

KEYRUS
biopharma



insight into value

INNOVATIVE CONTRACT
RESEARCH ORGANIZATION

PRACTICAL APPLICATION OF THE EMA 0070 POLICY

on publication of clinical data for medicinal product for human use



Stéphane Chollet (*Manager Innovation, Tools & Applications*)

June 20th, 2018 | Paris, France



AGENDA

✓ Regulatory context

- *GDPR & Data Transparency*
- *The challenge*

✓ Policy 0070 application

- *Guidance's proposed process*
- *Risk assessment / Threshold determination*

✓ Implementing a de-identification strategy

- *Special cares*
- *CSR's data encryption*

✓ Conclusion



AGENDA

- ✓ **Regulatory context**
- ✓ **Policy 0070 application**
- ✓ **Implementing a de-identification strategy**
- ✓ **Conclusion**





APPLICABLE REGULATION

Transparency



Regulations imply that individual data must remain available and exchangeable (or will soon become available)

[may. 2001] 1049/2001 EC: **Access to European** Parliament, Council and Commission **documents**



The policies say that we need to protect anonymity

[dec. 2010] *Policy 0043*: EMA policy on **access to documents** (related to **medicinal products for human and veterinary use**)

[jun. 2014] 536/2014 EC: “**Disclosure rules** describing the practical **implementation** of the **transparency** rules”

[oct. 2014] *Policy 0070*: mandatory **publication of clinical data** for **medicinal products** for **human use**

[apr. 2016] 2016/679 EC: Global Data Protection Regulation (**right to access the data**)

[oct. 2018] *Policy 0070 step 2* enforcement: **Publication** of all **Clinical Trials-related information**

[dec. 2000] 45/2001 EC: European Data Protection **Regulation**

[apr. 2001] 2001/20 EC: **Clinical Trials** must **follow GCP** → **patient anonymity**

[apr. 2003] 45 CFR 164.514: HIPAA* **privacy rules**

Rules of anonymization have been defined for the data exchange (HIPAA, Transcelerate, CDISC ...)

CSR cannot contain any patient identification information.

[apr. 2016] 2016/679 EC: Global Data Protection Regulation (**pseudonymization**)

