Pseudalis

Use and share your documents without exposing the personal data in them

Pseudalis detects the personal data in your documents and replaces each value with a tag that holds it in encrypted form. Only the key of your computer can restore the document, or any text derived from it.

Version **1.0.11** Platform **Windows 11, x64, without driver**

License **Boost Software License 1.0**

Example *With tags*

```
Subject: contract renewal

Dear [[PERSON:Rnifypn Eouiitlrwky#EKR]],

Following our call of [[DATE:28/08/2018#VLC]], please
find the renewal for [[ORG:Xkekwmufwmcj Eouiitlrwky et
Wldwdst SARL#T4N]], [[LOCATION:Saint-Ntaahgp#QWB]].
The refund will be made to the account [[IBAN:GB45 NWBK
6152 7186 9288 16#K9P]].

You can reach me at [[PHONE:06 07 09 00 95#M2X]] or
[[EMAIL:rnifypn.eouiitlrwky@pspbshklkspj.com#7HD]].
```

0.946
F1 of the detector on a test set in five languages (version 1.0.11)

41,000
characters read per second on two cores

40
countries whose national numbers are verified

17 MB
installer that carries everything, without driver

At a glance

A tag carries its value
Each value becomes a tag that holds it encrypted with FF1 over AES-256, with a check. No correspondence table is kept.

With or without tags
The values can be written as their substitutes alone, so that the file reads as an ordinary document for a prompt, test data or a file sent outside.

Imposed terms and columns
Administrators impose up to 100 terms and 100 columns of CSV files, replaced before anything else is read, so that the joins between exported tables hold.

Pictures described by Albireo
With Albireo on the same computer, each picture is replaced by a description that names no person.

Documents keep their layout
Word, Excel, PowerPoint, PDF, text, CSV, Markdown, HTML, XML, JSON and YAML files keep their layout, and a PDF keeps its pages.

Names and properties
File names, archive entries and document properties take substitutes too, and come back at the restoration.

Originals kept encrypted
The originals go to an encrypted vault bounded in time and space, and come back on demand from any copy of the anonymized file.

Consistent substitutes
The same value takes the same substitute everywhere, a protected IBAN still passes its check, and the dates move together by a secret shift.

Built into the tools you already use

A drive
A file or a folder copied to the drive of Pseudalis is anonymized once written, its name included. No driver is installed.

Explorer commands
Anonymize and Restore on files, folders and selections of up to 100 items, with or without tags, in place or as a copy.

A web application
Drop files, archives and folders, or paste a text, and get them back anonymized or restored.

A local interface
A documented HTTPS interface on the loopback address for other programs: files, texts, names, records and settings.

Four detection devices

Rules spot the values that have a structure; two small trained models spot, from the context, the ones the rules miss.

Validated scanners

E-mail and IP addresses, telephone numbers, IBAN, cards and the national numbers of about forty countries, accepted when their check digits verify.

Structural SVM tagger

A sequence labeler over words that reads names, places, addresses and organisations from their context.

Word tables

44 tables in five languages: civilities, field labels, legal forms, given names, public bodies and about two thousand organisations.

Byte-level transformer

A span model of 116,791 parameters that reads any spelling and reports where a value starts and ends.

Measures of the detector

Whole detector, exact spans, every category (Pseudalis 1.0.5)

Test set	Precision	Recall	F1
Test, five languages	0.955	0.935	**0.945**
Values never seen in training	0.951	0.935	**0.943**
Six times the noise	0.941	0.897	**0.918**
Whole documents	0.961	0.943	**0.952**
Places abroad only	0.960	0.939	**0.949**

Categories of personal data

Category	State
Names, nicknames, logins	● Covered
Addresses, cities, postal codes with their city	● Covered
Telephone, e-mail, dates	● Covered
Civility, sex, nationality	● Covered
Identifiers with a check: social security, IBAN, cards, about 40 countries	● Covered
IP, MAC, UUID, IMEI, plates, VIN, GPS	● Covered
Organisations and named public bodies	● Covered
Identity card, driving license, residence permit	● Partly
Customer, case, contract and employee numbers	● Partly
Health, beliefs or criminal facts written in whole sentences	○ Not covered
Text inside pictures	○ Not covered

The limits of Pseudalis

- Detection is probabilistic in the learned categories: a protected document should be read again before it is shared.
- The trust model is the computer: every user of the computer shares its key, its originals and its administration.
- A scanned page without text layer is kept as a gray box: no OCR is performed.
- A document in a language other than the five that the recognizers read keeps more values readable.

Characteristics

Documents	Word, Excel and PowerPoint (.docx, .xlsx, .pptx and their macro and template variants), PDF
Texts	More than thirty extensions, among them .txt, .csv, .md, .json, .xml, .html, .yaml and .sql, and any file whose content is text
Languages read	French, English, Italian, Spanish and German
Encryption of the values	FF1 over AES-256, each tag authenticated with HMAC-SHA-256
Vault of the originals	AES-256-CTR with HMAC-SHA-256; master key protected by DPAPI; keys derived with HKDF-SHA-256
Local interface	https://127.0.0.1:8444, documented with Swagger
Dependencies	None: the installer carries everything, and the programs import the libraries of Windows alone

Standards and references

NIST SP 800-38G	FF1 format-preserving encryption of the values
AES-256, HMAC-SHA-256	Encryption and authentication of the tags and of the vault
HKDF-SHA-256, PBKDF2-SHA-256	Derivation of the keys and protection of the exported key
ECDSA P-256	Signature of the update feed
GDPR, Articles 25 and 32	Pseudonymization and encryption as measures of protection by design and of security
EDPB Guidelines 01/2025	Pseudonymisation, read as they stood in November 2026

More on
cydral-technology.eu/en/products/pseudalis/

Downloads
cydral-technology.eu/en/downloads/

Contact
cydral-technology.eu/en/contact/

Free and open source under the Boost Software License 1.0. © 2026 Cydral Technology, France.