CAPÍTULO 2

Ética no tratamento de dados

1. Introdução

Definido de forma simples, *ética* são princípios de comportamento baseados em ideias de certo e errado. Princípios éticos geralmente focam em ideias como justiça, respeito, responsabilidade, integridade, qualidade, confiabilidade, transparência e confiança. A ética de manipulação de dados está relacionada a como obter, armazenar, gerenciar, usar e descartar dados de maneiras que estejam alinhadas com princípios éticos. Manipular dados de forma ética é necessário para o sucesso a longo prazo de qualquer organização que queira obter valor de seus dados. O manuseio antiético de dados pode resultar na perda de reputação e clientes, porque coloca em risco pessoas cujos dados são expostos. Em alguns casos, práticas antiéticas também são ilegais.18 Em última análise, para profissionais de gerenciamento de dados e as organizações para as quais trabalham, a ética de dados é uma questão de responsabilidade social.

A ética do tratamento de dados é complexa, mas se concentra em vários conceitos principais:

- **Impacto nas pessoas:** como os dados representam características de indivíduos e são usados para tomar decisões que afetam a vida das pessoas, é fundamental gerenciar sua qualidade e confiabilidade.
- **Potencial de uso indevido: o** uso indevido de dados pode afetar negativamente pessoas e organizações, portanto, há um imperativo ético para evitar o uso indevido de dados.
- **Valor econômico dos dados:** Dados têm valor econômico. A ética da propriedade de dados deve determinar como esse valor pode ser acessado e por quem.

As organizações protegem dados com base, em grande parte, em leis e requisitos regulatórios. No entanto, como os dados representam pessoas (clientes, funcionários, pacientes, fornecedores, etc.), os profissionais de gerenciamento de dados devem reconhecer que há razões éticas (bem como legais) para proteger os dados e garantir que eles não sejam mal utilizados. Mesmo dados que não representam diretamente indivíduos ainda podem ser usados para tomar decisões que afetam a vida das pessoas.

Há um imperativo ético não apenas para proteger os dados, mas também para gerenciar sua qualidade. As pessoas que tomam decisões, bem como aquelas impactadas pelas decisões, esperam que os dados sejam completos e precisos. Tanto de uma perspectiva comercial quanto técnica, os profissionais de gerenciamento de dados têm a responsabilidade ética de gerenciar os dados de uma forma que reduza o risco de que eles possam ser deturpados, mal utilizados ou mal compreendidos. Essa responsabilidade se estende por todo o ciclo de vida dos dados, da criação à destruição dos dados.

Figura 12 Diagrama de Contexto: Ética no tratamento de dados

Data Handling Ethics

Definition: Data handling ethics are concerned with how to procure, store, manage, interpret, analyze / apply and dispose of data in ways that are aligned with ethical principles, including community responsibility.

Goals:

1. To define ethical handling of data in the organization.
2. To educate staff on the organization risks of improper data handling.
3. To change/instill preferred culture and behaviors on handling data.
4. To monitor regulatory environment, measure, monitor, and adjust organization approaches for ethics in data.

Business Drivers

Inputs:

- Existing and Preferred Organization Ethics
- Business Strategy & Goals
- Organizational Structure
- Business Culture
- Regulations
- Existing Corporate Policies

Activities:

1. Review Data-Handling Practices (P)
2. Identify Principles, Practices, and Risk Factors (P)
3. Create and Ethical Data Handling Strategy
4. Address Practices Gaps (D)
5. Communicate and Educate Staff (D)
6. Monitor and Maintain Alignment (C)

Deliverables:

- Current Practices and Gaps
- Ethical Data Handling Strategy
- Communication Plan
- Ethics Training Program
- Ethical Corporate Statements on Data
- Awareness to Ethical Data Issues
- Aligned Incentives, KPIs, and Targets
- Updated Policies
- Ethical Data Handling Reporting

Suppliers:

- Executives
- Data Stewards
- Executive Data

Participants:

- Data Governance Bodies
- CDO / CIO
- Executives

Consumers:

- Employees
- Executives
- Regulators

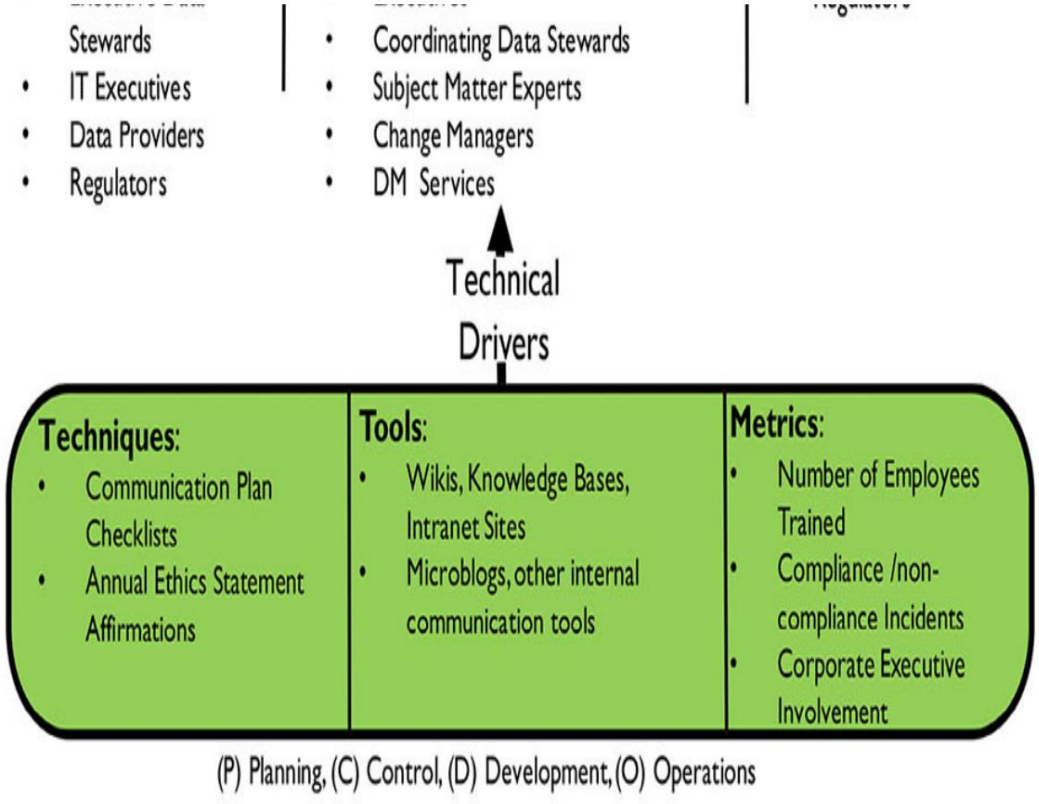

Figure 12 Context Diagram: Data Handling Ethics

Infelizmente, muitas organizações falham em reconhecer e responder às obrigações éticas inerentes ao gerenciamento de dados. Elas podem adotar uma perspectiva técnica tradicional e professar não entender os dados; ou elas assumem que se seguirem a letra da lei, não terão risco relacionado ao manuseio de dados. Esta é uma suposição perigosa.

