# O Uso do Endereço de Email pelos *Sites* mais Acessados pelo Público Brasileiro e os Possíveis Impactos na Privacidade de Seus Usuários

Emerson Ribeiro de Mello
Instituto Federal de Santa Catarina – IFSC
mello@ifsc.edu.br

Shirlei Aparecida de Chaves
Instituto Federal de Santa Catarina – IFSC
shirlei.chaves@ifsc.edu.br

## ABSTRACT

More and more online businesses are using the email for a purpose other than its original one, which was to enable communication between people. Email address is used as a user account identifier and systems send transactional messages to their user's email addresses. This paper presents an analysis of the user's account registration process in the main online businesses used by Brazilian people in 2018. It has been found that the exclusive use of email as account identifier is the most used option and that the great majority does not verify if the email informed during the registration process really belongs to the requesting user. Along with the findings discussion, we outline some measures that could be used by these businesses to reduce the negative impact to users privacy.

## KEYWORDS

Identidade, Privacidade, Usabilidade

## 1 INTRODUÇÃO

Desde a sua invenção em 1971, praticamente duas décadas antes da invenção da *World Wide Web* [1] e da expansão mundial da Internet, o uso do email sofreu diversas transformações. De acordo com Ray Tomlinson [2], o email foi criado com o intuito de permitir a comunicação entre pessoas usando a rede de computadores, no caso a ARPANET.

Atualmente o uso do email vai além de permitir a comunicação entre pessoas. Ducheneaut e Bellotti [3] indicaram como o uso do email estava sendo sobrecarregado em relação ao seu projeto original, especialmente como ferramenta de gerenciamento de informações pessoais, no que eles convencionaram chamar de *email como habitat*.

Segundo Chadwick [4], o identificador de um usuário consiste em um conjunto de caracteres, dígitos e símbolos que são usados para identificar de forma única uma entidade em um determinado sistema. Segundo Jin, *et. al.* [5], muitos *sites* adotaram o endereço de email como o identificador de seus usuários. Como não existe a possibilidade de duas ou mais pessoas possuírem o mesmo endereço de email, seu uso como identificador único é justificado e interessante também do ponto de vista da usabilidade. Isto é, o usuário não precisaria criar um identificador único, e possivelmente diferente, para cada *site* com que for interagir.

Se não fosse feito uso do endereço de email como identificador, então o usuário teria que respeitar as semânticas específicas de cada *site* (p. ex. identificador formado somente com letras), bem como restrições de espaço de nomes. Neste último caso, por exemplo, se o usuário tentar criar uma conta de usuário em um novo *site* e o identificador comumente usado por ele em outros *sites* (p. ex. `meuLogin`) já estiver em uso por um outro usuário, então ele precisará criar um identificador diferente. Isso pode resultar no aumento da lista de identificadores que esse usuário precisará gerenciar.

Uma outra vertente ainda na qual o email vem sendo usado é a de confirmação de transações. Sempre que um usuário interage com algum negócio *online*, por exemplo, comprando um produto, assinando um serviço, etc., a confirmação dessa transação é enviada por email (p. ex. nota fiscal eletrônica, recibo de confirmação de pagamento ou de processamento de pedido, atualização sobre o rastreamento e entrega de produtos enviados via correios, etc.). Ou seja, neste caso o email também não é usado para comunicação entre pessoas, mas sim para o envio de informações de um sistema para uma pessoa, sendo geralmente uma comunicação em um único sentido. Isso porque, normalmente não é dada ao usuário a possibilidade de responder o email recebido para, por exemplo, solicitar esclarecimentos, tirar dúvidas, etc. Tipicamente cada email transacional é único e sob demanda, em resposta a um gatilho ou transação específicos [6]. Como um cenário exemplo, considerar o caso citado acima de uma compra em um *site* de comércio eletrônico, no qual a confirmação de pagamento pela operadora de crédito dispara uma chamada a uma *Application Programming Interface* (API) de envio de email, geralmente contratada com um terceiro, para enviar email para o comprador informando que a compra foi confirmada.

Com a promulgação da Lei Geral de Proteção de Dados [7] brasileira, motivada também pelo Regulamento Geral de Proteção de dados europeu (*General Data Protection Regulation* - GDPR)[1], de caráter extraterritorial e portanto com implicações também no Brasil, a coleta e tratamento de dados pessoais passa a ser regida por princípios como a coleta de dados mínima necessária para a finalidade do negócio, e a adoção da metodologia *Privacy by Design* ou "Privacidade desde a concepção". Desse modo, espera-se que a privacidade seja considerada nas estruturas tecnológicas adotadas pelas organizações, incluindo as precauções necessárias para que não haja por exemplo divulgação indevida de dados pessoais.

Nesse trabalho partiu-se do pressuposto de que o uso de algo para um fim para o qual não foi projetado pode resultar em efeitos colaterais não previstos ou mesmo indesejados. Sendo assim, elaboraram-se as seguintes questões: *O uso do email para algo além da comunicação entre pessoas gera prejuízos ou riscos para as empresas? Esse uso fere a privacidade de seus usuários?*

Esse artigo apresenta um estudo sobre o uso do endereço de email pelos *sites* mais acessados pelo público brasileiro em abril de 2018. Buscou-se verificar a prática desses *sites* com relação ao uso do endereço de email como identificador de conta de usuário, bem como o uso de emails transacionais. Considerando os resultados

[1] http://europa.eu/dataprotection

obtidos, fez-se uma pequena discussão sobre os impactos dessas práticas na usabilidade e na privacidade dos usuários desses *sites*.

As empresas presentes na amostra foram agrupadas em duas categorias: empresas com entrega de bens e serviços no mundo físico; e empresas com entrega de bens e serviços no mundo digital. Entendeu-se que essa categorização seria necessária, pois cada categoria possui requisitos diferentes em relação a quantidade de informação pessoal que precisa coletar de seus usuários.

