Ministério da Justiça

Termo de Cooperação/Projeto:

Acordo de Cooperação Técnica FUB/CDT e MJ/SE Registro de Identidade Civil – Replanejamento e Novo Projeto Piloto

Documento:

RT Infraestrutura Tecnológica: Geração de Número RIC

Data de Emissão:

09/02/2015

Elaborado por:

Universidade de Brasília – UnB
Centro de Apoio ao Desenvolvimento Tecnológico – CDT
Laboratório de Tecnologias da Tomada de Decisão – LATITUDE.UnB

MINISTÉRIO DA JUSTIÇA

José Eduardo Cardozo
Ministro

Marivaldo de Castro Pereira
Secretário Executivo

Helvio Pereira Peixoto
Coordenador Suplente do Comitê Gestor do SINRIC

UNIVERSIDADE DE BRASÍLIA

Ivan Marques Toledo Camargo
Reitor

Paulo Anselmo Ziani Suarez
Diretor do Centro de Apoio ao Desenvolvimento Tecnológico – CDT

Rafael Timóteo de Sousa Júnior
Coordenador do Laboratório de Tecnologias da Tomada de Decisão – LATITUDE

EQUIPE TÉCNICA
Ana Maria da Consolação Gomes Lindgren
Andréa Benoliel de Lima
Celso Pereira Salgado
Delluiz Simões de Brito
Elaine Fabiano Tocantins
Fernando Saliba Oliveira
Fernando Teodoro Filho
Guilherme Braz Carneiro
Joaquim de Oliveira Machado
José Alberto Sousa Torres
Marcelo Martins Villar
Raphael Fernandes de Magalhães Pimenta
Rodrigo Borges Nogueira
Rodrigo Gurgel Fernandes Távora
Sara Lais Rahal Lenharo

EQUIPE TÉCNICA
Flávio Elias Gomes de Deus (Pesquisador Sênior)
William Ferreira Giozza (Pesquisador Sênior)
Ademir Agustinho de Rezende Lourenço
Adriana Nunes Pinheiro
Alessandro Zimmer
Alysson Fernandes de Chantal
Amanda Almeida Paiva
Andréia Campos Santana
Antônio Claudio Pimenta Ribeiro
Carolinne Januária de Souza Martins
Caio Rondon Botelo de Carvalho
Daniela Carina Pena Pascual
Danielle Ramos da Silva
Diogenes Ferreira Reis Fustinoni
Eduarda Simões Veloso Freire
Fábio Lúcio Lopes Mendonça
Fábio Mesquita Buiati
Glaidson Menegazzo Verzeletti
Heverson Soares de Brito
Johnatan Santos de Oliveira
José Carneiro da Cunha Oliveira Neto
José Elenilson Cruz
Kelly Santos de Oliveira Bezerra
Luciano Pereira dos Anjos
Luciene Pereira de Cerqueira Kaipper
Luiz Antônio de Souto Evaristo
Luiz Claudio Ferreira
Marcos Vinicius Vieira da Silva
Marco Schaffer
Mirele Maria Cavalcante Rocha
Pedro Augusto Oliveira de Paula
Renata Elisa Medeiros Jordão
Roberto Mariano de Oliveira Soares
Sandro Augusto Pavlik Haddad
Sergio Luiz Teixeira Camargo
Soleni Guimarães Alves
Suzane Lais De Freitas
Valério Aymoré Martins
Vera Lopes de Assis
Vinicius de Morais Alves
Wladimir Rodrigues da Fonseca

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

HISTÓRICO DE REVISÕES

Data	Versão	Descrição
09/02/2015	0.1	Versão inicial.

latitude
Laboratório de tecnologias da tomada de decisão

Universidade de Brasília – UnB
Campus Universitário Darcy Ribeiro - FT – ENE – Latitude
CEP 70.910-900 – Brasília-DF
Tel.: +55 61 3107-5598 – Fax: +55 61 3107-5590

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

SUMÁRIO

1 INTRODUÇÃO 6
2 HISTÓRICO DAS ESTRATÉGIAS DE NUMERAÇÃO DE IDENTIDADES CIVIS 8
2.1 No Brasil 8
2.1.1 CPF 8
2.1.2 Título Eleitoral 9
2.1.3 Carteira Nacional de Habilitação 9
2.1.4 Documento de Identidade emitido pelas Unidades da Federação 9
2.2 No Mundo 9
2.2.1 África do Sul 10
2.2.2 Áustria 10
2.2.3 Bélgica 10
2.2.4 Canadá 11
2.2.5 Coréia do Sul 11
2.2.6 Estados Unidos 11
2.2.7 Estônia 12
2.2.8 Índia 12
2.2.9 México 13
3 REQUISITOS DO NÚMERO RIC 14
3.1 Codificação 14
3.2 Longevidade 15
3.3 Formatação 16
3.4 Dígito Verificador 17
3.5 Privacidade 18
3.5.1 Pseudônimos 20
3.6 Regras de Exclusão 21
3.7 Aleatoriedade 22
3.8 Definição de Estados (*status*) 22
3.9 O Número RIC e Chaves em Bancos de Dados 23
3.10 Estratégias de Geração do Número RIC 24
3.10.1 Distribuição de Números RIC 24
3.10.2 Geração Descentralizada do Número RIC 25
4 GENERALIZAÇÃO DO NÚMERO RIC 25
4.1 Um Identificador Para Entidades 26
4.2 Reserva de Faixa Numérica 26

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

4.3 Identificadores Temporários 27
4.4 Considerações sobre os Domínios dos Identificadores 28
5 CONCLUSÃO 29
6 REFERÊNCIAS 30

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

1 INTRODUÇÃO

A Secretaria Executiva (SE/MJ), vinculada ao Ministério da Justiça (MJ), é responsável por viabilizar o desenvolvimento e a implantação do Registro de Identidade Civil, instituído pela Lei nº 9.454, de 7 de abril de 1997, regulamentado pelo Decreto nº 7.166, de 5 de maio de 2010.

Atualmente, a República Federativa do Brasil conta com sistema de identificação de seus cidadãos amparado pela Lei nº 7.116, de 29 de agosto de 1983. Essa lei assegura validade nacional às Carteiras de Identidade, ou Cédulas de Identidade; confere também autonomia gerencial às Unidades Federativas no que concerne à expedição e controle dos números de registros gerais emitidos para cada documento. Essa condição de autonomia, ao contrário do que pode parecer, fragiliza o sistema de identificação, já que dá condições ao cidadão de requerer legalmente até 27 (vinte e sete) cédulas de identidades diferentes. Com essa facilidade legal, inúmeras possibilidades fraudulentas se apresentam de maneira silenciosa, pois, na grande maioria dos casos, os Institutos de Identificação das Unidades Federativas não dispõem de protocolos e aparato tecnológico para identificar as duplicações de registro vindas de outros estados, ou até mesmo do seu próprio arquivo datiloscópico. Consoante aos fatos, os Institutos de Identificação não trabalham interativamente para que haja trocas de informações de dados e geração de conhecimento para manuseio inteligente e seguro para individualização do cidadão em prol da sociedade.

Com foco na busca de soluções para tais problemas, o Projeto RIC prevê a administração central dos dados biográficos e biométricos dos cidadãos no Cadastro Nacional de Registro de Identificação Civil (CANRIC) e ABIS (do inglês *Automated Biometric Identification System*), respectivamente. A previsão desse novo modelo sustenta a não duplicação de registros e a consequente identificação unívoca dos cidadãos brasileiros natos e naturalizados. O Projeto RIC, portanto, visa otimizar o sistema de identificação e individualização do cidadão brasileiro nato e naturalizado com vistas a um perfeito funcionamento da gestão de dados da sociedade, agregando valor à cidadania, à gestão administrativa, à simplificação do acesso aos serviços disponíveis ao cidadão e à segurança pública do país.