O ambiente de dados está evoluindo rapidamente. As organizações estão usando dados de maneiras que não imaginariam até mesmo alguns anos atrás.
Embora as leis codifiquem alguns princípios éticos, a legislação não consegue acompanhar os riscos associados à evolução do ambiente de dados.

As organizações devem reconhecer e responder à sua obrigação ética de proteger os dados a elas confiados, promovendo e sustentando uma cultura que valoriza o tratamento ético das informações.

2. Motoristas de negócios

Como as declarações de W. Edward Deming sobre qualidade, ética significa “fazer

certo quando ninguém está olhando.” Uma abordagem ética ao uso de dados está sendo cada vez mais reconhecida como uma vantagem competitiva de negócios (Hasselbalch e Tranberg, 2016). O manuseio ético de dados pode aumentar a confiabilidade de uma organização e os resultados de dados e processos da organização. Isso pode criar melhores relacionamentos entre a organização e suas partes interessadas. Criar uma cultura ética envolve introduzir governança adequada, incluindo a instituição de controles para garantir que os resultados pretendidos e resultantes do processamento de dados sejam éticos e não violem a confiança ou infrinjam a dignidade humana.

O manuseio de dados não acontece no vácuo, e clientes e stakeholders esperam comportamento e resultados éticos de empresas e seus processos de dados. Reduzir o risco de que dados pelos quais a organização é responsável sejam mal utilizados por funcionários, clientes ou parceiros é uma razão primária para uma organização cultivar princípios éticos para o manuseio de dados. Há também uma responsabilidade ética para proteger dados de criminosos (ou seja, para proteger contra hackers e potenciais violações de dados. (Consulte o Capítulo 7.)

Diferentes modelos de propriedade de dados influenciam a ética do manuseio de dados. Por exemplo, a tecnologia melhorou a capacidade das organizações de compartilhar dados entre si. Essa capacidade significa que as organizações precisam tomar decisões éticas sobre sua responsabilidade de compartilhar dados que não pertencem a elas.

As funções emergentes de Diretor de Dados, Diretor de Riscos, Diretor de Privacidade e Diretor de Análise estão focadas no controle de riscos por meio do estabelecimento de práticas aceitáveis para o tratamento de dados.
Mas a responsabilidade se estende além das pessoas nessas funções. Lidar com dados eticamente requer o reconhecimento de toda a organização dos riscos associados ao uso indevido de dados e o comprometimento organizacional em lidar com dados com base em princípios que protegem indivíduos e respeitam os imperativos relacionados à propriedade de dados.

3. Conceitos Essenciais

3.1 Princípios éticos para dados

Os princípios aceitos da bioética, que focam na preservação da dignidade humana, fornecem um bom ponto de partida geral para princípios de ética de dados. Por exemplo, os Princípios de Belmont para pesquisa médica podem ser adaptados em disciplinas de Gerenciamento de Informação (US-HSS, 1979).

- **Respeito pelas Pessoas:** Este princípio reflete o requisito ético fundamental de que as pessoas sejam tratadas de uma forma que respeite sua dignidade e autonomia como indivíduos humanos. Ele também requer que, em casos em que as pessoas tenham "autonomia diminuída", cuidados extras sejam tomados para proteger sua dignidade e direitos.

Quando consideramos dados como um ativo, temos em mente que os dados também afetam, representam ou tocam as pessoas? Dados pessoais são diferentes de outros "ativos" brutos, como petróleo ou carvão. O uso antiético de dados pessoais pode influenciar diretamente as interações das pessoas, oportunidades de emprego e lugar na comunidade. Projetamos sistemas de informação de uma forma que limita a autonomia ou a liberdade de escolha? Consideramos como o processamento de dados pode afetar pessoas com deficiências mentais ou físicas? Levamos em conta como elas acessarão e utilizarão os dados? O processamento de dados ocorre com base em consentimento informado e válido?

- **Beneficência:** Este princípio tem dois elementos: primeiro, não causar danos; segundo, maximizar os possíveis benefícios e minimizar os possíveis danos.

O princípio ético de "não causar dano" tem uma longa história na ética médica, mas também tem uma aplicação clara no contexto de gerenciamento de dados e informações. Os profissionais de dados e informações éticos devem identificar as partes interessadas e considerar os resultados do processamento de dados e trabalhar para maximizar os benefícios e minimizar o risco de danos causados pelos processos projetados. Um processo é projetado de uma forma que

assume um resultado de soma zero em vez de uma situação ganha-ganha? O processamento de dados é desnecessariamente invasivo e há uma maneira menos arriscada de atender aos requisitos da necessidade do negócio? O tratamento de dados em questão carece de transparência de uma forma que pode esconder possíveis danos às pessoas?

- **Justiça:** Este princípio considera o tratamento justo e equitativo das pessoas.

Algumas perguntas que podem ser feitas sobre esse princípio: Pessoas ou grupos de pessoas estão sendo tratados de forma desigual em circunstâncias semelhantes? O resultado de um processo ou algoritmo resulta em efeitos que beneficiam ou prejudicam desproporcionalmente um determinado grupo de pessoas? O aprendizado de máquina está sendo treinado usando conjuntos de dados que contêm dados inadvertidamente reforçando preconceitos culturais?

O Relatório Menlo do Departamento de Segurança Interna dos Estados Unidos adapta os Princípios Belmont à Pesquisa em Tecnologia da Informação e Comunicação, adicionando um quarto princípio: Respeito à Lei e ao Interesse Público (US-DHS, 2012).