Para cada uma das empresas presente na amostra foi criada uma conta de usuário e buscou-se verificar: os dados pessoais solicitados durante o processo de criação de conta; o tipo de identificador único de usuário; se o email informado durante o registro era confirmado como de propriedade do usuário que estava realizando o cadastro; e, por fim, se havia e como era o procedimento para exclusão de conta de usuário. Notou-se que, dessa amostra, o endereço de email foi a opção mais adotada como identificador da conta de usuário e que a grande maioria não confirma se o endereço de email informado durante a criação da conta de usuário realmente pertence ao usuário em questão. Por fim, a maioria das empresas com entrega no mundo físico não possui interfaces que permitam excluir as contas de usuário.

Esse artigo está organizado da seguinte forma. Na Seção 2 é feita uma revisão sobre o processo de criação de conta de usuários e identificadores de conta de usuário usados pelas empresas com negócios *online*. A metodologia utilizada durante o desenvolvimento da análise investigativa é apresentada na Seção 3. Na Seção 4 são apresentados os resultados da análise, bem como uma discussão sobre os mesmos. Por fim, na Seção 5 são feitas as considerações finais desse trabalho.

## 2 CONTEXTUALIZAÇÃO

Para criar uma conta de usuário em um negócio *online*, geralmente é necessário que se escolha um identificador único de usuário e uma senha, podendo eventualmente fornecer outros dados pessoais que o negócio *online* considerou como relevantes.

A criação dessa conta normalmente se dá por meio de um autocadastro realizado pelo próprio usuário, em funcionalidade própria fornecida pelo *site*. Dentre as diversas questões a serem definidas pela empresa quando da disponibilização dessa funcionalidade de autocadastro, está a questão sobre a semântica do identificador único de usuário a ser utilizado. Por exemplo, se o mesmo deverá ser um identificador único no âmbito local ou global [8]. As possibilidades para identificador único comumente usadas pelas empresas brasileiras com negócios *online* são: nome de usuário, CPF, endereço de email e número de telefone celular.

O **nome de usuário** consiste de um identificador único de âmbito local. Ou seja, garante-se neste caso que dois usuários distintos, não possuirão o mesmo nome de usuário em um mesmo *site*. Contudo, é possível que usuários distintos possuam o mesmo nome de usuário em *sites* distintos. Dessa forma, não existe a garantia que um determinado usuário poderá deter o mesmo **nome de usuário** em diferentes *sites*. Isso é algo ruim do ponto de vista da usabilidade e pode resultar na fadiga digital, uma vez que o usuário precisará lembrar de diversos nomes de usuários diferentes [9]. A natureza de alguns negócios *online* determinam o **nome de usuário** como a única solução possível, como é o caso de provedores de serviços de email.

O Cadastro de Pessoa Física (**CPF**) é um banco de dados gerenciado pela Receita Federal do Brasil cujo objetivo é manter informações cadastrais de contribuintes. Como o CPF é único para cada pessoa e como praticamente todo brasileiro possui, ou deverá possuí-lo em algum momento, muitas empresas adotaram o CPF como identificador único de âmbito global. Do ponto de vista da usabilidade, o uso do CPF como identificador traz vantagens para o usuário, pois esse poderia usar o mesmo CPF em diferentes *sites*, evitando assim a fadiga digital. Devido a natureza de alguns serviços entende-se que o CPF é uma informação cadastral obrigatória, p. ex. para emissão de nota fiscal. Porém, para alguns outros serviços o CPF é uma informação cadastral irrelevante, p. ex. para uma rede social. Dessa forma, o usuário pode se sentir desconfortável quando for criar uma conta de usuário e for solicitado seu número de CPF.

O **endereço de email** consiste de um identificador único de âmbito global, ou seja, garante-se aqui que dois indivíduos distintos não possuirão o mesmo endereço de email. O uso do endereço de email como identificador apresenta a vantagem de ser uma informação que, em teoria, o usuário consegue lembrar mais facilmente e poderia usar o mesmo endereço em diferentes *sites*. Assim como no caso do CPF, o usuário pode ficar desconfortável em usar seu **endereço de email** como identificador em situações nas quais a natureza do serviço que está sendo acessado não precisar do endereço de email como uma informação cadastral crucial. Por exemplo, um *site* para compra e venda de veículos.

Do ponto de vista da empresa, o uso do **endereço de email** como identificador traz facilidades para a condução de atividades administrativas da conta (p. ex. recuperação da senha) e para envio de comunicação, seja essa ofertas de produtos e serviços ou informação sobre transações que o usuário fez (p. ex. envio de nota fiscal). Ou seja, fazer com que o usuário informe seu endereço de email como identificador garante a empresa que essa já terá de antemão um canal de comunicação com seus usuários.

Aos emails enviados por esse canal de comunicação, para as atividades administrativas da conta do usuário e algumas das situações de comunicação citadas acima, se convencionou chamar de *emails transacionais*. Emails transacionais são emails automáticos os quais, segundo Nielsen Norman Group [10], são um dos pontos de contato mais importantes para informar o usuário sobre suas transações. Em [6, 10] são definidas as seguintes atividades como exemplos comuns de emails transacionais: criação e ativação de contas; mensagens de boas vindas; convites e compartilhamentos de usuários; alertas de segurança da conta; redefinição de senhas e de segundo fator de autenticação; recibos de compras e notificações de envio; avisos legais. São, portanto, emails que podem conter dados pessoais do usuário, além de possibilitar operações importantes como a troca e/ou recuperação de senha.

