PRIVITAR

UCAS

# Supporting an evidence-based approach to data access requests

PRIVITAR.COM

# 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.

It offers a wide range of data products and services through its public website. Each year, UCAS publishes nearly three million aggregate data points in a variety of formats to download for public use.

The data 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.

UCAS has a progressive and forward-thinking approach to data privacy, and is aware of the need to respond to these new types of threats to maintain the confidentiality of the data it safeguards.

As a team of privacy risk experts, Privitar Labs partnered with UCAS to assist them in managing data releases, providing objective evidence founded in privacy best practices and advanced privacy risk assessment techniques.

# Answering data requests with informed decisions

## UCAS balances support for research and innovation with their commitment to safeguard the privacy of individuals.

When releasing data, UCAS are driven by two competing factors: their uncompromising commitment to safeguarding the privacy of their service users, and their ability to enable high-quality, innovative research that drives positive change. UCAS's research remit is to enhance public understanding of patterns in higher education, and promote equality of access to higher education.

Consideration of inherent risk within a data release is a primary component of privacy best practice. Accurate and considered responses to this risk allow organisations such as UCAS to make informed decisions on how to appropriately handle data and respond to the associated risks. Privitar's expertise with risk analysis in aggregate statistics empowered UCAS to take an evidence-based approach to managing data access requests.

Additionally, UCAS leveraged Privitar's privacy risk analysis expertise to confidently evaluate whether statistical disclosure controls could be applied to the dataset to enable safe data releases.

# Assessing the risk of releasing data

## Privitar's risk analysis tools support informed decisions by identifying privacy vulnerabilities in large datasets.

This project focused on risk analysis for a specific data request. UCAS was asked to release tabular data based on a slice of the 2019 end of cycle data, considering only UK applicants aged 18. The candidate data release would contain statistics aggregated from more than 200,000 unique applicants and broken down by a number of dimensions, including those deemed by UCAS to be sensitive, such as gender, ethnic group and attainment scores.

The granularity of the requested statistics raised immediate concerns for UCAS. Their existing heuristic measures determined that the dataset was too disclosive to release. However, UCAS lacked direct evidence to support this decision, leaving them open to criticism for being too cautious in the balance between safeguarding applicant data and fulfilling their commitment to open data.

Privitar's risk analysis tools use a standardised, automated process to analyse large datasets and identify privacy vulnerabilities to differencing and reconstruction attacks. Such attacks allow sensitive information about individuals to be discerned from aggregate statistics.

The candidate data release was analysed in a secure environment using Privitar's risk analysis software and was considered in isolation. This approach allowed us to demonstrate both the reconstruction risk inherent solely to the candidate release, and also highlight that the true risk of reconstruction was likely to be even greater in practice. UCAS already publicly releases sets of tables derived from the same underlying data source, such as their 2019 end of year cycle report. The combined reconstruction risk of the candidate release plus these public tables is likely to be higher than the candidate release considered alone. A second version of the candidate release was also analysed. In this case, UCAS had applied the traditional statistical disclosure control method of rounding the aggregate statistics to the nearest 5 counts.

# An accurate picture of vulnerabilities

Privitar's risk analysis technology discovered multiple vulnerabilities in the candidate data release. The fine granularity of the requested dimensions meant that aggregation of the data alone did not protect the applicants' sensitive information. Without further protection, the ethnic group of 8,310 applicants, the gender of 1,847 applicants, and the attainment scores of 8,702 applicants could be reconstructed from the aggregate statistics.

Moreover, the investigation found that applying the statistical disclosure control of rounding to the nearest 5 counts was not sufficient to protect all of the vulnerabilities. Rounding reduced the number of vulnerabilities, but over 20% of vulnerabilities remained for each sensitive attribute.

Further sources of risk were also identified. Most importantly, the UCAS 2019 end of year cycle report had already been published, containing high level aggregate statistics derived from the same data source as the candidate release. It is very likely that if the candidate release were to be provisioned in parallel to the UCAS 2019 end of year cycle report, these two releases taken together would present a higher risk of disclosure. The analysis also underestimates potential privacy threats from actors with access to detailed background information about applicants. For example, an actor with full knowledge of the data from a particular school could subtract that school's data from the aggregate statistics, potentially revealing further vulnerabilities for other applicants.

| Attribute | Number of vulnerabilities detected | |
|---|---|---|
| | Raw statistics | Rounded statistics |
| Ethnic Group | 16,749 | 5,071 |
| Gender | 3,400 | 976 |
| Attainment scores | 12,996 | 2,785 |

# Conclusion

This project demonstrated that Privitar's automated risk analysis tools can be used to support UCAS's risk assessment process for a candidate data release. As one component of risk assessment, a quantitative analysis of privacy vulnerabilities helps data controllers make evidence-based decisions in response to data requests.

For a specific data request, Privitar's risk analysis demonstrated that aggregation of the data was not sufficient to protect the privacy of applicants. Sensitive information belonging to more than 4% of applicants would be made vulnerable by the candidate release. Moreover, the traditional statistical disclosure control of rounding the aggregate statistics did not protect these vulnerabilities. Our method of analysis further empowered UCAS in their decision to withhold the data, as it also allowed us to highlight the additional risk of the candidate release when considered with public releases from the same source dataset, such as the 2019 end of year cycle report.

UCAS's non-disclosure decision was legitimised, with privacy risk analysis evidence at the centre, demonstrating to the data requester, the UCAS board of trustees, governance teams and, ultimately, UCAS service users that privacy is paramount in UCAS's undertakings.

# About Privitar Labs

## The practical application of Privacy Enhancing Technologies (PETs) is advancing rapidly.

Many promising new techniques are emerging, but the challenge lies in matching these to the right business use cases, and developing them into products that are easy to understand and to use correctly, at scale. This is Privitar's mission.

Within Privitar Labs, we're driving the creation of practical solutions using PETs. We work in close partnership with our strategic customers to apply leading privacy techniques to enable new uses of data.

Privitar is the leader in modern data provisioning, empowering organizations to maximize business value by using data safely, with speed and at scale. We make data highly accessible, through the use of privacy enhancing technologies, so organizations can optimize business and customer outcomes. Only Privitar has the right combination of technology and expertise to create a safe data provisioning ecosystem to enable clients to share data and unlock new data insights while keeping data safe and businesses compliant. Founded in 2014, Privitar is headquartered in London, with regional headquarters in Boston and locations throughout the US and Europe. For more information, please visit www.privitar.com

PRIVITAR.COM