

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA, TECNOLOGIA E  
SOCIEDADE

ALINE GRASIELE CARDOSO DE BRITO

**SISTEMÁTICA PARA MAPEAR ÁREAS DO CONHECIMENTO DA  
CIÊNCIA BRASILEIRA: O CASO DA NANOTECNOLOGIA**

São Carlos  
2016

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA, TECNOLOGIA E  
SOCIEDADE

ALINE GRASIELE CARDOSO DE BRITO

**SISTEMÁTICA PARA MAPEAR ÁREAS DO CONHECIMENTO DA  
CIÊNCIA BRASILEIRA: O CASO DA NANOTECNOLOGIA**

Tese apresentada ao Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade do Centro de Educação e Ciências Humanas da Universidade Federal de São Carlos, para obtenção do título de Doutora em Ciência, Tecnologia e Sociedade.

Orientação: Prof. Dr. Luc Quoniam

Linha de pesquisa: Gestão Tecnológica e Sociedade Sustentável

São Carlos  
2016

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar  
Processamento Técnico  
com os dados fornecidos pelo(a) autor(a)

B862s Brito, Aline Grasielle Cardoso de  
Sistemática para mapear áreas do conhecimento da  
ciência brasileira : o caso da nanotecnologia /  
Aline Grasielle Cardoso de Brito. -- São Carlos :  
UFSCar, 2017.  
178 p.

Tese (Doutorado) -- Universidade Federal de São  
Carlos, 2016.

1. Indicadores em Ciência e Tecnologia.  
2. Mineração de dados. 3. Acesso aberto. 4.  
Currículos Lattes. 5. Nanotecnologia. I. Título.



---

Folha de Aprovação

---

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de tese de doutorado do(a) candidato(a) **Aline Grasielle Cardoso de Brito** realizada em 07/12/2016

Prof. Dr. Leandro I. Lopes de Faria - PPGCTS/UFSCar

---

Prof. Dr. Luc Marie Quoniam - PPGCTS/UFSCar

---

Profa. Dra. Lillian M. Araújo de R. Alvares - UnB

---

Prof. Dr. Jésus Paschual Mena Chaco - UFABC

Prof. Dr. Daniel Rodrigo Leiva - UFSCar

Certifico que a sessão de defesa foi realizada com a participação à distância dos membros **Prof. Dr. Luc Marie Quoniam, Prof. Dr. Jésus Paschual Mena Chaco e Profa. Dra. Lillian M. Araújo de R. Alvares** e, depois das arguições e deliberações realizadas, o participante à distância está de acordo com o conteúdo do parecer da comissão examinadora redigido no relatório de defesa do(a) aluno(a) **ALINE GRASIELE CARDOSO DE BRITO**.

Prof. Dr. Leandro I. Lopes de Faria  
Presidente da Comissão Examinadora  
(PPGCTS/UFSCar)

## **DEDICATÓRIA**

*Aos meus avós, Deusmira (in memorian) e Gervázio (in memorian), Maura (in memorian) e Vitor.*

## **AGRADECIMENTOS**

Primeiramente e acima de tudo, agradeço a Deus, pela minha vida e por todas as oportunidades de aperfeiçoamento de meu espírito.

Agradeço ao orientador, Luc Quoniam, pela paciência, valioso auxílio e dedicação.

Aos professores membros da banca, pelas preciosas contribuições.

Aos amigos do NIT e do PPGCTS, pelos momentos de amizade, aprendizados e alegrias – especialmente à Vera, Roni, Leandro, Daniel, Lucas, Bráulio, Douglas, Danilo, Nayara, Isadora...

À minha mãe, Beatriz e às minhas irmãs Nathalie, Paloma e Pérola, por serem as bases de amor e motivação. Ao meu noivo, Anderson, pela compreensão, amor e carinho. À amiga-irmã, Jacqueline, força do dia a dia.

Aos meus professores.

Ao NIT/Materiais.

À UFSCar.

E à CAPES pela bolsa concedida.

Muito obrigada!

## Resumo

Nos últimos anos os estudos sociais da ciência têm revelado um especial interesse na extração de conhecimentos em bases de dados, o que possibilita o tratamento e análise de conjuntos de dados disponíveis sobre a produção científica e tecnológica de pesquisadores, instituições e países. O objetivo desta pesquisa foi desenvolver e aplicar uma sistemática para coleta, tratamento e análise de informações sobre áreas do conhecimento científico no Brasil, utilizando essencialmente dados de acesso aberto e ferramentas livres. O método foi compreendido por um estudo de caso, a unidade caso foi a área prioritária nanotecnologia e a fonte de informação compreendeu as bases: Base de Currículos Lattes; Diretório de Grupos de Pesquisa; Biblioteca Digital Brasileira de Teses e Dissertações; e a Espacenet do Escritório Europeu de Patentes. Os resultados alcançados foram 1) a criação e apresentação de uma sistemática para coleta, tratamento e análise de informações 2) indicadores para a área tratada. Conclui-se que as Informações em Ciência e Tecnologia, com acesso aberto, permitem uma democrática geração de conhecimentos para este contexto. As fontes de informação e o conjunto de ferramentas levantadas possibilitaram a elaboração de indicadores capazes de elucidar múltiplas dimensões da produção científica e tecnológica – de pesquisadores e grupos de pesquisa – além do ensino e pesquisa na pós-graduação. Os indicadores obtidos e sua aplicação reforçam o objetivo do processo de data mining, no sentido de transformar dados em informação a ser empregada no processo decisório das organizações que se dedicam à preservação e a inovação do conhecimento.

**Palavras-chave:** Indicadores em Ciência e Tecnologia. Mineração de dados. Acesso Aberto. Currículos Lattes. Grupos de Pesquisa. Espacenet. BDTD. Nanotecnologia.

## Abstract

In recent years the social studies of science have revealed a special interest in the extraction of knowledge in databases, which enables the treatment and analysis of sets of data available on the scientific and technological production of researchers, institutions and countries. The objective of this research was to develop and apply a systematics for the collection, treatment and analysis of information about areas of scientific knowledge in Brazil, using essentially open access data and free tools. The method was a case study, the unit case was the priority area nanotechnology and the source of information comprised the bases: Lattes Curriculum Base; Directory of Search Groups; Brazilian Digital Library of Theses and Dissertations; And the Espacenet of the European Patent Office. The results achieved were 1) the creation and presentation of a systematic for the collection, treatment and analysis of information 2) indicators for the treated area. It is concluded that the Information in Science and Technology, with open access, allow a democratic generation of knowledge in this context. The sources of information and the set of tools made possible the elaboration of indicators capable of elucidating multiple dimensions of scientific and technological production - of researchers and research groups - in addition to teaching and research in postgraduate studies. The indicators obtained and their application reinforce the objective of the data mining process, in the sense of transforming data into information to be used in the decision-making process of organizations that are dedicated to the preservation and innovation of knowledge.

**Keywords:** Indicators of Science and Technology. Data mining. Open Access. Lattes resumes. Research Groups. Espacenet. BDTD. Nanotechnology.

