

Ferramentas e Bases de Dados Open Science para Pesquisa em Inovação.

André Moraes Santos¹, Cláudia Terezinha Kniess², Luc Quoniam³, Emerson Antonio Maccari⁴

¹ Universidade do Vale do Itajaí (UNIVALI) - amsantos@univali.br

² Universidade Nove de Julho (UNINOVE) - kniesscl@yahoo.com.br

³ Aix Marseille Université/ France - mail@quoniam.info

⁴ Universidade Nove de Julho (UNINOVE) - emersonmaccari@gmail.com

PALAVRAS-CHAVE

Inovação;
Big Data;
Ciência aberta.

Received 20.11.2017

Revised 24.11.2017

Accepted 28.11.2017

ISSN 1980-4431

Double blind review



RESUMO

Este artigo tem como objetivo propor um conjunto de ferramentas para a recuperação e análise de informações em bases de dados públicas, úteis ao estudo da inovação. Como contexto de aplicação, foram selecionadas duas bases públicas: (a) Base Nacional do Diretório de Grupos de Pesquisa; (b) base internacional de patentes Espacenet. O modelo de sistema de inovação de hélice quádrupla foi escolhido para ilustrar as possibilidades de análise. A pesquisa seguiu os princípios da ciência aberta (Open Science), sendo todos os dados, artefatos e resultados obtidos disponibilizados de forma livre e aberta. O método que orientou o desenvolvimento da pesquisa foi o “Design Science Research”. Utilizou-se as ferramentas Patent2Net, ScriptGP2 e Gephi. Com base nos artefatos e bases de dados, foi proposto um framework para a extração e análise das informações. Demonstrou-se que os recursos de dados, frameworks e artefatos identificados e desenvolvidos nesta pesquisa permitem estudar diferentes relações e aspectos da inovação, tanto no âmbito nacional quanto internacional.

KEYWORDS

Design Science;
Big Data;
Innovation.

ABSTRACT

This article aims to propose a set of tools for the retrieval and analysis of information in public databases, useful for the study of innovation. As an application context, two public databases were selected: (a) National Database of the Directory of Research Groups; (b) International patent base Espacenet. The quintuple propeller innovation model was chosen to illustrate the possibilities of analysis. The research followed the principles of Open Science, with all the data, artifacts and results obtained being made freely and openly available. The method that guided the development of the research was the "Design Science Research". The tools Patent2Net, ScriptGP2 and Gephi were used. Based on the artifacts and databases, a framework was proposed for the extraction and analysis of information. It has been demonstrated that the data resources, frameworks and artifacts identified and developed in this research allow us to study different relations and aspects of innovation, both nationally and internationally.

1 Introdução

O desenvolvimento tecnológico instrumental é central para o avanço do conhecimento científico. Novos equipamentos, softwares e metodologias permitem ao pesquisador coletar e analisar dados, até então impossíveis ou extremamente difíceis de serem obtidos. As novas tecnologias para o sequenciamento de DNA, por exemplo, permitiram grandes avanços na área da genética e medicina, assim como os aceleradores de partículas permitiram o estudo das estruturas subatômicas no campo da física (Goodwin, McPherson, & McCombie, 2016).

No campo científico da administração, também experimentou-se significativos avanços em função de novos métodos de pesquisa e ferramentas computacionais, principalmente com a abordagem de *BigData* (George, Osinga, Lavie, & Scott, 2016). Combinando conhecimentos de programação de computadores, métodos científicos e estatística, foi possível desenvolver novas ferramentas para explorar grandes bases de dados. Acompanhando o desenvolvimento de novas ferramentas, a comunidade científica também tem estimulado a Ciência Aberta, do inglês, *Open Science*, o que significa tornar público e acessível todos os produtos e processos envolvidos em uma pesquisa científica (OECD, 2016).

As pesquisas sobre inovação também podem se beneficiar dos avanços nos instrumentais tecnológicos e nas polícias de Ciência. As bases mundiais de patentes são excelentes fontes para estudos sobre inovação e competitividade, tanto em mercados desenvolvidos quanto em econômicas emergentes (J. Kim & Lee, 2015). Da mesma forma, bases nacionais de pesquisa científica e inovação, como a Plataforma de Grupos de Pesquisa Lattes, por exemplo, podem ser úteis em função da qualidade e grande volume de dados. Ambas são bases abertas, disponíveis ao acesso público. Combinando estas fontes de dados com ferramentas específicas para extração e análise, é possível explorar novas possibilidades de dados para pesquisa.

Este artigo demonstra como utilizar ferramentas computacionais, baseadas em código de software livre (*open source*), para extrair e analisar grandes volume de dados para pesquisas em inovação. Tendo como fonte as bases internacionais de patentes e plataforma nacional de

grupos de pesquisa é possível extrair e analisar dados que seriam extremamente difíceis ou custosos de serem obtidos utilizando-se apenas os processos normais. Espera-se que a solução proposta que possa utilizada por pesquisadores em problemas reais voltados à operacionalização de suas pesquisas.

Para ilustrar as possibilidades de uso e aplicação das ferramentas e dados, tomou-se como base o modelo de inovação de hélice quádrupla. O modelo de hélice quádrupla deriva do modelo de tripla hélice e define um sistema de inovação como sendo um conjunto de relações entre governo, empresas, universidades e sociedade, pautadas pela responsabilidade ambiental (Carayannis & Rakhmatullin, 2014; Casaramona, Sapia, & Soraci, 2015). Para que se possa estudar o modelo, é necessário que existam informações disponíveis sobre governos, empresas, universidades, tecnologias e impacto social envolvidos no sistema. Também é importante identificar as relações existentes entre os atores capazes de promover a inovação por meio da produção, comercialização e uso de novos conhecimentos e tecnologias (Casaramona et al., 2015; Etzkowitz, 2003; Nelson, 1993; OECD, 1997).

O artigo está estruturado em quatro seções principais onde será apresentada a base teórica, método, resultados e discussões finais.

2 Base Teórica

Nesta seção será apresentada a base teórica e conceitual utilizada para o *framework* de extração e análise de informações proposto por este artigo. O Quadro 1 apresenta cada um dos principais tópicos e sua aplicação no contexto deste artigo.

Quadro 1. Tópicos e avaliação do contexto do estudo

Base teórica/conceitual	Aplicação nesta pesquisa
Modelo de hélice quádrupla	Modelo teórico que dá suporte a identificação de atores (universidades, empresas, governo e sociedade) e suas relações em um sistema de inovação dentro de um contexto de sustentabilidade.
Ciência Aberta e Responsabilidade social da pesquisa	Base teórica que justifica e dá suporte ao uso de ferramentas abertas e acesso incondicional aos dados e resultados da pesquisa científica e justifica a necessidade de aplicabilidade e

	retorno social da pesquisa acadêmica.
Transdisciplinaridade	Base teórica que justifica e dá suporte à análise da participação de diferentes disciplinas científicas em um problema
Artefatos tecnológicos	Base teórica que justifica a relevância científica da construção de artefatos que permitam explorar informações para a solução de problemas do mundo real.
Análise de redes sociais	Modelo teórico que dá suporte a análise das relações entre os atores de um sistema de inovação.
Modelo de Recuperação de informações	Modelo que indica a relação entre os termos de busca e as informações disponíveis na base de dados.

