ZLTECH

ZL GDPR SOLUTIONS
Fulfilling Subject Access Requests

Since the General Data Protection Regulation has gone into effect as of May 25, 2018, organizations that manage EU resident data are now required to swiftly respond to Subject Access Requests—an individual's request for access to or deletion of their personal data.

At present there are two fundamental technological obstacles preventing organizations from effectively fulfilling a Subject Access Request, which few solutions are able to address:

1. Scalability
2. Complexity

Scalability

Most large organizations have accumulated a variety of data silos over the decades, and the ability to bridge them is becoming one of the greatest IT challenges of our time. As the number of data repositories continues to grow, so does the number of places to search for personal data.

If a Subject Access Request came in tomorrow, how many repositories would an organization have to

search? A large organization may have data spanning dozens or even hundreds of silos, including:

» Format-based silos (email, file share, SharePoint, IM, ECM, etc.)

» Geographical location-based silos

» Management-based silos (in place vs. archive, cloud vs. hybrid)

» Function-based silos (eDiscovery, regulatory compliance, records management)

Searching each of these repositories individually is unfeasible and each search engine likely has its own limitations, making the ability to search across all silos from a centralized platform essential.

Complexity

Beyond the silo problem, fulfilling Subject Access Requests still carries complications, namely the inability to accurately and defensible identify documents containing personal data, while ensuring it pertains to the respective subject. A limited search engine will prevent a defensible response to Subject Access Requests for two primary reasons:

1. It will return large quantities of irrelevant and non-personal data, multiplying the time needed to review documents prior to release or deletion. Review is necessary to ensure each document:
   - Pertains to the subject and contains their personal data.
   - Doesn't contain other sensitive data that should be withheld or redacted.
   - Doesn't contain important information that cannot be deleted due to "legitimate interests" such as business, records management, legal, or regulatory purposes.
2. It may not be capable of locating all documents containing the subject's personal data, resulting in an incomplete search.

Consider the variations that must be searched for in this simple name and address combination alone:
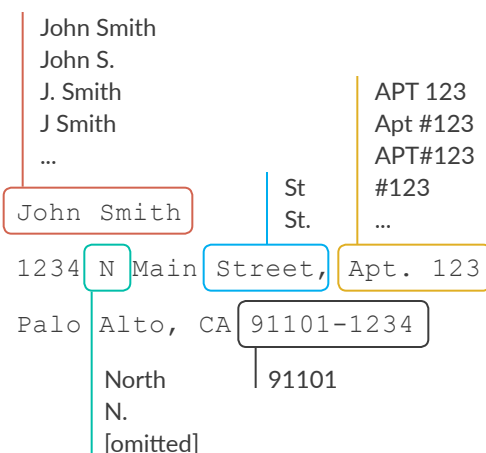
The operations this search requires to ensure a complete and accurate return are not simple. The user may need to search multiple keywords with multiple variations and spellings, keywords that appear within a particular distance from the subject's name (to ensure they pertain to the subject), numerical values that appear in multiple formats, and numerous other criteria that would be missed by most search engines. For example, "Street" could appear in other forms, such as "St." or "St". The same applies to the word "North," which may appear as "N" or be omitted entirely. It's also possible that a subject's complete address appears in a document, but not their full name, or vice versa.

Most search engines do not account for each variation in which an address may appear, let alone all the possible variations for each type of personal data that will need to be found. Furthermore, time limitations require the user to search for all variables simultaneously. Otherwise, the combinations for each variable will require many discreet searches, which will be multiplied across each repository. Under current approaches, organizations may have to run hundreds of individual searches for a single Subject Access Request.

Solving Scalability and Complexity

While many solutions enable organizations to perform complex searches across five million documents, change this number to five billion and many platforms on the market will fail.

Similarly, some products attempt to provide an enterprise-wide search, yet massive data volumes prohibit their architecture from offering the depth of search and analysis required for GDPR.