Nesse contexto, o termo de cooperação entre MJ/SE e FUB/CDT define um projeto que objetiva identificar, mapear e desenvolver parte dos processos e da infraestrutura tecnológica necessária para viabilizar a implantação do número único de Registro de

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

Identidade Civil – RIC no Brasil.

Resultante de um subconjunto das atividades previstas para inicialização da cooperação MJ/SE e FUB/CDT, o presente documento apresenta uma primeira abordagem para o problema da geração de um identificador para a identidade nacional única a ser proposta pelo RIC. Analisaremos diferentes aspectos que podem ser considerados na escolha de uma estratégia de numeração, apresentando diretrizes para a subsidiar essa escolha.

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

2 HISTÓRICO DAS ESTRATÉGIAS DE NUMERAÇÃO DE IDENTIDADES CIVIS

No Brasil e no mundo, diferentes mecanismos foram e são utilizados para a geração de identificadores civis nacionais. Nesta seção apresentamos alguns exemplos e/ou fatos que suportem escolhas no projeto de um identificador único para o RIC.

2.1 NO BRASIL

Atualmente, existem no Brasil diversos tipos de documentos e cadastros. A criação de cada um deles é, em geral, independente entre os órgãos que os propõem e a maioria visa atender às necessidades destes órgãos ou identificar civilmente uma classe profissional, por exemplo.

Embora exista um documento isento de finalidade classista ou de destinação negocial específica, o RG, *Registro Geral*, criado especificamente para a identificação civil, com validade nacional e emitido por cada Unidade da Federação, não figura como a única chave de identificação de uma pessoa referenciada em sistemas manuais ou computacionais de entidades públicas e privadas. A falta de padronização nos números do RG, aliada à possibilidade de se obter múltiplos documentos válidos, ou com validade não passível de verificação, influenciaram a ampla adoção do CPF, *Cadastro de Pessoa Física*, como um dos identificadores preferenciais em diferentes bases de dados.

O CPF, criado para identificação de contribuintes junto à Receita Federal, é um número de identificação amplamente utilizado no Brasil pela população junto a provedores de serviços, sejam estes prestados digitalmente através da internet ou fisicamente. [1]

Os números utilizados pelo CPF são divididos da seguinte forma:

- algarismos da primeira à oitava posição: algarismos gerados pela Receita Federal formando um número único por pessoa física, em uma região fiscal;
- algarismo na nona posição: região fiscal, de um total de dez regiões fiscais;
- algarismos na décima e décima primeira posição: dígitos verificadores, calculados por algoritmo módulo 11.

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

2.1.2 Título Eleitoral

O mecanismo em vigor de composição do número do Título Eleitoral Brasileiro é regulamentado pela Resolução nº 21.538, de 14 de outubro de 2003 [2], sendo da seguinte forma:

- os oito primeiros algarismos serão sequenciados;
- os dois algarismos seguintes serão representativos da unidade da Federação de origem da inscrição;
- os dois últimos algarismos constituirão dígitos verificadores, calculados por algoritmo módulo 11.

2.1.3 Carteira Nacional de Habilitação

Conforme a Resolução nº 192, de 30 de março de 2006 [3], do CONTRAN, o Número de Registro que compõe a Carteira Nacional de Habilitação (CNH) é formado por:

- nove algarismos que constituem um número único para cada habilitação;
- dois algarismos que constituem dígitos verificadores, calculados por algoritmo módulo 11.

2.1.4 Documento de Identidade emitido pelas Unidades da Federação

Pela legislação em vigor, Lei nº 7.116, de 29 de agosto de 1983 [4], cada UF, Unidade da Federação, é responsável pela emissão da cédula de identidade com um *número de registro geral* próprio, com garantia de unicidade apenas na UF emissora. Por este motivo, esta é geralmente informada juntamente com o número de registro geral, quando da utilização da cédula de identidade, para que o par (UF; número de registro geral) seja único no Brasil.

Cada Unidade da Federação pode estabelecer regras próprias para a geração do número de registro geral.

2.2 NO MUNDO

Os países a serem analisados nesta seção podem emitir diversos diferentes documentos civis, mas citaremos apenas o mais utilizado em cada país. Como o objetivo deste RT é a criação de uma proposta de identificador único para o RIC, nos restringiremos a mostrar apenas os identificadores utilizados nos países relacionados e/ou fatos relevantes que nos suportem na tomada de decisões relativas a como construir um identificador único RIC.

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

2.2.1 África do Sul

Durante o período do *Apartheid*, a África do Sul introduziu um documento de identidade para maiores de dezoito anos, no qual o número de identificação trazia, dentre outros dados, informação relativa ao *grupo racial* do qual uma pessoa fazia parte. Essa informação, após o fim do regime de segregação, não aparece mais nos documentos emitidos atualmente.

O formato padrão do número da identidade sul-africana, o qual possui 13 dígitos, é da forma $YYMMDDGSSSCAZ$ [5], onde:

- $YYMMDD$: data de nascimento;
- G: sexo;
- SSS: número sequencial;
- C: informa se a pessoa é ou não um cidadão sul-africano;
- A: valor predefinido que, geralmente, é 8 ou 9. Este campo já foi usado, durante o *Apartheid*, para codificar o grupo racial;
- Z: dígito verificador, calculado segundo o *algoritmo de Luhn* ou *módulo 10*.

O *Registro Central de Residentes* mantém os registros e número de identidade associado. Este número é utilizado para gerar um outro número de identificação que fica armazenado em um cartão de identidade com *chip*. Este último número não é utilizado, servindo apenas como semente para a geração de números de identificação específicos (pseudônimos) para diferentes setores governamentais, tais como, saúde, educação, impostos, etc. [6]

O uso de pseudônimos em diferentes tipos de serviços foi uma solução necessária devido à resistência da população em aceitar a utilização de um número único nas diferentes organizações públicas, o que poderia acarretar o relacionamento de informações entre os diferentes órgãos do governo.

Entretanto, o modelo austríaco é frequentemente questionado acerca de sua complexidade e até mesmo do retorno sobre o investimento [7].

Na Bélgica, o número identificador é conhecido como *National Registry Number*, possui onze dígitos, sendo composto pela data de nascimento do cidadão, um número sequencial

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

que também codifica o sexo e o dígito verificador [8].

O número é considerado uma informação sensível à privacidade, pois contém a data de nascimento que pode ser imediatamente lida no número nacional de registro.

O *Social Insurance Number* (*SIN*) possui formato $999\text{-}999\text{-}999$, sendo que os identificadores iniciados com 9 são destinados a residentes temporários. O primeiro número do SIN é a região onde se fez o registro e o último é o dígito verificador [9].

2.2.5 Coréia do Sul

O *Número de Registro de Residente* é da forma $YYMMDD\text{-}GRRRRSZ$, sendo:

- $YYMMDD$: data de nascimento;
- G: codifica o sexo e século de nascimento;
- $RRRR$: região de nascimento;
- S: número serial;
- Z: dígito verificador, calculado utilizando aritmética modular.

