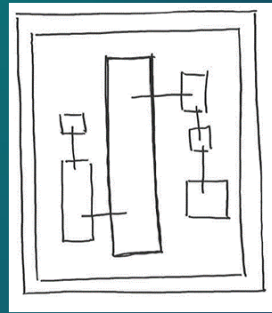


INFRAESTRUTURA NACIONAL DE DADOS



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL



ENCONTROS SAMA²⁰²⁰

02 OUT

11h15-13h00
LISBOA

ama AGÊNCIA PARA A
MODERNIZAÇÃO
ADMINISTRATIVA

COMPETE
2020

PORTUGAL
2020

UNIÃO EUROPEIA
Fundo Europeu
de Desenvolvimento Regional

Infraestrutura Nacional de Dados :: IND

ENCONTROS
SAMA 2020

- Sobre o projeto SAMA – Execução
- Objetivos alcançados
- O que é?
- Parcerias
- Overview de processos de tratamento
- Vantagens
- Desenvolvimentos em curso
- Alguns casos de uso

IND :: Sobre o projeto SAMA - Execução

ENCONTROS
SAMA 2020

Custo total elegível: 987 061,50 €

Apoio financeiro da UE: 562 131,52 € (56,95%)

Apoio financeiro público nacional: 424 929,97 € (43,05%)

Apoio

● Aprovado

556 230,93 €

● Validado

556 230,93 € (100%)

● Pago

556 230,93 € (100%)

IND :: Objetivos alcançados

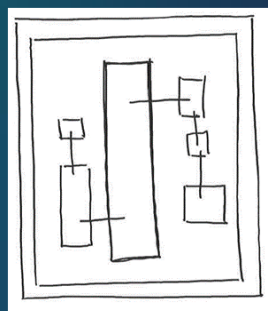
ENCONTROS SAMA²⁰²⁰

uso mais intensivo e integrado dos dados

ganhar recursos, espaço para intensificar a inovação

alargamento substancial dos domínios cobertos

modernização da infraestrutura tecnológica



aplicação de novas técnicas de extração de informação

aproveitar toda a cadeia produtiva

intensificação da apropriação e utilização de dados administrativos e de outras fontes

maior retorno à sociedade

IND :: O que é?

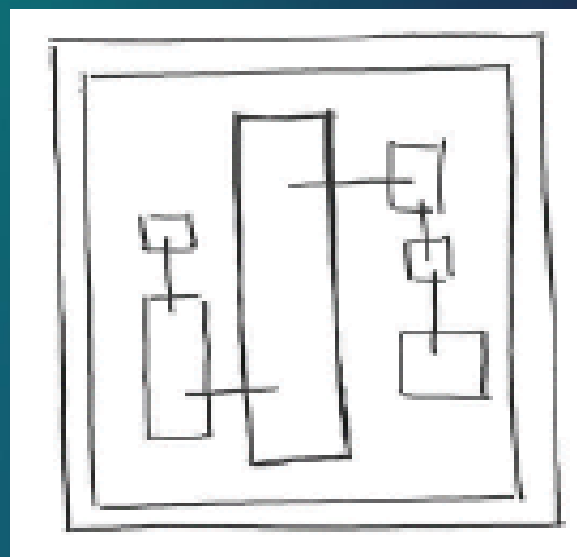
- múltiplas fontes administrativas ou não;
- centralizar/armazenar um maior volume de dados, estruturados e não estruturados;
- um único ponto de entrada de dados (validação, limpeza, transformação, qualidade, análise de coerência, enriquecimento com novas variáveis, ...);
- garantia de relações de integridade entre conjuntos de dados;
- número de entregáveis variado: conjuntos de dados, documentação, ...
- um único ponto de acesso para aceder e explorar dados;
- usabilidade mais simples e uma exploração eficiente;
- elevado grau de otimização de recursos na exploração da informação;
- facilitar a análise de dados pelos responsáveis do estudo de fenómenos estatísticos e criação de novos produtos;
- uma visão simplificada, unificada, integrada e documentada...
- ...

