

O que é o DataExtract?

Extração e processamento de informação constante em documentos em formato PDF

Através deste desafio, pretende-se a extração da informação dos DIFs dos PRIIPs notificados previamente à ASF, com vista à validação dentro do prazo legal de 2 dias úteis e ao tratamento dos dados extraídos para identificação de tendências e riscos.

Os PDF notificados referem-se tanto a novos PRIIPs como a PRIIPs já existentes que sejam atualizados, podendo variar consideravelmente em complexidade, desde produtos simples a produtos com várias opções de investimento.

Notas

- / **PRIIP** (*Packaged Retail and Insurance-based Investment Products*) são produtos de investimento de retalho ou produtos de investimento com base em seguros;
- / **DIF** (Documento de Informação Fundamental) é o documento pré-contratual normalizado que o produtor do PRIIP é obrigado a elaborar e disponibilizar ao investidor não profissional antes da subscrição;
- / A ASF não tem poderes de aprovação de produtos. A validação do DIF consiste na aceitação da respetiva notificação, mediante a verificação da conformidade do documento com os requisitos regulamentares aplicáveis.

Estado atual

A entidade notifica previamente o DIF à ASF através de uma plataforma eletrónica dedicada, preenchendo um conjunto de campos num *template* estruturado e fazendo *upload* do ficheiro PDF do documento.

No caso das alterações, é enviado um PDF “limpo” e um PDF com as alterações registadas (“*tracked changes*”).

A partir da submissão, a ASF dispõe de um prazo de 2 dias úteis para aferir a conformidade do DIF com os requisitos regulamentares.

Cada DIF é analisado individualmente por um técnico, com recurso a uma *checklist* de verificação, sem qualquer ferramenta de automatização ou de apoio à análise.

Com base na análise realizada, a ASF valida ou rejeita o documento, devolvendo-o à entidade notificante para correção, quando aplicável. Após validação, a entidade fica obrigada a disponibilizar o DIF aquando da comercialização do produto, passando o documento a constituir informação pública.

a. Limitações & riscos associados

- / Desadequação entre o prazo para análise (2 dias úteis) e o volume de DIFs recebidos (743, em 2025), sobretudo em períodos de maior afluxo de notificações;
- / Validação inteiramente manual, morosa e sujeita a variabilidade de critério entre técnicos;
- / Ausência de extração da informação substantiva contida no DIF para aproveitamento analítico;
- / Inexistência de uma base de dados estruturada que agregue o conteúdo dos DIFs ao longo do tempo.

Objetivos a alcançar

a. Resultados funcionais

- / Aumentar a capacidade e eficiência da análise da documentação, através da identificação automática de incumprimentos de requisitos regulamentares nas notificações recebidas;
- / Assegurar a uniformização de análise, reduzindo risco operacional.

b. O que se espera da solução tecnológica?

- / Extração automática do conteúdo do DIF a partir do PDF, independentemente de variações de *layout* e de tamanho;
- / Estruturação dos dados extraídos em base de dados;
- / Identificação automática de incumprimento de requisitos regulamentares e outros critérios e regras parametrizáveis, definidos e disponibilizados pela ASF;
- / Repositório das inconformidades identificadas;
- / Cálculo de indicadores a partir da informação extraída;
- / Apresentação gráfica dos resultados, com vista à identificação de tendências e riscos;
- / A solução deverá permitir a validação e edição dos resultados gerados (por exemplo, indicadores calculados, gráficos produzidos e verificação de não cumprimento) pelo técnico.

c. Que resultados seriam úteis?

Criação de um *dashboard* dinâmico, com informação atualizada em tempo real, que disponibilize alertas automáticos em caso de não conformidade ou incoerência detetada no processo e indicadores calculados com base na informação extraída de forma sistemática.

Indicadores de desempenho

- / Tempo médio de tratamento por DIF/técnico
Base: Para um DIF simples, relativo a um novo PRIIP e que cumpra integralmente todos os requisitos, o técnico necessita de cerca 2 a 3 horas de análise
Meta: Processamento e a validação do DIF em tempo real, após a notificação;
- / Volume e qualidade da informação disponível para análise
Base: Não é recolhida informação
Meta: Recolha de informação relevante do DIF, p. ex.: tipo produto, público-alvo, nível de risco, prazo recomendado, custos e cenários de desempenho, entre outros.

Enquadramento legal & regulamentar a considerar

a. Legislação & regulamentação aplicáveis

- / Regulamento (UE) 1286/2014 (Regulamento PRIIPs);
- / Regulamento Delegado (UE) 2017/653, que estabelece as normas técnicas de regulamentação sobre apresentação, conteúdo, reexame e revisão do DIF, alterado pelo Regulamento Delegado (UE) 2021/2268 (aplicável desde 1 de janeiro de 2023);
- / Lei n.º 35/2018, de 20 de julho — Anexo II, que adapta a ordem jurídica interna ao Regulamento PRIIPs.

b. Requisitos RGPD: N.a.

Dados

a. Dados disponíveis

- / DIFs em PDF submetidos através da plataforma de notificação prévia;
- / Informação registada nos campos do *template* da plataforma de notificação prévia;
- / Dados não estruturados com os resultados da validação.

b. Dados necessários: DIFs em PDF submetidos através da plataforma de notificação prévia.

c. Restrições de acesso & confidencialidade: Será necessário definir condições específicas de confidencialidade e de acesso à informação nas fases de notificação prévia e de validação do DIF, dado que o documento só se torna informação pública após a conclusão deste processo.

d. Tratamento de dados pessoais: N.a.

O que esperamos?