Em uma publicação feita pelo site *Mail Online* [10], foi noticiada uma grave crise de vazamento sistemático de dados, na qual o documento de identidade sul-coreano aparece como uma das causas. O número de identificação pessoal, que contém diversas informações de seu proprietário, em conjunto com o nome deste, permitia que um adversário habilitasse números telefônicos e até mesmo contas bancárias.

2.2.6 Estados Unidos

O *Social Security Number*, SSN, é um identificador de nove dígitos composto de três partes, com formato $999\text{-}99\text{-}9999$. Os três grupos numéricos, na primeira versão do SSN, significavam, respectivamente, o número de área, número de grupo e um número serial [11].

Em 2009, o jornal *Washington Post* [12] mostrou pesquisadores da *Carnegie Mellon University* afirmando que o SSN estava comprometido como uma informação sensível pessoal, já que poderia ser predito a partir de dados públicos. Além disso, a notícia acrescentou que diversos estados americanos já tinham legislação relativa à reedição ou remoção do SSN em documentos públicos.

Em 2011, a *Social Security Administration,* SSA, alterou a forma de se gerar o SSN, passando a utilizar um método chamado de *randomização* (do inglês *randomization*). Os

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

principais motivos para a mudança são a preservação da integridade do número e a melhor utilização do espaço de números possíveis para atribuição. Neste último caso, regiões onde o crescimento populacional é maior, emitem mais documentos, tendendo a esgotar mais rapidamente as possibilidades numéricas, uma vez que a região era codificada como parte do SSN.

A Estônia se destaca no cenário internacional pela avançada infraestrutura de serviços eletrônicos disponibilizados a partir da implantação de seu projeto de e-ID [13].

O número de identificação pessoal estoniano apresenta apenas caracteres numéricos e é da forma $GYYMMDDSSSC$, onde:

- G: campo que codifica o sexo e o século de nascimento, sendo que G é ímpar para o sexo masculino e par para feminino. Assim, para nascidos no século 19 $G \in \{1, 2\}$, século 20 $G \in \{3, 4\}$ e século 21 $G \in \{5, 6\}$;
- $YYMMDD$: ano, mês e dia de nascimento;
- SSS: número serial;
- C: dígito verificador, calculado por uma operação do tipo $mod\ 11$.

Como o número de identidade estoniano tem em sua composição informações pessoais, a construção do número pode ser facilmente deduzida a partir do conhecimento de tais informações. Em uma publicação feita pelo *Washington Post* [14], foi noticiada que uma ferramenta foi implementada para viabilizar a construção de números de identidade e sua consulta em sites de busca.

O projeto de identidade pessoal indiano é um dos maiores do mundo em número de pessoas a serem identificadas e um consistente trabalho de escolha da estratégia de numeração do identificador único foi descrito em [15].

No Projeto de identidade única da Índia, o UID (*Unique ID*) possui doze algarismos, $a_1a_2a_3a_4a_5a_6a_7a_8a_9a_{10}a_{11}a_{12}$, onde:

- a_1: número de versão, com valores 0 e 1 reservados e valores 2 a 9 escolhidos aleatoriamente para compor o UID. O valor 0 pode ser utilizado, por exemplo, como um caractere sinalizador para um novo tamanho do UID. Já o valor 1 foi reservado para entidades;

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

- $a_2a_3a_4a_5a_6a_7a_8a_9a_{10}a_{11}$: valores gerados aleatoriamente;
- a_{12}: dígito verificador, calculado pelo esquema de Verhoeff.

2.2.9 México

O documento de identidade nacional mexicano é a CURP, *Clave Única de Registro de Población*. Este se utiliza de uma sequência de dezoito caracteres alfanuméricos para compor a chave única de identificação da população.

Uma característica peculiar da identidade mexicana, não encontrada em outros países estudados, é a utilização de letras do nome e do local de nascimento na composição do identificador, resultando em um esquema que utiliza pelo menos dez letras [16].

$$NNNNYYMMDDGLLNNNHH$$

Sendo:

- $NNNN$ e NNN: estes dois grupos, que aparecem separados na composição do identificador, são formados por **letras** específicas do nome;
- $YYMMDD$: valores numéricos, sendo, respectivamente, ano, mês e dia de nascimento;
- G: sexo, sendo M para mulheres e H para homens;
- LL: local de nascimento, codificado por valor alfabético;
- HH: valores atribuídos pela autoridade de identificação para tratar ambiguidades.

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

3 REQUISITOS DO NÚMERO RIC

Privacidade está entre os principais requisitos do e-Gov, como mostra estudo do *Oxford Internet Institute*, que elencou, dentre sete barreiras chave do e-Gov, a falta de confiança e salvaguardas inadequadas de segurança e privacidade [17]. Um usuário de um documento de identidade precisa ter, tanto quanto possível, garantias de que sua privacidade é considerada importante desde o projeto de sistemas, diretrizes, leis, e muitos outros elementos nos quais a privacidade pode ser tratada. Este é um tema que vai muito além do escopo deste relatório, uma vez que privacidade não é tão somente um aspecto técnico que trata de mecanismos de proteção de dados. Neste documento, entretanto, o objetivo é considerar a não exposição de informações pessoais a partir do conhecimento de um identificador único a ser proposto.

Nos tópicos a seguir, elencamos diversos aspectos considerados na escolha da estratégia de construção do identificador único para o RIC. Cada característica descrita contribui de alguma forma com as decisões de projeto para a estratégia de geração desse identificador.

3.1 CODIFICAÇÃO

Um identificador em geral pode ser codificado de diversas formas, usando diferentes alfabetos e/ou representações numéricas. A fim de encontrar uma codificação apropriada para o identificador único RIC, consideremos que este se destina a ser usado e memorizado pelas pessoas no cotidiano, em um grande número de aplicações e situações. Assim, reconhecemos a importância de se manter o identificador o mais curto possível.

Dentre as situações mais comuns de utilização de um identificador único, podemos exemplificar:

- apresentação do identificador impresso, de uma pessoa para outra, seja em um documento de identidade oficial, na tela de um telefone celular ou em uma anotação manual;
- por telefone ou qualquer outro canal via comunicação verbal, no qual o identificador é falado para a outra parte;
- pela digitação em um teclado, seja de telefone ou em um terminal bancário de autoatendimento, por exemplo.

Embora a utilização de um identificador alfanumérico possa resultar em uma sequência mais curta para memorização, a codificação apenas com números $(0,1,2,3,4,5,6,7,8,9)$ é

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

mais vantajosa, pois torna mais fácil a comunicação do identificador entre duas partes. Por exemplo, a digitação de sequências numéricas em teclados telefônicos ou terminais com teclados numéricos é significativamente mais simples que informar letras.

A representação numérica é mais fácil de ser reconhecida pela população em geral. Letras podem representar dificuldades para analfabetos funcionais. Além disso, a simples verbalização de letras pode também ser um obstáculo, seja pelas diferenças de sotaques das diversas regiões do país, seja pela semelhança fonética de letras como “P”, “B” e “D”, por exemplo. Esse obstáculo de comunicação é, inclusive, reconhecido e contornado por soluções como o *Voice Procedure*, com um conjunto de técnicas que visam padronizar comunicações faladas para fins diversos, de modo que os interlocutores possam se beneficiar da comunicação falada de um modo mais simples e claro. Particularmente, o entendimento de letras individuais é tratado pelo uso de métodos como o *Alfabeto de Radiotelefonia* da *ICAO* [18], no qual cada letra é pronunciada como uma palavra pré-definida de modo a clarificar o processo de comunicação verbal. Exemplo: A → Alpha, B → Bravo, etc.

