

Data Discovery & Classification for GDPR

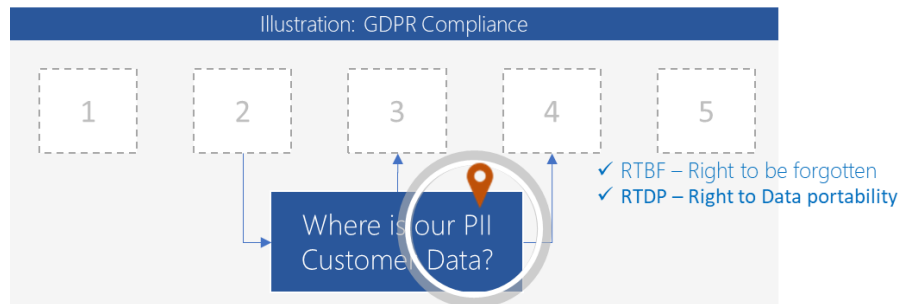
the promise of understood Customer data...



Getting traction on GDPR today!

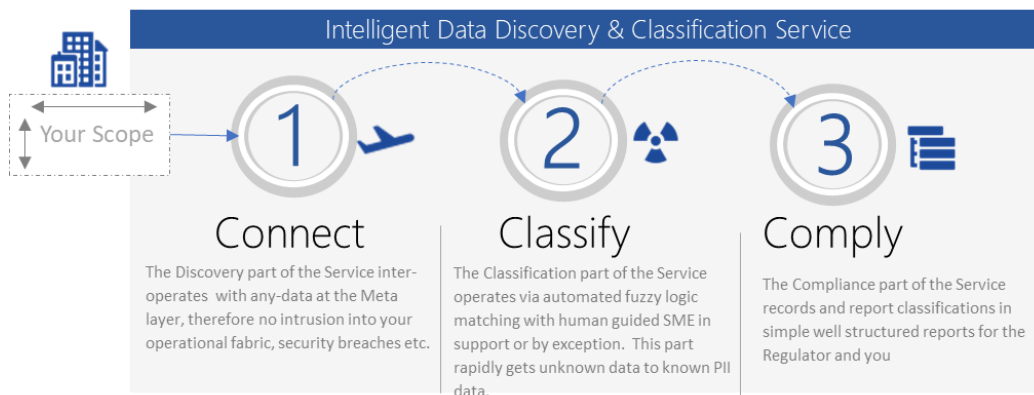
Getting to GDPR (General Data Protection Regulation) compliance means a number of things for your Organisation. One of these, is pin-pointing **PII (personal identifiable information)** in your Enterprise.

The SENYA Intelligent Data Discovery & Classification service (iDX) does that using a combination of **automated intelligence** supported by **human guided technology** to locate and accurately **tag** PII data at scale. Locating sensitive PII is essential to protecting it thus complying with GDPR.



Help me see personal Customer Data clearly!

The Senya iDX service allows you to connect any data element within your Organisation with a GDPR **understood data type**. This rapidly allows you to know **what** personal identifiable Customer data you have, **where it is located** and in **what context**. Using the Service is a simple 1,2,3 step process.



What makes the Senya iDX service unique is our ability to inter-operate at the Meta-Data Layer and only by exception dip into actual data. Benefits of this approach may not be immediately obvious, so let's illustrate by example:

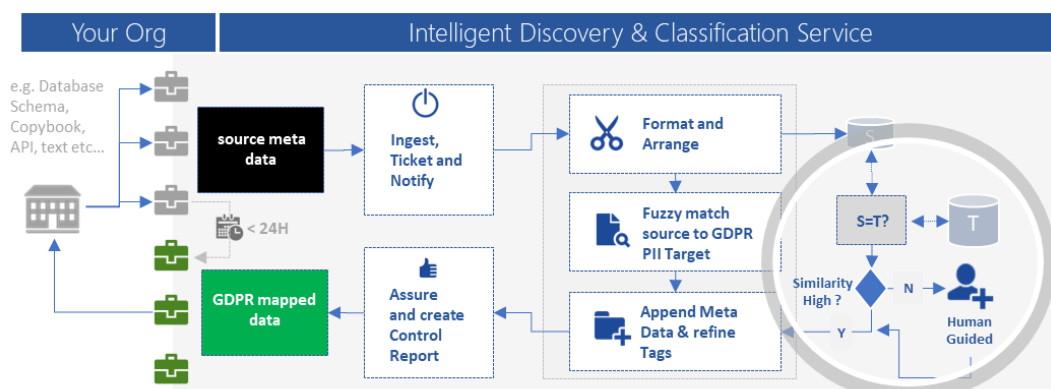
Imagine having to review a set of HOME Magazines to find all of the articles dealing with Kitchen renovations. Of course, one way is paging through every magazine looking for the relevant articles, much in the same way many solutions scan and look for PII patterns at the raw data level. A cumbersome and time-consuming process.

