Proposta de um Modelo de Transparência de Dados e Informações do Processo de Venda e Emissão de Certificado Digitais para a Governança do ITI

Proposal of a Transparency Model of Data and Information about the Process of Sales and Emission of Digital Certificates for ITI Governance

Fernanda Gomes
Universidade Federal de Santa Catarina
Florianópolis, SC, Brasil
fernandaoliveiragomess@gmail.com

Jean Martina
Universidade Federal de Santa Catarina
Florianópolis, SC, Brasil
jean.martina@ufsc.br

RESUMO

Um problema muito comum nas empresas é a falta de dados organizados, os quais quando armazenados em fontes distintas complicam a extração de informação dos dados. A centralização dos dados facilita a busca por informações que podem auxiliar a tomada de decisão das empresas. A utilização da ferramenta de Data Warehouse ajuda a armazenar esses dados de forma íntegra, consistente, temporal e serve de apoio à tomada de decisão. Além dessa centralização dos dados, a transparência deles contribuí para a governança, que é uma série de práticas as quais alinham os interesses de todos os *stakeholders*, de uma organização seja ela pública ou privada. Neste trabalho, é proposto melhorar a governança e proporcionar a necessária transparência de dados relacionados ao mercado de certificação digital no Brasil. O resultado desse projeto pode ser utilizado pelo Instituto Nacional de Tecnologia da Informação (ITI), autarquia federal responsável pela Infraestrutura de Chaves Públicas Brasileira (ICP-Brasil). Para isso, foi realizado um estudo do mercado e proposto um modelo de coleta e divulgação de dados. O modelo foi implementado e testado, concluindo-se que ele é viável e fornece um panorama das informações do mercado que podem ajudar sua compreensão.

Palavras-Chave

Data Warehouse; Transparência; Governança; Certificado Digital; ICP-Brasil.

ABSTRACT

A common problem in companies is the lack of organized data mostly stored in different sources. Centralized data facilitates the search for information. That information can be helpful for the decision make of the companies. In addition to this centralization of data, their transparency contributes to governance. Governance is a series of practices that align the interests of all stakeholders of an organization, whether public or private. In this work is proposed to improve governance and provide the necessary transparency of data related to the digital certification market in Brazil. The result of this project can be used by the National Institute of Information Technology (ITI), federal authority responsible for the Brazilian Public Key Infrastructure (PKI-Brazil). Moreover, a market study was carried out and a data collection and dissemination model was proposed. The model was implemented and tested in a fictitious scenario. With that, it was possible to conclude that the proposal is feasible and provides an overview of the market information that may help its understanding.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SBSI 2018 June 4^{th} – 8^{th}, 2018, Caxias do Sul, Rio Grande do Sul, Brazil

Copyright SBC 2018.

CCS Concepts

•**Information systems** → **Extraction, transformation and loading;**

Keywords

Data Warehouse; Transparency; Governance; Digital Certificate; PKI-Brazil.

1. INTRODUÇÃO

Organizar grandes quantidades de dados é uma tarefa comum em várias organizações. Essa tarefa muitas vezes pode se tornar um obstáculo. O problema acontece pela existência de dados de diversas fontes (ex. planilhas, documentos, arquivos, entre outros) que se encontram espalhados entre os setores de uma organização. Grande parte das organizações sofre com uma abundância de dados redundantes e inconsistentes. Em geral, o problema não é a falta de dados mas a dificuldade em administrar com eficiência e assim utilizá-los no apoio à tomada de decisão [6].

Dados consistentes, organizados e centralizados geram informações de qualidade que podem servir de suporte à tomada de decisão nas organizações. Para que exista real compreensão de um mercado é necessário saber em quais produtos investir, o que ele solicita de tecnologia, o perfil do cliente deste mercado, dentre outras perguntas estratégicas

que podem gerar vantagem competitiva a quem retém as respostas.

O Instituto Nacional de Tecnologia da Informação (ITI), autarquia federal responsável pela Infraestrutura de Chaves Públicas Brasileira, tem informações sobre o mercado de certificação digital. A Instrução Normativa nº 14 regulamenta que as Autoridades Certificadoras enviem os dados presentes nos anexos 1 e 2 da normativa [5].

Além dos dados não serem concentrados em uma base de dados, a transparência disponibiliza poucos dados resultando em poucas informações. Vale ressaltar que essas informações não são somente de cunho financeiro, como as de divulgação obrigatória regida pela Lei da Transparência.