3.2 LONGEVIDADE

Como um identificador civil nacional, o número RIC deve possuir um domínio grande o suficiente para identificar toda a população atual e acompanhar o crescimento populacional por um longo período de tempo, o que consideraremos na ordem de séculos. Ao mesmo tempo, o tamanho do número deve ser conveniente para o uso pelas pessoas, inclusive para memorização.

Conforme dados do IBGE [19], a população brasileira é composta por mais de duzentos milhões de habitantes, com projeção de mais de duzentos e vinte milhões para o ano de 2030. Podemos estabelecer uma estimativa que, mesmo sendo imprecisa, possa subsidiar uma apropriada escolha da quantidade de números RIC disponíveis para atribuição aos residentes ou cidadãos brasileiros de modo único. Para tanto, consideremos a taxa bruta de natalidade no Brasil que, em 2000, segundo o IBGE, foi de 20,86 nascidos por mil habitantes, com queda anual, chegando a 14,47 em 2014. Para fins de estimativa da ordem de grandeza de números RIC necessários em um período arbitrário de 100 anos, consideraremos uma taxa bruta de natalidade constante e igual a 20 nascidos por mil habitantes, tomando como população base o número de habitantes atual, que é de aproximadamente duzentos milhões de pessoas. Assim, o número total de indivíduos

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

nascidos no Brasil, de 2015 a 2115, será de menos de um bilhão e meio de pessoas, considerando neste grupo também a população atual. Ressaltamos que o cálculo não é para *estimar* a população do Brasil em 2115, mas sim o número bruto de nascidos somado à população atual. Adicionalmente, pelo superdimensionamento dado aos parâmetros, consideramos ser razoável que números relativos à imigração também estejam já considerados.

$$n = p_{atual}.(1+i)^a = 200.000.000 \times (1+0{,}02)^{100} = 1.448.929.223$$

Onde:

n: número de nascidos mais população atual até 2115;

p_{atual}: número aproximado de habitantes no Brasil em 2015;

i: taxa de crescimento populacional que, neste caso, consideramos uma aproximação da taxa bruta de natalidade;

a: período de tempo de 100 anos.

Desta forma, desconsiderando o dígito verificador, consideramos que seja uma escolha adequada que o número RIC contenha dez dígitos, o que garantiria dez bilhões de diferentes possibilidades. A utilização de nove dígitos também é possível, pois, vale notar que, por se tratar de um número, a expansão do identificador para um domínio maior é viável, bastando acrescentar um dígito quando a quantidade de números utilizados ultrapassar uma certa margem de utilização. Além disso, a compatibilidade com sistemas, quando da alteração na quantidade de dígitos, é relativamente simples de ser mantida.

Uma vantagem na utilização de dez dígitos, ao invés de nove, é que a geração de números fica mais esparsa, ou seja, reduz-se a possibilidade de se encontrar um número RIC válido e ativo ao acaso.

3.3 FORMATAÇÃO

Na notação numérica é de conhecimento geral a representação de números grandes com os algarismos agrupados três a três, como em $999.999.999$, sendo que cada grupo de algarismos é separado por pontos. Tipicamente em língua inglesa, os grupos de milhares são separados por vírgulas, como em $999{,}999{,}999$. A representação numérica em grupos de quatro algarismos também é comumente conhecida, como em números telefônicos e de cartões de crédito, por exemplo.

Deste modo, utilizando o senso comum a respeito de representações numéricas gerais e para evitar confusão com a representação atual de outros documentos brasileiros, em

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

especial, do CPF, no formato 999.999.999-99, sugerimos que o número RIC seja representado em grupos de no máximo quatro algarismos, separados por hífen para um melhor destaque dos grupos numéricos, como mostrado a seguir:

9999-9999-999

A escolha por grupos de quatro algarismos pode ser considerada apropriada também do ponto de vista da psicologia, em estudos acerca da memória de curta duração, especialmente a partir de Cowan, 2001 [20], cujo trabalho diz, por exemplo, que a disposição de números de telefone em grupos de três ou quatro algarismos pode ser um indício de quantos elementos podem ser confortavelmente mantidos como foco de atenção ao mesmo tempo. Assim, considerando o formato proposto acima, podemos dizer que, ao se comunicar verbalmente um número RIC (o que é um caso de uso comum em um documento de identificação), o receptor terá, na grande maioria das vezes, condições de entender, transcrever, memorizar e analisar cada grupo numérico. Além disso, a verbalização do número RIC é mais rápida ao se considerar um grupo numérico de quatro algarismos ao invés de três, uma vez que se reduz o número de pausas entre grupos. Por exemplo, a comunicação verbal de um número no formato 99-99-99-99-99-9, geralmente, resulta na utilização de cinco pausas, enquanto com o número RIC se utilizam duas pausas.

3.4 DÍGITO VERIFICADOR

O dígito verificador (DV) é um algarismo que introduz informação redundante a uma sequência de algarismos, com a finalidade de permitir a verificação da integridade desta sequência, detectando a existência de erros diversos, mas com limitações que variam com o tipo de esquema utilizado.

No contexto de um documento de identificação, o dígito verificador tem a função de detectar erros de digitação, mas também pode auxiliar na detecção de fraudes que utilizem um dígito verificador inconsistente. Entretanto, é importante salientar que esse dígito não deve ser considerado uma forma confiável de identificação de fraudes, pois o método de cálculo é de domínio público.

O valor do dígito verificador é computado e existem atualmente diversos algoritmos disponíveis, sendo mais comum, nos documentos brasileiros, a utilização de esquemas baseados em aritmética modular com bases 10 ou 11. A escolha destes esquemas foi realizada em um período onde a capacidade computacional era bastante limitada

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

relativamente aos dias atuais. Mesmo que outras opções mais eficazes existissem, a simplicidade no funcionamento e implementação eram requisitos determinantes.

Para o RIC, recomendamos a adoção do esquema de Verhoeff [21] que, embora seja mais complexo que os demais esquemas utilizados nos documentos brasileiros, atualmente pode ser seguramente utilizado, considerados o poder computacional disponível e acesso a implementações gratuitas em linguagens de programação diversas.

Os detalhes e fundamentos relativos ao esquema de Verhoeff e outros podem ser encontrados em [22]. Mostraremos apenas as propriedades que suportam nossa escolha de cálculo do dígito verificador do número RIC. É relevante considerar que o projeto indiano de identidade nacional única também selecionou o esquema de Verhoeff [15].

Um adequado dígito verificador para uso em um número de identificação civil deve detectar os erros mais comuns cometidos quando da digitação de uma sequência numérica [23]. Na tabela a seguir, estão elencados os tipos de erros mais comuns, suas frequências e a resposta do esquema de Verhoeff a cada tipo.

Tipo de erro	Forma	Frequência	Frequência de detecção com Verhoeff
Dígito simples	$a \rightarrow b$	79,1%	100%
Transposição adjacente	$ab \rightarrow ba$	10,2%	100%
Transposição alternada	$abc \rightarrow cba$	0,8%	94%
Gêmeos	$aa \rightarrow bb$	0,5%	95%
Gêmeos alternados	$aca \rightarrow bcb$	0,3%	95%

Tabela 1: Frequências de erros de digitação e detecção pelo algoritmo de Verhoeff

3.5 PRIVACIDADE

Um identificador civil pode ser considerado como um atributo pessoal, assim como a data de nascimento, nome e outros números de documentos que identificam uma pessoa [8]. Desta forma, a informação contida no número RIC deve ser, preferencialmente, independente da informação contida em outros atributos pessoais, para que cada atributo seja utilizado sempre que necessário, sem revelar nenhuma informação não solicitada de outro atributo. Por exemplo, conforme mostrado na seção 2, existem identificadores civis que trazem em suas composições um conjunto de informações pessoais que nem mesmo o portador do número de identidade pode estar ciente. Como descrito na seção 2.2.6, nos Estados Unidos, o *Social Security Number* poderia ser deduzido a partir de informações

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

públicas, o que potencializa as ocorrências de fraudes, onde um adversário utilizaria o número de um documento oficial associado a outras informações pessoais, mesmo sem ter acesso legítimo à informação.