Com o advento da popularização dos *smartphones* e do acesso móvel à Internet, o **número de telefone celular** começou a ser usado por alguns *sites* como identificador de usuário. O número de celular é um identificador de âmbito global, considerando que este número não é compartilhado por mais de uma pessoa. Do ponto de vista da usabilidade, usar o número do celular pode ser benéfico para o usuário, pois usaria o mesmo número em diferentes *sites*. Contudo, isso também poderia ser negativo, caso o usuário venha

a trocar de número (p. ex. mudança de estado ou país), pois esse teria que lembrar de atualizar seu cadastro em todos os *sites* onde usou o celular como identificador. Do ponto de vista da privacidade o número de celular poderia ser ainda pior que usar o endereço de email, pois permitiria campanhas publicitárias mais agressivas contra o usuário (p. ex. ligações telefônicas e envio de SMS).

Em suma, a natureza e o público de cada empresa determinam qual a melhor opção para identificador de conta de usuário. Cabe ainda a cada empresa, em suas escolhas, garantir o melhor balanceamento entre privacidade e usabilidade.

## 3 METODOLOGIA

De acordo com Gerhardt e Silveira [11], a pesquisa realizada neste trabalho tem objetivo exploratório, com o intuito de tornar o problema investigado mais explícito. A abordagem utilizada é predominantemente qualitativa, uma vez que analisa as informações coletadas de modo mais intuitivo, embora de forma organizada, com enfoque maior na interpretação do objeto. Quanto à natureza, pode ser entendida como pesquisa aplicada, pois visa gerar conhecimento para aplicação prática. Os procedimentos técnicos utilizados envolvem pesquisa bibliográfica e levantamento por amostra.

A pesquisa exploratória realizada nesse trabalho buscou verificar o uso do email pelas empresas que oferecem negócio *online* para o público brasileiro, em dois diferentes cenários. No primeiro cenário está o **uso do endereço de email como identificador de conta de usuário** (Veja Seção 2). No segundo cenário está o **uso do email transacional**, no qual o endereço de email obtido como informação complementar de cadastro é usado como canal de comunicação para informar sobre as transações executadas pelo sistema de informação.

Ciente que diferentes regras de negócio podem impor diferentes requisitos funcionais para o desenvolvimento de um sistema, optou-se por classificar as empresas a serem analisadas em duas categorias: **(a) empresas que fornecem produtos ou serviços com entrega no mundo físico**; **(b) empresas que fornecem produtos ou serviços com entrega exclusiva no mundo digital**.

Dentro da categoria (a) estão empresas que entregam ao usuário um bem físico, por exemplo um tênis, ou um serviço que usuário obrigatoriamente terá uma interação presencial no mundo real, por exemplo uma hospedagem em um hotel, uma refeição ou um curso. Por outro lado, na categoria (b) estão empresas que entregam produtos ou serviços exclusivamente digitais e que não necessitam de uma interação presencial entre o usuário e o fornecedor do produto ou serviço. Por exemplo, venda *online* de música de forma que o usuário pode baixar o arquivo de música e passar a ter a propriedade desse arquivo, ou um serviço de assinatura para ouvir música sob demanda, no qual interromper a assinatura, implica obrigatoriamente em perder o acesso ao serviço.

Schaar [12] afirma que garantir a privacidade dos usuários deveria ser uma premissa quando se projeta um sistema de informação. Por exemplo, um sistema só deveria coletar dados de usuário essenciais para atender as regras de funcionamento de seu negócio. A Lei Geral de Proteção de Dados Pessoais (LGPD) [7] trata como um princípio que se limite o tratamento de dados ao mínimo necessário para a realização de suas finalidades, abrangendo os dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados. Ou seja, que se busque coletar o mínimo possível de dados pessoais, se e somente se necessário.

Considerando tal afirmação, assume-se aqui que a quantidade de informação pessoal obrigatória que deve ser coletada pelas empresas da categoria (a) será maior do que das empresas da categoria (b). Por exemplo, para uma loja *online* que vende tênis é essencial conhecer o nome completo e o CPF, bem como o endereço de entrega. Contudo, o endereço de email, se não for usado como identificador da conta de usuário, não é crucial para permitir que o usuário consiga comprar o tênis. Por outro lado, para um serviço de *streaming* de música, o endereço de email seria obrigatório e suficiente, pois esse seria usado como identificador único do usuário. Ou seja, para essa categoria de serviço, informações como o endereço do usuário, filiação, etc seriam irrelevantes.

Ciente que há um grande número de empresas em cada uma dessas categorias, buscou-se por uma amostra representativa dos principais serviços acessados no Brasil. Para a seleção dessa amostra foi feito uso do serviço de *ranking* de *sites* Alexa Internet[2], da Amazon. Foi obtida uma lista com os 500 *sites* mais acessados no Brasil em 12/04/2018. Primeiramente, fez-se uma classificação manual dos domínios de *sites* conhecidos e que se enquadravam nas categorias de interesse, daqueles que não se enquadravam (*sites* de busca como Google, de compartilhamento de *torrents*, pornografia, e etc). Em uma segunda iteração, os *sites* não conhecidos dos autores ou cujo domínio não revelava a sua natureza foram visitados um a um, a fim de verificar seu enquadramento nas categorias objeto desse trabalho. Além dessas considerações, outras foram utilizadas para o refinamento da amostra:

- Para *sites* de uma mesma empresa e com diferentes domínio nacional de nível superior (*country code Top-Level Domain* – ccTLD), considerou-se analisar somente o ccTLD `.br`. Por exemplo, para amazon.com e amazon.com.br foi mantido na amostra apenas o domínio amazon.com.br, simplesmente por se tratarem da mesma empresa;
- Foram excluídos da amostra *sites* de bancos, que notoriamente exigem uma interação no mundo físico (p. ex. abertura de uma conta corrente) ou que façam uma coleta mais ampla de dados como, fornecimento de foto e documentos digitalizados, além da necessidade de aprovação da conta mediante análise por exemplo de crédito. Ou seja, não há aqui uma criação de conta de usuário puramente *online*;
- Serviços de pagamento *online* como PayPal e PagSeguro foram excluídos, pois entende-se que esses são intermediadores de pagamento. Ou seja, não vendem produtos ou serviços;
- Foram removidos *sites* que são subserviços de outro serviço, já contido na amostra, como o blogger.com e o blogspot.com, pois a criação de conta de usuário acontece no Google, sendo assim analisar o Google já contemplaria seus subserviços.