Para alcançar uma boa governança, além de existir a transparência, que é obrigatória por lei, faz-se necessária a divulgação de informações da organização. Essa divulgação traz confiança e clareza da situação do setor para com os *stakeholders* [3].

O enfoque do trabalho é a transparência dos dados e informações de certificados e mídias armazenadoras do mercado de certificação digital brasileiro, que são produto para a extração de conhecimento e tomada de decisão. É proposto um modelo de transparência de dados e informações para a Infraestrutura de Chaves Públicas do Brasil (ICP-Brasil) mantido pelo ITI, o qual é fruto de um estudo de mercado aplicado junto aos conhecimentos tecnológicos. O intuito desta proposta é encontrar uma solução para o problema de falta de dados e informações divulgados sobre o mercado nacional de certificação digital, visto que a maioria das informações publicadas são especulações visto a pouca informação divulgada. Outra contribuição é o envio automático dos dados.

O trabalho foi dividido em cinco seções, sendo que o primeiro traz a introdução. Na seção dois será apresentada toda a fundamentação teórica. A seção três mostra a proposta do trabalho. Na seção quatro é apresentado os resultados da simulação da proposta. A última seção apresenta a conclusão e possíveis trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Governança Corporativa

A governança corporativa é o sistema pelo qual as empresas e demais organizações são dirigidas, monitoradas e incentivadas, envolvendo os relacionamentos entre sócios, conselho de administração, diretoria, órgãos de fiscalização e controle e demais partes interessadas [3]. Pode-se definir a governança como um conjunto de valores, princípios, propósitos e regras que regem o sistema de poder entre os mecanismos de gestão das corporações, buscando a maximização da riqueza dos acionistas e o atendimento dos direitos de outras partes interessadas, minimizando oportunismos conflitantes com esse fim [1]. O IBGC, considera como princípios básicos da governança: Equidade, *Accountability* (prestação de contas), Responsabilidade Corporativa e Transparência [3]. O trabalho em questão utiliza dos conceitos de transparência para seu desenvolvimento.

2.2 Certificados Digitais

Um certificado é o elemento mais básico de uma Infraestrutura de Chaves Públicas (ICP), também conhecido como certificado de chave pública. Ele contém a chave pública que associa o usuário com a chave privada correspondente [2]. Essa chave privada é utilizada para assinar digitalmente, a mesma é de conhecimento único e exclusivo do seu dono. Assinar uma mensagem digitalmente é basicamente pegar a mensagem e aplicar funções matemáticas para resumi-la, funções de *hash*, e utilizar sua chave privada no resultado dessa mensagem resumida para cifrar. Sendo um objeto puramente digital, é possível carregar os certificados em um computador. Um certificado é composto pelo nome do proprietário, chave pública, organização ao qual o dono pertence, endereço de correio eletrônico, data de emissão e de expiração, nome da parte confiável que emitiu o certificado, número de série adicionado pelo emissor, dentre outras informações. Por fim, esse conteúdo é assinado pelo emissor e adicionado ao certificado. O certificado digital pode ser comparado a um RG (Carteira de Identidade), que contém uma assinatura, as informações do proprietário e é emitida por um órgão responsável que garante sua veracidade, só que usado para identificar-se virtualmente.

2.3 ITI e ICP-Brasil

O Instituto Nacional de Tecnologia da Informação (ITI) é uma autarquia federal vinculada à Casa Civil da Presidência da República, cujo objetivo é manter a ICP-Brasil. Ele é a Autoridade Certificadora-Raiz (AC-Raiz) da cadeia de certificação da ICP-Brasil. Com a Medida Provisória 2.200-2 de 24 de agosto de 2001 foi criada legalmente a ICP-Brasil [4]. A primeira função de uma ICP é permitir, por meio das Autoridades Certificadoras (AC), a distribuição e o uso de chaves públicas e certificados com garantia de segurança. Junto com à atividade das Autoridades Registradoras (AR), que tem a função de verificar a identidade de uma pessoa antes de emitir seu certificado [4]. Normalmente essa verificação é feita presencialmente em um ponto de atendimento e os documentos, como de identificação e de residência, do indivíduo são averiguados para garantir que o certificado emitido é realmente da pessoa que diz querer o emitir. A Infraestrutura de Chaves Públicas Brasileira, a ICP-Brasil, foi instituída pela Medida Provisória no 2.200, de 28 de junho de 2001. ICP-Brasil é uma cadeia hierárquica e de confiança que viabiliza a emissão de certificados digitais para identificação virtual do cidadão [4]. O ITI também tem as funções de credenciar e descredenciar os demais participantes da cadeia, supervisionar e realizar auditoria dos processos [4].

