Danae Pla Karidi

ATHENA Research Center

AMNESIA

Data anonymization

amnesia

Danae Pla Karidi- 2021

@openaire_eu

Unlock the information

- **Research and studies need statistical information and properties**

- **Personal identification is not necessary in most fields**

- **Low reduction in data quality is tolerable**

  - **Or can be mitigated by using larger amounts of data**

- Anonymization unlocks the valuable information in data
  - The anonymized data are **different from the original data**
  - Anonymization is a one-way transformation of data
  - Original data cannot be retrieved

- Pseudo-Anonymization is not Anonymization
  - In Pseudo-Anonymized data there is a way to retrieve the original data
  - Pseudo – Anonymized data are still personal data
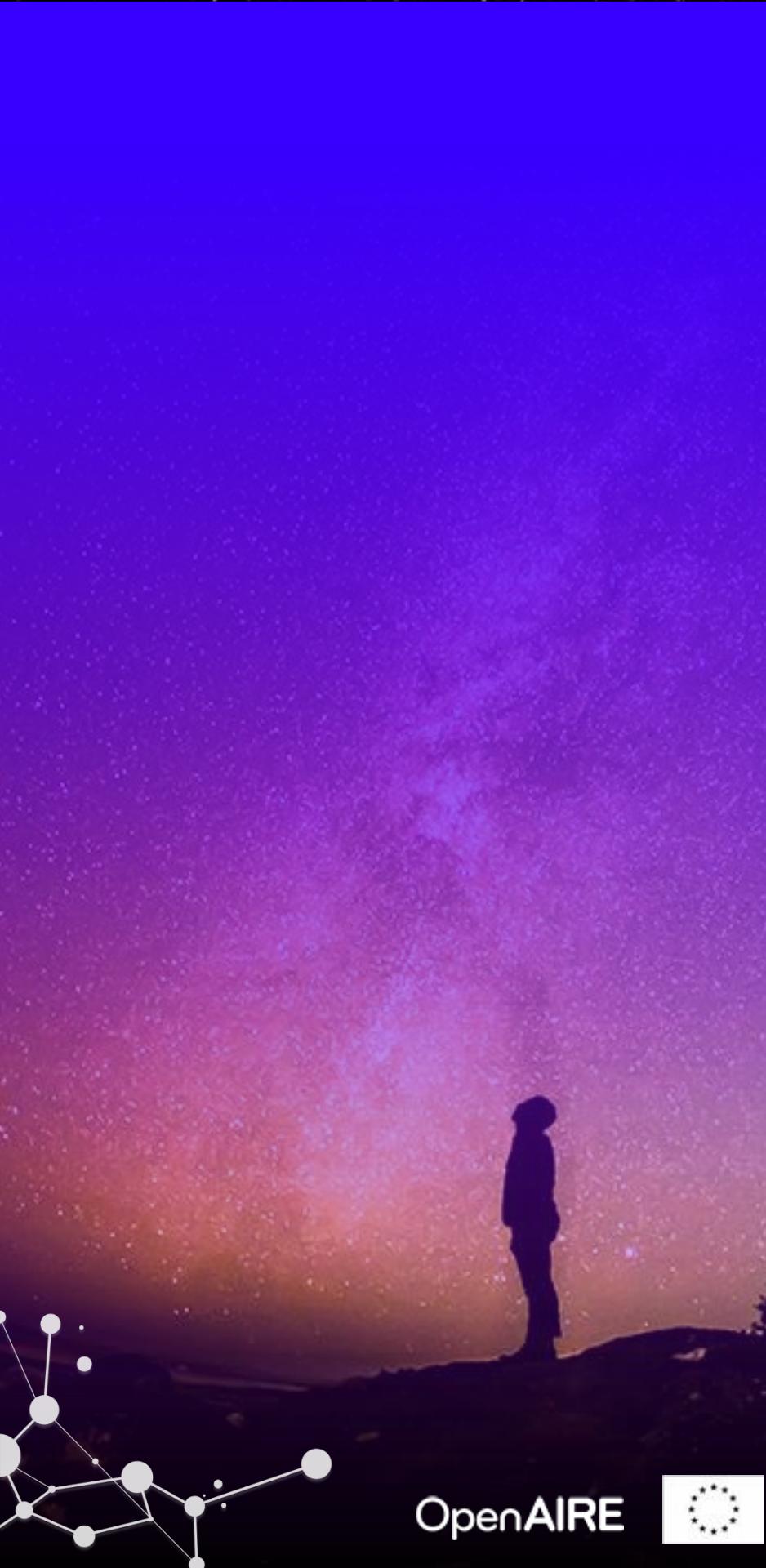
Why anonymize?