Em consonância com legislações e convenções internacionais, por exemplo, [24], [25], [26], [27], o Brasil tem desenvolvido o suporte legal que estabelece princípios, regras e limites no uso de dados pessoais, como mostra o *Anteprojeto de Lei de Proteção de Dados Pessoais* [28]. Dentre os princípios, destacamos o *Princípio da Minimização de Dados*, que envolve a *necessidade*, *propósito* e *consentimento* relativos à abertura de uma informação pessoal, que pode ser utilizada se for **necessária** em uma operação qualquer e apenas para a **finalidade** para a qual a informação foi provida de modo **consciente** pelo seu proprietário.

Quanto ao número RIC, considerada a tendência mundial de se regulamentar e apresentar diretivas para o uso de dados pessoais, recomenda-se que o número não tenha em sua estrutura qualquer tipo de significado semântico. Caso contrário, conforme exemplificado na seção 2.2, problemas na privacidade dos dados pessoais podem ser agravados pelo uso de informações de um indivíduo na construção de um número de identidade.

A presença de informações como data de nascimento, sexo ou local de nascimento na formação do número RIC implica na impossibilidade de o usuário de um serviço qualquer avaliar se esses dados pessoais poderão ser ou não revelados, quando da solicitação do número RIC para operações de autenticação, por exemplo. Uma vez que o princípio da *necessidade* da informação de um dado pessoal se concretize como determinação legal no Brasil, o uso de informações pessoais na formação do identificador se torna uma proibição, já que a utilização do número RIC nos mais diversos serviços nem sempre vai requerer as informações supostamente contidas no número.

Ao se utilizar informações pessoais na construção de um identificador, um outro impacto emerge, que é a provável necessidade de se empregar uma maior quantidade de algarismos que compõem o número, uma vez que informações com significado semântico tendem a não utilizar todo o espaço de possibilidades numéricas para sua representação, reduzindo, assim, o tamanho do domínio de todos os números de identidade possíveis com tamanho $n \in \mathbb{N}$. Por exemplo, se a data de nascimento está presente, é necessária a existência de uma parte aleatória/sequencial no número de identidade para distinção de todos os indivíduos nascidos em um dia qualquer. Essa parte variável precisa ser grande o

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

suficiente para que seja desprezível a probabilidade de nascimento de mais indivíduos que o tamanho projetado para a parte variável pode suportar. Neste exemplo, quanto ao tamanho do domínio de identificadores, três problemas são evidentes, a saber.

1. Os números que poderiam compor a parte variável, mas que não foram utilizados em um dia qualquer, serão perdidos.
2. A informação *data*, especialmente o ano, relativamente à quantidade de identificadores gerados no tempo, varia de modo lento, possivelmente fazendo com que quase todos os números de identidade ativos em um ponto do tempo possuam um "*prefixo*", ou seja, por exemplo, os algarismos aparecerão em todos os números gerados em um espaço de 100 anos. [20]
3. A partir do conjunto de informações pessoais que constam do número de identidade, pode-se variar apenas a pequena parte do número que é aleatória/sequencial e tornar relativamente simples a busca por números válidos. Um exemplo foi apresentado na seção 2.2.7, onde um site na internet permite a construção de números da identidade estoniana e a consulta deste número em sites de busca [14].

Considerando definições da *Teoria da Informação* [29], a inserção de informação com significado semântico reduz a *entropia* de uma mensagem, neste caso representada pelo número RIC. A entropia, intuitivamente, quantifica a imprevisibilidade do conteúdo de uma informação. Assim, um número aleatório possui entropia maior que um número do mesmo tamanho codificando uma data, por exemplo.

3.5.1 Pseudônimos

No contexto de um número de identificação civil, o pseudônimo é um número que oficialmente pode substituir o identificador real, de modo temporário, revogável ou persistente, possivelmente com restrições de privacidade mais flexíveis que o número de identidade real.

Existe uma vasta literatura que trata do projeto, uso, políticas e implementação de pseudônimos em soluções de identidade eletrônica [7] [30]. Diferentes casos de uso, legislação, infraestrutura disponível, etc., é que vão determinar a solução mais adequada para o uso de pseudônimos.

Ao se utilizar um pseudônimo, um dos objetivos principais é a preservação da identidade real do usuário ou inviabilizar o relacionamento entre diferentes bases de dados para descoberta de padrões ou histórico de ação de um usuário. Comumente, o provedor

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

de um serviço explicita as condições nas quais um pseudônimo pode ser usado e é igualmente explícito o uso deste em lugar de um identificador real do usuário.

Para o uso de pseudônimos no RIC, recomendamos que seja utilizado um domínio de números específico para essa finalidade com quinze dígitos mais o DV. Assim, dispõe-se de um amplo espaço de possíveis números para uso massivo com possibilidade de alteração frequente, além de não criar um pseudônimo que se assemelhe ao número RIC. Com dezesseis dígitos, temos:

$$9999\text{-}9999\text{-}9999\text{-}9999$$

Para uma noção do tamanho proposto para o domínio de pseudônimos, consideremos a hipótese de uma população de **um bilhão de pessoas** alterando seus pseudônimos **diariamente** por um período de **cem anos**. Neste caso, ao fim deste período, estariam utilizados, aproximadamente, $36{,}5\%$ dos números com **dezesseis** dígitos, sendo que um destes dígitos é o DV.

$$10^9 \text{ pessoas} * 365 \text{ dias} * 100 \text{ anos} = 3{,}65 * 10^{14} \text{ pseudônimos utilizados}$$

$3{,}65 * 10^{14}$ é um número com quinze algarismos significativos, logo, sendo o identificador de pseudônimos composto por quinze dígitos mais o DV, apenas $36{,}5\%$ do espaço de números possíveis são utilizados no cenário considerado.

3.6 REGRAS DE EXCLUSÃO

Dependendo do tipo de gerador pseudoaleatório a ser utilizado e de como os números RIC são amostrados, há a probabilidade de se gerar números específicos que podem ser considerados indesejados. Por exemplo, $1111\text{-}1111\text{-}11x$ ou $0000\text{-}2222\text{-}22x$, onde x é o dígito verificador, são identificadores que podem ser facilmente utilizados em diversas finalidades, como campanhas publicitárias, testes de sistemas ou mesmo como tentativas de fraude. Por isso, recomenda-se a utilização de testes de aleatoriedade para filtrar cada número RIC gerado, evitando aqueles números facilmente deduzidos.

As regras de eliminação de números podem variar no tempo de acordo com a conveniência e devem ser pensadas de modo a equilibrar o espaço de números disponíveis e as razões que levaram à elaboração da regra. Dentre estas razões, podem fazer parte, por exemplo, critérios legais, ordens judiciais ou até mesmo critérios culturais e religiosos. Para exemplificar, o *Social Security Number*, documento norte-americano, não atribui o número 666 na área reservada ao código de área no número do SSN [31]. Números

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

popularmente conhecidos por expressarem, metaforicamente, algum conceito ou pensamento depreciativo também podem ser evitados.