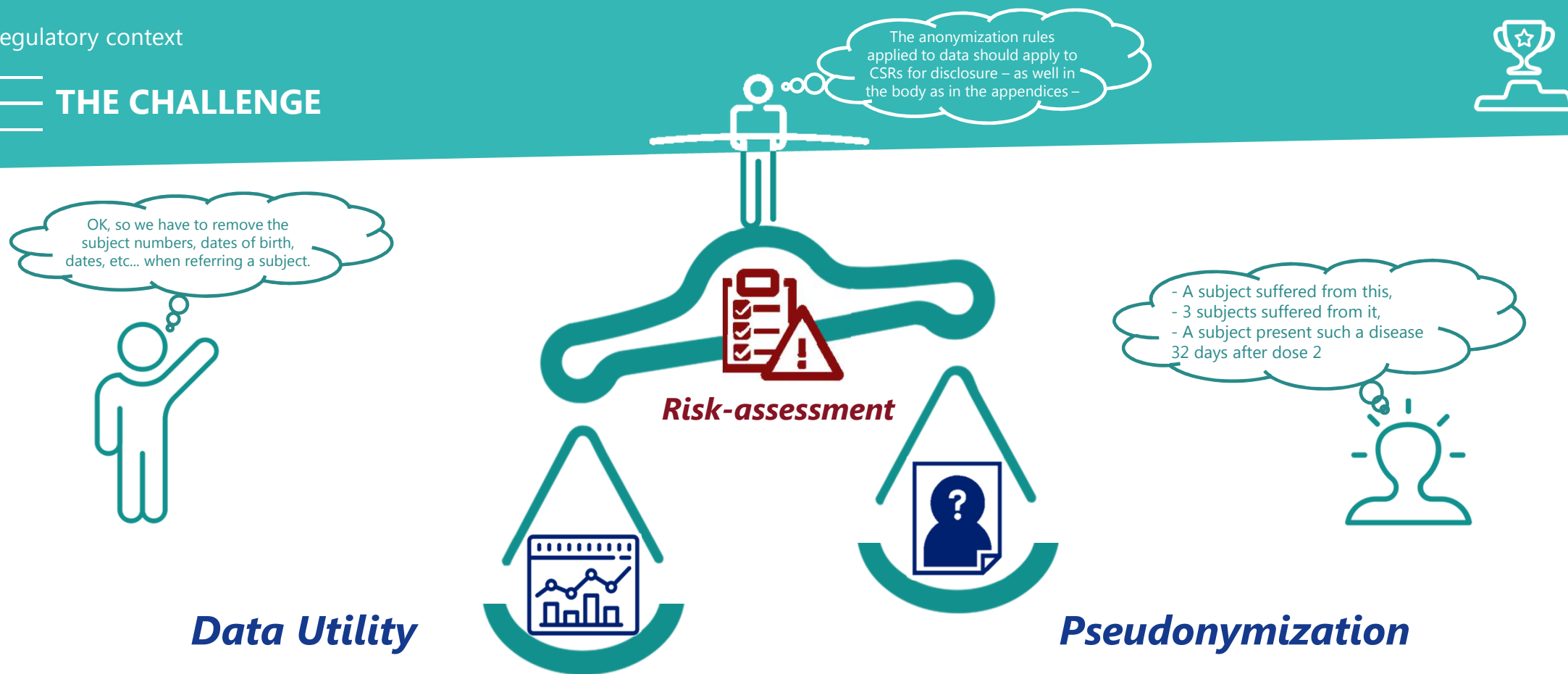
[may 2018] GDPR enforcement

In case of individual data disclosure, these data must be pseudonymized



Privacy

THE CHALLENGE



→ Find the **best compromise** between **data utility** & **data protection**

AGENDA

- ✓ Regulatory context
- ✓ **Policy 0070 application**
- ✓ Implementing a de-identification strategy
- ✓ Conclusion





MAIN FEATURES

Prerequisites:



Terms of use clarifying that users **shall not, IN ANY CASE, attempt to re-identify** trial participants or other individuals.



Data controller must **continuously follow** the **developments** in **re-identification techniques**, and if necessary **re-assess** the **risk** of **re-identification**.

Description:



Scope = **trial participants** & personal data in relation to **investigators, sponsors & applicants/MAHs**



Purpose = **remove values** of information allowing a **direct** or **indirect identification**



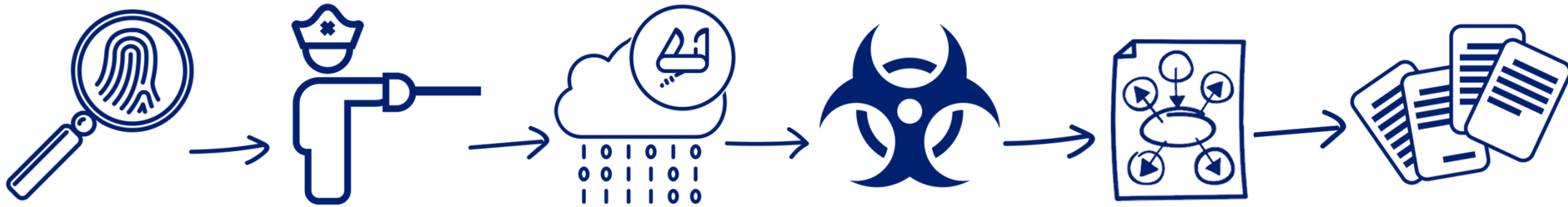
Main methods = **Masking, Randomization & Generalization** (*k-anonymity*), ...



Specific areas of interest = **Identifiers / Quasi Identifiers, Dates, Geographical** Locations



PROPOSED PROCESS



- 1- **Identifiers** determination
- 2- Identification of **possible adversaries** & **plausible attacks** on the data
- 3- **Data utility** considerations
- 4- Determining the **risk of re-identification** (*threshold & actual risk*)
- 5- **Anonymization methodology**
- 6- **Documenting** the methodology and the process

LET'S START WITH AN EXAMPLE



Can identify the user's detailed location from latitude and longitude.



When location is blurred, it becomes impossible to identify the user.