## Lista de figuras

Figura 1 – Esquema da avaliação em ciência.....	33
Figura 2 – Composição dos grupos de pesquisa em nanotecnologia .....	53
Figura 3 – Variáveis controladas e com ruído de um sistema .....	80
Figura 4 – Procedimentos para coleta de ID .....	92
Figura 5 – Visualização dos resultados do SciplLattes.....	97
Figura 6 – Pesquisadores do core em nanotecnologia e seus Grupos de Pesquisa.....	99
Figura 7 – Resumo das principais informações de uma busca P2N.....	102
Figura 8 – Análises on-line de ferramentas do P2N.....	102
Figura 9 – Dados para tratamentos em outras ferramentas .....	103
Figura 10 – Representação dos procedimentos realizados .....	107
Figura 11 – Rede da relação entre os termos da expressão de busca .....	114
Figura 12 – Rede geral em nanotecnologia .....	116
Figura 13 – Visualização do pesquisador José Arana Varela.....	117
Figura 14 – Visualização do pesquisador Elson Longo da Silva .....	118
Figura 15 – Informações dos nós selecionados.....	119
Figura 16 – Infográfico com o perfil do pesquisador em nanotecnologia.....	120
Figura 17 – Classes dos vocabulários em nanotecnologia .....	132
Figura 18 – Visualização dos vocabulários em classes .....	133
Figura 19 – Nuvem de palavras dos resumos de teses em nanotecnologia.....	134
Figura 20 – Demonstrativo de uma rede de palavras-chave de teses da UFSCar .....	135
Figura 21 – Mapa dos países depositantes, com proteção e do inventor.....	140
Figura 22 – Conexões da tecnologia H01M4.....	141
Figura 23 – Mapa conceitual por classificação das patentes analisadas .....	142
Figura 24 – Detalhamento das áreas E e F.....	143
Figura 25 – Nuvem de palavras com os abstracts das patentes H01G11/36 .....	144

## Lista de gráficos

Gráfico 1 – Crescimento do patenteamento em nanotecnologia entre 2000 e 2014 .....	53
Gráfico 2 - Investimentos em nanotecnologia no Brasil de 2004 a 2014.....	54
Gráfico 3 – Comportamento com o acréscimo de termos .....	88
Gráfico 4 – Termos mais frequentes da expressão de busca nos artigos.....	110
Gráfico 5 – Áreas ligadas a nanotecnologia.....	111
Gráfico 6 – Distribuição dos pesquisadores em nanotecnologia por estados.....	112
Gráfico 7 – Distribuição dos pesquisadores em nanotecnologia por regiões .....	113
Gráfico 8 – Número das teses e dissertações em nanotecnologia .....	130
Gráfico 9 – Estatística das frequências das palavras dos resumos em nanotecnologia .....	131
Gráfico 10 – Evolução das patentes relacionadas a tecnologia tratada (H01G11/36).....	137

## Lista de quadros

Quadro 1 – Indicadores de publicação .....	41
Quadro 2 – Situações relevantes para diferentes estratégias de pesquisa. ....	77
Quadro 3 – Esquema da condução do estudo de caso .....	82
Quadro 4 – Comparação de funcionalidades WoS versus Currículo Lattes .....	84
Quadro 5 – Operadores reconhecidos pela Plataforma Lattes.....	86
Quadro 6 – Levantamento dos termos da área .....	87
Quadro 7 – Expressão de busca.....	90
Quadro 8 – Estrutura lista final. ....	91
Quadro 9 – Critérios para busca na PL .....	92
Quadro 10 – Saída do ScriptKeyword.....	109
Quadro 11 – Evolução dos grupos de pesquisa dos pesquisadores em nanotecnologia.....	124
Quadro 12 – Distribuição do core em nano por estado .....	125
Quadro 13 – Instituições do “core nano” em São Paulo .....	127
Quadro 14 – Instituições do “core nano” em Minas Gerais .....	128
Quadro 15 – Nível máximo de detalhamento dos grupos de pesquisa em nanotecnologia.....	129
Quadro 16 – Países onde as patentes estão protegidas .....	138
Quadro 17 – Patentes sobre H01G11/36 protegidas no Brasil .....	138

## **Lista de tabelas**

Tabela 1 – Artigos publicados encontrados nos Currículos Lattes .....	64
Tabela 2 – Censo 2014 da distribuição dos grupos de pesquisa.....	65

## Lista de abreviaturas e siglas

ABDI	Agência Brasileira de Desenvolvimento Industrial
BCL	Base de Currículos Lattes
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
C&T	Ciência e Tecnologia
CAGR	Compound Annual Growth Rate
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CGEE	Centro de Gestão e Estudos Estratégicos
CHD	Classificação Hierárquica Descendente
CIP	Classificação Internacional de Patentes
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CTS	Ciência, Tecnologia e Sociedade
DGP	Diretório de Grupo de Pesquisa
EPO	Escritório Europeu de Patentes
ESCT	Estudos Sociais da Ciência e da Tecnologia
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FINEP	Financiadora de Estudos e Projetos
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
IBN	Iniciativa Brasileira de Nanotecnologia
INCT	Institutos Nacionais de Ciência e Tecnologia
INPI	Instituto Nacional de Propriedade Intelectual
IPC	International Patent Classification
ISI	Institute for Scientific Information
JCR	Journal Citation Reports
LAI	Lei de Acesso à Informação
MCTI	Ministério da Ciência, Tecnologia e Inovação
NIT	Núcleo de Informação Tecnológica
OAI	Open Archives Initiative
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
OMPI	Organização Mundial da Propriedade Intelectual
P&D	Pesquisa e Desenvolvimento
PCT	Política Científica e Tecnológica
PIB	Produto Interno Bruto
PL	Plataforma Lattes
SciELO	Scientific Electronic Library Online
SRI	Sistema de Recuperação da Informação
UNESCO	Organização das Nações Unidas para a Educação e Cultura
WoS	Web of Science

## PUBLICAÇÕES

### ARTIGOS EM PERIÓDICOS

AMARAL, R. M.; BRITO, A. G. C.; ROCHA, K. G. S.; QUONIAM, L.; FARIA, L. I. L. Panorama da inteligência competitiva no Brasil: os pesquisadores e a produção científica na plataforma Lattes. **Perspectivas em Ciência da Informação**, v.21, n.4, p.3-17, out./dez. 2016.

BRITO, A. G. C.; QUONIAM, L.; MENA-CHALCO, J. P. Exploração da Plataforma Lattes por assunto: proposta de metodologia. **Transinformação** [online]. 2016, vol.28, n.1, pp.77-86.

### ARTIGOS COMPLETOS EM EVENTOS

BRITO, A. G. C.; AMARAL, R. M.; QUONIAM, L. A plataforma lattes como fonte de informação. *In: III Encontro Internacional – Dados, Tecnologia e Informação (III DTI)*, 2016. Marília:- SP. Anais... Marília: UNESP, 2016.

BRITO, A. G. C.; AMARAL, R. M.; FARIA, L. I. L. QUONIAM, L.; VIEIRA, C. Visibilidade científica na Plataforma Lattes e Portal da Inovação. *In: ENCONTRO NACIONAL EM CIÊNCIA DA INFORMAÇÃO E BIBLIOTECONOMIA – ENANCIB*, 17, 2016, Salvador – BA. *Anais...* Salvador: UFBA, 2016.

AMARAL, R. M.; ROCHA, K. G. S.; BRITO, A. G. C.; QUONIAM, L.; FARIA; L. I. L. Análise da produção científica brasileira em inteligência competitiva. *In: ENCONTRO NACIONAL EM CIÊNCIA DA INFORMAÇÃO E BIBLIOTECONOMIA – ENANCIB*, 16, 2015, João Pessoa – PB. *Anais...* João Pessoa: UFPB, 2015.

BRITO, A. G. C.; QUONIAM, L. Análise das publicações sobre a Plataforma Lattes. *In: ENCONTRO NACIONAL EM CIÊNCIA DA INFORMAÇÃO E BIBLIOTECONOMIA – ENANCIB*, 15, 2014, Belo Horizonte – MG. *Anais...* Belo Horizonte: ECI/UFMG, 2014. p. 3080-3709.

BRITO, A. G. C.; MILANEZ, D. H.; QUONIAM, L. Análise das green patents em nanotecnologias ligadas à área de biocombustíveis. *In: ENCONTRO DE CIÊNCIA E ENGENHARIA DE MATERIAIS DE SÃO CARLOS: ENERGIA, SUSTENTABILIDADE E INOVAÇÃO - ECEM*, 1, 2013, São Carlos - SP. *Anais...* São Carlos: ECEM/SanCas 2013.