Anonymized data are outside the scope of GDPR

Anonymization provides a statistical guaranty about the risk of information leakage

It is the most suitable way to give information to third parties, without revealing personal data

k-anonymity

Each entry becomes indistinguishable from other *k*-1 entries

id	Zipcode	Age	National.	Disease
1	13053	28	Russian	Heart Disease
2	13068	29	American	Heart Disease
3	13068	21	Japanese	Viral Infection
4	13053	23	American	Viral Infection
5	14853	50	Indian	Cancer
6	14853	55	Russian	Heart Disease
7	14850	47	American	Viral Infection
8	14850	49	American	Viral Infection
9	13053	31	American	Cancer
10	13053	37	Indian	Cancer
11	13068	36	Japanese	Cancer
12	13068	35	American	Cancer

id	Zipcode	Age	National.	Disease
1	130**	<30	*	Heart Disease
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	1485*	≥40	*	Cancer
6	1485*	≥40	*	Heart Disease
7	1485*	≥40	*	Viral Infection
8	1485*	≥40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

OpenAIRE | NEXUS

Generalization Hierarchy
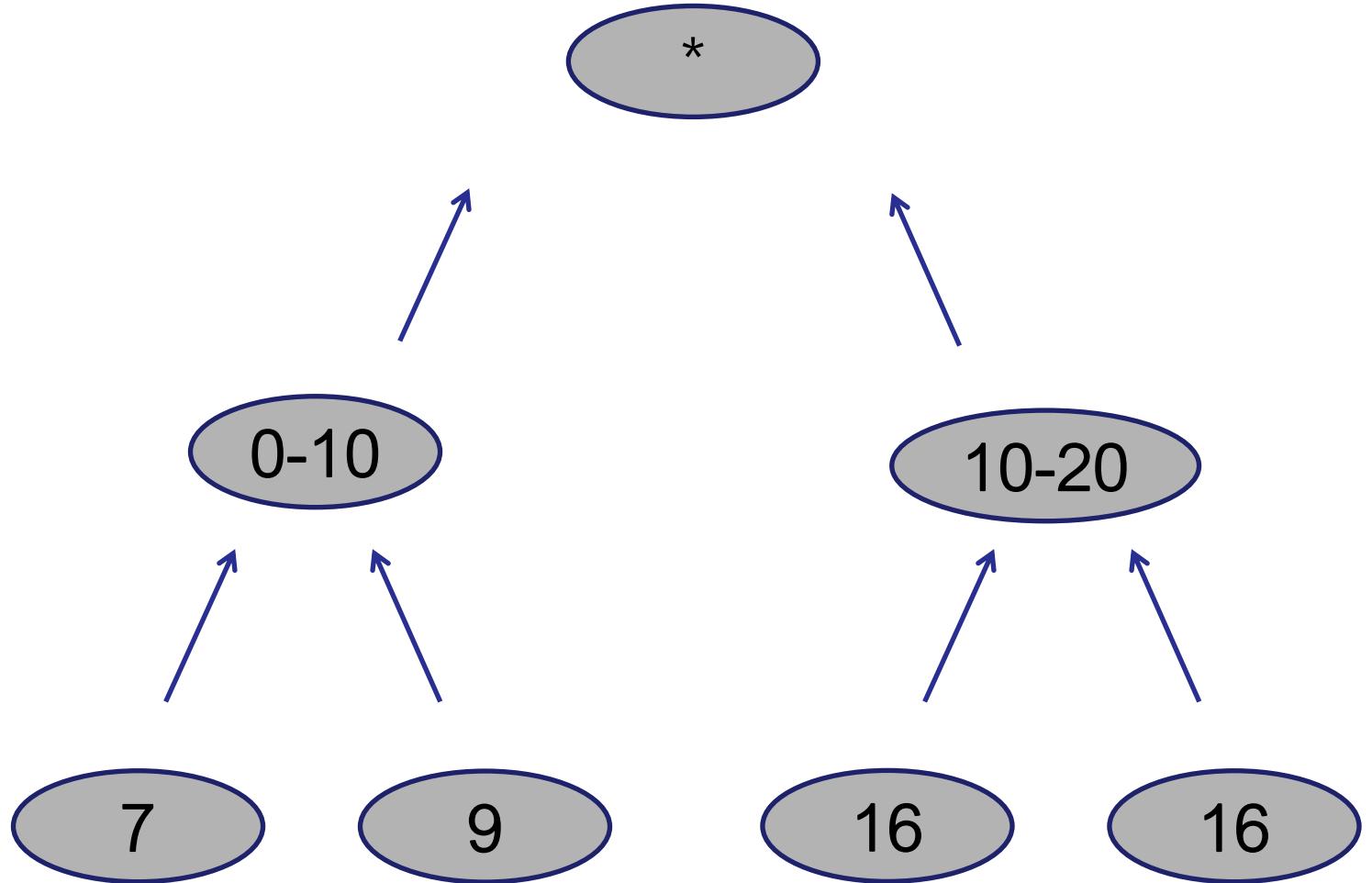
km-anonymity

	Fruits	Meat	Vegetables	Fish
Vassilis	X	X		
Manolis	X	X	X	
Eleni			X	
Maria		X	X	
Kostas	X			X

	Fruits	Meat	Other food
Vassilis	X	X	
Manolis	X	X	X
Eleni			X
Maria		X	X
Kostas	X		X

- 2^2-anonymous
- Any combination of m items will not appear less than k times

Why Amnesia

📱 User friendly

🛢 Works locally, no data transfer risk

👤 Allows users to customize the solution

🛠 The only tool to offer anonymization for set-valued data

⚒ The only tool to support k^m-anonymity

✓ Easy to incorporate to third party information systems

THANK YOU

Danae Pla Karidi

danae@athenarc.gr

https://amnesia.openaire.eu

@AmnesiaTool