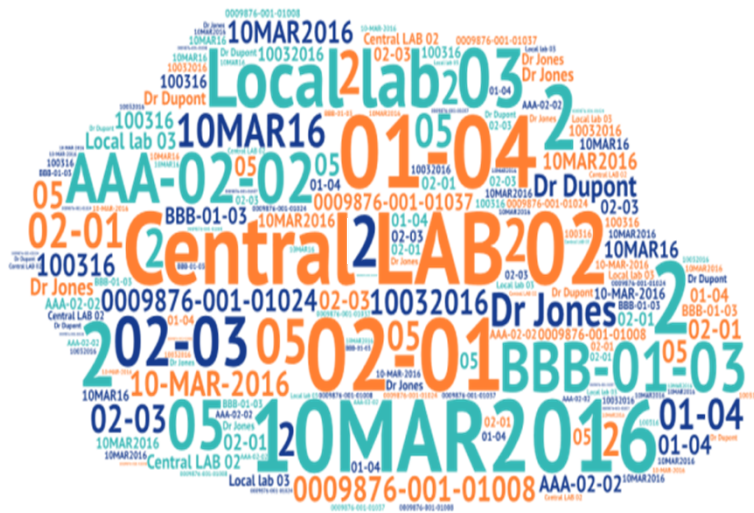


Which action between:
1- aggregating the data
2- remove the location?

User	Loc.
A	110°N 440°E
B	120°N 450°E
C	100°N 460°E

User	Loc.
A	~100°N ~450°E
B	~100°N ~450°E
C	~100°N ~450°E

Data has been aggregated instead of deleted to retain its utility



→ **Quasi-identifiers**

→ Combinations of items

f_j = distinct values patient frequency

- Do we have to **compute all the combinations** from **1C1 to (n-1)Cn***?
- **re-identification probability** \equiv **maximum probability** for a **patient to be re-identified** based on the information present on the data base: **$\max_{\text{re-id}} \mathbf{P} = 1 / \min(\mathbf{f}_i)$**



... AND ASSESS THE THRESHOLD ...

Acceptable re-identification risk (computed for the combinations of j identifiers):

$$reIDP_j = \frac{1}{cell\ size_j} \rightarrow \tau = \max_j reIDP_j$$

Cell size (τ value)	<3 (>0,33)	3 (0,33)	5 (0,2)	11 (0,09)	20 (0,05)
Confidence Level	Identifiable data	Highly untrusted data disclosure	Trusted data disclosure	EMA requirements	Highly trusted data disclosure

Special care:

→ Studies having **study population less than the required cell size**

(e.g. Phase I study in rare disease have 10 patients enrolled)

→ Number of meta data (e.g. CSR parameters) when transforming the data:

{70; 72; 75; 77; 80; 82; 85; 87; 92} + mean_{age} = 80

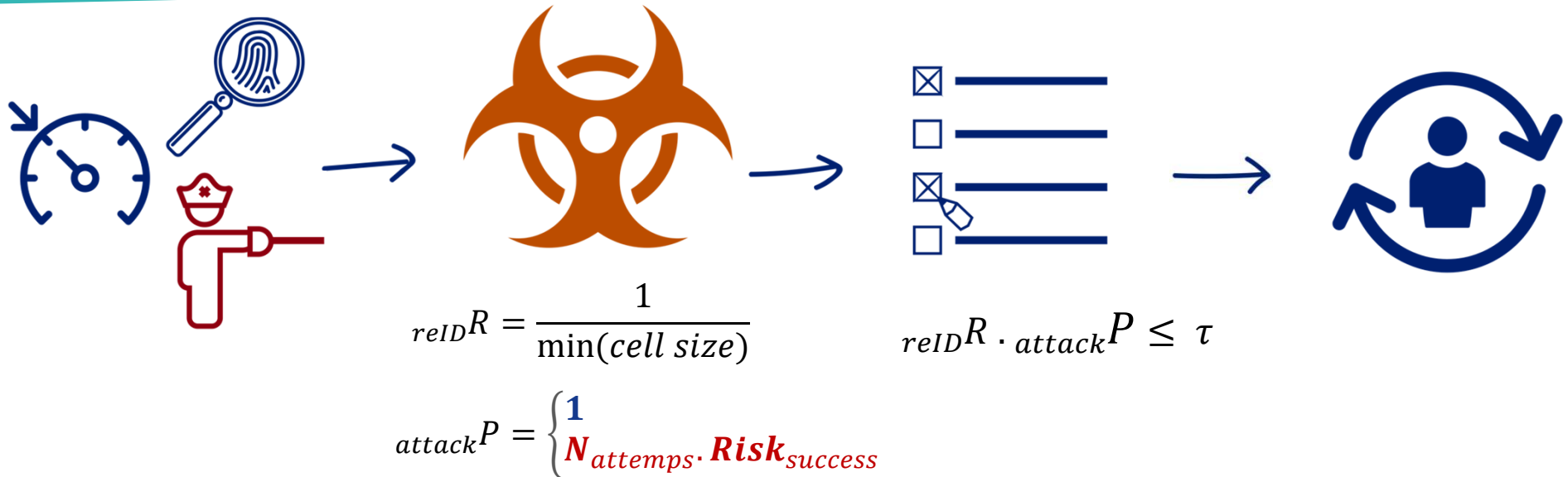
→ ~~{70; 72; 75; 77; 80; 82; 85; 87; >89}~~ + mean_{age} = 80

