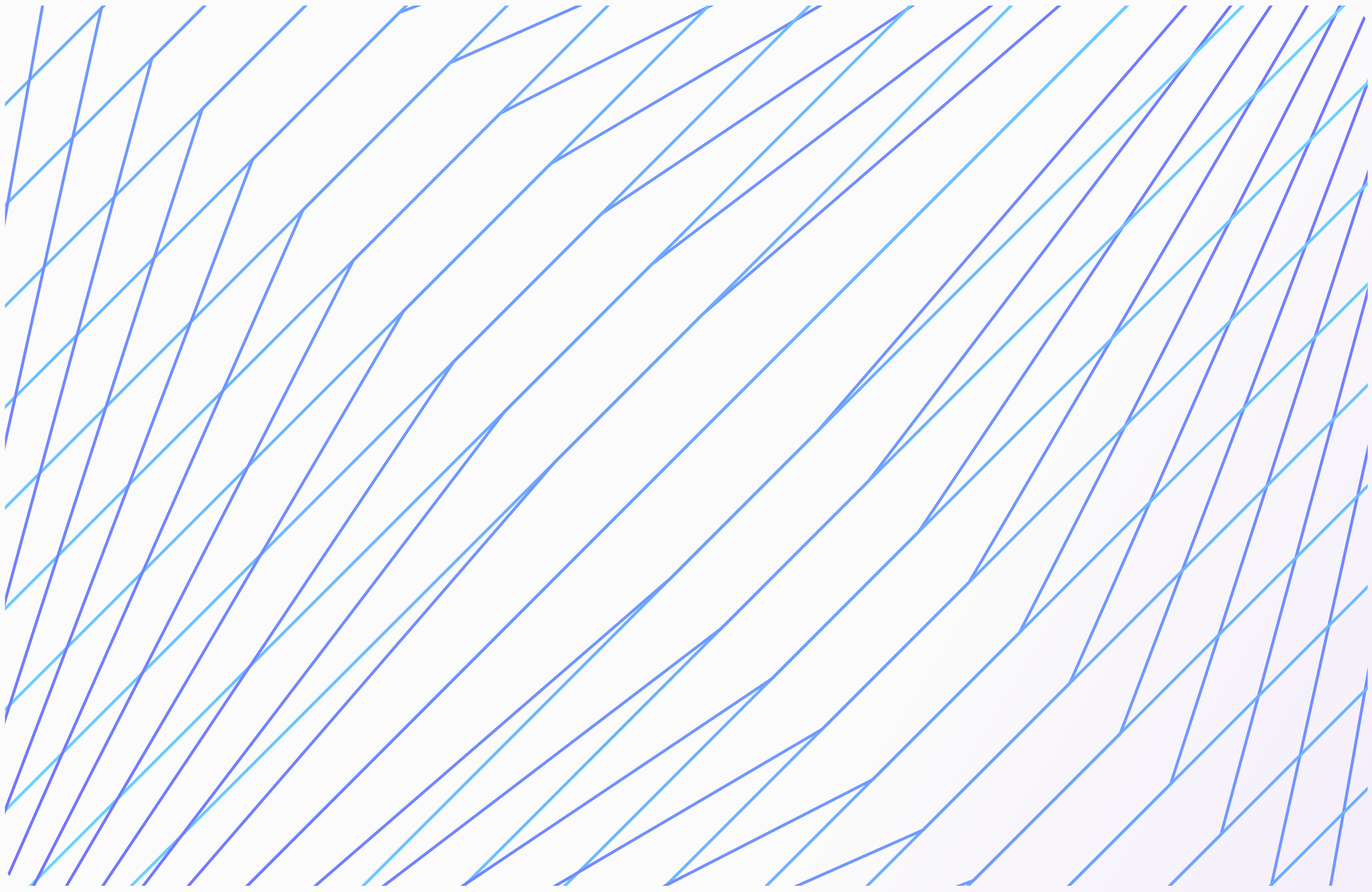
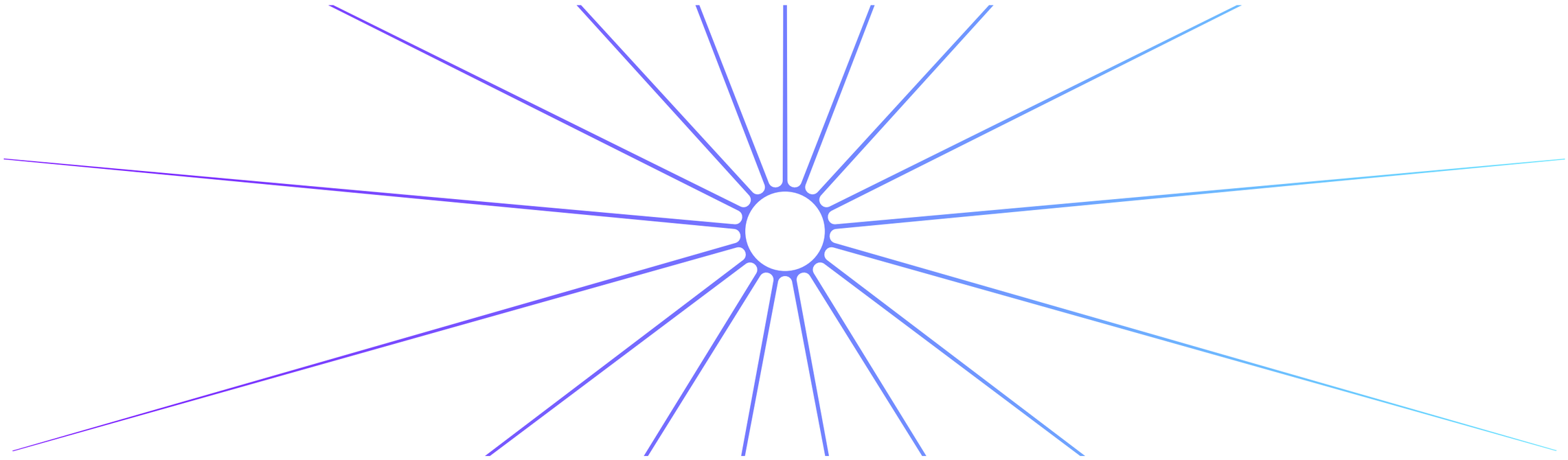
TONIC

