ÉTICA, PRIVACIDADE e BIG-DATA

PAULO CESAR MASIERO

SUMÁRIO

- Definição de privacidade
- Propaganda Direcionada
- Exemplos e casos
- Identidade
- Reputação
- Propriedade
- Anonimidade

PRIVACIDADE: Definição

- Privacidade (Webster): O estado de estar sem companhia (ou sozinho) ou estar sem ser observado.
- Direito à Privacidade: poder estar livre de intrusão (ou intromissão)
- Por que um conceito tão simples provoca tanto debate?
- Uma possível resposta é que a simplicidade permite várias interpretações e as interpretações são moldadas pela nossa cultura, história, visão do mundo e tecnologia disponível. Isto é, um contexto.

PRIVACIDADE: Definição

- Privacidade (Webster): O estado de estar sem companhia (ou sozinho) ou estar sem ser observado.
- Direito à Privacidade: poder estar livre de intrusão (ou intromissão)
- Por que um conceito tão simples provoca tanto debate?
- Uma possível resposta é que a simplicidade permite várias interpretações e as interpretações são moldadas pela nossa cultura, história, visão do mundo e tecnologia disponível. Isto é, um contexto.

O QUE PRIVACIDADE SIGNIFICA PARA VOCÊ?

- CONSIDERE QUE VOCÊ SEJA:
  - Um candidato a um emprego.
  - Um oficial da polícia.
  - Um político no Brasil.☺
- **Indivíduos** julgam a privacidade pelo perigo(dano) percebido se alguma informação importante para um certo contexto se tornar de conhecimento público.

PRIVACIDADE NA ERA MODERNA

- Privacidade das Comunicações
  - Muitos de nós acreditamos que nossas conversas pessoais, e-mails, telefonemas, mensagens intantâneas etc são privadas. Mas…
  - … é muito fácil monitorar todas essas comunicações.
- Privacidade do Comportamento:
  - Antes era dificil descobrir e compartilhar como agiamos dentro de casa, o que comprávamosmos, onde íamos etc.
  - Hoje é possível capturar digitalmente muito do nosso comportamento e usar esses dados para prever o que faremos.
- Privacidade da nossa Pessoa:
  - O direito de nos mantermos relativamente anônimos na sociedade, se assim quisermos. Isso era possível.
  - Hoje tudo mudou com a proliferação de câmaras, fotos digitais e rastreamento de localização.

PRIVACIDADE – Categorias Principais

- Privacidade Física: estar livre de intromissão em sua pessoa física, espaço ou propriedades.
  - Em casa, na rua, ou no trabalho.
- Privacidade Informacional: sua expectativa de privacidade quando informação pessoal é coletada, armazenada e compartilhada de forma digital ou em algum outro formato.
- Privacidade Organizacional: Agencias governamentais, organizações e empresas esperam manter seus segredos e atividades sem serem revelados a terceiros.

Back in 1993...

"On the Internet, nobody knows you're a dog."

Bem vindo à era do Big Data!

- Hoje trabalhamos online, socializamos online, vemos os shows favoritos online, pagamos impostos, jogamos, namoramos, vamos ao banco ..... online.
- Essas atividades todas geram uma enorme quantidade de dados que são armazenados em gigantescos *data centers* ao redor do mundo.
- Há mais de um trilhão de URLs no índice do Google, e mais de 3,5 bilhões de buscas todo dia.
- Há milhões de fotos, vídeos, blogs, tweets e usuários do facebook ...
- Além de dados geoespaciais, de governos, de universidades, etc.

Bem vindo à era do Big Data!

- Um número enorme de empresas rastreia o comportamento das pessoas na Web, geralmente usando cookies, e armazena esses dados. A maior parte dessa informação não identifica os indivíduos.
- A maior parte das empresas coleta dados com o objetivo de fazer ou vender propaganda direcionada a potenciais compradores.
- Algumas vendem os dados coletados a terceiros, como a eXelate.
- Outras agregam dados de várias fontes para obter um bom entendimento do que você faz e quem você é, como a SpoKeo.

Bem vindo à era da invasão à Privacidade!

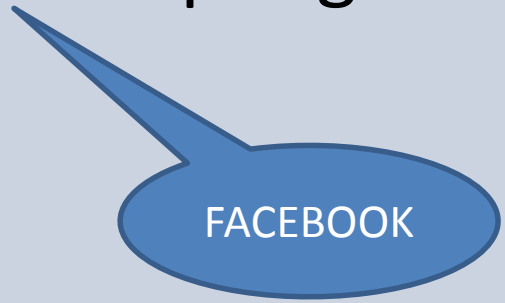
- A princípio não há nada moralmente errado em propaganda direcionada, desde que o consumidor esteja consciente dela.
- Se o seu dado é coletado e usado apenas com o propósito de propaganda, o impacto é relativamente benigno. Há até argumentos favoráveis.
- Outras questões mais importantes são as seguintes:
  - Quem está usando nossos dados?
  - Por que (ou para quê) estão usando nossos dados?
  - Como podemos nos proteger de invasões à privacidade?
  - Como saber quem está nos espiando?

