Defensible Preservation of Data: A Checklist

The following checklist provides a roadmap to potential steps a company might take to promote "defensible" preservation of this data. Not all steps are necessary or relevant, but the consequences of each should be considered:

Outside Communications and Guidelines

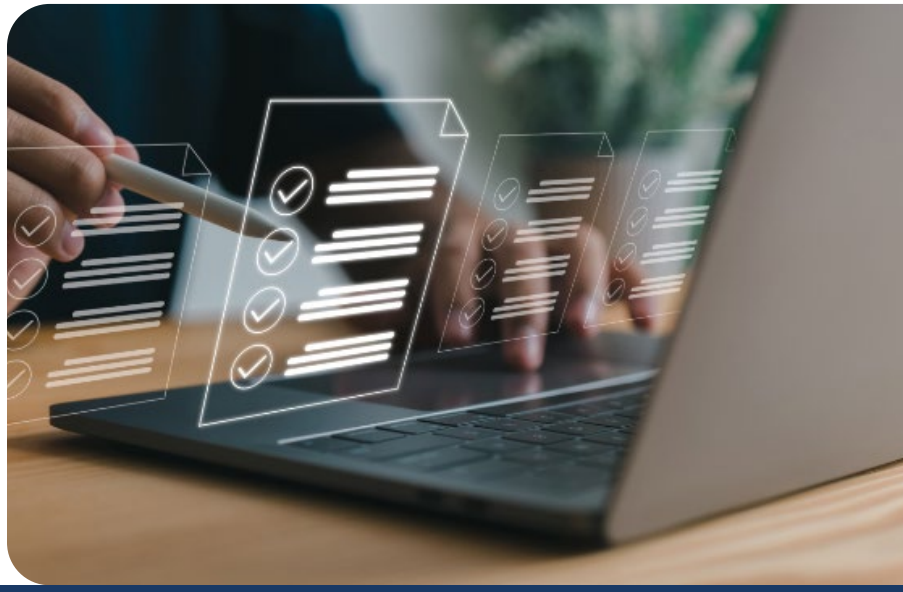
- ☐ What assumptions have been made about the nature of the queries or reports that the requesting party might need?
- ☐ What discussions, if any, with other parties have been conducted? What assumptions have been made in the absence of a meeting?
- ☐ Have any of these systems been produced in litigation before? To any regulatory body? What were the earlier formats and approaches? How is this extract different?
- ☐ Has anyone been asked to prepare an affidavit regarding any aspect of the extraction / documentation process?
- ☐ Are these systems considered "systems of record / sources of truth," especially if maintained under a regulatory framework?
- ☐ In general, what industry regulations would apply to the extract?

Data Retention

- ☐ What of the various industry retention time frames would be applicable to the extracts?
- ☐ Describe any litigation holds that have been applied previously and their specific implementation method?
- ☐ How long will the data be available on the source system until it is decommissioned?
- ☐ Does archive data, a prior system, litigation snapshots, long term backups, etc., exist for any of these systems?
- ☐ How much of the data is duplicated in other current systems?
- ☐ Do these systems contribute to a data warehouse / data lake, etc.? Please specify.

Process and Documentation

- ☐ Has a formal risk assessment been conducted on the application and its data sources?
- ☐ Has a formal project manager or lead been appointed, reporting to a risk-based steering committee?
- ☐ What are the full descriptions of each system, including history, business purpose, user and system documentation, report formats, export specifications, etc.?
- ☐ Is the system custom vs. commercial/packaged?
- ☐ Does the system's descriptors use internal company aliases vs. external "jargon?"
- ☐ What applicable standards have been used to guide the preservation method?
- ☐ What is the methodology for the extract (e.g., ASCII dump, database backup, point-in-time snapshot, migration to new system, third party archiving service, report output)
- ☐ What is the nature of the available documentation?
- ☐ In terms of verification and validation methods – what audit reports of the extract are being created?
- ☐ Is there an available data dictionary/schema for the items being extracted?
- ☐ Which business and technical owners have been consulted to determine system content and extract approach? Which vendors have been consulted?
  - ☐ What documentation is available to discuss the decisions and tradeoffs?
  - ☐ What will be included in the documentation package?



Privacy and Personally Identifiable Information (PII)

- ☐ How is PII and other sensitive data identified and protected in the extract?
- ☐ What applicable regulatory standards apply to personal data (e.g., GDPR, CCPA/CPRA, NYS, etc.)
- ☐ Has a data classification model been applied to the extracts (e.g., confidential/proprietary, private, sensitive/internal, public)
  - ☐ If so, how are these levels denoted in the extracts?
  - ☐ How are source access rights realized in the migration?
- ☐ How will sensitive or proprietary data be protected in the extract format?

Source Data Review & Responsibilities

- ☐ What is the signoff process for each system?
- ☐ Who actually performs the migrations and how are results evaluated/tested/compared?
- ☐ Are revisions to the migration process and results of test migrations documented?
- ☐ Was a complete schema available for the source system? What other documentation is available?
- ☐ Did the source system use virtual or computed fields?
- ☐ Are any source tables subject to additional third-party licensing/proprietary requirements?
- ☐ Were any local extracts made outside the database that might need to be incorporated?
- ☐ Were any other supporting systems / websites / data feeds providing input or validation to any of these systems?
- ☐ What steps have been taken to assess source data quality?

Mechanics of the Transfer

- ☐ Are error logs/exception logs available?
- ☐ Are checksums, hash values, record counts, overall size metrics provided?
- ☐ What audit trails or logs have made part of the extract?
- ☐ Does the documentation include – data entry screens, business roles, report formats/outputs?
- ☐ What field mappings have been chosen for each extract?
- ☐ Are data transformations occurring in terms of field size, data type, remapping, concatenation, etc.?
- ☐ What other business rules were in place in the legacy system and how are they represented?
- ☐ Were there database views in place that have been migrated?
- ☐ What time stamp information is available to demonstrate when various data elements were migrated?
  - ☐ Is there a possibility that different tables would be migrated at different times and not be in sync?
- ☐ Is an ETL configuration available?
- ☐ How are long text/memo/BLOB entries handled as opposed to fixed length fields? How are binary characters addressed?
- ☐ Is there any EBCDIC or other non-ASCII coded data? Foreign language characters?
- ☐ How are one-to-many or many-to-many relationships treated in the extract process?
- ☐ Will specifications be provided as to the extract format, e.g., CSV, tab-delimited, use of special text characters, continuation lines, etc.?
- ☐ Are code entries expanded by lookup tables or left as codes?
- ☐ Is data “de-normalized” by the extraction, i.e., unique values are repeated for the sake of visibility and reporting?
- ☐ Is referential integrity maintained through the extract (e.g., is any data referenced by joins included)?
- ☐ What data has NOT been migrated (e.g., system tables, history files, temporary tables, lookup/validation tables)
- ☐ Is the source database migrated from a “dump” or live copy?
- ☐ How does the migrated system handle queries? reports? purging?