Playbook

Centralized vs decentralized data de-identification

Published in 2025





Centralized vs decentralized data de-identification

When implementing data privacy compliance across enterprise organizations, having a robust strategy and framework for your approach to data de-identification is key to success. Perhaps you've purchased software to de-identify your data but don't yet have a process in place for integrating the software's use within your company. In this playbook, we'll outline three high-level strategies for managing compliant data de-identification across teams and use cases, including best practices for successfully rolling out each approach.

The problem:

Identifying a software solution to meet your data transformation needs is just the first step in defining your approach to compliant data de-identification. How you leverage that solution within your company surfaces a number of new questions to consider:

- Who will be responsible for managing the data de-identification solution?
- What roles or permissions will your users have?
- How will you govern specific de-identification techniques?
- What security and audit controls will you put in place?

Determining a management plan is critical to maximizing the solution's value across your use cases and ensuring its scalability across your organization.

The solution:

Define from day one whether you'll implement a centralized or decentralized approach to managing your data de-identification software, and designate the resources you'll need, including people resources. Identify built-in capabilities to support your approach, such as RBAC or audit trails.

Using the Tonic.ai platform as an example, we'll outline what this could look like at your organization. Read on for key considerations, recommendations, and the functionality that your team needs to successfully manage and maximize the value of data de-identification at your organization.

Three steps for successful data privacy compliance:

Step 1

Identify an organizational model for data de-identification that works for you.

Enterprise organizations can use one of the following options to ensure compliant data de-identification across software testing, development and AI model training:

Centralized management

Some organizations manage data access and de-identification through a centralized team. This centralized team:

- Evaluates new use cases
- Connects to and manages the required tooling for compliant data de-identification
- Determines data sensitivity and configures the required transformations, such as data masking, redaction, or synthesis techniques
- Generates de-identified data on a one-time or recurring basis on behalf of other stakeholders

The benefits of a centralized model include tighter security, faster execution, and more standardized de-identification practices.

However, this model requires putting a strong framework in place and, depending on the scale of your data de-identification needs, might necessitate additional team members to accommodate demand.