Fonte: Elaborada pelos autores (2017).

2.1 Modelos para análise da inovação: Hélice tripla, quádrupla e quádrupla

No final da década de 90, Leydesdorff e Etzkowitz (1996) observaram que a transição da economia industrial para a economia do conhecimento havia causado uma alteração nos papéis tradicionais do governo, empresas e universidades. Os autores perceberam que para compreender a dinâmica da inovação era necessário estudar como os próprios atores do sistema se reconfiguravam e redefiniam suas funções (Etzkowitz & Leydesdorff, 2000). A dupla de pesquisadores lançou mão do conceito elaborado por Sabato e Botana (1968) sobre a tríade de atores de um Sistema Nacional de Inovação (SNI), universidade-empresa-governo, e acrescentou a possibilidade de sobreposição e até alteração das funções destes atores, chamando este modelo de hélice tripla, do inglês, “triple-helix”. (Leydesdorff, 2012).

O modelo de hélice tripla sustenta que, na economia do conhecimento, a inovação está ligada a um papel mais atuante da universidade e na hibridação das funções da universidade, indústria e governo para gerar novos formatos institucionais e sociais para a produção, a transferência e aplicação do conhecimento (Chung, 2014; Etzkowitz & Leydesdorff, 2000; Ivanova & Leydesdorff, 2014; Villarreal & Calvo, 2015).

O modelo de hélice quádrupla foi proposto Carayannis e Campbell (2009) com base no

modelo de hélice tripla e inclui a sociedade como o quarto elemento ativo no processo de inovação. A sociedade participa ativamente na cocriação e coprodução de inovações. Ideias inovadoras podem surgir do meio social e serem financiadas socialmente, como nos crescentes exemplos de *crowdfunding* e *crowdsourcing* (Afuah & Tucci, 2012; Carayannis & Campbell, 2012; Mollick, 2014). Além disso, o estilo de vida, valores culturais, manifestações artísticas e outras características de uma sociedade, expressadas através dos meios de comunicação, indicam as prioridades e desejos sociais que devem ser considerados pelos atores de um sistema de inovação.

Aprofundando a pesquisa sobre a complexidade das relações entre agentes de um sistema de inovação de hélice tripla e quádrupla, Carayannis e Campbell (2010) exploraram a relação que os sistemas de inovação tem com o ambiente natural (meio ambiente, sob a perspectiva ecológica) e propuseram uma extensão ao modelos anteriores adicionando uma quinta hélice.

O modelo de hélice quádrupla é “um modelo que se baseia, e se especializa, no conjunto das interações sociais e intercâmbios acadêmicos em um Estado (estado-nação) com o objetivo de promover e evidenciar um sistema cooperativo de conhecimentos, habilidades e inovações para um desenvolvimento mais sustentável” (Carayannis & Campbell, 2010, p. 62). O modelo de hélice quádrupla considera que a desenvolvimento econômico baseado na inovação depende da interação entre universidade, empresas, governo e sociedade civil, em um contexto de desenvolvimento sustentável (Carayannis, Barth, & Campbell, 2012). A quinta hélice representa a visão transdisciplinar para a geração de inovações compatíveis com os princípios da sustentabilidade ambiental (Carayannis & Campbell, 2010). O equilíbrio entre o desenvolvimento econômico e o ambiente natural tornou-se imperativo para a própria existência humana, sendo endossado por todos os países membros Organização das Nações Unidas como prioridade mundial (Griggs et al., 2013). Como exemplo, Casaramona, Sapia e Soraci, (2015) utilizaram o modelo de hélice quádrupla para analisar a cooperação internacional na busca de soluções inovadoras para energias renováveis em países da região do mediterrâneo. O modelo de hélice quádrupla também foi utilizado por, Bifulco et al. (2016) para analisar as inovações

tecnológicas na construção de cidades inteligentes. Além de acrescentar o elemento ambiental, o modelo de hélice quádrupla também destaca a transdisciplinaridade e interdisciplinaridade do conhecimento como forma de abordar a complexidade dos problemas ambientais. Por serem problemas que envolvem o ambiente e a sociedade, simultaneamente, é necessário o envolvimento de várias áreas do conhecimento, como ciências naturais, engenharias, ciências sociais e humanas (Carayannis & Rakhmatullin, 2014).

2.2 Ciência Aberta (*Open Science*) e Responsabilidade Social da Pesquisa

Ciência aberta, do inglês, *Open Science*, é uma perspectiva que significa tornar público e acessível todos os produtos e processos envolvidos em uma pesquisa científica, utilizando, preferencialmente, a World Wide Web como forma de divulgação e acesso (Kraker, Leony, Reinhardt, & Beham, 2011). A iniciativa para a ciência aberta vem sendo adotada por diversos países, instituições e pesquisadores (OECD, 2016). Em 2012, a Comissão Europeia recomendou a todos os Estados-Membros que disponibilizassem o acesso livre aos resultados de pesquisas financiadas com recursos públicos, a fim de tornar a ciência melhor e reforçar a economia baseada no conhecimento (Kuchma & others, 2014). O princípio que norteia a premissa de “Ciência Aberta” está no direito fundamental de acesso ao conhecimento pela sociedade. O conhecimento é compreendido como um fator essencial de valorização, mobilidade social, democratização e desenvolvimento. Assim, os atores envolvidos na produção e gestão do conhecimento devem estar comprometidos com a promoção, valorização, divulgação e compartilhamento do conhecimento junto à sociedade.

A abertura da ciência também enseja a discussão sobre a responsabilidade social da pesquisa científica em proporcionar a justa aplicação do conhecimento na solução de problemas da inclusão social, desenvolvimento econômico e social e da defesa do meio ambiente (Faria & Sauerbronn, 2008). Este movimento de ênfase na responsabilidade social da pesquisa científica também vem sendo apoiado por diversos países, os quais estimulam a crítica e o acesso livre aos resultados de pesquisas financiadas com

recursos públicos a fim de promover a avaliação e difusão do conhecimento gerado (Kuchma & others, 2014).

