

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE CIÊNCIAS MATEMÁTICAS



Desenvolvimento de uma solução para dados abertos no portal do INE- Portugal

Rita Gonçalves Rodrigues

Mestrado em Matemática Aplicada à Economia e Gestão

Trabalho de Projeto orientado por:
Professor Tiago Miguel Dias Domingues

Agradecimentos

Em primeiro lugar, gostaria de expressar a minha profunda gratidão aos meus pais e ao meu irmão, pelo apoio constante, pela paciência inabalável e pela confiança que sempre depositaram em mim.

Agradeço igualmente aos meus dois orientadores, pelo papel fundamental que desempenharam neste projeto. Ao Professor Tiago Miguel Dias Domingues, da Faculdade de Ciências da Universidade de Lisboa, agradeço a orientação académica rigorosa, os contributos críticos e a disponibilidade constante. Ao Eng.º Carlos Dias, do Instituto Nacional de Estatística, deixo um agradecimento muito especial, pela generosidade com que partilhou o seu conhecimento, pela atenção aos detalhes e por todo o apoio técnico e humano que me prestou ao longo do estágio. O seu acompanhamento próximo foi verdadeiramente determinante para a concretização deste trabalho.

Deixo também uma palavra de agradecimento à Professora Teresa Alpuim, coordenadora do mestrado, cujo empenho e dedicação foram determinantes para a celebração do protocolo com o Instituto Nacional de Estatística, sem o qual este estágio não se teria concretizado.

Gostaria também de agradecer ao meu namorado, Micael Gonçalves, por toda a paciência, compreensão e apoio ao longo do meu percurso académico. A sua presença constante foi essencial, mesmo nos momentos mais exigentes.

A todos os que, de alguma forma, contribuíram para esta etapa, seja com apoio, palavras amigas ou simplesmente com a sua presença, o meu sincero obrigado.

Concluir esta tese representa, para mim, mais do que o fim de um percurso académico: é também a celebração de um caminho feito com pessoas que deixaram uma marca.

Resumo

O presente relatório descreve o trabalho desenvolvido no âmbito do estágio curricular realizado no Instituto Nacional de Estatística (INE), no contexto do Mestrado em Matemática Aplicada à Economia e à Gestão. O principal objetivo consistiu em propor uma solução para a criação de uma área de dados abertos no portal do INE (<https://dados.ine.pt>), alinhada com a Diretiva (UE) 2019/1024 e com as boas práticas europeias de Governo Aberto.

Para atingir esse objetivo, foram definidas três etapas fundamentais:

- i. análise do ponto de situação nacional e internacional sobre a disponibilização de dados abertos e o envolvimento dos Institutos Nacionais de Estatística (INEs);
- ii. estudo comparativo das arquiteturas, plataformas e modelos de metadados mais utilizados na União Europeia, com ênfase em países com elevada maturidade digital como França, Irlanda e Espanha;
- iii. proposta de uma solução técnica e funcional para o INE-Portugal, incluindo a definição da arquitetura do sistema, catálogo de dados, modelo de metadados, processo de carregamento e difusão dos dados da plataforma.

A metodologia aplicada envolveu revisão documental, análise comparativa de portais europeus e avaliação de conformidade com o modelo de metadados DCAT-AP, em especial no que se refere à integração com os conjuntos de dados de elevado valor (*High-Value Datasets* – HVDs). A proposta final recomenda a adoção de uma arquitetura baseada na plataforma CKAN, integrada com o atual Banco de Dados de Difusão (BDD) do INE, permitindo a disponibilização automatizada, gratuita e em formatos abertos dos dados estatísticos.

Conclui-se que a solução proposta permitirá reforçar a transparência institucional, promover a reutilização dos dados públicos por diferentes setores da sociedade e contribuir para o posicionamento do INE como uma entidade de referência na transformação digital e na promoção de uma cultura de dados abertos em Portugal.

Palavras-chaves: Dados abertos; Interoperabilidade; Reutilização da informação; Transparência.

Abstract

This report describes the work developed during the curricular internship at Statistics Portugal (INE), as part of the Master's degree in Applied Mathematics to Economics and Management. The main goal was to propose a solution for creating an open data section on the INE portal (<https://dados.ine.pt>), aligned with Directive (EU) 2019/1024 and European best practices on Open Government.

To achieve this goal, the work was structured into three main stages:

- i. analysis of the national and international status of open data initiatives and the involvement of National Statistical Institutes (NSIs);
- ii. comparative study of the most widely used architectures, platforms, and metadata models across the European Union, with a focus on countries with high digital maturity such as France, Ireland, and Spain;
- iii. proposal of a technical and functional solution for INE-Portugal, including the definition of system architecture, data catalog, metadata model, data upload process, and data dissemination strategy.

The methodology included document review, comparative analysis of European open data portals, and evaluation of compliance with the DCAT-AP metadata standard, especially regarding integration with High-Value Datasets (HVDs). The final proposal recommends adopting architecture based on the CKAN platform, integrated with INE's current Data Dissemination Database (BDD), allowing for automated, free, and open-format access to statistical data.

The proposed solution is expected to strengthen institutional transparency, promote the reuse of public data across different sectors of society, and support INE's position as a reference entity in digital transformation and the promotion of open data culture in Portugal.

Keywords: Open data; Interoperability; Data Reuse; Transparency.

Índice

Agradecimentos.....	i
Resumo.....	ii
Abstract.....	iii
Lista de Abreviaturas.....	vi
1. Introdução.....	1
2. Dados abertos.....	2
2.1. Enquadramento.....	2
2.2. Estudo da Maturidade de dados abertos.....	3
2.2.1. Resultados Gerais da Maturidade dos dados abertos de 2024.....	4
2.2.2. Análise Comparativa de Casos Relevantes.....	7
2.3. Dados abertos – Portugal.....	11
2.4. Conjuntos de Dados de Elevado Valor (High-Value Datasets - HVDs).....	13
3. Arquitetura/Plataformas.....	15
3.1. Catálogo de dados (Modelos).....	16
3.2. Plataformas.....	18
3.3. Modelo de Metadados.....	24
4. Proposta de Solução para o INE.....	28
4.1. Arquitetura.....	28
4.2. Catálogo de Dados.....	35
4.3. Plataforma/Software.....	36
4.4. Modelo de Metadados.....	37
4.5. Processo de Carregamento/ Disponibilização dos Dados.....	38
4.6. Difusão dos Dados.....	40
5. Conclusão.....	45
6. Bibliografia.....	46

Índice de Figuras

Figura 2.1.: Esquema do funcionamento do portal data.europa.eu.....	3
Figura 2.2.: Pontuações gerais de maturidade de 2024.....	6
Figura 2.3.: Países classificados por categoria.....	7
Figura 2.4.: Portal do INE Portugal.....	13
Figura 3.1.: Modelo 1 - Plataforma Única.....	16
Figura 3.2.: Modelo 2 - Servidores Separados.....	17
Figura 3.3.: Modelo 3 - Catálogos Federados.....	17
Figura 4.1.: Base de dados do portal do INE.....	28

Figura 4.2.: Portal de França	29
Figura 4.3.: Portal de Irlanda	30
Figura 4.4.: Portal de Espanha	31
Figura 4.5.: Diagrama de arquitetura	33
Figura 4.6.: Arquitetura do sistema de difusão de dados do INE (Documento interno – INE)	39
Figura 4.7.: Filtros de pesquisa do portal de Espanha	41
Figura 4.8.: Indicadores do portal do Pordata	42
Figura 4.9.: Secção dos HVDs no portal do INE.....	43

Índice de Tabelas

Tabela 2.1.: Análise do estudo por Dimensão (2024)	5
Tabela 3.1.: Comparação dos portais de dados abertos, plataformas tecnológicas e grau de conformidade com o modelo de metadados DCAT-AP	26

Lista de Abreviaturas

UE	União Europeia
INE	Instituto Nacional de Estatística
BDD	Banco de Dados de Difusão
API	<i>Application Programming Interface</i>
REST	<i>Representational State Transfer</i>
HVD	<i>High-Value Datasets</i>
IoT	<i>The internet of things</i>
TI	Tecnologias de Informação
CKAN	<i>Comprehensive Knowledge Archive Network</i>
SaaS	<i>Software as a Service</i>
DCAT	<i>Data Catalog Vocabulary</i>
DCAT-AP	<i>DCAT Application Profile for data portals in Europe</i>

1. Introdução

A crescente digitalização das sociedades contemporâneas tem vindo a transformar profundamente a forma como os dados são produzidos, disseminados e utilizados. No seio desta evolução, os dados abertos destacam-se como um instrumento fundamental para reforçar a transparência institucional, incentivar a participação cívica e promover a inovação social e económica. Esta abordagem relativa à gestão da informação pública tem ganho particular relevo em contextos governativos que privilegiam políticas de acesso livre e reutilização da informação do setor público.

A União Europeia tem impulsionado esta transformação através de várias iniciativas estratégicas, sendo a [Diretiva \(UE\) 2019/1024](#) um marco regulatório que define orientações claras para os Estados-Membros na adoção de práticas de dados abertos, com especial ênfase nos Conjuntos de Dados de Elevado Valor (*High-Value Datasets* – HVDs). Portugal tem acompanhado este movimento, com progressos evidentes na maturidade das suas políticas de dados abertos, nomeadamente através da plataforma [dados.gov.pt](#). Contudo, ainda existem desafios significativos em termos de integração, normalização e abrangência na disponibilização de informação estatística em formatos abertos.

Neste panorama, o Instituto Nacional de Estatística (INE) assume um papel central e estratégico no ecossistema nacional de dados. Enquanto organismo responsável pela produção e difusão oficial de estatísticas em Portugal, o INE não só é um fornecedor de informação essencial à definição de políticas públicas, como também um agente promotor da literacia estatística, da investigação científica e do desenvolvimento económico. A disponibilização dos seus dados em formatos abertos e reutilizáveis representa, por isso, uma oportunidade para ampliar o impacto social e económico da informação estatística, potenciar o desenvolvimento de soluções inovadoras e reforçar o valor público dos dados produzidos com recursos do Estado.

É neste enquadramento que se insere o presente trabalho, realizado no âmbito do estágio curricular do Mestrado em Matemática Aplicada à Economia e à Gestão, e que se debruça sobre a conceção de uma solução para dados abertos no Portal do INE. A proposta é sustentada por uma análise crítica das práticas internacionais mais avançadas, bem como por um levantamento das exigências técnicas e funcionais necessárias à operacionalização de uma plataforma moderna, alinhada com as diretrizes da União Europeia e com os princípios do Governo Aberto.

2. Dados abertos

2.1. Enquadramento

Os dados abertos são definidos, de acordo com a *Open Definition*, como "dados que qualquer pessoa pode aceder, utilizar, modificar e partilhar, para qualquer propósito". Esta definição sublinha a liberdade de acesso e reutilização como elementos essenciais para que a informação pública possa ser plenamente explorada por todos os setores da sociedade.

O movimento dos dados abertos, que é parte integrante das políticas dedicadas ao Governo Aberto (*Open Government*), combina os princípios da transparência, participação e colaboração entre o Estado e os cidadãos, assim como o potencial de desenvolvimento económico que o digital trouxe.

A grande maioria desses dados já são, por lei, considerados públicos. No entanto, o principal desafio reside em tornar esses dados acessíveis, compreensíveis e reutilizáveis. Isto implica a disponibilização dos dados em formatos estruturados, legíveis por máquinas e suportados por ferramentas abertas, para que possam ser reutilizados, transformados ou integrados por qualquer cidadão ou entidade.

A relevância dos dados abertos manifesta-se na diversidade dos seus beneficiários. Os cidadãos passam a ter um acesso mais direto e eficiente à informação pública, o que reforça a transparência e a responsabilização do Estado. As instituições governamentais, por sua vez beneficiam de uma maior eficiência administrativa e de uma melhor articulação entre organismos públicos. O setor empresarial pode utilizar estes dados como matéria-prima para o desenvolvimento de novas aplicações, plataforma ou serviços com elevado potencial comercial. Adicionalmente, setores como o jornalismo de dados, a investigação universitária ou mesmo organizações não-governamentais com preocupações cívicas.

Importa distinguir entre dados abertos e dados meramente disponibilizados ao público. Apenas os primeiros estão sujeitos a licenças abertas que permitem a sua reutilização sem restrições significativas. Dados disponibilizados sem tais licenças ou sob condições restritas não podem ser considerados verdadeiramente abertos.

A nível europeu, a União Europeia tem vindo a promover a política de dados abertos como um pilar do governo aberto, alinhada à [Diretiva de dados abertos \(2019/1024\)](#). Especificamente, a diretiva de dados abertos incentiva os Estados-Membros da UE a tomarem o máximo de informações publicamente acessíveis disponíveis para reutilização, com o objetivo de facilitar a sua reutilização por parte de cidadãos, empresas, instituições e outras entidades.

As políticas de livre acesso aos dados visam promover uma ampla disponibilização e reutilização da informação gerada pelo setor público, seja para fins privados, comerciais ou sociais. Estas políticas procuram eliminar, ou pelo menos reduzir significativamente, as restrições legais, técnicas ou financeiras que possam limitar o acesso e a reutilização da informação. Ao favorecer a circulação livre e aberta desses dados, não só se criam oportunidades para os agentes económicos desenvolverem novos produtos e serviços, como também se reforça o envolvimento dos cidadãos na vida pública e se estimula a inovação social.

Assim, os dados abertos representam não apenas uma questão de eficiência administrativa ou inovação tecnológica, mas também um instrumento crucial para a democratização da informação e o fortalecimento das instituições públicas no contexto da sociedade digital.

2.2. Estudo da Maturidade de dados abertos

O estudo sobre a maturidade dos dados abertos constitui uma referência essencial para avaliar o nível de desenvolvimento alcançado na área dos dados abertos na Europa.

Neste contexto, a iniciativa *European Data Portal*, promovida pela Comissão Europeia (<https://data.europa.eu/>), assume-se como o ponto central de acesso aos dados abertos europeus provenientes de portais internacionais, da União Europeia, bem como de plataformas nacionais, regionais, locais e de dados geográficos.

Como se pode ver pelo esquema da Figura 2.1., este portal tem como principais objetivos:

- Facilitar o acesso e promover a reutilização dos dados abertos europeus entre cidadãos, empresas e organizações;
- Incentivar e apoiar a divulgação de mais metadados e dados de melhor qualidade, tanto pelas instituições, agências e outros organismos da UE como pelos países europeus;
- Sensibilizar e capacitar cidadãos e entidades para as oportunidades associadas à disponibilidade e utilização de dados abertos.

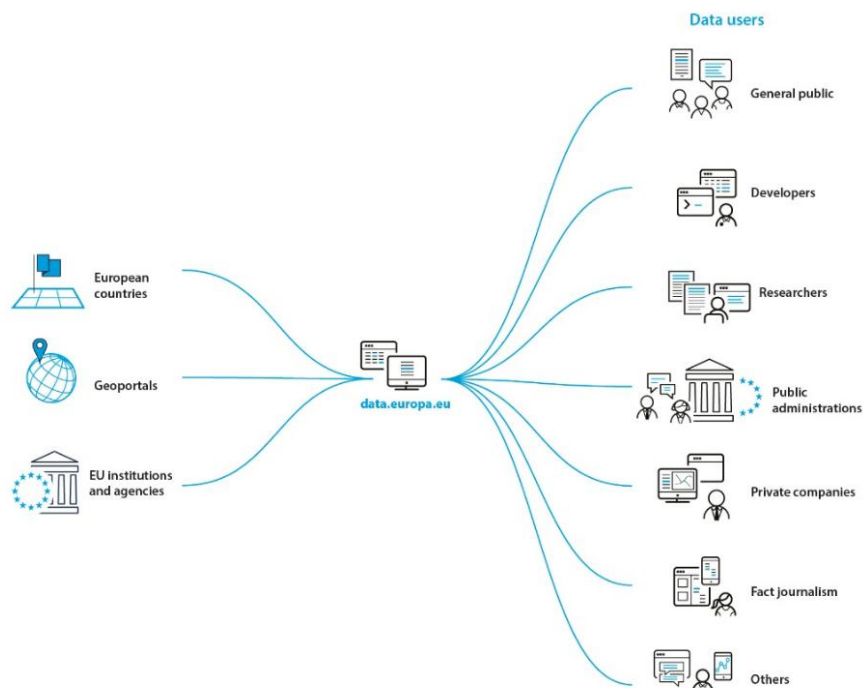


Figura 2.1.: Esquema do funcionamento do portal *data.europa.eu*

A Estratégia Europeia de Dados, presente na Figura 2.1., visa promover um mercado único de dados, com benefícios claros para empresas, investigadores e administrações públicas. Neste contexto, o portal europeu de dados abertos continuará a desempenhar um papel central na promoção da acessibilidade e da interoperabilidade dos dados, facilitando a sua reutilização e potenciando a criação de valor económico e social.

A implementação de portais de dados abertos nos vários países europeus tem evoluído a ritmos distintos, refletindo diferentes níveis de maturidade e compromisso com a agenda dos dados abertos. Com o

objetivo de monitorizar este progresso, a Comissão Europeia, através da iniciativa *European Data Portal*, contrata anualmente à empresa CapGemini a realização de um estudo que avalia o progresso dos países europeus na disponibilização e reutilização dos dados abertos do setor público.