Partly centralized management

Other organizations manage data access and de-identification through a centralized team, but allow data owners and consumers to:

- Connect to and manage the required tooling for compliant data de-identification of the datasets they are responsible for
- Configure their own transformations, including data masking, redaction, or synthesis techniques
- Generate de-identified data on demand, leveraging tooling or automated workflows via an API

The benefits of a partly centralized model include more efficient onboarding of new teams and use cases.

However, individual users might need more training. Less centralization can also mean less standardized de-identification and a need for tighter user management.

Decentralized management

In this model, anyone at an organization can:

- Connect to and manage the required tooling for compliant data de-identification of the datasets they are responsible for
- Configure their own transformations, including data masking, redaction, or synthesis techniques
- Generate de-identified data on demand, leveraging tooling or automated workflows via an API

Organizations with a decentralized model often still designate a team or user for dedicated data de-identification support. If using a commercial software solution for data de-identification, at least two administrators are recommended for user management.

A decentralized model maximizes flexibility and efficiency with regard to new teams and use cases. However, organizations must closely monitor data de-identification practices and data security.

Step 2

Develop a process for expanding data de-identification to new use cases, workflows, and teams.

Whether you opt for a centralized or decentralized model, you need a strong framework for implementing data privacy compliance across workflows and new use cases, as they arise.

For a centralized model, consider:

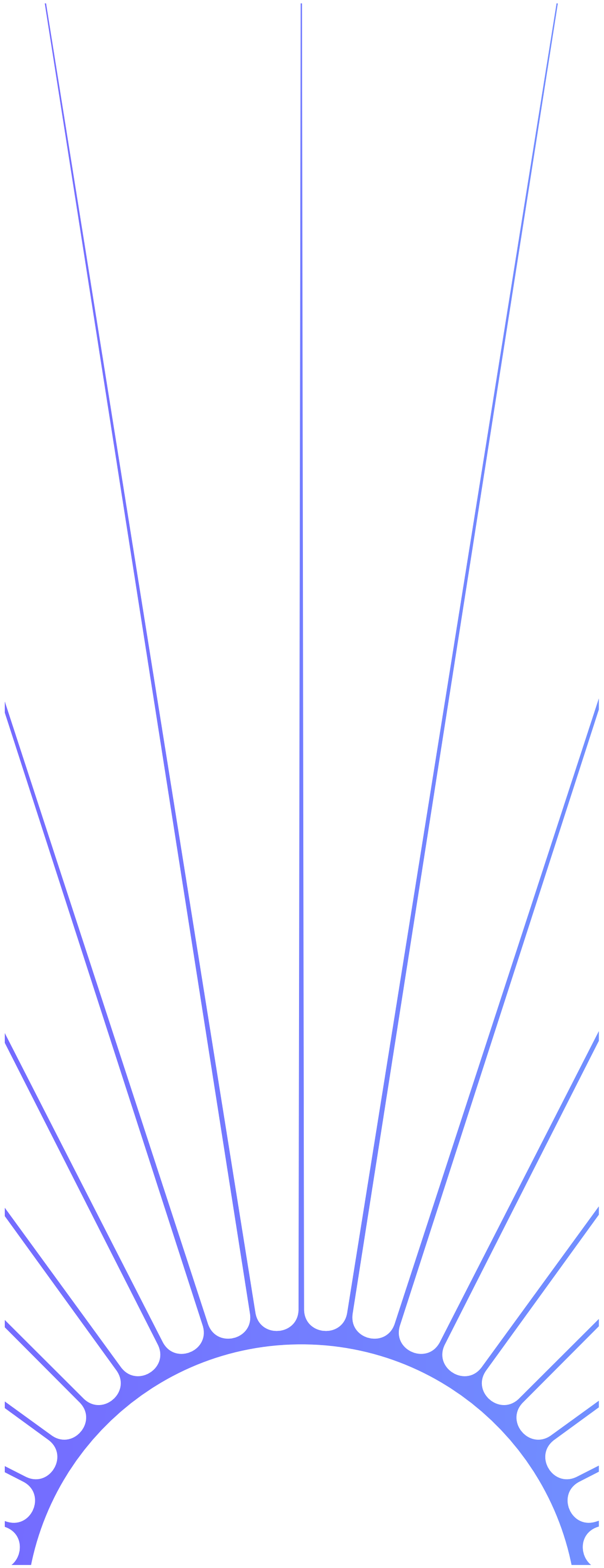
- How new teams in need of data de-identification support interface with your centralized team
- What information they need to share to onboard a new use case for data de-identification
  - For structured data, this can include database connection details, tables to be de-identified or truncated, subsetting specifications, and data generation cadence (one-time or recurring)
  - For unstructured data, this can include access to dataset files, defining which sensitive entities should be redacted and which should be replaced with synthetic values, and identifying custom entity types to be de-identified

