**ROTEIRO DO PROJECTO  
Automação de Processos GEP**

*Excel + PDF + E‑mail + Relatório + FillForm*

Data: 17-09-2025

# 1. Sumário Executivo

Pretende‑se automatizar o ciclo de trabalho associado aos processos GEP, desde a recepção do PDF, extracção estruturada de dados e geração de artefactos, até ao envio de e‑mail e preparação do preenchimento automático no portal GEPProperty. Todo o projecto deverá residir na unidade H:, sem dependências do disco do sistema (C:).

## Objectivos

* Criar um ‘driver’ simples em Excel (.xlsm) com botões para cada etapa do processo.
* Extrair texto de PDFs com o utilitário Poppler/pdftotext instalado na unidade H:.
* Normalizar e guardar informação em ficheiros de trabalho (doc.txt, data.json) e gerar um relatório final (PDF).
* Abrir um rascunho de e‑mail no Outlook com dados do segurado preenchidos automaticamente.
* Preparar dados para preenchimento automático (fillform) no portal GEPProperty.

# 2. Arquitectura de Alto Nível

A solução organiza‑se em componentes independentes e substituíveis, orquestrados por macros VBA no Excel.

## Componentes Principais

|  |  |  |
| --- | --- | --- |
| Componente | Função | Tecnologia/Local |
| Excel .xlsm (UI) | Orquestração, botões e estado | Excel + VBA (H:) |
| Extracção PDF | Converter PDF→TXT | Poppler/pdftotext (H:\Programas instalados\Ferramentas\Poppler\bin) |
| (Opcional) OCR | Ler PDFs imagem | Tesseract (H:\Programas instalados\Ferramentas\Tesseract) |
| Processamento texto | Regex e normalização | VBA (VBScript.RegExp) |
| E‑mail | Abrir rascunho para envio | Outlook (Automation) |
| Word | Ler DOCX da peritagem e gerar relatório | Word (Automation) |
| Automação Web | Preencher GEPProperty | Power Automate Desktop (ou Playwright) |

# 3. Estrutura de Pastas (Unidade H:)

A estrutura seguinte isola cada processo, facilita auditoria e permite reprocessamento.

* H:\PROCESSOS\_GEP\\_templates\ — modelos de e‑mail e relatório
* H:\PROCESSOS\_GEP\\_logs\ — registos globais
* H:\PROCESSOS\_GEP\\_config\ — config.ini (caminhos utilitários)
* H:\PROCESSOS\_GEP\apn\_<ID>\origem\ — PDF original e anexos (DOCX da peritagem)
* H:\PROCESSOS\_GEP\apn\_<ID>\trabalho\ — doc.txt, data.json, relatorio\_fillform.pdf, log.txt
* H:\PROCESSOS\_GEP\apn\_<ID>\output\ — artefactos finais, e‑mails exportados, anexos

# 4. Fluxo do Processo (Etapas)

## ETAPA 0 — Setup (uma só vez)

* Instalação do Poppler/pdftotext em H:.
* Criação da estrutura H:\PROCESSOS\_GEP\ e activação de Macros/Localizações Fidedignas.
* Outlook e Word instalados e configurados.

## ETAPA 1 — Criar Processo e Importar PDF

1. Escolher o PDF no botão “Escolher ficheiro”.
2. Criar subpasta do processo (apn\_<ID>) e copiar o PDF para \origem\.
3. Converter PDF→TXT (pdftotext). Em caso de falha e detecção de imagem, opção de OCR (fase 2).
4. Guardar \trabalho\doc.txt e registar \trabalho\log.txt.

## ETAPA 2 — Extracção de Campos (regex)

Aplicar padrões regulares ao doc.txt e guardar resultado em data.json.

* Identificação do Segurado: Nome, Morada, Localidade, Código‑Postal, Telefone, E‑mail, NIF/NIPC.
* Dados do risco/apólice: Nº de Apólice, datas relevantes, outras referências.
* Validações mínimas (CP, e‑mail, NIF).

## ETAPA 3 — E‑mail ao Segurado (Outlook)

* Ler data.json e preencher placeholders (nome/email/apólice).
* Abrir rascunho em Outlook (Para, Assunto, Corpo HTML).
* Utilizador revê e envia.

## ETAPA 4 — Integração com o Word da Peritagem

* Depositar o DOCX com Descrição, Causas e Conclusões em \origem\.
* Ler o DOCX via Word Automation e normalizar para texto.