### RESUMOS EM EVENTOS

BRITO, A. G. C.; QUONIAM, L. Uso da plataforma lattes por assunto: recuperação automática de identificadores dos currículos Lattes. *In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA PARA O PROGRESSO DA CIÊNCIA – SBPC*, 67, 2015, São Carlos – SP. *Anais...* São Carlos, 2015.

## Sumário

<b>1 INTRODUÇÃO .....</b>	<b>17</b>
<b>2 ESTUDOS SOCIAIS DA CIÊNCIA E ACESSO ABERTO EM INFORMAÇÃO, CIENTÍFICA E TECNOLÓGICA .....</b>	<b>25</b>
2.1 Sociologia da ciência .....	27
2.2 Avaliação da ciência .....	32
2.3 Política científica e tecnológica.....	44
2.3.1 A nanotecnologia, uma área prioritária? .....	49
2.3.1.1 A Nanotecnologia no cenário mundial.....	50
2.3.1.2 A nanotecnologia no Brasil.....	51
2.4 Acesso aberto à Informação, Científica e Tecnológica.....	54
2.4.1 Plataforma Lattes.....	59
2.4.1.1 A Base de Currículo Lattes.....	61
2.4.1.2 O Diretório de Grupos de Pesquisa .....	65
2.4.2 Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) .....	66
2.4.3 As patentes e a Espacenet.....	66
2.4.4 Acesso à Informação no Brasil.....	69
<b>3 MÉTODO E DESENVOLVIMENTO DA PESQUISA .....</b>	<b>75</b>
3.1 O delineamento da pesquisa.....	76
3.2 Procedimentos específicos à Base de Currículos Lattes .....	83
3.2.1 Exploração das especificidades .....	83
3.2.2 Elaboração da expressão de busca .....	85
3.2.3 Extração dos ID dos Currículos Lattes.....	90
3.2.4 Criação dos filtros .....	92
3.3 Desenvolvimento e aplicação da sistemática para mapeamento de áreas prioritárias	94
3.3.1 A coleta dos dados e informações e as ferramentas utilizadas.....	95
3.3.1.1 O ScriptLattes .....	96
3.3.1.2 O ScriptGP .....	98
3.3.1.3 A extração BDTD .....	99
3.3.1.4 O Patent2Net.....	100

<b>4 RESULTADOS E DISCUSSÃO.....</b>	<b>105</b>
4.1 Sistemática para mapeamento de áreas prioritárias .....	105
4.2 Mapeamento e análise de uma área prioritária: a nanotecnologia .....	108
4.2.1 Indicadores da Base de Currículos Lattes .....	109
4.2.2 Indicadores Diretório de Grupos de Pesquisa.....	123
4.2.3 Indicadores BDTD .....	130
4.2.2 Indicadores Espacenet.....	136
<b>5 CONCLUSÃO.....</b>	<b>148</b>
<b>Referências.....</b>	<b>153</b>
APÊNDICE A – Esquema para a construção da expressão de busca .....	165
ANEXO 1 – Petição para a retirada dos captchas dos currículos da Plataforma Lattes .....	169
ANEXO 2 – Interface da busca avançada e recuperação dos Currículos Lattes .....	172
ANEXO 3 – Apresentação do MCTI sobre nanotecnologia.....	173
ANEXO 4 – Arquivo de configuração para o ScriptLattes .....	174
ANEXO 5 – Arquivo de configuração para o Patent2Net .....	177

# 1 INTRODUÇÃO

A ciência é definida amplamente como um conhecimento sistematizado do mundo natural, que permite a humanidade a esperança de uma vida melhor, graças aos avanços científicos conquistados em todos os períodos históricos. Pode-se afirmar que a ciência, de algum modo, tem ajudado a salvar milhões de vidas de situações como as doenças, a fome e a pobreza.

A descoberta da penicilina, o desenvolvimento de sementes de alta produtividade e a distribuição de eletricidade são três exemplos na qual a ciência contribuiu para o bem-estar social no século XX. Em muitos países, tais avanços tiveram impacto ainda maior, estimulando o crescimento da economia e reforçando a criação de uma extensa classe média, que para muitos teóricos, é uma condição fundamental para a democracia. No entanto, outros países não obtiveram benefícios semelhantes, onde se verificam imensas desigualdades no que tange a distribuição e geração de recursos.

Desde o nascimento da ciência moderna no século XIV, os ganhos da aplicação do conhecimento científico têm sido distribuídos de forma desigual, contribuindo para o fosso cada vez maior entre os países desenvolvidos e os em desenvolvimento. Apesar da difusão acelerada dos dados, informações e conhecimento, o mundo da ciência está longe de ser plano (WAGNER, 2008).

O compartilhamento dos resultados das pesquisas entre os pesquisadores se tornou essencial ao desenvolvimento Científico e Tecnológico (C&T) dos países. Nesse contexto, a atividade científica deveria se tornar mais e mais sociabilizada, uma vez que, o cientista isolado dá lugar ao pesquisador inserido em grupos e conseqüentemente na comunidade científica, entrando em contato permanente com seus pares, que dele exige competitividade e produtividade (TARGINO; NEYRA, 2006).

Dessa forma, as redes auto organizadas que abrangem todo o globo são a característica mais notável da ciência hoje. Essas redes constituem os colégios invisíveis de pesquisadores, que não colaboram entre si porque lhes é dito, mas sim porque querem trabalhar juntos. Não porque eles compartilham um laboratório ou mesmo uma disciplina, mas porque eles podem oferecer uns aos outros *insights*, conhecimentos ou habilidades complementares (WAGNER, 2008). Como em um ciclo, a tendência histórica, bem como as funções e os processos dominantes na era da informação estão cada vez mais organizados em torno de redes. “Embora a forma de organização social em redes tenha existido em outros tempos e espaços,

o novo paradigma da tecnologia da informação fornece a base material para sua expansão penetrante em toda a estrutura social” (CASTELLS, 2010).

O exponencial crescimento de fontes de informação na internet apresenta a necessidade da utilização de diversas ferramentas que facilitem as buscas, recuperação e análises dos dados disponíveis na rede de computadores. E como resposta ao acúmulo do volume de informações, vários mecanismos de armazenamento, recuperação, processamento, análise e visualização têm sido implementados nos últimos anos (BOUTET; QUONIAM, 2012; BRIN; PAGE, 1998).

Não obstante, os sistemas de informação, atualmente quase sempre formados por informações em ambiente *web*, nem sempre são facilmente visualizados, compreendidos e recuperados, necessitando de análises e ferramentas específicas para criação de novos conhecimentos (HICKS, 2015). Assim as ferramentas de análise e mensuração dos dados registrados e as técnicas de construção de indicadores, principalmente utilizando bases de dados automatizadas, facilitam o gerenciamento e a visualização de passado, presente e futuro das agendas de pesquisas e do comportamento da ciência.

Diversos autores de estudos sociais da ciência, (BOURDIEU, 1975; EPSTEIN, 1990; HOCHMAN, 1994; MERTON, 1970) em especial sobre a sua avaliação (BAUMGARTEN; LIMA, 2013; EPSTEIN, 1990; MERTON, 1970; ZIMAN, 2000), como por exemplo, nos estudos bibliométricos de Faria *et. al.* (2011) Gregolin *et. al.* (2005) têm utilizado bases de dados internacionais para a elaboração de indicadores. Regalado (2010) apresenta dados, compilados em conjunto pelo MCTI e *ISI/Thomson Reuters*, indicando que o número de publicações de autores brasileiros em periódicos catalogados na *Web of Science (WoS)* foi de 6.038, em 1997, para 32.100, em 2007. Tal aumento foi acompanhado também por um aumento na fração que as publicações brasileiras representam nessa base: de 0,8% para 2,7% no mesmo período. A fim de comparação, um levantamento realizado na Plataforma Lattes sobre o corpo de conhecimento de uma área apenas, a nanotecnologia (2.520 pesquisadores), identificou 96.840 artigos completos publicados em periódicos<sup>1</sup> (VLAB4U, 2015).

