😎 Duplicate of SMART REDACTION

CONTENTS

Search

Important links

Resource	Link	Details
Jira Epic	DBT-35044: Smart redaction for GDPR-compliance DONE	
Figma design concept		
Semantic space	Smart Redaction - SEMANTIC - Confluence (atlassian.net)	
Development updates	Smart Redaction Notes - SEMANTIC -	weekly updates from project team

	[Confluence (atlassian.net)]()	
Annotation guidelines	[semantic.gitlab-pages.drooms.com]()	Guidelines for data annotation for smart redaction

Project team (contact people)

- PM Lead: @Johannes Graf
- Semantic Lead:
- BE Lead:
- FE: Lead:
- QA Lead:
- Design Lead:

Background & Research

Legacy information

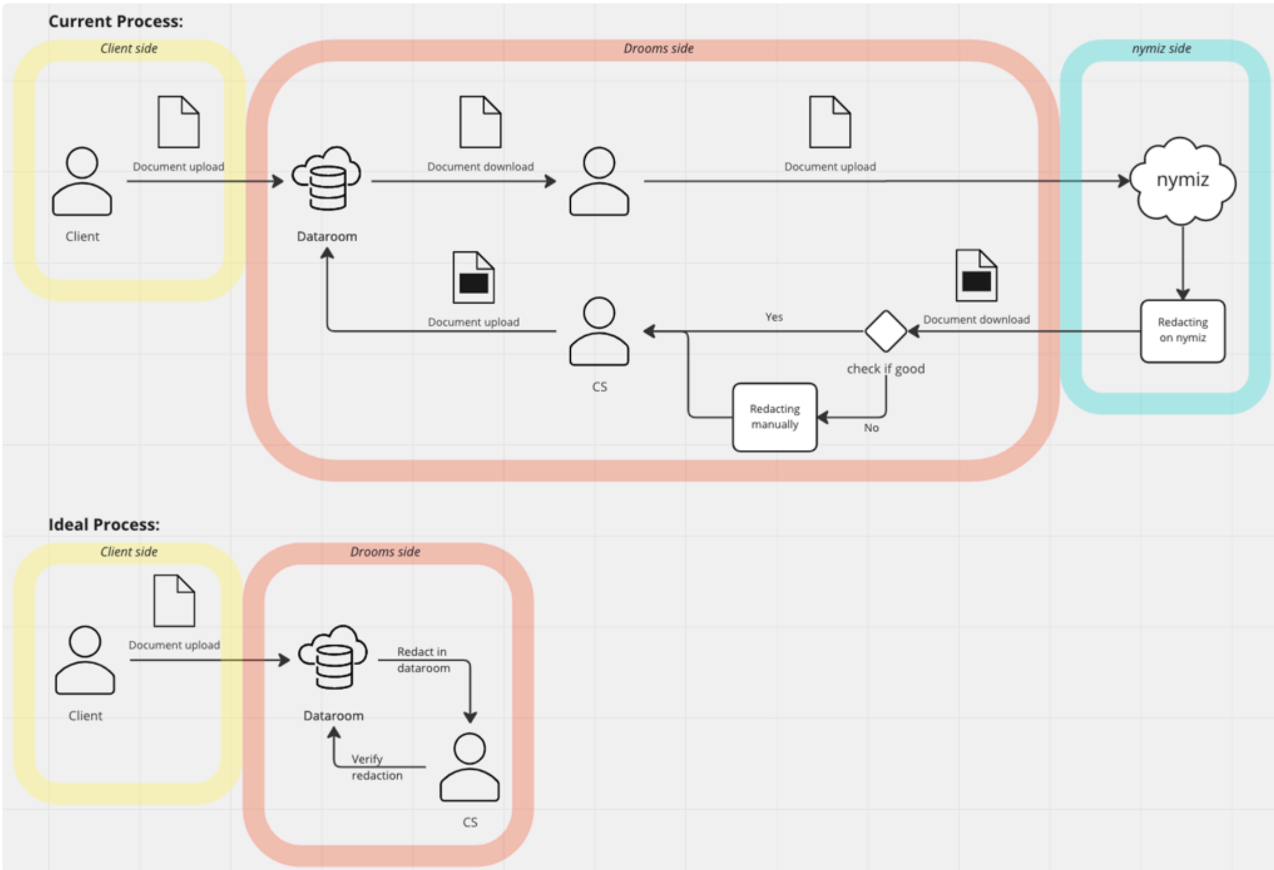
Redaction gained a lot of demand due to new regulations for GDPR compliance. Especially in the Spanish business of Drooms for Non-performing loans (NPL) transactions, there is a lot of sensitive personal data (PII like IBANs, addresses, IDs) that must be redacted. A lot of customers reach out to our Customer Success team to help them with redacting documents to make the DR GDRP-compliant.