3.7 ALEATORIEDADE

A área de estudos relativa à *aleatoriedade* é bastante ativa, ainda com diversos problemas em aberto. O tema é de fundamental importância em áreas como a *criptografia*.

Existem na literatura diferentes modelos para mensurar a aleatoriedade de uma sequência de *bits*, por exemplo. Uma fonte de *bits* aleatórios pode ser qualificada, tipicamente, pelas propriedades da *imprevisibilidade* e *incompressibilidade*, sendo a primeira a mais intuitiva [32]. Um processo aleatório qualquer, como o lançamento de uma moeda, é dito imprevisível se o conhecimento de n amostras não ajuda na predição da amostra $n+1$. O mesmo processo produz amostras incomprimíveis se cada amostra não pode ser descrita por uma sequência de *bits* menor que a própria sequência ou, ainda, se apenas uma fração das sequências produzidas podem ser comprimidas dentro de limites estabelecidos.

Um gerador de números RIC deve ser cuidadosamente escolhido para prover níveis aceitáveis de imprevisibilidade, mas não necessariamente incompressibilidade. Por exemplo, um número RIC do tipo $nnnn\text{-}abcd\text{-}nnx$, onde x é o dígito verificador e $(n,a,b,c,d \mid n \neq a \neq b \neq c \neq d)$ são dígitos quaisquer, pode ser comprimida, uma vez que há a repetição do algarismo n e, portanto, a repetição de um padrão de *bits*. Entretanto, espera-se que o número exemplificado não tenha sido previsto por um observador com acesso a amostras prévias produzidas pelo gerador utilizado.

Em termos práticos, existem especificações, padrões e técnicas diversos que estabelecem critérios mínimos para geradores de sequências aleatórias [33]. A escolha adequada é dependente da aplicação e pode variar desde funções presentes em *frameworks* e especificações suportados por linguagens de programação como a linguagem Java, por exemplo, até a utilização de dispositivos do tipo *HSM*, *Hardware Security Modules*.

3.8 DEFINIÇÃO DE ESTADOS (*STATUS*)

O identificador projetado para o RIC pode assumir diferentes estados, que são passíveis de alteração com o tempo a partir de novos requisitos. Inicialmente, podem ser considerados os seguintes.

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

- *Válido*: identificador segue regras de formação e dígito verificador está calculado corretamente.
- *Inválido*: identificador não segue regras de formação e/ou dígito verificador está calculado erroneamente.
- *Ativo*: identificador está associado a um indivíduo e pode ser utilizado para transações diversas.
- *Inativo*: identificador está associado a um indivíduo, mas não está apto para uso. As situações que podem levar a este estado são diversas como, por exemplo, falecimento do indivíduo ou não utilização do identificador único RIC por um período arbitrário.

Recomenda-se manter apenas um pequeno conjunto de possíveis estados, para maximizar o potencial de interoperabilidade e manutenabilidade dos sistemas e políticas de uso. Se estados relativos a negócios específicos são introduzidos, a complexidade de outros sistemas/negócios passa a ser parte do número RIC, levando a maiores dificuldades e custos de manutenção. Por exemplo, não é recomendado que um estado como `<RIC clonado>` faça parte da lista de estados, por remeter a lógica de negócio não afeta à proposta de um número *agnóstico*, ou seja, com potencial de uso diverso e independente do tipo de negócio.

3.9 O NÚMERO RIC E CHAVES EM BANCOS DE DADOS

O número RIC foi projetado para ter significado próprio como um atributo pessoal e, idealmente, será imutável. Entretanto, esta característica não pode ser assegurada em alguns casos, já que o atributo está sujeito a regras próprias de geração e utilização, assumindo um papel no cotidiano das pessoas. Por isso, apesar de não carregar significado semântico de outros atributos pessoais, o número RIC possui sua própria semântica.

Assim, para representar unicamente um indivíduo e referenciá-lo em sistemas computacionais e bases de dados, recomenda-se a utilização de uma *chave substituta* (*surrogate key*) [34] que, por definição, tem como características básicas:

- nenhum significado semântico;
- imutável;
- não manipulável por pessoas ou pela aplicação;
- gerada automaticamente sem estar submetida a regras negociais;

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

- não reutilizável;
- única em todo o sistema.

O número RIC não compartilha com as chaves substitutas, pelo menos, as quatro primeiras características acima, motivo pelo qual o número não é adequado para ser utilizado como chave de uma entidade em um banco de dados relacional, por exemplo.

Questões relativas à segurança do número RIC em seu armazenamento e transmissão, pelo uso de métodos criptográficos e outros aplicáveis, serão tratadas em documentos específicos.

3.10 ESTRATÉGIAS DE GERAÇÃO DO NÚMERO RIC

3.10.1 Distribuição de Números RIC

Propomos um método para distribuição de números RIC válidos, gerados previamente, em cenários onde seja necessário atribuir o identificador de modo *off-line*, imediatamente após um procedimento de cadastro, por exemplo. Consideramos como requisitos a existência de um operador que solicita um conjunto de números RIC, bem como a existência de certificados digitais válidos do operador e da central que enviará o conjunto de números solicitados.

Seja um conjunto de números RIC $N_{RIC} = \{n_1, n_2, \dots, n_m\}$, válidos e gerados aleatoriamente, ou seja, $n_i \in_R D_p$, com $i \in 1 \dots m$, onde D_p é o domínio de todos os números RIC únicos ainda não atribuídos. Para distribuir N_{RIC}, propomos que cada n_i seja enviado na forma de uma mensagem cifrada, assinada pela central e endereçada ao operador que solicitou o pacote de números RIC. Técnicas criptográficas comuns estão disponíveis para a concretização do mecanismo descrito, como, por exemplo, as descritas pela RFC 4880 [35], seções 2.1 e 2.2.

Após a decifragem de um n_i qualquer, este deve ser associado ao cadastro correspondente para atualização posterior da base de dados centralizada.

Regras e políticas deverão ser definidas para estabelecimento de limites no uso do mecanismo de distribuição de números RIC descrito. Recomendam-se que os números não utilizados pelo operador sejam descartados, embora seja possível que o reaproveitamento aconteça após um período de tempo arbitrário. Políticas adequadas podem aceitar o descarte temporário de identificadores até um limite aceitável, de forma que o sistema como um todo não seja impactado.

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

3.10.2 Geração Descentralizada do Número RIC

Embora seja recomendável a geração centralizada do número RIC, pode-se conseguir a descentralização de duas maneiras, a saber.

1. *Geradores cíclicos de números pseudoaleatórios com a utilização de uma mesma semente*: Geradores pseudoaleatórios como o LSFR, por exemplo, são iniciados pela utilização de uma *semente* de tamanho k cuidadosamente escolhida. O *período* desse tipo de gerador é $2^k - 1$, podendo ser utilizado para a geração de números RIC, de modo concorrente com múltiplos geradores ou não.

   Sejam m geradores pseudoaleatórios cíclicos iniciados com uma mesma semente. Para a produção concorrente descentralizada de números RIC, basta que cada gerador $G_i \mid i \in \{1 \dots m\}$ produza uma saída a cada m números RIC n_i gerados, ou seja, G_i terá como saída o conjunto $\{n_i, n_{i+m}, n_{2i+m}, \dots, n_{2j+m}\}$ com $j \in 1 \dots \infty$. Assim, por exemplo, ao se utilizar quatro geradores $(m = 4)$, o gerador G_3 produz na saída os números RIC $n_3, n_7, n_{11}, \dots, n_{3j+m}$.