ENCONTROS
SAMA₂₀₂₀

IND :: Parcerias



ENCONTROS SAMA 2020



Agência para a Energia

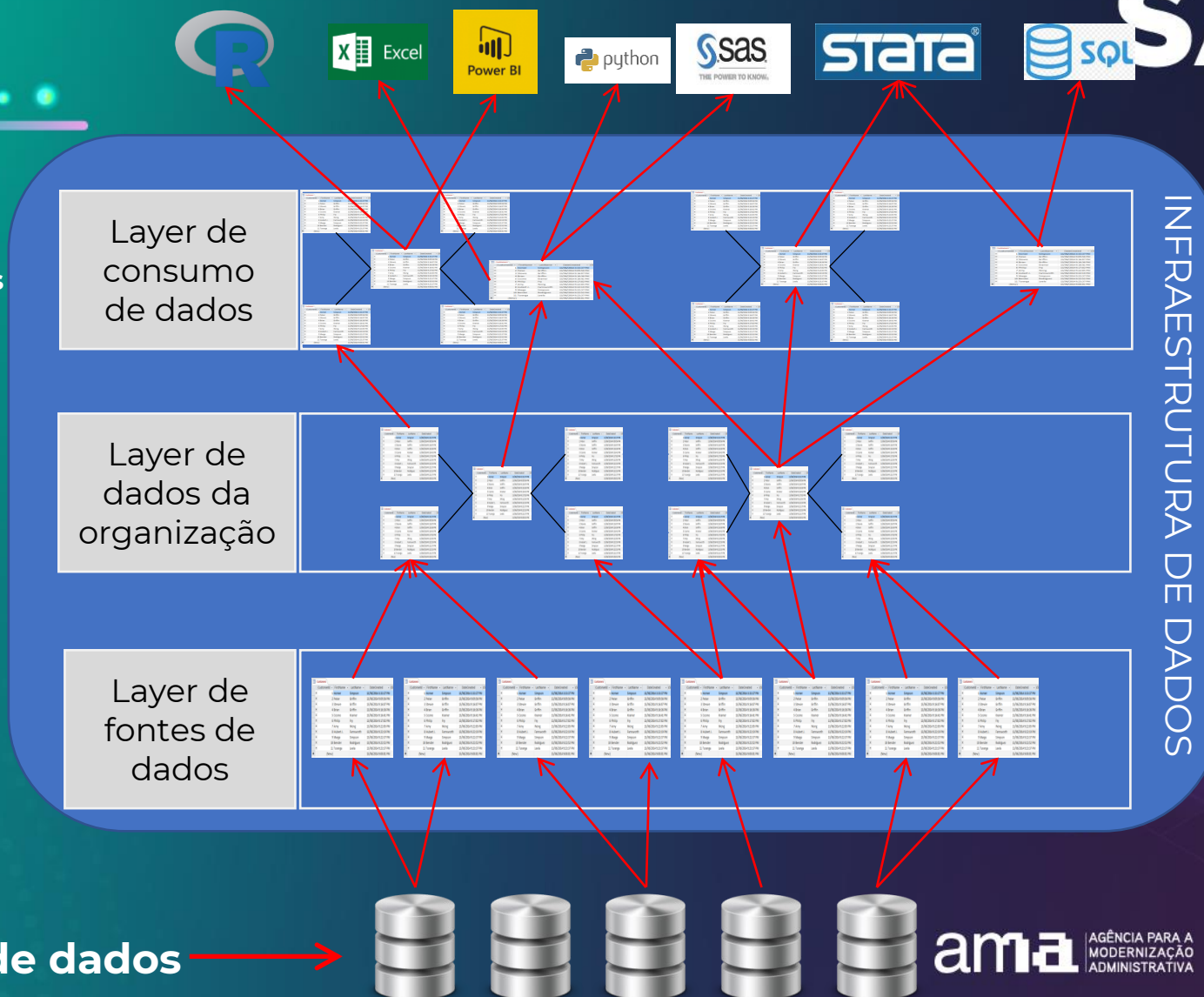
IND :: Overview de processos de tratamento

