Lauro Gripa Neto

LGPD na prática

MAGRATHEA LABS

www.magrathealabs.com

MAGRATHEA LABS

Lauro Gripa Neto
Engenheiro de Software

lauro@magrathealabs.com laurogripa

Governança de dados

Fonte: https://www.orioninc.com/blog/putting-together-data-governance-strategy/

MEDIDAS IMEDIATAS (CURTO PRAZO)

Utilizando GDPR como base

- LGPD é baseada na lei europeia **G**eneral **D**ata **P**rotection **R**egulation
- Leis modernas e baseadas em princípios
- Incentivam mudança na cultura (pessoas)
- Existem diversas experiências que servem de base
- Procure por “GDPR compliance/compliant”

Links:

- https://gdprchecklist.io/
- https://github.com/erichard/awesome-gdpr
- https://github.com/bakke92/awesome-gdpr

Hosting

- AWS: https://devcenter.heroku.com/articles/gdpr
- Azure: https://azure.microsoft.com/en-us/blog/protecting-privacy-in-microsoft-azure-gdpr-azure-policy-updates/
- Heroku: https://devcenter.heroku.com/articles/gdpr
- DigitalOcean: https://www.digitalocean.com/legal/gdpr/

Exemplos de aplicações

RAILS

- https://github.com/prey/gdpr_rails
- https://github.com/ankane/ahoy

REDIS

- https://redislabs.com/blog/dont-worry-happy-redis-labs-ready-gdpr/

POSTGRES

- https://www.enterprisedb.com/

Anomização

“Utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo”

Anomização

0. Sampling

```
SELECT * FROM people TABLESAMPLE
BERNOULLI(20);
```

Pró: simples

Contra: não é anonimização!

Anomização

1. Suppression

```
UPDATE people SET name = '<CONFIDENTIAL>';
UPDATE people SET address = NULL;
```

Pró: simples

Contra: quebra restrições

Anomização

2. Random Substitution

```
UPDATE people SET name = md5(random()::text);
UPDATE people SET salary = 100000*random();
```

Pró: simples e não quebra restrição de nulidade

Contra: ruim para analytics e integração contínua

Anomização

3. Variance

```
UPDATE people
SET salary = salary *  (1+ (2 * random() - 1 ) * 0.25) ;
```

Pró: Mantém os dados números significativos

Contra: Permite reidentificação dos dados

Anomização

4. Encryption

```
CREATE EXTENSION pgcrypto;
UPDATE people
SET name = SELECT crypt('name', gen_salt('md5'));
```

Pró: bom para texto com restrição de unicidade

Contra: chaves podem ser roubadas se forem estáticas

Anomização

5. Shuffling

```
WITH p1 AS (
   SELECT row_number() over (order by random()) n,
         salary AS salary1
   FROM people
),
p2 AS (
   SELECT row_number() over (order by random()) n,
         id AS id2
   FROM people
)
UPDATE people
SET salary = p1.salary1
FROM p1 join p2 on p1.n = p2.n
WHERE id = p2.id2;
```

Pró: mantém integridade e valores reais

Contra: maior complexidade e ineficiência para valores pouco distintos (e.g. booleanos)

Anomização

6. Faking / Mocking

```
UPDATE people
SET address = fake_address();
```

Pró: ótimo para integração contínua

Contra: ruim para analytics e relativamente complexo para criar “fakes” significativos (utilizar Faker)

Anomização

7. Partial Suppression

```
UPDATE people
SET phone = 'XX XX XX ' || substring(phone FROM 9
FOR 5 );
```

Pró: determinístico e o dono consegue identificar

Contra: funciona somente somente com texto

Anomização

8. Generalization

```
CREATE TABLE anonymous_people
AS (SELECT
        id,
        '*' AS name,
        int4range(age/10*10, (age/10+1)*10) AS age
    FROM people);
```