2.3 Transdisciplinaridade

A emergência de problemas complexos nas últimas décadas, como por exemplo a mudança climática global ou a pesquisa para o câncer, têm exigido novas abordagens de produção de conhecimento e práticas de pesquisa (Groß & Stauffacher, 2014; Podestá, Natenzon, Hidalgo, & Toranzo, 2013; Rosenfield, 1992). Para tratar problemas complexos, os grupos científicos estão buscando arranjos de forma multi, inter ou transdisciplinar. De uma forma geral, a transdisciplinaridade busca a articulação dos conhecimentos de diferentes campos científicos visando a solução de problemas complexos (Gibbons & Nowotny, 2001). Para obter a transdisciplinaridade, cientistas, profissionais e demais atores sociais precisam se organizar em grupos capazes de promover a integração de diferentes perspectivas de identificação, formulação e busca de soluções para um problema em comum (Stokols, Hall, Taylor, & Moser, 2008). Os grupos transdisciplinares têm sido utilizados por governos, agências de fomento e instituições privadas como uma abordagem útil na solução de problemas complexos de natureza ambiental, social, econômica e de saúde (Huutoniemi, Klein, Bruun, & Hukkinen, 2010; Wagner et al., 2011).

2.4 Artefatos Tecnológicos

Os artefatos são as criações da mente humana, ou seja, artificiais, pois não existem na natureza (Simon, 1996). No contexto do *Design Science Research* os artefatos são objetos de pesquisa, desenvolvimento e construção para a solução de problemas de interesse de organizações, grupos ou indivíduos. Ainda que existam diferentes classificações para os artefatos, para Dresch et al. (2015) os quatro tipos mais comumente utilizados nas pesquisas de DSR são os propostos por March e Smith (1995): constructo, modelo, método ou instanciação. No campo das ciências da gestão, por exemplo, os métodos de PERT/COM podem ser considerados artefatos úteis para solucionar problemas de gestão de projetos (Lacerda, Dresch, Proença, & Antunes Júnior, 2013). Um artefato do tipo

instanciação é aquele capaz de ser executado no ambiente de contexto do problema. Neste caso, a execução do artefato e as regras para o seu funcionamento fazem parte da instanciação. Um exemplo são os *software* para computadores (Venable & Baskerville, 2012).

2.5 Análise de Redes Sociais

A análise de redes sociais (ARS) é o nome do campo de estudo que tem por objetivo a análise das relações existentes entre diferentes atores de uma rede (Wasserman, 1994). Por meio da representação e análise das relações é possível identificar padrões estruturais, grupos, atores principais, entre outras características (Zancan, dos Santos, & Campos, 2012). Na área de administração é bastante conhecido a influência das redes sociais nos negócios e organizações (Fracassi, 2017). No contexto desta pesquisa, a análise de redes sociais é importante para compreender a dinâmica dos relacionamentos entre os diferentes atores e elementos previsto no modelo de hélice quádrupla de inovação.

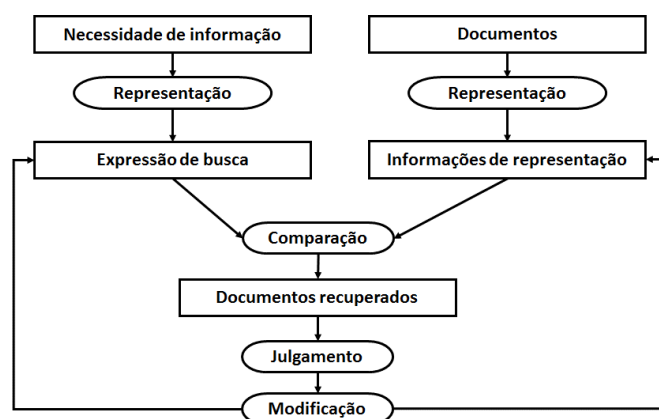
Uma rede é feita de nós e laços, também chamados de arestas, que conectam estes nós. Em uma representação simples de rede temos um conjunto de nós interligados por laços. Representações mais complexas também podem incluir variações de tamanhos para nós e laços, indicando diferentes valores para cada elemento. A partir destas relações é possível analisar aspectos de centralidade, fluxo de informações e conectividade da rede (Marteleto, 2001). Pesquisadores na área de redes sociais tem desenvolvido diferentes métricas para a avaliação e caracterização da dinâmica de uma rede sendo as mais comumente utilizadas: centralidade de informação (*degree* ou *degree centrality*), centralidade de proximidade (*closeness centrality*) e centralidade de intermediação (*betweenness centrality*) (Freeman, 2004; Y. Kim, Choi, Yan, & Dooley, 2011).

2.6 Modelo de Recuperação de informações

Recuperar as informações pertinentes em bases de dados é um dos grandes desafios quando se trabalha com grandes volumes de dados. Em geral, as bases de dados oferecem mecanismos de busca que permitem ao usuário indicar palavras ou termos que serão utilizados para encontrar

resultados que correspondam a estes termos. No modelo tradicional de recuperação de informações (Figura 1), Belkin et al. (1993) identificam dois componentes principais: o usuário e o sistema de recuperação. Do ponto de vista do usuário, este deve definir corretamente quais termos definem adequadamente o tema que está sendo buscado. Já o sistema de busca deve representar e indexar adequadamente os documentos para que os resultados atendam aos critérios de busca do usuário.

Figura 1. Modelo de recuperação de informações



Fonte: Belkin (1993, p. 3).

3 Método

Esta pesquisa tem como objetivo a proposição do uso de um conjunto de artefatos para a solução de um problema prático inerente ao avanço científico. Assim, buscou-se apoio na perspectiva conhecida como *Design Science Research*.

A *Design Science Research* (DSR), tem se apresentado como uma abordagem alternativa a dominância da pesquisa comportamental nas áreas de administração, sistemas de informações e contabilidade, por exemplo (Dresch et al., 2015). Na visão de pesquisa em *Design Science*, o foco está no uso de teorias para o projeto e desenvolvimento de soluções para problemas do mundo real, geralmente na forma de objetos, métodos, como constructos, modelos, métodos e instâncias de soluções tecnológicas (March & Smith, 1995; Van Aken, 2005).

3.1 Etapas da *Design Science Research*

Diversas formas de desenvolvimento prático da pesquisa em *Design Science* podem ser

encontrados na literatura de diferentes áreas como, por exemplo, saúde (Abelson et al., 2003), gestão (Van Aken, 2005), engenharia (Lacerda et al., 2013). e sistemas de informações (Arnott & Pervan, 2012; Gregor & Hevner, 2013; A. Hevner & Chatterjee, 2010; A. R. Hevner, March, Park, & Ram, 2004). Dresch et al. (2015) revisaram 13 artigos principais que apresentam formas de desenvolvimento de pesquisas área de Design Science e destacaram quatro etapas principais comuns a maioria das propostas: Definição e contexto do problema; Desenvolvimento da solução (artefato); Avaliação da solução (artefato); Comunicação dos resultados.

3.1.1 Definição e Contexto do problema

A fase inicial de uma pesquisa se deve especificar o problema em seu contexto e justificar a relevância de uma solução (Peppers et al., 2007). Nesta etapa, o pesquisador deve atentar para a definição do problema e para as características desejáveis de uma solução satisfatória (Dresch et al., 2015).