Diante disso, a presente pesquisa parte da premissa de que para uma visão holística da ciência brasileira, a base de dados da *WoS* não seja a fonte de informação mais indicada, pois sua composição reflete uma seleta parte da produção científica mundial e por conseguinte,

---

<sup>1</sup> 1.358 livros publicados/organizados ou edições; 5.006 capítulos de livros publicados; 5.218 textos em jornais de notícias/revistas; 51.033 trabalhos completos publicados em anais de congressos; 15.615 resumos expandidos publicados em anais de congressos; 126.248 resumos publicados em anais de congressos; 2.101 artigos aceitos para publicação; 39.000 apresentações de trabalho e; 3.203 demais tipos de produção bibliográfica. Somando o total de 345.622 registros de produção bibliográfica.

apenas uma parte disso é o que se refere a produção brasileira, ou seja, pode ser considerada uma amostra da amostra.

Nessa perspectiva, o Ministério da Ciência, Tecnologia e Inovação (MCTI) em conjunto com o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) investiu em um relevante e inovador sistema de informação curricular, que permite o registro da produção científica e tecnológica de cada pesquisador cadastrado, a Plataforma Lattes (PL).

Atualmente a Plataforma Lattes é adotada pela maioria das instituições de fomento, universidades e institutos de pesquisa do Brasil. A divulgação estatística dos dados científicos é realizada mediante indicadores macros, iniciativas nesse sentido são as análises estatísticas disponíveis na Plataforma Aquarius e Painel Lattes (CNPQ, 2015), que não permitem análises profundas e pormenorizadas. “Por sua riqueza de informações e sua crescente confiabilidade e abrangência, a PL se tornou elemento indispensável e praticamente compulsório à análise de mérito e competência dos pleitos de financiamentos na área de ciência e tecnologia” (CNPQa, 2016).

Contudo, a Plataforma Lattes também vem sendo utilizada principalmente para avaliações isoladas de grupos ou instituições (BRITO; QUONIAM, 2014; DIGIAMPIETRI, 2012), o que dificulta, quando não inviabiliza, o mapeamento de assuntos ou pólos de conhecimento nas diversas áreas de pesquisa. Este comportamento reforça a necessidade de utilização de seus dados em análises temáticas e não só por instituição, programa de pós-graduação, grupo ou lista de pesquisadores.

No Brasil, o Estado é reconhecidamente o organismo responsável pela alocação dos recursos em ciência e tecnologia, assim a necessidade de avaliação da atividade científica, torna-se fundamental, pois os investimentos destinados à ciência são limitados e competem com outros investimentos dos setores públicos (DAVYT; VELHO, 2000). As iniciativas de avaliação da ciência se concentram na elaboração de metodologias apropriadas para a construção de indicadores, que surgem da medição dos insumos (inputs) e resultados (outputs) da instituição científica, país ou região geográfica. Sendo que insumo seria a combinação de fatores que viabilizam a produção de determinada quantidade de resultado científico (SPINAK, 1998). Portanto, a avaliação é parte do processo de construção do conhecimento científico, através dela que se define o rumo da ciência e das instituições vinculadas (DAVYT; VELHO, 2000).

Desse modo, as avaliações quantitativas das informações científicas, através dos estudos métricos (OKUBO, 1997; SPINAK, 1998; FARIA et al., 2010) podem apontar caminhos para a alocação de recursos destinados a C&T. Os indicadores são utilizados pelos governos e pelas agências de fomento no planejamento, execução e acompanhamento das políticas públicas em ciência e tecnologia, pela comunidade científica como forma de compreender o sistema no qual ela está inserida. Já as empresas buscam compreender as contingências tecnológicas do seu ambiente competitivo. Assim, os indicadores são um meio para compreender e avaliar as dinâmicas das atividades científicas e tecnológicas e sua relação com a sociedade (VANTI, 2002). Gregolin et al., (2005, p. 5) definem os indicadores “como dados estatísticos usados para medir algo intangível, que ilustram aspectos de uma realidade multifacetada”.

A elaboração de indicadores normalmente trata dados de um único tipo de informação, ou de artigos publicados, ou sobre patentes, ou de pessoas ou de grupos de pessoas. Nota-se a necessidade de uma elaboração de indicadores integrada, que forneça visão holística ao se utilizar de todos esses tipos de informação, possibilitando análises multidimensionais de um dado contexto, pois “boas práticas indicam o uso de indicadores múltiplos para fornecer um quadro mais consistente e próximo do real” (HICKS, 2015, p. 2).

Para Dowbor (2003) grande parte da gigantesca quantidade de informação é lixo informativo que desorienta. Na ausência de informações articuladas, que permitam uma ação racional, a informação se constitui em um vasto recurso subutilizado. Não pela sua insuficiência, mas em parte pelo seu excesso, pois a informação útil, quando espalhada e sem objetividade, não interessa, não pode ser utilizada. Ou seja, a informação deve ser relevante e intencional para o que se faz. De certo modo é exatamente isso o que os indicadores fazem, tornam as decisões mais racionais e sustentáveis na elaboração, implementação, acompanhamento e avaliação das políticas públicas em Ciência e Tecnologia (C&T).

Em 2008 o Ministério da Ciência, Tecnologia e Inovação (MCTI) elegeu os “Programas prioritários para os setores portadores de futuro” que compreendeu diversas temáticas<sup>2</sup>: Dessa maneira, por ser considerada uma área prioritária e devido a sua importante

---

<sup>2</sup> Tecnologias da Informação e Comunicação; Fármacos e Complexo Industrial da Saúde; Petróleo e Gás; Complexo Industrial da Defesa; Aeroespacial Nuclear; Fronteiras para a Inovação (Biotecnologia e Nanotecnologia); Fomento da Economia Verde (Energia renovável, Biodiversidade, Mudanças climáticas, Oceanos e zonas costeiras); Ciência, Tecnologia e Inovação para o Desenvolvimento Social (Popularização da C,T&I e melhoria do ensino de ciências, Inclusão produtiva e social Tecnologias para cidades sustentáveis) (MCTI, 2012).

relevância para o desenvolvimento científico e tecnológico do país e sua contribuição à competitividade das indústrias, a nanotecnologia foi eleita área objeto da presente pesquisa.

Diante do exposto, cabe ressaltar que em linhas gerais as pesquisas não podem ser consideradas lineares e uma tese de doutoramento menos ainda. A hipótese inicial desta pesquisa tentava descobrir se “*era possível democratizar o uso e acesso a informação científica e tecnológica para geração de indicadores micro, meso e macro para mapear uma área prioritária do desenvolvimento nacional*”.

Após a qualificação desta tese, no segundo semestre de 2015, a hipótese não foi mais plenamente favorecida, pois a Base de Currículos Lattes recebeu modificações ligadas aos seus requisitos de segurança para extração dos dados. O CNPq (2016a) implementou o dispositivo de segurança *captcha* que impediu, o uso do principal *software* utilizado para extração e tratamento dos dados dos currículos, o ScriptLattes.

Desse modo, não foi mais viável insistir na hipótese de pesquisa inicial, pois a instituição responsável pela Base de Currículos Lattes (BCL) se posicionou contra a retirada do *captcha*, alegando fazê-lo para proteger as informações dos pesquisadores cadastrados (ver Anexo 1). Houve mobilização dos pesquisadores que estudam Informação, Científica e Tecnológica (ICT) para a criação de uma petição *on line*<sup>3</sup> com mais de 3.300 assinaturas, mas não foi suficiente para revogação da decisão do CNPq (GALILEU, 2016). O responsável pela petição foi o Professor Jesús Pascual Mena-Chalco, criador do ScriptLattes, a carta elaborada pode ser encontrada no Anexo 1.