A amostra analisada foi composta por 78 *sites*, dos quais 49 foram classificados na categoria de empresas com entrega de produtos ou serviços no mundo físico e 29 na categoria de empresas com entrega de produtos ou serviços no mundo digital. A lista com todos os sites analisados nesse trabalho pode ser vista no Apêndice A.

[2]https://www.alexa.com

Assim, para cada um dos *sites*, contido na amostra, foi criada uma conta de usuário. Durante esse procedimento de criação de conta observou-se: (1) quais informações poderiam ser usadas como identificador da conta de usuário; (2) quais eram os dados mínimos obrigatórios para permitir a criação da conta, ou seja, sem os quais não seria possível concluir o processo de criação da conta; (3) qual era o procedimento para recuperação de senha; (4) se era possível excluir completamente a conta criada e assim remover todos os dados pessoais associados a essa conta.

## 4 RESULTADOS E DISCUSSÃO

No processo de criação de contas foi observado qual informação era solicitada para atuar como o identificador de conta de usuário e a compilação dos resultados é apresentada pelos gráficos da Figura 1. Constatou-se que o uso exclusivo de **endereço de email** como identificador foi a opção mais usada em ambas as categorias de empresas (55% e 48%). O **endereço de email** também era tido como alternativa para os casos onde era possível usar CPF, número de telefone celular ou mesmo, nome de usuário. Por fim, o uso exclusivo do **nome de usuário** como identificador foi a opção menos adotada pelas empresas que fazem entrega de bens ou serviços no mundo físico (apenas 4%), e também teve pouca adoção por empresas do mundo digital (10%). Ou seja, pode-se concluir que na amostra analisada, o **endereço de email** é usado como identificador de conta de usuário por 96% das empresas que fazem entrega de bens ou serviços no mundo físico e por 90% das empresas do mundo digital.

Observou-se que há casos em que, apesar do endereço de email ter sido usado como identificador de conta de usuário, um nome de usuário também foi criado de forma automática e implícita pelo sistema. Ou seja, o usuário não pôde escolher qual seria seu nome de usuário durante a criação, porém os sistemas permitiam ao usuário alterá-lo posteriormente. Para esses casos notou-se que o objetivo primário do nome de usuário era para atuar como um pseudônimo para o usuário durante suas interações com outros usuários no sistema, evitando assim a exposição de seu endereço de email nessas interações.

Sabe-se que humanos estão propensos a erros [13] e isso pode ser um dos motivos para que os procedimentos para criação de conta de usuários, ou mesmo procedimentos de troca de senha, exijam que o usuário forneça a senha desejada e depois confirme a senha digitada. De forma análoga, pode-se pressupor que usuários poderiam errar quando estiverem fornecendo seu endereço de email durante a criação de sua conta de usuário. Os sistemas geralmente possuem rotinas para garantir que o endereço de email fornecido seja um endereço válido, por exemplo, por meio de expressões regulares. Contudo, somente com expressões regulares não seria possível determinar se o endereço de email fornecido, realmente pertence ao usuário em questão.

Se o sistema não possuir uma rotina para confirmar que o endereço de email fornecido realmente pertença ao usuário, então expõe-se o usuário a uma série de riscos. Por exemplo, sistemas de recuperação de senha geralmente enviam uma nova senha para o endereço de email do usuário. Nesse caso, o usuário não terá mais como recuperar sua senha e como consequência, perderá o acesso a sua conta no *site*. Porém, o risco pode ser ainda maior se o endereço de email fornecido pertencer a uma outra pessoa. Nesse caso, esse terceiro poderia ter total acesso a conta do usuário e aos seus dados pessoais.

Tal situação, onde um usuário forneça um endereço de email válido, que não seja seu e que pertença a outra pessoa, pode ocorrer em cenários específicos[3], e que atendam os seguintes critérios:

- Se a pessoa que estiver criando a conta de usuário não possuir afinidade com tecnologia e que não entenda como funciona o sistema de email; e
- Se o endereço de email fornecido pertencer a uma outra pessoa; e
- Se o usuário que está criando a conta e a outra pessoa, a qual teve seu endereço de email usado inadvertidamente, forem homônimos; e
- Se o real detentor do email possuir um endereço de email comum em um provedor de email que seja popular, como o Gmail. Por exemplo, nome@gmail.com, ou sobrenome@ ou mesmo nome.sobrenome@.

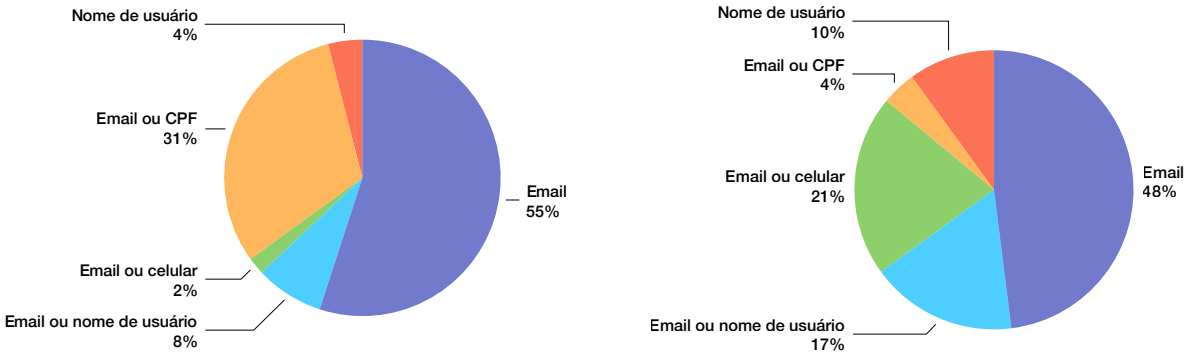
Mark Weiser [14] apresentou a visão de um mundo onde a computação do século 21 seria móvel e ubíqua. Com o advento dos *smartphones* tal visão tornou-se realidade, pois esses levaram a computação para a rotina do usuário de um modo integrado, ao contrário da dinâmica em torno do computador pessoal, que mantinha a computação como uma atividade separada e não integrada ao dia-a-dia da pessoa. Isto ajuda a explicar como os telefones inteligentes conseguiram atingir um público que o computador pessoal até então não tinha conseguido.