A metodologia deste estudo assenta em quatro dimensões fundamentais:

Política

Esta dimensão examina as estratégias e políticas nacionais de dados abertos, incluindo os modelos de governança adotados e os mecanismos implementados para garantir a sua efetiva execução. São considerados critérios como a existência de legislação específica, incentivos à publicação de dados e estruturas de coordenação entre os diferentes níveis da administração pública.

Portal

A dimensão portal analisa a funcionalidade e a eficácia dos portais nacionais de dados abertos, determinando a facilidade de acesso, navegação e reutilização dos dados. São avaliados aspetos como a usabilidade, a cobertura temática dos dados, a adoção de tecnologias que facilitem a pesquisa e extração da informação, bem como as estratégias de sustentabilidade a longo prazo.

Qualidade

Nesta dimensão é avaliada a qualidade dos dados disponibilizados, com especial atenção na precisão, atualidade e integridade dos dados e metadados. Considera-se também a conformidade com normas e padrões internacionais, como o **DCAT-AP**, e a disponibilização dos dados em formatos legíveis por máquinas e tecnicamente reutilizáveis.

Impacto

Por fim, a dimensão impacto procura medir os efeitos reais da disponibilização dos dados abertos. Esta análise incide sobre a sua reutilização por diferentes setores, a documentação de casos de sucesso e os métodos utilizados para quantificar os impactos económicos, sociais e ambientais da abertura dos dados.

A abordagem baseada nessas quatro dimensões permite uma análise detalhada do grau de maturidade dos dados abertos em cada país. Este diagnóstico é essencial para identificar boas práticas, destacar áreas de melhoria e apoiar a formulação de políticas mais eficazes e sustentadas, promovendo uma cultura de transparência, inovação e valorização dos dados abertos enquanto recurso estratégico.

2.2.1. Resultados Gerais da Maturidade dos dados abertos de 2024

A maturidade dos dados abertos na Europa continua a evoluir positivamente. Em 2024, a pontuação média dos 34 países analisados atingiu os 80%, representando um aumento de 1 ponto percentual (pp) em relação ao ano anterior (79%). No caso específico dos Estados-Membros da União Europeia, a média manteve-se estável nos 83%.

A Eslováquia subiu 4 pp, alcançando o quarto lugar (96%). A República Checa entrou no top 10 pela primeira vez, ocupando a oitava posição (94%). Em termos globais, 18 países registaram melhorias, enquanto 15 apresentaram uma descida.

Relativamente aos Estados-Membros, como mencionado anteriormente, mantiveram-se nos 83% devido a uma combinação de avanços em algumas áreas e recuos noutras. Verificaram-se progressos nas dimensão de impacto (+4 pp) e política (+2 pp), enquanto se registaram quedas na dimensão de portal (-3 pp) e qualidade (-3 pp). Letónia (+10 pp), Croácia (+9 pp) e República Checa (+6 pp) destacaram-se pelo crescimento, fruto de investimentos em infraestruturas digitais e em iniciativas que promovem o uso de dados abertos por empresas e cidadãos. Em contraste, Bulgária (-13 pp), Alemanha (-10 pp) e Finlândia (-5 pp) apresentaram as maiores quedas, associados a dificuldades na implementação de novas regulamentações, na garantia da qualidade dos dados e na participação ativa de *stakeholders* na reutilização da informação.

Tabela 2.1.: Análise do estudo por Dimensão (2024)

Dimensão	Resultado médio UE	Variação
<i>Política</i>	91%	+2 pp
<i>Portal</i>	86%	-3 pp
<i>Qualidade</i>	79%	-3 pp
<i>Impacto</i>	80%	+4 pp

Através do estudo da maturidade dos dados, realizado em 2024, disponível na plataforma data.europa.eu, observa-se na Tabela 2.1. que a análise por dimensão permite verificar o seguinte:

- **Política** manteve-se como a dimensão mais madura (91%), evidenciando o fortalecimento das estratégias nacionais;
- **Portal**, apesar de ser uma das dimensões mais maduras, registou uma queda (-3pp), relevando um desfasamento entre a disponibilização dos dados e a evolução tecnológica dos portais;
- **Qualidade**, com 79%, foi a dimensão menos madura, embora com melhorias nos mecanismos automáticos de recolha de metadados e na conformidade com o DCAT-AP;
- **Impacto** ultrapassou pela primeira vez a dimensão da qualidade, atingindo 80%, o que demonstra um crescente esforço dos países em avaliar os benefícios concretos da reutilização dos dados.

Na Figura 2.2. observa-se os resultados do estudo da maturidade dos dados de diversos países, relativo ao ano 2024, realizado pela empresa CapGemini.

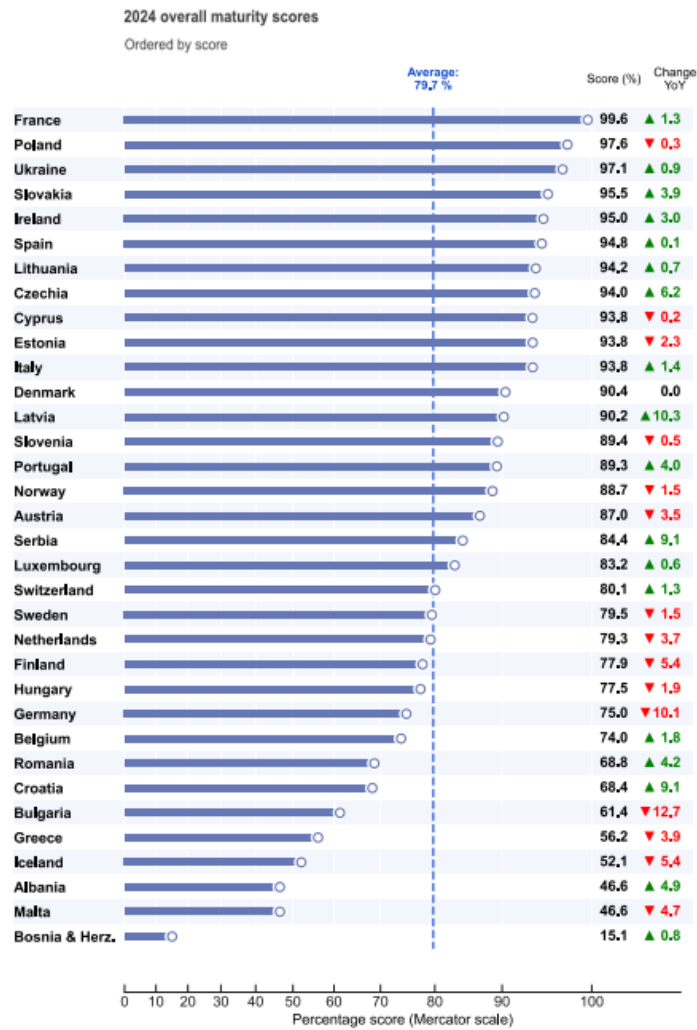


Figura 2.2.: Pontuações gerais de maturidade de 2024

Verifica-se que se destacam a França, Polónia e Ucrânia, como os países com os melhores rankings, respetivamente 1º, 2º e 3º.

De acordo com os resultados do estudo, os países são classificados em 4 categorias:

- **Definidores de tendências** (*Trend-setters*);
- **Localizadores rápidos** (*Fast-trackers*);
- **Seguidores** (*Followers*);
- **Iniciantes** (*Beginners*).

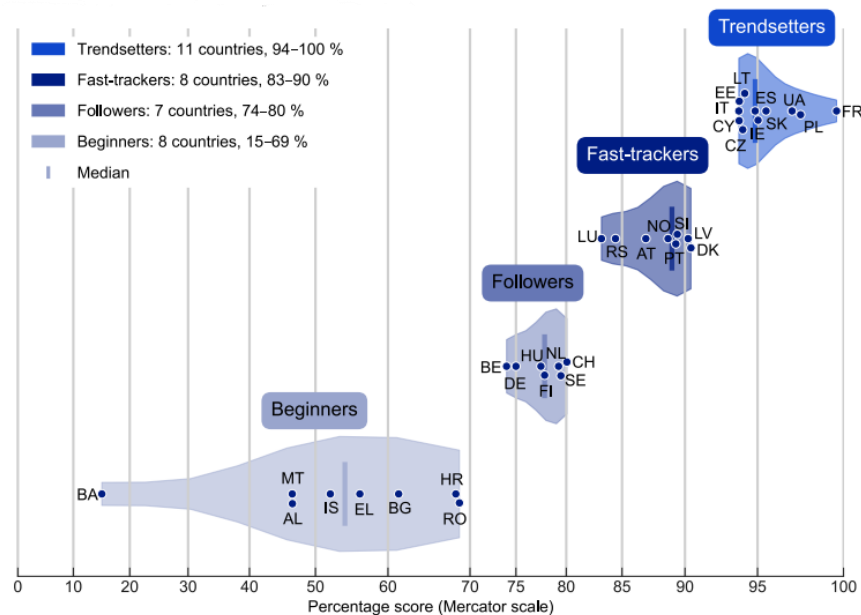


Figura 2.3.: Países classificados por categoria

Conforme identificado na Figura 2.3., retirada do relatório [Open Data Maturity](#), a distribuição dos países de acordo com a classificação:

- **Definidores de Tendências (94-100%):** França, Polónia, Ucrânia*, Eslováquia, Irlanda, Espanha, Lituânia, República Checa, Chipre, Estónia e Itália.
- **Localizadores Rápidos (83-90%):** Dinamarca, Letónia, Eslovénia, Portugal, Noruega*, Áustria, Sérvia* e Luxemburgo.
- **Seguidores (74-80%):** Suíça*, Suécia, Países Baixos, Finlândia, Hungria, Alemanha e Bélgica.
- **Iniciantes (15-69%):** Roménia, Croácia, Bulgária, Grécia, Islândia*, Albânia*, Malta e Bósnia e Herzegovina*.

Nota: Países marcados com um asterisco (*) não fazem parte da UE27

2.2.2. Análise Comparativa de Casos Relevantes

A análise comparativa entre diferentes países permite identificar distintas abordagens no que diz respeito à maturidade dos dados abertos, evidenciando tanto estratégias bem-sucedidas como os principais desafios enfrentados. Este tipo de análise é fundamental para compreender como diferentes contextos políticos, administrativos e tecnológicos influenciam o desenvolvimento e a eficácia das políticas de dados abertos.

Para esta análise, foram selecionados os países melhor posicionados em cada uma das quatro categorias definidas no estudo de maturidade (Definidores de Tendências - **França**, Localizadores Rápidos - **Dinamarca**, Seguidores - **Suíça** e Iniciantes - **Roménia**). Esta escolha metodológica visa proporcionar uma perspectiva abrangente e representativa das diversas realidades existentes no domínio dos dados abertos. Ao centrar a análise nos líderes de cada grupo, torna-se possível identificar com maior clareza

os fatores que influenciam o desempenho de cada dimensão avaliada, bem como destacar boas práticas que poderão servir de referência para países em fases menos avançadas de desenvolvimento nesta área.

França

A França mantém uma das pontuações mais elevadas na maturidade geral dos dados abertos, com um desempenho de 100% em quase todas as dimensões.

Dimensão da Política (100%)

A tradição francesa no acesso à informação remonta à Declaração dos Direitos do Homem e do Cidadão de 1789. O país tem vindo a implementar a diretiva europeia de dados abertos através de legislação robusta, como a [Lei da República Digital](#) e a [Lei sobre a Orientação da Mobilidade](#), que promove, entre outros, o acesso a dados de transporte em tempo real. A sua estratégia visa promover a abertura da informação pública através de ações estruturadas e inovadoras, em colaboração com entidades públicas e privadas.

Dimensão do Impacto (100%)

A França identificou 4 áreas de impacto para dados abertos: científica, económica, democrática e política. São amplamente documentados os casos de reutilização de dados abertos, com destaque para o aumento da eficiência governamental, o combate às alterações climáticas e a redução das desigualdades sociais. Há também um foco significativo na medição do impacto da reutilização de dados de elevado valor.

Dimensão do Portal (98,5%)

O portal destaca-se pela sua forte orientação para o utilizador e pelas funcionalidades tecnológicas avançadas. Qualquer pessoa pode publicar dados, sendo os conjuntos de dados de fontes oficiais identificados com um ícone específico. O portal oferece ferramentas gratuitas para apoiar os fornecedores na melhoria da qualidade dos dados, recolhe *feedback* dos utilizadores e disponibiliza um sistema de pesquisa avançado, com tecnologias que otimizam a descoberta de dados.

Dimensão da Qualidade (100%)

A França adotou uma estratégia de qualidade baseada no modelo das 5 estrelas de Berners-Lee, incluindo aspetos como licenciamento, leitura por máquina e formatos estruturados. O país garante a interoperabilidade dos dados através da conformidade com as diretrizes da UE e valida automaticamente os metadados segundo o padrão DCAT-AP. A equipa promove *workshops*, seminários e guias para apoiar os fornecedores e a plataforma suporta diversos formatos de metadados, como DCAT¹, CKAN e GeoNetwork.

Outros Aspetos:

A promoção dos dados abertos passa também pelo uso de redes sociais, com o objetivo de atingir um público mais jovem. Existem iniciativas específicas para setores-chave e conjuntos de dados direcionados. O plano de ação francês inclui a expansão da publicação de dados abertos, o reforço da qualidade e a interoperabilidade, bem como a priorização de conjuntos de dados de elevado valor (HVDs).

Dinamarca

¹ DCAT (*Data Catalog Vocabulary*) é um vocabulário usado para descrever catálogos de dados de forma padronizada, facilitando a interoperabilidade e a partilha de metadados entre plataformas de dados na web.

A Dinamarca destaca-se como um dos países mais avançados entre os Localizadores Rápidos, com uma pontuação global de 90,4%.

Dimensão da Política (90,6%)

A sua estratégia de dados abertos está integrada num dos pilares da Estratégia Nacional para a Digitalização 2024-2027. O país demonstra um elevado nível de implementação do regulamento relativo a conjuntos de dados de elevado valor (HVDs).

Dimensão do Impacto (96,6%)

A Dinamarca obteve a pontuação máxima num dos indicadores que medem a Dimensão do Impacto, “consciência estratégica”, o que demonstra um compromisso forte com a reutilização e impacto dos dados. Recolhe e classifica casos de reutilização em diversas áreas, incluindo a governação, o ambiente e a economia.

Dimensão do Portal (82,8%)

Este portal é altamente funcional e está constantemente em modernização. Promove a utilização de licenças padronizadas e formatos como CSV, JSON, XML e GeoJSON para facilitar a partilha de dados. Destacam-se também dados em tempo real, como, por exemplo, os relativos a estações de carregamento de veículos elétricos.

Dimensão da Qualidade (91,7%)

A Dinamarca atingiu 100% no indicador de “qualidade de implementação e dados vinculados”, um subconjunto da dimensão da qualidade. Desenvolveu um modelo próprio de avaliação e criou uma linguagem comum para qualidade de dados. Oferece regularmente formação online e *webinars* para garantir metadados e conjuntos de dados de alta qualidade.

Outros Aspetos:

É líder na disponibilização dos indicadores dos HVDs, de acordo com a Diretiva de dados abertos (2019/1024), nomeadamente dados estatísticos, meteorológicos e geoespaciais. Adota um modelo centralizado de governança de dados e incentiva a reutilização por empresas e cidadãos, apoiando o desenvolvimento de serviços inovadores.

Suíça

A Suíça, classificada como seguidora, apresenta um nível de maturidade geral de 80,1%, com uma abordagem própria e descentralizada à gestão dos dados abertos.

Dimensão da Política (89,8%)

A Suíça dispõe de uma política nacional de dados abertos e incentiva a publicação de dados em tempo real. A governança de dados é descentralizada, permitindo que regiões administrativas e municípios tenham autonomia na gestão dos seus dados. O país exige que órgãos públicos mantenham um inventário atualizado dos dados, incluindo aqueles que não podem ser publicados abertamente.

Dimensão do Impacto (75%)

Este país aplica uma metodologia estruturada para avaliar o impacto dos dados abertos, cobrindo aspetos governamentais, sociais, ambientais e económicos. Documenta casos de reutilização nas áreas de inovação tecnológica, empreendedorismo e eficiência administrativa, promovendo ativamente eventos e formações sobre o tema.

Dimensão do Portal (72,4%)

O portal oferece suporte a dados tabulares e geoespaciais com funções de pré-visualização. O seu desenvolvimento é orientado pelo *feedback* dos utilizadores e há um esforço contínuo para alinhar os dados com o modelo DCAT-AP, atingindo mais de 90% de conformidade.

Dimensão da Qualidade (83%)

A Suíça impõe padrões exigentes para os metadados, com licenças claras e aderência a diretrizes como as *Creative Commons*. Realiza atividades regulares de apoio aos publicadores e utiliza sistemas automáticos para verificar a qualidade, como a deteção de *links* quebrados ou falhas de conformidade.

Outros Aspetos:

A estratégia suíça dá ênfase ao envolvimento do setor privado, promovendo o uso de dados em soluções inovadoras. Esta colaboração tem contribuído para reforçar a relevância dos dados abertos enquanto recurso estratégico nacional.

Roménia

A Roménia representa os Iniciantes nesta análise, com uma pontuação global de 68,8%. Apesar dos desafios, tem vindo a implementar medidas estruturadas para avançar na maturidade dos dados abertos.