Todavia, vale ressaltar que até o presente momento não se tem conhecimento de nenhuma publicação que fez uso destes dados com má fé e que possam ter denegrido ou exposto qualquer indivíduo, uma vez que os pesquisadores e profissionais que se utilizavam do ScriptLattes para realizarem suas pesquisas, se pautam em uma pesquisa ética, como por exemplo, seguindo as diretrizes descritas no manifesto de Leiden (HICKS, 2015) - subseção 2.2.1.

Para esta tese, a alternativa foi expandir o trabalho de pesquisa, visando agora o Acesso Aberto e apresentando um problema de pesquisa ainda mais abrangente que reuniu dados e informações de quatro diferentes fontes de informações com acesso aberto ou acesso livre (*open access*). De acordo com o manifesto de *Budapest Open Access Initiative*, o acesso aberto é definido pela disponibilização livre e permanente, na internet, de cópias gratuitas e sem restrições de acesso, de literatura de caráter científico ou acadêmico (BOAI, 2002).

---

<sup>3</sup> Disponível em: <https://www.change.org/p/peti%C3%A7%C3%A3o-para-a-retirada-dos-captchas-dos-curr%C3%ADculos-da-plataforma-lattes>. Acesso em: 21 out 2016.

Com base nessas reflexões e nas lacunas de indicadores de Ciência, Tecnologia e Inovação no Brasil, *seria possível incrementar a construção e identificação dos indicadores de áreas prioritárias a partir de base de dados de acesso aberto e ferramentas livres?*

O objetivo principal da pesquisa foi *desenvolver e aplicar uma sistemática para coleta, tratamento e análise de informações sobre áreas do conhecimento científico no Brasil, utilizando essencialmente dados de acesso aberto e ferramentas livres*. Para alcançar este objetivo o método de pesquisa utilizado foi o estudo de caso acompanhado de um levantamento documental, desenvolvido mediante análises quantitativas. A amostra analisada compreendeu um conjunto de informações relacionadas a nanotecnologia e extraídas das bases de dados: Currículo Lattes, Diretório de Grupos de Pesquisa, Biblioteca Digital Brasileira de Teses e Dissertações e Espacenet (*European Patent Office*). O objetivo principal pode ser desdobrado nos seguintes objetivos específicos:

- Levantar as ferramentas computacionais de uso livre adequadas para a coleta de dados e informações das bases de dados utilizadas;
- Elaborar formas válidas de recuperação da informação para coletar dados e informações do Currículo Lattes, Diretório de Grupos de Pesquisa, Biblioteca Digital Brasileira de Teses e Dissertações e Espacenet, sobre a Nanotecnologia;
- Esquematizar os procedimentos da sistemática para coleta, tratamento e análise de dados e informações, da Base de Currículos Lattes, Diretório de Grupos de Pesquisa, Biblioteca Digital Brasileira de Teses e Dissertações e Espacenet relacionadas à nanotecnologia;
- Realizar a aplicação da sistemática mediante a construção de indicadores sobre nanotecnologia;

A proposta desta tese é inovadora por permitir elaborar indicadores micro, médio e macro por intermédio das buscas por assunto na Base de Currículos Lattes, Diretório de Grupos de Pesquisa, BDTD e Espacenet o que até então, não tem sido realizado de forma integrada. Além disso, grande parte das análises bibliométricas realizadas atualmente, utilizam bases de dados internacionais proprietárias e que não externalizam o comportamento e a real representatividade da ciência brasileira.

Desta forma, os resultados da ciência puderam ser acessados, neste trabalho: pela produção da informação científica com a BDTD e a Base de Currículos Lattes; pela produção da informação tecnológica com a Espacenet; nas teses e dissertações (ensino e pesquisa) com

a BDTD; e pelos pesquisadores (pessoas) com a BCL. As bases de dados selecionadas, além de serem em acesso aberto constituem fontes de informação específicas:

- Base de Currículo Lattes: abrange a trajetória de pessoas;
- Diretório de Grupos de Pesquisa: abrange agrupamentos temáticos de pessoas;
- Biblioteca Digital Brasileira de Teses e Dissertações: abrange resultados das pesquisas acadêmicas;
- Espacenet: abrange documentos de patentes de mais de 90 países.

Almeja-se que os indicadores aqui realizados possam ser utilizados para a avaliação em ciência e tecnologia ao oferecer múltiplas dimensões da realidade. Bem como, proporcionar subsídios para uma tomada de decisão mais racional e sustentável no que tange a formação das Políticas em Ciência e Tecnologia no Brasil, favorecendo melhor gerenciamento da agenda de Inovação do país.

Mugnaini (2015), como outros pesquisadores corroboram com essa afirmação:

[...] fazer qualquer avaliação exige, primordialmente, acesso aos dados. Por essa razão temos o primeiro problema: a produção brasileira é pouco representada nas bases de maior reconhecimento internacional. O caminho [...] é no sentido de separar indicadores nacionais de internacionais. A produção brasileira tem importância relativa em cada um desses contextos [...] O contexto nacional precisa ser considerado, e para isso fontes de informação nacionais precisam ser utilizadas.

Os indicadores disponibilizados atualmente sobre BCL, DGP (Diretório de Grupos de Pesquisa) e BDTD permitem apenas análises macro, que podem sinalizar algumas respostas, entretanto sem especificações do conhecimento não elucidam as características proeminentes de cada área. Já a Espacenet não oferece possibilidades de interação *on line* com os dados recuperados. As análises por assunto oportunizam possibilidades de conhecer aspectos mais reveladores do avanço científico que interessam aos pares e aos tomadores de decisão das Políticas Públicas (representantes do Estado).



## **2 ESTUDOS SOCIAIS DA CIÊNCIA E ACESSO ABERTO EM INFORMAÇÃO, CIENTÍFICA E TECNOLÓGICA**

No modo ocidental de se fazer ciência verifica-se um propósito de dominação (relacionado ao empirismo), pois ao se ter informação se tem poder. Nesse contexto, a indução perde espaço para dedução e as “coisas” são o que dizem que são de si mesma e do outro. Assim, a modernidade tem na ciência a sua alavanca forte. Contudo, a ciência não é autônoma, os homens se apropriaram da ciência e atribuíram um fim a ela, provavelmente quando o capitalismo descobriu o poder vindo do conhecimento (FEYERABEND, 1989).

Em seu livro “Contra o Método”, Feyerabend (1989) defende a ideia de que não há regras metodológicas que devam sempre ser usadas pelos cientistas. Afirma que a fundamentação prescritiva do método científico limita as atividades dos cientistas e dessa maneira restringe o progresso científico. A ciência se beneficiaria mais com um viés libertário, que ele chamou de anarquismo teórico. O autor também sugere que o anarquismo teórico é preferível também por ser mais humanizado do que outros sistemas de organização, pois não impõe regras tão rígidas aos cientistas.

Entretanto, a ciência também pode ser considerada uma instituição, embora não tenha uma constituição escrita, em essência se comporta como uma organização social, formada por diversos grupos de pessoas, que dependem de laços de confiança pessoal e institucional para alcançarem seus objetivos comuns (ZIMAN, 2000). Uma das formas de estudar e compreender a sua atuação é mediante as contribuições dos teóricos dos Estudos Sociais da Ciência.

A posição de Feyerabend é geralmente vista como radical na filosofia da ciência, pois implica que a filosofia não consegue propiciar uma descrição geral da ciência, nem permite descobrir um método de diferenciação entre produtos da ciência e entidades não científicas como os mitos. Isso também reflete que as recomendações filosóficas podem ser ignoradas pelos cientistas, se o seu objetivo é o progresso.

