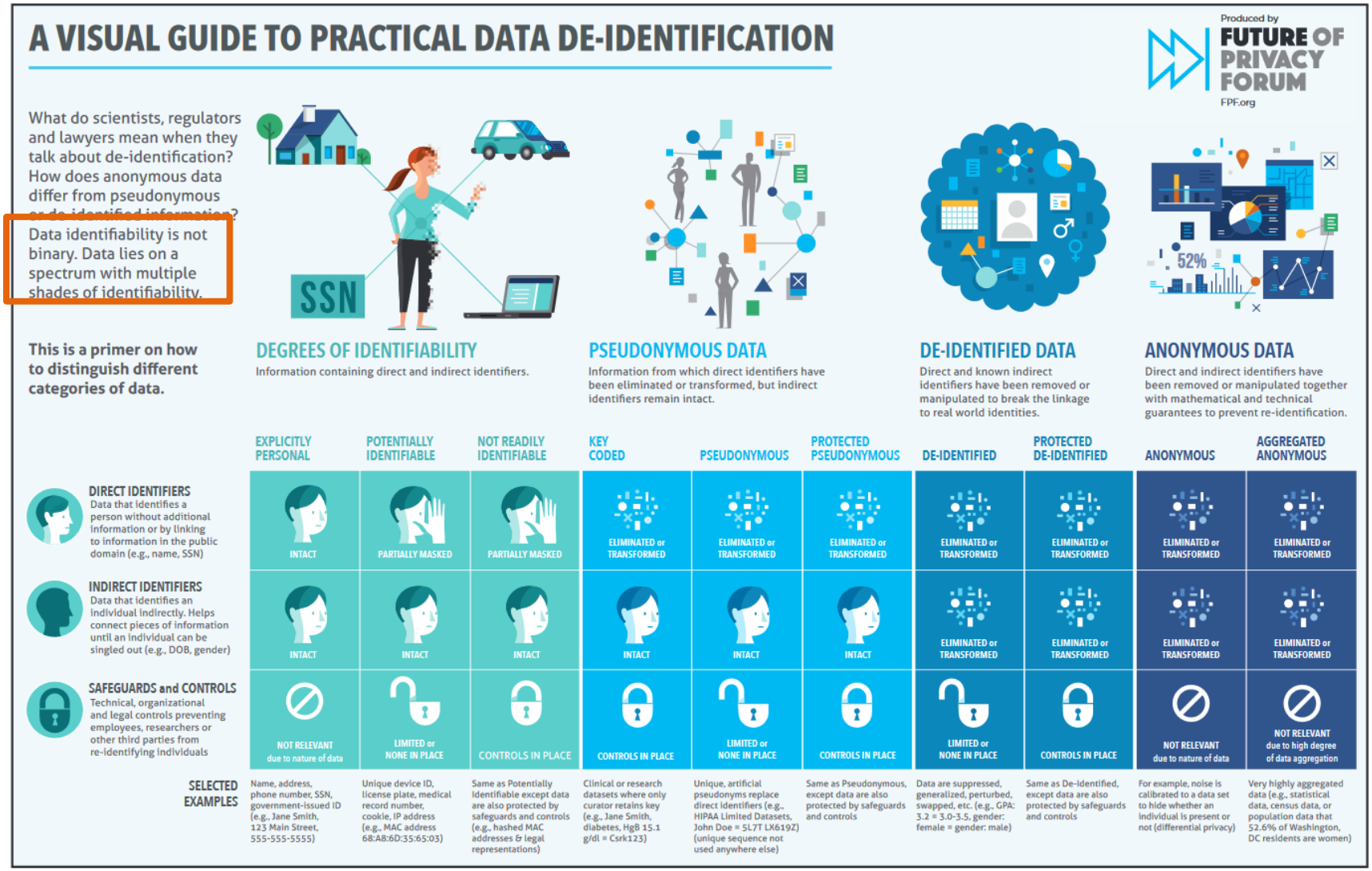
Anonymisation vs de-identification

De-identification is "the process used to **prevent a person's personal identity from being revealed**".

There are different **degrees** of identification of data (in a nutshell):



We have tried to simplify something rather complex!



Source: [future of privacy forum]

cartong

MINISTÈRE DE L'EUROPE ET DES AFFAIRES ÉTRANGÈRES

Techniques of de-identification

Pseudonymization is a method of replacing, at the individual level, identifiable data with a **reversible** and consistent value (e.g. a beneficiary code), provided that **this value is kept separate and safely**.

Aggregation is the merging or **grouping of data** from individuals into groups (which may in turn be pseudonymised or anonymised).

Anonymization is the process of **removing or modifying all personal identifiers**. Anonymization techniques are **irreversible**.