3.1.2 Desenvolvimento da solução (o artefato)

Com base nas especificações e conhecimentos sobre o problema, o pesquisador deve iniciar a busca de uma solução. Uma vez que o problema está claramente definido e possíveis soluções são apontadas, o pesquisador deverá propor o uso, adaptação ou aperfeiçoamento de um artefato já existente ou especificar um novo artefato. Esta é uma etapa marcada por intenso esforço criativo e geralmente envolve o uso de conhecimentos transdisciplinares (Dresch et al., 2015). Em seguida, o pesquisador especifica as características do artefato e os procedimentos para sua construção e avaliação. Os resultados esperados do artefato precisam garantir uma solução satisfatória para o problema em questão (Dresch et al., 2015).

3.1.3 Avaliação da solução (artefato)

A quarta fase consiste na avaliação do artefato. Nela o pesquisador precisa demonstrar evidências do uso do artefato na solução dos problemas identificados. Também é importante descrever as limitações do artefato em alcançar os objetivos inicialmente previstos (Dresch et al.,

2015). A avaliação do artefato é importante para a **validade pragmática** da pesquisa. A validade pragmática é, segundo Drescher (2015), a confirmação da utilidade prática da solução proposta para o problema. A solução deve ser viável dentro das restrições que se apresentam no ambiente do problema a ser solucionado. A validade pragmática busca assegurar que a solução proposta para resolver determinado problema de pesquisa de fato funcione, garantindo que os resultados esperados sejam alcançados (Van Aken, 2005).

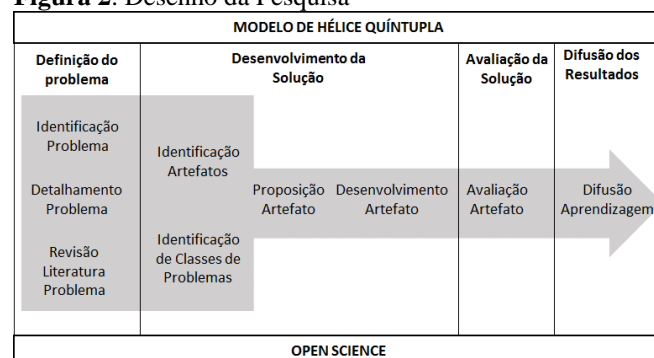
3.1.4 Comunicação dos resultados

A última fase é composta pela comunicação dos resultados e explicitação da aprendizagem gerada com o desenvolvimento do artefato. Além disso, as limitações, encaminhamentos futuros e outros fatos relevantes sobre o desenvolvimento do artefato são formalizados. Segundo Dresch et al. (2015, p. 133) “o objetivo dessa etapa é assegurar que a pesquisa realizada possa servir de referência e como subsídio para a geração de conhecimento, tanto no campo prático quanto no teórico”.

3.2 Desenho de Pesquisa

O desenho da presente pesquisa, ilustrado na Figura 2, tomou como base a proposta do *Design Science Research* para a identificação dos problemas relativos a questão de pesquisa, o desenvolvimento dos artefatos para solucioná-los e a validação destes artefatos. O desenvolvimento da pesquisa também tem como elemento norteador, o modelo de hélice quádrupla, para o qual deseja-se identificar elementos e relações nas bases de dados. Por fim, toda a pesquisa está baseada no conceito de ciência aberta.

Figura 2. Desenho da Pesquisa



Fonte: Elaborada pelos autores (2017).

3.3 Contexto de aplicação

3.3.1. Base internacional de Patentes

Uma das principais fontes mundiais de dados sobre tecnologia e inovação é o registro de patentes. A análise do conjunto de patentes pode servir como uma base de dados estratégica para o desenvolvimento de políticas públicas, geração de novas tecnologias e criação de novos conhecimentos científicos e tecnológicos (Jeong & Yoon, 2014; Tekic, Drazic, Kukulj, & Vitas, 2014). Bases internacionais de patentes ou indicadores de inovação são particularmente úteis como fontes secundárias para as pesquisas que analisam economias emergentes e cenários internacionais. (Cândido, Wielewick, & Zimmermann, 2016). Com relação às patentes, diversas bases tem sido criadas e disponibilizadas, como, por exemplo, Google Patent Search (www.google.com/patents), Patentscope (<http://www.wipo.int/patentscope/en/>), USPTO (<http://www.uspto.gov/>) e Espacenet (<http://worldwide.espacenet.com/>) (Ferraz, Quoniam, Reymond, & Maccari, 2016). Destas, escolheu-se a *Espacenet*, uma base mantida pelo *European Patent Office* (EPO), pois dispõe de serviços livres e gratuitos além de ser uma das bases internacionais com maior cobertura de dados (Jürgens & Herrero-Solana, 2015).

3.4.2 Plataforma Lattes - O Diretório dos Grupos de Pesquisa

O Diretório dos Grupos de Pesquisa no Brasil é uma parte integrante da Plataforma Lattes e pode ser considerado como o mais completo inventário dos grupos de pesquisa em atividade no país (Ramos & Machado, 2014). Dentre as informações constantes na base, identificam-se os recursos humanos constituintes dos respectivos grupos, as linhas de pesquisa propriamente dita, os setores de atividades envolvidos, as especialidades do conhecimento, a produção científica, tecnológica e artística e os padrões de interação com o setor produtivo. Além disso, as informações individuais de cada participante estão vinculadas aos seus respectivos currículos Lattes. No que tange às atualizações, elas são realizadas de forma compulsória pelos líderes de grupos, pesquisadores, estudantes e dirigentes de pesquisa das instituições participantes, pois constituem

critério de avaliação e elegibilidade pela Capes, CNPq e outros órgãos de fomento à pesquisa (BRASIL, MCTI., 2013).

4 Análise dos Resultados

Seguindo a perspectiva da *Design Science Research*, os resultados serão apresentados na sequência das quatro etapas principais: Definição e contexto do problema; Desenvolvimento da solução (artefato); Avaliação da solução (artefato); Comunicação dos resultados.

4.1 O problema

O principal problema encontrado para se extrair e analisar informações da base de patentes e da base de grupos de pesquisa é a limitação do processo de busca e a carência de ferramentas *open source* para a análise do grande volume de dados extraído. Em geral, os mecanismos de busca apenas permitem a recuperação das informações e algumas análises simples, mas não possuem funcionalidades para analisá-las em sua complexidade.