Para comprovar a posição de que as regras metodológicas geralmente não contribuem para o sucesso científico, Feyerabend mostra contraexemplos de que a boa ciência opera de acordo com um certo método fixo. Ele examina alguns exemplos de episódios da ciência que são geralmente relacionados como instâncias inegáveis de progresso (por exemplo, a revolução Copernicana), e mostra como todas as regras prescritivas comuns da ciência são violadas nestas circunstâncias.

Iniciando com a suposição de que um método científico universal historicamente não existe, Feyerabend afirma que a ciência não merece o *status* privilegiado que detém na sociedade ocidental. Uma vez que os pontos de vista científicos não surgem de um método universal que possa garantir conclusões de alta qualidade. Não há justificção para a valorização científica reivindicada sobre outras ideologias, como as religiões, por exemplo.

O autor também afirma que conquistas científicas como a chegada do homem à lua, por exemplo, não são razão suficiente para dar à ciência um *status* especial. Em sua opinião, não é justo utilizar suposições científicas para determinar quais problemas merecem ser resolvidos e para julgar o mérito de outras ideologias. Além disso, o sucesso dos cientistas está tradicionalmente envolvido com elementos não-científicos, tais como inspiração a partir de pensamentos míticos ou de fontes religiosas.

Baseado nestes argumentos, Feyerabend defendeu a ideia de que a ciência deve ser separada do estado, da mesma maneira que a religião é separada na moderna sociedade. Ele vislumbrou uma sociedade livre na qual todas as tradições têm iguais direitos e igual acesso aos centros de poder. Por exemplo, os pais devem ser capazes de determinar o contexto ideológico da educação de seus filhos, em vez de terem suas opções limitadas pelos padrões científicos.

De acordo com Ziman (2003), a presença da ciência domina a sociedade atualmente. E o fazer científico, se apresenta como uma tentativa de compreensão do mundo, às vezes bem-sucedida, outras nem tanto, é o como controlar as coisas, de ter domínio sobre “nós mesmos”, de seguir um caminho que se julga seguro (SAGAN, 2006). E atualmente, tem-se a impressão de que a ciência está mais próxima do cotidiano, pois presencia-se um momento de grande influência midiática, onde os meios de comunicação em massa não têm receio de tratar as questões científicas e tecnológicas. Nessa lógica, também se vivencia uma onipresença da ciência no coletivo.

Nessa perspectiva, a democratização da ciência e sua comunicação pública, apresenta-se a seguir quatro modelos relacionados aos estudos sociais da ciência e da tecnologia. Tais modelos focam duas grandes tendências: unidirecional com o modelo de déficit cognitivo e modelo contextual – e dialógica com o modelo da expertise leiga e modelo de participação pública (LEWENSTEIN, 1992).

No modelo de *déficit* cognitivo os cientistas são considerados os especialistas, que detêm o conhecimento e o público é visto como carente (ou com um *déficit*) de conhecimentos de fatos relevantes de ciência e tecnologia. Essa ausência de conhecimento é

identificada como analfabetismo científico. O modelo contextual assume que os indivíduos não recebem a informação como recipientes vazios. Valoriza as experiências culturais e os saberes prévios, reconhecendo que quando os conhecimentos científicos ou tecnológicos fazem parte do contexto e do entorno de quem está se aproximando deles, o processo de compreensão é facilitado. No modelo de expertise leiga o conhecimento existente na sociedade é considerado relevante e é valorizado nas soluções de problemas científicos e tecnológicos. Por último, o de participação pública, confere a participação do público e dos cientistas (em assuntos de C&T e na formulação de políticas científicas e tecnológicas) nas mesmas condições e em espaços propícios para isso como fóruns, debates e conferências de consenso (LEWENSTEIN, 1992).

De acordo com Feyerabend (1989), a ciência deve também estar sujeita ao controle democrático: não apenas determinar que assuntos devam ser pesquisados através de eleição popular, as suposições e conclusões científicas também devem ser supervisionadas por comitês de pessoas leigas. Segundo ele, os cidadãos devem utilizar seus próprios princípios ao tomar decisões a respeito do que realmente importa. Assim, a ideia de que as decisões devam ser racionalistas é elitista, pois presume que filósofos ou cientistas estão em posição de determinar os critérios pelos quais as pessoas em geral devem tomar suas decisões.

## **2.1 Sociologia da ciência**

A Sociologia da Ciência como um ramo de estudo da sociologia está dentro do campo da sociologia do conhecimento, estuda a influência de fatores externos no desenvolvimento da ciência. Após o grande impulso de Thomas Kuhn e com a radicalização da posição kuhniana, surgiram estudos cada vez mais radicais que pensavam a verdade científica como algo puramente conformado por fatores sociais, como as posições da Escola de Edimburgo e seu Programa Forte de Sociologia, a antropologia da ciência de Bruno Latour, e toda uma vertente de estudos pós-kuhnianos e pós-modernos (VESSURI, 1991).

As reflexões da corrente clássica da Sociologia da Ciência com a intenção de avaliação da ciência se iniciaram com as contribuições do norte americano Robert K. Merton. Para ele, a ciência é uma instituição social sustentada por um conjunto de normas próprias, tornando-a uma comunidade científica. Merton compreendia as características da investigação científica, como sendo de dois tipos: internas e institucionais. Nas internas estão incluídas a coerência lógica e a realização empírica. Nas institucionais, estão as especificidades institucionais da

ciência que descendem das internas e se resumem no *ethos* da ciência. O *ethos* da ciência é composto por uma série de normas não escritas e caracterizado por quatro imperativos institucionais: o comunalismo, o universalismo, o desinteresse e o ceticismo organizado. Propostos em 1942, esses quatro imperativos ficaram conhecidos sob o acrônimo de CUDOS (MERTON, 1970).

Em 1957 Merton intenta melhorar sua proposta e acrescenta outras duas normas – originalidade e humildade – publicando o acrônimo CUDOSH, cujo S refere-se a ceticismo em inglês (*skepticism*). Para Merton esses princípios garantiriam o que ele considerava “boa ciência”, pois sendo aprovada pela sociedade, também seria neutra e livre para continuar progredindo. Nesse contexto, a ciência estaria acima dos conflitos sociais e a serviço da sociedade (HAYASHI, 2013; MERTON, 1970).

Na década de 60 a obra de Thomas Kuhn merece destaque, pois evidenciou a ruptura da visão da ciência como um sistema autônomo de produção de verdades. Em 1962 após a publicação de sua obra “A estrutura das revoluções científicas”, a visão clássica e mais ingênua da ciência sofre intensas críticas. Para Kuhn a ciência não é linear e cumulativa, mas cíclica em relação aos períodos de normalidade e crise, que geram as mudanças de paradigmas, ou seja, para Kuhn havia uma espécie de alternância entre a ciência normal e ciência revolucionária. E a comunidade científica vista como autônoma, isolada e auto reprodutora, com cientistas neutros e interessados somente no progresso da sua área. Desse modo não se pode afirmar que Kuhn explicou a dinâmica das práticas científicas na sociedade moderna (HOCHMAN, 1994).

Dentre outras contribuições que decorreram as de Kuhn, destacam-se os pressupostos do Programa Forte, a análise de controvérsias, o desenvolvimento das abordagens antropológicas e os estudos que propõem apreenderem uma mesma dinâmica o universo dos artefatos e dos homens.

As análises das epistemologias lógicas fundadas na filosofia da linguagem foram retomadas pela sociologia do conhecimento, inspirando, nos anos 70, o chamado "Programa Forte de Sociologia". Com base na noção de construção linguística como construção convencional, indissociável dos processos de interação social entre os indivíduos e de suas necessidades práticas, Bloor, Barnes e outros derivam a concepção de ciência desta noção de construção. A perspectiva construtivista, que inclui não somente a noção de linguagem, método e objeto construídos, mas também a ideia de que a legitimação dos conhecimentos científicos se constrói social e historicamente, parece impor-se cada vez mais no decorrer de nosso século (PORTOCARRERO, 1994, p. 27).