You might also need to institute a feedback loop to allow data owners and consumers to validate that the de-identified data meets their needs.

In a partly centralized or decentralized model, new teams and users of your de-identification processes need training on how to use the chosen solutions. Many organizations also provide guidelines such as:

- How data should be de-identified
- How often to refresh the de-identified data

Administrators of your data de-identification solutions need to monitor new user signups and to regularly audit usage. You might also need to establish a framework for internal support.



Step 3

Leverage built-in capabilities to manage data privacy compliance and de-identification

Robust solutions for data de-identification and synthesis include capabilities designed for scalability, customization, and control.

Platforms like Tonic Structural and Tonic Textual provide built-in functionality to empower enterprise organizations to exert nuanced, flexible control over access and de-identification practices. By leveraging these capabilities, organizations can manage myriad data de-identification use cases in a way best suited to their operational needs.

For software testing and development use cases, organizations often rely on test data management software. Tonic Structural is a modern test data management platform for transforming sensitive production data into safe, high-fidelity test data that preserves your data's utility. Capabilities within Structural that maximize the software's implementation and management include:

Role-Based Access Controls

To provision access to new users, Structural administrators can use built-in roles, or can create custom global (instance) or workspace permission sets.

Structural's role-based access controls provide enterprise organizations with maximum flexibility and control to manage how users interface with the product.

Workspace Sharing

The creator of a workspace can share that workspace with any user, who then has access to that workspace based on their assigned workspace permission set (in other words, full or limited access).

Structural administrators always have visibility into all workspaces, and can if needed give themselves full access to any workspace.

Custom Sensitivity Rules

Structural's built-in Privacy Scan identifies sensitive columns in your database, but custom sensitivity rules allow organizations greater control over which sensitive data is flagged and which generators Structural recommends.

For centralized teams, custom sensitivity rules can save time and effort when connecting to a new database.

For decentralized teams, custom sensitivity rules can help standardize data de-identification across an organization.

Privacy Report

For teams whose access to Structural is limited, downloading and sharing the Privacy Report gives stakeholders outside of Structural visibility into what data is sensitive and how that data is being de-identified.

Privacy Reports are often a key component of feedback loops for centralized teams, and internal audits for decentralized teams.

Generator Presets

De-identifying data shouldn't be a slog. Many organizations use Structural's robust preset generators to make the de-identification process more efficient and standardized, eliminating the need for users to reselect configuration settings across multiple columns, tables, and workspaces.

For AI model training and LLM privacy proxy use cases, organizations typically require a solution for redacting and synthesizing unstructured data. Tonic Textual is a platform powered by proprietary multilingual models for detecting and de-identifying sensitive entities within unstructured data at scale to ensure its safe use in model training and AI development. Its access and management capabilities include:

