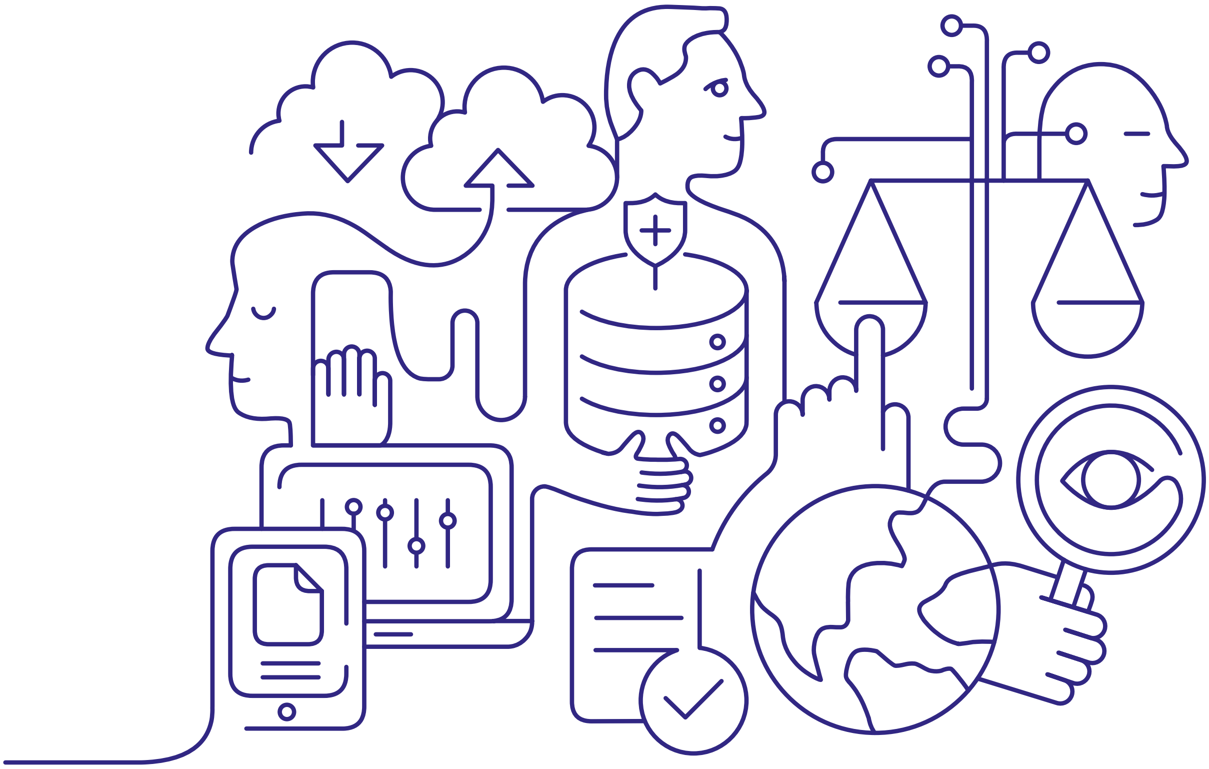
PRIVITAR

August 2019

Privitar Lens at UCAS Case Study



UCAS

PRIVITAR

Introduction

UCAS, the UK's Universities and Colleges Admissions Service, provides insights about admissions to higher education to contribute to public debate about education, access and social mobility. UCAS offers a wide range of data products and services through their public website. Each year, UCAS publishes nearly three million aggregate data points in a variety of formats to download for public use.

> *UCAS, as a thought leader in data privacy, is aware of the need to respond to these new types of threats to maintain the confidentiality of the data it is safeguarding.*

The data that UCAS works with is sensitive, personal information about student applicants. As an independent charity, UCAS is committed to use this information in a responsible way. Publishing this information poses challenges for managing the privacy risks resulting from aggregate data releases. Even though data aggregation itself partially mitigates the risk of re-identification, it can still leave the data vulnerable to differencing and reconstruction attacks that lead to the exposure of sensitive information about individuals.

New data privacy threats

Traditionally, statistical disclosure control techniques such as data swapping, cell perturbation, rounding and small count suppression are used to prevent the disclosure of individuals' information from aggregate data. Often, released datasets are coarsely aggregated due to the growing awareness of the privacy risks that arise from publishing data with high granularity.

In addition, manual checks by human privacy experts are usually carried out to ensure the confidentiality of data publications. However, recent research has shown that in many cases even the combination of these controls is insufficient in providing protection against more sophisticated privacy attacks such as differencing and reconstruction attacks.

UCAS, as a thought leader in data privacy, is aware of the need to respond to these new types of threats to maintain the confidentiality of the data it is safeguarding. UCAS always seeks to modernise its disclosure control methodology and to explore alternative privacy-enhancing technologies, such as differential privacy, to protect against increased risk. As an organisation committed to the open-data promise, UCAS hopes that the use of new techniques will permit stronger guarantees on the privacy of any data published and, in future, to extract even more value from data than current privacy restrictions allow.