Em 2015, o Supervisor Europeu de Proteção de Dados publicou um parecer sobre ética digital destacando as “implicações de engenharia, filosóficas, legais e morais” dos desenvolvimentos no processamento de dados e Big Data. Ele pediu um foco no processamento de dados que defenda a dignidade humana e estabeleceu quatro pilares necessários para um ecossistema de informações que garanta o tratamento ético dos dados (EDPS, 2015):

- Regulamentação orientada para o futuro do tratamento de dados e respeito pelos direitos à privacidade e à protecção de dados
- Controladores responsáveis que determinam o processamento de informações pessoais
- Engenharia e design conscientes da privacidade de produtos e serviços de processamento de dados
- Indivíduos empoderados

Esses princípios mapeiam amplamente o princípio estabelecido no Relatório Belmont, com foco na promoção da dignidade humana e da autonomia. O EDPS afirma que a privacidade é um direito humano fundamental. Ele desafia os inovadores a ver a dignidade, a privacidade e a autonomia como uma plataforma na qual um ambiente digital sustentável é moldado, em vez de um obstáculo ao desenvolvimento, e exige transparência e comunicação com as partes interessadas.

A Governança de Dados é uma ferramenta vital para garantir que esses princípios sejam considerados ao decidir quem pode fazer o quê com quais dados e sob quais circunstâncias o processamento é apropriado ou necessário. Os impactos éticos e os riscos do processamento de dados em todas as partes interessadas devem ser considerados pelos profissionais e gerenciados de maneira semelhante à qualidade dos dados.

3.2 Princípios por trás da lei de privacidade de dados A

política pública e a lei tentam codificar o certo e o errado com base em princípios éticos. Mas elas não podem codificar todas as circunstâncias. Por exemplo, as leis de privacidade na União Europeia, Canadá e Estados Unidos mostram diferentes abordagens para codificar a ética de dados. Esses princípios também podem fornecer uma estrutura para a política organizacional.

A lei de privacidade não é nova. Privacidade e privacidade de informações como conceitos estão firmemente ligados ao imperativo ético de respeitar os direitos humanos. Em 1890, os acadêmicos jurídicos americanos Samuel Warren e Louis Brandeis descreveram a privacidade e a privacidade de informações como direitos humanos com proteções na common law que sustentam vários direitos na constituição dos EUA. Em 1973, um código de Fair Information Practice foi proposto, e o conceito de privacidade de informações como um direito fundamental foi reafirmado no US Privacy Act de 1974, que afirma que “o direito à privacidade é um direito pessoal e fundamental protegido pela Constituição dos Estados Unidos”.

Na sequência das violações dos direitos humanos durante a Segunda Guerra Mundial, a Convenção Europeia dos Direitos Humanos (1950) estabeleceu tanto o direito geral à privacidade como o direito específico à privacidade da informação (ou o direito à protecção dos dados pessoais) como direitos humanos fundamentais para defender o direito à liberdade de expressão.

Dignidade. Em 1980, a Organização para Cooperação e Desenvolvimento Econômico (OCDE) estabeleceu Diretrizes e Princípios para Processamento Justo de Informações que se tornaram a base para as leis de proteção de dados da União Europeia.

Os oito princípios fundamentais da OCDE, os Fair Information Processing Standards, têm como objetivo garantir que os dados pessoais sejam processados de uma maneira que respeite o direito dos indivíduos à privacidade. Eles incluem: limitações na coleta de dados; uma expectativa de que os dados sejam de alta qualidade; o requisito de que, quando os dados são coletados, isso seja feito para um propósito específico; limitações no uso de dados; salvaguardas de segurança; uma expectativa de abertura e transparência; o direito de um indivíduo de contestar a precisão dos dados relacionados a si mesmo; e responsabilidade das organizações para seguir as diretrizes.

Os princípios da OCDE foram substituídos por princípios subjacentes ao Regulamento Geral de Proteção de Dados da UE (GDPR, 2016). Veja a Tabela 1.

Tabela 1 Princípios do RGPD

Table 1 GDPR Principles

GDPR Principle	Description of Principle
Fairness, Lawfulness, Transparency	Personal data shall be processed lawfully, fairly, and in a transparent manner in relation to the data subject.
Purpose Limitation	Personal data must be collected for specified, explicit, and legitimate purposes, and not processed in a manner that is incompatible with those purposes.
Data Minimization	Personal data must be adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed.
Accuracy	Personal data must be accurate, and where necessary, kept up-to-date. Every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purpose for which they are processed, are erased or rectified without delay.
Storage Limitation	Data must be kept in a form that permits identification of data subjects [individuals] for no longer than is necessary for the purposes for which the personal data are processed.
Integrity and Confidentiality	Data must be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures.
Accountability	Data Controllers shall be responsible for, and be able to demonstrate compliance with [these principles].

Esses princípios são equilibrados e dão suporte a certos direitos qualificados que os indivíduos têm sobre seus dados, incluindo os direitos de acesso, retificação de dados imprecisos, portabilidade, o direito de se opor ao processamento de dados pessoais que podem causar danos ou sofrimento e apagamento. Quando o processamento de dados pessoais é feito com base no consentimento, esse consentimento deve ser uma ação afirmativa que seja dada livremente, específica, informada e inequívoca. O GDPR exige governança e documentação eficazes para permitir e demonstr

conformidade e mandatos Privacidade desde o Design.

A lei de privacidade canadense combina um regime abrangente de proteção de privacidade com a autorregulamentação da indústria. O PIPEDA (Personal Information Protection and Electronic Documents Act) se aplica a todas as organizações que coletam, usam e disseminam informações pessoais no curso de atividades comerciais. Ele estipula regras, com exceções, que as organizações devem seguir no uso das informações pessoais dos consumidores. A Tabela 2 descreve as obrigações estatutárias com base no PIPEDA.19 No Canadá, o Federal Privacy Commissioner tem a responsabilidade exclusiva de ---

lidar com reclamações de privacidade contra organizações.

No entanto, eles desempenham um papel de ombudsman; suas decisões são apenas recomendações (não juridicamente vinculativas e sem valor precedente, mesmo dentro do gabinete do comissário).

