

ROTEIRO DO PROJECTO — Automação de Processos GEP (Excel + PDF + E-mail + FillForm) Data: 17-09-2025

1) SUMÁRIO EXECUTIVO

Objectivo

- Automatizar o ciclo “processo GEP” desde a recepção do PDF, extracção de dados, criação de artefactos (doc.txt/relatório), envio de e-mail ao segurado e preparação de dados para preenchimento automático no portal GEPPROPERTY.
- Tudo residente na unidade H:, sem depender do sistema (C:).

Pontos-chave

- Aplicação ‘driver’: um ficheiro Excel .xlsm com interface simples (botões) e macros VBA.
- Extracção de texto a partir de PDF com Poppler/pdftotext (instalado em H:).
- (Opcional / Fase 2) OCR com Tesseract, apenas quando o PDF for imagem.
- Geração de artefactos normalizados: doc.txt (texto integral), data.json (campos extraídos), Relatório.pdf (dados + descrições).
- Envio de e-mail via Outlook (rascunho) com dados puxados do data.json.
- (Fase 3) Preenchimento automático do site GEPPROPERTY (Relatórios GEP) com Power Automate Desktop ou Playwright.

2) ARQUITECTURA GERAL

Componentes

- Excel .xlsm (UI + orquestração): botões “Escolher ficheiro”, “Abrir pasta”, “Limpar”, “Email ao Segurado”, “Elaborar Relatório (Dados para fillform)”, “Elaborar Orçamento” (futuro).
- Motor de extracção:
 - Poppler: pdftotext.exe (H:\Programas instalados\Ferramentas\Poppler\bin\pdftotext.exe).
 - (Opcional) Tesseract OCR (H:\Programas instalados\Ferramentas\Tesseract\tesseract.exe).
- Processamento de texto: regex em VBA (VBScript.RegExp) para capturar campos.
- Office: Outlook (enviar rascunho) e Word (abrir/generar docx vindo dos GPTs).
- Automação Web (fase 3): Power Automate Desktop (recomendado para Windows) ou Playwright (Node/Python), controlado a partir do Excel.
- Registo e configuração: ficheiro config.ini (paths), log.txt por processo.

Fluxo de Alto Nível

PDF → (pdftotext / OCR) → doc.txt → parsing (regex) → data.json → Relatório.pdf → (Outlook: e-mail) → (GEPPROPERTY: fillform).

3) ESTRUTURA DE PASTAS (H:)

H:\PROCESSOS_GEP\

_templates\ -> modelos de e-mail e relatório

_bin\ -> atalhos/auxiliares (opcional)

_logs\ -> logs globais

_config\

config.ini -> caminhos (pdftotext, tesseract, etc.)

apn_25MR052881\ -> UM PROCESSO

origem\ -> cópia do PDF original e anexos

apn_25MR052881.pdf

peritagem_descricao_causas_conclusoes.docx

trabalho\

doc.txt -> texto extraído integral

data.json -> campos estruturados

relatorio_fillform.pdf -> relatório final para apoio ao site

log.txt -> registo de passos/erros do processo
output\
email_rascunho.msg -> (opcional) export de rascunho
anexos\ -> artefactos a anexar em e-mails (se houver)

4) UI NO EXCEL (.XLSM)

Folha RESUMO

- Secção 'Caminhos' (B3/B4).
- Secção 'Campos' (B6: Nome/Segurado, B7: Código Postal, B8: NIF/NIPC ... extensível).
- Barra de botões:
[Escolher ficheiro] → lança importação e extracção; cria estrutura do processo
[Abrir pasta] → abre a subpasta do processo actual
[Limpar] → limpa B2:B10
[Email ao Segurado] → abre rascunho no Outlook com e-mail pré-preenchido
[Elaborar Relatório (Dados para fillform)] → gera Relatório.pdf com dados extraídos + descrições Word
[Elaborar Orçamento] → (fase futura)

5) PIPELINE POR ETAPAS (detalhe)

ETAPA 0 — Setup (uma vez)

- Instalar Poppler (pdftotext.exe) em H:. Guardar caminho em config.ini ou Const no módulo.
- Criar H:\PROCESSOS_GEP\ e subpastas base.
- Activar macros (e pôr H:\PROJECTOS\... como Localização Fidedigna).
- Outlook instalado e configurado.

