

ΥΠΟΛΟΓΙΣΤΙΚΕΣ  
ΕΠΙΣΤΗΜΕΣ



VIRTUAL EVENT



**ΑΘΗΝΑ**

Έρευνα & Καινοτομία  
Τεχνολογίες Πληροφορίας

# Ημέρες Δράσης



23-24-25  
ΙΟΥΝΙΟΥ  
2021



ΕΥΦΥΗ  
ΒΙΟΜΗΧΑΝΙΚΑ  
ΣΥΣΤΗΜΑΤΑ

ΤΕΧΝΟΛΟΓΙΕΣ  
ΔΕΔΟΜΕΝΩΝ  
ΚΑΙ ΑΛΛΗΛΕΠΙΔΡΑΣΗΣ

# Amnesia



## Data Anonymization Made easy



**Manolis Terrovitis**

Senior Researcher

Information Systems Management Institute



**AONNA** Έρευνα & Καινοτομία  
Τεχνολογίες Πληροφορίας

Τα προαναφερθέντα έργα υλοποιήθηκαν στο πλαίσιο της Πρόσκλησης με κωδικό 031 με τίτλο «ΔΡΑΣΗ ΣΤΡΑΤΗΓΙΚΗΣ ΑΝΑΠΤΥΞΗΣ ΕΡΕΥΝΗΤΙΚΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΚΩΝ ΦΟΡΕΩΝ», που χρηματοδοτείται από το Επιχειρησιακό Πρόγραμμα «Ανταγωνιστικότητα, Επιχειρηματικότητα και Καινοτομία», ΕΣΠΑ 2014-2020 ΕΣΠΑ 2014-2020, με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης (Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης). Αναλυτικά: "Υπολογιστικές Επιστήμες και Τεχνολογίες Δεδομένων, Περιεχομένου και Αλληλεπίδρασης" (MIS 5002437) με 3 υποέργα: Υποέργο 1 «Γλωσσικές Τεχνολογίες Ανάλυσης Περιεχομένου και Αλληλεπίδρασης», Υποέργο 2 «Τεχνολογίες Ανάλυσης Περιεχομένου στον Πολιτισμό», Υποέργο 3 «ΜΕΔΑ: Μεγάλα Δεδομένα – Προκλήσεις, Μέθοδοι και Τεχνικές Αποδοτικής Διαχείρισης» και "Καινοτόμος εφαρμογή του βιομηχανικού διαδικτύου των πραγμάτων σε ευφυή περιβάλλοντα" (MIS 5002434) με Υποέργο 1: "Καινοτόμος εφαρμογή του βιομηχανικού διαδικτύου των πραγμάτων σε ευφυή περιβάλλοντα".



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

- GDPR limits the usage of personal data
  - according to law and contracts
  - Consent
  - Can be used for research
- Using Personal data
  - Consent might not be given or withdrawn
  - Difficult to manage
  - Usage for research purposes requires strict internal processes
  - Cannot share with third parties

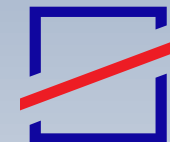




# Unlock the information



- **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



**ATHENA** Research & Innovation  
Information Technologies

- 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





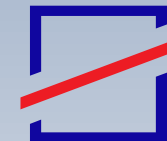
# Why Anonymize?



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



**AOHNA**

Έρευνα & Καινοτομία  
Τεχνολογίες Πληροφορίας



# Limitations of Anonymization



Anonymized data have lost some information

- The key idea of a good anonymization algorithm is to minimize this loss and limit it in the least important information

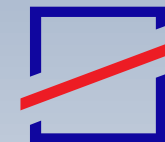
There are gray boundaries between anonymized and pseudo-anonymized data

Formal privacy guarantees provide a statistical guaranty for the anonymized data

- This is only an interpretation of the notion of “privacy”

It cannot easily be fully automated

- User input is needed



**AOHNA**

Έρευνα & Καινοτομία  
Τεχνολογίες Πληροφορίας



# When to anonymize



When you are a practitioner, and you want to share data with researchers and third parties without compromising the privacy of the user