4.2 Desenvolvimento da solução

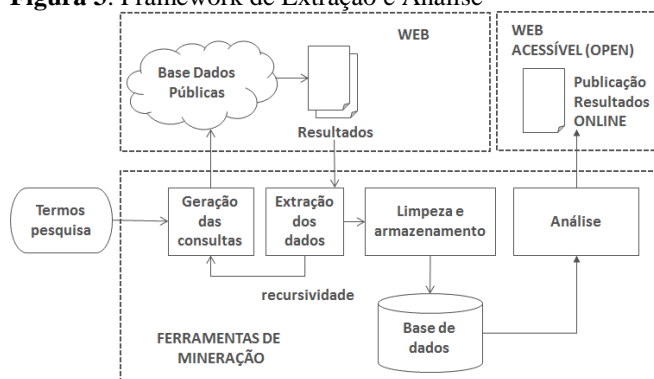
Para a extração e análise nas bases de patentes, identificou-se ferramentas comerciais - como, Intellixir (<http://www.intellixir.com/>); Matheo Patent (<http://www.matheo-software.com/en/>) e Patent Inspiration (<http://www.patentinspiration.com/>), todas comerciais, e gratuitas, baseadas em código aberto, como Lens software (<https://www.lens.org/lens/>) e o Patent2net (<http://patent2net.vlab4u.info/>) (Ferraz et al., 2016). As ferramentas comerciais possuem custos de uso e licenciamento bastante altos, o que dificulta sua difusão e uso, principalmente por estudantes, pesquisadores, pequenas e médias empresas e países emergentes ou em desenvolvimento. Como alternativa, comparou-se os dois sistemas gratuitos e abertos, Lens e Patent2Net, para verificar qual deles seria capaz de atender os requisitos para a solução satisfatória do problema. O sistema Patent2Net demonstrou-se mais adequado, atendendo à um número maior de requisitos e, principalmente, permitindo a exportação dos resultados. O sistema Lens, embora permita a exportação de dados, limita-os a apenas 1000 resultados por pesquisa.

Para a extração e análise nas bases de dados do Diretório de Grupos de Pesquisa da Plataforma Lattes, foi utilizada a ferramenta ScriptGP2, de código aberto, disponível em <https://bitbucket.org/vlab4u/scriptgp>. O ScriptGP2 já foi utilizada com sucesso para a extração e análise de grande volume de dados do diretório de grupos de pesquisa (Magalhães, Quoniam, Mena-Chalco, & Santos, 2014; Santos, Kono, & Quoniam, 2014).

4.2.1 Proposta de framework

Para realizar a extração e análise de dados nas bases de patentes e diretório de grupos de pesquisa foi proposto um *framework* que inclui uma série de etapas que vão desde a definição dos termos de pesquisa, configuração do ambiente computacional, extração dos dados e mineração e análise, conforme a Figura 3.

Figura 3. Framework de Extração e Análise



Fonte: Elaborada pelos autores (2017).

As ferramentas Patent2Net e ScriptGP2 geram um conjunto substantivo de resultados em formato html, com recursos de tabelas dinâmicas (*Pivot Tables*) e gráficos interativos. Além dos resultados tabulares, ambas as ferramentas geram arquivos no formato de grafos, podendo ser importados por softwares específicos para a análise de redes. No caso desta pesquisa, utilizou-se o software Gephi, o qual é uma ferramenta gratuita e de código aberto desenvolvida para a análise de redes gráficas (Bastian, Heymann, Jacomy, & others, 2009). Ele pode ser obtido gratuitamente no endereço <https://gephi.org/>. O software apresenta recursos básicos e avançados para a análise são as funções de visualização, layout e estatísticas.

4.3 Avaliação da solução

A solução proposta, composta pelos artefatos selecionados e *framework* de execução, foi testada com a aplicação em dois temas: patentes em reciclagem de terras raras e grupos de pesquisa em educação a distância. A escolha dos temas foi induzida por interesses dos pesquisadores envolvidos e pela ligação com o propósito de ciência aberta e sustentabilidade.

Foi avaliada a capacidade de extração e análise de informações que possibilitassem a identificação de elementos e relações para o modelo de hélice quintupla. Conforme o modelo de hélice quintupla, buscou-se informações e relações entre governo, empresas, universidades e sociedade, pautadas pela responsabilidade ambiental (Carayannis & Rakhmatullin, 2014; Casaramona et al., 2015).

4.3.1 Extração e análise de informações em patentes sobre reciclagem de terras raras

Terras raras é o nome dado à um grupo de elementos químicos composto pelos lantanídeos mais o escândio e ítrio, com diversas aplicações na indústria, sendo considerado estratégico para os setores automobilístico, aeroespacial, eletrônico, militar e médico (Dutta et al., 2016). A reciclagem de terras raras tem se tornado um tema de interesse acadêmico e comercial em razão da crescente demanda para uso em novas tecnologias eletrônicas (Massari & Ruberti, 2013), da escassez na natureza e impacto ambiental para extração (Bandara, Mantell, Darcy, & Emmert, 2015).

A expressão de consulta da base ESPACENET contemplou as patentes que continham referências à reciclagem de terras raras. Transformada para a linguagem de consulta do ESPACENET, a expressão de busca ficou no seguinte formato:

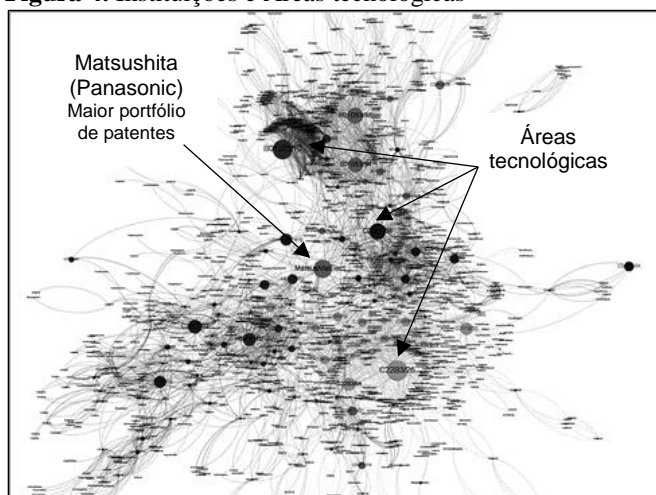
```
((ta="rare earth*" or ta="rare earths*") and (ta=recy*)).
```

Esta expressão utiliza a lógica booleana e permite recuperar patentes que tenham em seu título ou abstract os termos desejados. A expressão de busca foi inserida no software Patent2Net e resultou em 1603 patentes, registradas entre 1908 até 2014, totalizando 70Mb de dados brutos.