ETAPA 1 — Criar Processo e Importar PDF

Entrada: PDF (apn*.pdf).

Passos:

1. Utilizador clica "Escolher ficheiro".
 2. Macro cria subpasta H:\PROCESSOS_GEP\apn_XXXX\ (nome do ficheiro, normalizado).
 3. Copia o PDF para origem\ e também para Desktop\Uploads (opcional).
 4. Converte PDF→TXT (pdftotext). Se falhar e for imagem, caminho alternativo: OCR (Tesseract) → TXT.
 5. Guarda trabalho\doc.txt (texto integral). Escreve trabalho\log.txt.
- Saídas: doc.txt, log inicial.

ETAPA 2 — Extração de Campos (regex)

- Lê doc.txt e aplica padrões:
Nome/Segurado, Morada, Localidade, Código Postal, Telefone, E-mail,
Nº de Contribuinte (NIF/NIPC), Nº de Apólice, Data(s) relevantes, etc.
- Validações mínimas: e-mail (regex), CP (formato XXXX-XXX), NIF (9 dígitos).
- Grava trabalho\data.json com o dicionário de campos + timestamps.
- Actualiza Folha RESUMO (B6..).

ETAPA 3 — Email ao Segurado (Outlook)

- Lê data.json → obtém {{nome}}, {{email}} e outros placeholders.
- Constrói um rascunho (HTML) via Outlook.Automation:
Para: email
Assunto: "Processo GEP – {{nome}} / {{apólice}}"
Corpo: template (_templates\email_template.html) com placeholders.
- Abre com .Display (utilizador revê e clica Enviar).

ETAPA 4 — Integração com o Word da Peritagem

- O utilizador deposita em origem\ o DOCX gerado pelos GPTs ("descrição, causas, conclusões").

- Macro valida presença desse DOCX e extrai o texto (via automação Word).
- Guarda conteúdo em trabalho\peritagem.txt (ou incorpora directamente na geração do relatório).

ETAPA 5 — Relatório (Dados para fillform)

- Agrega data.json + peritagem.txt num único artefacto “relatorio_fillform.pdf”:
Cabeçalho do processo; bloco “Segurado/Interessado”; tabela de campos;
secções: Descrição, Causas, Conclusões (do DOCX).
- Alternativa de formato técnico: gerar também “fillform.json” (chave/valor) para consumo pela automação do site.

ETAPA 6 — Automação GEPPProperty (Fase 3)

- Tecnologia recomendada: Power Automate Desktop (PAD) — robusto para Windows/Edge e sem codificação pesada.
- Alternativa: Playwright (Node/Python) com script chamável a partir do Excel.
- A automação lê fillform.json e preenche os campos/combo no portal, faz uploads e submete.
- Regista screenshot/ficheiro de confirmação em output\.

ETAPA 7 — Encerramento do Processo

- Macro “Fechar Processo” (opcional): valida que relatório foi emitido, e-mail enviado e portal preenchido.
- Move artefactos finais para output\, fecha logs e prepara folha para novo processo (Limpar + reset estado).