De acordo com o Censo 2010 [15], existem mais de 11 milhões de pessoas com o nome "Maria" no Brasil. Considerando esse novo público, trazido para a Internet com o advento dos telefones inteligentes, seria factível imaginar que pessoas chamadas Maria e que não possuam afinidade com tecnologia, poderiam de forma descuidada, informar o endereço de email maria@gmail.com durante o processo de criação de sua conta de usuário em um *site*.

Nesse trabalho preocupou-se em determinar o percentual de empresas que validam se o endereço de email fornecido realmente pertence a pessoa que está criando a conta. Conforme apresentado na Figura 1, mais de 90% das empresas da amostra adotaram o endereço de email como identificador de conta de usuário. Ou seja, para um usuário conseguir passar pelo processo de autenticação desses *sites* será necessário que este forneça o mesmo endereço de email e senha, usados durante a criação de sua conta de usuário. Contudo, apesar de grande dependência sobre o endereço de email, somente 16% das empresas com entrega de bens e serviços no mundo físico e 28% das empresas com entrega no mundo digital, fizeram de uso de mecanismos para garantir que o endereço de email fornecido durante a fase de registro realmente pertencia a pessoa em questão.

Segundo Maqbali e Mitchel [16], o procedimento de recuperação de senha geralmente é composto por três passos: (1) pedido de recuperação – usuário acessa uma página específica; (2) validação que o pedido não foi feito por um *bot* – p. ex. uso de CAPTCHA [17]; (3) reestabelecimento da senha – sistema ajuda o usuário a lembrar

[3]Os autores desse trabalho tiveram seus endereços de email usados por diversos homônimos achando que os endereços lhes pertenciam.




(a) Empresas com entrega de produtos ou serviços no mundo físico

(b) Empresas com entrega de produtos ou serviços no mundo digital

**Figura 1: Identificador usado para criação de conta de usuário**

da senha (p. ex. mantendo cópia de todas as senhas já usadas, perguntas de segurança, etc.) ou gerando uma nova senha temporária e enviando a mesma para o endereço de email ou para o telefone do celular (via SMS) do usuário. O procedimento de recuperação de senha de todas as empresas da amostra depende do endereço de email.

O processo de criação de conta em algumas empresas também exigiu informações complementares do usuário (biográficas e endereço), além do identificador de conta e senha. Entende-se que a natureza do negócio pode justificar a solicitação desses dados extras. Por exemplo, a venda de produtos que precisam ser entregues fisicamente exige que o usuário forneça o endereço de entrega e o CPF, necessários para emissão da nota fiscal.

O gráfico da Figura 2 indica quais informações extras foram solicitadas, agrupadas em: empresas que validaram o endereço de email (p. ex. envio de um *link* por email para ativação da conta); e empresas que não validaram o email. A intenção aqui estava em verificar quais informações pessoais poderiam ser expostas, caso o real detentor do endereço de email, usado como identificador de conta de usuário, viesse a fazer uso de interfaces de recuperação da senha para assim ter acesso a conta criada pelo usuário descuidado.

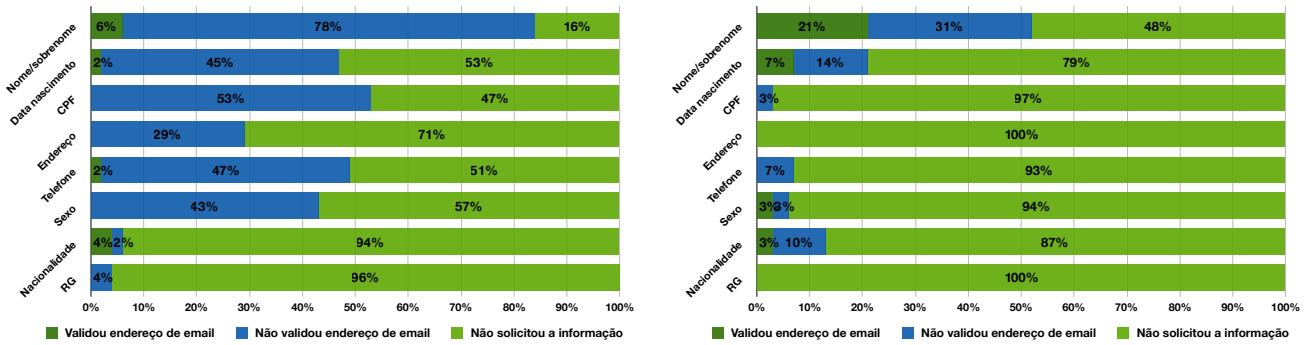
Como pode ser observado na Figura 2, as empresas com entregas no mundo físico solicitaram mais informações complementares, como esperado. Contudo, um grande percentual dessas empresas não validou o endereço de email durante a criação da conta de usuário. Todas as empresas que solicitaram o RG (4% da amostra) e CPF (53% da amostra) não validaram o endereço de email. O telefone foi solicitado por 49% das empresas da amostra, porém somente 2% validaram o endereço de email. Percentuais similares em relação à data de nascimento, em que 47% da amostra solicitaram essa informação e apenas 2% validaram o endereço. O nome e sobrenome foram as informações mais solicitadas, por 84% da amostra, contudo somente 6% da amostra validou o endereço de email.