Name	Pseudonymisation	De-identification	Anonymisation	Aggregation
Noah	123	Age : 44 ans, Village : XX, Size HH : 12	XXXXXXXX	Groupe A
Pierre	456	Age : 22 ans, Village : YY, Size taille : 2	XXXXXXXX	Groupe A
Maar-Dipha	789	Age : 28 ans, Village : ZZ, Size taille : 5	XXXXXXXX	Groupe B

However, keep in mind that:

Pseudonymization identifiable data with beneficiary code), provided that **this value is kept separate and safely**.

Good practice, but it is STILL considered personal data, as it can still be re-identified.

Aggregation is t groups (which m

Can still be demographically or community identifiable data… and can therefore still be sensitive / cause harm to individuals or groups

Anonymization **identifiers**. Anor

No longer personal data, but requires difficult techniques. Humanitarian organisations therefore almost NEVER work with anonymised data

Name	Pseudonymisation	De-identification	Anonymisation	Aggregation
Noah	123	Age : 44 ans, Village : XX, Size HH : 12	XXXXXXXX	Groupe A
Pierre	456	Age : 22 ans, Village : YY, Size taille : 2	XXXXXXXX	Groupe A
Maar-Dipha	789	Age : 28 ans, Village : ZZ, Size taille : 5	XXXXXXXX	Groupe B

cartong

MINISTÈRE DE L'EUROPE ET DES AFFAIRES ÉTRANGÈRES

Field testimony

From HI – On its mission in Syria, about the practise of pseudonymisation of databases

Questions & Answers

Do you have questions on the the topic of deidentification?

Break

Group exercise

Group discussion !

Your group will have a **topic attributed**
You are expected:
- for the topic, to **list Do's and Don'ts in terms of good practices**
- to **answer your « nut cracking » question** related to the theme

Group numbers:

1. Anonymisation- opportunity or technical challenge?
2. data as a key to better conversations between teams
3. Data destruction as the forgotten step
4. Feedback to communities to close the loop
5. Learning from the experiences of the project
6. Data sharing in questionable contexts

Anonymisation- opportunity or technical challenge?!

- **Don't plan a big anonymisation exercice on each dataset you have-** it's quite a heavy enterprise that should only be done when you have planned proper resources

- **Being creative about how to deidentify data** when you can (pseudonymisation, aggregation etc) is already a great step forward

Solutions:	OK	Full Redaction	Partial Redaction	Chunk Up/Zoom Out	Assign Alternative ID	Add Noise	Average
Variables:	Release w/o alteration	Do not release in any form	Incomplete release (remove certain part of the data)	Collapse observations into groups (age example: 16-15, 16-20, 21-25)	Assign you own number to each observation at random (example: 031, 852, 083, 364)	Add random numbers to a piece of numeric data	Collapse units of observation through averaging (average ages of 5 people to create a single observation)
super-variable	Likely to appear across a range of data sets						
Names							
first name	John	X			X		
last name	Smith	X			X		
middle name	Brian	X			X		
username	J55	X			X		
petition signature		X					
Non-Bio Characteristics							
job title	X (unless small sample size allows for identification based on this characteristic)						
Relationships							
Guardian's profession	Doctor						
Spouse's name	Jane						
doctor's name	Ethan						
Identification Numbers							
ID (student, School)		X			X		
donor information, including bank account #							
passport #							
IP address		X					
platform specific identifier #							
electronic signature from a 3rd party app							
official government-provided ID (such as social security number)		X			X		
identifier number from a 3rd party application							
Geographic							
Physical address	9*09*						
Resource Location							

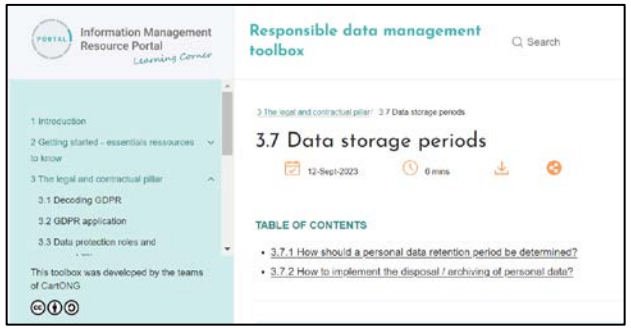
Data as a key to better conversations between teams

- Remember that "**Data is a team sport**" (School of Data)
- Create environments in which **people are more valued than data**, working together to use data as a vector for understanding & operational efficiency

IFRC | solferino academy

Data destruction as the forgotten step

- That final step of **deleting/archiving the data is essential** to ensure that when you might be gone on to other projects/missions, that nobody is left behind with data they don't know what they are supposed to do with