Tabela 2 Privacidade Canadense Obrigações estatutárias

Table 2 Canadian Privacy Statutory Obligations

PIPEDA Principle	Description of Principle
Accountability	An organization is responsible for personal information under its control and must designate an individual to be accountable for the organization's compliance with the principle.
Identifying Purposes	An organization must identify the purposes for which personal information is collected at or before the time the information is collected.
Consent	An organization must obtain the knowledge and consent of the individual for the collection, use, or disclosure of personal information, except where inappropriate.
Limiting Collection, Use, Disclosure, and Retention	The collection of personal information must be limited to that which is necessary for the purposes identified by the organization. Information shall be collected by fair and lawful means. Personal information shall not be used or disclosed for purposes other than those for which it was collected, except with the consent of the individual or as required by law. Personal information shall be retained only as long as necessary for the fulfillment of those purposes.
Accuracy	Personal information must be as accurate, complete, and up-to-date as is necessary for the purposes for which it is to be used.
Safeguards	Personal information must be protected by security safeguards appropriate to the sensitivity of the information.
Openness	An organization must make specific information about its policies and practices relating to the management of their personal information readily available to individuals.
Individual Access	Upon request, an individual shall be informed of the existence, use, and disclosure of his or her personal information, and shall be given access to that information. An individual shall be able to challenge the accuracy and completeness of the information and have it amended as appropriate.
Compliance Challenges	An individual shall be able to address a challenge concerning compliance with the above principles to the designated individual or individuals accountable for the organization's compliance.

Em março de 2012, a Comissão Federal de Comércio dos EUA (FTC) emitiu um relatório recomendando que as organizações projetassem e implementassem seus próprios

programas de privacidade baseados nas melhores práticas descritas no relatório (ou seja, Privacy by Design) (FTC 2012). O relatório reafirma o foco da FTC em Fair Information Processing Principles (ver Tabela 3).

Tabela 3 Privacidade dos Estados Unidos Critérios do programa

Table 3 United States Privacy Program Criteria

Principle	Description of Principle
Notice / Awareness	Data collectors must disclose their information practices before collecting personal information from consumers.
Choice / Consent	Consumers must be given options with respect to whether and how personal information collected from them may be used for purposes beyond those for which the information was provided.
Access / Participation	Consumers should be able to view and contest the accuracy and completeness of data collected about them.
Integrity / Security	Data collectors must take reasonable steps to assure that information collected from consumers is accurate and secure from unauthorized use.
Enforcement / Redress	The use of a reliable mechanism to impose sanctions for noncompliance with these fair information practices.

Esses princípios são desenvolvidos para incorporar os conceitos nas Diretrizes de Processamento Justo de Informações da OCDE, incluindo ênfase na minimização de dados (limitação razoável de coleta) e limitação de armazenamento (retenção sólida), precisão e o requisito de que as empresas devem fornecer segurança razoável para dados do consumidor. Outros focos para práticas justas de informação incluem:

- Escolha simplificada do consumidor para reduzir o fardo imposto a consumidores
- A recomendação de manter um procedimento abrangente de gestão de dados em toda a informação

vida útil

- Opção Não Rastrear
- Requisitos para consentimento expresso afirmativo
- Preocupações quanto às capacidades de coleta de dados de grandes provedores de plataforma; transparência e políticas e avisos de privacidade claros
- Acesso de indivíduos aos dados
- Educar os consumidores sobre práticas de privacidade de dados
- Privacidade por Design

Há uma tendência global em direção ao aumento da proteção legislativa da privacidade das informações dos indivíduos, seguindo os padrões definidos pela legislação da UE. Leis ao redor do mundo colocam diferentes tipos de restrições à movimentação de dados através de fronteiras internacionais.

Mesmo dentro de uma organização multinacional, haverá limites legais para compartilhar informações globalmente. Portanto, é importante que as organizações tenham políticas e diretrizes que permitam que a equipe siga os requisitos legais, bem como use dados dentro do apetite de risco da organização.

3.3 Dados on-line em um contexto ético

Existem agora dezenas de iniciativas e programas emergentes projetados para criar um conjunto codificado de princípios para informar comportamentos éticos online nos Estados Unidos (Davis, 2012). Os tópicos incluem:

- **Propriedade dos dados:** Os direitos de controlar os dados pessoais em relação a sites de mídia social e corretores de dados.

  Agregadores de dados pessoais podem incorporar dados em perfis profundos dos quais os indivíduos não têm conhecimento.
- **O Direito de Ser Esquecido:** Ter informações sobre um indivíduo apagadas da web, particularmente para ajustar a reputação online. Este tópico faz parte da retenção de dados

práticas em geral.

- **Identidade:** Ter o direito de esperar uma identidade e uma identidade correta, e de optar por uma identidade privada.
- **Liberdade de expressão online:** expressar as próprias opiniões versus intimidação, incitação ao terror, trollagem ou insultos.

3.4 Riscos de Práticas Antiéticas de Manipulação de Dados A maioria das pessoas que trabalham com dados sabe que é possível usar dados para deturpar fatos. O livro clássico *How to Lie with Statistics* de Darrell Huff (1954) descreve uma série de maneiras pelas quais os dados podem ser usados para deturpar fatos enquanto criam um verniz de factualidade. Os métodos incluem seleção criteriosa de dados, manipulação de escala e omissão de alguns pontos de dados. Essas abordagens ainda estão em ação hoje.

Uma maneira de entender as implicações do tratamento ético de dados é examinar práticas que a maioria das pessoas consideraria antiéticas.
O tratamento ético de dados envolve um dever positivo de tratar dados de acordo com princípios éticos, como confiabilidade. Garantir que os dados sejam confiáveis pode incluir medi-los em relação às dimensões de Qualidade de Dados, como precisão e pontualidade. Também inclui um nível básico de veracidade e transparência – não usar dados para mentir ou enganar, e ser transparente em relação às fontes, usos e intenções por trás do tratamento de dados de uma organização. Os cenários a seguir descrevem práticas de dados antiéticas que violam esses princípios, entre outros.

3.4.1 Timing É possível mentir por omissão ou inclusão de certos pontos de dados em um relatório ou atividade com base no timing. A manipulação do mercado de ações por meio de negociações de ações no "fim do dia" pode aumentar artificialmente o preço de uma ação no fechamento do mercado, dando uma visão artificial do valor da ação. Isso é chamado de market timing e é ilegal.

A equipe de Business Intelligence pode ser a primeira a notar anomalias. Na verdade, eles agora são vistos como jogadores valiosos no mercado de ações

centros do mundo recriando padrões de negociação procurando por tais problemas, bem como analisando relatórios e revisando e monitorando regras e alertas. A equipe de Business Intelligence Ética pode precisar alertar funções de governança ou gestão apropriadas sobre tais anomalias.