Dimensão da Política (80,5%)

O país ainda não possui uma estratégia de dados abertos, mas está a trabalhar para aplicar o [Regulamento de Execução da Comissão \(UE\) 2023/138](#) relativo aos HVDs. A coordenação é realizada por agências especializadas, responsáveis por questões mais amplas de governação digital e de dados, através de estruturas interministeriais e interdepartamentais.

Dimensão do Impacto (49,1%)

Apesar da pontuação baixa, a Roménia tem diretrizes metodológicas que obrigam as instituições a monitorizar a reutilização de HVDs e tem investido na compreensão das necessidades dos reutilizadores.

Dimensão do Portal (82,8%)

O portal nacional deixou de permitir a publicação direta por parte dos utilizadores. No entanto, tem sido promovido através de *webinars*, reuniões públicas com instituições de ensino e o setor privado, sessões informativas focadas nos estudantes e apresentações em eventos públicos. Esta abordagem visa aumentar a visibilidade e o envolvimento da comunidade.

Dimensão da Qualidade (62,7%)

A atualização dos metadados apresenta atrasos superiores à média europeia, podendo demorar até um mês. A Roménia utiliza APIs para facilitar a recolha e metadados e integra ferramentas da UE, como a Metodologia de Avaliação da Qualidade dos Metadados (MQA), para avaliar e monitorizar a qualidade dos metadados, permitindo, também, o *feedback* direto dos utilizadores.

Outros Aspetos:

A Roménia adotou as recomendações da *Diretiva INSPIRE*² para garantir a interoperabilidade transfronteiriça, com iniciativas avançadas relacionadas com dados geoespaciais, ambientais e de

² *Diretiva INSPIRE* – Diretiva europeia que visa garantir a interoperabilidade e o acesso harmonizado a dados geoespaciais na UE, especialmente para fins ambientais e de planeamento.

observação da Terra. Implementou também mecanismos de monitorização para recolher e classificar casos de reutilização, apostando numa evolução sustentada.

2.3. Dados abertos – Portugal

À semelhança dos restantes países europeus, também Portugal tem vindo a aderir à iniciativa dos dados abertos. O dados.gov.pt constitui a principal plataforma nacional de publicação e agregação de dados abertos, disponibilizando informação produzida pela Administração Pública em formatos abertos, legíveis e reutilizáveis por qualquer cidadão. O seu principal objetivo é facilitar o acesso à informação pública, promover a transparência da Administração Pública e ajudar na criação de serviços públicos eletrónicos por parte da sociedade civil.

Ao promover a reutilização da informação pública, o dados.gov.pt pretende também estimular a existência de novos serviços destinados à população em geral. O portal é um projeto da responsabilidade da Agência para a Modernização Administrativa (AMA, IP), que conta com a participação de inúmeros organismos da Administração Pública Central, Regional e Local como parceiros. Atualmente, o portal agrega mais de 9.700 conjuntos de dados, provenientes de 159 organizações, abrangendo áreas como População e Sociedade, Economia e Finanças, entre outras.

A disponibilização destes dados tem impulsionado a criação de diversas aplicações e visualizações que beneficiem os cidadãos e demonstrem a transparência governamental. Entre os exemplos mais relevantes destacam-se:

- **Fogo.pt:** Uma plataforma que fornece informação em tempo real sobre incêndios florestais em Portugal, auxiliando na prevenção e resposta a emergências;
- **Ocorrências em Aberto:** Uma aplicação que disponibiliza informações sobre ocorrências ativas, contribuindo para a segurança pública;
- **COVID Insights:** Uma plataforma que reúne e analisa dados sobre a pandemia em Portugal, fornecendo perspetivas nas áreas da saúde pública, economia e impacto social.

No estudo do *European Data Portal*, que monitoriza a evolução dos dados abertos nos Estados-Membros, Portugal, como se observa na Figura 2.2., foi classificado na 14ª posição em 2024 (89,3%), mantendo-se na mesma posição face a 2023, mas com um crescimento de 4 pp. Registou uma mudança significativa ao transitar para a categoria 2 no grupo dos Localizadores Rápidos.

Dimensão da Política (97,7%)

Portugal apresentou, em 2024, a maior melhoria nesta dimensão (+19pp), resultado da implementação recente de uma estratégia de dados abertos, com políticas que incentivam a publicação e o acesso a dados à informação gerada pelos cidadãos. Estas políticas promovem igualmente a descoberta de dados abertos na plataforma data.europa.eu, além de descreverem medidas concretas para apoiar a reutilização por parte do setor privado.

O país promove a reutilização de dados abertos através da organização de eventos públicos, *workshops* e oportunidades de *networking*. Além disso, permite que as autoridades locais utilizem os portais nacionais como plataforma de partilha e dados, dispensando o desenvolvimento de infraestruturas

próprias. É ainda de salientar que Portugal revê periodicamente as suas políticas e estratégias, de modo a acompanhar os avanços tecnológicos.

Dimensão do Portal (73,9%)

O dados.gov.pt é uma plataforma fundamental para promover a transparência e o acesso à informação pública. Através deste portal, cidadãos e organizações podem aceder a um vasto conjunto de dados, incentivando a sua reutilização. Os utilizadores podem subscrever notificações sobre atividades específicas de conjuntos de dados, através da funcionalidade de marcação por estrela³, mantendo-se atualizados sobre novas publicações.

O portal também oferece apoio técnico e formativo às entidades públicas para facilitar a publicação de dados, disponibilizando diretrizes, tutoriais *online* e sessões de esclarecimento. Esse suporte estimula uma maior participação institucional e fortalece a cultura de dados abertos no país.

Dimensão da Qualidade (85,7%)

Portugal adota uma abordagem proativa para garantir a qualidade dos metadados. O sistema dispõe de uma seção dedicada à "qualidade" na área de administração, que oferece uma visão geral da estrutura dos metadados e sugestões para a sua melhoria, como, por exemplo, descrições mais detalhadas, inclusão de etiquetas (*tags*) adicionais e utilização de formatos abertos e legíveis por máquina.

Além disso, Portugal utiliza a plataforma Udata, compatível com o modelo DCAT-AP, e possui ferramentas de mapeamento para outros *frameworks*⁴, garantindo a publicação de dados reutilizáveis e de elevada qualidade, beneficiando o ecossistema de dados abertos.

Dimensão do Impacto (100%)

Em Portugal, os organismos públicos têm desenvolvido métodos sistemáticos para classificar casos de reutilização de dados abertos, demonstrando um compromisso com a transparência e a inovação.

Recentemente, Portugal passou a disponibilizar dados sobre o impacto da utilização em desafios ambientais, sublinhando o papel dos dados abertos na formulação de soluções sustentáveis. Esta iniciativa não apenas fortalece a colaboração entre diferentes setores, mas também reforça a capacidade de utilizar informações de maneira eficaz para abordar os desafios que o país enfrenta neste campo.

INE – Portugal

O Instituto Nacional de Estatística de Portugal (INE-PT) é a entidade oficial responsável pela produção e divulgação de dados estatísticos essenciais para a compreensão da realidade económica, social e demográfica do país. O portal do INE-PT apresenta, em junho de 2025, um catálogo de dados com 12.615 indicadores estatísticos, organizados por diferentes temas. Estes indicadores encontram-se armazenados no Banco de Dados de Difusão (BDD) e são atualizados regularmente através de mecanismos automáticos. A consulta aos dados pode ser feita diretamente através do portal do INE (<http://ine.pt>). Na Figura 2.4. apresenta-se a página inicial do portal do INE.

³ Marcação por estrela

⁴ *Framework* é uma estrutura de software que fornece uma base para o desenvolvimento de aplicações e sistemas.



Figura 2.4.: Portal do INE Portugal

Além da consulta tradicional via portal, o INE disponibiliza também o acesso aos seus indicadores através de uma API ⁵(*Application Programming Interface* - Interface de Programação de Aplicação), que permite a criação/extração automática do catálogo de indicadores disponíveis no BDD do Portal, bem como da respetiva metainformação. Esta API, de tipo *REST Request*⁶, permite a consulta e extração de dados estatísticos por meio do uso de filtros, utilizando o código do indicador e as respetivas dimensões.

A informação técnica dos indicadores encontra-se disponível em: <http://smi.ine.pt>.

Esta API gera um ficheiro no formato JSON para ser utilizado por aplicações informáticas para difusão de dados estatísticos ou para a produção de conjuntos de dados personalizados, para aplicações móveis, produção de gráficos, etc. As principais características deste serviço incluem:

- Atualização em tempo real;
- Disponível em Português e Inglês;
- Acesso livre e sem custos.

Embora o portal dados.gov.pt também aceda aos indicadores disponibilizados pelo INE através desta API, importa referir que apenas uma pequena fração está atualmente acessível por essa via – cerca de 5,09%, o que corresponde a 740 indicadores, valor consultado no portal dados.gov.pt, face ao total de 14.533 disponíveis no BDD.

2.4. Conjuntos de Dados de Elevado Valor (High-Value Datasets - HVDs)

Na era digital, os dados abertos assumem uma função estratégica e cada vez mais determinante na governação, na economia e na sociedade. A disponibilização gratuita e acessível de dados públicos, sem restrições de uso ou de redistribuição, constitui um instrumento essencial para a promoção da transparência, da inovação e da eficiência na utilização dos recursos públicos. Neste contexto, os Conjuntos de Dados de Elevado Valor (*High-Value Datasets* - HVDs) surgem como elementos

⁵ API – conjunto de regras e especificações que permite a comunicação entre diferentes aplicações de software, facilitando o acesso automático a dados e funcionalidades.

⁶ REST (*Representational State Transfer*) – pedido de um estilo de arquitetura para construção de serviços web.

prioritários, destacando-se pelo seu potencial significativo para gerar impactos positivos e transformadores em múltiplas áreas.

A sua importância é reconhecida a nível europeu, sendo formalmente regulamentados pelo Regulamento de Execução nº2023/138, ao abrigo da Diretiva (UE) 2019/1024, referida na página 2, sobre dados abertos e reutilização da informação do setor público. Este regulamento estabelece que os HVDs devem ser disponibilizados de forma aberta, gratuita, em formatos legíveis por máquinas e com interfaces programáticas de acesso (APIs), garantindo a sua ampla utilização e integração por qualquer entidade pública ou privada.

Os HVDs estão organizados em seis categorias temáticas:

1. Geoespacial;
2. Observação da Terra e do ambiente;
3. Meteorologia;
4. Estatística;
5. Empresas e Propriedade de Empresas;
6. Mobilidade.

A relevância destes conjuntos de dados reside na sua elevada capacidade de fomentar a inovação, apoiar a formulação de políticas baseadas em evidência e potenciar o desenvolvimento de soluções tecnológicas com impacto real na vida dos cidadãos. Ao serem amplamente reutilizados, os HVDs permitem que as empresas otimizem os seus processos, criem produtos e serviços inovadores e melhorem a competitividade económica. No domínio da investigação científica, proporcionam acesso a dados robustos e atualizados, facilitando a realização de estudos avançados e a produção de conhecimento.

Ao nível da Administração Pública, estes dados sustentam decisões mais eficazes e sustentadas no conhecimento, promovendo uma gestão mais transparente, responsiva e centrada nas necessidades da população. A sua disponibilização aberta e acessível permite às organizações aceder a informação estratégica, potenciando análises mais robustas e sustentadas em diversos contextos.

No domínio da análise preditiva, os HVDs possibilitam a antecipação de tendências e a identificação de oportunidades com maior grau de precisão, contribuindo para a eficácia das estratégias organizacionais. Esta capacidade é particularmente relevante nos serviços públicos e na formulação de políticas públicas, onde os dados orientam a implementação de soluções mais eficazes e ajustadas a realidades concretas, com impacto direto na qualidade de vida dos cidadãos.

Além disso, os HVDs permitem a personalização de produtos e serviços, com efeitos positivos na experiência e satisfação dos consumidores. São também essenciais para a definição de estratégias ambientais sustentáveis, nomeadamente no que respeita à conservação de recursos naturais e à proteção de ecossistemas.

Adicionalmente, estes dados contribuem para o reforço da segurança pública, apoiam o planeamento e a implementação de cidades inteligentes e promovem a integração de tecnologias emergentes, como a Internet das Coisas (IoT). No setor da educação e da formação profissional, os HVDs possibilitam o alinhamento das competências adquiridas com as exigências do mercado de trabalho, promovendo maior empregabilidade e capacitação da força laboral.

Em suma, os HVDs constituem um instrumento essencial para enfrentar os desafios atuais e futuros. A sua reutilização promove a colaboração intersetorial, incentiva a inovação orientada por dados e impulsiona a construção de uma sociedade mais informada, inteligente, sustentável e equitativa.

HVDs em Portugal: o caso do INE

Em Portugal, o INE tem desempenhado um papel de destaque na disponibilização e valorização dos HVDs, contribuindo ativamente para a concretização das orientações europeias nesta matéria. O INE centra a sua atuação em duas categorias fundamentais: dados estatísticos e dados geoespaciais.

Atualmente, o INE disponibiliza 56 indicadores estatísticos, acessíveis por nome, tema, geografia e periodicidade, bem como quatro indicadores de natureza geoespacial. Os dados estatísticos são de enorme relevância para a análise de fenómenos sociais, económicos e demográficos, permitindo identificar padrões, tendências e correlações que sustentam a tomada de decisões fundamentadas em múltiplos setores. Estes dados potenciam a eficiência administrativa, apoiam investigadores e decisores políticos na formulação de estratégias eficazes e contribuem para o crescimento económico sustentado.

Os dados geoespaciais, por sua vez, oferecem uma perspetiva territorial essencial, ao incorporarem informações associadas a localizações específicas, como mapas, imagens de satélite, coordenadas GPS e limites administrativos. Estes dados são amplamente utilizados em áreas como o planeamento urbano, a gestão ambiental, a resposta a desastres e os sistemas de navegação. Através da análise geoespacial, é possível compreender com maior profundidade a distribuição espacial dos fenómenos complexos, facilitando a tomada de decisões estratégicas baseadas no território.

A disponibilização destes HVDs pelo INE demonstra não só um compromisso com a transparência e o serviço público, mas também com a promoção do valor económico e social dos dados. Ao assegurar o acesso a conjuntos de dados com elevado potencial de reutilização, o INE contribui ativamente para uma sociedade mais informada, inovadora e participativa, alinhada com os princípios do Governo Aberto e com os objetivos da transformação digital europeia.

3. Arquitetura/Plataformas

Nos últimos anos, a divulgação dos dados governamentais tem sido uma prioridade para a UE, com o objetivo de promover a transparência, inovação e reutilização de informações públicas. A disponibilização desses dados exige a adoção de modelos de arquitetura robustos e plataformas tecnológicas eficientes, garantindo que os dados sejam acessíveis, interoperáveis e seguros.

A escolha do modelo de arquitetura influencia diretamente como os dados são armazenados, organizados e distribuídos, afetando a sua usabilidade, governança e escalabilidade.

Geralmente, um repositório de dados funciona como a peça central de um projeto de dados abertos, servindo como um local central para encontrar dados, padronizar práticas e demonstrar o uso desses dados. Nesse contexto, o repositório funciona como um local de pesquisa onde qualquer pessoa, dentro ou fora do governo, pode aceder aos dados.

Alguns softwares de repositório oferecem funcionalidades que permitem a conversão automática de dados de um formato para outro, como CSV para XML, JSON ou Excel, ampliando a acessibilidade. Além disso, há softwares que possibilitam a visualização de conjuntos de dados diretamente no navegador, permitindo que os utilizadores mapeiem, classifiquem e combinem os dados sem necessidade de conhecimentos de programação. Outros também oferecem funcionalidades de redistribuição, permitindo que as organizações integrem automaticamente esses dados.

Essas capacidades reforçam a importância de uma arquitetura de dados bem estruturada, pois garantem que os dados não apenas sejam disponibilizados, mas também sejam facilmente acessíveis, reutilizáveis e integráveis por diferentes plataformas e utilizadores.

3.1. Catálogo de dados (Modelos)

É possível organizar um sistema de catálogo de dados abertos segundo três modelos distintos. O objetivo é mostrar de que forma os vários elementos e serviços se relacionam entre si, bem como ilustrar as transformações do sistema em diferentes escalas.

Modelo 1: Plataforma Única

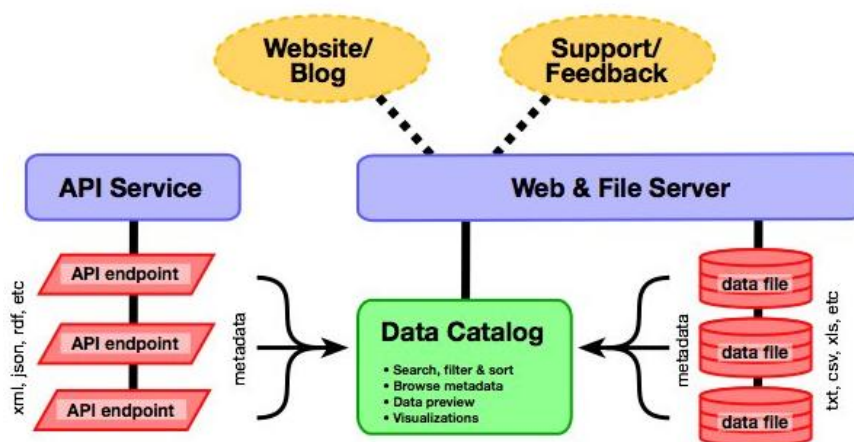


Figura 3.1.: Modelo 1 - Plataforma Única

Este modelo, Figura 3.1.⁷, demonstra uma infraestrutura de Tecnologias de Informação (TI) simples onde o catálogo de dados e os arquivos de dados são alojados num único ambiente de servidor. O servidor pode ser gerido internamente pela organização ou pode ser alojado na nuvem (*cloud*). Conjuntos de dados baseados em API, se houver, podem ser geridos separadamente de acordo com os requisitos da tecnologia subjacente.