Feedback to communities & learning from the experiences of the project

Feedback to communities to close the loop

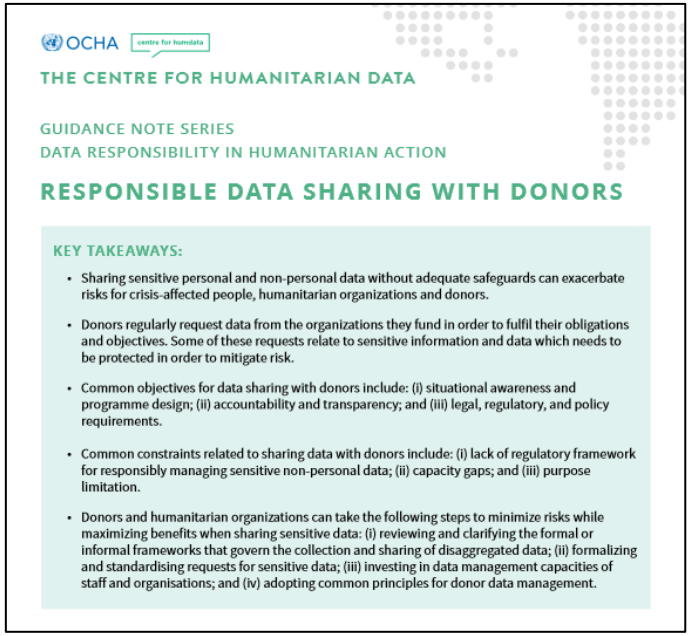
- When you are part of a survey, you like feeling that your time hasn't been wasted by seeing the result, even if it's just a few findings- think in the same way for affected populations!

Learning from the experiences of the project

- Each organisation has its own way of learning from projects, but **think about what can work best in terms of sharing of capitalisation**: learning events, peer exchange, practical traiings, awareness raising etc

Data sharing in questionable contexts

- Think of the **context**, the donor's potential **agenda**
- Discuss it with him, what you are afraid of, if you can have these types of discussions
- **Challenge** him on his practices
- **Refer** him to his organisational policies



OCHA centre for humdata

THE CENTRE FOR HUMANITARIAN DATA

GUIDANCE NOTE SERIES
DATA RESPONSIBILITY IN HUMANITARIAN ACTION

RESPONSIBLE DATA SHARING WITH DONORS

KEY TAKEAWAYS:

- Sharing sensitive personal and non-personal data without adequate safeguards can exacerbate risks for crisis-affected people, humanitarian organizations and donors.
- Donors regularly request data from the organizations they fund in order to fulfil their obligations and objectives. Some of these requests relate to sensitive information and data which needs to be protected in order to mitigate risk.
- Common objectives for data sharing with donors include: (i) situational awareness and programme design; (ii) accountability and transparency; and (iii) legal, regulatory, and policy requirements.
- Common constraints related to sharing data with donors include: (i) lack of regulatory framework for responsibly managing sensitive non-personal data; (ii) capacity gaps; and (iii) purpose limitation.
- Donors and humanitarian organizations can take the following steps to minimize risks while maximizing benefits when sharing sensitive data: (i) reviewing and clarifying the formal or informal frameworks that govern the collection and sharing of disaggregated data; (ii) formalizing and standardising requests for sensitive data; (iii) investing in data management capacities of staff and organisations; and (iv) adopting common principles for donor data management.

Peer-to-peer local exchanges

Local practises discussion in group !

You will be grouped together by **geographic zone**, to discuss common issues in responsible data responsibility:

- introduce yourselves by tackling a specific M&E topic in your geographical area,
- together, identify 1 topic of common interest among those shared, and exchange views on it

Groups are:

1. South East Asia (Philippines, Vietnam, Laos, Cambodia, Thaïland)
2. Bangladesh / Myanmar
3. Afghanistan/Pakistan/India/Nepal
4. Iraq
5. Lebanon
6. Yemen
7. Syria
8. Northern Africa / Middle East (Egypt, Lybia, Jordan)
9. Western Africa (Nigeria, Cameroun, Liberia, Sierra Leone)
10. East and South Africa (Burundi, Rwanda, Mozambique, Zimbabwe, Kenya, Uganda, Ethiopia, South Sundan)
11. Eastern Europe (Romania, Albania, Kosovo, Rep. Moldova, Ukraine)

Conclusion

Key messages to remember

- Much of what is required in terms of responsible data management is defined at the **planning /designing stage,** here you should mostly be **following the analysis plan/ RAD plan / SOPs/ data sharing agreements etc** defined uphill

- **There are many templates/ressources to help you** along your way / help your organisation structure/harmonise its procedures- it is not an insurmountable mountain, you will see many next week

- Some topics still lack off the shelf resources / procedures- but **follow your « responsible data » common sense** in the meantime !

Questions & Answers

Do you have questions from this session?

Homework for the next session

Thank you for your attention!

info@cartong.org

www.cartong.org