3.4.2 Visualizações enganosas Gráficos

e tabelas podem ser usados para apresentar dados de maneira enganosa.
Por exemplo, mudar a escala pode fazer uma linha de tendência parecer melhor ou pior. Deixar pontos de dados de fora, comparar dois fatos sem esclarecer sua relação ou ignorar convenções visuais aceitas (como a de que os números em um gráfico de pizza representando porcentagens devem somar 100 e somente 100) também pode ser usado para enganar as pessoas a interpretar visualizações de maneiras que não são suportadas pelos dados em si.20

3.4.3 Definições pouco claras ou comparações inválidas Um

meio de comunicação dos EUA relatou, com base em dados do US Census Bureau de 2011, que 108,6 milhões de pessoas nos EUA recebiam assistência social, mas apenas 101,7 milhões de pessoas tinham empregos de tempo integral, fazendo parecer que uma porcentagem desproporcional da população geral recebia assistência social.21 A Media Matters explicou a discrepância: O número de 108,6 milhões para o número de "pessoas recebendo assistência social" vem de uma análise de participação em testados por meios, que incluem ··· programas de contas do Census Bureau "qualquer pessoa que resida em uma casa na qual uma ou mais pessoas receberam benefícios" no quarto trimestre de 2011, incluindo, portanto, indivíduos que não receberam benefícios do governo. Por outro lado, o número de “pessoas com emprego em tempo integral”... incluía apenas indivíduos que trabalhavam, não indivíduos que residiam em uma casa onde pelo menos uma pessoa trabalhava.22 A coisa ética a fazer, ao apresentar informações, é fornecer contexto que informe seu ---

significado, como uma definição clara e inequívoca da população que está sendo medida e o que significa estar “recebendo assistência social”. Quando o contexto necessário é deixado de fora, a superfície da apresentação pode implicar um significado que os dados não suportam.

Se esse efeito é obtido pela intenção de enganar ou por meio de

simplesmente falta de jeito, é um uso antiético de dados.

Também é simplesmente necessário, de uma perspectiva ética, não fazer mau uso das estatísticas.

A 'suavização' estatística de números ao longo de um período pode mudar completamente a percepção do número. 'Data mining snooping' é um termo cunhado recentemente para um fenômeno em investigações estatísticas de mineração de dados onde correlações exaustivas são realizadas em um conjunto de dados, essencialmente sobretreinando um modelo estatístico. Devido ao comportamento da 'significância estatística', é razoável esperar alguns resultados estatisticamente significativos que são, na verdade, resultados aleatórios. Os não treinados podem ser enganados. Isso é comum nos setores financeiro e médico (Jensen, 2000; ma.utexas.edu, 2012).23

3.4.4 Viés

Viés refere-se a uma inclinação de perspectiva. No nível pessoal, o termo é associado a julgamentos ou preconceitos irracionais. Em estatística, viés refere-se a desvios de valores esperados. Estes são frequentemente introduzidos por meio de erros sistemáticos na amostragem ou seleção de dados.24 O viés pode ser introduzido em diferentes pontos do ciclo de vida dos dados: quando os dados são coletados ou criados, quando são selecionados para inclusão na análise, por meio dos métodos pelos quais são analisados e em como os resultados da análise são apresentados.

O princípio ético da justiça cria um dever positivo de estar ciente de possíveis vieses que podem influenciar a coleta, o processamento, a análise ou a interpretação de dados. Isso é particularmente importante no caso de processamento de dados em larga escala que pode afetar desproporcionalmente grupos de pessoas que foram historicamente submetidas a preconceito ou tratamento injusto. Usar dados sem abordar as maneiras pelas quais o viés pode ser introduzido pode agravar o preconceito ao mesmo tempo em que reduz a transparência no processo, dando aos resultados resultantes o verniz de imparcialidade ou neutralidade quando não são neutros. Existem vários tipos de viés:

- **Coleta de dados para resultado pré-definido:** O analista é

pressionados a coletar dados e produzir resultados para chegar a uma conclusão pré-definida, em vez de um esforço para tirar uma conclusão objetiva.

- **Uso tendencioso de dados coletados:** Dados podem ser coletados com viés limitado, mas um analista é pressionado a usá-los para confirmar uma abordagem pré-determinada. Dados podem até ser manipulados para esse fim (ou seja, alguns dados podem ser descartados se não confirmarem a abordagem).
- **Palpite e pesquisa:** o analista tem um palpite e quer satisfazê-lo, mas usa apenas os dados que o confirmam e não leva em conta outras possibilidades que os dados podem surgir.

- **Metodologia de amostragem tendenciosa:** A amostragem é frequentemente uma parte necessária da coleta de dados. Mas o viés pode ser introduzido pelo método usado para selecionar o conjunto de amostras. É virtualmente impossível para humanos amostrar sem viés de algum tipo. Para limitar o viés, use ferramentas estatísticas para selecionar amostras e estabelecer tamanhos de amostra adequados. A conscientização do viés em conjuntos de dados usados para treinamento é particularmente importan
- **Contexto e cultura:** os preconceitos geralmente são baseados em cultura ou contexto, portanto, é necessário sair dessa cultura ou contexto para ter uma visão neutra da situação.

Questões de viés dependem de muitos fatores, como o tipo de processamento de dados em questão, as partes interessadas envolvidas, como os conjuntos de dados são preenchidos, a necessidade comercial que está sendo atendida e os resultados esperados do processo. No entanto, nem sempre é possível ou mesmo desejável remover todo o viés. O viés comercial contra clientes pobres (clientes com os quais não se busca mais negócios) é uma peça fundamental para muitos cenários construídos por analistas comerciais; eles são desmarcados de amostras ou ignorados na análise. Nesse caso, os analistas devem documentar os critérios que usaram para definir a população que estão estudando. Em contraste, algoritmos preditivos que determinam o "risco criminal" de indivíduos ou policiamento preditivo

enviar recursos para bairros específicos teria um risco muito maior de violar princípios éticos de justiça ou equidade, e deveria ter maiores precauções para garantir a transparência e a responsabilização algorítmica e para combater o preconceito em conjuntos de dados que treinam quaisquer algoritmos preditivos.25

3.4.5 Transformando e Integrando Dados A

integração de dados apresenta desafios éticos porque os dados são alterados à medida que se movem de um sistema para outro. Se os dados não forem integrados com cuidado, eles apresentam risco de manipulação de dados antiética ou mesmo ilegal. Esses riscos éticos se cruzam com problemas fundamentais no gerenciamento de dados, incluindo:

- **Conhecimento limitado da origem e linhagem dos dados:** se uma organização não sabe de onde os dados vieram e como eles mudaram ao serem movidos entre sistemas, ela não pode provar que os dados representam o que ela afirma que representam.