6) ESPECIFICAÇÃO DE DADOS

Ficheiro trabalho\data.json (exemplo)

```
{
  "processo_id": "apn_25MR052881",
  "ficheiro_origem": "apn_25MR052881.pdf",
  "extraido_em": "2025-09-17T12:00:00Z",
  "segurado": {
    "nome": "ADM COND EDIFICIO GALERIAS STO ANTONIO",
    "morada": "R COMBATENTES GRANDE GUERRA 21",
    "localidade": "SÃO JOÃO DA MADEIRA",
    "codigo_postal": "3700-087",
    "telefone": "256807170",
    "email": "GERAL@KASACONDOMINIOS.COM",
    "nif_nipc": "900974460"
  },
  "apolice": {
    "numero": "MR23889151",
    "companhia": "GEP",
    "datas": {
      "emissao": "2025-09-16"
    }
  }
}
```

Campos mínimos a suportar (onde houver no PDF):

- Nome/Segurado, Morada, Localidade, Código Postal, Telefone, E-mail, NIF/NIPC
- Nº de Apólice, Datas relevantes (emissão/ocorrência), Referências internas

7) REQUISITOS / DEPENDÊNCIAS

- Windows 10/11 com Office (Excel, Outlook, Word) instalado.
- Poppler (pdftotext.exe) em H:\Programas instalados\Ferramentas\Poppler\bin\
- (Fase 2) Tesseract OCR em H:\Programas instalados\Ferramentas\Tesseract\

- Permitir macros no Excel; adicionar a pasta do projecto como Localização Fidedigna.
- Conta de Outlook configurada.
- Espaço livre na unidade H:.
- (Automação Web) Power Automate Desktop ou Playwright + Microsoft Edge/Chrome.

8) ERROS / LOGGING / RECUPERAÇÃO

- Log por processo: trabalho\log.txt (cada etapa regista sucesso/erro e tempo).
- Mensagens claras no Excel: falha converter PDF, ficheiro não encontrado, campo não extraído, Outlook indisponível, etc.
- Retentativas: se pdftotext falhar, oferecer OCR (fase 2).
- Validações de campos: realçar em amarelo valores ausentes ou inválidos.

9) SEGURANÇA & PRIVACIDADE

- Tudo em H: (dados locais). Não enviar dados para a nuvem, excepto o e-mail via Outlook.
- Evitar guardar credenciais. Automação web sem armazenar passwords em claro.
- Possibilidade de encriptação/ZIP protegido em output\ (opcional) antes de partilha.

10) PLANO DE IMPLEMENTAÇÃO (fases) Fase 1 (MVP) - Importar PDF → doc.txt - Extrair 6–10 campos principais → data.json - UI Excel: 5 botões chave - E-mail Outlook (rascunho) com Nome/Email do segurado - Geração de Relatório.pdf (dados + descrições do DOCX) - Estrutura de processos em H:

Fase 2

- OCR automático (Tesseract) quando pdftotext detectar texto nulo
- Melhoria de regex/normalização (datas/CP/telefone)
- Templates configuráveis (email/relatório) na pasta _templates

Fase 3

- Automação GEPPROPERTY (PAD/Playwright) a partir do fillform.json
- Captura de evidências (screenshots / recibos) em output\
- Botão “Fechar Processo” (arquivo e limpeza)

11) CRITÉRIOS DE ACEITAÇÃO (por fase) F1-1: Ao seleccionar PDF, cria-se H:\PROCESSOS_GEP\ com origem\, trabalho\, output\ F1-2: trabalho\doc.txt contém o texto integral do PDF. F1-3: trabalho\data.json inclui nome/email/NIF/CP/apólice quando constarem no PDF. F1-4: Botão “Email ao Segurado” abre rascunho com Para/Assunto/Corpo preenchidos. F1-5: “Elaborar Relatório” cria output trabalho\relatorio_fillform.pdf com campos + secções do DOCX.

12) NOTAS TÉCNICAS (nível alto) - VBA em late binding (sem referências fixas), módulos: importação, extracção, email, relatório, utilitários. - Regex centralizadas (PATTERN_*) com fallback e testes. - Uso do Word.Application para ler DOCX (descrição/causas/conclusões). - Geração do PDF do relatório: • via Word (modelo .dotx) preenchido e exportado como PDF, ou • via biblioteca PDF (mais técnico; Word é mais rápido e robusto). - Automação Web: PAD recomendada pela integração nativa Windows/Edge.

FIM