3. PROPOSTA

Inicialmente foi realizado um estudo sobre a transparência de dados nos *sites* das principais agências reguladoras, também autarquias federais, brasileiras. Foram estudados os mecanismos de transparência de dados de mercado de cada agência, procurando entender quais dados são capturados, como são capturados e disponibilizados. Os critérios para avaliação da transparência das autarquias foram relacionados a facilidade de encontrar o conteúdo no *site* e a usabilidade no processo de coleta e divulgação dos dados. As duas autarquias que se destacaram positivamente foram a ANCINE (Agência Nacional do Cinema) e SUSEP (Superintendência de Seguros Privados) servindo de inspiração para este trabalho.

Como pontos fortes de algumas autarquias, foram destacados os relatórios anuais. Esses mostram de maneira clara, com gráficos e dados estatísticos, como o mercado se comportou no ano. Outro ponto foi a divulgação dos dados,

atuais e históricos, do mercado. Esses são disponibilizados em forma de gráficos e consultas o que facilita o entendimento das informações pelos usuários.

O principal ponto fraco encontrado diz respeito a coletados dos dados. Todas as agências pesquisadas realizam a coleta de uma maneira manual, por envio através de planilhas, arquivos XML, dentre outros meios. A coleta dos dados é normalmente realizada mensalmente. Essa coleta manual pode ocasionar problemas na qualidade e inconsistência dos dados além de um trabalho repetitivo. Então, foi observado que o ITI, mesmo não sendo uma agência reguladora, realiza a transparência de dados do mercado de certificação digital em seu portal. Pensando em melhorar e sugerir a regulação deste foi proposto um modelo de transparência de mercado de certificação digital da ICP-Brasil para o ITI.

O trabalho focou nos produtos certificado digital e mídia de armazenamento. Mas vale ressaltar que o mercado contempla outros produtos, tais como: *software* de assinatura, prestação de serviços, seguros, auditoria, *datacenters*, consultoria, aplicações (ex. nota fiscal eletrônica, sistema de coleta de assinaturas, *GED/Workflow*), site seguro, sigilo de documentos eletrônicos, preservação de documentos eletrônicos, *hardware* criptográficos incluindo *HSMs*, carimbo do tempo, dentre outros.

Para realizar a coleta automática dos dados de venda e emissão foi necessário entender o processo.

3.1 Coleta

O processo começa em um *site* de uma autoridade certificadora privada (que cobra pelo seu serviço de certificação), da escolha do usuário, registrada pelo ITI. O *site* apresenta diversos certificados (ex. certificados para pessoa física ou jurídica, servidores, advogados, contadores, entre outros). Esse certificado ainda pode ser vendido junto a uma mídia de armazenamento (ex. *token* ou cartão). Após a escolha, o certificado vai para o carrinho de compras do *site* e para finalizar a compra um cadastro é realizado. Neste cadastro são coletadas informações do comprador. Após o pagamento, uma visita a um ponto atendimento de uma Autoridade Registradora ou domiciliar é agendada para que seja realizada a validação dos documentos do cliente e emissão do certificado. Caso aconteça algum problema, no caso de alguma irregularidade, o certificado não é emitido.

Desse processo foram selecionados diversos dados para captura, tais como: para qual fim o certificado será utilizado, qual mídia foi utilizada para armazenar o certificado, qual o valor do produto, qual autoridade certificadora e registradora participou desse processo, que tipo de atendimento foi realizado para a emissão do certificado, validade, qual a política de certificação utilizada, que tipo de pessoa realizou a compra, física ou jurídica, além de informações demográficas do emissor, da AC e da AR. Vale ressaltar, que os dados dos clientes que são enviados não expõem sua identidade, sendo apenas os dados de cidade, estado e tipo de pessoa (física ou jurídica), e a partir deles não é possível identificar o titular desse certificado.

Foi percebido que o local ideal para a coleta das informações seria no próprio *site* da autoridade certificadora. Esse é o local onde são encontradas todas as informações do processo. Com isso, foi proposto o modelo de automatização da coleta e disponibilização dos dados.

3.2 Armazenamento

Os dados foram armazenados em um *Data Warehouse* (DW), que inicialmente contém um *Data Mart* (DM), que representa o processo de negócio venda e emissão de um certificado digital. A escolha do DW no modelo *bottom-up* (formado por um conjunto de DMs) é ideal pois este pode ser expandido criando novos *Data Marts* que modelam outros processos de negócio envolvendo outros produtos.