Instead, our way is to scan the indices of each magazine and use fuzzy logic to determine the probability of whether the index headline is about Kitchen renovations. Not only that but the fuzzy logic also further understands what the Kitchen renovation is about from the headlines. Consequently, this approach accelerates the process of finding the article and scales well compared to the page by page effort that dramatically increases as we add more Magazines. Also note that nobody else can read the Magazine while we are paging through it, another limiting factor.

Now imagine this set of Magazines as being the Data in your Enterprise and Kitchen renovations, the PII data. Having to scan every system at the raw data level to identify PII could take a while, and not provide the anticipated outcomes. Using fuzzy logic to scan the metadata of each system and then dip into the actual PII data when necessary, will save valuable time and effort, and ultimately cost, with arguably better results.

Align disparate data with GDPR for compliance!

The service allows you to align any data element from any format to a GDPR target element. Typically, this means any data source; Big Data, Small Data, structured or unstructured with any format e.g. semi structured (spreadsheets, word docs, log files etc.), structured (WSDL, JSON, XML etc.), Database (Schemas), Application related (API structures, WSDL, copybooks etc.) All things being equal you can **match GDPR items**, with the required compliance in half the time and at least half the price compared to current services that are out there.



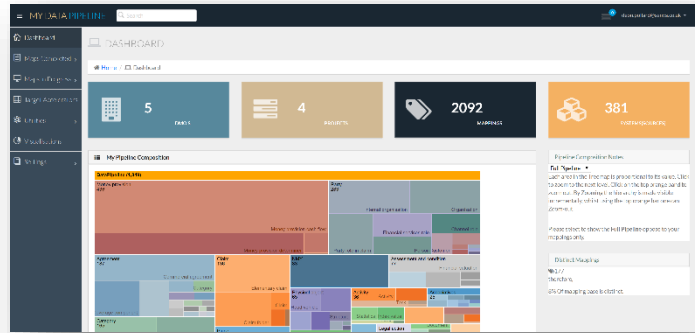
Fuzzy Logic Mapping uses Jaccard similarity, which is defined as the size of the set intersection divided by the size of the set union for two sets of objects. For example, the sets {a, b, c} and {a, c, d} have a Jaccard similarity of $2/4 = 0.5$ because the intersection is {a, c} and the union is {a, b, c, d}. The more that the two sets have in common, the closer the Jaccard similarity will be to 1.0.

the Senya Solution

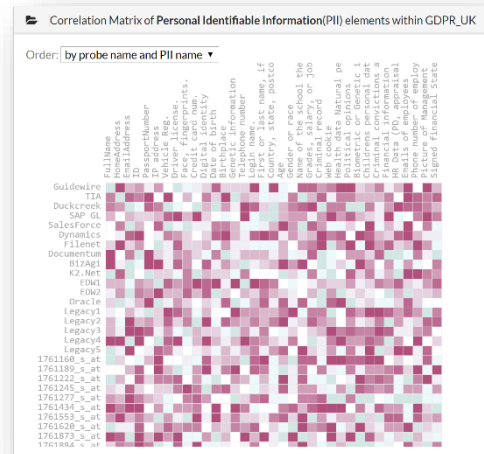
You can send any of your **meta-data** to a secure service and within 24hours get your **GDPR-aligned data back**. Each of your data elements are ranked via a similarity index to indicate a degree of surety around finding/matching your data to the GDPR categories. Senya can also help you to harvest Meta-Data.