Uma cultura de Self Service de dados; Fácil de usar – utilização de ferramentas familiares

Um único ponto de acesso para aceder e explorar dados

Modelação e integração de dados; redução de cópia de dados; qualidade de dados; enriquecimento de dados

Um único ponto de entrada de dados; validação; segurança; confidencialidade; proteção



ENCONTROS SAMA 2020



Fontes de dados



ama AGÊNCIA PARA A MODERNIZAÇÃO ADMINISTRATIVA

COMPETE 2020

PORTUGAL 2020

UNIÃO EUROPEIA Fundo Europeu de Desenvolvimento Regional

IND :: Overview de processos de tratamento

‘Confortável’



ENCONTROS SAMA 2020

IND :: Vantagens

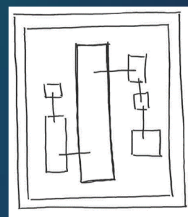
ENCONTROS SAMA²⁰²⁰

Disponibilizar um **serviço** de recolha, armazenamento, tratamento e análise de dados

Capacitação e mais **competências** dos seus profissionais

Padrões de **qualidade internacionais**

Adoção de nomenclaturas e **uniformização dos conceitos**



Identificação facilitada de **áreas de redundância** na recolha de informação, **reduzindo custos** ao erário público e **carga sobre os fornecedores** de informação

Economias de escala nos recursos humanos, desenho das bases de dados, plataformas e hardware de suporte

Criação de **massa crítica** para potenciar o processo de **inovação** no INE

Simplificar e **racionalizar procedimentos**

IND :: Desenvolvimentos em curso

ENCONTROS
SAMA 2020

- Melhorar o acesso a dados privados;
- Aumentar o número de dados transmitidos ao INE para produzir mais estatísticas;
- Documentar mais e melhor os dados de origem administrativa (e outros), pela definição formal de um Documento Metodológico e introdução de melhorias no catálogo de dados;
- Produzir dashboards com informações relevantes para os nossos parceiros (Retorno) ;

IND :: Alguns casos de uso



.pt Domains



eFatura



StatsLab
Estatísticas em Desenvolvimento



IRS para a
Comunidade
Científica

IND :: StatLab Estatísticas em Desenvolvidos

ENCONTROS
SAMA 2020



- [Censos com dados Administrativos \(Versão Html\)](#)
- [Remuneração bruta mensal média por trabalhador – cálculos do INE com base na informação da Segurança Social e da Caixa Geral de Aposentações](#)
- [Indicadores de mobilidade da população ao nível regional - uma leitura a partir da informação da iniciativa "Data for Good" da Meta](#)
- [Estatísticas do Rendimento ao nível local – indicadores de rendimento declarado no IRS](#)
- [A atividade económica regional no contexto da pandemia COVID-19 E-fatura 2019 – 2020](#)
- [Infraestrutura Nacional de Dados: Utilização de Dados Administrativos - Janeiro de 2016 a dezembro de 2019 - O nascimento do primeiro filho: O que muda no padrão de despesas das famílias?](#)
- [Remuneração bruta mensal média por trabalhador: a distribuição das remunerações brutas por trabalhador no setor das Administrações Públicas e no Setor Privado 2021](#)



