Pseudonymity:
An Answer to Assessment Privacy Concerns?

Learnosity

caveon

Contents

This paper is jointly prepared by Learnosity and Caveon to explain what pseudonymous personal data is and how it can be useful in the field of assessments. It was written to support a session of the same name at the ATP Innovations in Testing conference 2022.

Learnosity and Caveon are independent entities and no partnership is implied by this document.

This document is not legal advice; it should be regarded as general information only and you should obtain your own legal advice for your own circumstances.

Introduction

Pseudonymity is a way of storing electronic data where names or other information to identify a person are stored separately from the data about them. For example, an assessment result could be associated with a numeric ID representing the person who took the assessment rather than with the person's name. When data is pseudonymized, there is a separate index that allows matching the numeric ID to the name, stored separately.

A key benefit of pseudonymity is that if there is a data breach or other leakage, the data leaked may not include information that actually identifies the people involved. Pseudonymity can reduce the risks involved with processing personal data and often strikes a good balance between allowing data to be used and protecting people's privacy.

This document explains what pseudonymous data is, explores the benefits to the assessment community of pseudonymity in the processing of assessment results and other personal data used in testing and examinations, and describes how Caveon and Learnosity utilize pseudonymous data in the provision of their respective services.

Identified, pseudonymous, and anonymous data

Let's start by defining some commonly used terms:

- **Personal data** (also called personal information) is often defined as being any data or information related to an identified or identified natural person.
- **Anonymous data** is data that does not relate to an identified or identifiable person, either because identifying information was not captured in the first place or if it has been anonymized or de-identified with the intent that the data cannot be associated with any person again. Anonymous data needs to be unidentifiable[1].
- **Pseudonymous data** is data associated with a particular person, or persons, where additional information is needed to identify the specific people. Often this is created by replacing someone's name with a system generated ID or reference number, where the key to associate the ID to a person is held separately. The separately held information also needs to be kept secure to prevent it from being used to identify individuals.

The table below shows three examples. A common example of personal data in the assessment context is a list of names of people and their scores in an assessment.

The left column shows the full personal data — the name and score achieved.
The middle column shows anonymous data — essentially just a list of scores.
The right-hand column shows pseudonymous data — the names of people have been replaced with IDs.

Personal data		Anonymous data	Pseudonymous data	
Jose Singh	95%	95%	123456	95%
Ann Brown	85%	85%	364611	85%
Yasmin Patel	75%	75%	837135	75%
David Lazarus	70%	70%	809184	70%
James Mahon	65%	65%	634114	65%
Melinda Moss	62%	62%	981741	62%
Max Gutner	61%	61%	908173	61%
Raymond Baron	60%	60%	761045	60%

[1] It's outside the scope of this document, but readers should be aware that creating real-world anonymousdata is very difficult. There is a risk that by looking at details of the remaining fields (e.g. timestamps, patterns of activities) that investigation can piece together who the individuals are.

Identified, pseudonymous, and anonymous data (Cont.)

In most assessment use cases, it is important to be able to associate particular data with the people that generated the data. Data cannot be anonymous as you need to know who passed or failed a test so you can take the appropriate action(e.g., give a certificate or provide notification of failure to pass).

However, making data pseudonymous is a useful measure with assessment data. It still allows data to be associated with people when needed, but identity is masked for other processing. For a lot of tasks, pseudonymous data is sufficient and personal data from which an individual's data can be identified or is readily identifiable is not needed to achieve the organization's purposes.

A well-established example is the manual grading of essays. It's common practice to mask the name of the test taker to graders so they will not be influenced by any knowledge of the test taker, but, of course, the system requires the identity of the test taker to be able to assign the score in the master records.

The European GDPR law advocates pseudonymization and says[2]:

> **"The application of pseudonymization to personal data can reduce the risks to the data subjects concerned and help controllers and processors to meet their data-protection obligations."**

Let's move on to look at the security and legal compliance benefits of pseudonymization for personal assessment data.