Ultimately, while some solutions offer complexity of search and others may approach scalability, providing both of these components simultaneously is a fundamental problem that few solutions have been able to solve. One need only look to the eDiscovery world to see dozens of examples of solutions that offer one but not the other, which is why the more involved processing and analysis phases of eDiscovery must be reserved for later in the EDRM process once data volumes have been reduced.

Despite the similarities, GDPR Subject Access Requests depart from eDiscovery in several key areas that increase the level of difficulty:

» Subject Access Requests are expected to occur significantly more frequently and in greater numbers than is typical for eDiscovery cases. As such, organizations will have less time to perform Subject Access Requests than they do an eDiscovery search.

» GDPR Subject Access Requests require complex search operations to be performed across enormous volumes of data, as opposed to across data belonging to specific custodians.

» Unlike eDiscovery cases which rely on a "best effort" standard for searches, GDPR requires organizations to find all personal data regardless of location in order to avoid challenges by the requesting data subject should their personal data not be found.

» GDPR requires control over every duplicate of a document, rather than only a single copy.

» The scale of search required by GDPR is higher, spanning all enterprise repositories.

Without the ability to search all silos and account for all variations of data entries across billions of documents in a timely manner, Subject Access Requests will be indefensible and expose the organization to financial sanctions.

ZL UNIFIED SEARCH

ZL GDPR Solutions address the challenges of both scalability and complexity in one platform. Leveraging the scalable and flexible BigDB architecture, ZL is capable of searching and analyzing the world's largest enterprise data environments, often including billions of documents and hundreds of unique data sources, including:

» File Shares
» Email
» SharePoint
» ECM Systems
» SMS & MMS
» Social Media
» IM

With ZL's parallel query engine, organizations can enter customized, enterprise-wide searches to identify and remediate relevant personal data. Revisiting the previous sample search for a name and address, a user might enter the following query within ZL's search engine to search enterprise repositories for documents containing John Smith's name and address:

```
(John) w/2 (Smith) w/3
(1234 w/2 (Main w/2 (Street OR St. OR St))
w/3 (Ap* w/2 123) w/2 Palo Alto
w/2 CA w/2 91101*
```

This query's conditions can be broken into the following components:

1. **Wildcard Expansions "*":** Identify keywords that have multiple variations. "Ap*" will therefore return documents containing "Apt," "Apartment," etc.

2. **Proximity Rules "w/[number]":** Documents subject to these conditions must feature keywords within a given distance of one another. This condition can be used to ensure a keyword's relevance to the subject, or that two keywords are related to one another.

3. **Fuzzy Logic:** Returns documents with misspelled keywords and similar words.

4. **Boolean Logic "AND/OR":** Enables the greatest search flexibility.

As demonstrated, small samples of keywords and conditions require high levels of complexity to provide a complete search. This complexity increases exponentially as the number of keywords and conditions increase, as they will for a typical Subject Access Request. Combine this with the requirement of searching across potentially billions of documents, and very few solutions can handle the load.

ZL's parallel query engine functions by running multiple searches in parallel, each one for a different condition of the query. The documents that respond to each condition are then cross referenced with one another to display documents that respond to each of the required conditions. Users can then choose to tag these documents, apply retention policies, export, or delete them.

Built from the ground up to support enterprise data for the largest corporations on earth, ZL's vastly scalable architecture stands alone in its ability to provide the required complexity of search on a global scale.

ZL GDPR Solutions ultimately offer a true path to find and act on personal data for GDPR compliance, no matter where it lies.

Awards

2018 STEVIE GOLD WINNER
INTERNATIONAL BUSINESS AWARDS

Gold Stevie® Award in 2018 International Business Awards for Best Content Analytics Solution

KM World Trend-Setting PRODUCTS OF 2018

KMWorld Trend-Setting Products of 2018

2017 STEVIE GOLD WINNER
AMERICAN BUSINESS AWARDS

Gold Stevie® Award in 2017 American Business Awards for Best Big Data Solution