The redaction is required all main Drooms languages:

- German
- English
- French
- Spanish
- French

Discussion with CS/Sales/Users

The current workflow refers to the Customer Success (CS) team at Drooms. The client provides and uploads the documents in the Drooms dataroom. The CS team then downloads these documents and uploads them to the external redaction platform *nymiz.* Then, they select categories that must be redacted in *nymiz* and let the tool process these. The CS employee then downloads the redacted documents and checks for errors



Meeting with clients in Zurich and RES (5th of May 2023)

Big advisory)

- They are not using a redaction tool inside a VDR, they do it manually with Adobe (issue with not deleting redacted data, able to search for redacted terms)
- Redaction is collaborative with Juniors doing it first and Seniors reviewing it
- They liked our simple UI and the different redaction methods
- Additional entities they require:
  - additional entities: price sensitive figures
  - Wages / salary
  - Positions / Title
  - Currencies + Symbol

investment boutique):

- After redaction they manually review it
- Buyside - extraction functionality in demand (reverse Redaction)
- new entity: nationality / Domizil

- Comparison with Intralinks: IL - only singular documents can be redacted at a time. They need to filter each time again. Better would be if they could store and select different templates
- long load time for redacted document at Intralinks

(big investment bank):
- Once document is redacted it needst to be edited
- Whitelist names: redact all names but Andy Chen
- Follow up pages → select pages manually
- Use Case: redact entities and terms across multiple documents at the same time
- Signature: when detected rather redact too much than too little. Entie signature plus Names
- Want to create their own list of names to redact/not redact in DR
- Shared audit trail in sandbox + TDR
- Extraction: content preview of docs
- Schnittstelle SharePoint für Kunden

Meeting Notes with Customer Success (12th of January 2023)

Redaction started in Spain with NPL (debt, so need to redact debtee), also relevant for M&A for GDPR compliance

- Corporate debt

1. Personal data most urgent for GDPR (most common)
2. Companies want to obfuscate info for investors (when non binding offer, initial selling phase)
3. IDs are the most critical thing to redact (primary keys)

Meeting with RES (6th of January 2023)

- Almost every DR provider has Redaction
- Task management should be reflected in Redaction
- Original and redacted document need to be separately saved
- Which categories are interesting? Customer success does it manually (specific Spain -
- GDPR compliance as top level requirement, categories as sublevel (most interesting!!!!)
- Many industries use redaction (real estate, lawyers, Corporate finance) and have different categories to redact
- Do we need to hide numbers? Such as rent?
- How does CS do redaction? The could create a labeled dataset for us ⟶ They use nymiz hosted by Drooms
- Nymiz has a list of categories ⟶ check access (done)

- Images can be classified, but not high priority
- Timeline: Depends on Design

How good is nymiz:

- Good on most categories,
- What is missing: account number (non IBAN), Signatures should be a category
- Issue that random chunks of text are redacted:
    - Numbers and codes that are critical are redacted but shouldn’t be
- Some Street adressess are names so there is overlap

What type of documents need to be redacted?

- ALL Contracts (service providers, rental contracts, maintenance)
- Payroll data and other HR stuff
- Board meetings, memorandums (for M&A)
- For NPL max 10 doc types (but shittons of instances per type)
    - Passport and ID (as images)

Competitors have multiple versions of redacted versions (Architrave) → Versioning is a general featuere request

- Use Case: Different versions for different stakeholder

Feature: Admins should be able to unredact sth. To compensate for mistakes by AI.

Data Types

- 99% of redacted documents are scanned documents
- 1% are Excel data ⟶ need a redact the Excel (formatted as Image?)

What entities in Nymiz are NOT needed?

- Currencies (investor needs number)
- IP

- NOR
- Location (city and country nothing personal), Only adress I GDPR relevant, redact streets but not cities (but Lucie in France have request to redact cities)
- City should be its own category!!!

