

Amnesia - Anonymize your data before publishing

- **What it is**

[Amnesia](#) is a flexible data anonymization tool that allows to remove identifying information from data. Amnesia transforms relational and transactional data to anonymized datasets where formal privacy guaranties hold. It does not only remove direct identifiers like names, SSNs, etc., but also transforms secondary identifiers like birth date and zip code so that **individuals cannot be identified in the data**, by linking them to other sources of information.

- **Technical Requirements**

Amnesia is a web based or desktop application that uses a Java backend to implement the data anonymization algorithm. It runs both in Windows and Linux and requires Java version 8 or greater.

- **What it does**

Amnesia implements data anonymization techniques from the field of Privacy Preserving Data Publishing (PPDP). The key idea in anonymization is that identifying information is removed from the published data, so that sensitive information cannot be attributed to a person.


Anonymization techniques present descriptive information in an obscure or generalized way, to guarantee that such associations cannot take place. Anonymization algorithms transform the data into a form that provides a privacy guarantee with the minimum possible distortion of the original data. A significant challenge for every anonymization method is to provide the best trade-off between privacy guaranty strength and anonymized data quality.

Amnesia supports k-anonymity and k^m -anonymity, two privacy guaranties that make each record indistinguishable from other $k-1$ records. Original data are anonymized by using generalization and suppression. Generalization is the substitution of a value in the original file, e.g., "Athens" with a more abstract one, e.g., "Greece". Substitutions take place according to a predefined hierarchy of values, e.g., "Athens" < "Greece" < "Europe", which is user defined or it can be automatically created by the tool.

Amnesia offers algorithms that take advantage of modern hardware architectures, that feature multiple computer cores. Moreover, Amnesia allows the user to guide the anonymization process by visualizing the candidate solutions and allowing the user to choose and customize the most convenient one.

ID	Age	Zipcode	Diagnosis
1	28	13053	Heart Disease
2	29	13068	Heart Disease
3	21	13068	Viral Infection
4	23	13053	Viral Infection
5	50	14853	Cancer
6	55	14853	Heart Disease
7	47	14850	Viral Infection
8	49	14850	Viral Infection
9	31	13053	Cancer
10	37	13053	Cancer
11	36	13222	Cancer
12	35	13068	Cancer

k-anonymization



ID	Age	Zipcode	Diagnosis
1	[20-30]	130**	Heart Disease
2	[20-30]	130**	Heart Disease
3	[20-30]	130**	Viral Infection
4	[20-30]	130**	Viral Infection
5	[40-60]	148**	Cancer
6	[40-60]	148**	Heart Disease
7	[40-60]	148**	Viral Infection
8	[40-60]	148**	Viral Infection
9	[30-40]	13***	Cancer
10	[30-40]	13***	Cancer
11	[30-40]	13***	Cancer
12	[30-40]	13***	Cancer

Original and anonymized dataset with k=4

The user can explore the quality of the data with ad hoc queries and a visual representation of the value distribution. He can choose to suppress outliers in order to avoid unnecessary generalization of the original data and information loss. Amnesia focuses on usability and flexibility to allow the user to understand and guide the anonymization process. Since anonymization methods have not been extensively used in practice, it is essential that users will be able to tailor the anonymization processes and especially the information loss in the anonymized data to their needs. For more details, you can check our [documentation](#).

- **How can I use it?**

Amnesia is available both as an [online service](#) and as a [local application](#).

- **More information**



Data anonymization made easy

Amnesia: <https://amnesia.openaire.eu/>