4.3.1.1 Análise dos dados para o modelo de hélice quintupla

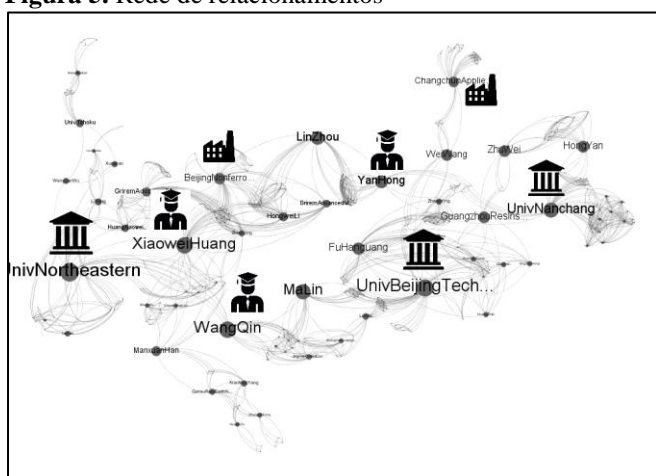
Os resultados fornecidos pelo Patent2Net, adicionados aos recursos de análise de redes da ferramenta Gephi, permitiram identificar relações entre empresas, universidades e pesquisadores em função das informações patentárias. Para exemplificar os resultados obtidos, selecionou-se a empresa Matsushita, uma empresa do grupo Panasonic, a qual possuía o maior número de patentes em terras raras. A partir desta empresa, mapeou-se as relações com as áreas tecnológicas de suas patentes, representadas na Figura 4. Por se tratarem de redes complexas, a forma mais adequada para sua visualização é no formato eletrônico, preferencialmente em mídias com alta resolução.

Figura 4. Instituições e Áreas tecnológicas



Fonte: Dados da pesquisa.

Figura 5. Rede de relacionamentos



Fonte: Dados da pesquisa.

Em outra análise, foi possível identificar redes que interligam diferentes inventores, universidades e empresas, conforme demonstrando

na Figura 5. A Figura 5 exemplifica as possibilidades de análises proporcionadas pela presente pesquisa em um recorte de interações de patentes chinesas. Nela é possível verificar que certos inventores estão ligados a diferentes universidades e empresas, o que representa um forte indício de relacionamento entre estes atores. Também é possível verificar relacionamentos entre depositantes quando ambos depositaram conjuntamente uma patente.

4.3.2 Extração e análise de informações dos Grupos de Pesquisa em EAD

O EAD é uma das formas de ensino inovador que mais cresce no contexto nacional e internacional, representando uma importante ferramenta para a democratização do conhecimento (Barak, Watted, & Haick, 2016). Desta forma, buscou-se identificar os elementos do modelo de hélice quádrupla a partir das informações dos Grupos de Pesquisa em EAD.

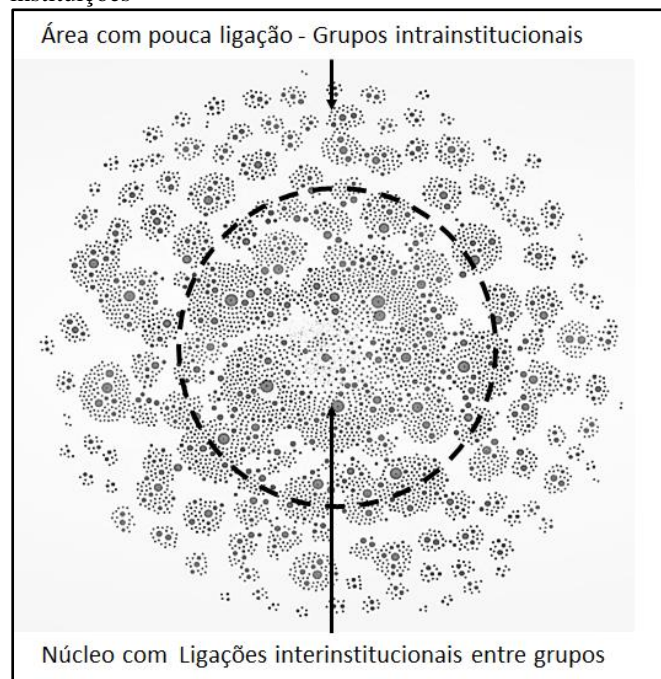
A expressão de busca formulada para a pesquisa na base de Grupos de Pesquisa continha 109 termos. Isto foi necessário devido ao amplo conjunto de palavras que podem estar relacionadas ao contexto de pesquisa EAD. Por razões de espaço a mesma não é reproduzida neste artigo. A expressão de busca foi inserida na ferramenta ScriptGP2 e executada em 2015 a partir do diretório de grupos de pesquisa do CNPq (<http://dgp.cnpq.br/>). Como resultado foram identificados 608 grupos de pesquisa e 4974 pesquisadores associados aos grupos.

4.3.2.1 Análise dos dados para o modelo de hélice quádrupla

A partir dos dados gerados pelo ScriptGP2 foi possível identificar as relações entre os diferentes grupos de pesquisa, interligados pelos seus pesquisadores e instituições. A rede formada pelas ligações entre grupos de pesquisa, pesquisadores e instituições é composta por 6029 nós e 12173 ligações. A análise da rede formada por estas relações demonstrou existir um grupo central com muitas conexões interinstitucionais e diversas “ilhas” isoladas, conforme ilustrado na Figura 6. O grupo central apresenta relações interinstitucionais entre os grupos, ou seja, pesquisadores participam em grupos de pesquisa de mais de uma instituição. Já os agrupamentos

intrainstitucionais são formados por grupos de pesquisa que não possuem nenhum membro externo à instituição ou que não possuem nenhum membro interno participando de um grupo de pesquisa de outra instituição. Também foram identificadas 316 relações entre instituições, sendo 150 destas relações do tipo público-privado, evidenciando as relações previstas no modelo de hélice quádrupla.

Figura 6. Redes entre pesquisadores, grupos de pesquisa e instituições



Fonte: Dados da pesquisa.

5 Considerações Finais

A contribuição esperada desta pesquisa é a geração de novos conhecimentos sobre os processos e artefatos tecnológicos para a recuperação e análise de informações em bases de dados públicas, úteis ao estudo da inovação por meio do modelo de hélice quádrupla. Espera-se que o *framework* proposto seja útil e possa ser aplicado para a prospecção tecnológica e científica, inovação e apoio a formação de políticas públicas. A pesquisa também reafirma que o uso de ferramentas *open source* contribuem para a *Open Science* por meio do acesso democrático e livre de importantes informações para o desenvolvimento científico e tecnológico, além de permitirem o desenvolvimento incremental das soluções já existentes.

De acordo com o princípio de *Open Science*, todos os dados e conteúdo desta pesquisa estarão

disponíveis on-line ou poderão ser solicitados aos autores. A não inclusão de um hiperlink diretamente neste artigo visa atender aos requisitos de não identificação de autoria.

6 Implicações e Pesquisas Futuras

Ao expandir as possibilidades de acesso e análise a fontes de dados secundários, espera-se novas contribuições, instrumentais e metodológicas na coleta, armazenamento, processamento, análise e interpretação dos dados. A crescente adoção de políticas públicas de transparência de dados também contribui para o aumento de fontes secundárias disponíveis.