   Com essa abordagem, todos os geradores estarão produzindo os mesmos números RIC, mas cada um apenas considera como sua saída os números produzidos após uma quantidade de amostras pré-definida.

   Diversas questões de segurança devem ser avaliadas antes de se adotar uma solução como a apresentada. Por exemplo, pode ser mais apropriado o uso de HSM específico e pré-configurado, ao invés de implementações em software cuja semente do gerador será distribuída nas unidades descentralizadas.

Descentralização virtual: Utilizando a técnica descrita na seção 3.10.1, pode ser realizada a distribuição de pacotes de números RIC para cada unidade descentralizada. Sempre que necessário, um número RIC é decifrado e utilizado.

4 GENERALIZAÇÃO DO NÚMERO RIC

Podemos considerar a generalização do número RIC para uso em outros contextos que não só a identificação civil ou em casos particulares de identificação temporária, por exemplo. Isso se justifica a partir da identificação dessas necessidades em outros países ou mesmo no Brasil.

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

4.1 UM IDENTIFICADOR PARA ENTIDADES

Com a palavra *entidade* nos referimos a qualquer instituição pública, privada ou mista que necessite interagir com sistemas e estruturas negociais, políticas, legais, etc., criadas para ou compatíveis com o número RIC, que é definido apenas para o propósito de identificação **civil**. Além disso, uma entidade pode designar diferentes abstrações que não sejam, necessariamente, pessoas jurídicas. Por exemplo, uma seção de um órgão público, uma tribo indígena ou um cargo de chefia em um ministério podem ser designados como entidades e possuírem um identificador próprio nas bases do RIC.

As possibilidades de uso e o gerenciamento e implantação de um identificador para entidades devem ser tratadas em mais detalhes, inclusive o suporte legal, em documentos específicos. Neste RT, apenas recomendaremos o formato do identificador, seguindo o embasamento já construído no Capítulo 3 para o número RIC. Assumimos que uma entidade não necessita dos mesmos requisitos de privacidade considerados para o indivíduo, por isso, a inclusão de informação semântica no identificador de entidades é possível, observado o uso racional do domínio de números reservado.

Segundo o site *Empresômetro* [36], existem atualmente mais de dezessete milhões de empresas ativas no Brasil. Conforme metodologia descrita no site, entende-se por empresa todos os tipos jurídicos, como sociedades anônimas, associações, igrejas, entidades públicas, etc. O crescimento anual no número de empresas, de 2012 para 2013 e de 2013 para 2014, foi de aproximadamente 12%.

Assim, é suficiente reservar nove dígitos significativos para a identificação de empresas. Entretanto, números de nove dígitos estão incluídos no domínio de números RIC para identificação civil, que são os números de dez dígitos iniciados com 0. Por isso, recomendamos a utilização de algarismos adicionais, com significado semântico ou não, para o identificador de entidades, de forma que totalize doze dígitos significativos.

Uma opção de formato está exemplificada a seguir. O dígito verificador foi adicionado, resultando em um identificador com treze algarismos:

9999-9999-9999-9

4.2 RESERVA DE FAIXA NUMÉRICA

Propomos a reserva de identificadores com onze dígitos significativos mais DV para aplicações futuras. Esta faixa pode ser utilizada para finalidades diversas ou para extensão

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

do domínio do número RIC, caso o domínio previsto de dez dígitos mais DV seja considerado insuficientemente esparso, por exemplo.

A faixa numérica em questão pode ser subdividida para atender a diferentes requisitos, a partir da fixação de algarismos específicos na formação do número. Um exemplo será identificado na seção a seguir.

4.3 IDENTIFICADORES TEMPORÁRIOS

A exemplo do que ocorre com o *Social Insurance Number* do Canadá, mostrado na seção 2.2.4, podemos definir um domínio de identificadores para uso temporário, em aplicações como, por exemplo, estrangeiros com permissão de trabalho no Brasil ou matrículas de servidores públicos. Um identificador temporário pode ou não estar vinculado a um número RIC válido e ativo.

A previsão de domínios temporários tem utilidade na preservação da generalidade do número RIC, projetado para ser um atributo pessoal e, por isso, idealizado para atender a requisitos de privacidade conforme legislação em vigor e tendências internacionais relativas à proteção de dados pessoais. O identificador temporário pode acomodar casos não previstos ou incompatíveis com o conceito do número RIC, seja pela proposição de novas aplicações ou mudanças na legislação.

Vale notar que o propósito de um identificador temporário é diferente do pseudônimo descrito na seção 3.5. No pseudônimo, os requisitos de privacidade podem ser vistos com menos rigor, devido a suas propriedades previstas, como a revogabilidade a qualquer tempo e a critério do usuário. Assim, sugerimos que o identificador temporário não seja utilizado com a finalidade já proposta para os pseudônimos.

Para a definição do domínio de identificadores temporários, propomos a utilização de um subdomínio da faixa reservada, descrita na seção anterior, com números de onze dígitos significativos mais DV, com o prefixo dado por um ou dois algarismos, conforme previsão de quantidade de números a serem necessários. Ou seja, caso sejam previstos dez bilhões de números possíveis para o identificador temporário, com, por exemplo, o algarismo 1 como prefixo, teremos:

1999-9999-9999

Ou, para o caso de um bilhão de números reservados com prefixo 11:

1199-9999-9999

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

4.4 CONSIDERAÇÕES SOBRE OS DOMÍNIOS DOS IDENTIFICADORES

Dentre os casos de números de identificação pessoal apresentados na seção 2, pudemos verificar a utilização de informações pessoais na formação numérica de um identificador e também casos, nos quais o número é gerado aleatoriamente.

O modelo adotado pela Índia é o que mais se aproxima das direções recomendadas por esse RT. Entretanto, no *Aadhaar* existe a reserva de dois algarismos, 0 e 1, o que é desnecessário no caso do número RIC, pelos seguintes motivos.

Consideremos o domínio D_p de todos os identificadores pessoais. Queremos um domínio D_e de identificadores para entidades de forma que $D_p \cap D_e = \emptyset$, para que haja uma explícita separação entre os domínios de números atribuídos a pessoas de números atribuídos a entidades.

No caso indiano, $D_e^{Índia}$ foi derivado do $D_p^{Índia}$ original, pela reserva de um décimo dos números. Além disso, outro décimo foi reservado para números iniciados com 0 para uso futuro. Ainda que o tamanho de $D_p^{Índia}$ continue sendo suficiente e adequado, uma desvantagem desse modelo é a semelhança visual entre os números de $D_e^{Índia}$ e $D_p^{Índia}$.

No RIC, estamos definindo D_p^{RIC} como sendo todos os números naturais com dez ou menos dígitos significativos e D_e^{RIC} como números naturais de 12 algarismos significativos. Dessa maneira, conseguimos uma clara distinção que beneficia a construção dos identificadores pessoais e de entidades.

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

5 CONCLUSÃO

Apresentamos argumentos e referências que fundamentaram a escolha de características relativas ao número RIC. Ao considerar aspectos de privacidade e usabilidade, pudemos construir um identificador único nacional que pode ser utilizado pelas pessoas em suas interações com o governo, entes privados e com a sociedade em geral.

Além da construção do número RIC, algumas estratégias simples foram apresentadas para viabilizar processos como a pré-distribuição de números para atribuição *off-line* e a descentralização da geração dos identificadores.