- / Extração estruturada da informação a partir de PDFs independentemente de variações de *layout*;
- / Estruturação dos dados extraídos em base de dados;
- / Possibilidade de consulta das referências legais e regulamentares associadas à identificação de requisitos desconformes;
- / Capacidade de adaptação a novos enquadramentos legais, regulamentares;
- / Repositório das inconformidades identificadas na revisão;
- / Auditabilidade e rastreabilidade do resultado gerado, dado o seu impacto na decisão de aprovação ou rejeição do documento;
- / Cálculo automático de indicadores;
- / Construção de um *dashboard* para identificação de tendências e riscos;
- / Integração com a plataforma de notificação prévia existente e cumprimento do prazo de 2 dias úteis em vigor;
- / Garantia de supervisão humana, devendo a solução funcionar como instrumento de apoio à decisão e não como substituto da análise especializada.

What is DataExtract?

Extraction and processing of information included in PDF

The purpose of this challenge is to extract the information contained in the KIDs of PRIIPs previously notified to ASF, with a view to validating them within the 2-business-day legal deadline, while processing the extracted data to identify trends and risks.

The notified PDFs relate both to new PRIIPs and to existing PRIIPs being updated, and can vary considerably in complexity, ranging from simple products to products with multiple investment options.

Notes

- / **PRIIP** (Packaged Retail and Insurance-based Investment Products) are retail investment products or insurance-based investment products;
- / **DIF** (Key Information Document) is the standardised pre-contractual document that the PRIIP manufacturer is required to prepare and make available to the non-professional investor prior to subscription;
- / ASF does not have product approval powers. The validation of the KID consists of accepting its notification after verifying that the document complies with the applicable regulatory requirements.

As-is

The entity pre-notifies the KID to ASF through a dedicated electronic platform, filling in a set of fields in a structured template and uploading the PDF file of the document.

In the event of amendments, both a "clean" PDF version and a tracked changes version are submitted.

From the submission date, ASF has a 2-business-day deadline to assess the KID's compliance with regulatory requirements.

Each KID is reviewed individually by a technician, using a verification checklist, without any automation tool or analytical support.

Based on the review carried out, ASF validates or rejects the document, returning it to the notifying entity for correction where applicable. Once validated, the entity is required to make the KID available when the product is marketed, at which point the document becomes public information.

a. Limitations & associated risks

- / Mismatch between the review deadline (2 business days) and the volume of KIDs received (743 received in 2025), particularly during peak notification periods;
- / Validation is entirely manual, time-consuming and subject to variability in criteria between technicians;
- / No extraction of the substantive information contained in the KID for analytical purposes;
- / No structured database aggregating KID content over time.

Goals to achieve

a. Functional results

- / Increase the capacity and efficiency of document analysis by automatically identifying non-compliance with regulatory requirements in received notifications;
- / Ensure consistency and standardization of the analysis process, reducing operational risk.

b. What is expected from the technological solution?

- / Automatic extraction of KID content from the PDF, regardless of layout and size variations;
- / Structuring of extracted data into a database;
- / Automatic identification of the regulatory requirements that are not being duly complied and other parameterizable criteria defined and made available by the ASF;
- / Repository of non-conformities identified;
- / Calculation of indicators based on the extracted information;
- / Graphical presentation of results, to identify trends and risks;
- / The solution shall allow validation and editing of the outputs (e.g. indicators computed, visuals produced and non-compliance checks) by the staff.

c. Which outputs would be desirable?

Creation of a dynamic dashboard, with real-time updated information, providing automatic alerts in cases of non-compliance or inconsistency detected during the process, and indicators systematically calculated from the extracted information.

Success metrics

- / Average processing time per KID/technician
Baseline: For a simple KID, relating to a new PRIIP that fully complies with all requirements, the technician needs around 2 to 3 hours of review
Target: Real time processing and validation of the KID upon notification;
- / Volume and quality of information available for analysis
Baseline: No information is currently collected
Target: Collection of relevant KID information, e.g. product type, target market, risk level, recommended holding period, costs and performance scenarios, among others.

Legal & regulatory framework to consider

a. Applicable legislation & regulation

- / Regulation (EU) 1286/2014 (PRIIPs Regulation);
- / Commission Delegated Regulation (EU) 2017/653, laying down regulatory technical standards on the presentation, content, review and revision of the KID, as amended by Commission Delegated Regulation (EU) 2021/2268 (applicable since 1 January 2023);
- / Law No. 35/2018, of 20 July — Annex II, transposing the PRIIPs Regulation into national law.

b. GDPR requirements: N.a.

Data

a. Dados disponíveis

- / KIDs in PDF format submitted through the prior notification platform;
- / Information recorded in the fields of the prior notification platform template;
- / Unstructured data containing validation outcomes.

- b. Necessary data:** KIDs in PDF format submitted through the prior notification platform.
- c. Access restrictions & confidentiality:** Specific confidentiality and information access conditions will need to be defined for the prior notification and KID validation stages, given that the document only becomes public information once this process is complete.
- d. Personal data processing:** N.a.

Call for action

- / Structured extraction of information from PDFs regardless of layout variations;
- / Structuring of extracted data into a database;
- / Ability to consult the legal and regulatory references associated with identified non-compliant requirements;
- / Capacity to adapt to new legal and regulatory frameworks;
- / Repository of non-conformities identified during revision;
- / Auditability of automatic validations, given their impact on the decision to accept or reject the document;
- / Automatic calculation of indicators;
- / Construction of a dashboard for identifying trends and risks;
- / Integration with the existing prior notification platform and compliance with the current 2 business day deadline;
- / Human oversight must be ensured; the solution should function as a decision-support tool, and not as a substitute for expert analysis.