No grupo das empresas com entregas no mundo digital, percebe-se que menos informações complementares foram solicitadas. Por exemplo, nenhuma empresa solicitou o endereço do usuário ou RG e um percentual menor, comparado com as empresas do mundo físico, solicitou dados como telefone e CPF. Contudo, todas as empresas que solicitaram o telefone (7% da amostra) e CPF (2% da amostra) não validaram o endereço de email. Sendo assim, pode-se concluir que, para grande parte da amostra analisada, o *design* falho do processo de criação de conta de usuários realmente permitiria a exposição de informações pessoais.

Considerando a Lei Geral de Proteção de Dados Pessoais [7] do Brasil e o Regulamento Geral de Proteção de Dados da União Europeia (*EU General Data Protection Regulation* – GDPR), que garantem aos usuários o direito ao esquecimento de seus dados, buscou-se verificar quais *sites* permitiriam ao usuário excluir sua conta. Como pode ser observado na Figura 3, das empresas que não confirmaram o endereço de email, 68% das empresas com entrega no mundo físico e 24% com entrega no mundo digital, não oferecem qualquer interface para permitir a exclusão da conta de usuário.

Das empresas que não confirmaram o endereço de email, 19% com entrega no mundo digital e 10% com entregas no mundo físico, não permitem aos usuários a exclusão de sua conta, mas somente a desativação. Ou seja, os dados pessoais dos usuários ainda são mantidos pela empresa, não respeitando assim as legislações atuais. Em suma, a maioria das empresas com entregas no mundo físico não oferece interfaces para exclusão da conta de usuário. Para as empresas com entrega no mundo digital a situação é inversa, ou seja, a maioria das empresas oferece alguma interface para exclusão ou desativação da conta de usuário.

Por fim, verificou-se também que todos os cadastros solicitaram email, mesmo os poucos que utilizaram exclusivamente nome de usuário como identificador único (4% no mundo físico e 10% no mundo digital). Isso indica que em algum momento o email poderá ser utilizado com propósitos transacionais. E, de fato, observou-se o uso transacional para recuperação de senha, caso de 100% da amostragem conforme já apresentado anteriormente. Pôde-se observar também que o uso transacional para ativação de conta, o qual serviria para validar a propriedade do email pelo usuário, não teve utilização tão ampla, pois como apresentado anteriormente, mesmo na amostra das empresas com entrega no mundo digital, a qual teve maior percentual de validação, apenas 28% validou. Os




(a) Empresas com entrega de produtos ou serviços no mundo físico (b) Empresas com entrega de produtos ou serviços no mundo digital

**Figura 2: Dados pessoais complementares solicitados durante a criação de conta de usuário**

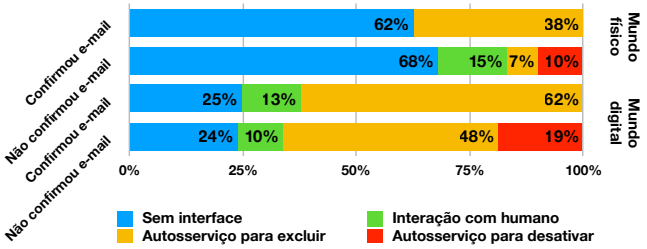



**Figura 3: Interfaces para exclusão de conta de usuário**

demais usos, como envio de notas fiscais e confirmação de pagamento, para a amostragem, não foram analisados, pois envolveriam de fato realizar compras nos referidos serviços.

## 5 CONSIDERAÇÕES FINAIS

O uso do endereço de email como identificador de conta de usuário traz benefícios do ponto de vista da usabilidade, contudo pode gerar riscos para a privacidade dos usuários. O *email como habitat*, conforme definido em [3], possibilitaria um cenário de ataque que bastaria comprometer a conta de email de um usuário alvo e a partir disso seria possível personificá-lo, ou ter acesso aos seus dados pessoais, em todos os demais *sites* que este possua conta, usando por exemplo, procedimentos de recuperação de senha. Em [18] é apresentado um conjunto de hipóteses que mostra como a experiência negativa com a violação de dados pessoais pode tornar o usuário mais sensível à questão da privacidade. Por exemplo, uma pessoa que já teve seus dados pessoais revelados em um incidente de privacidade, estará menos propensa a confiar nas organizações novamente e terá uma preocupação maior com seus dados pessoais.

De acordo com Schaar [12], é essencial projetar sistemas destinados a processar dados pessoais considerando a privacidade por padrão. Sendo assim, procedimentos para criação de contas de usuários deveriam validar se os dados informados são realmente do usuário em questão. A não validação do endereço de email pode expor indevidamente dados de uma pessoa mesmo em cenários onde essa pessoa não interagiu diretamente com qualquer sistema de informação, como é o caso da Nota Fiscal Eletrônica (NF-e), Nota Fiscal de Serviços Eletrônica (NFS-e) ou Nota Fiscal de Consumidor Eletrônica (NFC-e) [19], ou mesmo contratação de serviços por telefone. Em [20] é argumentado ainda que como os consumidores são vulneráveis tanto pela falta de informação como de controle, as organizações têm também um dever moral, muitas vezes negligenciado, de tomar precauções com os dados dos usuários, indo além do dever legal. Com a LGPD entrando em vigor em agosto de 2020, além da boa fé na atividade de tratamento de dados pessoais, as organizações terão ainda que observar como princípio legal que situações acidentais de comunicação ou difusão de dados pessoais sejam prevenidas por medidas técnicas ou administrativas.