Activate negative exceptions to exclude instances from a category!!! (eg., hide all names except XYZ)

In NPL lots of images of passports (need to be redacted could be done by AI)

Signatures:

- Some docs need signatures
- Index Structure for fake IDs
  - Not needed for our redaction tool
  - Would be great to remove names from Index (nice to have but not urgent)

Current workflow

- Get data from Client
- Open new DR (finance reason, billing)
- Upload in nymiz
- Upload back to DR

Ideal Workflow according

- Select a document or folder (better)
- Right click and redact
- Admin should be able to choose doc and visibility

Should be able to deliver original document in archives (after transaction to buyer)

Technical research

Smart Redaction - SEMANTIC - Confluence (atlassian.net)

Project description

Simple user story

As a user of Drooms, I want to use the smart redaction feature to highlight all categories of sensitive data like IBAN, Personal names, IDs, and passports to make my document GDPR-compliant without spending hours of manual work

Detailed description

Meeting summaries

Meeting minutes recorded and tracked in below pages.

Smart Redaction Notes - SEMANTIC - Confluence (atlassian.net)

Tasks

Project Completion Checklist

✅ **Before starting each project, make sure you have checked the below common tasks. Kick-Off Call preparation** NOT RELEVANT DONE IN PROGRESS	✅ **Research and Preparation Phase – Check all points below for completion** NOT RELEVANT DONE IN PROGRESS	✅ **Ready for development Phase – Check all points below for completion (phase can be overlapped with previous one to speed up and be agile)** NOT RELEVANT DONE IN PROGRESS	✅ **In Work – Check all points below for completion, this includes Rollout preparation by PMs** NOT RELEVANT DONE IN PROGRESS
☑ Create an Epic in Jira, link it to this project space	☑ Extend competitor research and complete Competitor, Market, Trend research Template	☑ QA ticket estimates	☐ All BE tickets completed
☑ Set estimated Start and End dates of Epic for Research	☑ Define clear project requirements for design	☐ QA testing ticket creation	☐ All FE Tickets completed

Preparation status	and developers make use of Jobs to be done or simple user story framework		
☑ Create [Project Documentation page]() in Confluence	☑ Create User flows and link Miro board	☑ FE and BE Technical task creation was completed	☐ QA testing completed
☑ Add general background information	☐ Gather relevant data on time, define metrics in exchange with Developers and stakeholders	☑ FE estimated all User stories	☐ Marketing and Design briefing prepared (can be done earlier as well)
☑ Add all links, including feature requests	☑ User Journey and final flow (Lo-fi/Wireframes) finalized	☑ initial Release date estimate available	☐ Alignment and final plan for marketing activities and campaigns
☐ Complete the [Product Brief Template]()	☑ Technical Research by developers completed		☐ Feature announcement in Product announcement prepared and developed
☑ Plan a kick-off call including all features leads, CS Leads, and other relevant stakeholders (Marketing, Finance, Sales) and bring project docs and product brief as the foundation	☑ Hi-fi/final designs created		☐ Effect on guided tours, updated or new one created, tested, code added from devs and released
☑ Plan recurring meetings with the core team and extended stakeholder team in the kickoff call	☑ Milestone meeting with all project involved people and relevant stakeholders and team leads		☐ Training material created and [published in Confluence]()
☑ Get rough estimate by Design, BE, FE regarding project effort, feasibility and resource outlook and add it to [corresponding Discovery page]().	☑ All Assets created by Design		☐ Training session planned and completed with Sales and CS
☑ Create a project chat involving only key stakeholders (PM, BE, FE, QA, DES, CSM)	☑ Write all user stories		☐ Release notes prepared
	☑ Add final links to the Epic and confluence project page (miro, figma...)		☐ Help Center article created or updated with CS
	☑ QA review of the tickets		☐ Marketing and relevant Design Material for Release announcements ready

	☑ Initiated pricing strategy with finance (when applicable)		☑ Pricing strategy is ready for release, finance articles created, contract updated
	☑ Initiated legal evaluation, impact and task (when applicable)		☑ Legal adjustment are ready for release like T&C adjustments
	☑ Confident with Epic scope: Review epic for first release and evaluate value compared to effort, is MVP fine or do we need more value or iterations		☐ Website adjustment are ready for release
			☐ Included in release candidate: add <release candidate number> here
			☐ Finalize documentation and close relevant tickets (DBT, RICE)

Recurring tasks during a project

- [x] Status update page
- [x] Exchange with developers for epic end estimates based on current project scope and status

After release

- [x] Monitor Metrics regularly, publish them in confluence and communicate them
- [x] Bug fixes
- [x] Follow next iterations based on your priorities

Project/other tasks

Specific tasks you need to do in this exact project

- [ ] Task_1

Tasks from Meetings

Adjust “Task report” for your project. Click on “Task report”, click editing button (pencil), and from “Space(s) and page(s)” select your project/project meetings.

Task report

Looking good, no incomplete tasks.

Sub-pages