[2] Within recital 28

How pseudonymous data practically increases security

When considering the security of assessment data, it is helpful to identify risks and threats to data and potential countermeasures. A common approach to analyzing security risk is to consider the impact and probability of each potential risk to the confidentiality, integrity, and availability of data and put in place measures to reduce the impact and probability.

In general, pseudonymization reduces the impact or consequences of many risks, for example:

- **Impact of data breach.** A security risk that concerns all organizations is data being leaked or otherwise exposed in a breach, for example on the Internet. However, if data is pseudonymous, this is much less of a loss to confidentiality than if fully identified data is leaked. That is because unless the breach also contains the index of IDs to names, it will be difficult or impossible to identify the real people behind the leaked pseudonymous data, and, as such, the leakage or breach is much less of a concern for individuals.

- **Insider risk.** To run their operations, most organizations need to give employees some level of access to personal data. Insider risk is where such a trusted individual breaks trust and accesses or uses the data inappropriately, such as for nefarious or otherwise prohibited purposes. Providing most employees that need access to data with pseudonymous data only reduces this risk and the associated impact.

- **Use of processors.** It's common in assessment programs to use several different service providers for delivery, scoring, and analysis of assessments. In privacy parlance, these companies are commonly referred to as "processors." The risks of security problems are obviously increased the more organizations have access to and process data. However, many processors and sub-processors do not need to know the identity of test-takers and can perform the processing required to deliver their services using pseudonymous data. In such a case, there is a much lower impact if they have a security failure.

Data that is pseudonymous, where the index or keys to the pseudonymized data is held separately, is in general much more secure than identified data.

How pseudonymous data reduces legal and compliance risk

There are significant benefits to pseudonymity under many privacy laws. These laws vary by geography so here is an overview for some countries and territories.

Europe (including the UK)

Under the GDPR, pseudonymized personal data is still personal data and therefore processing needs to comply with GDPR[3]. The same is true under UK data protection law after Brexit. Although individuals working with the data may not know the identity of the test takers, the testing organization is still able to link the individual records back to individual data subjects.

However, there are significant benefits to pseudonymization under the GDPR and UK data protection law:

1. Security Measures

Under these laws, organizations are required to implement technical and organizational security measures that are appropriate to the risk, considering the personal data and processing involved. Pseudonymization is recognized as a strong measure to secure data and if in place, other measures may be less needed.

2. Breach Notification

These laws have strong requirements around notification in the event of a data breach and can also result in fines following data breaches. A breach of genuinely pseudonymous data involves much less risk than a breach of identified data, and typically it would not be necessary to notify data subjects of a pseudonymous data breach and it might not even be required to notify the supervisory authority. (However, this depends on a risk analysis, and needs to be evaluated on a case-by-case basis.) It's also much less likely to result in a fine If a data breach happens.

3. Data Protection by Design

These laws encourage use of pseudonymization as part of the recommended "data protection by design." In many interactions with regulators there will be some benefit given for implementing pseudonymization. For example, if you can achieve a purpose using pseudonymous data rather than identified data, then it is likely expected that you do this, including to conform with personal data minimization expectations.

4. International Transfers

The European Union and the UK have strict rules regarding the transfer of personal data to other geographies which are beyond the scope of this document. The key point here is that when considering the possible impact on individuals of the processing of their personal data outside of the EU or UK, pseudonymization may be an appropriate safeguard to allow processing/transfer to continue.

5. Processing Purposes

If you have some data collected for one purpose and want to use it for a secondary purpose, then if you can do the secondary purpose with pseudonymous data, you may be able to proceed without going back to the test taker. There are other factors that need to be considered in determining compatibility of purposes of processing data, so you should review carefully with a privacy expert beforehand.