Autores como David Bloor, Harry Collins, Steve Woolgar, Callon, Bruno Latour, Shapin e Schaffer são destaques desse processo (PORTOCARRERO, 1994).

Pierre Bourdieu no final dos anos 80 elucida nos estudos sociais da ciência a compreensão da especificidade social além da cognitiva. Além de reconhecer na história da ciência as suas relações com a sociedade e o seu papel nos processos sociais de dominação. Bourdieu propõe o conceito de campo científico, um espaço metafórico de luta política pela dominação científica onde se manifestam relações de poder e no qual há uma distribuição desigual de um tipo específico de capital (HAYASHI, 2013).

O Campo Científico na visão de Bourdieu (1975) é um sistema de relações objetivas entre posições adquiridas em batalhas realizadas anteriormente. O campo científico é local de competição no qual está em jogo especificamente o monopólio da autoridade científica, definida, de modo inseparável, como a capacidade técnica e o poder social, ou, de outra maneira, o monopólio da competência científica, no sentido da capacidade - reconhecida socialmente - de um agente falar e agir legitimamente em assuntos científicos. O campo científico, assim como a sociedade, está dividido entre dois pólos: a) o dos dominantes, que ocupam a hierarquia superior na distribuição de capital científico e que podem impor a definição de ciência que se conforma com seus interesses; e b) o dos dominados, com pouco ou nenhum capital, situando-se na hierarquia inferior do campo (BOURDIEU, 1975).

Dito isto, crédito científico pode ser considerado um capital simbólico, mas não-monetário, composto pela autoridade e/ou competência científica. Um capital "que pode ser acumulado, transmitido e até reconvertido, sob certas condições, em outros tipos de capital", isto no mercado específico da produção do conhecimento científico (BOURDIEU, 1975, p. 25).

Bourdieu ao propor a noção de campo científico fez uma crítica ao interacionismo de Latour e Merton, pois para ele, não é um espaço de interação entre iguais, mas sim um espaço estruturado e já apropriado de antemão. Os agentes pertencentes ao campo têm mais ou menos capital acumulado e os que o possui tem condições de dominar o campo. No Brasil pode-se ver os dominantes dos campos em posições de poder como nos Comitês da Capes e do CNPq.

Em contrapartida, Latour crítica a visão de Bourdieu por ser panorâmica, uma espécie de sobrevoos, pois aponta o óbvio, uma vez que dominantes há em todo lugar. Para Latour cada laboratório tem formas diferentes de atuação, relacionadas às especificidades das áreas científicas. Para esta seção acredita-se que é a posição que permite a escolha de estratégias,

nesse sentido quanto mais um campo científico acumula créditos mais ele impõe isso aos outros campos.

Por fim, a analogia entre campo científico e mercado capitalista foi além, pois vê a ciência como um mercado particular dentro da ordem econômica capitalista. Ao olhar para a comunidade científica “autônoma e neutra”, Bourdieu (1975) descobre o mercado.

Latour analisou o laboratório, investigando como a ordem científica é criada a partir do caos, em um processo no qual, o observador é tão construtor de fatos quanto o cientista observado. O seu livro *Laboratory Life* apresenta uma fotografia do funcionamento das práticas científicas e da própria ciência, bem diferente das análises que privilegiam as macro dimensões da vida social (HOCHMAN, 1994).

Nesse estudo, cientistas e grupos de trabalho aparecem como estrategistas, negociadores, calculadores, mobilizadores de recursos de todos os tipos, em permanente competição. Fazem parte de um contexto onde existem apenas dois tipos de consenso: o primeiro diz respeito à tradição em que se inserem, quanto ao passado da disciplina e à sua base conceitual; e segundo ao fato de que os recursos utilizados na competição devem ser apresentados e reconhecidos por todos como científicos (HOCHMAN, 1994).

Latour e Woolgar (1979) sugerem a extensão do conceito de crédito para credibilidade. Essa ampliação sustenta, como em Bourdieu (1975), elementos econômicos em que o cientista avalia as oportunidades do campo. Só que, agora, o objetivo principal, da atividade científica, é o reinvestimento dos recursos acumulados, formando um ciclo de credibilidade, uma evidente associação entre o ciclo do cientista e o ciclo de investimento de capital. Ou seja, o comportamento do cientista é similar ao do investidor de capital.

A proposta do ciclo de credibilidade torna possível compreender a conversão de dinheiro, equipamentos, informações, prestígio, credenciais, áreas de estudo, argumentos, *papers*, livros, prêmios, o que vincula o cientista ao mundo exterior ao laboratório, por exemplo, com agências de fomentos a pesquisa, fornecedores etc. (HOCHMAN, 1994).

Latour e Woolgar (1979), ao observarem grupos de pesquisas, apontam o principal pesquisador, chefe do laboratório (o *chairman*), como um empresário capitalista que contrata e despede, tem técnicos e cientistas trabalhando para ele, podendo aumentar seu capital sem estar engajado diretamente na atividade. Sua posição será mantida se continuar administrando seu laboratório e fazendo com que produza informações consideradas relevantes em áreas importantes, obtendo credibilidade, recursos e colaboração para uma conversão acelerada de um tipo de credibilidade a outro.

Vessuri (1991) resume de forma excelente o comportamento da ciência formal ao apresentar as perspectivas recentes dos estudos sociais, ressalta como marco inicial o crescimento lento e firme dos estudos de Robert K. Merton entre as décadas de 1940 e 1960, que evidenciaram o contraponto dos estudos em sociologia da ciência e sociologia do conhecimento. As abordagens utilizadas em sociologia para compreender o comportamento da ciência individualmente se debruçam sobre aspectos de uma vertente de estudos pós-kuhnianos e pós-modernos e a partir de uma visualização macro estabelece explicações e reflexões sobre o comportamento da ciência formal, sendo elas:

- A primeira abordagem é a chamada de “Programa Forte da Sociologia do Conhecimento Científico” preocupada em explicar que as contribuições convergiram em um esforço para provar que a ciência deixou de ser uma atividade autônoma, regida por suas próprias leis, estando determinada por fatores sociais, mas o que começou como uma promessa de programa forte que revolucionaria a sociologia da ciência com a sociologia do conhecimento, resultou em um produto instável, sem composição firme.
- A abordagem da “Exploração da Dimensão Microsociológica” também foi abordada, mediante os estudos em laboratórios de Karin Knorr Cetina e também a obra de Latour, que constitui uma das propostas mais recentes e brilhantes e talvez a mais frustrante, para a investigação empírica e teórica da sociologia da ciência.
- A “sociologia da inovação” é tratada por Michel Callon e John Law - no desenvolvimento da ANT - Actor Network Theory (Teoria ator-rede) que, ao analisar a atividade científica, considera tanto os atores humanos como os não humanos, estes últimos devido à sua vinculação ao princípio de simetria generalizada. Atores em rede são entidades heterogêneas que constituem uma rede e incluem, entre outros, catalisadores, acumuladores, usuários, investigadores, fabricantes e departamentos governamentais que definem e implementam regulações que afetam a tecnologia em uma trama sem costura que não distingue o animado do inanimado, por um lado razões científicas e técnicas e por outro, razões sociais e políticas.
- As “Disciplinas Científicas - como a infra-estrutura da ciência”, diferem da abordagem tradicional da história da ciência, pois no passado constituíram referências para histórias naturais descritivas do conhecimento, eram descritas por praticantes ativos que acumulavam conhecimentos cada vez mais perfeitos, diante de seus próprios pontos de vistas. Hoje as disciplinas são territórios acadêmicos que distribuem

privilégios e responsabilidades de conhecimento e estruturam as pretensões sobre os recursos.

