

O Catálogo de Dados Corporativo

Aprimore a descoberta de dados, garanta a governança de dados e impulsione a inovação

Ole Olesen-Bagneux

Authorized Portuguese translation of the English edition of **The Enterprise Data Catalog** ISBN 9781492098713 © 2023 Ole Olesen-Bagneux. This translation is published and sold by permission of O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

Tradução em português autorizada da edição em inglês da obra **The Enterprise Data Catalog** ISBN 9781492098713 © 2023 Ole Olesen-Bagneux. Esta tradução é publicada e vendida com a permissão da O'Reilly Media, Inc., detentora de todos os direitos para publicação e venda desta obra.

© Novatec Editora Ltda. [2023].

Todos os direitos reservados e protegidos pela Lei 9.610 de 19/02/1998. É proibida a reprodução desta obra, mesmo parcial, por qualquer processo, sem prévia autorização, por escrito, do autor e da Editora.

Editor: Rubens Prates

Tradução: Márcio Martins

Revisão gramatical: Lia Gabriele Regius

ISBN do impresso: 978-85-7522-864-7

ISBN do ebook: 978-85-7522-865-4

Histórico de impressões:

Setembro/2023 Primeira edição

Novatec Editora Ltda.

Rua Luís Antônio dos Santos 110

02460-000 – São Paulo, SP – Brasil

Tel.: +55 11 2959-6529

Email: novatec@novatec.com.br

Site: <https://novatec.com.br>

Twitter: twitter.com/novateceditora

Facebook: facebook.com/novatec

LinkedIn: linkedin.com/in/novatec

BAR20230905

Sumário

Apresentação	11
Prefácio	15
Parte I = Como organizar dados para que você possa pesquisá-los	23
Capítulo 1 = Introdução a catálogos de dados	24
A funcionalidade principal de um catálogo de dados	24
Criar uma visão geral do cenário de TI	25
Organize os dados	27
Possibilite a pesquisa dos dados da empresa	31
Descoberta de dados	35
A equipe de descoberta de dados	37
Arquitetos de dados	37
Engenheiros de dados	40
Configuração da equipe de descoberta de dados	41
Funções e responsabilidades dos usuários finais	42
Resumo	44
Capítulo 2 = Organização de dados: como desenvolver uma arquitetura robusta para realizar pesquisas	46
Organização de domínios no catálogo de dados	46
Arquitetura de domínio em um catálogo de dados	47
Compreensão dos domínios	48
Processos e capacidades	52
Fontes de dados	56
Inserção de ativos no catálogo de dados	58
Pull	59
Push	60
Organização de ativos nos domínios	61
Metadados do ativo	61
Qualidade dos metadados	68
Classificação	71
Resumo	75

Capítulo 3 = Compreensão da pesquisa: conceitos, recursos e mecânica76

Por que realizar pesquisas em um catálogo de dados?.....	76
Recursos de pesquisa em um catálogo de dados.....	78
Diferença entre pesquisar nos dados e pesquisar por dados.....	79
Como pesquisar em um catálogo de dados?.....	83
Linguagem de consulta de catálogo de dados.....	84
Explicação dos recursos de pesquisa em um catálogo de dados.....	85
Pesquisar por tudo?.....	95
A mecânica da pesquisa.....	96
Recall e precisão.....	96
Lei de Zipf.....	100
Serendipidade.....	101
Resumo.....	103

Capítulo 4 = Aplicação da pesquisa: de padrões simples a avançados..... 105

Pesquisa como um bibliotecário – não como um cientista de dados.....	106
Padrões de pesquisa.....	108
Pesquisa simples básica.....	108
Pesquisa simples detalhada.....	110
Pesquisa simples flexível.....	111
Pesquisa por intervalo.....	113
Pesquisa por blocos.....	114
Pesquisa por instrução.....	117
Padrões de navegação.....	118
Navegação por glossários.....	118
Navegação por domínios.....	119
Navegação por linhagem.....	119
Navegação por grafos.....	120
Pesquisa em um catálogo de dados baseado em grafos.....	122
Resumo.....	123

Parte II = Democratização de dados com um catálogo de dados.....125**Capítulo 5 = Descoberta de dados: como capacitar usuários finais e envolver stakeholders 126**

Um catálogo de dados é uma rede social.....	126
Metadados ativos.....	128
Garanta o engajamento dos stakeholders.....	131
Envolve os líderes de governança de dados.....	131
Envolve os líderes de análise de dados.....	135
Envolve líderes de domínio.....	137
Visualização de todos os dados através de uma única lente.....	139
O backbone operacional e a plataforma de dados.....	139
Resumo.....	142

Capítulo 6 = Acesso a dados: fatores-chave para uma implementação bem-sucedida 143

Escolha de um catálogo de dados	143
Análise dos fornecedores	143
Alguns fornecedores importantes	145
Catálogo de catálogos	147
Como acessar os dados	149
Provedores de dados e consumidores de dados	149
Abordagem centralizada	151
Abordagem descentralizada	153
Abordagem híbrida	156
Construção de domínios	158
Questionário nº 1: Descrição do domínio e ativos pelo proprietário de domínio	158
Questionário nº 2: Descrição dos ativos no domínio pelo administrador de ativos	159
Questionário nº 3: Descrição dos termos do glossário dos ativos pelo administrador de ativos	159
Resumo	160

Capítulo 7 = Gerenciamento de dados: como melhorar o gerenciamento do ciclo de vida 162

O valor do gerenciamento do ciclo de vida dos dados e por que o catálogo de dados é um divisor de águas	162
Diversos ciclos de vida	165
Ciclo de vida dos dados	165
Uso do catálogo de dados para gerenciamento do ciclo de vida dos dados	167
O ciclo de vida dos ativos de dados no catálogo de dados	168
Ciclo de vida dos termos de glossário	174
Ciclo de vida das fontes de dados	177
Influência e suporte dos ciclos de vida	178
Pesquisa aplicada com base em ciclos de vida	179
Pesquisa aplicada para conformidade regulatória	180
Melhores práticas de manutenção	182
Manutenção dos dados fora do catálogo de dados	183
Manutenção de metadados dentro do catálogo de dados	184
Gerenciamento aprimorado do ciclo de vida dos dados	186
Resumo	187

Parte III = Futuro dos catálogos de dados 189**Capítulo 8 = Futuro: o mecanismo de pesquisa corporativo e a melhoria do gerenciamento de
dados 190**

O mecanismo de pesquisa corporativo	190
O mecanismo de pesquisa corporativo na Hugin & Munin	192
Dos dados ao conhecimento	197

Uma reflexão da Teoria do Meio sobre o mecanismo de pesquisa corporativo	201
O mecanismo de pesquisa corporativo é uma nova tecnologia?	205
O mecanismo de pesquisa corporativo se tornará realidade?	206
Resumo	207
Posfácio ■ Considere a implementação de um catálogo de dados	212
Apêndice ■ Linguagem de consulta de catálogo de dados	214
Índice remissivo	220