## ETAPA 5 — Relatório “Dados para FillForm”

* Agregação de data.json + texto do DOCX num relatório final.
* Exportação para PDF (Word → PDF), com cabeçalho do processo e secções bem marcadas.
* Opcional: gerar também fillform.json (chave/valor) para automação Web.

## ETAPA 6 — Automação do GEPProperty (Fase 3)

* Tecnologia recomendada: Power Automate Desktop (PAD) sobre Microsoft Edge.
* Alternativa técnica: Playwright (Node/Python).
* A automação consome fillform.json e preenche campos/combos, faz uploads e submete.

## ETAPA 7 — Encerramento do Processo

* Fecho de logs e movimentação de artefactos finais para \output\.
* Limpeza da folha Excel e preparação para novo processo.

# 5. Dados e Ficheiros Produzidos

|  |  |  |
| --- | --- | --- |
| Ficheiro | Conteúdo | Uso |
| doc.txt | Texto integral extraído do PDF | Fonte para regex e auditoria |
| data.json | Campos estruturados (chave/valor) | E‑mail, Relatório e FillForm |
| relatorio\_fillform.pdf | Dados + Descrição/Causas/Conclusões | Apoio humano e evidência |
| log.txt | Eventos/erros por processo | Diagnóstico e rastreabilidade |

# 6. UI no Excel (.xlsm)

* [Escolher ficheiro] — cria processo, converte PDF→TXT, actualiza RESUMO.
* [Abrir pasta] — abre a pasta do processo activo.
* [Limpar] — limpa B2:B10 e estado.
* [Email ao Segurado] — gera rascunho no Outlook com dados do data.json.
* [Elaborar Relatório (Dados para fillform)] — agrega data.json + DOCX e emite PDF.
* [Elaborar Orçamento] — reservado para fase futura.

# 7. Regras de Extracção (exemplos de padrões)

Todos os padrões são case‑insensitive e poderão ter variantes. Exemplos indicativos:

* Nome/Segurado: ^(Nome|Segurado)\s\*[:\-]\s\*(?P<nome>[^\r\n]+)
* Código‑Postal: (?P<cp>[0-9]{4}-[0-9]{3})
* E‑mail: (?P<mail>[A-Za-z0-9.\_%+\-]+@[A-Za-z0-9.\-]+\.[A-Za-z]{2,})
* NIF/NIPC: (?P<nif>\b[0-9]{9}\b)
* Nº Apólice: (?i)N[ºo]\s\*de\s\*Apólice\s\*[:\n]\s\*(?P<apol>[A-Z0-9]+)

Os padrões finais serão ajustados com base em amostras reais do doc.txt.

# 8. Requisitos e Dependências

* Windows 10/11 com Excel, Outlook e Word.
* Poppler/pdftotext em H:\Programas instalados\Ferramentas\Poppler\bin
* (Fase 2) Tesseract OCR em H:\Programas instalados\Ferramentas\Tesseract
* Macros activas e pasta do projecto como Localização Fidedigna no Excel.
* Conta de Outlook configurada.

# 9. Erros, Logging e Recuperação

* Mensagens claras no Excel (conversão falhada, campos não extraídos, Outlook indisponível).
* Registo por processo em trabalho\log.txt e registo global em \_logs\.
* Fallback para OCR quando o PDF não contém texto (fase 2).

# 10. Segurança e Privacidade

* Todos os dados residentes na unidade H:.
* Não são guardadas credenciais; automação Web sem passwords em claro.
* Possibilidade de encriptação/ZIP de saída para partilha externa (opcional).

# 11. Plano de Implementação e Critérios de Aceitação

## Fase 1 — MVP

* PDF→doc.txt; extracção de 6–10 campos; data.json.
* UI com 5 botões; e‑mail Outlook; relatório PDF com DOCX.
* Estrutura de processos em H:.

## Fase 2 — Robustez

* OCR automático; regex refinadas; templates configuráveis em \_templates.

## Fase 3 — Automação Web

* Preenchimento no GEPProperty (PAD/Playwright) a partir do fillform.json; evidências (screenshots).

## Critérios de Aceitação (amostra)

1. Ao seleccionar PDF, cria‑se H:\PROCESSOS\_GEP\apn\_<ID>\ com origem\, trabalho\, output\.
2. trabalho\doc.txt contém o texto integral; data.json guarda campos obrigatórios.
3. ‘Email ao Segurado’ abre rascunho com Para/Assunto/Corpo preenchidos.
4. ‘Elaborar Relatório’ gera relatorio\_fillform.pdf com dados + texto do DOCX.