IND :: eFatura

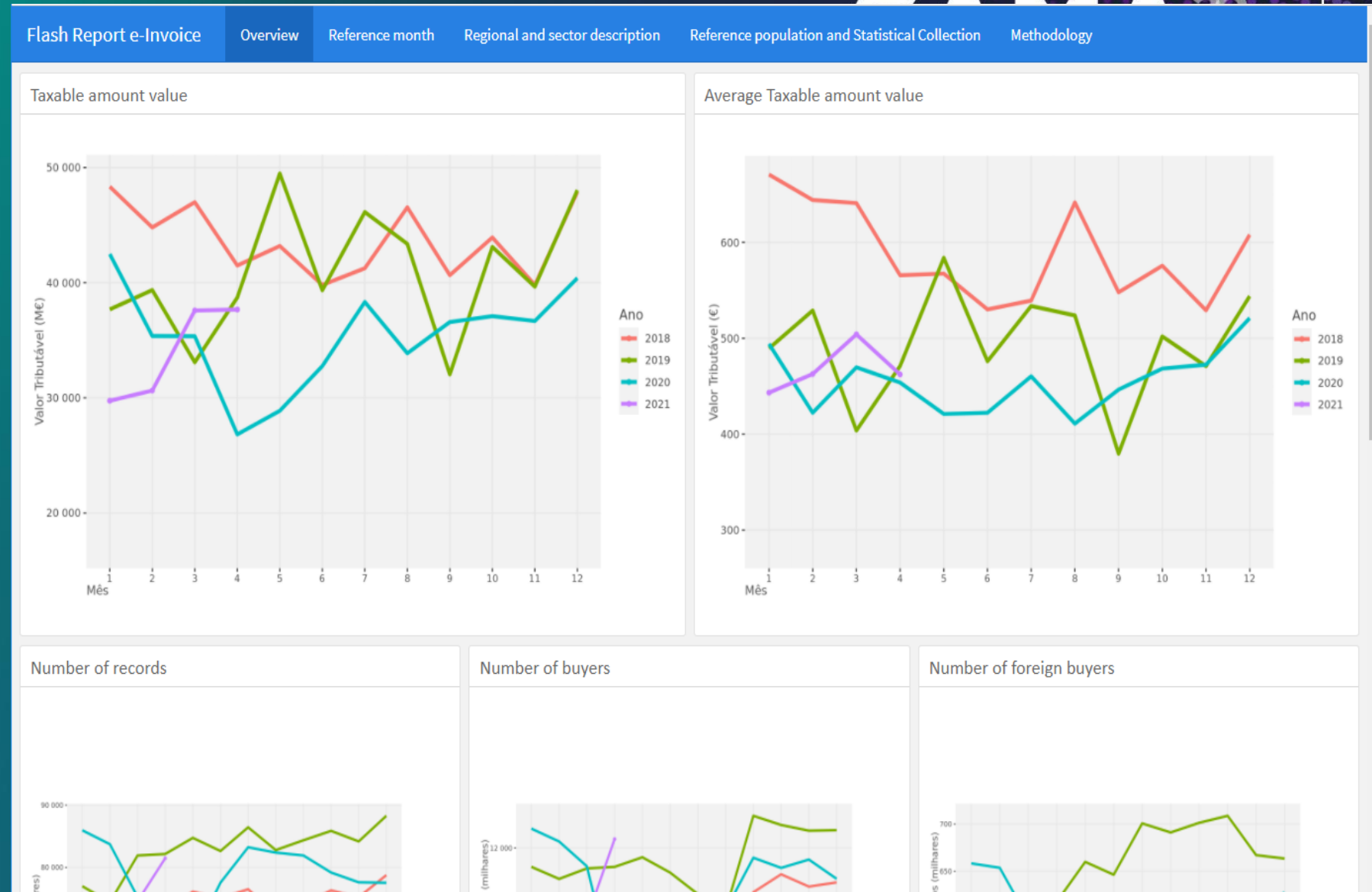
ENCONTROS
SAMA 2020

- **80 milhões** de registos por mês, desde de 2020
- Um conjunto de procedimentos foram implementados, com elevados ganhos no tratamento de dados brutos
- Os dados originais são tratados e **enriquecidos** com:
 - Inclusão de novos atributos provenientes de outras fontes
 - Um referência temporal para as variáveis de caracterização
 - Identificação e imputação de valores anormais
 - Criação de novas tabelas agregadas
 - Implementação de planos alternativos
- Definidos **níveis de serviço** para os utilizadores finais