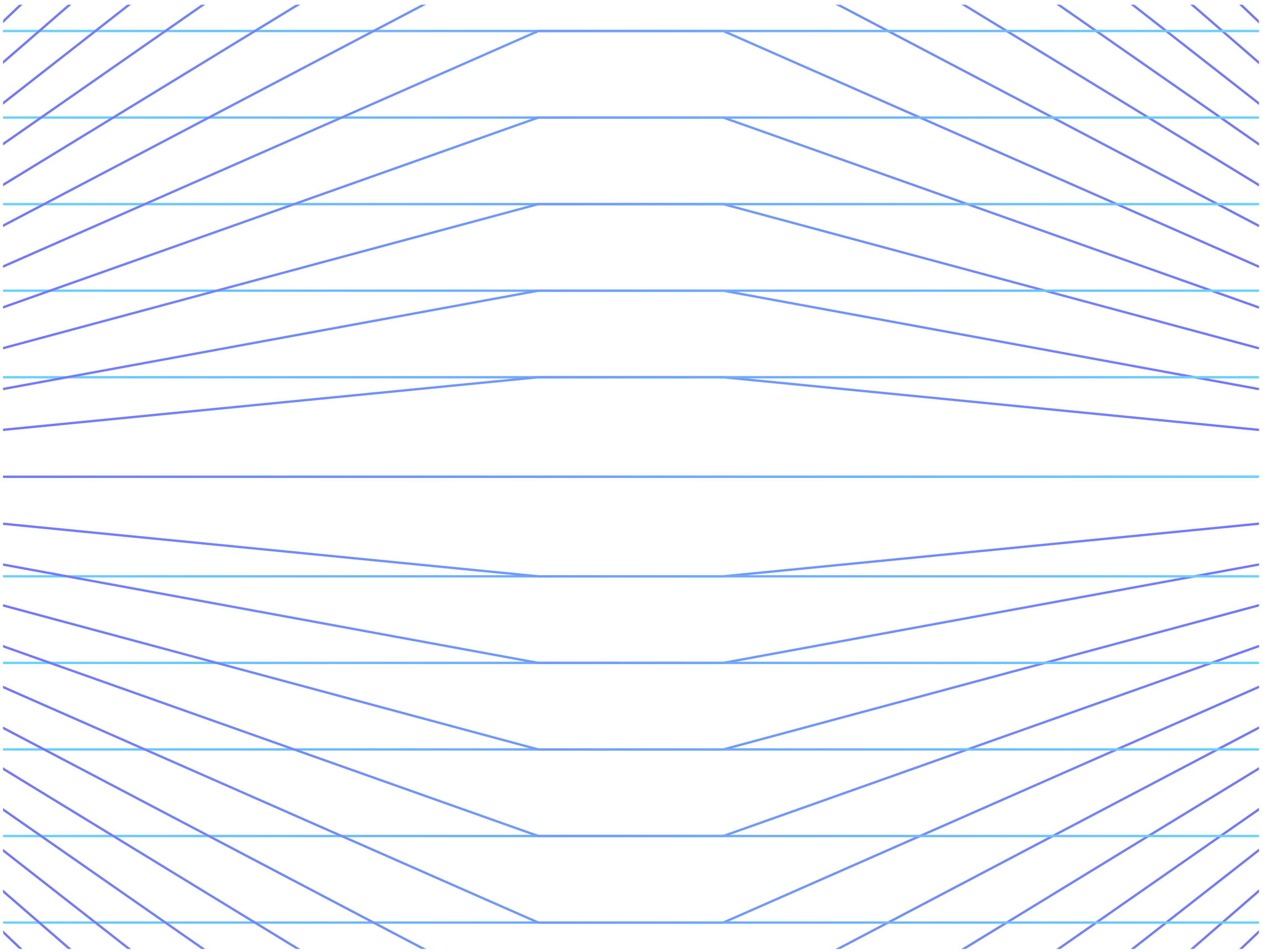
Role-Based Access Controls

To provision access to new users, Textual administrators can use built-in roles or create custom global (instance) or dataset permission sets.

Textual's role-based access controls provide enterprise organizations with maximum flexibility and control to manage how users interface with the product.

Usage and Access Monitoring

Tracking who has access to Textual, what actions they take, and how they're using Textual is simple with built-in monitoring and downloadable reports and logs.



TONIC
THE FAKE DATA COMPANY

Optimized data de-identification with Tonic.ai

The choice of a data de-identification solution is just the beginning of your journey toward achieving data privacy compliance within software and AI development. The software you choose should come with comprehensive data access and governance capabilities, and the way you manage it within your company will ultimately determine your success.

If your team needs help in deciding whether a centralized, partly centralized, or decentralized model works best for your organization, or understanding how to use Tonic's functionality to manage access and security processes, connect with our de-identification experts for a consultation. Our team is well-versed in supporting enterprise organizations as they deploy and scale data privacy compliance processes, and can help your team to architect a custom solution that fits your needs.

Scalable, realistic data de-identification doesn't have to be a CI/CD pipeline dream. Take your first step toward test data success today with Tonic.ai.

hello@tonic.ai
www.tonic.ai
@tonicfakedata