>89 = 80*9 - (70+72+75+77+80+82+85+87) = 92

→ {70; 72; 75; 77; 80; 82; 85; >86; >86} + mean_{age} = 80



... TO PERFORM THE RISK ASSESSMENT



Risk computed on the identifiers, hence the importance of the:

- **Identifiers determination**
- Disclosure **context** (*public** or *not***)
- **Available** re-identification **technologies*****

* *answering EMA Policy 0070*

** *answering EMA Policy 0043*

*** *periodic quasi-identifiers re-evaluation*

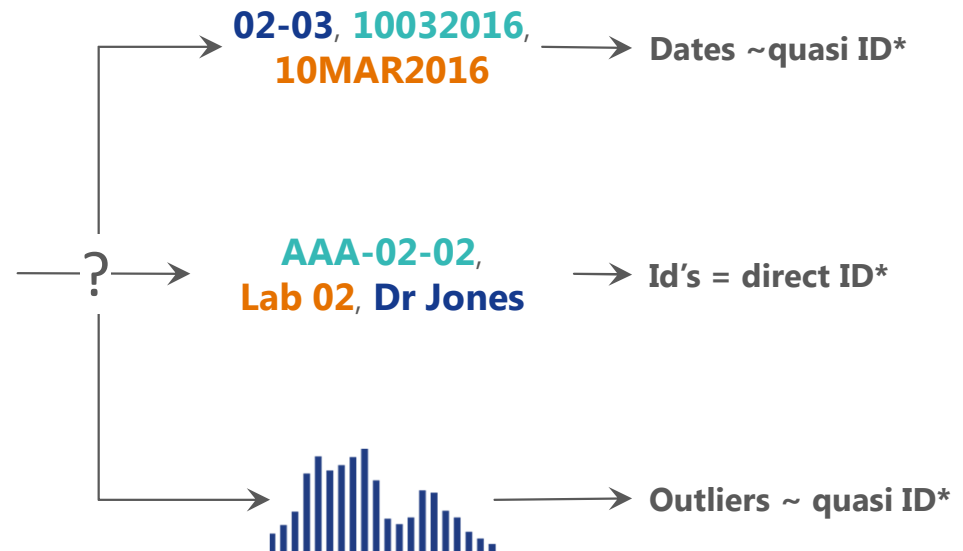
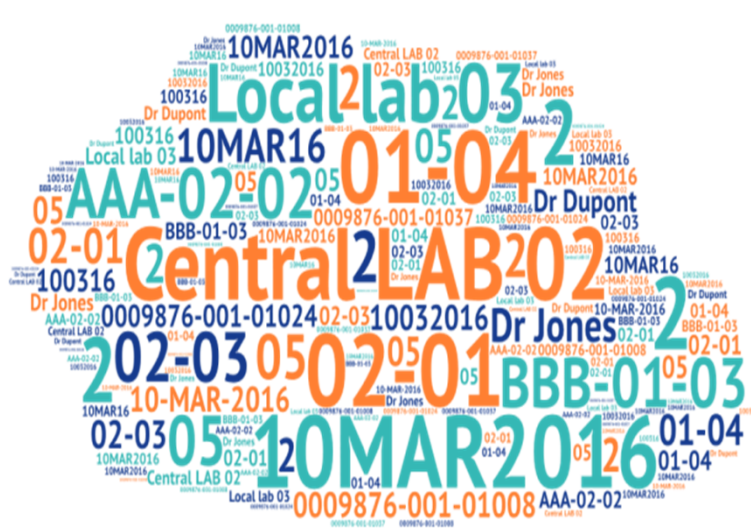
AGENDA