For more detailed information on the GDPR implications, this IAPP document has some useful guidance: **https://iapp.org/media/pdf/resource_center/PA_WP2-Anonymous-pseudonymous-comparison.pdf.**[4]

[3] This can be contrasted with anonymous data, which is not personal data and therefore is completely outside the scope of the GDPR.
[4] Readers who are looking to delve more into the subject may also find recordings of an EDPS workshop on pseudonymous data and risk mitigation useful, see **https://edps.europa.eu/ipen-webinar-2021-pseudonymous-data-processing-personal-data-while-mitigating-risks_en**

How pseudonymous data reduces legal and compliance risk (Cont.)

USA

The US has a patchwork of federal and state privacy laws, the former being largely sector-specific (at least when it comes to the commercial sector) and the latter still relatively few in number and existing alongside state data breach notification laws.

At the state level, only California, Virginia, and Colorado presently have privacy laws of general application. The California Consumer Privacy Act recognizes pseudonymization as a concept and pseudonymous data may have certain benefits in research contexts, although this may become clearer over time. The Virginia Consumer Data Protection Act, which becomes operative on January 1, 2023, goes a bit further, excluding pseudonymous data from the scope of some obligations, like responding to consumer rights requests, while providing that certain requirements have to be followed, such as separately storing data to preserve the pseudonym and using effective technical and organizational measures to prevent unauthorized access. Similarly, the Colorado Privacy Act, which becomes effective on July 1, 2023, exempts pseudonymous data from some consumer rights and the obligation to comply with certain requests. It is likely that other US state privacy laws will similarly provide incentives to pseudonymize personal data.

All fifty US states, Washington D.C., and Puerto Rico now have their own breach notification laws. Although these laws differ with respect to specific details, all create incentives for organizations to follow good security measures. A common theme among these laws is that the definition of personal information involves a combination of identified data. Only breaches of specified, identifiable personal information trigger reporting requirements under all US state laws—for example, a name and an email address, or a social security number. Because pseudonymous data is not personal information and is not capable of identifying a person without the key, if only pseudonymous data is leaked, there is no breach that would trigger a reporting obligation under state breach notification laws.

Other jurisdictions

An ever-increasing number of other countries are enacting privacy laws, often taking inspiration from the European GDPR. A comprehensive overview is not possible in this document, but examples of these laws that include provisions on pseudonymous data are:

Australia

Under the Australian privacy principles, entities are required to give individuals the option of not identifying themselves, or of using a pseudonym (unless an exception applies).

Canada

According to Canadian federal law applicable to the commercial sector, breaches of security safeguards involving personal information are only required if there is a real risk of significant harm. The only Canadian province with a mandatory breach reporting obligation, Alberta, has a very similar requirement. It is reasonable to consider that under these laws, breaches involving pseudonymous data only would not have to be reported. An amendment to the Quebec Privacy Act that will come into effect this year will allow for personal information to be used without consent for necessary study, research, or the production of statistics, if it is de-identified, the meaning of which in this case is aligned with pseudonymization. The provincial privacy laws of Nova Scotia and British Columbia also place certain restrictions on public-sector entities transferring personal information outside of Canada and it may be possible to meet the requirements of these laws by only sharing pseudonymous data, such as a token or ID number, with a supplier based in other countries.

Japan

A law that will come into effect in Japan in April 2022 introduces the concept of pseudonymous information and such information will be exempt from data subject rights and mandatory breach notifications, subject to certain requirements being met regarding how the pseudonymous information is produced.

South Korea

In 2020 an amendment to existing privacy laws introduced the concept of pseudonymous data and provided for the ability to use and transfer data that is pseudonymized where such activities were previously not allowed without consent from the individual. There are various requirements and guidelines about this, with an expected benefit that businesses can combine pseudonymous data from various industries in order to add value.