- **Dados de baixa qualidade:** as organizações devem ter padrões claros e mensuráveis para a qualidade dos dados e devem medir seus dados para confirmar se eles atendem aos padrões de qualidade. Sem essa confirmação, uma organização não pode garantir os dados e os consumidores de dados podem correr riscos ou colocar outras pessoas em risco ao usar os dados.
- **Metadados não confiáveis:** Os consumidores de dados dependem de Metadados confiáveis, incluindo definições consistentes de elementos de dados individuais, documentação da origem dos dados e documentação da linhagem (por exemplo, regras pelas quais os dados são integrados). Sem Metadados confiáveis, os dados podem ser mal compreendidos e potencialmente mal utilizados. Em casos em que os dados podem se mover entre organizações e especialmente onde podem se mover através de fronteiras, os Metadados devem incluir tags que indiquem sua procedência, quem os possui e se eles requerem proteção específica.

- **Nenhuma documentação do histórico de correção de dados:** As organizações também devem ter informações auditáveis relacionadas às formas como os dados foram alterados. Mesmo que a intenção da correção de dados seja melhorar a qualidade dos dados, isso pode ser ilegal. A correção de dados deve sempre seguir um processo formal e auditável de controle de alterações.

3.4.6 Ofuscação/ Redação de Dados Ofuscação ou

redação de dados é a prática de tornar informações anônimas ou remover informações sensíveis. Mas a ofuscação sozinha pode não ser suficiente para proteger dados se uma atividade downstream (análise ou combinação com outros conjuntos de dados) puder expor os dados.

Este risco está presente nas seguintes situações:

- **Agregação de dados:** Ao agregar dados em algum conjunto de dimensões e remover dados de identificação, um conjunto de dados ainda pode servir a uma finalidade analítica sem a preocupação de divulgar informações de identificação pessoal (PII).
  Agregações em áreas geográficas são uma prática comum (ver Capítulos 7 e 14).

- **Marcação de dados:** A marcação de dados é usada para classificar a sensibilidade dos dados (secretos, confidenciais, pessoais, etc.) e para controlar a divulgação para comunidades apropriadas, como o público ou fornecedores, ou mesmo fornecedores de determinados países ou outras considerações da comunidade.
- **Mascaramento de dados:** O mascaramento de dados é uma prática em que somente dados enviados de forma apropriada desbloquearão processos.
  Os operadores não conseguem ver quais são os dados apropriados; eles simplesmente digitam as respostas fornecidas a eles e, se essas respostas estiverem corretas, outras atividades são permitidas.
  Os processos de negócios que usam mascaramento de dados incluem call centers terceirizados ou subcontratados que devem ter apenas acesso parcial às informações.

O uso de conjuntos de dados extremamente grandes em análises de Ciência de Dados levanta preocupações práticas, em vez de meramente teóricas, sobre a eficácia da anonimização. Dentro de grandes conjuntos de dados, é possível combinar dados de maneiras que permitam que indivíduos sejam identificados especificamente, mesmo que os conjuntos de dados de entrada tenham sido anonimizados. A primeira preocupação quando os dados chegam a um data lake é analisá-los em busca de dados sensíveis e aplicar métodos de proteção aceitos. No entanto, estes por si só podem não oferecer proteção suficiente; é por isso que é vital que as organizações tenham uma governança forte e um compromisso com o tratamento ético de dados. (Consulte o Capítulo 14.)

3.5 Estabelecendo uma Cultura de Dados Ética Estabelecer

uma cultura de tratamento ético de dados requer entender as práticas existentes, definir comportamentos esperados, codificá-los em políticas e um código de ética, e fornecer treinamento e supervisão para impor comportamentos esperados. Assim como outras iniciativas relacionadas à governança de dados e à mudança de cultura, esse processo requer uma liderança forte.

O tratamento ético de dados obviamente inclui seguir a lei, mas também influencia como os dados são analisados e interpretados, bem como como são alavancados interna e externamente. Uma cultura organizacional que valoriza claramente o comportamento ético não só terá códigos de conduta, mas garantirá que controles claros de comunicação e governança estejam em vigor para dar suporte aos funcionários com consultas e caminhos de escalonamento adequados para que, se os funcionários tomarem conhecimento de práticas antiéticas ou risco ético, eles sejam capazes de destacar o problema ou interromper o processo sem medo de retaliação. Melhorar o comportamento ético de uma organização em relação aos dados requer um processo formal de Gestão de Mudanças Organizacionais (OCM). (Consulte o Capítulo 17.)

3.5.1 Revise as práticas atuais de tratamento de dados O

primeiro passo para a melhoria é entender o estado atual. O objetivo da revisão das práticas existentes de tratamento de dados é entender o grau em que elas estão direta e explicitamente conectadas aos drivers éticos e de conformidade. Esta revisão também deve identificar como

bem os funcionários entendem as implicações éticas das práticas existentes na construção e preservação da confiança de clientes, parceiros e outras partes interessadas. O resultado da revisão deve documentar os princípios éticos que fundamentam a coleta, o uso e a supervisão de dados da organização, durante todo o ciclo de vida dos dados, incluindo atividades de compartilhamento de dados.

3.5.2 Identificar Princípios, Práticas e Fatores de Risco O propósito de formalizar práticas éticas em torno do manuseio de dados é reduzir o risco de que os dados possam ser mal utilizados e causar danos a clientes, funcionários, fornecedores, outras partes interessadas ou à organização como um todo. Uma organização que tenta melhorar suas práticas deve estar ciente dos princípios gerais, como a necessidade de proteger a privacidade dos indivíduos, bem como preocupações específicas do setor, como a necessidade de proteger informações financeiras ou relacionadas à saúde.

A abordagem de uma organização à ética de dados deve estar alinhada com os requisitos de conformidade legal e regulatória. Por exemplo, organizações que operam globalmente precisam ter um amplo conhecimento dos princípios éticos na base das leis dos países em que operam, bem como conhecimento específico dos acordos entre países. Além disso, a maioria das organizações tem riscos específicos, que podem estar relacionados à sua pegada tecnológica, sua taxa de rotatividade de funcionários, os meios pelos quais coletam dados de clientes ou outros fatores.

Os princípios devem ser alinhados com os riscos (coisas ruins que podem acontecer se os princípios não forem seguidos) e práticas (as maneiras certas de fazer as coisas para que os riscos sejam evitados). As práticas devem ser apoiadas por controles, conforme ilustrado no exemplo a seguir:

- **Princípio orientador:** As pessoas têm direito à privacidade com relação às informações sobre sua saúde. Portanto, os dados pessoais de saúde dos pacientes não devem ser acessados, exceto por pessoas que estão autorizadas a acessá-los como parte do atendimento aos pacientes.

