Data Filter Tool

When I am writing, Huna is thinking and working with me to provide expert advice. As Huna fact-checks, thinks, and identifies patterns for me, where does my data go? How is my data protected and secured? What is going on under the hood?

We share those same concerns with our users. We take this matter very seriously. We believe that some things that type should never have the opportunity to leave your direct control. That is why we have taken several measures to protect you and your data.

In addition to implementing Secure DevOps for our construction of Huna, as well as meeting regulatory requirements, we added another layer of assurance for customers by providing a **Data Filter** tool that lets you identify the type of sensitive data that you don't want Huna to use.

The following description explains the steps and measures we have taken to protect your data, including the **Data Filter** and its place in the broader secure architecture. It also explains the controls we are providing the administrators and users to provide the additional level of control they desire and need. We realize the description that follows is a bit technical, but we want to provide as much clarity as possible on this important topic.

Industry best practice sources identify two primary data loss types:

- **Disappearance** is loss of integrity or availability, which we have addressed through a resilient cloud-based design.
- **Leakage** is loss of confidentiality, both intentional and unintentional, which is the focus of our multi-faceted approach described in this document.

Even in highly disciplined organizations with mature information security cultures, all users might not think to protect certain types of data. In the middle of a busy workflow, anyone could overlook or forget certain aspects of data protection. These perfectly commonplace, human lapses result in data **leakage**. Our **Data Filter** tool is designed to give enterprise administrators the fine-grained control necessary to stop data from leaking.

As always, technical controls are no substitute for a company culture which teaches and enforces secure practices as the norm. Nor does this one layer compensate for a broader defense-in-depth strategy, which we implement seriously at Huna to include

- Abiding government and industry requirements and regulations (NIST, OWASP, CMMC)
- Encrypting data at rest and in transit
- Configurable read/write permissions of libraries
- Centralized means of limiting data-at-endpoints
- Secure architecture and infrastructure
- Code review and analysis procedures
- Security checks and controls in CI/CD workflows
- Secure DevOps capability and culture building

The **Data Filter** is not meant to be a replacement nor substitute for any of the above; rather, it is meant to be complementary assistance for organizations who are also already implementing defense-in-depth strategies of their own. To assist such an organization in maintaining control over *prioritized loss modes* that the implementing organization has identified as sensitive. We offer ALL of the following:

- Organizations have complete freedom to choose what goes into libraries. All libraries from all authors are private by default and can only made public (e.g. organization-wide or group-wide) with administrative review and acceptance.
- Choose who has access and what kind: Role-Based Access Control (RBAC) with column-level granularity ensures that only approved individuals have access.
- A combination of built-in and configurable regex filters allow character-level control over the blocking of information that should not *by any means* be transmitted. This information never leaves the library or the laptop.

We recognize that certain personally identifiable information should never leave the originator's control. This information includes things like social security numbers, salaries, birthdays, etc. Huna prevents searches against a list of regular expressions. Content that matches these expressions never leaves the local machine as part of a Huna query.

In an example, we see a social security number: 000-00-0000. By default, the filter is configured to block transit as soon as it recognizes a SSN. That is, once the end user types 000-00, the remaining information is filtered out. This filter is configurable up to whichever degree of granularity the administrator prefers. If even the existence of a single digit presents too much risk, it is possible (though not altogether recommended) to filter any 1 to N length collection of digits from search.

Furthermore, we recognize that organizations have sensitive data in addition to common PII data that we may not be familiar with. That is why we provide you with additional controls for meeting your specific needs. Authorized personnel can add expressions to this list to meet their organizations standards, guidelines, and policies as needed. However, *regardless* of how you choose to configure your filter, all queries are encrypted using modern techniques with a TLS-encrypted queries to a RESTful API, which queries against an encrypted-at-rest database.

This feature is a first line of defense and is meant to be used in conjunction with other secure practices as are standard in the art. We are exercising many of these practices to provide a secure end-to-end handling of data. As recognized in *Data Loss Prevention* by Simon Liu and Rick Kuhn (US National Institute of Standards and Technology), "Identifying and blocking all sensitive data is neither possible as an outcome nor wise as a goal." This feature is designed to give you *more control* to fully enact business and mission-informed Data Loss Prevention.