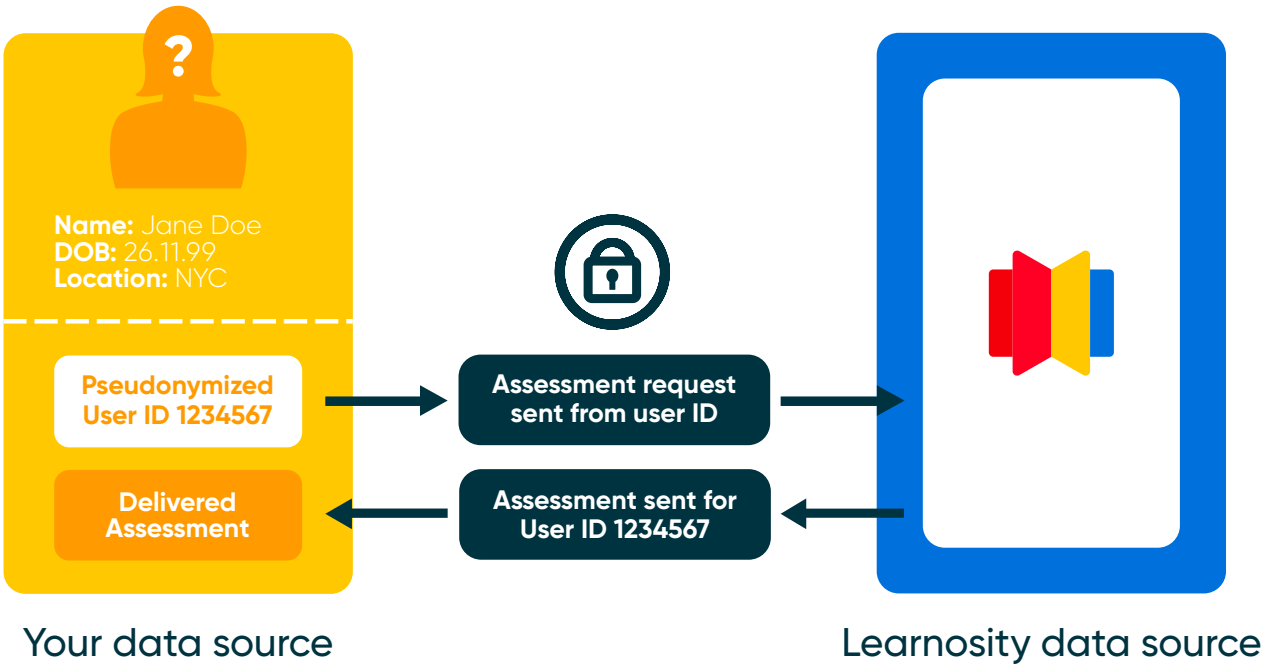
How Learnosity uses pseudonymous data

Learnosity is the global leader in assessment solutions. Serving over 700 customers and more than 40 million learners, our mission is to advance education and learning worldwide with best-in-class technology.

Learnosity Assessment Engine APIs make it easy for modern learning platforms to quickly launch fully-featured products, scale on-demand, and always meet fast-evolving market needs.

Pseudonymity is at the core of the Learnosity Assessment Engine APIs. Learnosity doesn't use or need learners' personal identities. Therefore, following the principles of privacy by design and data minimization, Learnosity requires that customers using its services pass a nameless user ID. Learnosity then delivers an assessment to that unknown learner and passes back the results. Learnosity has no knowledge of the learner's identity. Only the customer can map the user ID back to that individual.

For example, in the diagram below, the learner is called Jane Doe, but an ID is generated "1234567" and Learnosity only knows that and doesn't know her name, address, or date of birth. It delivers the assessment and passes back the result.



The advantage of using pseudonymity for Learnosity customers is that they can use Learnosity as a processor with much less concern about the privacy of their learners than with a processor that has the identities of learners. This reduces both security and compliance risk and is a good example of how privacy by design benefits all stakeholders—Learnosity, its customers, and learners.