Website, suporte ao utilizador e *feedback* são elementos essenciais do envolvimento do utilizador numa iniciativa de dados abertos e muitas vezes podem ser fornecidos pela mesma infraestrutura ou por uma infraestrutura semelhante à usada pelo próprio catálogo. Conceptualmente, no entanto, são sistemas separados que estão apenas vagamente conectados ao catálogo de dados.

Este modelo é adequado quando há um pequeno número de conjuntos de dados (menos de 200) no catálogo de dados, os conjuntos de dados são pequenos (menos de 100Mb) e uma única entidade desempenha um papel importante na coordenação do catálogo de dados e na gestão da infraestrutura de TI.

⁷ Fonte: <https://opendatatoolkit.worldbank.org/en/data/opendatatoolkit/technology#catalog-look-like>

Modelo 2: Servidores Separados

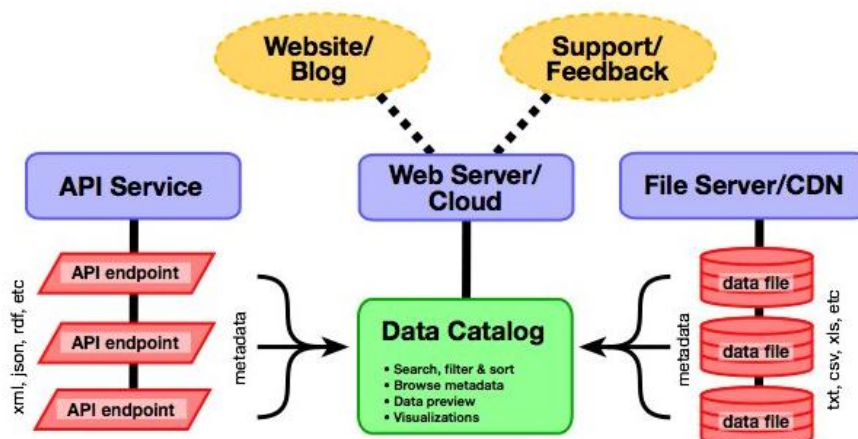


Figura 3.2.: Modelo 2 - Servidores Separados

Este modelo, Figura 3.2.⁸, ilustra como o catálogo de dados e o servidor de arquivos podem ser geridos separadamente, utilizando infraestrutura alojada internamente ou na nuvem. Esta abordagem é ligeiramente mais sofisticada do que o Modelo 1, sendo mais adequada para catálogos e conjuntos de dados de maior dimensão. A utilização de infraestrutura na nuvem é frequentemente uma solução económica, embora possa não ser viável em regiões com largura de banda limitada ou em cenários onde há a necessidade de um maior controlo sobre a informação.

Modelo 3: Catálogos Federados

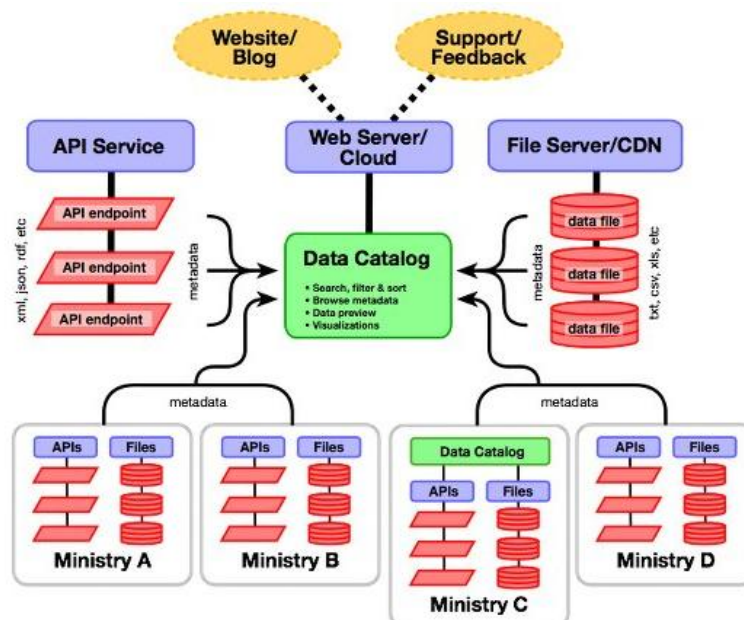


Figura 3.3.: Modelo 3 - Catálogos Federados

Este modelo, Figura 3.3.⁹, demonstra como a gestão do catálogo de dados abertos pode ser descentralizada, envolvendo uma ou mais entidades contribuintes. Nesta abordagem, determinados arquivos de dados e/ou serviços de API são geridos por entidades distintas, enquanto os metadados ainda

⁸ Fonte: <https://opendatatoolkit.worldbank.org/en/data/opendatatoolkit/technology#catalog-look-like>

⁹ Fonte: <https://opendatatoolkit.worldbank.org/en/data/opendatatoolkit/technology#catalog-look-like>

são fornecidos ao catálogo central, permitindo a pesquisa e acesso entre entidades. Uma entidade pode, inclusive, operar o seu próprio catálogo (por exemplo, dados geoespaciais ou estatísticas educacionais) mantendo, simultaneamente, suporte ao catálogo central. Este modelo é especialmente adequado quando uma ou mais entidades possuem a capacidade técnica e a experiência necessária para gerir os seus próprios dados abertos. As entidades que não dispõem dessa capacidade podem optar por confiar na entidade implementadora, conforme ilustrado nos anteriores Modelo 1 ou Modelo 2.

3.2. Plataformas

As plataformas de dados abertos são soluções tecnológicas que facilitam a publicação, gestão e distribuição de dados abertos na *web*.

Para os administradores, essas plataformas oferecem ferramentas para organizar e disponibilizar conjuntos de dados de forma estruturada, garantindo padronização e acessibilidade. Para os utilizadores, proporcionam uma experiência consistente e intuitiva, permitindo a descoberta, visualização e reutilização dos dados de maneira eficiente.

Além de centralizarem informações, essas plataformas promovem a transparência e a interoperabilidade, tomando os dados abertos mais acessíveis e utilizáveis.

Plataformas mais conhecidas:



CKAN (*Comprehensive Knowledge Archive Network*)

URL: <http://ckan.org/>

CKAN é uma plataforma de código aberto amplamente utilizada para a publicação e gestão de dados abertos. Desenvolvida pela *Open Knowledge Foundation*, pode ser instalada em servidores Linux, inclusive em ambientes de computação em nuvem.

A plataforma é escrita em *Python* e foi projetada para permitir a publicação e a gestão de dados tanto por meio de uma interface *web* quanto através de uma API, facilitando a integração com outros sistemas.

O CKAN pode ser instalado em alojamento próprio ou fornecido em modo de serviço *hosting*¹⁰ por uma taxa mensal.

Esta plataforma possui uma arquitetura modular através da qual recursos adicionais ou personalizados podem ser adicionados. Um exemplo é a extensão *DDI Importer*, que possibilita a importação de metadados de catálogos de microdados, ampliando as capacidades de catalogação da plataforma.

Exemplos:

- Brasil (<http://dados.gov.br/>);
- Estados Unidos (<http://catalog.data.gov/>);
- Africa Open Data (<http://africaopendata.org/>);

¹⁰ Serviço *hosting* é quando uma empresa externa cuida de toda a parte técnica de um sistema, como a instalação, configuração e manutenção, e o mantém a funcionar num servidor próprio. Assim, o utilizador pode usar a plataforma sem precisar preocupar-se com servidores ou configurações técnicas.

- Espanha (<http://datos.gob.es>).

Vantagens:

- **Código aberto:** Permite personalizações e adaptações conforme as necessidades específicas de cada organização;
- **Ampla adoção:** Utilizada por diversos governos e organizações globais, o que evidencia a sua robustez e confiabilidade;
- **Comunidade ativa:** Possui uma comunidade vibrante que contribui com *plugins*, extensões e suporte contínuo.

Desvantagens:

- **Complexidade na instalação e manutenção:** Requer conhecimentos técnicos avançados para a implementação e gestão, o que pode ser um desafio para equipes com recursos limitados;
- **Interface de utilizador básica:** Embora funcional, pode necessitar de customizações para atender a requisitos estéticos ou de usabilidade específicos.



DKAN

URL: <https://getdkan.org/>

A Plataforma DKAN foi desenvolvida com foco na compatibilidade com o CKAN, possuindo uma API praticamente idêntica. Isso permite que sistemas e aplicações originalmente criados para o CKAN funcionem no DKAN sem necessidade de modificações.

Diferentemente do CKAN, que é baseado em *Python*, o DKAN foi construído sobre o Drupal, um dos sistemas de conteúdos (CMS) mais populares, desenvolvido em PHP. Essa característica torna o DKAN uma escolha estratégica para organizações que já utilizam infraestrutura baseada em Drupal, facilitando a sua implementação e integração com sites institucionais existentes.

O Drupal possui a sua própria arquitetura modular e conta com milhares de módulos disponíveis, além de permitir a criação de módulos personalizados. Também dispõe de uma comunidade ativa de desenvolvedores, o que contribui para a constante evolução da plataforma.

Exemplo:

- Colónia – Alemanha (<http://www.offenedaten-koeln.de/>)

Vantagens:

- **Baseado em Drupal:** Aproveita a flexibilidade e os recursos do Drupal, facilitando a integração com sites institucionais que já utilizam essa plataforma;
- **API compatível com CKAN:** Permite a reutilização de aplicações e ferramentas desenvolvidas pela CKAN, oferecendo versatilidade na escolha da plataforma;
- **Arquitetura modular:** Possui milhares de módulos disponíveis para download, permitindo extensões e customizações conforme as necessidades da organização.

Desvantagens:

- **Dependência do Drupal:** Organizações que não estejam familiarizadas com o Drupal podem enfrentar desafios de adaptação e implementação;
- **Performance:** Pode apresentar desafios de desempenho em comparação com soluções mais especializadas, dependendo da configuração e do volume de dados.



Socrata

URL: <http://www.socrata.com/>

A Socrata é uma plataforma de catálogo de dados abertos baseada na nuvem, que opera no modelo *Software as a Service* (SaaS), fornecendo ferramentas para a gestão, visualização e disponibilização de dados públicos. Destinada principalmente a governos e organizações, a plataforma permite a publicação de conjuntos de dados de forma acessível e interativa, promovendo a transparência e a inovação.

Uma das funcionalidades distintivas da Socrata é a possibilidade de os utilizadores criarem visualizações personalizadas diretamente a partir dos dados publicados. Essas visualizações podem ser salvas e reutilizadas por outros utilizadores, o que favorece o trabalho colaborativo e a disseminação de informação. Além disso, a plataforma oferece APIs que possibilitam a integração com outros sistemas e o desenvolvimento de aplicações baseadas nos dados abertos.

A Socrata é disponibilizada exclusivamente como uma solução hospedada, sem a opção de instalação em servidores próprios. Ainda assim, a plataforma fornece uma versão de código aberto da sua API, o que facilita a migração de dados e aplicações caso a organização deseje deixar o modelo SaaS.

Reconhecida como um dos principais fornecedores no setor de repositórios de dados abertos, a Socrata apresenta uma solução robusta e rica em funcionalidades. No entanto, o seu custo elevado pode representar um obstáculo, com valores que podem atingir centenas de milhares de dólares por ano.

Exemplos:

- U.S. Medicare (<https://data.medicare.gov/>);
- City of Los Angeles (<https://data.lacity.org/>);
- Global Island Partnership (<https://impact.glispa.org/>);
- Carbon Disclosure Project (<https://data.cdp.net/>).

Vantagens:

- **Solução completa:** Oferece ferramentas integradas para gestão, visualização e análise de dados, facilitando a publicação e o acesso a informações públicas;
- **Interface amigável:** Proporciona uma experiência de utilizador intuitiva, com recursos avançados de visualização e criação de *dashboards*;
- **Suporte e atualizações:** Por ser uma solução comercial, oferece suporte técnico dedicado e atualizações regulares da plataforma.

Desvantagens:

- **Custo elevado:** O preço da solução pode ser um desafio para organizações com orçamento restrito, podendo atingir valores significativos anuais;
- **Dependência do fornecedor:** A natureza proprietária da plataforma pode limitar a flexibilidade e a capacidade de personalização, além de gerar dependência em relação ao fornecedor.



OpenDataSoft

URL: <http://www.opendatasoft.com/>

O OpenDataSoft é uma plataforma SaaS baseada na nuvem, que oferece uma solução robusta para a publicação, gestão e visualização de dados abertos.

Uma das suas principais características é a disponibilização de um *front-end* de código aberto, o que proporciona flexibilidade para personalizações e adaptações conforme as necessidades dos utilizadores e das organizações.

A plataforma suporta diversos formatos abertos amplamente utilizados, como CSV, JSON e XML, além de formatos geoespaciais como KML, OSM e SHP. Isso torna a plataforma altamente compatível com diferentes tipos de dados, incluindo dados geográficos e tabulares, o que é essencial para organizações que trabalham com grandes volumes de dados multidimensionais e geoespaciais.

A plataforma destaca-se ainda pela interface de pesquisa intuitiva, suporte multilíngue e geração automática de APIs a partir dos conjuntos de dados carregados, facilitando a integração com outras ferramentas e aplicações externas. A gestão centralizada de dados também permite que diferentes departamentos de uma mesma organização tenham um ponto único de acesso e administração, reduzindo redundâncias e promovendo maior consistência informacional.

Exemplos:

- Kapsarc – Arábia Saudita (<https://datasource.kapsarc.org/page/home/>);
- Paris Data (<https://opendata.paris.fr/pages/home/>);
- City of Cary – USA (<https://data.townofcary.org/pages/homepage/>).

Vantagens:

- **Front-end de código aberto:** Permite personalizações e ajustes conforme as necessidades dos utilizadores e das organizações;
- **Suporte a múltiplos formatos:** Compatível com diversos formatos de dados, incluindo geoespaciais, ampliando as suas aplicações;
- **Interface de pesquisa intuitiva:** Facilita a descoberta e a navegação pelos conjuntos de dados disponíveis;
- **Facilidade na partilha de dados:** Permite que organizações disponibilizem e distribuam os seus dados de forma acessível, promovendo a transparência e a inovação;
- **APIs automatizadas:** A plataforma gera APIs automaticamente a partir dos conjuntos de dados carregados, facilitando a integração com outras ferramentas e aplicações externas;
- **Gestão centralizada de dados:** Permite que diferentes departamentos dentro de uma organização tenham um ponto único de acesso e gestão de dados, reduzindo redundâncias e inconsistências.

Desvantagens:

- **Modelo SaaS:** Pode não ser adequado para organizações que preferem ou necessitam de soluções hospedadas internamente;
- **Custos associados:** Embora ofereça funcionalidades avançadas, os custos podem ser um fator limitador para algumas instituições;

- **Limitações na personalização avançada:** Apesar do *front-end* ser de código aberto, modificar a plataforma além das configurações padrão pode exigir conhecimentos técnicos mais especializados;
- **Dependência da infraestrutura do fornecedor:** Como é um serviço baseado na nuvem, as organizações dependem do fornecedor para atualizações, suporte e manutenção, o que pode não ser ideal para todos os casos;
- **Gestão de permissões pode ser complexa:** Para organizações com requisitos rigorosos de controle de acesso, a configuração de permissões pode ser trabalhosa e exigir ajustes constantes.



Udata

URL: <https://udata.ca/>

O **Udata** é uma plataforma de código aberto destinada à gestão e publicação de dados abertos, amplamente utilizada por governos e organizações com o objetivo de promover a transparência e a reutilização de dados públicos. Desenvolvida e mantida pela **Etalab**, entidade vinculada ao governo francês responsável pela estratégia nacional de dados e governo digital, o Udata oferece funcionalidades de catalogação, indexação e visualização de conjuntos de dados, facilitando o seu acesso por cidadãos, empresas e desenvolvedores.

Com arquitetura modular e extensível, a plataforma suporta metadados detalhados e disponibiliza APIs que possibilitam a integração com outros sistemas. A plataforma também permite a adição de funcionalidades por meio de *plugins*, o que a torna altamente flexível e adaptável às necessidades institucionais. A interface intuitiva favorece a pesquisa, exploração e reutilização dos dados, incentivando a colaboração entre utilizadores.

Ao contrário de soluções SaaS, que requerem um fornecedor externo, o Udata pode ser instalado e gerido em infraestruturas próprias, garantindo maior controlo sobre os dados e a sua administração. A sua adoção tem vindo a crescer entre administrações públicas que procuram uma alternativa independente, personalizável e de confiança para a gestão de dados governamentais.

Exemplos:

- Portugal (<https://dados.gov.pt/>)
- França (<https://www.data.gouv.fr/>)
- Luxemburgo (<https://data.public.lu/>)

Vantagens:

- **Código aberto:** Permite flexibilidade para personalizações e adaptações conforme as necessidades específicas de cada organização;
- **Arquitetura modular:** Permite a adição de novas funcionalidades via *plugins*, tornando-a uma solução adaptável a diferentes contextos;
- **Implementação interna:** Pode ser instalada e gerida em infraestruturas próprias, garantindo maior controlo sobre os dados e a sua administração.