IND :: eFatura

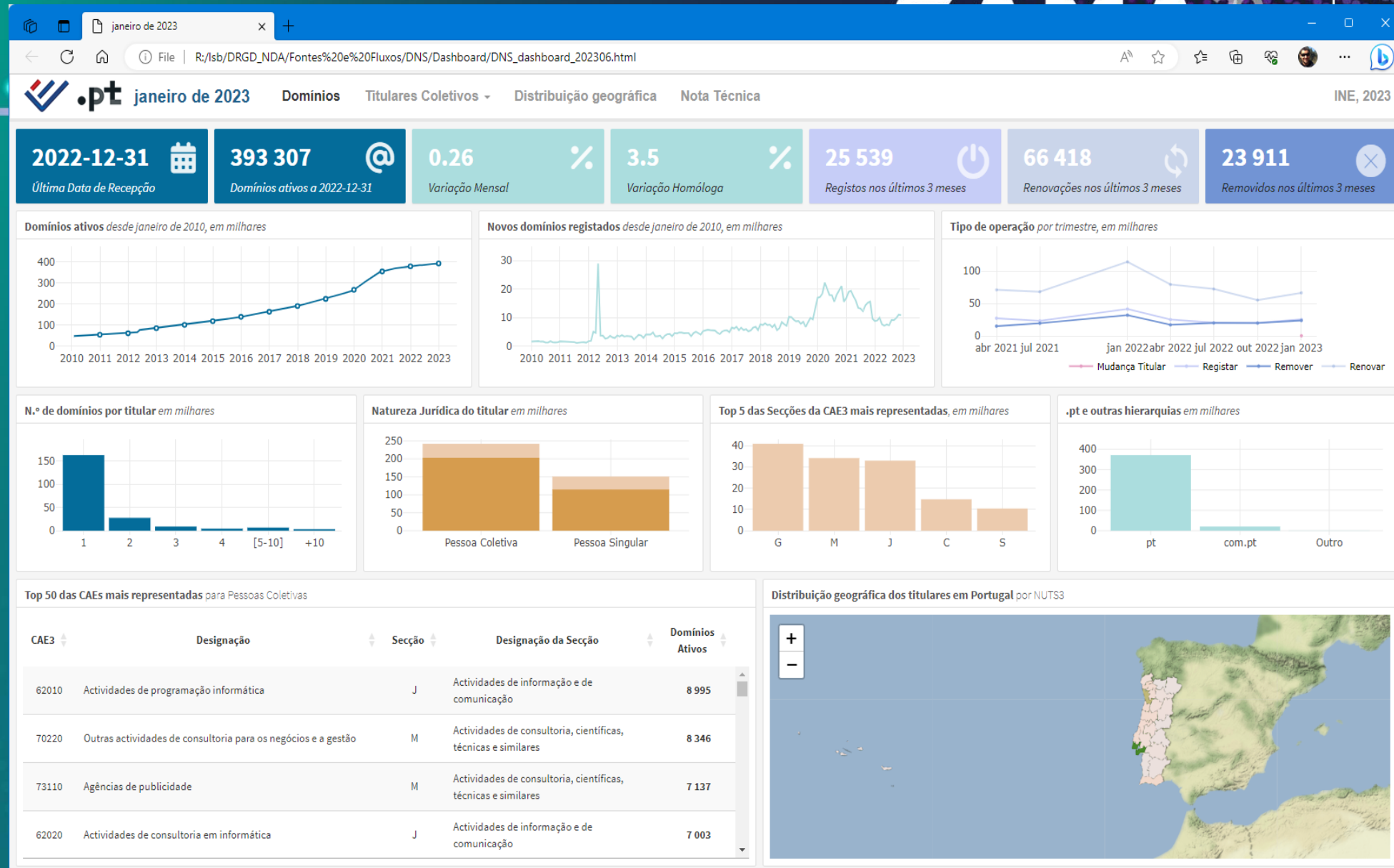
- Um exemplo de um Flash Report criado mensalmente
- Os dados e FR são disponibilizados em 30 horas após a transmissão

