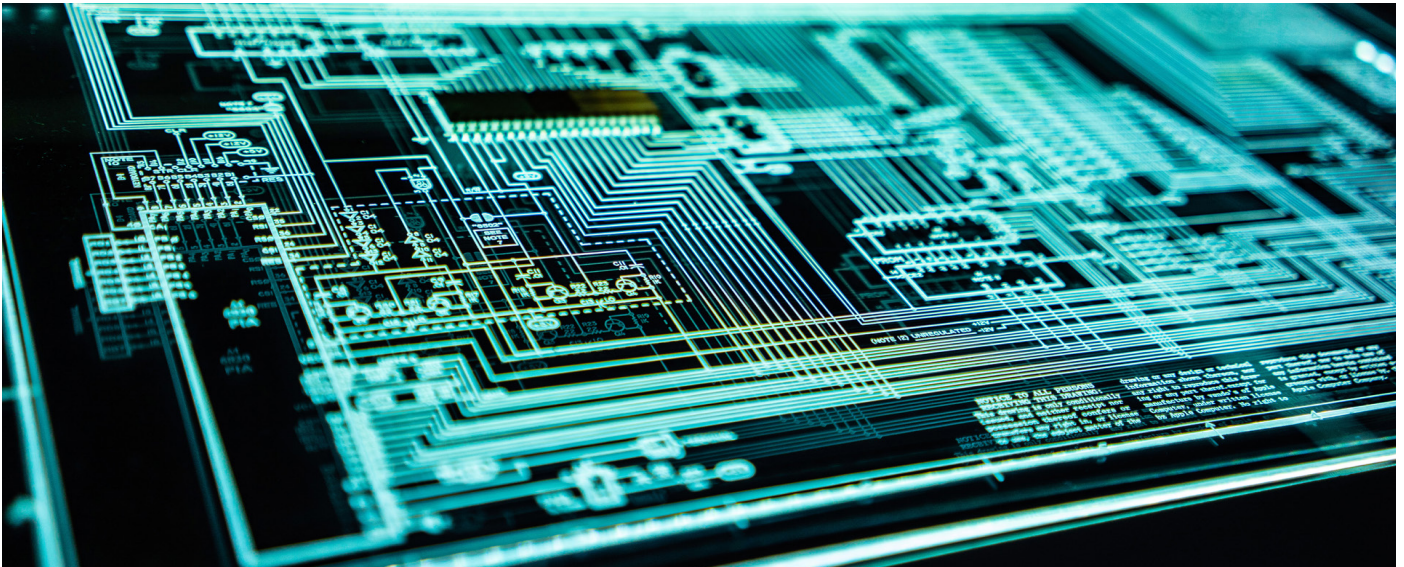




## ■ GDPR Document Analyzer

The automated solution for the evaluation and recognition of data



Each company receives, produces and stores documents of various types. The Italian legislation has already created the specific requirements, and more recently the General Protection Regulation of Data (GDPR, General Data Protection Regulation - EU Regulation 2016/679), make the data controllers responsible for the safety and security of the private citizen's data management.

The new European Regulation requires the introduction of corporate IT tools that support the work of the Supervisory Board, the Data Protection Officer and the Office Risks or Security Officers. All these offices must monitor the location of personal data on company storage, on the media provided to employees, and/or on services in the cloud used for operations. They must also detect the presence of personal data in the communications from and to the outside, or between the various company functions.

### The GDPR Document Analyzer tool

The RES Suite tool that can solve these needs is the GDPR Document Analyzer (G-DA), which allows the user to constantly monitor the location of documents containing personal information. The tool takes advantage of the most advanced systems for analyzing textual information and of images, regardless of their format. Within each document, personal information is identified and includes calculations of the probability index that the document contains such information. Later, these are made available through the generation of reports, or optionally they can be integrated into the dashboard tools already in use as part of the compliance function.

### Features of the GDPR Document Analyzer

The tool consists of 3 individual, yet related, components:

- Crawler: records the documents present in corporate or cloud storage
- Extractor: given the file containing the document, extracts the texts and images contained

- Analyzer: evaluates text and images to identify personal data which may be contained within the document being analyzed.

Depending on the needs, different crawlers compatible with the GDPR can be used. The Document Analyzer (G-DA), including the one integrated into the RES-IAD solution, which is part of the RES Suite.

The extractor makes use of the powerful Apache Tika library, the de facto standard for this application. Thanks to the unique capabilities of the tool, the GDPR Document Analyzer is capable of analyzing most of the most common documents, including:

- HTML
- XML and derivatives
- Microsoft Office (xls, doc, xlsx, docx, pptx)
- OpenDocument documents (odf)
- iWorks documents
- WordPerfect documents
- PDF (both textual and scanned)
- Ebooks (epub, fiction)
- RTF
- Compressed archives containing documents in analyzable formats
- Textual documents
- Email

The analyzer is the most innovative part of the GDPR Document Analyzer, it is based on Machine Learning algorithms, whose first appearances in scientific publications date back to 2013. By combining classical approaches based on rules and dictionaries, along with artificial intelligence algorithms, it increases the robustness of the personal data detector by increasing the accuracy and greatly reducing the likelihood of error. Thanks to these characteristics, the G-DA analyzer is able to work immediately on any type of document and to improve its accuracy over time. G-DA is able to learn from operator feedback reporting any evaluation errors, customizing each installation according to the most documents and formats commonly present in the company.

The characteristics of the analyzer:

- analyzes text in Italian and English (other languages possible on request);
- detects personal information (name, surname, addresses, social security numbers, telephone numbers, photos with faces, ...);
- does not require the text in any particular formats;
- does not require particular phrases or formulas to identify personal elements;
- can be used as a service, including SaaS in the Cloud, separately from the other components of the GDPR Document Analyzer solution.

GDPR Document Analyzer provides as output a detailed report indicating, for each document analyzed, the probability that it contains personal information. The tool also allows the user to obtain aggregate views by storage or directory of the amount of personal documents present.



[www.res-group.eu/en](http://www.res-group.eu/en)

Headquarters:  
Viale Piero e Alberto Pirelli, 6  
20126 Milano, Italy