Desvantagens:

- **Comunidade limitada:** Embora esteja em crescimento, a comunidade de utilizadores e desenvolvedores é menor em comparação com outras plataformas, o que pode limitar o suporte e a disponibilidade de recursos;
- **Processo de aprendizagem:** Pode requerer um período de adaptação para equipas não familiarizadas com a plataforma ou com tecnologias relacionadas.

Plataformas Geoespaciais:



ArcGIS Open Data

URL: <http://opendata.arcgis.com/>

O ArcGIS Open Data é uma plataforma SaaS baseada na nuvem, onde os utilizadores podem explorar dados espaciais e não espaciais por meio de uma interface consistente e acessível. Integrado no ArcGIS Online, o serviço é disponibilizado gratuitamente para dados abertos e aproveita toda a infraestrutura da plataforma ArcGIS, amplamente utilizada em ambientes de análise geoespacial.

A plataforma permite a extração de recursos específicos e o *download* dos dados em diversos formatos abertos, bem como o acesso via APIs. Um dos elementos tecnológicos centrais do ArcGIS Open Data é o *Koop*, um mecanismo ETL (*Extract, Transform, Load*) de código aberto, que converte automaticamente serviços *web* em formatos acessíveis e reutilizáveis. A solução integra-se com centenas de aplicações de código aberto para dispositivos móveis, *web* e *desktop*, ampliando significativamente o seu potencial de uso.

Exemplos:

- City of Los Angeles (<http://geohub.lacity.org/>)
- US Department of Homeland Security (<https://hifld-dhs-gii.opendata.arcgis.com/>)
- Global Forest Watch (<http://data.globalforestwatch.org/>)

Vantagens:

- **Integração com ArcGIS:** Aproveita os serviços e a infraestrutura do ArcGIS *Online*, que é uma plataforma amplamente utilizada para dados espaciais;
- **Facilidade de Uso:** Interface consistente e fácil de usar para explorar e compartilhar dados;
- **API e Personalização:** Oferece APIs para integração com outras plataformas e desenvolvimento de aplicações personalizadas;
- **Capacidade de Extração e Download:** Permite o *download* dos dados em vários formatos abertos, facilitando a reutilização.

Desvantagens:

- **Custos:** Embora a plataforma seja gratuita para os dados abertos, o ArcGIS Online e outros serviços do ArcGIS podem ter custos elevados, conforme o volume de utilização e do número de recursos;
- **Dependência do Ecossistema:** A plataforma é fortemente integrada ao ArcGIS, o que pode tornar difícil a integração com sistemas fora desse ecossistema;

- **Requisitos de Infraestrutura:** A integração total com serviços e API do ArcGIS pode exigir uma infraestrutura adicional, especialmente em grandes volumes de dados.



GeoNode

URL: <https://geonode.org/>

O GeoNode é uma plataforma de código aberto destinada ao desenvolvimento de sistemas de informação geoespacial e para a implementação de infraestruturas de dados espaciais. Projetado com foco na extensibilidade, o GeoNode permite adaptações conforme as necessidades institucionais e pode ser integrado com plataformas já existentes, incluindo outras ferramentas de Sistemas de Informação Geográfica (SIG) e portais de dados abertos.

Exemplos:

- UN World Food Programme (<https://geonode.wfp.org/>)
- Haiti Data (<https://haitidata.org/>)

Vantagens:

- **Código aberto:** Permite personalização completa e adaptação às necessidades específicas da organização;
- **Integração com outras plataformas:** Compatível com sistemas externos, como soluções de dados abertos e outras ferramentas do SIG;
- **Flexibilidade:** A plataforma é projetada para ser estendida, permitindo que os desenvolvedores a modifiquem de acordo com os seus requisitos;
- **Comunitária e Suporte Ativo:** Como uma plataforma de código aberto, conta com uma comunidade ativa de desenvolvedores e uma boa documentação.

Desvantagens:

- **Requisitos Técnicos:** A instalação, configuração e manutenção podem exigir um conhecimento técnico significativo, especialmente em ambientes personalizados;
- **Potencial crescimento:** A expansão da plataforma pode enfrentar limitações dependendo do volume e da complexidade dos dados, exigindo uma infraestrutura adequada;
- **Interface e Usabilidade:** A interface de utilizador pode não ser tão intuitiva quanto soluções comerciais, exigindo treino para utilizadores menos experientes.

3.3. Modelo de Metadados

A evolução dos modelos de metadados nos projetos de dados abertos a nível internacional tem seguido um percurso progressivo, adaptando-se às necessidades de interoperabilidade e reutilização da informação. Essa evolução pode ser segmentada em diferentes fases, refletindo o nível de maturidade dos sistemas de catalogação de dados.

Standard Interno

A fase inicial da adoção de metadados ocorre quando uma entidade define a sua própria estrutura de descrição da informação. Nesta etapa, cada organização estabelece um modelo descritivo próprio, incluindo os campos de metadados, descrições e formatos para estruturar os conjuntos de dados. Apesar de oferecer um nível básico de organização, esta abordagem pode originar dificuldades na partilha e reutilização de dados entre diferentes entidades, devido à falta de uniformização.

Dublin Core

Um avanço significativo na padronização dos metadados é representado pela adoção do modelo Dublin Core, um esquema amplamente utilizado para descrever objetos digitais, incluindo textos, imagens, vídeos e páginas *web*. Baseado em XML e RDF (*Resource Description Framework*), este modelo foi desenvolvido pela *Dublin Core Metadata Initiative* (DCMI), uma organização internacional que apoia a inovação em design de metadados e melhores práticas em toda a ecologia de metadados. O principal objetivo do *Dublin Core* é facilitar a pesquisa e a recuperação de informação, garantindo uma descrição mais estruturada dos conteúdos digitais.

DCAT – Data Catalog Vocabulary

Com a crescente necessidade de interoperabilidade entre catálogos de dados públicos, surge o DCAT (*Data Catalog Vocabulary*), um modelo mais avançado e abrangente, amplamente adotado a nível europeu. O DCAT foi desenvolvido para permitir a descrição uniforme de *datasets* em catálogos de dados governamentais, promovendo a pesquisa e o intercâmbio de informação entre diferentes sistemas e países. A sua implementação visa tornar os dados públicos mais acessíveis e reutilizáveis em diversos contextos, incluindo administração pública, investigação científica e desenvolvimento tecnológico.

DCAT-AP

Como evolução do modelo DCAT, a União Europeia desenvolveu o DCAT-AP (*DCAT Application Profile for data portals in Europe*), uma especificação criada para facilitar a descrição e a interoperabilidade de dados setoriais em toda a Europa. A principal vantagem do DCAT-AP reside na possibilidade de os conjuntos de dados publicados por diferentes países e setores possam ser facilmente pesquisáveis e interligados, garantindo um elevado grau de harmonização na publicação de dados públicos. Isso pode ser alcançado pela troca de descrições de conjuntos de dados entre portais de dados.

Desde 2015, a Comissão Europeia, através do programa ISA (*Interoperability Solutions for Public Administrations*), tem promovido a revisão e atualização do DCAT-AP, assegurando que o modelo responda às necessidades emergentes da governação digital e da economia baseada em dados. A adesão a este modelo tem sido fortemente incentivada entre os Estados-Membros da União Europeia, com o objetivo de fortalecer a transparência, a colaboração internacional e a reutilização de dados públicos.

De acordo com o relatório *Open Data Maturity* de 2024, verifica-se que 89% dos Estados-Membros, com exceção da Bulgária, da Grécia e de Malta, adotam a estrutura do modelo DCAT-AP ou garantem a sua interoperabilidade. Adicionalmente, na maioria dos países da UE, a percentagem de conjuntos de dados com metadados em conformidade com a especificação DCAT-AP ultrapassa os 90%.

Esta realidade é evidenciada na Tabela 3.1, onde se apresenta um levantamento comparativo dos portais de dados abertos, das plataformas tecnológicas utilizadas e dos níveis de adoção do modelo DCAT-AP entre diversos países.

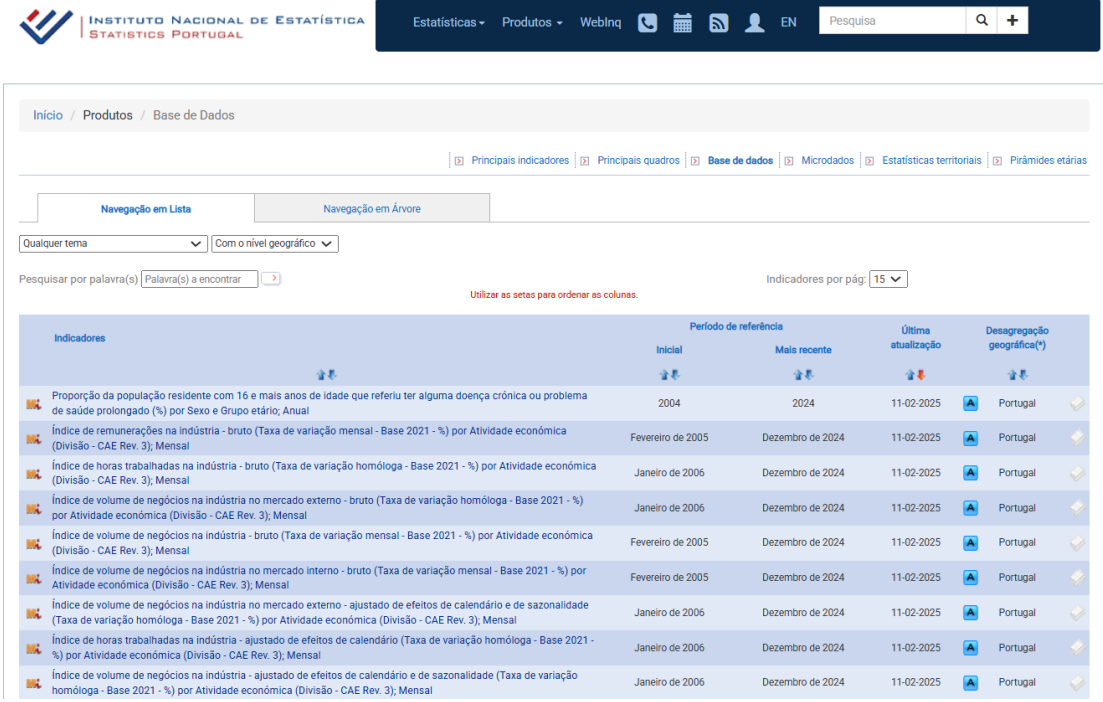
Tabela 3.1.: Comparação dos portais de dados abertos, plataformas tecnológicas e grau de conformidade com o modelo de metadados DCAT-AP

País	Portal de dados abertos (open government)	Plataforma open data	Modelo de metadados
Bósnia e Herzegovina	Nenhum		
Albânia	https://opendata.gov.al/en	DKAN	DCAT-AP (<10%)
Sérvia	https://data.gov.rs/	Udata	DCAT-AP (>90%)
Ucrânia	https://data.gov.ua/en	CKAN	DCAT-AP (>90%)
Islândia	https://opingogn.is/	CKAN	DCAT-AP (<10%)
Noruega	https://data.norge.no/	Custom, including React, Java/Kotlin, Python, Jena, Fuseki, Elasticsearch, PostgreSQL, MongoDB and Kafka	DCAT-AP (61%-70%)
Suíça	https://opendata.swiss/en/	CKAN	DCAT-AP (>90%)
União Europeia	https://data.europa.eu/pt	CKAN	
Bélgica	https://data.gov.be/en	Custom back end with Drupal front end	DCAT-AP (81%-90%)
Bulgária	https://data.egov.bg/	Custom, including Fluentd, Elasticsearch node, MariaDB and Graylog	DCAT-AP (21%-30%)
República Checa	https://data.gov.cz/english	Custom, including LinkedPipes	DCAT-AP (>90%)
Dinamarca	http://www.datavejviser.dk/	CKAN back end with a DCAT plug-in and a front end designed with React	DCAT-AP (>90%)
Alemanha	https://www.govdata.de/	Custom, including the CMS Typo3 for editorial content and Piveau for data storage	DCAT-AP (51%-60%)
Estónia	https://avaandmed.eesti.ee/	Custom, including Typescript, PostgreSQL and Solr	DCAT-AP (>90%)
Irlanda	https://data.gov.ie/	CKAN	DCAT-AP (71%-80%)
Grécia	https://data.gov.gr/ http://repository.data.gov.gr/	Custom based on CKAN	DCAT-AP (81%-90%)
Espanha	https://datos.gob.es/en	CKAN and Drupal	DCAT-AP (>90%)

França	https://www.data.gouv.fr/	Udata	DCAT-AP (>90%)
Croácia	https://data.codefor-croatia.org/	CKAN	DCAT-AP (61%-70%)
Itália	https://dati.gov.it/	CKAN and Drupal	DCAT-AP (>90%)
Chipre	https://www.data.gov.cy/	EKAN	DCAT-AP (>90%)
Letónia	https://data.gov.lv/en	CKAN	DCAT-AP (81%-90%)
Lituânia	https://data.gov.lt/?lang=en	Self-developed solution in Python	DCAT-AP (>90%)
Luxemburgo	https://data.public.lu/en	Udata	DCAT-AP (81%-90%)
Hungria	https://kozadatportalo.hu/	CKAN	DCAT-AP (>90%)
Malta	https://open.data.gov.mt/	Custom	DCAT-AP (31%-40%)
Países Baixos	https://data.overheid.nl/en	CKAN	DCAT-AP (61%-80%)
Áustria	https://www.data.gv.at/en	CKAN and Wordpress	DCAT-AP (81%-90%)
Polónia	https://dane.gov.pl/	Custom, including Falcon, Django, RDFLib, Wagtail CMS, Typescript	DCAT-AP (>90%)
Portugal	https://dados.gov.pt/en/	Udata	DCAT-AP (>90%)
Roménia	https://data.gov.ro/en	CKAN	DCAT-AP (51%-60%)
Eslovénia	https://podatki.gov.si/	CKAN	DCAT-AP (81%-90%)
Eslováquia	https://data.gov.sk/en	Custom, including LinkedPipes	DCAT-AP (>90%)
Finlândia	https://www.avoindata.fi/en	CKAN	DCAT-AP (>90%)
Suécia	https://www.dataportal.se/en	Custom, including EntryScape, the Strapi CMS and NodeBB	DCAT-AP (81%-90%)

4. Proposta de Solução para o INE

O Banco de Dados de Difusão (BDD) do INE constitui uma base estruturada composta atualmente por cerca de 12.600 indicadores, atualizados regularmente. Este banco de dados alimenta o produto "Base de Dados", disponível para consulta no Portal do INE, apresentado na Figura 4.1.. A plataforma oferece acesso a 25 temas distintos, organizados tanto em formato de lista quanto em estrutura hierárquica (árvore), além de disponibilizar um mecanismo de pesquisa por palavra-chave, facilitando a navegação e a recuperação de informação estatística.



The screenshot displays the INE Data Portal interface. At the top, there is a navigation bar with the INE logo and menu items: Estatísticas, Produtos, WebInq, and a search bar. Below the navigation bar, there are tabs for 'Principais indicadores', 'Principais quadros', 'Base de dados', 'Microdados', 'Estatísticas territoriais', and 'Pirâmides etárias'. The main content area shows a list of indicators with columns for 'Indicadores', 'Período de referência' (Initial and Most recent), 'Última atualização', and 'Desagregação geográfica(*)'. The list includes indicators such as 'Proporção da população residente com 16 e mais anos de idade que referiu ter alguma doença crónica ou problema de saúde prolongado (%) por Sexo e Grupo etário, Anual' and 'Índice de remunerações na indústria - bruto (Taxa de variação mensal - Base 2021 - %) por Atividade económica (Divisão - CAE Rev. 3); Mensal'.

Indicadores	Período de referência		Última atualização	Desagregação geográfica(*)
	Inicial	Mais recente		
Proporção da população residente com 16 e mais anos de idade que referiu ter alguma doença crónica ou problema de saúde prolongado (%) por Sexo e Grupo etário, Anual	2004	2024	11-02-2025	Portugal
Índice de remunerações na indústria - bruto (Taxa de variação mensal - Base 2021 - %) por Atividade económica (Divisão - CAE Rev. 3); Mensal	Fevereiro de 2005	Dezembro de 2024	11-02-2025	Portugal
Índice de horas trabalhadas na indústria - bruto (Taxa de variação homóloga - Base 2021 - %) por Atividade económica (Divisão - CAE Rev. 3); Mensal	Janeiro de 2006	Dezembro de 2024	11-02-2025	Portugal
Índice de volume de negócios na indústria no mercado externo - bruto (Taxa de variação homóloga - Base 2021 - %) por Atividade económica (Divisão - CAE Rev. 3); Mensal	Janeiro de 2006	Dezembro de 2024	11-02-2025	Portugal
Índice de volume de negócios na indústria - bruto (Taxa de variação mensal - Base 2021 - %) por Atividade económica (Divisão - CAE Rev. 3); Mensal	Fevereiro de 2005	Dezembro de 2024	11-02-2025	Portugal
Índice de volume de negócios na indústria no mercado interno - bruto (Taxa de variação mensal - Base 2021 - %) por Atividade económica (Divisão - CAE Rev. 3); Mensal	Fevereiro de 2005	Dezembro de 2024	11-02-2025	Portugal
Índice de volume de negócios na indústria no mercado externo - ajustado de efeitos de calendário e de sazonalidade (Taxa de variação homóloga - Base 2021 - %) por Atividade económica (Divisão - CAE Rev. 3); Mensal	Janeiro de 2006	Dezembro de 2024	11-02-2025	Portugal
Índice de horas trabalhadas na indústria - ajustado de efeitos de calendário (Taxa de variação homóloga - Base 2021 - %) por Atividade económica (Divisão - CAE Rev. 3); Mensal	Janeiro de 2006	Dezembro de 2024	11-02-2025	Portugal
Índice de volume de negócios na indústria - ajustado de efeitos de calendário e de sazonalidade (Taxa de variação homóloga - Base 2021 - %) por Atividade económica (Divisão - CAE Rev. 3); Mensal	Janeiro de 2006	Dezembro de 2024	11-02-2025	Portugal

Figura 4.1.: Base de dados do portal do INE

A disponibilização de dados abertos constitui uma importante ferramenta para facilitar o acesso e a reutilização da informação estatística por parte dos utilizadores. Contudo, o principal desafio reside precisamente na criação de mecanismos que promovam esse acesso de forma eficiente e estruturada.