DASHBOARD



IND :: .pt domains

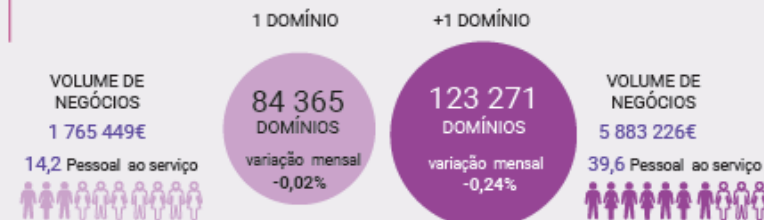
Exemplo de um DASHBOARD



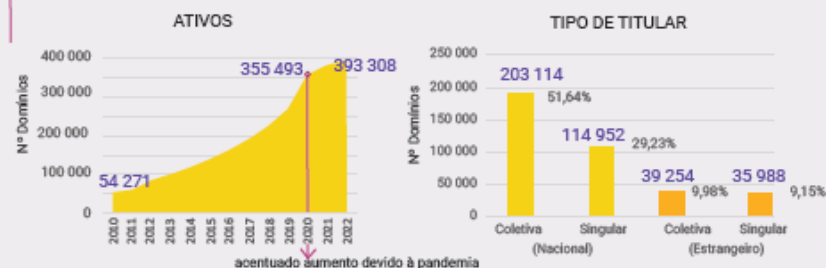
DOMÍNIOS.PT EM JANEIRO 2023



TITULARES NACIONAIS COLETIVOS

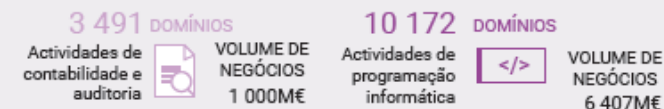


DOMÍNIOS ATIVOS E TIPO DE TITULAR



Titulares, por dimensão da empresa

Dimensão	Nº Domínios	%
Micro	18 784	82,74%
Pequena	2 464	10,85%
Média	1 045	4,6%
Grande	409	1,8%



ATIVIDADE ECONOMICA MAIS REPRESENTADA

DOMÍNIOS, POR PAÍSES DOS TITULARES

TOP 10



DOMÍNIOS POR SECÇÃO DA CAE

TOP 5



Exemplo de uma ficha de caracterização (metainformação)

- 1 Objectivo
- 2 Informações Principais
 - 2.1 Fonte, Versão, Unidade de Registo e Chave de Ligação
 - 2.2 Dicionário de Variáveis
 - 2.3 Variáveis em Falta
- 3 Agregados Familiares Fiscais

2 Informações Principais

2.1 Fonte, Versão, Unidade de Registo e Chave de Ligação

Impresso IRS Mod. 3 2021	Portaria 8/2021, de 7 de janeiro
Tabela	IRS_2021_ROSTO
Versão	Setembro N+1
Unidade de registo das variáveis	Declaração (HID) ou Individuo (NIF_TITULAR)
Chave de ligação com Anexos e NL	HID ou HID NIF_TITULAR

2.2 Dicionário de Variáveis

2.2.1 Descrição das Variáveis

Variavel	Tipo	Unidade	Descricao	N	Media
ANO	Continua	HID	R_Q1_02: ano dos rendimentos	9.976.976	–
VERSAO	String	HID	R_VARDERIV: variável derivada - versão dos dados recebidos	9.976.976	–
HID	String	HID	R_VARDERIV: variável derivada - identificador da declaração (coincide com o NIF da posição A)	9.976.976	–

Infraestrutura Nacional de Dados :: IND

ENCONTROS
SAMA 2020

Obrigado.

jorge.magalhaes@ine.pt

Lisboa, 2 de outubro de 2023



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

ama AGÊNCIA PARA A
MODERNIZAÇÃO
ADMINISTRATIVA

COMPETE
2020

PORTUGAL
2020

UNIÃO EUROPEIA
Fundo Europeu
de Desenvolvimento Regional