No cenário com a NFC-e e a NF-e, que consiste de uma instância de email transacional, o consumidor ao comprar um simples café ou mesmo um carro, pode informar seu email verbalmente para o vendedor e assim receberá o documento fiscal por email. Nesse caso, entende-se que uma interface para validar se o email fornecido de fato pertence ao usuário impactaria na usabilidade, dada as circunstâncias em que essa informação é solicitada. Contudo, entende-se que os atuais sistemas poderiam ser alterados para não enviarem o documento fiscal diretamente para o email, mas sim para enviarem um *link*. Assim, o usuário teria que acessar a página desse *link* e fornecer alguma informação que estaria presente no documento fiscal, da qual ele teria conhecimento, para só então ter acesso ao documento completo.

Em situações vividas pelos autores, que não foram consideradas na amostra, constatou-se problemas no *design* de sistemas de empresas de telecomunicação e de fabricantes de veículos. No caso da empresa de telecomunicação, o endereço de email fornecido durante a contratação do serviço (isso pode ser feito por telefone), é usado para que o usuário possa gerenciar completamente sua assinatura a partir do portal do assinante. No portal do assinante o usuário pode: ver todos os dados pessoais (fornecidos verbalmente durante a contratação); trocar o nome da rede sem fio ou a senha do equipamento que está na casa do cliente; etc. Ou seja, nesse caso o *design* falho pode ir além de ferir a privacidade do usuário, pois permitiria ao real detentor do email gerar prejuízos financeiros para o cliente da empresa, por exemplo, contratando um plano mais caro, ou mesmo gerar transtornos para o cliente e custos para a empresa, com a troca da senha da rede sem fio.

No caso do fabricante de veículos, o veículo adquirido possuía um sistema de rastreamento por satélite que permite o bloqueio remoto do veículo. Novamente, o endereço de email, fornecido verbalmente durante a aquisição do veículo, foi usado para criar uma conta de usuário no sistema de rastreamento. Ou seja, aqui a privacidade do comprador do veículo estava ainda mais em risco, pois o real detentor do email recebeu inicialmente a NF-e, que contém dados pessoais e dados do veículo adquirido, e ainda permitiria a esse rastrear constantemente o deslocamento diário desse veículo, podendo inclusive, solicitar o bloqueio do veículo remotamente.

Para os casos descritos acima, não existia qualquer canal de comunicação que pudesse ser usado pelo real detentor do email para excluir seu endereço dos sistemas de informação das empresas. Contudo, entende-se que nesse cenário a validação do email no momento da compra também impactaria na usabilidade e talvez o envio de um *link* de ativação por email, para acessar um *site* e fornecer alguma informação de conhecimento do real comprador, pudesse mitigar o problema.

Por fim, conclui-se que os *sites* da maioria das empresas com negócios *online* mais acessadas pelo público brasileiro não consideraram a privacidade por padrão durante sua concepção. Constatou-se que a não validação do endereço de email, apesar de ter um impacto positivo do ponto de vista da usabilidade, pode sim expor dados pessoais dos usuários desses sistemas e, em alguns casos específicos, pode inclusive gerar prejuízos financeiros para os usuários e para as empresas.

Sendo assim, considerando os questionamentos levantados no início deste trabalho, sobre o uso do email para algo além da comunicação entre pessoas poder gerar prejuízos ou riscos para as empresas e ferir a privacidade de seus usuários, conclui-se que usar o email como identificador único de usuário em contas de negócios *online* pode ferir a privacidade dos usuários ao expor dados pessoais acidentalmente. Isto também poder gerar prejuízos financeiros às empresas, seja por questões de imagem, seja por questões legais, especialmente por causa da LGPD[7], a qual prevê diversas sanções, incluindo multas, no caso de infrações.

Importante salientar que não se advoga pelo não uso do email como identificador único ou como email transacional, mas sim que medidas técnicas sejam adotadas para evitar que dados pessoais sejam expostos indevidamente. Medidas técnicas tão simples como a confirmação da propriedade do endereço de email informado.

## REFERÊNCIAS

[1] Tim J Berners-Lee. The world-wide web. *Computer networks and ISDN systems*, 25(4-5):454–459, 1992.
[2] Ray Tomlinson. The first network email. *Ray Tomlinson website*, 2009.
[3] Nicolas Ducheneaut and Victoria Bellotti. E-mail as habitat: An exploration of embedded personal information management. *interactions*, 8(5):30–38, September 2001. ISSN 1072-5520. doi: 10.1145/382899.383305. URL http://doi.acm.org/10.1145/382899.383305.
[4] David Chadwick. Federated identity management. *Foundations of Security Analysis and Design V*, pages 96–120, 2009.
[5] L. Jin, H. Takabi, and J. B. D. Joshi. Security and privacy risks of using e-mail address as an identity. In *2010 IEEE Second International Conference on Social Computing*, pages 906–913, Aug 2010. doi: 10.1109/SocialCom.2010.134.
[6] SparkPost. Transactional email - benchmark report, 2018. URL https://www.sparkpost.com/resources/white-papers-guides/transactional-email-benchmark-report.
[7] Presidência da República. Lei Geral de Proteção de Dados Pessoais (LGPDP) - Nº 13.709, agosto 2018.
[8] Maarten van Steen and Andrew S Tanenbaum. *Distributed systems*. Maarten van Steen, 3 edition, fev 2017. ISBN 978-90-815406-2-9.
[9] Audun Jøsang, Muhammed Al Zomai, and Suriadi Suriadi. Usability and privacy in identity management architectures. In *Proceedings of the fifth Australasian symposium on ACSW frontiers-Volume 68*, pages 143–152. Australian Computer Society, Inc., 2007.
[10] NN/g Nielsen Norman Group. Transactional email and confirmation messages, 2020. URL https://www.nngroup.com/reports/ecommerce-transactional-email-confirmation-message.
[11] Tatiana Engel Gerhardt and Denise Tolfo Silveira, editors. *Métodos de pesquisa*. Editora da UFRGS, 2009. URL http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf. ISBN: 978-85-386-0071-8.
[12] Peter Schaar. Privacy by design. *Identity in the Information Society*, 3(2):267–274, 2010.
[13] James Reason. *Human error*. Cambridge university press, 1990.
[14] Mark Weiser. The computer for the 21st century. *Scientific american*, 265(3): 94–104, 1991.
[15] IBGE. Nomes mais populares no censo demográfico 2010, 2010. URL https://censo2010.ibge.gov.br/nomes/#/ranking.
[16] Fatma Al Maqbali and Chris J Mitchell. Web password recovery—a necessary evil? *arXiv preprint arXiv:1801.06730*, 2018.
[17] Luis Von Ahn, Manuel Blum, Nicholas J Hopper, and John Langford. Captcha: Using hard ai problems for security. In *International Conference on the Theory and Applications of Cryptographic Techniques*, pages 294–311. Springer, 2003.
[18] Juliana Graciela dos Santos, Alexandre Cappellozza, and Alberto Luiz Albertin. Antecedents of perceived benefits of compliance towards organizational data protection policies. *IEEE Latin America Transactions*, 16(3):891–896, 2018.
[19] MF. Ajuste sinief 19, nota fiscal de consumidor eletrônica, modelo 65, 2016. URL https://www.confaz.fazenda.gov.br/legislacao/ajustes/2016/AJ_019_16.
[20] Mary J. Culnan and Cynthia Clark Williams. How ethics can enhance organizational privacy: Lessons from the choicepoint and tjx data breaches. *MIS Quarterly*, 33:673–687, 2009.