Pró: Bom para analytics

Contra: Modifica o tipo dos dados e quebra integração contínua

Anomização

Strategy	Data Types	When to use
Suppression	All	Useless attributes
Random Substitution	All	Useless attributes with integrity constraints
Variance	Numeric / Dates	Analytics
Encryption	Text	Analytics / UNIQUE attributes
Shuffling	All	Analytics
Faking / Mocking	All	Dev / CI / Functional Testing
Partial Suppression	Text	Direct Identifiers
Generalization	Numeric / Dates	Analytics

NOSSAS EXPERIÊNCIAS

Desafios

- Obtenção de consentimento
  - Adquirir consentimento em e-mail marketing
    i. Clientes de clientes
- Acessibilidade
  - Developers não tem mais acesso a contas de usuários
- Direito de ser esquecido
  - Remoção dos dados (em até 7 dias)
  - Construção de interface para deleção ou download dos dados
    i. Clientes de clientes

Preocupações

- Imagem/marketing
- Relações comerciais
- Produtividade
- Desempenho
- Escalabilidade

Soluções

- Obtenção de consentimento
  - Double Opt-in para conversão de Leads
  - Cookie consent para rastreamento
- Acessibilidade
  - Anonimização dos dados
- Direito de ser esquecido
  - Foco em qualidade
  - Gerenciar consentimento
  - Existe o “legitimo interesse”, porém, sempre busque o consentimento.
  - Nem todos os dados precisam deletados

Dicas

- Pensamento sistêmico
- Princípios
- Governança de dados
- Foco em qualidade
- Desenvolver pensando em dados
- Processos bem definidos
- Transparência e boa vontade de estar em conformidade

Novamente, governança de dados

Fonte: https://www.orioninc.com/blog/putting-together-data-governance-strategy/

PRIORIZANDO PRIVACIDADE (LONGO PRAZO)

PRIVACY BY DESIGN / PRIVACY BY DEFAULT

Recomendado pelo GDPR

- https://gdpr-info.eu/issues/privacy-by-design/

Criticado por ser muito “vago”

- https://www.esat.kuleuven.be/cosic/publications/article-1542.pdf

Engineering Privacy by Design

Seda Gürses, Carmela Troncoso, and Claudia Diaz

K.U. Leuven/IBBT, ESAT/SCD-COSIC
firstname.lastname@esat.kuleuven.be

LOCAL-FIRST

- fast
- multi-device
- offline
- collaboration
- longevity
- privacy
    - cloud apps are fundamentally non-private
- user control

CRDT: conflict-free replicated data types

- https://en.wikipedia.org/wiki/Conflict-free_replicated_data_type

Fonte: https://www.inkandswitch.com/local-first.html

#OwnYourData

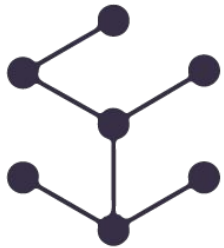
Campanha por Brittany Kaiser (Denunciante da Cambridge Analytica)

- https://twitter.com/ownyourdatanow

SELF-SOVEREIGN IDENTITY (BLOCKCHAIN)

own your identity (https://medium.com/@Cristf/want-to-own-your-data-start-by-owning-your-identity-e5eb6d54793d)

- centralized vs federated vs decentralized: https://identitypraxis.com/2019/04/06/getting-ready-for-decentralized-identity-the-personal-information-economy/
- siloed vs third-party vs p2p: https://medium.com/evernym/the-three-models-of-digital-identity-relationships-ca0727cb5186
- self-sovereign identity: https://medium.com/metadium/introduction-to-self-sovereign-identity-and-its-10-guiding-principles-97c1ba603872
- zero-knowledge proofs: https://hackernoon.com/wtf-is-zero-knowledge-proof-be5b49735f27
- crypto shredding: https://www.thoughtworks.com/radar/techniques/crypto-shredding
- verifiable credentials data model: https://www.w3.org/TR/vc-data-model/
- sovrin: https://sovrin.org
- jolocom: https://jolocom.io



Blockchain Joinville

meetup.com/blockchain-joinville

MAGRATHEA LABS

www.magrathealabs.com/careers

contact@magrathealabs.com

Obrigado!

MAGRATHEA LABS

www.magrathealabs.com