Privacidade vs Segurança e Proteção

- Muitas empresas têm violado a *permissão* (implícita) de coletar dado para propaganda.
- Colocam cookies em seu computador sem seu conhecimento e sua permissão.
- A tecnologia torna a coleta cada vez mais fácil e eficiente.

Privacidade vs Segurança e Proteção (Cont.)

- É razoável esperar que o Governo pode proteger os usuário por meio de leis, agências e polícia?
- Portanto, uma questão importante é: como equilibrar privacidade, segurança e proteção em um mundo cada vez mais transparente e perigoso?



Atores

- **Coletadores de Dados**
  - Empresas de TI (Websites) que coletam, armazenam, usam e repassam dados obtidos quando um usuário usa um dispositivo computacional ou aciona leitores de cartão de crédito, RFID, CCTV etc.
- **Mercadores de dados (Agregadores, Data Markets)**.
  - São plataformas em que os usuários (indivíduos, organizações de marketing, agencias governamentais ou outras empresas) podem buscar por conjuntos de dados específicos que atendam às suas necessidades e então baixá-los (pagando ou de graça).

Atores

- **Usuários de dados**
  - Pessoas ou organizações que compram ou têm acesso livre a conjuntos de dados específico que atendem às suas necessidades.
- **Reguladores/Protetores**
  - Agências ou organizações privadas que monitoram questões de privacidade sob vários pontos de vista e que estão envolvidas em políticas autorregulatórias e seu cumprimento pela indústria.
  - Governo: ministérios, congresso nacional, autarquias, justiça…
  - Fundações privadas
  - Especialistas, blogueiros, etc.

Iniciativas para regulamentar a propaganda direcionada

- De 2009 a 2012 a Federal Trade Commission (FTC) dos Estados Unidos liderou iniciativas para estabelecer um programa voluntário para as indústrias regulamentarem a coleta e uso de dados pessoais.
- Um grupo das empresas de Web, Networking Advertising Initiative, também trabalhou neste assunto e propôs princípios autorregulatórios, que são amplamente compatíveis com os do FTC.
- A discussão continua até o momento e há dois pontos principais de divergência.

Princípios consensuais

- As empresas devem adotar “privacy by design” e construir produtos e serviços que protegem a privacidade.
- As empresas devem obter consentimento dos consumidores e fornecer opções claras para que eles possam concordar ou não com a coleta de seus dados (opt in/opt out)
- As empresas devem aumentar a transparência de suas práticas em relação aos dados.
- Consequência: veja a política de privacidade de websites como Facebook, LinkedIn, Google, Amazon e Tweeter.

Princípios não consensuais

- Defensores da privacidade querem:
  - uma política de opt in/opt out para todos os sites,
  - uma lista “branca” nacional de não-rastreio (Do not track list).
- Estão em andamento discussões sobre como proteger o consumidor em ambientes de *marketplaces* de dispositivos móveis. Duas ideias são discutidas:
  - Maior transparência,
  - Controle do usuário (especialmente tornando a opção opt out default).

Caso Google vs Apple/Safari

… three individuals in the UK who found out a few years ago that Google had been using some clever tricks to bypass the privacy settings on the Safari browser used on their Apple computers.

The result of Google's ingenuity was that by tracking people's browsing habits and noting the websites they had visited, advertisers using Google's services were able to target advertising at individuals in much more personalised ways than would otherwise have been possible.

Google stopped this practice after a furore in the US led to it being fined millions of dollars by the authorities. But the three individuals - Judith Vidal-Hall, Robert Hann and Marc Bradshaw - decided to press on and sue Google in the UK for the distress that its actions had caused them.

http://www.theinquirer.net/inquirer/opinion/2403552/fallout-from-googles-safari-privacy-scandal-continues

Caso do Carrinho de Compras

Caso relatado por um profissional que trabalha num loja de moda feminina por e-comércio. Esse tipo de negócio exige um esforço permanente para entender o comportamento do cliente e seu perfil e oferecer serviços que atendam às suas expectativas.

Essa empresa foi abordada por um representante de uma empresa de "e-mail remarketing" (shopback.com.br) com a proposta de capturar o e-mail de pessoas que navegam pelo site sem se cadastrar e não concluem a compra. Com isso podem enviar e-mails com ofertas para esse possível cliente.