Considerando que apenas uma fração dos indicadores do INE se encontra atualmente incorporada no portal dados.gov.pt, recomenda-se a implementação de uma área específica no Portal do INE dedicada à disponibilização de dados abertos (<https://dados.ine.pt>).

4.1. Arquitetura

A seleção da arquitetura mais adequada para o portal de dados abertos do INE foi fundamentada numa análise das melhores práticas e das soluções mais eficazes implementadas em países com um elevado grau de maturidade na disponibilização de dados abertos. Para esta análise, foi levado em consideração o estudo sobre a maturidade dos dados abertos promovido pela iniciativa *European Data Portal* da Comissão Europeia, além da atratividade e eficiência dos portais de dados abertos de diversos países da União Europeia. Dentro deste grupo, destacaram-se os portais de França, Irlanda e Espanha.

A escolha destes países foi motivada não apenas pela sua experiência consolidada na gestão de dados abertos, mas também pela existência de infraestruturas tecnologicamente avançadas e pela adesão rigorosa às diretrizes europeias, nomeadamente à [Diretiva \(UE\) 2019/1024](#), que regula a reutilização de informações do setor público e a disponibilização de dados abertos.

Portal de França

O portal de dados abertos de França, que se observa na Figura 4.2., é mantido pela Etalab, uma unidade governamental francesa dedicada à coordenação de iniciativas relacionadas com dados abertos e governo digital. Este portal utiliza a plataforma Udata, uma solução de código aberto desenvolvida especificamente para a gestão e distribuição de catálogos de dados abertos.



Figura 4.2.: Portal de França

Principais características

Como referido anteriormente, a infraestrutura técnica do portal da França é suportada pela plataforma Udata, desenvolvida para armazenar, documentar e distribuir dados abertos de forma eficiente e segura. O modelo de gestão adotado é descentralizado, permitindo que diferentes organismos da administração pública publiquem diretamente os seus conjuntos de dados, o que assegura maior autonomia e agilidade no processo de disponibilização da informação.

Uma das particularidades do portal é o suporte a *Linked Data*, que facilita a interconexão entre conjuntos de dados e promove a interoperabilidade entre diferentes fontes de informação. Além disso, o portal disponibiliza *APIs REST* que possibilitam o acesso automatizado aos dados, promovendo a reutilização em aplicações e serviços externos. A sua arquitetura é extensível, pois permite a adaptação e integração com outras ferramentas tecnológicas, embora o seu uso seja predominantemente vinculado às especificações definidos pelo governo francês.

Vantagens

Entre as principais vantagens do portal francês destacam-se a sua arquitetura flexível e modular, que permite personalizações adaptadas às necessidades específicas de cada entidade fornecedora de dados.

O suporte nativo a metadados e a forte aderência aos padrões europeus garantem elevada interoperabilidade e conformidade com as boas práticas definidas a nível comunitário. Por fim, a interface do portal é reconhecida pela sua usabilidade, sendo intuitiva e interativa, o que facilita o acesso e exploração dos dados por diversos perfis de utilizadores.

Desvantagens

Apesar das suas qualidades, o portal apresenta algumas limitações. Uma das principais diz respeito à dependência da plataforma Udata, cuja adoção fora do contexto francês ainda é limitada. Esta especificidade pode representar um desafio para iniciativas que visem maior alinhamento internacional ou integração com outras soluções amplamente difundidas, como o CKAN.

Portal de Irlanda

O portal de dados abertos da Irlanda, apresentado na Figura 4.3., é gerido pelo departamento de Turismo, Cultura, Arte, Desporto e Media da Irlanda, e serve como um repositório central para os dados abertos do governo irlandês.

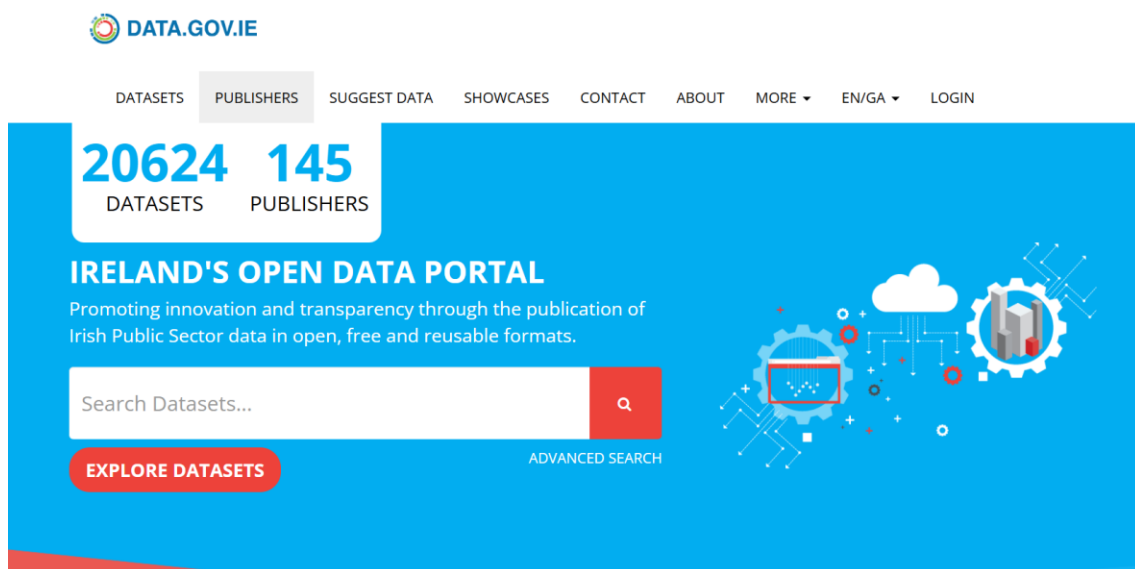


Figura 4.3.: Portal da Irlanda

Principais características

O portal de dados abertos da Irlanda baseia-se na plataforma CKAN, uma das soluções de código aberto mais amplamente adotadas internacionalmente para a gestão e publicação de dados governamentais. A utilização desta plataforma garante a compatibilidade com padrões técnicos amplamente reconhecidos e facilita a interoperabilidade com outras iniciativas globais de dados abertos.

Uma das principais características do portal irlandês é a disponibilização de APIs, que permitem o acesso automatizado aos conjuntos de dados. Esta funcionalidade incentiva a reutilização da informação por terceiros, nomeadamente investigadores, empresas e organizações da sociedade civil. O portal destaca-se também pelo forte compromisso com a acessibilidade e a inclusão digital, seguindo as diretrizes europeias nesta matéria, de forma a garantir que todos os cidadãos, independentemente das suas limitações, possam aceder aos dados públicos.

Adicionalmente, a interface do portal é intuitiva e orientada para o utilizador, incorporando funcionalidades de pesquisa avançada e categorização temática dos conjuntos de dados, o que permite uma navegação simples e eficaz direcionada às necessidades de informação dos utilizadores.

Vantagens

Entre as principais vantagens do portal irlandês destaca-se o elevado nível de organização e estrutura proporcionado pela plataforma CKAN, que torna a gestão e o acesso aos dados particularmente funcionais e acessíveis. A arquitetura da solução está em plena conformidade com os padrões europeus, tanto em termos de acessibilidade digital quanto de reutilização de dados, o que garante a adesão às melhores práticas definidas pela União Europeia. A interface amigável, complementada com filtros de pesquisa bem definidos, contribui para uma experiência de utilização eficaz e fluida.

Desvantagens

Apesar dos seus pontos fortes, o portal irlandês pode apresentar limitações no que respeita à necessidade de desenvolvimento adicional para a implementação de funcionalidades específicas que não estejam contempladas de forma nativa na plataforma CKAN. Este fator pode implicar recursos técnicos e operacionais adicionais, sobretudo em contextos que exijam soluções altamente personalizadas.

Portal da Espanha

O portal de dados abertos da Espanha, que se pode observar na Figura 4.4., é um dos mais bem documentados e estruturados da Europa, com uma arquitetura baseada na plataforma CKAN.



Figura 4.4.: Portal da Espanha

Principais características

O portal de dados abertos da Espanha também usa a plataforma CKAN, uma infraestrutura compatível com os principais padrões técnicos do setor, o que facilita a interoperabilidade com outros sistemas.

Uma das funcionalidades centrais do portal é o motor de agregação, que permite a sincronização de catálogos de dados provenientes de diversas entidades governamentais, promovendo assim uma visão integrada dos dados disponíveis. O portal apresenta também uma abordagem multilíngue, assegurando acessibilidade para diferentes idiomas, o que facilita o uso por cidadãos de diversas regiões e origens linguísticas.

Adicionalmente, o portal disponibiliza um ponto de consulta SPARQL (*SPARQL Protocol and RDF Query Language*), que possibilita a realização de consultas semânticas sobre os dados publicados, facilitando a exploração avançada da informação. A disponibilização de uma API baseada em *Linked Data* reforça ainda mais a interoperabilidade, permitindo que os dados governamentais possam ser facilmente integrados com outros sistemas ou plataformas externas.

Vantagens

O portal espanhol destaca-se pela arquitetura modular e extensível da plataforma CKAN, que proporciona uma elevada flexibilidade na implementação de novas funcionalidades e na personalização do portal às necessidades específicas dos utilizadores institucionais. A existência de uma infraestrutura bem documentada facilita tanto a implementação como a manutenção contínua do sistema, reduzindo o tempo e os custos associados.

Além disso, o elevado grau de maturidade e robustez da plataforma assegura uma reutilização eficaz dos dados, permitindo que os conjuntos de dados sejam utilizados de forma eficiente por uma diversidade de utilizadores.

Desvantagens

O portal pode apresentar algumas limitações, nomeadamente no que respeita à necessidade de desenvolvimento adicional para dar resposta a requisitos específicos de visualização e interação com os dados. Tal situação pode exigir recursos técnicos complementares, sobretudo em contextos que exijam interfaces altamente especializadas ou experiências interativas mais ricas.

Com base na análise efetuada, conclui-se que a arquitetura adotada pelo portal de dados abertos de Espanha constitui a opção mais indicada para a implementação do portal de dados abertos do INE. Esta escolha justifica-se pela robustez, flexibilidade e conformidade com os padrões internacionais de dados abertos que caracterizam o modelo espanhol. A adoção desta abordagem oferece uma base sólida para assegurar a transparência, a acessibilidade e a reutilização dos dados públicos, elementos fundamentais para a consolidação de um ecossistema digital verdadeiramente aberto e interoperável.

Arquitetura e Tecnologia da plataforma do portal da Espanha

Do ponto de vista tecnológico, a arquitetura do portal de dados abertos é composta principalmente por uma ferramenta de gestão de conteúdos, um módulo multilíngue que permite uma navegação simétrica para cada um dos idiomas, um gestor de catálogo de dados, um motor de agregação para integrar todos os catálogos de dados das entidades editoras, uma solução API de dados vinculados e um ponto de consulta SPARQL.

De seguida, na Figura 4.5., apresenta-se o diagrama de arquitetura de alto nível e uma breve descrição de cada uma das camadas, conforme disponibilizado no portal de dados abertos de Espanha.

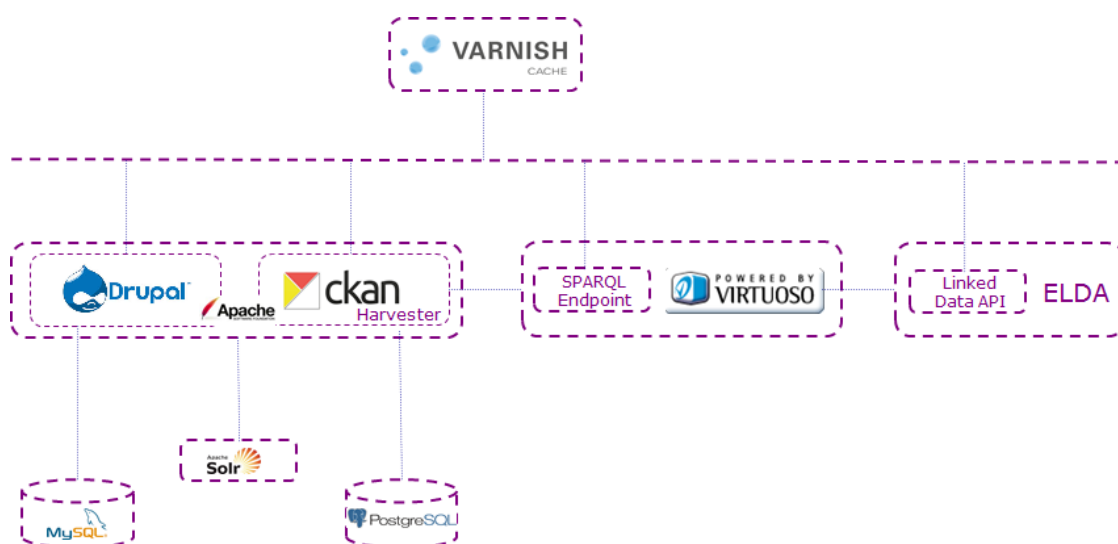


Figura 4.5.: Diagrama de arquitetura

1. Reverse Proxy

O sistema recorre ao *Varnish* como mecanismo de encaminhamento inteligente para os quatro serviços essenciais do projeto, nomeadamente o portal de conteúdos, o catálogo de dados, o ponto de consulta SPARQL e a API de dados vinculados. Esta solução permite um balanceamento eficaz da carga, garantindo um desempenho otimizado e uma experiência de utilização fluida.

2. Portal de Conteúdos

Desenvolvido com base na plataforma de gestão de conteúdos Drupal, este componente desempenha um papel crucial na administração e apresentação do conteúdo editorial do projeto, permitindo uma gestão dinâmica e flexível da informação disponibilizada aos utilizadores.

3. Catálogo de Dados

Baseado na plataforma CKAN, este módulo contém o repositório central de metadados e integra o agregador avançado que permite a sincronização de catálogos provenientes de todas as entidades de publicação. Esta funcionalidade assegura uma maior uniformização e acessibilidade dos dados publicados.

4. Camada de Persistência

A infraestrutura de armazenamento de dados apoia-se em bases de dados robustas, nomeadamente *MySQL* e *PostgreSQL*, garantindo a integridade e a fiabilidade da informação armazenada tanto no portal de conteúdo como no catálogo de dados.

5. Mecanismo de Indexação

Para reforçar a capacidade de pesquisa e recuperação de informação, foi implementado o *Apache Solr*, um motor de indexação altamente eficiente que oferece funcionalidades avançadas de pesquisa tanto para o portal de conteúdo como para o catálogo de dados, melhorando significativamente a experiência dos utilizadores;

6. Ponto de Consulta SPARQL

A solução Virtuoso, da *OpenLink Software*, foi adotada como sistema de armazenamento triplo (*RDF triplestore*), permitindo a modelação e gestão semântica dos dados. Este componente

desempenha um papel central na construção de um gráfico de conhecimento que enriquece o catálogo de dados e possibilita a execução de consultas avançadas em SPARQL.

7. Linked Data API

Desenvolvida com base no *framework ELDA da Epimorphics*, esta API *RESTful* funciona como uma interface de acesso ao ponto de consulta SPARQL. Garante a interoperabilidade dos dados abertos e facilita a sua integração com outras aplicações e serviços externos, promovendo uma reutilização eficaz da informação.

A solução tecnológica adotada inclui o desenvolvimento de mecanismos avançados com o objetivo de responder a dois dos requisitos mais complexos do projeto:

- **Agregador de catálogos**

A funcionalidade de agregação foi concretizada através do desenvolvimento de uma extensão específica para a plataforma CKAN, baseada no mecanismo de agregação *ckanext-harvest*. Esta solução permite a recolha automatizada de dados provenientes de múltiplos catálogos, assegurando a sua integração contínua e eficiente no sistema central.

- **Funcionalidade multilíngue**

Tendo em consideração a necessidade de proporcionar uma experiência de navegação uniforme para utilizadores que acedem ao portal em diferentes idiomas, foi desenvolvida uma solução para Drupal que replica simetricamente a estrutura de navegação em todas as línguas suportadas. Esta abordagem garante a coerência da experiência de utilização e assegura o cumprimento rigoroso dos requisitos de acessibilidade digital de nível AA¹¹.