- After the data is anonymized, you do not need consent

When you want to give data to recipients you do not fully trust

- Encryption will reduce the risks of data leaks to unauthorized third parties, it will do nothing for untrusted recipients

When you want to openly publish data and you are not fully aware of the audience

When reduction in information quality is acceptable





# Why Amnesia



User friendly



Works locally, no data transfer risk



Allows users to customize the solution



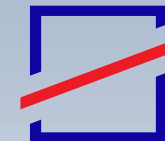
The only tool to offer anonymization for set-valued data



The only tool to support  $k^m$ -anonymity



Easy to incorporate to third party information systems



**AOHNA**

Έρευνα & Καινοτομία  
Τεχνολογίες Πληροφορίας

# Statistics

<https://amnesia.openaire.eu>

68k visitors

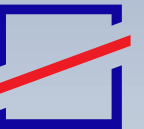
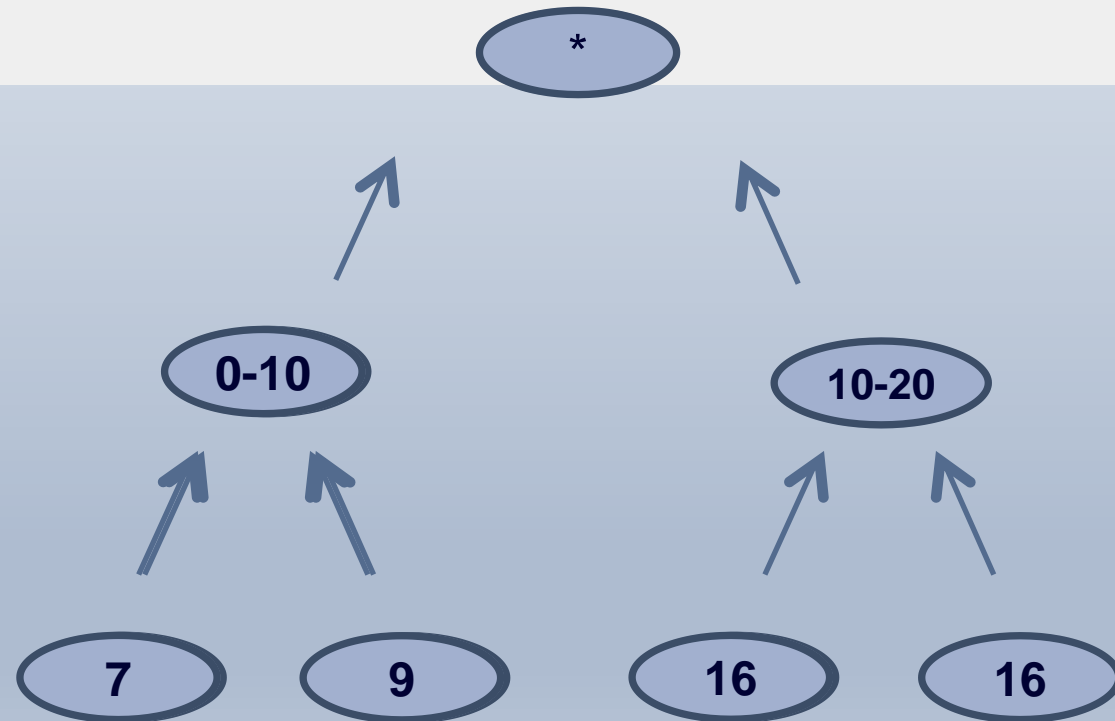
207k page views

4k unique downloads

40k Uses of the on-line service



# Generalization Hierarchy



# Amnesia Limitations

- **Users are not familiar with anonymization techniques**
- **The process is novel and requires effort from the user's part**
- **Amnesia cannot decide on privacy parameters**
- **K-anonymity does not protect from every type of attack**



Manolis Terrovitis

[mter@athenarc.gr](mailto:mter@athenarc.gr)

<https://amnesia.openaire.eu>



VIRTUAL EVENT

# Ημέρες Δράσης

23-24-25  
ΙΟΥΝΙΟΥ  
2021



Thank you!