A questão é decidir se isso deve ser feito ou não, pois se uma pessoa não se registrou na loja, deveria ela ser incomodada com o envio de e-mails? Como decidir o que fazer?

Caso FindFace

FindFace is a service that can search for VK.com (a Russian Social Network) accounts on the base of a portrait photo of a person. 30 search attempts are free, and then you'll have to pay. It was developed by two young nerds: Kukarenko and Kubakov. The service has mobile apps for iOS and Android, complemented by a website version. Users can take a photo and immediately use it to search with FindFace. The app shows profile photos of the potential matches. You can click on each photo to look through all public images on the user's account.

The program has successfully found 9 of 10 test "victims" in an experiment carried by a journalist. If you take photos of strangers on the streets or in the subway sneakily, accuracy decreases two or even three times. And if you upload images taken from a long distance, the service often becomes unable to find a human in the photo. Still, if you zoom or crop the image, FindFace will work again. In the daylight it's not hard to take a photo of a pedestrian with an average smartphone that would be good enough for Findface. In the subway you'll need to use tripod or a good camera.

Caso FindFace (Cont.)

It does not take a wild imagination to come up with sinister applications in this field too, as for example authoritarian regimes able to tag and identify participants in street protests. Kabakov and Kukharenko said they had not received an approach from Russia's FSB security service, but "if the FSB were to get in touch, of course we'd listen to any offers they had".

FindFace describes itself as a dating service. For example, you see an attractive person, take a photo and browse through their account — ok, now you've to a topic to make a pass. In fact, this service can let you make a lot of more useful — and strange — things.

They believe the real money-spinner from their face-recognition technology will come from law enforcement and retail. Kukharenko and Kabakov have recently returned from the US, and Kabakov was due to travel to Macau and present the technology to a casino chain. The pair claims they have been contacted by police in Russian regions, who told them they started loading suspect or witness photographs into FindFace and came up with results. "It's nuts: there were cases that had seen no movement for years, and now they are being solved," said Kabakov.

https://www.theguardian.com/technology/2016/may/17/findface-face-recognition-app-end-public-anonymity-vkontakte
https://blog.kaspersky.com/findface-experiment/11916/;
http://www.theatlantic.com/technology/archive/2016/05/find-face/483962/
https://advox.globalvoices.org/2016/04/07/the-russian-art-of-meta-stalking/

Há uma geração ou duas uma pessoa poderia facilmente desaparecer dentro de seu país, agora não. http://www.wired.com/vanish/2009/11/ff_vanish2/

Um arcabouço (framework) para a ética em Big Data

- Quatro elementos comuns definem o que pode ser eticamente considerado para indivíduos e organizações:
  - Identidade: Qual é a relação entre sua identidade offline e online?
  - Privacidade/Anonimidade: Quem deve controlar o acesso ao dado?
  - Propriedade: Quem é o dono dos dados, quem tem direitos sobre sua transferência e quais são as obrigações das pessoas que geram e usam os dados?
  - Reputação: Como podemos determinar que um dado é confiável? O aumento de informações e formas de interação criadas por BD aumenta a complexidade de gerir como somos percebidos e julgados.

IDENTIDADE

- O nosso entendimento histórico do significado de identidade está sendo transformado por BD.
- Entender nossos valores com respeito ao conceito de identidade melhora e expande nossa habilidade para determinar as ações que são apropriadas ou inapropriadas.
- BD permite a terceiros a habilidade de resumir muito facilmente, agregar ou correlacionar vários aspectos da nossa identidade.

IDENTIDADE

- Se BD está evoluindo o significado do conceito de identidade, então está também evoluindo nossa responsabilidade ética para com o conceito que a palavra representa.
- Portanto:

Quanto mais nossas ações forem alinhadas com a evolução e expansão da identidade, mais poderemos entender de forma completa e explícita os valores que as motivam.