Disponibilidade do Código-Fonte

No âmbito do compromisso com a transparência e a partilha de conhecimento, o código-fonte do projeto encontra-se integralmente disponível no *GitHub*¹². Todas as extensões desenvolvidas para o CKAN foram publicadas, bem como os módulos de contribuição para o Drupal, incluindo funcionalidades específicas e temas personalizados. Esta iniciativa promove não apenas a reutilização e evolução da solução por outras entidades, como também reforça os princípios de código aberto e a cultura de colaboração no domínio da gestão de dados públicos.

Deste modo, a implementação da arquitetura do portal de dados abertos de Espanha para o desenvolvimento do portal de dados abertos do INE-PT constitui uma decisão estratégica, fundamentada em critérios técnicos e operacionais sólidos. A solução resultante constitui uma plataforma inovadora, eficiente e alinhada com as melhores práticas internacionais no que respeita à disponibilização e reutilização de dados abertos.

¹¹ Os requisitos de acessibilidade digital de nível AA, definidos pelas *Web Content Accessibility Guidelines* (WCAG), visam garantir que conteúdos digitais sejam acessíveis a pessoas com diferentes tipos de deficiência, sendo o nível AA padrão mais comum em contextos institucionais e públicos.

¹² *GitHub* é uma plataforma online de alojamento e colaboração em projetos de software amplamente utilizada para desenvolvimento e partilha de código.

4.2. Catálogo de Dados

Atualmente, o INE utiliza o Modelo 2, referido no capítulo 3.1, designado como “Servidores Separados”, para a gestão do seu catálogo de dados abertos. Esta abordagem baseia-se na separação entre o catálogo de dados e o servidor de armazenamento de ficheiros, permitindo uma gestão mais eficiente e estruturada da infraestrutura tecnológica.

Após uma análise comparativa dos 3 modelos apresentados anteriormente, conclui-se que a manutenção desta estrutura representa a solução mais adequada e vantajosa para o INE, sustentada pelos seguintes fatores fundamentais:

- **Capacidade de expansão e eficiência**

O Modelo 2 permite uma maior flexibilidade na organização e gestão dos dados, garantindo que o catálogo de dados possa ser ampliado sem comprometer a infraestrutura existente. A separação do armazenamento dos ficheiros permite otimizações específicas para diferentes tipos de dados, garantindo melhor desempenho e acessibilidade.

- **Infraestrutura já consolidada**

A manutenção deste modelo justifica-se também pelo facto de o INE já operar sob esta estrutura com sucesso, sem registo de problemas técnicos ou operacionais significativos. A adoção de um novo modelo implicaria custos adicionais, incluindo investimentos em adaptação, migração de dados e formação técnica, sem garantias de benefícios proporcionais que justificassem uma transição potencialmente disruptiva.

- **Maior controlo sobre a informação**

A decisão de manter o segundo modelo está igualmente relacionada com a necessidade de assegurar um controlo rigoroso e contínuo sobre a informação disponibilizada, nomeadamente no que diz respeito à segurança dos dados e à conformidade com diretrizes nacionais e europeias relacionadas com a gestão de dados públicos. Esta abordagem é particularmente relevante para um organismo como o INE, responsável pela produção e divulgação de estatísticas oficiais, onde a integridade, a fiabilidade e a transparência da informação são fundamentais. A estrutura de “Servidores Separados” confere uma camada adicional de segurança, ao permitir um melhor isolamento e proteção dos dados contra acessos não autorizados ou potenciais vulnerabilidades.

- **Compatibilidade com a infraestrutura tecnológica existente**

A continuidade deste modelo evita potenciais dificuldades na integração com diversos sistemas internos e externos, mantendo a compatibilidade com os processos já estabelecidos. Além disso, a separação entre catálogo de dados e armazenamento dos ficheiros confere uma flexibilidade acrescida, permitindo ao INE optar por soluções híbridas que combinem alojamento em serviços internos e serviços em nuvem, conforme as necessidades futuras de crescimento e modernização tecnológica.

- **Eficiência de custos**

A manutenção do Modelo “Servidores Separados” revela-se a alternativa mais económica. A adoção de modelos alternativos, como o Modelo 3, “Catálogos Federados”, exigiria uma descentralização da gestão dos dados para diferentes entidades, o que implicaria investimentos substanciais em infraestruturas, recursos humanos especializados e mecanismos de coordenação mais complexos.

Dado que o sistema atualmente em vigor já se demonstra funcional e eficiente, não se justifica a alocação de recursos adicionais para uma mudança estrutural que, na prática, não proporcionaria ganhos suficientemente expressivos para compensar os custos e riscos envolvidos.

Considerando todos estes aspetos, conclui-se que a manutenção do modelo “Servidores Separados” constitui a estratégia mais coerente para o INE, assegurando um equilíbrio entre eficiência operacional, segurança, compatibilidade tecnológica e sustentabilidade financeira. Assim, recomenda-se a continuidade desta estrutura para a gestão de dados abertos, garantindo a estabilidade da plataforma, a continuidade dos processos e a capacidade de adaptação a eventuais desafios futuros no domínio da difusão de informação estatística oficial.

4.3. Plataforma/Software

A escolha da plataforma para um portal de dados abertos constitui uma decisão estratégica de extrema relevância, influenciando diretamente a acessibilidade, a interoperabilidade e a eficiência na disponibilização de informações públicas. Considerando as opções analisadas em detalhe anteriormente, a escolha recai sobre o CKAN, uma vez que se destaca como a solução mais consolidada e amplamente adotada para a disponibilização de dados abertos, tanto no contexto da União Europeia como em diversos países desenvolvidos fora do bloco europeu, conforme demonstrado na Tabela 3.1.

O CKAN é uma das plataformas mais utilizadas para a gestão e partilha de dados abertos, com implementações bem-sucedidas em diversos portais governamentais como os da Eslováquia, Irlanda e Espanha. Essa ampla adoção confere-lhe um ecossistema sólido e maduro, extensivamente testado e suportado por uma comunidade ativa de desenvolvedores e utilizadores. Em contraste, algumas alternativas como a Udata, embora tecnicamente viáveis, apresentam uma adoção mais restrita, concentrada essencialmente em França, Portugal e em algumas administrações locais, o que poderá limitar a sua aplicabilidade a nível internacional.

Um dos principais fatores determinantes na escolha do CKAN prende-se com o facto de se tratar de uma solução *open-source*, eliminando assim os custos de licenciamento associados a plataformas proprietárias como a Socrata ou a OpenDataSoft. Além disso, o CKAN oferece suporte a APIs RESTful, proporcionando a investigadores, empresas e cidadãos a possibilidade de consumir os dados de forma programática, sem necessidade de interação manual com a interface do portal.

Outro aspeto essencial na escolha do CKAN é a sua compatibilidade com o padrão DCAT-AP, que garante a interoperabilidade com outros portais de dados abertos da União Europeia. A adesão a este padrão facilita significativamente a partilha de metadados entre diferentes plataformas, promovendo a harmonização e normalização dos dados a nível europeu, melhorando, assim, a acessibilidade e reutilização da informação disponibilizada.

Além destas vantagens, o CKAN distingue-se pelo seu conjunto robusto de funcionalidades, concebidas para facilitar a organização, pesquisa e visualização dos dados. Entre estas, destacam-se:

- a organização eficiente da informação por categorias e etiquetas (*tags*);
- mecanismos de pesquisa avançada com filtros configuráveis;
- visualizações interativas que permitem aos utilizadores explorar os dados diretamente na plataforma, sem necessidade de descarregar ficheiros;

- suporte a múltiplos formatos de dados (CSV, JSON, XML, dados geoespaciais), proporcionando uma maior flexibilidade na disponibilização da informação.

Adicionalmente, o CKAN possui uma vasta comunidade de desenvolvedores e utilizadores que contribuem ativamente para a sua evolução e aperfeiçoamento contínuo. Este aspeto assegura não apenas um suporte técnico abrangente e uma documentação extensa, mas também um fluxo constante de melhorias e atualizações. A existência de uma comunidade ativa reduz significativamente o risco de obsolescência tecnológica e facilita a resolução de eventuais problemas técnicos que possam surgir ao longo do tempo.

Apesar da escolha do CKAN, é importante destacar que outras alternativas poderiam ser consideradas, nomeadamente a OpenDataSoft e a Udata. A OpenDataSoft é uma solução que se destaca pelos seus recursos avançados de visualização de dados e *dashboards* interativos, proporcionando uma experiência mais rica e intuitiva para os utilizadores. No entanto, os seus custos elevados podem torná-la uma solução economicamente inviável para o INE, uma vez que implicaria uma dependência direta de fornecedor externo para suporte e evolução da plataforma. Por outro lado, a Udata, embora também seja uma plataforma *open-source*, tem uma adoção significativamente mais limitada, o que poderá representar desafios no que concerne ao suporte técnico e à compatibilidade com padrões internacionais.

Deste modo, a escolha do CKAN para a modernização do portal de dados do INE Portugal baseia-se na sua ampla adoção internacional, na sua natureza *open-source*, na conformidade com padrões internacionais, na sua capacidade de integração simplificada com outros sistemas e nas funcionalidades avançadas que disponibiliza. A sua flexibilidade, escalabilidade e o suporte comunitário contínuo garantem que a plataforma poderá evoluir de forma sustentável, alinhando-se às necessidades do INE e às crescentes exigências da sociedade no que se refere ao acesso à informação pública.

4.4. Modelo de Metadados

Tendo em consideração os modelos de metadados apresentados anteriormente, a escolha recai sobre o DCAT-AP, uma vez que este se apresenta como a solução mais consolidada e amplamente adotada para a descrição de catálogos de dados abertos no contexto da União Europeia. Esta decisão fundamenta-se não apenas na sua robustez e elevada aceitação institucional, mas também na capacidade que este modelo possui de garantir uma interoperabilidade eficaz entre múltiplos portais de dados, promovendo a padronização e a acessibilidade da informação pública.

A adoção do DCAT-AP para o portal do INE assenta, essencialmente, na necessidade de assegurar que os metadados publicados possam ser facilmente integrados e reutilizados tanto a nível nacional como internacional. Desenvolvido no âmbito do programa ISA da Comissão Europeia, este modelo proporciona uma estrutura harmonizada para a descrição de conjuntos de dados, assegurando a sua compatibilidade com padrões europeus e facilitando a troca eficiente de informação entre organismos governamentais, instituições de investigação e a sociedade civil.

A sua ampla implementação é evidenciada pelo facto de, conforme referido no relatório sobre a maturidade dos dados abertos dos países e anteriormente citado, 89% dos Estados-Membros na União Europeia adotarem a sua estrutura ou garantirem a sua interoperabilidade com modelos equivalentes, conferindo-lhe um carácter quase universal no domínio da catalogação de dados governamentais.

Quando comparado com outros modelos de metadados, o DCAT-AP destaca-se pela sua abrangência e especialização na organização de dados governamentais. Enquanto modelos como o *Dublin Core* fornecem um esquema descritivo genérico, adequado para uma ampla gama de recursos digitais, o DCAT-AP foi concebido com um propósito mais direcionado: a gestão, a estrutura e a disseminação de catálogo de dados públicos. Esta estrutura especializada traduz-se num elevado padrão de qualidade da informação, o que, por sua vez, facilita a pesquisa, a recuperação e a reutilização dos dados.

Adicionalmente, um dos aspetos que mais reforça a adoção do DCAT-AP pelo INE Portugal é a sua facilidade de integração com plataformas europeias de dados abertos, nomeadamente o portal data.europa.eu. A interoperabilidade assegurada por este modelo permite que os metadados publicados pelo INE sejam automaticamente indexados e disponibilizados a nível europeu, ampliando significativamente o alcance e o impacto dos dados disponibilizados pela instituição. Esta compatibilidade promove, ainda, um ecossistema de conhecimento mais dinâmico, colaborativo e interconectado.

Outro fator de destaque prende-se com a extensibilidade proporcionada pelo DCAT-AP, permitindo que a sua estrutura base seja adaptada às necessidades específicas do INE sem comprometer a conformidade com padrões internacionais. Esta flexibilidade possibilita a incorporação de elementos adicionais e atributos específicos, ajustando-se de forma eficiente às exigências institucionais e legislativas, enquanto mantém a integridade e a consistência dos metadados no contexto europeu.

A implementação do DCAT-AP no portal do INE fortalece a transparência da instituição, promovendo uma difusão mais eficaz dos dados públicos. A padronização dos metadados não só facilita a interoperabilidade e a reutilização da informação por diversos setores, incluindo a administração pública, as instituições académicas e a indústria, como também impulsiona a inovação e o desenvolvimento de novas soluções baseadas em dados.

A adoção deste modelo reflete, portanto, um compromisso firme com a modernização da gestão da informação e com a consolidação de um ambiente digital mais acessível, inclusivo e colaborativo.

Desta forma, a escolha do DCAT-AP como modelo de metadados para o portal do INE Portugal encontra-se ancorada nos seus benefícios estruturais que proporciona, na sua compatibilidade com padrões europeus e na sua contribuição para a melhoria da acessibilidade, reutilização e visualização dos dados públicos. Esta decisão estratégica assegura que a informação disponibilizada pela instituição esteja plenamente alinhada com os requisitos internacionais, promovendo um ecossistema mais transparente, eficiente e interoperável, em sintonia com as melhores práticas adotadas a nível europeu.

4.5. Processo de Carregamento/ Disponibilização dos Dados

O Instituto Nacional de Estatística tem como missão produzir e disponibilizar informação estatística oficial de qualidade, desempenhando um papel fundamental na disseminação de dados essenciais para a sociedade. O processo de carregamento e disponibilização de dados no INE envolve várias etapas estruturadas, garantindo que a informação seja tratada com rigor e fiabilidade antes da sua difusão.

O Banco de Dados de Difusão constitui o repositório final de informação estatística agregada e estruturada para fins de difusão. Criado com o propósito de unificar dados previamente dispersos, o BDD armazena indicadores estatísticos, bem como a respetiva metainformação, assegurando a sua consistência e qualidade. A sua arquitetura segue um modelo multidimensional que permite uma gestão

eficiente dos dados e uma disseminação otimizada. Os principais fluxos de informação no BDD envolvem a receção de dados estatísticos agregados provenientes dos departamentos de matéria do INE ou do *Data Warehouse* (DW)¹³ e a disponibilização de indicadores estatísticos estruturados e respetiva metainformação para diversos produtos de difusão, incluindo o [Portal das Estatísticas Oficiais](#) e APIs para consulta externa.

A arquitetura do sistema de difusão de dados do INE encontra-se representada na Figura 4.6., evidenciando a forma como os dados são recebidos, organizados e disponibilizados aos utilizadores finais.



Figura 4.6.: Arquitetura do sistema de difusão de dados do INE
(Documento interno – INE)

O carregamento de dados no BDD realiza-se por duas vias principais. A primeira ocorre através de ficheiros XML, seguindo um esquema (XSD) definido pelo INE, sendo o carregamento efetuado por uma plataforma *web* que valida a estrutura dos ficheiros antes da sua inserção no sistema. Para facilitar este processo, é utilizado um *template* em *Excel* (*Package*) que permite gerar automaticamente ficheiros XML. A segunda via consiste na transferência direta de dados a partir do DW, utilizando *scripts* SQL, o que elimina a necessidade de gerar ficheiros XML e agiliza o processo.

Uma vez carregados no BDD, os dados passam por um processo de validação que verifica a metainformação associada e a coerência dos valores face às dimensões de análise definidas. Após a validação, a informação é disponibilizada internamente para revisão e aprovação, antes da sua publicação.

A fase de difusão representa a etapa final do ciclo de produção estatística, permitindo que a informação chegue ao público de forma acessível e organizada. O principal meio de difusão do INE é o Portal das Estatísticas Oficiais, onde os dados são consultados de forma gratuita. Os dados do BDD são disseminados através de diferentes canais, incluindo a exploração através do *BDDEplorer*, a API *APIJsonBDD* para consulta e extração de dados por entidades externas, *webservices* para integração

¹³ *Data Warehouse* (DW) é um sistema centralizado que armazena dados de várias fontes, estruturado para facilitar análises e apoiar a tomada de decisões.

com outros portais como o dados.gov.pt, bem como a criação de anuários estatísticos e outros relatórios institucionais.

Para garantir a integridade dos dados, o BDD opera em dois ambientes distintos: um ambiente interno, utilizado para o carregamento e validação de dados, e um ambiente externo, responsável pela difusão da informação ao público. A transferência de dados entre os ambientes é realizada diariamente através de um processo semi-automatizado, onde os novos indicadores ou atualizações são copiados para o ambiente de difusão, garantindo que a informação publicada esteja sempre atualizada.

Integração da plataforma CKAN

A integração do CKAN no portal do INE deverá ser realizada de forma a otimizar o acesso e a disponibilização dos dados abertos, garantindo compatibilidade com as infraestruturas já existentes, nomeadamente o BDD. Neste contexto, consideram-se duas abordagens possíveis para essa integração.