## A AMOSTRA DOS SITES ANALISADOS

**Tabela 1: Amostra dos sites ordenados pela posição do ranking do serviço Alexa Internet em abril de 2018**

| # | Empresa | Categoria |
|---|---|---|
| 4 | facebook.com | Mundo digital |
| 7 | mercadolivre.com.br | Mundo físico |
| 10 | instagram.com | Mundo digital |
| 13 | netflix.com | Mundo digital |
| 15 | olx.com.br | Mundo físico |
| 18 | twitter.com | Mundo digital |
| 20 | americanas.com.br | Mundo físico |
| 23 | microsoft.com | Mundo digital |
| 27 | aliexpress.com | Mundo físico |
| 44 | pinterest.com | Mundo digital |
| 53 | tumblr.com | Mundo digital |
| 54 | gearbest.com | Mundo físico |
| 61 | linkedin.com | Mundo digital |
| 76 | magazineluiza.com.br | Mundo físico |
| 80 | saraiva.com.br | Mundo físico |
| 83 | submarino.com.br | Mundo físico |
| 89 | amazon.com.br | Mundo físico |
| 97 | netshoes.com.br | Mundo físico |
| 101 | casasbahia.com.br | Mundo físico |
| 119 | vagas.com.br | Mundo físico |
| 123 | dropbox.com | Mundo digital |
| 124 | minhateca.com.br | Mundo digital |
| 128 | extra.com.br | Mundo físico |
| 136 | github.com | Mundo digital |
| 138 | kabum.com.br | Mundo físico |
| 140 | spotify.com | Mundo digital |

| | | |
|---|---|---|
| 145 | elo7.com.br | Mundo físico |
| 157 | apple.com | Mundo físico |
| 159 | ebay.com | Mundo físico |
| 161 | shoptime.com.br | Mundo físico |
| 175 | dafiti.com.br | Mundo físico |
| 177 | pontofrio.com.br | Mundo físico |
| 183 | decolar.com | Mundo físico |
| 185 | webmotors.com.br | Mundo físico |
| 199 | booking.com | Mundo físico |
| 216 | voegol.com.br | Mundo físico |
| 217 | infojobs.com.br | Mundo físico |
| 218 | walmart.com.br | Mundo físico |
| 222 | grancursosonline.com.br | Mundo digital |
| 228 | mercadopago.com.br | Mundo físico |
| 235 | serasaconsumidor.com.br | Mundo digital |
| 246 | estantevirtual.com.br | Mundo físico |
| 247 | drogaraia.com.br | Mundo físico |
| 252 | trello.com | Mundo digital |
| 258 | tripadvisor.com.br | Mundo digital |
| 265 | steampowered.com | Mundo digital |
| 266 | latam.com | Mundo físico |
| 269 | udemy.com | Mundo digital |
| 273 | popcash.net | Mundo digital |
| 274 | leagueoflegends.com | Mundo digital |
| 282 | natura.net | Mundo físico |
| 290 | avon.com.br | Mundo físico |
| 336 | wixsite.com | Mundo digital |
| 341 | shutterstock.com | Mundo digital |
| 346 | dell.com | Mundo físico |
| 347 | icarros.com.br | Mundo físico |
| 350 | carrefour.com.br | Mundo físico |
| 356 | viralcpm.com | Mundo digital |
| 364 | catho.com.br | Mundo físico |
| 369 | centauro.com.br | Mundo físico |
| 371 | zapimoveis.com.br | Mundo digital |
| 375 | avast.com | Mundo digital |
| 382 | descomplica.com.br | Mundo digital |
| 387 | samsung.com | Mundo físico |
| 401 | zattini.com.br | Mundo físico |
| 405 | issuu.com | Mundo digital |
| 419 | voeazul.com.br | Mundo físico |
| 421 | enjoei.com.br | Mundo físico |
| 429 | ricardoeletro.com.br | Mundo físico |
| 454 | alibaba.com | Mundo físico |
| 463 | airbnb.com.br | Mundo físico |
| 465 | ifood.com.br | Mundo físico |
| 472 | behance.net | Mundo digital |
| 474 | shopfacil.com.br | Mundo físico |
| 479 | bitly.com | Mundo digital |
| 488 | kanui.com.br | Mundo físico |
| 489 | pichau.com.br | Mundo físico |
| 495 | leroymerlin.com.br | Mundo físico |