GDPR Perspectives:
Operationalizing GDPR
Article 30 Record Keeping
Implementing GDPR Article 30 Record Keeping From Personal Data Records

BigID has pioneered an approach to documenting data processing activities and associated systems using actual discovered data versus surveys and interviews. Understanding the provenance of data from collection to processing and disposition is no longer a nice to have. With the EU General Data Protection Regulation it is now a must have - particularly in order to address Article 30 requirements that mandates covered organizations maintain a record of processing activities under their responsibility.

However, the combination of exploding data collection combined with growing data processing complexity has made inadequate, traditional approaches to documenting data processing. Building and maintaining records of data processing have not advanced much in two decades, relying on incomplete survey information to assemble data inventories and drawing tools like Visio to diagram data flows that may be only loosely connected to actual systems-level activity.

Email surveys based on interviews and stakeholder recollections can never equal data-driven discovery approaches that leverage data science to generate actual processing records. This Whitepaper examines how modern approaches to data discovery and mapping like BigID are trailblazing a new approach to capturing data lineage and flows to meet modern data processing record keeping - from initial mapping creation to ongoing maintenance of records.
Problem Overview: What Is Article 30 Record Keeping

For most subject organizations, addressing data subject rights along with the Directive's Article 30 record keeping provision represent two of the more vexing challenges. The Article 30 provision is designed to compel organizations to operationalize the Regulation's underlying accountability principle by demonstrating accurate knowledge of data collection and processing. However, previously, most organizations could only avail themselves of questionnaire-based surveys for finding and mapping their data. If surveys are completely detached from the data, they are fundamentally incapable of accurately representing how data is either collected or processed.

For organizations that hold EU citizen data, the obligation to maintain up-to-date and accurate records of data processing activity that regulators can request on demand, raises a set of fundamental questions:

- If you are reliant on human vs machine data accounting can your records of data processing ever be accurate?
- How do you keep human record keeping current with changes to data?
- Can labor-intensive and manual record keeping ever be cost efficient and non-disruptive to the business?

In specific terms, the European Union's General Data Protection Regulation (GDPR) Article 30 provision stipulates that covered entities must maintain an up-to-date and accurate record that can be presented in an electronic form of all processing activities under its responsibility. The question therefore is whether this is best achieved by man or machine - manual steps based on surveys or automated discovery, classification and data source profiling processes.
Surveys or Scans: How To Build Data-driven Data Accounting

Organizations have had to contend with two imperfect options to generate a record of data processing activities: either engage multiple stakeholders in a time-consuming and resource-intensive series of surveys and questionnaires to generate at best a moment in time approximation of the data processing flow; or, running a series of scans to accurately find and document data flows but without a real understanding of the data in the context of data subjects.

Survey-based approaches for data processing record keeping have several shortfalls:

- No reliable view of actual data processing;
- No comprehensive inventory of personally identifiable data across data sources and data formats;
- No capability to frame actual processing flows in the context of a specific data subject's consent and stated purpose of use;
- No monitoring of changes to the underlying processing or data collection steps.

Operationalizing Article 30 Compliance: From Data Records to Record Keeping

Organizations can both demonstrate they have operationalized accountability, and proactively address GDPR compliance by maintaining a consolidated data flow map view that combines business concepts and regulatory parameters, including consent and purpose of use, with real data insights.

Rather than hope that a theoretical view of data processing activities derived from surveys is accurate and up to date, stakeholders can effectively visualize, annotate and describe the steps of the data processing flow centered on data subjects and their attributes and informed by purpose of use and consent policies.
In basic terms, a data flow map should serve as a current and detailed representation of how data is captured, processed, stored and disposed in the course of a business process.

However, a data flow without collective business context doesn't capture the full purpose behind a data process. A data-driven approach to recordkeeping, therefore, also needs a way for organizations to capture business context like purpose of use and marrying that with the data perspective to provide a unified view of a data process that spans IT and business stakeholders.

A New Data-driven Approach to Data Processing Record Keeping

Meeting the Article 30 provision has taken on urgency for companies subject to GDPR requirements because it makes the Regulation's accountability principle for data processing a practical reality. Manual survey or interview-based processes can never satisfy an accounting-driven outcome since they fail to accurately document personal data records. They are representations captured at one point in time that rely on memory, over ongoing discovery. If the intent of GDPR is to safeguard personal data collected and processed by enterprises than companies, they need to strive for reality over fiction. A modern approach to Article 30 record keeping therefore needs to capture the following capabilities:

Data Flow Visualization
Visualization of data processing provides clear articulation and documentation of data processing activities easily understood by regulator and company alike. The visualization should clearly articulate data collection, processing and disposition. It should be sufficiently flexibility to capture a diverse set of processes. It should be easy to compose and output, to establish a point in time record.

Purpose of Use and Consent
Article 30 aims to establish that data processes have accurately collected purpose of use. Records therefore need to capture and maintain purpose of
use. Records therefore need to capture and maintain purpose of use. In an ideal world, the record keeping would also reflect appropriate consents and retention rules so they become both records and testaments of data processing. For tools that assist companies with documenting their data processing, this requirement to capture business context means that the tool needs a way to collect information from business functions and superimpose it over any data flows visualizations.

**Data-driven Data Inventory**
To be true records of data processing, Article 30 records also need to be accurate accountings of actual data collection and processing. This invalidates surveys or questionnaires as sustainable longer term solutions. People struggle to recall where they placed their car keys let alone which data they put in which file shares, for example. This concern becomes more pronounced if the GDPR data processing records are used for framing changes to data processing to mitigate privacy risk. Accuracy in accounting can’t be accomplished with compiled questionnaires. Data processing records should be data-driven and built around actual data inventories and flows.

**Toward Lineage**
Understanding data flow is rooted in understanding data lineage. A snapshot of data processing can if done right also provide clarity on data provenance and lineage. Understanding the time-series evolution of how data moves through an organization helps organizations both safeguard data by uncovering risk and also improving efficiency in processing.

**Continuous Compliance**
GDPR record keeping is intended to be an up-to-date snapshots, rather than a quarterly exercise. For organizations, therefore, there exists a challenge of how they make record keeping so that snapshots can accurately be taken at any point time. Surveys, by their nature, are one-shot undertakings that may be out of date even by the time that responses are produced. Scans, conversely, can be both dynamic and current. Moreover, if done right they can also provide perspective on change - alerting organizations to new processing activities or collection of new PI/PII that warrants attention before a compliance problem unfolds.
GDPR Article 30 asks organizations to provide evidentiary proof that every digital process requiring collection and processing of personal data is properly accounted for. But to make it truly accountable to individuals to whose data organizations collect, store and process, the data subjects that are the focus of GDPR provisions, record keeping of data processing needs to be based on actual data and processing records that are specific to each individual.

Addressing Article 30 - and by extension the GDPR in totality - requires the ability to build, maintain and automate data processing record keeping that integrates identity-centric data intelligence with workflow steps, business concepts and purpose of use tagging in an intuitive, interactive visualization.

Without a comprehensive, current and identity-centric accounting of whose data they process, and how the data is processed at a systems level in the context of both business processes and GDPR policies, maintaining data privacy and protection accountability is an elusive goal.

Using BigID, organizations can generate and maintain a consistent and accurate set of data processing flows which directly represent real processing flows built on direct data insights while simultaneously monitoring data activities for change in collection and processing.