- ✓ Regulatory context
- ✓ Policy 0070 application
- ✓ **Implementing a de-identification strategy**
- ✓ Conclusion





IDENTIFIERS DETERMINATION



Identifying outliers:

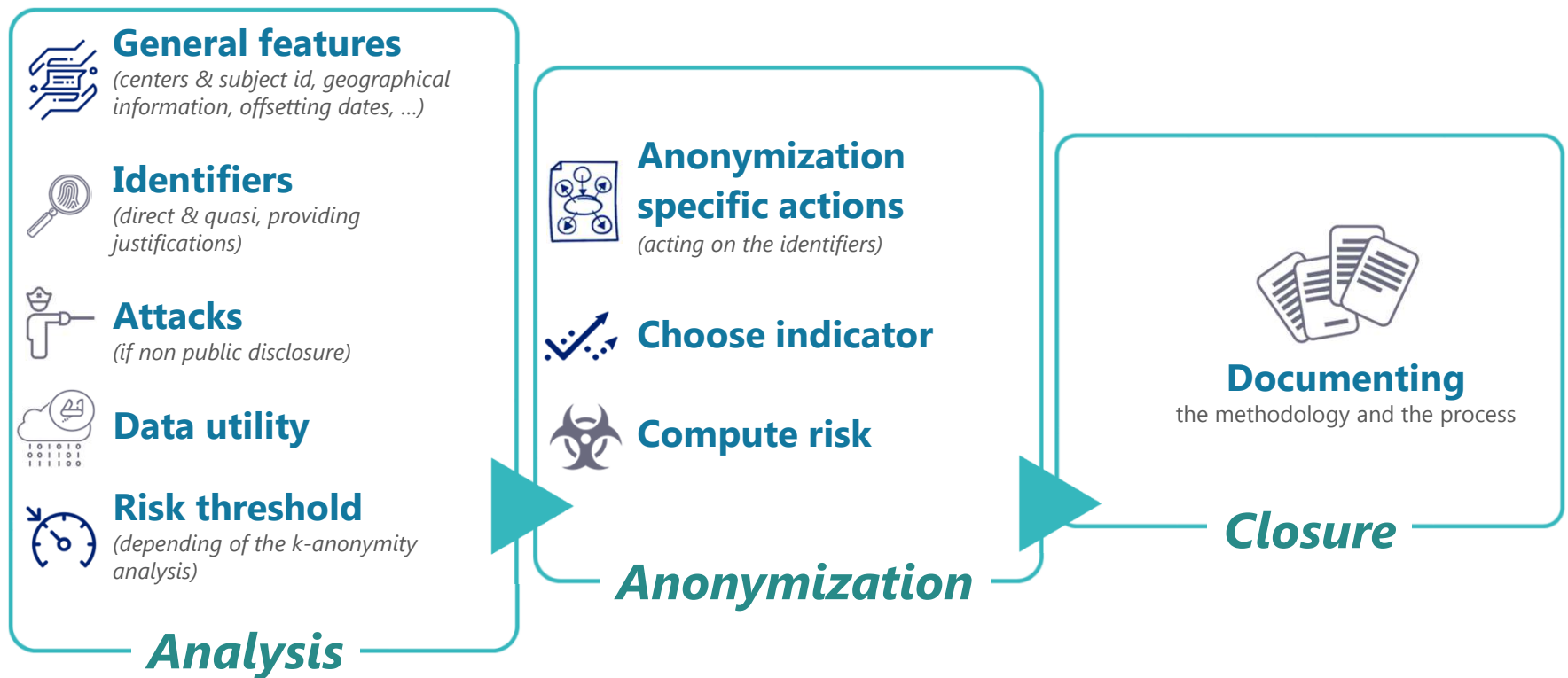
- Medical history (*rare disease*)
- Body Mass Index (<18 or >40)
- ...