Some call-out features:

1. Discover PII Data across small data, big data, structured and unstructured. Clearly see what data is used where in what context.



2. Correlate PII Data Types across your data sources. Very easy to see where your PII data is located and what potential risk impact they have.



3. Ability to fuzzy-logic match your data. Similarity index indicate a degree of surety around finding/matching your data to the GDPR categories.

Field	Description	ConversionRule	Similarity
Value	The value of the index that appIDecimal		54%
Status	Identifies the position of the <enum>		100%
The word or phrase that ident	String		100%
Value	The value of the index that appIDecimal		56%
The mutually exclusive cate	Identifier		88%
External reference	The number or code assigned to	String	53%
External reference	The number or code assigned to	String	94%
External reference	The number or code assigned to	String	100%
Full name	The complete aggregated name	String	100%
External reference	The number or code allocated to	String	79%
External reference	The number or code used for id	String	100%
The date at which the <event>	Date		100%
External reference	The number or code used for id	String	70%
Names	The city where the Risk Address is located.		88%
The location of the place; exp	String		95%
Opening date	The date at which this <claim>	Date	95%
Inception date	The date on which the agree	Date	100%
Planned end date	The planned end date of the <a	Date	94%
External reference	The number or code allocated to	String	0%
The mutually exclusive cate	Identifier		84%
The mutually exclusive cate	Identifier		100%

More questions about the SENYA solution

	SENYA Platform	Standard iDX Service	Extended iDX Service
Arrangement	Senya provide Platform and you do it yourself	Senya provide turn-key service with standard Service Level Agreement and accompanying services	Senya provide turn-key service with negotiated Service Level Agreement and accompanying custom services
Cost	Monthly subscription, pay-as-you-go;	Monthly subscription or price per Domain	Flexible; to-be-negotiated
Usage	Unlimited use within your Organization bounded to a Company Legal entity	Limited to agreed Domain/s. GDPR tag all data for agreed Domain/s	Limited to agreed Domain/s. GDPR tag all data for agreed Domain/s
Term	Month by Month	Month by Month	Annual Subscription
Human Guided Mapping SME's	To be provided by you	1 x SME	Flexible; to-be-negotiated
GDPR	PII Standard List 35 elements	PII Standard List 35 elements	PII, SHI, SPI, Other
Target Accelerators	○	●	●
Recordkeeping of Maps	○	●	●
Fuzzy Logic/ML	◐	●	●
Solution Hosting	In-Cloud/On-site	●	●

Supported Input Formats

Structured Data	Semi-Structured Data
<ol style="list-style-type: none"> 1. Data-in-place <ol style="list-style-type: none"> a. SQL type Database Schemas b. Cobol Copybooks c. CSV Headers d. Formats: Tabulated Text, SQL DDL or Excel 2. Data-in-motion <ol style="list-style-type: none"> a. API Parameters, Soap WSDL, Rest data parameters b. Message Headers and/or Payloads c. CSV Headers d. Formats: JSON, XML, Tabulated Text, CSV, or Excel 	<ol style="list-style-type: none"> 1. Data-in-place <ol style="list-style-type: none"> a. LOG File Headers b. Folder/File systems c. Formats: Tabulated Text, or Excel 2. Document Libraries <ol style="list-style-type: none"> a. Sharepoint b. Formats: Tabulated Text, or Excel 3. Meta-Data Libraries <ol style="list-style-type: none"> a. IGC Terms b. Formats: Tabulated Text, CSV, XML or Excel 4. Online Feeds / Social <ol style="list-style-type: none"> a. Twitter #Tags, HTML Tables b. Formats: Tabulated Text, XML or Excel
Unstructured Data	Big Data
<ol style="list-style-type: none"> 1. Data-in-place <ol style="list-style-type: none"> a. Email b. Images c. Office Docs (Word, PowerPoint, Excel) d. Formats: Folder/File properties & optionally actual emails/documents/images 	<ol style="list-style-type: none"> 1. Data-in-place <ol style="list-style-type: none"> a. HDFS Filesystem b. Hive Tables c. HBASE object structures d. Formats: JSON, XML, Tabulated Text, CSV, or Excel

Contact | info@senya.co.uk | www.senya.co.uk