Para a construção desse DW, algumas perguntas estratégicas foram criadas, tais como: quais regiões emitem mais certificados, quais são os certificados mais emitidos, qual mídia de armazenamento é a mais vendida, qual é o perfil do comprador de um certo tipo de certificado, dentre outras perguntas que ajudam a compreender melhor o mercado. O DM foi construído com um tabela de fato que representa o processo de venda/emissão de certificado contendo a chave de todas as tabelas de dimensão e o valor, em reais, que essa venda gerou.

As dimensões criadas foram a dimensão certificado, mídia de armazenamento, ponto de atendimento, pessoa e prestadora de serviço de certificação (que pode ser uma AR ou AC, sendo assim, é referenciada duas vezes na tabela de fato). Todas essas dimensões contêm informações demográficas e qualitativas como tipo de atendimento da prestadora de serviço, nome fantasia das ARs e ACs e na tabela pessoa também existe o atributo de tipo de pessoa, que pode ser física e jurídica. Por fim, a dimensão de tempo que informa os dados temporais sobre o dia da emissão.

Um exemplo de questão que pode ser resolvida tendo em mãos esses dados é a distribuição de pontos de atendimento nos lugares onde existe uma demanda maior de emissão de certificados. Para isso foi criado um site onde são disponibilizadas as informações capturadas em forma de consultas e gráficos. Com isso, todos os interessados no mercado podem ter acesso as informações que podem servir de apoio à tomada de decisão, base para estudo, curiosidade, entre outros objetivos.

3.3 Disponibilização

Para disponibilizar os dados e informações foi criado uma aplicação *web* de transparência. Esta apresenta gráficos mostrando a situação do mercado demograficamente e qualitativamente além de proporcionar ao usuário o *download* dos dados em formato CSV com filtros de sua escolha. Por exemplo, uma pessoa deseja fazer *download* de dados sobre um tipo de certificado em um período de tempo, ou deseja apenas dados de certificados que foram emitidos por autoridades certificadoras do Sul do país, como pode ser visualizado na figura 1.

Esses dados presentes no CSV podem ser analisados por diversas ferramentas de front-end de DW, também conhecidos como ferramentas de BI (*Business Intelligence*) onde o usuário pode fazer o download do arquivo CSV ou colocar esse arquivo em um banco de dados e fazer a conexão com à ferramenta. Como exemplo dessas ferramentas podem ser citadas o *Tableau* e *Metabase*. As ferramentas de *front-end* são voltadas ao usuário final. Essas focam na usabilidade do sistema para que qualquer pessoa de qualquer área de uma organização possa utilizar sem a necessidade de auxílio da área de TI (Tecnologia da Informação).

Essas ferramentas possibilitam a visualização das informações de maneira mais amigável por meio de gráficos. Com isso, essas ferramentas se tornam complementares ao website, que apresenta apenas alguns gráficos com pesquisas pré-

formadas, deixando o usuário livre para consultar os dados.

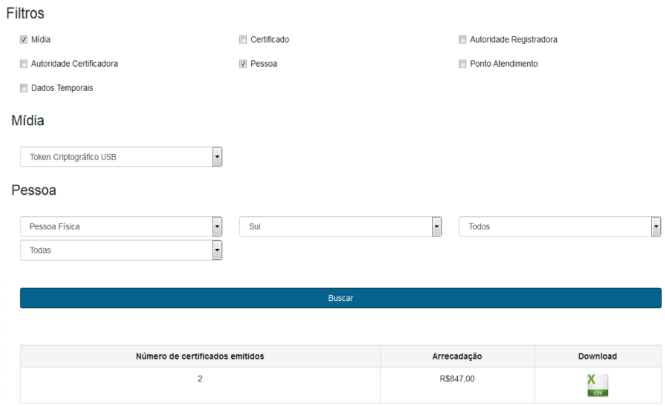

Figura 1: Exemplo de consulta com filtros aplicados pelo usuário.

O modelo de transparência sugerido por esta proposta também prevê a disponibilização, neste *site* de transparência, de relatórios anuais, mostrando como foi o ano para o mercado.