Non-identifying outliers:

- Heart rate, Blood pressure
- Fasting glucose
- ...

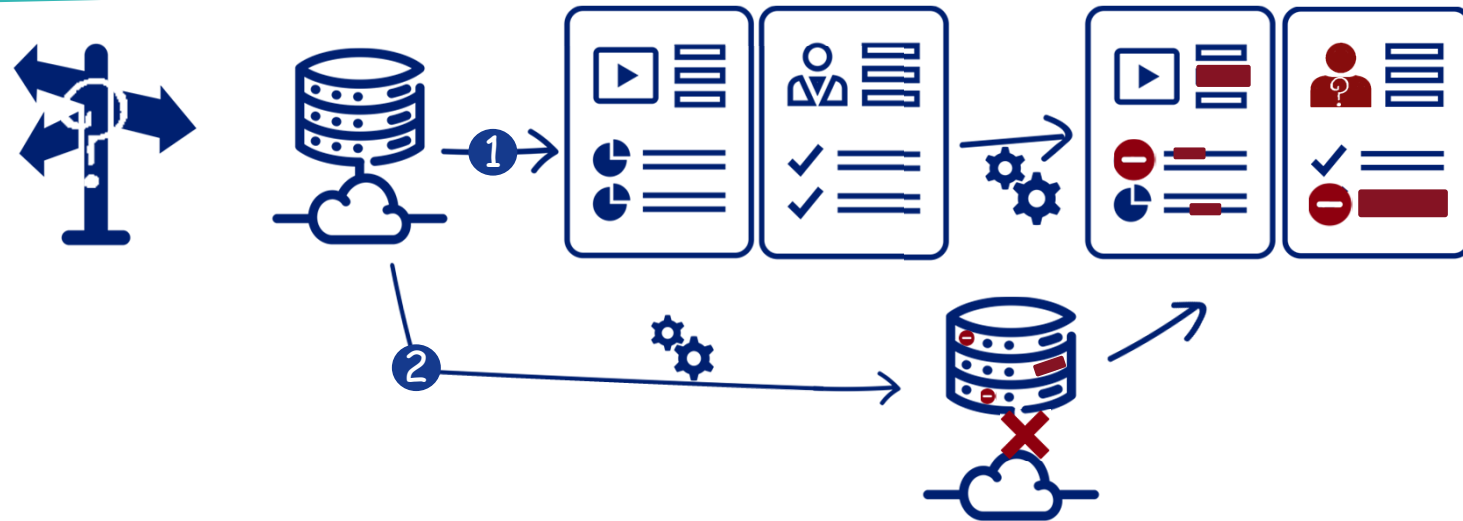


ANONYMIZATION PROCESS





WHICH WAY FOR THE CSR DATA ENCRYPTION?



- 1 **Quick** action, no need of the source data
- 2 **Consistent** information between CSR and source data

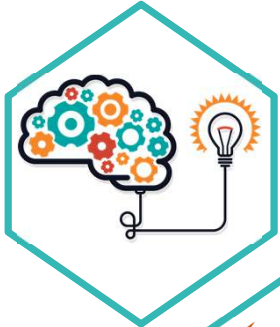


- 1 **Inconsistency** between source data and CSR information
- 2 **Need the programs** to regenerate the TFL's

AGENDA

- ✓ Regulatory context
- ✓ Policy 0070 application
- ✓ Implementing a de-identification strategy
- ✓ **Conclusion**





Take home messages

Policy 0043 → **reactive** demands, **private** disclosure ("on demand")

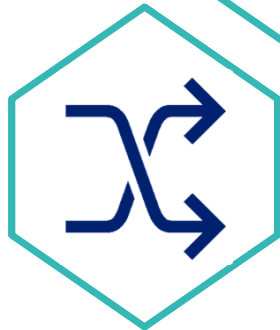
Policy 0070 → **proactive** demands, **public** disclosure



Risk assessment

Closely related to the determined **identifiers** (**direct** / **quasi**)

Valid for a **specific disclosure** and have to be **reassessed** in case of needs



Next step

Oct. 2018 - **Policy 0070** phase 2 (**patient's data** publication) enforcement



KEYRUS
biopharma