Pesquisas futuras podem se valer das novas possibilidades de dados para confirmar teorias existentes, testar novas teorias ou desenvolver aplicações práticas de análises preditivas.

A adoção da política de Ciência Aberta também deve contribuir para a transparência da pesquisa científica pública, tornando-a integralmente disponível, dos dados até os resultados.

Referências

- Abelson, J., Forest, P. G., Eyles, J., Smith, P., Martin, E., & Gauvin, F. P. (2003). Deliberations about deliberative methods: issues in the design and evaluation of public participation processes. *Social Science & Medicine*, 57(2), 239–251.
- Afuah, A., & Tucci, C. L. (2012). Crowdsourcing as a solution to distant search. *Academy of Management Review*, 37(3), 355–375.
- Arnott, D., & Pervan, G. (2012). Design science in decision support systems research: An assessment using the hevner, march, park, and ram guidelines. *Journal of the Association of Information Systems*, 13(11), 923–949.
- Bandara, H. M. D., Mantell, M. A., Darcy, J. W., & Emmert, M. H. (2015). Rare Earth Recycling: Forecast of Recoverable Nd from Shredder Scrap and Influence of Recycling Rates on Price Volatility. *Journal of Sustainable Metallurgy*, 1(3), 179–188. <https://doi.org/10.1007/s40831-015-0019-3>
- Barak, M., Watted, A., & Haick, H. (2016).

- Motivation to learn in massive open online courses: Examining aspects of language and social engagement. *Computers & Education*, 94, 49–60.
- Bastian, M., Heymann, S., Jacomy, M., & others. (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 8, 361–362.
- Belkin, N. J. (1993). Interaction with texts: Information retrieval as information seeking behavior. *Information retrieval*, 93, 55–66.
- Bifulco, F., Tregua, M., Amitrano, C. C., & D’Auria, A. (2016). ICT and sustainability in smart cities management. *International Journal of Public Sector Management*, 29(2), 132–147. <https://doi.org/10.1108/IJPSM-07-2015-0132>
- BRASIL, MCTI. (2013). Plataforma Lattes. Recuperado 2 de agosto de 2013, de <http://lattes.cnpq.br/>
- Cândido, A. C., Wielewick, P. F., & Zimmermann, R. A. (2016). Barriers and Drivers for Innovation in Times of Crisis. *Revista de Negócios*, 20(4), 01-10.
- Carayannis, E. G., Barth, T. D., & Campbell, D. F. (2012). The Quintuple Helix innovation model: global warming as a challenge and driver for innovation. *Journal of Innovation and Entrepreneurship*, 1(1), 1–12.
- Carayannis, E. G., & Campbell, D. F. J. (2009). “Mode 3” and “Quadruple Helix”: toward a 21st century fractal innovation ecosystem. *International Journal of Technology Management*, 46(3/4), 201. <https://doi.org/10.1504/IJTM.2009.023374>
- Carayannis, E. G., & Campbell, D. F. J. (2010). Triple Helix, Quadruple Helix and Quintuple Helix and How Do Knowledge, Innovation and the Environment Relate To Each Other?: A Proposed Framework for a Trans-disciplinary Analysis of Sustainable Development and Social Ecology. *International Journal of Social Ecology and Sustainable Development*, 1(1), 41–69. <https://doi.org/10.4018/jesd.2010010105>
- Carayannis, E. G., & Campbell, D. F. J. (2012). *Mode 3 Knowledge Production in Quadruple Helix Innovation Systems*. New York, NY: Springer New York.
- Carayannis, E. G., & Rakhmatullin, R. (2014). The Quadruple/Quintuple Innovation Helixes and Smart Specialisation Strategies for Sustainable and Inclusive Growth in Europe and Beyond. *Journal of the Knowledge Economy*, 5(2), 212–239. <https://doi.org/10.1007/s13132-014-0185-8>
- Casaramona, A., Sapia, A., & Soraci, A. (2015). How TOI and the Quadruple and Quintuple Helix Innovation System Can Support the Development of a New Model of International Cooperation. *Journal of the Knowledge Economy*, 1–17. <https://doi.org/10.1007/s13132-015-0253-8>
- Chung, C. J. (2014). An analysis of the status of the Triple Helix and university–industry–government relationships in Asia. *Scientometrics*, 99(1), 139–149.
- Dresch, A., Lacerda, D. P., & Antunes_Júnior, J. A. V. (2015). *Design Science Research: Método de Pesquisa para Avanço da Ciência e Tecnologia*. Porto Alegre: Bookman Editora.
- Dutta, T., Kim, K.-H., Uchimiya, M., Kwon, E. E., Jeon, B.-H., Deep, A., & Yun, S.-T. (2016). Global demand for rare earth resources and strategies for green mining. *Environmental Research*, 150, 182–190. <https://doi.org/10.1016/j.envres.2016.05.052>
- Etzkowitz, H. (2003). Innovation in Innovation: The Triple Helix of University-Industry-Government Relations. *Social Science Information*, 42(3), 293–337. <https://doi.org/10.1177/05390184030423002>
- Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and “Mode 2” to a Triple Helix of university–industry–government relations. *Research policy*, 29(2), 109–123.

- Faria, A., & Sauerbronn, F. F. (2008). A responsabilidade social é uma questão de estratégia? Uma abordagem crítica. *Revista de Administração Pública*, 42(1), 7–34.
- Ferraz, R. R. N., Quoniam, L., Reymond, D., & Maccari, E. A. (2016). Example of open-source OPS (Open Patent Services) for patent education and information using the computational tool Patent2Net. *World Patent Information*, 46, 21–31.
- Fracassi, C. (2017). Corporate Finance Policies and Social Networks. *Management Science*, 63(8), 2420–2438. <https://doi.org/10.1287/mnsc.2016.2433>
- Freeman, L. C. (2004). *The development of social network analysis: a study in the sociology of science*. Vancouver, BC : North Charleston, S.C: Empirical Press ; BookSurge.
- George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big Data and Data Science Methods for Management Research. *Academy of Management Journal*, 59(5), 1493–1507. <https://doi.org/10.5465/amj.2016.4005>
- Gibbons, M., & Nowotny, H. (2001). The potential of transdisciplinarity. In *Transdisciplinarity: joint problem solving among science, technology, and society* (p. 67–80). Springer.
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351. <https://doi.org/10.1038/nrg.2016.49>
- Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS quarterly*, 37(2), 337–355.
- Griggs, D., Stafford-Smith, M., Gaffney, O., Rockström, J., Öhman, M. C., Shyamsundar, P., ... Noble, I. (2013). Policy: Sustainable development goals for people and planet. *Nature*, 495(7441), 305–307.
- Groß, M., & Stauffacher, M. (2014). Transdisciplinary Environmental Science: Problem-oriented Projects and Strategic Research Programs. *Interdisciplinary Science Reviews*, 39(4), 299–306. <https://doi.org/10.1179/0308018814Z.00000000093>
- Hevner, A., & Chatterjee, S. (2010). *Design science research in information systems*. Springer. Recuperado de http://link.springer.com/chapter/10.1007/978-1-4419-5653-8_2
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly: Management Information Systems*, 28(1), 75–105.
- Huutoniemi, K., Klein, J. T., Bruun, H., & Hukkinen, J. (2010). Analyzing interdisciplinarity: Typology and indicators. *Research Policy*, 39(1), 79–88. <https://doi.org/10.1016/j.respol.2009.09.011>
- Ivanova, I. A., & Leydesdorff, L. (2014). Rotational symmetry and the transformation of innovation systems in a Triple Helix of university-industry-government relations. *Technological Forecasting and Social Change*, 86, 143–156. <https://doi.org/10.1016/j.techfore.2013.08.022>
- Jeong, Y., & Yoon, B. (2014). Development of patent roadmap based on technology roadmap by analyzing patterns of patent development. *Technovation*. <https://doi.org/10.1016/j.technovation.2014.03.001>
- Jürgens, B., & Herrero-Solana, V. (2015). Espacenet, Patentscope and Depatisnet: A comparison approach. *World Patent Information*, 42, 4–12. <https://doi.org/10.1016/j.wpi.2015.05.004>
- Kim, J., & Lee, S. (2015). Patent databases for innovation studies: A comparative analysis of USPTO, EPO, JPO and KIPO. *Technological Forecasting and Social Change*, 92, 332–345.