Lens and differential privacy

Lens is Privitar's technology for producing privacy-preserving aggregate data releases, made available to consumers as reports, cross-tabulations, or interactive dashboards. Lens provides strong protection against privacy breaches for aggregate data releases derived from highly sensitive datasets.

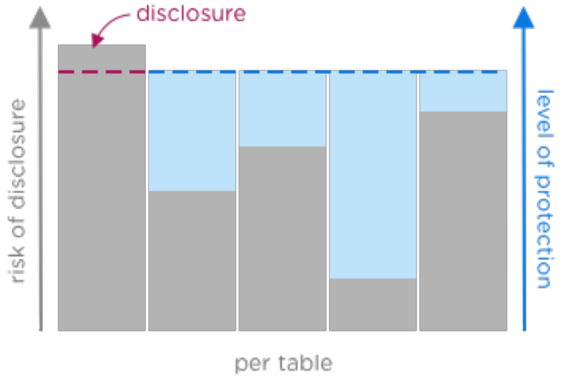
Lens addresses the risks of differencing and reconstruction attacks through the use of differential privacy techniques and state-of-the-art automated risk analysis. It helps data controllers gain confidence in the protections applied by providing a thorough and scalable analysis of potential privacy vulnerabilities in a planned release and demonstrating how any such attacks are defended against.

In a proof of concept project, Lens technology was used to produce a privacy-protected release of aggregate statistics from sensitive UCAS data. In a controlled session on-site at UCAS, Privitar used an extract from UCAS's undergraduate applications database together with a description of the required aggregate information to produce a result similar to a part of UCAS's annual End of Cycle report. Privitar, in conjunction with UCAS, identified sensitive variables in the data and defined a set of privacy criteria that must be met for aggregate data to be considered sufficiently protected.

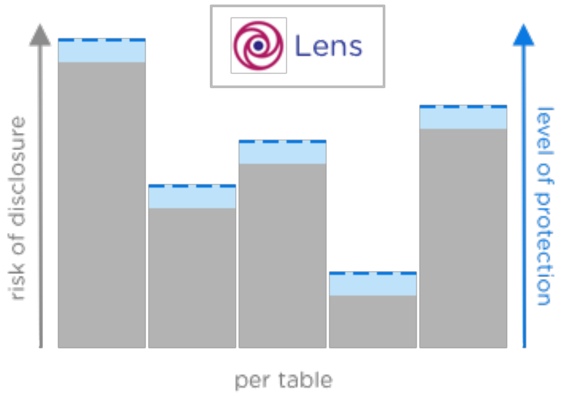
Using Lens technology deployed on internal UCAS infrastructure, a dataset of 600,000 applicants was successfully processed within a secure environment according to the report specification, and the collection of privacy-protected statistics was produced. Lens' automated privacy risk assessment detected potential vulnerabilities in the statistical release through targeted differencing attacks. These attacks were defended against through the addition of differentially private noise to the published aggregates.

Lens' built-in privacy analysis engine calibrated the magnitude of the differentially private noise in a data- and risk-dependent manner. It chose the appropriate level of distortion so that all vulnerabilities that were found were prevented from accurately reconstructing individual confidential values. In this way, strong data protection was achieved with minimised impact on the accuracy of the aggregates.

Compared to blanket approaches such as small count suppression or rounding, Lens' disclosure control mechanisms can achieve higher levels of protection while preserving more accuracy (see figure). This is because 'global' risk mitigation methods that do not take into account the specific data being published need to make conservative assumptions about the risk of successful privacy attacks. Such worst-case assumptions often result in a larger loss of accuracy than necessary for some parts of the data, while failing to defeat more powerful attacks.



'Global' risk mitigation methods add a sub-optimal amount of distortion and may still miss privacy vulnerabilities.



Lens's principled protection is specific to the data. All vulnerabilities are addressed with precisely set noise.

In a second step, it was investigated whether more granular information could be extracted without compromising data confidentiality or quality. Additional dimensions were added to the original specification. Lens reported additional vulnerabilities in the higher-resolution data and suggested increased levels of noise addition to mitigate against the newly detected risks. However, a large proportion of the added statistics was not significantly impacted by the noise addition.

Conclusion

In summary, the proof of concept demonstrated how Privitar Lens technology could add benefit to UCAS's existing data release process. Lens detected potential privacy vulnerabilities in a release similar to UCAS's aggregated data releases, based on real data, and automatically protected against these risks using differential privacy techniques.

Lens' automated risk analysis would help UCAS, as a data guardian, to gain confidence in the protections applied and to further increase trust with data subjects that their data is being used responsibly. At the same time, Lens' data- and risk-dependent approach to privacy risk mitigation could improve data quality and enable UCAS to strike the right balance between publishing valuable aggregate information with high granularity, while applying only the minimum distortion needed to prevent privacy attacks.

We are Privitar

We create software designed for enterprise-wide privacy protection.

Privitar's technology provides organisations with centralised data privacy governance and controls that can be easily deployed at scale, to create a contextual, policy-based privacy layer that works at the data level. Additionally, Privitar's software features an innovative combination of data privacy capabilities to enable enterprise-wide access, use and distribution of valuable privacy-enhanced data.

Get in touch.
Visit www.privitar.com or contact info@privitar.com