- **Risco:** Se houver amplo acesso aos dados pessoais de saúde dos pacientes, as informações sobre os indivíduos podem se tornar de conhecimento público, colocando em risco seu direito à privacidade.
- **Prática:** Somente enfermeiros e médicos terão permissão para acessar os dados pessoais de saúde dos pacientes e somente para fins de prestação de cuidados.
- **Controle:** Haverá uma revisão anual de todos os usuários dos sistemas que contêm informações pessoais de saúde dos pacientes para garantir que somente as pessoas que precisam ter acesso tenham acesso.

3.5.3 Crie uma estratégia e um roteiro de tratamento ético de dados Após uma revisão do estado atual e o desenvolvimento de um conjunto de princípios, uma organização pode formalizar uma estratégia para melhorar suas práticas de tratamento de dados. Essa estratégia deve expressar tanto os princípios éticos quanto o comportamento esperado relacionado aos dados, expressos em declarações de valores e um código de comportamento ético. Os componentes de tal estratégia incluem:

- **Declarações de valores:** Declarações de valores descrevem no que a organização acredita. Exemplos podem incluir verdade, imparcialidade ou justiça. Essas declarações fornecem uma estrutura para o tratamento ético de dados e tomada de decisões.
- **Princípios de tratamento ético de dados:** Os princípios de tratamento ético de dados descrevem como uma organização aborda os desafios apresentados pelos dados; por exemplo, como respeitar o direito dos indivíduos à privacidade. Princípios e comportamentos esperados podem ser resumidos em um código de ética e apoiados por uma política de ética. A socialização do código e da política deve ser incluída no plano de treinamento e comunicação.
- **Estrutura de conformidade:** Uma estrutura de conformidade

inclui fatores que impulsionam as obrigações organizacionais. Comportamentos éticos devem permitir que a organização atenda aos requisitos de conformidade. Os requisitos de conformidade são influenciados por preocupações geográficas e setoriais.

- **Avaliações de risco:** As avaliações de risco identificam a probabilidade e as implicações de problemas específicos que surgem dentro da organização. Elas devem ser usadas para priorizar ações relacionadas à mitigação, incluindo a conformidade dos funcionários com os princípios éticos.
- **Treinamento e comunicações:** O treinamento deve incluir a revisão do código de ética. O funcionário deve assinar que está familiarizado com o código e as implicações do manuseio antiético de dados. O treinamento precisa ser contínuo; por exemplo, por meio de um requisito para uma afirmação anual da declaração de ética. As comunicações devem atingir todos os funcionários.

- **Roteiro:** O roteiro deve incluir um cronograma com atividades que podem ser aprovadas pela gerência. As atividades incluirão a execução do plano de treinamento e comunicações, identificação e correção de lacunas nas práticas existentes, mitigação de riscos e planos de monitoramento. Desenvolva declarações detalhadas que reflitam a posição-alvo da organização sobre o manuseio apropriado de dados, inclua funções, responsabilidades e processos, e referências a especialistas para mais informações. O roteiro deve cobrir todas as leis aplicáveis e fatores culturais.

- **Abordagem para auditoria e monitoramento:** Ideias éticas e o código de ética podem ser reforçados por meio de treinamento. Também é aconselhável monitorar atividades específicas para garantir que elas estejam sendo executadas em conformidade com os princípios éticos.

3.5.4 Adote um modelo de risco ético socialmente responsável

Profissionais de dados envolvidos em Business Intelligence, análise e
A Ciência de Dados geralmente é responsável por dados que descrevem:

- Quem são as pessoas, incluindo seus países de origem e suas características raciais, étnicas e religiosas
- O que as pessoas fazem, incluindo atividades políticas, sociais e potencialmente criminosas
- Onde as pessoas vivem, quanto dinheiro elas têm, o que elas compram, com quem elas falam ou enviam mensagens de texto ou e-mails
- Como as pessoas são tratadas, incluindo resultados de análises, como pontuação e rastreamento de preferências que as marcarão como privilegiadas ou não para negócios futuros

Esses dados podem ser mal utilizados e contrariar os princípios subjacentes à ética de dados: respeito pelas pessoas, beneficência e justiça.

Executar atividades de BI, análise e ciência de dados de forma justa requer uma perspectiva ética que olhe além dos limites da organização para a qual as pessoas trabalham e leve em conta as implicações para a comunidade mais ampla. Uma perspectiva ética é necessária não apenas porque os dados podem ser facilmente mal utilizados, mas também porque as organizações têm uma responsabilidade social de não causar danos com seus dados.

Por exemplo, uma organização pode definir critérios para o que considera clientes "ruins" para parar de fazer negócios com esses indivíduos. Mas se essa organização tiver o monopólio de um serviço essencial em uma área geográfica específica, alguns desses indivíduos podem ficar sem esse serviço essencial e estarão em perigo por causa da decisão da organização.

Projetos que usam dados pessoais devem ter uma abordagem disciplinada para o uso desses dados. Veja a Figura 13. Eles devem levar em conta:

- Como eles selecionam suas populações para estudo (seta 1)
- Como os dados serão capturados (seta 2)

- Em quais atividades a análise se concentrará (seta 3)
- Como os resultados serão disponibilizados (seta 4)

Dentro de cada área de consideração, eles devem abordar potenciais riscos éticos, com foco particular em possíveis efeitos negativos sobre clientes ou cidadãos.

Um modelo de risco pode ser usado para determinar se o projeto deve ser executado. Também influenciará como executar o projeto. Por exemplo, os dados serão tornados anônimos, as informações privadas removidas do arquivo, a segurança nos arquivos reforçada ou confirmada, e uma revisão da lei de privacidade local e outras aplicáveis revisada com o jurídico.
A perda de clientes pode não ser permitida por lei se a organização for um monopólio em uma jurisdição e os cidadãos não tiverem outras opções de fornecedores, como energia ou água.