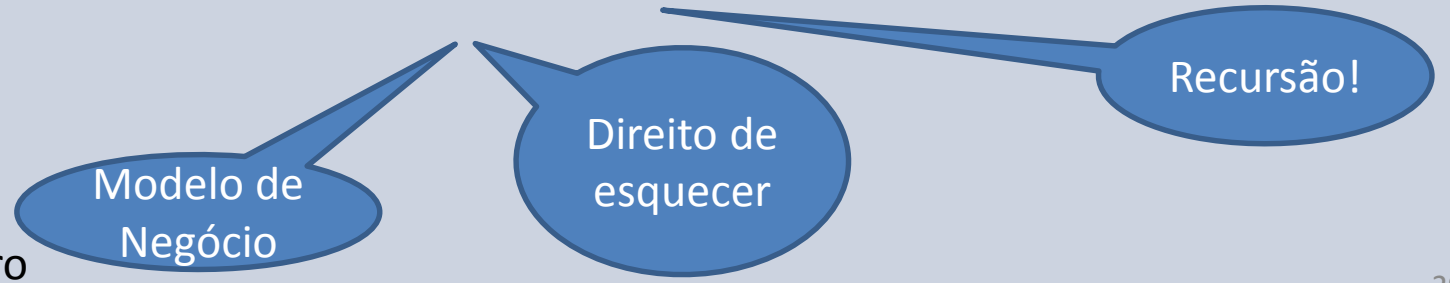
IDENTIDADE

- Um juiz de Nova York e outro da Flórida decidiram há poucos anos que um endereço IP não é evidência suficiente para identificar infratores de direito autoral.
  - https://torrentfreak.com/ip-address-not-person-140324/
- O Caso da Target.
- Walll-Mart, em 2011, foi acusado de usar uma comunidade online falsa para aumentar o apoio a novas lojas em áreas em que a ideia não era popular. Uma empresa de relações públicas teria sido a responsável.

REPUTAÇÃO

REPUTAÇÃO

- Até cerca de 20 anos atrás nossa reputação consistia principalmente de o que as pessoas, especialmente as que você conhecia e interagiam com você, sabiam sobre você.
- A menos que alguém seja famoso, a maioria de nós gerenciamos nossa reputação agindo corretamente (ou não!) com aqueles diretamente à nossa volta.
- Em alguns casos, pensamos também num segundo círculo que pode influenciar a reputação: o que as pessoas que conheceram você disseram sobre você para as pessoas que eles conheceram.



PROPRIEDADE

- O grau de propriedade que temos sobre informações específicas sobre nós varia largamente.
- Consideremos as seguintes situações no mundo offline:
  - Somos “proprietários” de fato do nosso peso e altura?
  - A nossa existência por si só constitui um ato criativo, sobre o qual temos direitos de propriedade ou outros direitos associados com criação?
  - Informações sobre a história da nossa família, genes, descrição física, time de futebol preferido, etc. constituem propriedades nossas?

PROPRIEDADE

- Há alguma distinção entre a qualidade da propriedade que temos sobre essas informações?
- Se sim, como esses direitos e privilégios offline, garantidos pela constituição e outras leis, aplicam-se à presença online da mesma informação?

Muitas entidades e governos em diferentes países vêm propondo leis para garantir a privacidade de indivíduos e controlar como informações pessoais são usadas online.

Anonimidade: o que é um “Dado Pessoal”

- Informação que identifica uma pessoa (IIP)
- Por muito tempo se entendeu que seriam dados como nome, rg, cpf, endereço,...
- Hoje, a distinção entre IIP e outro tipo de dado ocorre basicamente por restrições tecnológicas.
- É necessário ter uma outra definição que é considerada quando as pessoas estão preocupadas com privacidade.

- Definição de Davis e Paterson: qualquer dado gerado no decorrer das atividades de uma pessoa.

Anonimização

- Anonimização de dados é um tipo de *sanitização* da informação com o objetivo de proteger a privacidade.
- Pode ser obtida por "encriptação" ou por remoção de informação de identificação pessoal dos conjuntos de dados, de tal forma que a pessoa que gerou a informação permanece anônima.
- Desanonimização é o reverso de anonimização. É uma estratégia de data mining (BD) em que dados anônimos são cruzados com outras fontes de dados para reidentificar a fonte do dado anônimo.

Problemas com Anonimização: Netflix

- Em 2006 Netflix criou um concurso para procurar um algoritmo bom para fazer recomendações de filmes.
- Para testar o algoritmo liberou um banco de dados com 500K registros anônimos com notas dadas por assinantes a filmes.
- Dois pesquisadores da UT Austin entraram no concurso e criaram um algoritmo que consegue identificar assinantes do Netflix fazendo agregação e correlação com uma base de dados pública: Internet Movie Database (IMDb)

Problemas com Anonimização

- Nos anos 90, a pesquisadora Latanya Sweeney estudou os dados do censo e concluiu que:
  - 87.1% das pessoas dos EUA são unicamente identificadas pela combinação do Zip code (5 dgts), Data de Nascimento (c/ano) e sexo.
  - 53% das pessoas dos EUA são unicamente identificadas por sua cidade, data de nascimento e sexo
  - 18% por seu país, data de nascimento e sexo.
- Quando era ainda uma estudante de pós-graduação identificou dados de um banco de dados medicos liberados pelo governo de Massachussets para um pesquisador, usando esse tipo de informação.