- <https://doi.org/10.1016/j.techfore.2015.01.009>
- Kim, Y., Choi, T. Y., Yan, T., & Dooley, K. (2011). Structural investigation of supply networks: A social network analysis approach. *Journal of Operations Management*, 29(3), 194–211.
- Kraker, P., Leony, D., Reinhardt, W., & Beham, G. (2011). The case for an open science in technology enhanced learning. *International Journal of Technology Enhanced Learning*, 3(6), 643–654.
- Kuchma, I., & others. (2014). Policy Framework and Roadmap for Open Access, Open Research Data and Open Science. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, (IV), 333–352.
- Lacerda, D. P., Dresch, A., Proença, A., & Antunes_Júnior, J. A. V. (2013). Design Science Research: A research method to production engineering. *Gestao e Producao*, 20(4), 741–761. <https://doi.org/10.1590/S0104-530X2013005000014>
- Leydesdorff, L. (2012). The triple helix, quadruple helix,..., and an N-tuple of helices: Explanatory models for analyzing the knowledge-based economy? *Journal of the Knowledge Economy*, 3(1), 25–35.
- Leydesdorff, L., & Etzkowitz, H. (1996). Emergence of a Triple Helix of university—industry—government relations. *Science and public policy*, 23(5), 279–286.
- Magalhães, J. L., Quoniam, L., Mena-Chalco, J. P., & Santos, A. (2014). Extração e tratamento de dados na base lattes para identificação de core competencies em dengue. *Informação & Informação*, 19(3), 30–54. <https://doi.org/10.5433/1981-8920.2014v19n3p30>
- March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision support systems*, 15(4), 251–266.
- Marteleto, R. M. (2001). Análise de redes sociais: aplicação nos estudos de transferência da informação. *Ciência da informação*, 30(1), 71–81.
- Massari, S., & Ruberti, M. (2013). Rare earth elements as critical raw materials: Focus on international markets and future strategies. *Resources Policy*, 38(1), 36–43.
- Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of business venturing*, 29(1), 1–16.
- Nelson, R. R. (1993). National innovation systems: a comparative analysis. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*.
- OECD. (1997). *National innovation systems*. Paris: Organisation for Economic Co-operation and Development.
- OECD. (2016). Open science - OECD. Recuperado 15 de julho de 2016, de <https://www.oecd.org/sti/outlook/e-outlook/stipolicyprofiles/interactionsforinnovation/openscience.htm>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3), 45–77.
- Podestá, G. P., Natenzon, C. E., Hidalgo, C., & Toranzo, F. R. (2013). Interdisciplinary production of knowledge with participation of stakeholders: A case study of a collaborative project on climate variability, human decisions and agricultural ecosystems in the Argentine Pampas. *Environmental Science & Policy*, 26, 40–48. <https://doi.org/10.1016/j.envsci.2012.07.008>
- Ramos, R. R., & Machado, C. J. S. (2014). Uma análise espaço-temporal dos grupos de pesquisa do CNPQ: a dengue no Brasil. *Hygeia*, 10(18), 58–70.
- Rosenfield, P. L. (1992). The potential of

- transdisciplinary research for sustaining and extending linkages between the health and social sciences. *Social science & medicine*, 35(11), 1343–1357.
- Sábato, J. A., & Botana, N. R. (1968). La ciencia y la tecnología en el desarrollo futuro de América Latina. Apresentado em The World Order Models Conferenc, Bellagio, Italy, September 25-30.
- Santos, A. M. dos, Kono, C. M., & Quoniam, L. (2014). Análise dos Grupos de Pesquisa em Patentes no Brasil: um estudo a partir do Diretório de Grupos de Pesquisa do CNPQ. Apresentado em XXXVIII Encontro da ANPAD, Rio de Janeiro, RJ, Brasil.
- Simon, H. A. (1996). *The sciences of the artificial*. Cambridge, Mass. u.a.: MIT Press.
- Stokols, D., Hall, K. L., Taylor, B. K., & Moser, R. P. (2008). The Science of Team Science. *American Journal of Preventive Medicine*, 35(2), S77–S89. <https://doi.org/10.1016/j.amepre.2008.05.002>
- Tekic, Z., Drazic, M., Kukulj, D., & Vitas, M. (2014). From Patent Data to Business Intelligence – PSALM Case Studies. *24th DAAAM International Symposium on Intelligent Manufacturing and Automation, 2013*, 69(0), 296–303. <https://doi.org/10.1016/j.proeng.2014.02.235>
- Van Aken, J. E. (2005). Management research as a design science: articulating the research products of mode 2 knowledge production in management. *British journal of management*, 16(1), 19–36.
- Venable, J., & Baskerville, R. (2012). Eating our own cooking: Toward a more rigorous design science of research methods. *Electronic Journal of Business Research Methods*, 10(2), 141–153.
- Villarreal, O., & Calvo, N. (2015). From the Triple Helix model to the Global Open Innovation model: A case study based on international cooperation for innovation in Dominican Republic. *Journal of Engineering and Technology Management - JET-M*, 35, 71–92. <https://doi.org/10.1016/j.jengtecman.2014.10.002>
- Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., ... Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, 5(1), 14–26.
- Wasserman, S. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.
- Zancan, C., dos Santos, P. da C. F., & Campos, V. O. (2012). As Contribuições Teóricas da Análise de Redes Sociais (ARS) aos Estudos Organizacionais. *Revista Alcance*, 19(1), 62–82.