Como os projetos de análise de dados são complexos, as pessoas podem não ver os desafios éticos. As organizações precisam identificar ativamente os riscos potenciais. Elas também precisam proteger os denunciantes que veem riscos e levantam preocupações. O monitoramento automatizado não é proteção suficiente contra atividades antiéticas. As pessoas — os próprios analistas — precisam refletir sobre possíveis vieses. Normas culturais e ética no local de trabalho influenciam o comportamento corporativo — aprenda e use o modelo de risco ético.
A DAMA International incentiva os profissionais de dados a assumirem uma posição profissional e a apresentarem a situação de risco aos líderes empresariais que podem não ter reconhecido as implicações de usos específicos de dados e essas implicações em seu trabalho.

Figura 13 Modelo de Risco Ético para Projetos de Amostragem

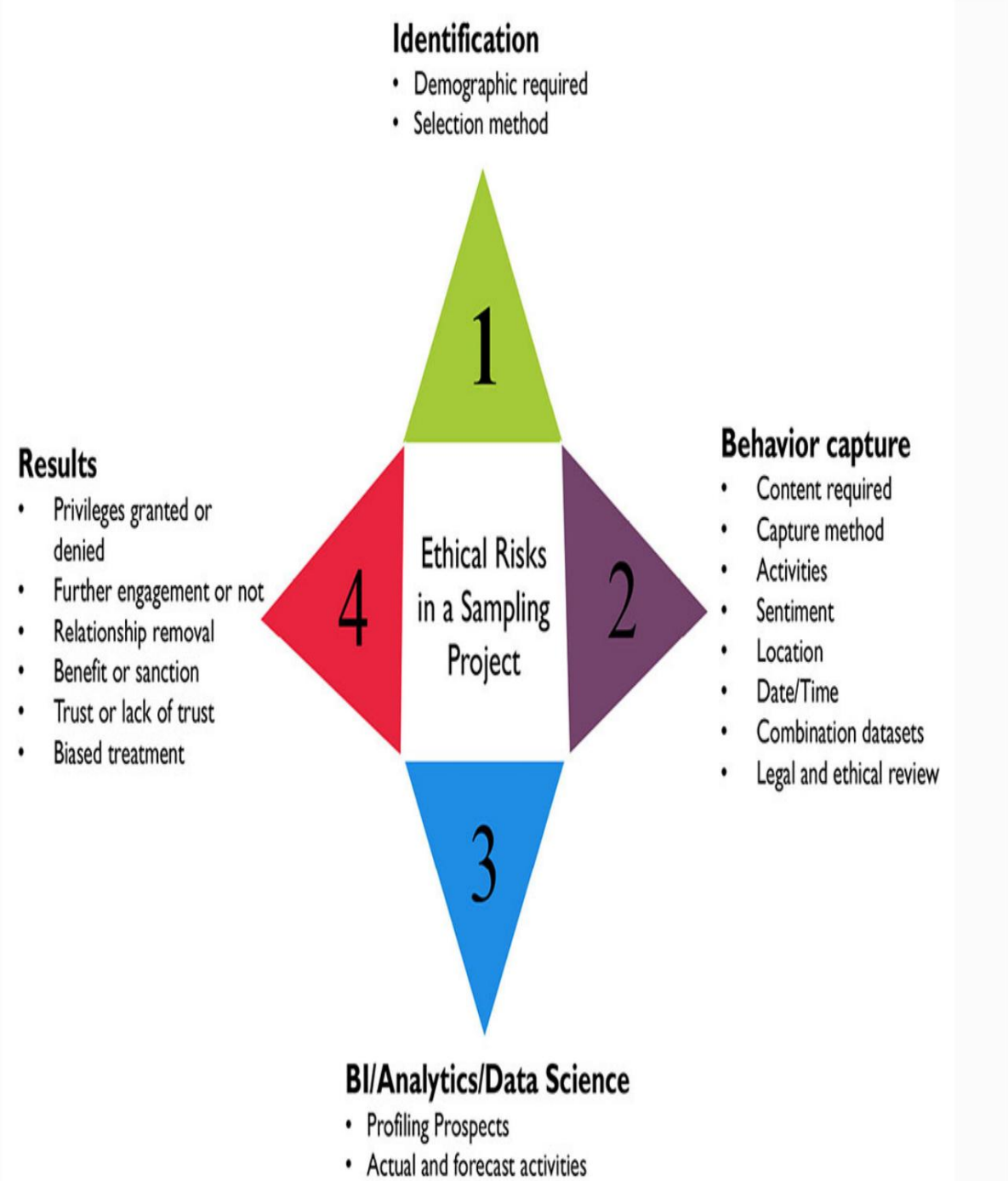

Figure 13 Ethical Risk Model for Sampling Projects

3.6 Ética e Governança de Dados

A supervisão do tratamento adequado de dados recai sobre a governança de dados e o aconselhamento jurídico. Juntos, eles são obrigados a manter-se atualizados.

atualizado sobre mudanças legais e reduzir o risco de impropriedade ética ao garantir que os funcionários estejam cientes de suas obrigações. A Governança de Dados deve definir padrões e políticas e fornecer supervisão das práticas de tratamento de dados. Os funcionários devem esperar tratamento justo, proteção contra possíveis violações de relatórios e não interferência em suas vidas pessoais. A Governança de Dados tem um requisito de supervisão específico para revisar planos e decisões propostas por estudos de BI, análise e Ciência de Dados.

A certificação Certified Data Management Professional (CDMP) da DAMA International exige que o profissional de gerenciamento de dados assine um código formal de ética, incluindo a obrigação de lidar com dados de forma ética para o bem da sociedade, além da organização que o emprega.

4. Trabalhos Citados / Recomendados

Blann, Andrew. *Manipulação e análise de dados.* Oxford University Press, 2015. Impresso. Fundamentos da ciência biomédica.

Conselho para Big Data, Ética e Sociedade (site) http://bit.ly/2sYAGAq.

Davis, Kord. *Ética do Big Data: equilibrando risco e inovação.* O'Reilly Media, 2012. Impresso.

Supervisor Europeu de Proteção de Dados (EDPS). Parecer 4/2015 “Rumo a uma nova ética digital: Dados, dignidade e tecnologia.” http://bit.ly/2sTFVlI.

Comissão Federal de Comércio, EUA (FTC). *Relatório da Comissão Federal de Comércio Protegendo a Privacidade do Consumidor em uma Era de Mudanças Rápidas.* Março de 2012. http://bit.ly/2rVqTxQ e http://bit.ly/1SHOpRB.

REGULAMENTO GDPR (UE) 2016/679 DO PARLAMENTO EUROPEU E DO CONSELHO de 27 de abril de 2016 relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral de Proteção de Dados).

Hasselbalch, Gry e Pernille Tranberg. *Ética de Dados: A Nova*