How Caveon uses pseudonymous data

Caveon is the world's only company exclusively dedicated to protecting the security and validity of important tests. Caveon provides a variety of professional and technology-enabled services to many of the most prominent high-stakes testing programs, to ensure that their tests are fair, reliable, and valid. One powerful service that Caveon has provided to clients since its inception is Data Forensics, which is statistical analysis of test response data to detect anomalies that indicate possible test fraud or prior exposure of test content.

Caveon's Data Forensics analyses are used by its clients to identify: (1) untrustworthy test scores that programs may invalidate, (2) patterns that indicate potential testing irregularities that require further monitoring or investigation, and (3) breached test content that programs should refresh or replace. In some cases, Caveon clients may use Data Forensics results to initiate investigations into the cause of the identified statistical anomalies, particularly when the results indicate significant potential test fraud.

To conduct a Data Forensics analysis, the client testing program — most often through its test delivery vendor — provides Caveon with test response data for all of the test instances to be included in the analyses. Although Caveon statistical analyses benefit from utilizing numerous data elements, Caveon subscribes to the principles of privacy by design and data minimization and respects the privacy of test-takers. Privacy by design and data minimization dictate that Caveon only receives the least amount of relevant data to perform its analyses and strives to avoid unnecessary collection and processing of personal data. Caveon Data Forensics balances comprehensive and informative analysis with privacy considerations. For example, Caveon may analyze data by the training location of examinees, but never analyzes data by race or gender, and so should not receive those data. In addition, because Data Forensics results sometimes lead to adverse program actions against a test taker (which could include score invalidation or the initiation of an investigation) both Caveon and the testing program are able to eliminate potential claims of bias if no personal or demographic data was included in the analyzed datasets that led to the adverse action.

Based on all of the foregoing reasons, Caveon recommends that its clients do not provide personal data to Caveon for Data Forensics analyses. However, even if a client provides test response data that includes personal data, Caveon will often pseudonymize the data using the process described below.

Caveon's process for achieving pseudonymization typically works in one of two ways: (1) the client or test delivery vendor strips out all personal identifiers from the test response data and assigns each test taker an alpha-numeric ID, but does not provide the key to Caveon; or (2) the client provides test response data with all personal data included and Caveon assigns a random, unique numeric ID to each test taker and stores the key separately.

In those instances where Caveon is required to generate and assign a unique identifier for each test instance, the Data Forensics team follows the following procedure:

1. Create a list that has the same number of rows as the number of test instances in the dataset;
2. Randomly order the list created;
3. Assign list element numbers to the randomly ordered rows in the list; and
4. Insert the random numbers into the client-provided dataset and delete the original data values (that included personal identifiers).

By employing pseudonymity in its handling of test response data for Data Forensics analyses as described here, Caveon protects the privacy of test-takers, shields itself from claims of bias in its analyses, and reduces its data management burdens, all while reducing its risk of potential legal liability in the event of a security breach.

Notwithstanding the great benefits of pseudonymity, there are situations where Caveon is required to receive and process personal data to provide Data Forensics analyses in accordance with its client's objectives and instructions. Thus, it should be acknowledged that pseudonymity, while useful in many regards, is not a one-size-fits-all solution for data and privacy protection and must be used in conjunction with other effective administrative, organizational, and technical data protection and privacy measures.

Conclusion

In this paper, we've introduced the concept of pseudonymization and explained how it is a useful measure both for reducing security risk and to aid with legal compliance. We've given examples of how it is used by Learnosity and Caveon and suggested why it is a useful practice in the assessment industry.

We hope that the explanations and illustrations provided here will be helpful to others in the assessment industry.

Authors:

John Kleeman, EVP Business Development & Industry Relations, Learnosity
Marc Weinstein, VP Audit and Response Solutions and Chief Privacy Officer, Caveon
Jamie Armstrong, Group Legal Counsel, Learnosity

Discover more at:

https://learnosity.com
https://caveon.com

Contact us:

marketing@learnosity.com
Info@caveon.com

Legal note: