Access Control and Collaborative Authoring on the Semantic Web

V. Moll, B. Wilges and R. C. Bastos

***Abstract*— Access control continues to be a major challenge in the information security area. Mechanisms such as OpenID and FOAF+SSL become necessary to supply decentralized authentication, to provide security and to assist in the management of users and their access to resources. Moreover, expectations of collaborative creation and editing still need support to promote Linked Data regardless of the application server. This article presents the mechanisms for decentralized access control and its application possibilities. Also featured are viable alternatives to implement collaborative creation and editing within the context of the Semantic Web.**

***Keywords*— *A*ccess Control, Collaborative Authoring, Semantic Web.**

I. Introdução

A *WORD Wide Web* (WWW) ou simplesmente *Web*, tem revolucionado a forma como que as pessoas se comunicam, compartilham e buscam informações na internet. Atualmente estão em progresso diversas pesquisas que visam descentralizar a informação com o objetivo de facilitar e aumentar a disponibilidade de acesso a mesma. Este e outros temas tem sido preocupações constantes para a W3C, que estabelece os padrões dos formatos (HTML, XHTML, CSS, XML, RDF) utilizados para geração de conteúdo na *Web*.

A *Web* oferece serviços de publicação de conteúdos baseados em Linguagem de Marcação de Hipertexto (*HyperText Markup Language*) ou gerando páginas que são amplamente disponibilizadas nos chamados sites e blogs. Para procurar informações de interesse, os usuários geralmente utilizam um navegador *Web* onde informam as palavras-chave a um motor de busca, que retornam os resultados de uma forma organizada, e também de maneira rápida e eficiente. Até mesmo o processo de busca envolve um esforço descentralizado, já que os motores de busca necessitam indexar uma quantidade imensa de informações distribuídas em vários servidores da internet.

Segundo a Royal Pingdom [1], empresa de monitoramento da internet, até dezembro de 2010 foram criados 21,4 milhões de novos sites, somando 255 milhões de páginas online. Em 2010, 152 milhões de blogs foram criados. Nessa perspectiva, percebe-se que existe, cada vez mais, um empenho em tentativas e soluções que atendam aos usuários de um modo descentralizado. Neste contexto, pode-se enxergar a *Web* semântica como uma extensão da atual *Web*, ou seja, na *Web* atual visa-se o desenvolvimento de sistemas que possibilitem espaços de autoria colaborativa, compartilhamento de informações, além de sistemas de controle de acesso descentralizados para garantir que as informações sejam acessadas e manipuladas por usuários com identidades únicas e que sejam devidamente autorizados. Segundo Hollenbach, Presbrey e Berners-Lee [2], exigir que os usuários tenham contas separadas para cada servidor em que eles editem dados pode se tornar cansativo, o ideal seria ter uma única identidade para o usuário.

Esse artigo apresenta pesquisas e análises de modelos de controle de acesso descentralizado, a proposta deste trabalho é estudar formas de implementar controles de acesso de modo seguro e potencialmente eficaz. As próximas sessões apresentam conceitos fundamentais da *Web* Semântica, também são apresentados diversos trabalhos nessa linha e relacionados ao tema. São apresentadas as tecnologias para o desenvolvimento de controle de acesso descentralizado, implementações que propõem garantias de privacidade, além de questões relacionadas a criação colaborativa. Na sequência são relatadas algumas soluções alternativas na adoção de tecnologias que possam se agregar, e se complementar, visando um maior potencial nas aplicações para *Web* semântica. Ao final do artigo são relatadas as considerações, as perspectivas inseridas dentro dessa proposta de controle descentralizado e criação colaborativa.

II. Web semântica

Na *Web* Semântica a informação possui um significado e os recursos são representados de forma padronizada para permitir o processamento em motores de busca. A representação das informações é feita através do uso do formato RDF - (*Resource Description Framework*) que é uma recomendação da W3C para padronizar a definição e utilização dos metadados, que servem para descrição de recursos da *Web* [3]. O RDF tenta trazer interoperabilidade ante a multiplicidade de formatos incompatíveis existentes.

RDF é utilizado com um padrão para troca de dados na *Web*, pois possui características que facilitam a fusão de dados, mesmo que os esquemas sejam diferentes. Além disso, o RDF amplia a estrutura de links da *Web* porque usa URIs para nomear a relação entre as coisas, bem como relações entre duas extremidades, o que é normalmente referenciado como uma "tripla". A base do RDF é uma tripla do tipo sujeito-predicado-objeto que representam afirmações. Com este modelo simples, ele permite que os dados estruturados e

V. Moll, Universidade Federal de Santa Catarina (UFSC), Florianópolis, Santa Catarina, Brasil, vmoll@das.ufsc.br

B. Wilges, Universidade Federal de Santa Catarina (UFSC), Florianópolis, Santa Catarina, Brasil, beaw@inf.ufsc.br

R. C. Bastos, Universidade Federal de Santa Catarina (UFSC), Florianópolis, Santa Catarina, Brasil, rogerio@inf.ufsc.br

semi-estruturados sejam misturados, expostos e compartilhados entre diferentes aplicações [4]. Pois essa estrutura forma um grafo dirigido, rotulado, representado pelos nós do grafo. Cada tripla representa uma ligação "nó-arco-nó". A Fig. 1 apresenta a forma gráfica da tripla RDF descrita na Tabela 1.

Tabela 1: Descrição da tripla RDF

Tripla	(sujeito,predicado,objeto)
Forma relacional	predicado(sujeito,objeto)
RDF/XML	<rdf:Description rdf:about="subject"> <ex:predicate> <rdf:Description rdf:about="object"/> </ex:predicate> </rdf:Description>

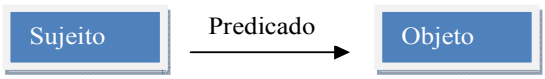

Figura 1. Forma gráfica de uma tripla RDF

III. Controle de Acesso e Personalização da Web

Em 2002 o W3C desenvolveu um sistema chamado W3C ACL *System*, no qual propôs o uso de RDF para controle de acesso a arquivos em servidores de páginas *Web*. Embora tivesse infraestrutura adequada para a *Web* semântica, o sistema proposto tinha algumas limitações: (i) as regras de concessão de acesso a arquivos eram definidas em RDF, (ii) os mecanismos de autenticação e autorização eram gerenciados em um banco de dados centralizado, (iii) as informações armazenadas não associavam um usuário a uma URI (*Uniform Resource Identifier*) e (iv) o compartilhamento das informações com outras organizações exigia o conhecimento prévio da lista de usuários e grupos do sistema.

Em contrapartida o sistema proposto por Hollenbach, Presbrey e Berners-Lee [2], os autores sugeriram melhorias em relação ao W3C ACL e destacam alguns requisitos de projeto importantes para sistemas da *Web* semântica: (i) gerência descentralizada de contas de usuários, (ii) armazenamento distribuído de dados, (iii) autorização baseada em regras, (vi) mecanismo de autenticação descentralizado com desempenho semelhante ao dos mecanismos tradicionais, (v) facilidade de manutenção das informações de controle de acesso. Para alcançar esse objetivo Hollenbach, Presbrey e Berners-Lee [2] utilizaram o FOAF+SSL [5], que provou ser muito útil no desenvolvimento de sistemas de controle de acesso baseado em RDF. No trabalho deles o desenvolvimento da implementação completa do controle de acesso foi através do servidor *Web* Apache. Esse sistema de controle de acesso é realizado em nível de documento, usando as regras escritas em RDF.

Além disso, segundo Heitmann et al. [6], a personalização na *Web* está se tornando um produto, no entanto, privacidade e personalização não estão em acordo. Segundo os autores os sistemas requerem que o usuário confie no serviço de personalização, considerando que não farão uso indevido de seus dados ou comércio dos mesmos. Para evitar isso, na pesquisa deles foi proposta uma arquitetura baseada em *Linked Data*, FOAF e WebIDs, combinando propriedades de privacidade que permitem personalização em arquiteturas centralizadas com portabilidade e descentralização de dados da *Web*. Eles apresentaram uma arquitetura que descreve como combinar a infraestrutura de dados existente na *Web*, bem como seus padrões para gerenciar identidades de modo descentralizado, a fim de disponibilizar a privacidade e portabilidade aos perfis de usuários.

Ainda segundo Heitmann et al. [6], em vez de trabalharmos com armazenamento de dados em sistemas fechados, trabalharemos em um ecossistema universal com perfis de usuários móveis. Ou seja, os perfis de usuários podem ser trocados entre serviços sociais ou, até mesmo, ser hospedado pelo próprio usuário. Essa pesquisa é baseada no trabalho de Hollenbach, Presbrey e Berners-Lee [2], onde a arquitetura proposta descreve como combinar *Linked Data*, WebIDs e vocabulário *Web Access Control* (WAC): *Linked Data*, o *Friend-of-a-Friend* (FOAF) e Semanticamente Interligar Comunidades Online (SIOC), permitindo a descrição de perfil de usuários independente do domínio deles.

Na pesquisa de Mühleisen, Kost e Freytag [7], também se considerou que aplicações sociais são uma das áreas que mais crescem na *Web*. No entanto, de acordo com os autores, questões de privacidade só acontecem se todas as informações de todos os usuários desses aplicativos estivessem armazenadas em um único servidor. Na proposta deles, com pequenas extensões das Tecnologias da *Web* Semântica e conceitos de *Linked Data*, é possível implementar uma abordagem distribuída para a *Web* social. Dessa forma, os usuários mantêm o controle sobre os seus dados e ainda são capazes de combinar os seus dados com usuários em diferentes sistemas. Nesse trabalho as políticas de acesso foram descritas pelo usuário com *Policy-enabled Linked Data Server* (PeLDS) e a autenticação foi realizada utilizando certificados SSL e o enfoque de verificação FOAF + SSL.

De acordo com Hollenbach [8], para realmente aproveitar os benefícios da *Web* Semântica, é preciso desenvolver instrumentos adequados para escrever aplicações *Web* que agreguem, visualizem e editem os diferentes dados que a *Web* Semântica disponibiliza. Em sua pesquisa de mestrado [8] é apresentado uma biblioteca *Widget* em JavaScript para criação de aplicações *Web* que possam ler e escrever na *Web* Semântica. Os *widgets* realizam operações de edição. As regras de controle de acesso a conteúdos são geradas pelo usuário e são suportados pelo FOAF + SSL, que é uma técnica de autenticação descentralizada, a qual permite aos usuários gerenciar de forma independente as restrições sobre seus dados.

Para Gamble e Goble [9], o que falta na *Web* é um mecanismo para suportar: "o que os cientistas compartilham", "e como eles compartilham". Para os autores questões que

envolvem o compartilhamento de dados científicos na *Web* devem ser vinculados. Sendo assim, os autores propõem técnicas de confiança social para compartilhar uma nova classe emergente de objetos digitais científicos (*Research Objects*). Eles sugerem um mecanismo para a introdução de métricas de confiança social na *Web* distribuída facilitando o controle de acesso na agregação de recursos de dados relacionados (*Linked Data*). Nesse trabalho é apresentada a métrica de confiança *Colleague of a Colleague* (Cocoa) para compartilhamento de conhecimento científico.

Em geral todas as abordagem e pesquisas relacionadas aos temas de controle de acesso descentralizado e personalização, se preocupam em atender demandas que estão acontecendo na atual *Web*. Ou seja, tem-se a preocupação de disponibilizar um controle de perfil único, que possa ser integrado entre diferentes sistemas. E dessa forma, pode-se buscar garantias de privacidade de dados. Além disso, existe um esforço em relacionar os dados (*Linked Data*), de forma que uma informação esteja literalmente relacionada com todas as suas possíveis referências na *Web*.

IV. Mecanismos para controle de acesso

Em relação ao controle de acesso, os modelos tradicionais são comumente divididos em: discricionários DAC (*Discretionary Access Control*), obrigatórios MAC (*Mandatory Access Control*) e baseados em papéis RBAC (*Role-based Access Control*) [10]. Todos estes originaram-se a partir da matriz de controle de acesso [11]. As listas de controle de acesso (ACLs) armazenam conjuntos de usuários e seus respectivos direitos de acesso sobre cada objeto do sistema [12].

O mecanismo de autenticação representa uma das etapas do controle de acesso a recursos. Na *Web tradicional*, mecanismos como OpenID e OAuth podem ser utilizados para fornecer autenticação única (*Single Sign-on*), descentralizada e personalizada, com o objetivo de proporcionar segurança e auxiliar no gerenciamento de usuários e seus acessos a recursos. No contexto da *Web* semântica, a autenticação descentralizada pode ser feita utilizando-se o Shibboleth que visa o compartilhamento de contas entre domínios, porém, depende de um processo rígido de federação entre as organizações envolvidas. Outra opção é utilizar o padrão OpenID [13], desenvolvido pela indústria para autenticação de usuários finais em determinada URI através do armazenamento de sua identidade digital em um formato padrão. No entanto, tanto Shibboleth quanto OpenID não cumprem plenamente os requisitos de arquitetura da *Web* (REST - *Representational State Transfer*). Uma terceira alternativa para a autenticação descentralizada é o protocolo FOAF+SSL [5]. FOAF é uma tecnologia que facilita o compartilhamento e uso de informação sobre pessoas e suas atividades. A descrição de uma pessoa é feita em um arquivo RDF para criar uma rede de relacionamentos entre serviços, plataformas estendendo-se a toda a *Web*. O FOAF+SSL é um protocolo de autenticação seguro que possibilita a construção de redes sociais distribuídas, abertas e seguras também conhecidas por *Web* social. A Fig. 2 descreve a arquitetura do processo de autenticação e autorização proposto por Hollenbach, Presbrey e Berners-Lee [2].

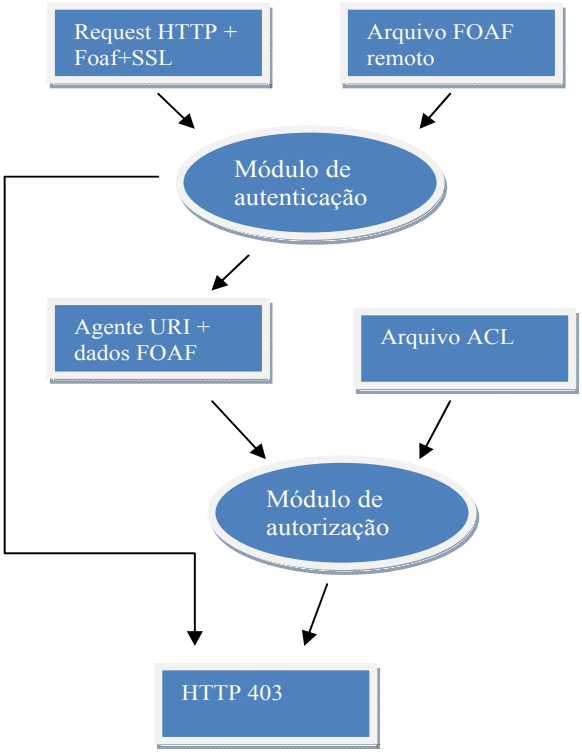

Figura 2. Processo de autenticação e autorização proposto por Hollenbach, Presbrey e Berners-Lee [2].

O servidor é implementado com dois módulos do apache, um para autenticação (mod_authn_webid) e outro para autorização (mod_authz_webid). Esse módulo recebe uma requisição HTTP juntamente com o arquivo FOAF+SSL, os dados do FOAF são comparados com um certificado de arquivo remoto FOAF, se o certificado bater a requisição de autenticação é concedida. No módulo de autorização é utilizado um arquivo de metadados N3 que contém uma lista de controle de acesso (ACL). O servidor direciona os clientes para a ACL de um determinado arquivo usando o cabeçalho do protocolo HTTP.

V. Soluções para criação colaborativa

O termo *Linked Data* refere-se fundamentalmente ao uso da própria *Web* para criar links tipificados de dados provenientes de diversas fontes. Os dois principais recursos utilizados são: (i) o modelo de dados RDF, utilizado para publicar dados estruturados na *Web* e (ii) os links RDF que servem para interligar dados de diferentes fontes conforme estão disponíveis na *Web*. A aplicação destes dois recursos resulta em um espaço comum de dados denominadode *Web data*.

A navegação na *Web* semântica é diferente da navegação feita na *Web* tradicional, isto porque enquanto na *Web* tradicional tem-se apenas os navegadores de páginas HTML

sem ligação entre os dados, na *Web* semântica temos navegadores orientados a *Linked Data*. Ou seja, os usuários podem realizar uma navegação entre os dados referenciados em fontes externas seguindo links RDF. Além disso, assim como as páginas tradicionais com links para outras páginas podem ser rastreadas, da mesma forma, os dados disponíveis na *Web* semântica podem ser rastreados através dos links RDF. Um link RDF indica que uma peça de dados tem algum tipo de relacionamento com outra peça em alguma outra fonte na *Web*.

Na *Web* semântica pode-se utilizar o protocolo SPARQL para acessar dados. Além de dados, pode-se consultar serviços descritos em WSDL 2.0 através do método *query* do SPARQL disponível tanto em HTTP quanto com SOAP para obter dados e serviços de outros sites.

No contexto da *Web* semântica, outra questão que ainda tem muito a ser desenvolvida por pesquisadores da área é a proposta de criação e edição colaborativa que sustente a ideia de *Linked Data*. As propostas atuais para isso, ainda são muito dependentes de uma configuração específica, e isso ocasiona uma série de limitações na implementação. O que existe para o desenvolvimento de criação colaborativa é: o WebDAV e o SPARQL/Update. O WebDAV (*Web-based Distributed Authoring and Versioning*) é uma extensão do protocolo HTTP definido na RFC 4918 [14] e permite aos usuários editar e gerenciar arquivos colaborativamente em servidores *Web*. O SPARQL/Update é uma linguagem que oferece suporte a operações de atualização de arquivos RDF e a *Linked Data*. No caso do sistema proposto em [2], foi desenvolvido um módulo para o *Apache Web Server* e utilizado o WebDAV para a manipulação dos Metadados dos recursos em ACLs. Conforme sugerido pelos autores, trabalhos futuros poderiam utilizar o SPARQL/Update para gerenciar os metadados dos recursos do servidor *Web* e ainda fornecer suporte a *Linked Data*.

Atualmente, para realizar a edição de dados diretamente na *Web*, um usuário pode utilizar extensões em seu *browser*. Por exemplo, o Tabulator é uma extensão para o navegador Firefox. Esta extensão oferece uma interface amigável para a navegação e edição de dados RDF.

Uma vez que tanto o WebDAV quanto o SPARQL/Update podem ser utilizados para realizar a edição de arquivos RDF, verifica-se que é necessário que os servidores *Web* ofereçam suporte a estes protocolos para facilitar o acesso aos dados publicados na *Web*.

Pesquisas recentes [15], tem proposto a extensão dos princípios de *Linked Data* para ligação de dados distribuídos através da *Web*, tais como a extensão aos princípios de modificação, adição e exclusão de recursos, o acesso a conteúdos de mídia, bem como recuperação de conteúdos e metadados do *Linked Data*, implementando o *Linked Media Framework* (LMF).

Esse projeto tem a finalidade de gerar uma plataforma de serviços de TV on-line inteligente o que tem sido um novo setor de crescimento da indústria do entretenimento. Eles já começaram a fazer uso da *Web* Social (Web 2.0) e utilizam os princípios da *Web* Semântica com técnicas, abordagens e ferramentas, dando assim origem à plataformas e serviços de TV na Semântica Social. Esse projeto é integrado com o projeto *Apache Stanbol* (em incubação), desenvolvido pelo projeto IKS que propõe suporte nativo ao *Linked Data*.

Outra proposta bastante interessante é o wiki.ontologi.es escrito em Perl, que permite a criação colaborativa tanto com o protocolo WebDAV quanto pelo SPARQL/Update.

Considerando os recursos fornecidos e o suporte a *Linked Data*, sugerimos a adoção do protocolo SPARQL/Update no mapeamento de recursos de um servidor *Web*, principalmente quando se utiliza tuplas RDF para fornecer suporte a criação colaborativa. Uma vez que o desenvolvimento de aplicações *Web* está associado com a linguagem de programação e plataforma utilizada, existem diversas maneiras de se implementar o suporte nativo a *Linked Data* e a edição de documentos RDF. Conforme apresentado na Fig. 3, uma possível forma de implementação seria utilizar a biblioteca dotnetRDF , utilizando a linguagem C# para publicação no servidor *Web* Microsoft IIS *(Internet Information System*).

Figura 3. Opções propostas para a implementação do suporte nativo a *Linked Data* e edição de documentos RDF em aplicações *Web*.

Uma implementação alternativa seria utilizar o container Glassfish ou TOMCAT e implementar o suporte ao protocolo SPARQL/Update utilizado o framework Jena - A *Semantic Web Framework* for Java. O Jena oferece recursos para manipulação de arquivos RDF, RDF *Schema*, OWL, SPARQL e inclui um motor de inferência baseado em regras.

VI. Conclusões

O suporte a criação colaborativa, edição de conteúdo e controle de acesso em servidores *Web* ainda, que apresente diversas alternativas e soluções que possam ser consideradas eficientes para implementação, têm muitos campos de pesquisa e necessitam de atenção por parte dos pesquisadores da área. Isto porque ainda não existe uma abordagem única e geral o suficiente para atender a todas as expectativas de criação e edição colaborativa e, ainda, controle de acesso descentralizado na perspectiva da *Web* Semântica.

Neste artigo, foram apresentadas algumas iniciativas e pesquisas relacionadas que visam atender as necessidades de fortalecer o desenvolvimento da atual *Web*. Também foram apresentados possíveis direcionamentos para implementação do suporte ao protocolo SPARQL/Update em diferentes servidores *Web*. Com a implementação desse protocolo fica garantida a relação entre os dados (*Linked Data*), um fator extremamente significativo para o desenvolvimento de aplicações na *Web* atual. O objetivo deste trabalho é ampliar a capacidade de publicações e pesquisas sobre dados na *Web* Semântica. Tentativas de implementação e a realização de experimentos poderiam ser objeto de trabalhos futuros.

REFERÊNCIAS

[1] Royal Pingdom. Pesquisa: veja os números da internet no mundo em 2010. Disponível em: <http://royal.pingdom.com/2011/01/12/internet-2010-in-numbers/> Acesso em: 27 mar. de 2011.

[2] Hollenbach, J., Presbrey, J., and Berners-Lee, T. (2009). Using rdf metadata to enable access control on the social semantic web. 2009

[3] Berners-Lee, T. (1998). Semantic Web Roadmap. URL http://www.w3.org/DesignIssues/Semantic.html.

[4] Resource description framework (rdf). URL http://www.w3.org/RDF/. 2011

[5] Story, H., Harbulot, B., Jacobi, I., and Jones, M. (2009). Foaf+ssl: Restful authentication for the social web.

[6] Heitmann, B. and Hayes, C. (2010). Achieving privacy-enabled user profile portability with webid and the web of data? 2010

[7] Muhleisen, H., Kost, M., and Freytag, J. (2010). Swrl-based access policies for linked data. In Proceedings of the ESWC 2010 2nd Workshop on Trust and Privacy on the Social and Semantic Web, Heraklion, Greece.

[8] Hollenbach, J. (2010). A Widget Library for Creating Policy-Aware Semantic Web Applications. PhD thesis, Massachusetts Institute of Technology.

[9] Gamble, M. and Goble, C. (2010). Standing on the shoulders of the trusted web: Trust, scholarship and linked data.

[10] Sandhu, R. S., Coyne, E. J., Feinstein, H. L., and Youman, C. E. (1996). Role-based Access Control Models. In IEEE Computer, pages 38–47.

[11] Lampson, B. (1974). Protection. In Proceedings of 5th Princeton Symp. on Information Science and Systems, pages 437–443. ACM.

[12] Barkley, J. (1997). Comparing simple role based access control models and access control lists. In Proceedings of the second ACM workshop on Role-based access control, RBAC '97, pages 127–132, New York, NY, USA. ACM.

[13] Recordon, D. and Reed, D. (2006). Openid 2.0: a platform for user-centric identity management. In Proceedings of the second ACM workshop on Digital identity management, DIM '06, pages 11–16, New York, NY, USA. ACM. URL http://doi.acm.org/10.1145/1179529.1179532.

[14] RFC 4918. HTTP Extensions for Web Distributed Authoring and Versioning (WebDAV), June 2007. Disponível em: <http://www.rfc-editor.org/info/rfc4918> Acesso em: ago de 2011.

[15] Violeta Damjanovic, Georg Güntner, Sebastian Schaffert, Thomas Kurz, Dietmar Glachs. Semantic Social TV. NEM Summit 2011, 28 September, Torino, Italy

Vinícius Moll received the master degree in Computer Science from the Federal University of Santa Catarina, Florianópolis, Brazil, in 2010, and he is a PhD student in Engineering and Automation Department. His current research interests includes security in networks, cloud computing, virtualization and identity management systems.

Beatriz Wilges received the master degree in Computer Science from the Federal University of Santa Catarina, Florianópolis, Brazil, in 2008, and is a PhD student in Engineering and Knowledge Management. Her current research interest includes modeling and data mining, knowledge management and multi-agent systems.

Rogério Cid Bastos has a PhD in Production Engineering. He is a professor in the Federal University of Santa Catarina, Florianópolis, Brazil. His current research interest are knowledge management, information technology and fuzzy logic.