A primeira consiste em manter o BDD como repositório principal e utilizar o CKAN exclusivamente como uma plataforma de apresentação. Neste caso, o BDD continuaria a ser a ferramenta interna do INE para armazenar e estruturar os dados estatísticos, enquanto o CKAN permitiria a pesquisa, visualização e descarregamento por utilizadores externos. A ligação entre os dois sistemas poderia ser feita através de APIs que permitiriam ao CKAN aceder diretamente aos dados do BDD sem necessidade de replicação, o que garante coerência e integridade da informação, além de minimizar alterações nos processos internos existentes.

A segunda abordagem passaria por migrar a gestão dos dados para o CKAN, com carregamento direto na plataforma, que inclui uma base de dados interna (normalmente *PostgreSQL*), e uma API robusta para a administração dos conjuntos de dados. No entanto, esta solução poderia implicar uma reformulação significativa dos processos internos do INE, nomeadamente no que diz respeito à validação e carregamento de metainformação, podendo implicar um investimento adicional na adaptação das infraestruturas e na formação dos utilizadores.

Tendo em conta os requisitos operacionais do INE, bem como a necessidade de preservar a estabilidade dos sistemas existentes, a primeira abordagem revela-se a mais adequada. Ao manter o BDD como sistema de referência e ao utilizar o CKAN como interface pública, evita-se a redundância e a inconsistência dos dados, maximizando o aproveitamento das infraestruturas consolidadas. Além disso, ao estabelecer uma ligação entre o CKAN e o BDD através de APIs, garante-se que a informação está sempre atualizada, assegurando um equilíbrio entre inovação tecnológica e continuidade operacional.

4.6. Difusão dos Dados

A proposta de solução para a difusão dos dados do INE visa melhorar a acessibilidade, transparência e reutilização da informação estatística, garantindo que os indicadores sejam disponibilizados de forma estruturada, eficiente e alinhada com as necessidades dos diversos perfis de utilizadores. Para tal, propõe-se uma abordagem baseada em boas práticas internacionais e alinhada com as diretrizes europeias para dados abertos, incluindo os HVDs. A adoção destas diretrizes assegurará uma maior interoperabilidade dos dados, promovendo o seu uso em contextos variados e incentivando a inovação através da sua reutilização em diferentes setores da sociedade.

Acesso

Para assegurar um acesso mais inclusivo e global aos dados estatísticos, recomenda-se que o portal do INE disponibilize a informação em, pelo menos, duas línguas: português e inglês. Esta iniciativa não só ampliará o alcance e a usabilidade dos dados por parte de investigadores, empresas e cidadãos estrangeiros, como também reforçará o compromisso do INE com a transparência e a internacionalização da informação estatística.

Adicionalmente, é essencial garantir a acessibilidade para utilizadores com deficiências visuais. Para tal, devem ser incorporadas descrições textuais detalhadas (*tags alt*) nas imagens, gráficos e tabelas, permitindo que softwares de leitura de ecrã interpretem corretamente o conteúdo. Além disso, recomenda-se a implementação de diretrizes rigorosas de acessibilidade digital, garantindo compatibilidade com tecnologias assistidas, como leitores de ecrã avançados e sistemas de conversão para Braille. Desta forma, assegura-se que todos os cidadãos, independentemente das suas limitações visuais, possam aceder e interpretar a informação disponibilizada pelo INE sem barreiras.

Navegação

A navegação no portal deve ser intuitiva, permitindo que os utilizadores encontrem rapidamente os dados necessários. Para isso, propõe-se a continuação da implementação de uma pesquisa avançada que permita a procura por palavra-chave e categoria, bem como a adição de filtros personalizados, como o formato dos documentos e a periodicidade dos dados, tal como disponibilizado no portal de dados abertos de Espanha como se pode observar na Figura 4.7.. O portal espanhol permite filtrar por diferentes critérios que facilitam a pesquisa e agilizam o acesso à informação, pelo que a adoção de uma funcionalidade semelhante no INE-PT contribuiria para uma melhor experiência do utilizador.

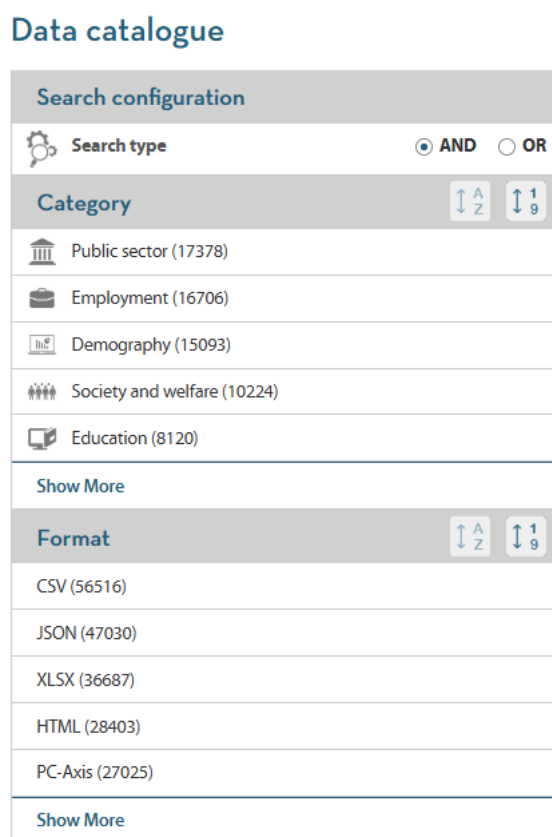


Figura 4.7.: Filtros de pesquisa do portal de Espanha

A organização dos indicadores deve seguir uma lógica estruturada, destacando listas de indicadores mais procurados, categorizando-os por temas e subtemas e garantindo uma estrutura hierárquica intuitiva. Para facilitar a compreensão e reutilização dos dados, sugere-se a criação de uma secção de documentação detalhada, onde serão explicados os conceitos estatísticos e as metodologias utilizadas na produção da informação.

Como se pode observar na Figura 4.8., o portal da Pordata (Portugal - <https://www.pordata.pt>) é um exemplo de uma boa organização de indicadores por categorias, permitindo a navegação estruturada por temas e a visualização de tendências estatísticas ao longo do tempo, facilitando a interpretação e a análise comparativa dos dados.

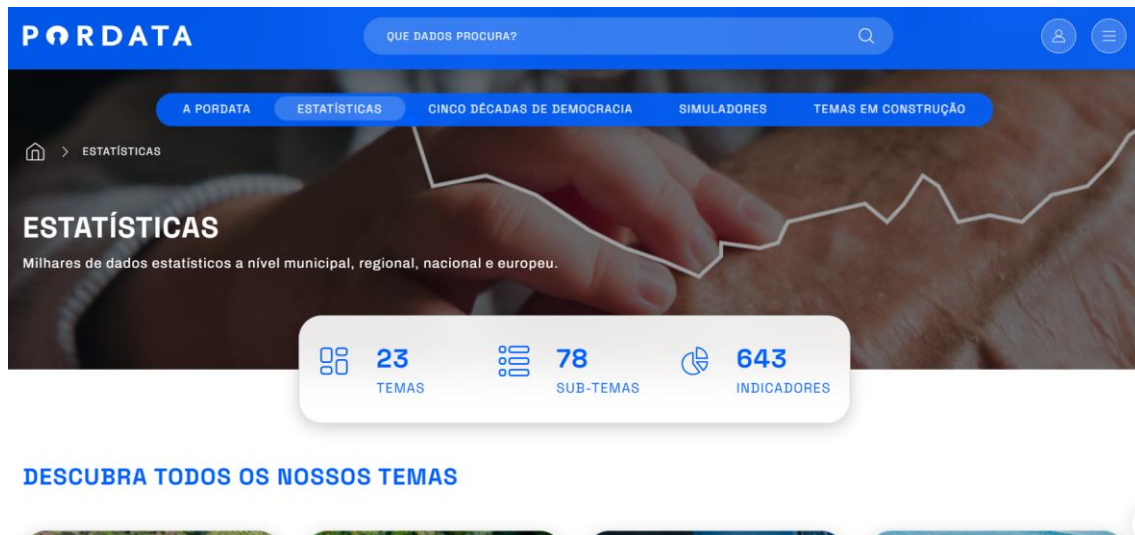


Figura 4.8.: Indicadores do portal do Pordata

Relativamente aos HVDs, é essencial destacar a sua importância no contexto da disponibilização dos dados abertos, uma vez que representam conjuntos de informação de elevado valor económico, social e ambiental. Como podemos observar na Figura 4.9., o INE já disponibiliza uma secção específica para estes dados, o que constitui uma boa prática que deve ser mantida e continuamente aprimorada para garantir conformidade com a [Diretiva \(UE\) 2019/1024](#). A relevância dos HVDs reside na sua capacidade de fomentar a inovação, facilitar a tomada de decisão baseada em dados e impulsionar o desenvolvimento de novas aplicações e serviços digitais.



Figura 4.9.: Secção dos HVDs no portal do INE

No portal de Espanha, os HVDs estão claramente identificados e categorizados, permitindo um acesso simplificado por parte dos utilizadores. De forma semelhante, o portal francês adota uma abordagem estruturada, garantindo que esses conjuntos de dados sejam destacados e facilmente acessíveis. A experiência do portal da Irlanda também demonstra boas práticas na documentação e organização dos HVDs, assegurando que os utilizadores encontrem a informação de forma eficiente.

Assim, recomenda-se que o INE continue a evidenciar os HVDs de forma clara e acessível, melhorando a usabilidade da secção dedicada e assegurando a implementação de um sistema de atualização automática. Esta medida permitirá que os utilizadores tenham sempre acesso às versões mais recentes dos dados, promovendo uma maior transparência, interoperabilidade e reutilização da informação estatística para fins académicos, empresariais e institucionais.

Visualização

A forma como os dados são apresentados influencia diretamente a facilidade de compreensão e análise por parte dos utilizadores. Recomenda-se que o INE disponibilize os resultados da consulta dos indicadores em diferentes formatos, incluindo tabelas interativas, gráficos dinâmicos e mapas com georreferenciação. A utilização de mapas permite uma análise espacial detalhada e facilita a interpretação da informação estatística em contexto territorial.

Plataformas como OpenDataSoft e CKAN já oferecem soluções para a organização e visualização de dados abertos, permitindo uma apresentação mais interativa e dinâmica dos indicadores estatísticos. A adoção de ferramentas semelhantes no portal do INE poderia melhorar significativamente a experiência do utilizador.

Metadados

Os metadados associados aos HVDs devem ser estruturados segundo o padrão DCAT-AP, um modelo europeu que garante que a informação sobre os conjuntos de dados seja descrita de forma padronizada. Esta estrutura permite que os dados do INE sejam facilmente encontrados, compreendidos e reutilizados por outras plataformas nacionais e internacionais, assegurando a interoperabilidade entre diferentes repositórios de dados na União Europeia. Além disso, o acesso a estes dados deve ser otimizado, permitindo *downloads* rápidos em vários formatos e oferecendo integração via API, para que os utilizadores, como investigadores e programadores, possam automatizar a obtenção e o uso dos dados. Para melhorar a gestão e a adaptação da oferta de dados, recomenda-se ainda a implementação de

ferramentas de monitorização de acessos. Estas ferramentas permitiriam ao INE analisar como e com que frequência os HVD são utilizados, possibilitando ajustes estratégicos na disponibilização dos dados, de acordo com as necessidades dos utilizadores.

Exportação

A disponibilização dos dados para descarregamento deve ser flexível e adaptada às necessidades dos diferentes utilizadores. Os dados devem estar disponíveis em múltiplos formatos abertos, incluindo CSV, JSON, XML, XLSX e TSV, garantindo compatibilidade com diversas ferramentas de análise e integração com sistemas externos. Além disso, recomenda-se a implementação de um sistema de descarregamento em lote, que permita aos utilizadores selecionar múltiplos conjuntos de dados e descarregá-los de uma só vez. O Portal de Espanha inclui essa funcionalidade, o que facilita a obtenção de grandes volumes de informação sem necessidade de múltiplos *downloads* individuais.

Outro aspeto relevante é a disponibilidade de APIs bem documentadas, permitindo que os utilizadores acedam automaticamente aos dados e aos metadados. No portal de Irlanda, a API é destacada como uma ferramenta essencial para a automação do acesso aos dados. A melhoria da documentação da API no INE e a inclusão de exemplos práticos de utilização ajudariam a fomentar a adoção desta ferramenta pelos utilizadores.

A implementação destas medidas permitirá ao INE não apenas modernizar a sua estratégia de difusão de dados, mas também consolidar-se como uma entidade de referência na disponibilização de informação estatística de qualidade. Ao assegurar um acesso mais eficiente, interativo e acessível à informação estatística, o INE promove uma maior transparência e democratização da informação, fomentando a sua reutilização por diversos setores da sociedade, desde investigadores e académicos até empresas e cidadãos em geral.

Além disso, ao alinhar-se com as melhores práticas internacionais e diretrizes europeias para dados abertos, o INE reforçará a sua posição no panorama global da governação de dados, contribuindo para a interoperabilidade e integração dos seus repositórios com outras plataformas nacionais e internacionais. Esta abordagem permitirá não só potenciar o impacto social e económico da informação estatística, mas também estimular a inovação e o desenvolvimento de novas soluções baseadas em dados, impulsionando a economia digital e o conhecimento científico.

Deste modo, ao investir continuamente na acessibilidade, usabilidade e atualização da informação estatística garantirá que o INE se mantenha na vanguarda da disponibilização de dados abertos, promovendo um ecossistema informacional mais dinâmico, transparente e centrado nas exigências da sociedade moderna.

5. Conclusão

Este trabalho teve como propósito principal o desenvolvimento de uma proposta concreta sólida, viável e tecnicamente fundamentada para a criação de uma área dedicada à disponibilização de dados abertos no portal do Instituto Nacional de Estatística (INE). Partindo do contexto europeu, onde a [Diretiva \(UE\) 2019/1024](#) define orientações claras para a reutilização da informação do setor público, foi possível identificar a necessidade e a oportunidade de modernização da estratégia de difusão estatística do INE.

Ao longo do trabalho, foi possível identificar, com base em literatura especializada e estudos recentes sobre maturidade digital, a necessidade de modernizar e diversificar os mecanismos de acesso aos dados estatísticos disponibilizados pelo INE. O diagnóstico realizado demonstrou que, apesar dos avanços já alcançados, persistem limitações no que respeita à acessibilidade plena, à reutilização automatizada da informação e à integração com o ecossistema nacional e europeu de dados abertos.

Com o intuito de superar essas limitações, procedeu-se à análise crítica de boas práticas adotadas por países com elevada maturidade digital, como França, Irlanda e Espanha, o que permitiu compreender diferentes abordagens técnicas e organizacionais na implementação de portais de dados abertos. Com base nessa análise, foi delineada uma solução técnica para o INE, baseada na plataforma CKAN, articulada com o Banco de Dados de Difusão (BDD) e alinhada com o modelo de metadados DCAT-AP.

Recomenda-se que a futura implementação desta proposta tenha em consideração não só os aspetos tecnológicos, como também a capacitação interna das equipas, a atualização contínua dos conjuntos de dados e a promoção ativa da sua reutilização por parte de cidadãos, investigadores e empresas. A interoperabilidade e a automatização devem ser prioridades, tal como a integração com o ecossistema nacional e europeu de dados abertos.

A principal contribuição deste trabalho reside na definição de uma solução viável, alinhada com padrões europeus, que poderá posicionar o INE como referência nacional no domínio da transparência, da ciência de dados e da inovação pública baseada em evidência.

Em síntese, a abertura dos dados estatísticos representa hoje uma oportunidade estratégica para transformar informação pública em valor económico, social e científico. Mais do que uma obrigação legal, trata-se de um passo decisivo rumo a uma cultura de dados mais madura, participativa e orientada para o bem comum. O sucesso desta transição dependerá do compromisso institucional e da capacidade de criar um ambiente de dados verdadeiramente acessível, confiável e útil.

6. Bibliografia

- *European Union*. Portal Europeu de Dados.
URL: <https://data.europa.eu/pt>
- Instituto Nacional de Estatística. Portal do INE.
URL: <https://www.ine.pt>
- Agência para a Modernização Administrativa. Dados.gov.pt – Portal de dados abertos da Administração Pública.
URL: <https://dados.gov.pt>
- *European Commission*. *Open Data Maturity Report 2024*.
URL: <https://data.europa.eu/en/publications/open-data-maturity/2024>
- *World Bank Group*. *Open Data Toolkit. Technology Options*.
URL: <https://opendatatoolkit.worldbank.org/en/data/opendatatoolkit/technology#catalog-look-like>
- *Dublin Core*. *About the Dublin Core Metadata Initiative or “BCMI”*.
URL: <https://www.dublincore.org/about/>
- *Digital Curation Centre*. *DCAT – Data Catalog Vocabulary*.
URL: <https://www.dcc.ac.uk/resources/metadata-standards/dcat-data-catalog-vocabulary>
- *Publications Office of the European Union*. *DCAT Application profile for data in Europe (DCAT-AP)*
URL: <https://op.europa.eu/pt/web/eu-vocabularies/dcat-ap>
- Udata. Soluções Udata.
URL: <https://udata.ca/>
- Iniciativa de dados abertos do Governo da Espanha. Portal de Espanha
URL: <https://datos.gob.es/en>
- Governo da Irlanda. Portal de dados abertos da Irlanda
URL: <https://data.gov.ie/>
- Governo de França. Portal de dados abertos de França
URL: <https://www.data.gouv.fr/>
- PORDATA. Base de Dados de Portugal Contemporâneo
URL: <https://www.pordata.pt>