Por fim, apresentamos propostas de identificadores relacionados ao RIC, mas que não são necessariamente utilizados para identificação civil, como os identificadores de empresas, por exemplo.

As atividades envolvidas nesta etapa observaram formalmente a execução dos passos da metodologia elencada para gestão do projeto, PMI/PMBoK.

A equipe da UnB considera que teve acesso a todas as informações necessárias à boa condução dos trabalhos e que a disponibilização dessas informações pela equipe do MJ, assim como as atividades conjuntas de análise e discussão, levou a etapa do projeto a bom termo.

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

6 REFERÊNCIAS

[1] Nóbrega, Cristóvão Barcelos da;, “HISTÓRIA DO IMPOSTO DE RENDA NO BRASIL, UM ENFOQUE DA PESSOA FÍSICA (1922-2013),” Secretaria da Receita Federal do Brasil, 2014. [Online]. Available: http://www.youblisher.com/p/997520-Historia-do-imposto-de-Renda-no-Brasil. [Acesso em 26 01 2015].

[2] Tribunal Superior Eleitoral, “Resolução nº 21.538, de 14 de outubro de 2003,” 14 Outubro 2003. [Online]. Available: http://www.tse.jus.br/legislacao/codigo-eleitoral/normas-editadas-pelo-tse/resolucao-nb0-21.538-de-14-de-outubro-de-2003-brasilia-2013-df. [Acesso em 27 Janeiro 2015].

[3] CONTRAN, “Resoluções do CONTRAN,” 30 Mar 2006. [Online]. Available: http://www.denatran.gov.br/download/Resolucoes/resolucao_192_06.doc.

[4] BRASIL, “LEI Nº 7.116,” 29 Ago 1983. [Online]. Available: http://www.planalto.gov.br/ccivil_03/leis/1980-1988/l7116.htm.

[5] “South African ID Number Check,” 2015. [Online]. Available: http://www.legalcity.net/Index.cfm?fuseaction=tools.idcheck. [Acesso em 2015].

[6] E. A. Whitley e G. Hosein, Global Challenges for Identity Policies, Palgrave Macmillan, 2009.

[7] N. Vandezande, “Identification numbers as pseudonyms in the EU public sector,” *European Journal of Law and Technology*, vol. 2, 2011.

[8] WP13, “D13.3: Study on ID number policies,” 14 Sep 2007. [Online]. Available: https://lirias.kuleuven.be/bitstream/123456789/205522/1/fidis-wp13-del13_3_number_policies_final.pdf. [Acesso em Feb 2015].

[9] Service Canada, “Social Insurance Number,” [Online]. Available: http://www.servicecanada.gc.ca/eng/sc/sin/index.shtml. [Acesso em Mar 2015].

[10] ASSOCIATED PRESS, “MailOnline,” 14 Oct 2014. [Online]. Available: http://www.dailymail.co.uk/wires/ap/article-2791953/South-Korea-identity-thefts-forces-ID-overhaul.html.

[11] Social Security Administration, “The SSN Numbering Scheme,” 2015. [Online]. Available: http://www.ssa.gov/history/ssn/geocard.html.

[12] Washington Post, “Researchers: Social Security Numbers Can Be Guessed,” 6 July 2009. [Online]. Available: http://www.washingtonpost.com/wp-dyn/content/article/2009/07/06/AR2009070602955.html.

[13] Estonia, Estonian e-ID, 2015.

[14] M. Sidorin, “Estonian ID card number generator,” 2015. [Online]. Available: http://www.uzzersite.eu/labs/estonian-id-card-generator/index.html.

[15] H. Kanakia, S. Nadhamuni e S. Sarma, “A UIDAI Numbering Scheme,” May 2010. [Online]. Available: http://uidai.gov.in/UID_PDF/Working_Papers/A_UID_Numbering_Scheme.pdf.

[16] RENAPO (Registro Nacional de Población e Identificación Personal), “Composição identificador CURP,” [Online]. Available: http://sistemas.uaeh.edu.mx/dce/admisiones/docs/guia_CURP.pdf.

[17] Oxford Internet Institute, “Breaking Barriers to eGovernment,” *MODINIS contract 291722, Deliverable 2, Prepared by: Oxford Internet Institute (and others) For: European Commission Directorate General for Information Society and Media.*, 2007.

[18] International Civil Aviation Organization, “Alphabet - Radiotelephony,” 2005. [Online].

Available: http://www.icao.int/Pages/AlphabetRadiotelephony.aspx.

[19] "Projeção da população do Brasil e das Unidades da Federação," 2015. [Online]. Available: http://www.ibge.gov.br/apps/populacao/projecao/.

[20] N. Cowan, "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," *Behavioral and Brain Sciences, 2, 97–185,* 2001.

[21] J. Kirtland, "Identification Numbers and Check Digit Schemes," *Mathematical Assoc. of America,* 2001.

[22] J. A. Gallian, "The Mathematics of Identification Numbers," *The College Mathematics Journal. Vol 22, No. 3.*, pp. 194-202, 1991.

[23] M. Gargano e M. Courtney, "Identification Numbers and the Mathematics of Check Digit Schemes for Computer Scientists," 2003.

[24] UK government, "Data Protection Act 1998," 1998. [Online]. Available: http://www.legislation.gov.uk/ukpga/1998/29/introduction.

[25] EUROPEAN UNION, "Protection of personal data," 1995. [Online]. Available: http://ec.europa.eu/justice/data-protection/.

[26] OECD, "OECD Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data," 2013.

[27] Privacy Commissioner of Canada, "Privacy Legislation in Canada," May 2014. [Online]. Available: https://www.priv.gc.ca/resource/fs-fi/02_05_d_15_e.asp.

[28] Congresso Nacional, "Anteprojeto de Lei de proteção dos dados pessoais," 2015. [Online]. Available: http://www.acessoainformacao.gov.br/menu-de-apoio/recursos-passo-a-passo/anteprojeto-lei-protecao-dados-pessoais.pdf.

[29] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal 27 (3),* p. 379–423, 1948.

[30] A. Pfitzmann e M. Hansen, "Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management – A Consolidated Proposal for Terminology," 15 Feb 2008.

[31] SSA/USA, "Social Security is Changing the Way SSNs are Issued," [Online]. Available: http://www.ssa.gov/kc/SSAFactSheet--IssuingSSNs.pdf. [Acesso em 13 Mar 2015].

[32] R. G. Downey e D. R. Hirschfeldt, Algorithmic Randomness and Complexity, Springer Science & Business Media, 2010.

[33] D. Eastlake III, J. Schiller e S. and Crokcer, RFC 4086: Randomness Requirements for Security, 2005.

[34] R. S. Wazlawick, Object-Oriented Analysis and Design for Information Systems: Modeling with UML, OCL, and IFML, Elsevier Inc., 2013.

[35] J. Callas, L. Donnerhacke, H. Finney, D. Shaw e R. Thayer, "Request for Comments: 4880," Nov 2017. [Online]. Available: http://www.ietf.org/rfc/rfc4880.txt.

[36] IBPT – Instituto Brasileiro de Planejamento e Tributação, "Empresômetro," 2015. [Online]. Available: http://www.empresometro.com.br/.

Confidencial.
Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

Universidade de Brasília – UnB

Centro de Apoio ao Desenvolvimento Tecnológico – CDT

Laboratório de Tecnologias da Tomada de Decisão – LATITUDE

www.unb.br– www.cdt.unb.br – www.latitude.eng.br