3.4 Fluxo dos dados

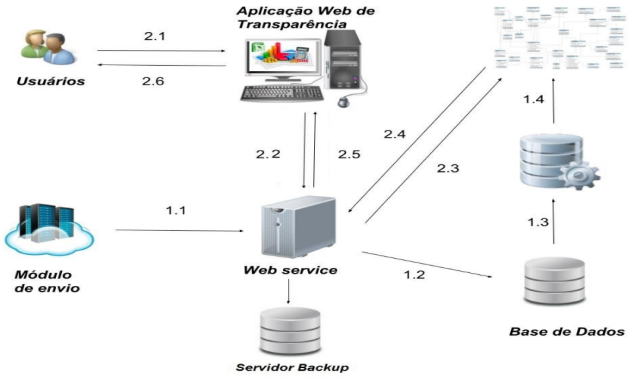

Figura 2: Fluxo de dados.

Na figura 2 em 1.1 é mostrado o fluxo de dados das vendas e emissões via *Web Service*. Quando um certificado é vendido esse ainda não foi emitido, pois apenas o agendamento com a autoridade registradora foi realizado, então esses dados de certificados vendidos ficam em uma tabela temporária no banco de dados, esperando que o certificado seja emitido, esse fluxo é ilustrado por 1.1, 1.2 e 1.3. Após o certificado ser emitido é alterada, também via *Web Service*, uma *flag* de emitido na tabela temporária. Um *script* é rodado diariamente para verificar quais dados de venda de certificados pertencem a certificados que foram emitidos, buscando na *flag* de emitido da tabela por tuplas com o campo emitido preenchido com verdadeiro. Este mesmo *script* é responsável pela limpeza e envio para o modelo dimensional, representado por 1.4.

A partir do momento que os certificados estão no modelo dimensional, esses já estão disponíveis para consulta, como mostra o fluxo 2.1 a 2.6. Os dados também são enviados para um servidor *backup* criando redundância para aumentar a disponibilidade. Outro ponto importante é a padronização dos dados que devem ser enviados, para isso foi criado um manual com a padronização de todos os campos, para que todas as ACs sigam. Um exemplo de erro que pode acontecer caso os dados não padronizados é uma pessoa digitar "S.Paulo" e outra São Paulo, a intenção é falar do mesmo local mas na hora de realizar consultas os dois serão interpretados de maneira distinta.

4. RESULTADOS

Toda proposta apresentada no fluxo de dados foi implementada. Foram criados sites de autoridades certificadoras fictícias para poder simular o funcionamento da proposta. Foram realizadas diversas compras e emissões para poder testar a arquitetura e sua integração. Com isso, o fluxo dos dados foi testado por completo e o funcionamento desse ocorreu como o esperado. Ao final da simulação, todos os dados estavam no *Data Mart* e ao acessar à aplicação *web* de transparência de dados podiam ser visualizados os gráficos populados e as consultas retornando os dados. Com isso a periodicidade do envio tornou-se diária e o envio automatizado. O DW respondeu todas as perguntas desejadas.

5. CONCLUSÃO

A modelagem do *Data Mart* do DW respondeu as perguntas criadas, sendo assim, a modelagem cumpriu com os objetivos. O resultado desse trabalho mostra que com uma arquitetura adequada é possível extrair informações das ACs sem a necessidade do trabalho humano contínuo, como acontece atualmente, utilizando-se da automatização do processo.

Como trabalhos futuros, sugere-se a aplicação dessa proposta em um ambiente real para que seja possível testar como ela se comporta com diversas requisições. Com isso, será possível encontrar problemas reais. O mercado de certificação digital possui outros produtos por isso, este trabalho pode ser estendido para contemplá-los, criando novos DMs, para que exista uma compreensão total do mercado. [5]

6. REFERÊNCIAS

[1] A. Andrade and J. P. Rossetti. *Governança corporativa: fundamentos, desenvolvimento e tendências*. Atlas, 2004.

[2] R. Housley and T. Polk. *Planning for PKI: best practices guide for deploying public key infrastructure*. John Wiley & Sons, Inc., 2001.

[3] IBGC. Código das melhores práticas de governança corporativa. http://www.ibgc.org.br/index.php/governanca/governanca-corporativa, 2015. Acessado em 11/04/2017.

[4] ITI. Instituto Nacional de Tecnologia da Informação. http://www.iti.gov.br/institucional, 2016. Acessado em 16/04/2017.

[5] ITI. INSTRUÇÃO NORMATIVA Nº 14. http://www.iti.gov.br/images/repositorio/legislacao/instrucoes-normativas/IN_14_2016_-_Envio_dos_certificados_digitais_com_biometriaAss.Dig.pdf, 2016. Acessado em 20/04/2017.

[6] H. S. Singh. Data warehouse: conceitos, tecnologias, implementação e gerenciamento. *Editora Makron Books. São Paulo*, 2001.