- A última abordagem tratada foi a “Cientometria e a Política Científica”, que fez uso sistemático de medições nos estudos sociais da ciência e obteve um desenvolvimento significativo que pode ser considerado como uma ferramenta indispensável para a administração da instituição e relevante para seu êxito. Em vista da unidade ciência e sociedade, não surpreende que os intentos de estender uma métrica da esfera natural e social e inclusive de medir a própria ciência tenham uma história rica, complexa e variada (VESSURI, 1991).

Por fim, não há dúvida de que as várias abordagens permitiram vislumbrar uma eventual síntese que, superando os excessos das modas retóricas passageiras, ainda se esforçam para iluminar contrastes e complementaridades, mas que facilitam uma compreensão mais adequada da dinâmica do avanço científico, pois os dados sociológicos da ciência e da tecnologia tem um amplo potencial para contribuir a uma maior compreensão do solo da ciência e da cultura humana.

As ideias de Kuhn, Bloor, Bourdieu e Latour se assemelham em alguns aspectos e discordam em outros. Nota-se que o pensamento do valor da ciência é bastante dicotômico, uma vez que vai desde sua compreensão como um sistema de conhecimento puramente teórico e neutro até a ideia de ciência como prática política. De acordo com Portocarrero (1994, p. 29), para uma nova história das ciências, “a distinção internalismo e externalismo e, todas as tentativas de sua síntese, são pouco produtivas e devem ser substituídas por uma reflexão sobre as condições históricas da formação destas duas categorias, sua genealogia, com o objetivo de ultrapassar tal dicotomia”.

## **2.2 Avaliação da ciência**

Os Estudos Sociais da Ciência e a Cientometria são campos de estudo voltados para a análise das práticas e dinâmicas das comunidades científicas e da avaliação da produção científica e tecnológica produzidas pelos cientistas em suas respectivas áreas de conhecimento.

A *priori* o tema da avaliação apresenta dois elementos intrínsecos, a qualidade e a atitude comparativa. Quanto à qualidade, é importante ressaltar as dificuldades associadas à falta de um debate mais aprofundado sobre as razões e fontes desse julgamento. Em geral, nas

discussões sobre o assunto, o que está em debate é o consenso e dissenso acerca do tema, entre a coletividade científica (BAUMGARTEN, 2004).

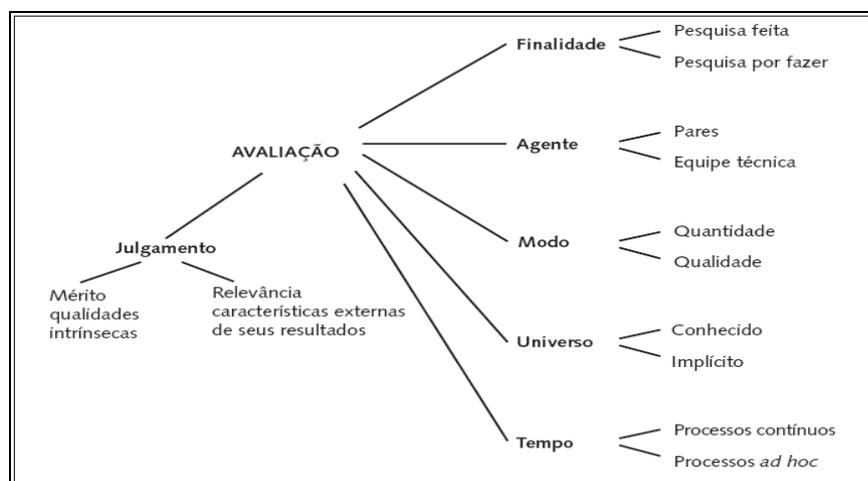
As origens da avaliação no processo de construção do conhecimento podem ser encontradas com o surgimento da própria ciência. Uma vez que, a ação avaliativa naturalmente exige o uso de valores e julgamentos subjetivos, incorporando uma série de elementos culturais e contextuais (DAVYT; VELHO, 2000).

De acordo com Guimarães (1994), os processos avaliativos no setor de Ciência e Tecnologia (C&T) do Brasil podem ser realizados a partir dos seguintes parâmetros e suas diversas possibilidades de combinações:

- a) Dependendo da finalidade da avaliação: incide sobre a pesquisa já feita ou por fazer;
- b) De acordo com o agente: a avaliação pode ser efetuada por pares da coletividade científica (*peer review*), ou por equipes internas da agência financiadora (corpo técnico);
- c) Quanto ao modo de avaliação: este pode ser qualitativo ou quantitativo;
- d) Quanto ao universo: pode-se trabalhar com um universo conhecido ou implícito; e
- e) Quanto ao tempo: os processos podem ser contínuos ou *ad hoc* (GUIMARÃES, 1994).

A Figura 1 ilustra os elementos compreendidos pelos processos avaliativos no setor de C&T do Brasil, cuja fase de julgamento antecede a própria avaliação e podem ser: por mérito com qualidades intrínsecas; e por relevância com características externas de seus resultados. No esquema há cinco critérios a serem observados e estão relacionados a: finalidade (pesquisa feita ou pesquisa por fazer); agente (pelos pares ou por equipe técnica); modo (quantidade ou qualidade); universo (conhecido ou implícito); e tempo (processos contínuos ou processos *ad hoc*).

Figura 1 – Esquema da avaliação em ciência



Fonte: Baumgarten (2004)

Como qualquer modelo, o esquema de Baumgartem (2004) é um esforço para compreender a realidade segmentando-a, o que pode ser bem mais complexo na prática. No Brasil, a gestão do setor de C&T é efetuada em conjunto pelo Estado e pelos cientistas, resultando na atual estrutura científica e tecnológica do país (BAUMGARTEM, 2004).

O sistema de avaliação da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) é bastante recente, tendo iniciado em fins da década de 1970, e sua estruturação decorreu do crescimento acentuado da pós-graduação no país a partir desse período. Hoje, além da sua atuação como agência de fomento, a Capes é responsável pela avaliação do Sistema Nacional de Pós-graduação.

A Capes, pauta sua principal avaliação mediante níveis de qualidade das publicações científicas realizada por meio do Qualis, implantado em 1998. As classificações do Qualis são publicadas quadrienalmente, a partir de critérios definidos pelos comitês de áreas, constituídos por membros da comunidade científica. Cada área possui os seus critérios na definição dos estratos dos periódicos que podem ser A1 (nível mais alto), A2, B1, B2, B3, B4, B5 e C (com peso zero) (FRIGERI; MONTEIRO, 2015).

Uma consequência desse sistema de avaliação é que o Qualis vem sendo utilizado como um indicador para:

concessão de financiamentos, para a inclusão de títulos em bibliotecas e indexadores, na orientação de pesquisadores e leitores durante a escolha de títulos, na submissão de trabalhos e na pesquisa de material bibliográfico, além de estimular os editores a elevar o padrão de qualidade dos seus periódicos a partir da adequação aos critérios da Capes. Há uma busca por criar estruturas que permitam a padronização dos periódicos, além da crescente indexação em bases internacionais. Muitos periódicos buscam atrair autores estrangeiros, a partir da pressão crescente para internacionalizar o conhecimento científico brasileiro (FRIGERI; MONTEIRO, 2015, p. 3).

Outra iniciativa é o SciELO, que exerce no Brasil um papel semelhante ao do ISI *Institute for Scientific Information*, indexando as melhores revistas brasileiras selecionadas por critérios de qualidade. A atividade principal do ISI consiste em coletar artigos publicados nas revistas mais prestigiosas do mundo, em todos os ramos da ciência, e desde o ano de 1992 o ISI passou a pertencer à empresa *Thompson Reuters*.

O SciELO foi concebido como um projeto e uma estratégia para superar o fenômeno conhecido como "ciência perdida", causado pela presença muito fraca dos periódicos de países

em desenvolvimento nos índices internacionais. Além da falta de visibilidade, este fenômeno também se manifestou na falta de comunicação, em muitas áreas, entre os pesquisadores de países desenvolvidos e em desenvolvimento, bem como entre os pesquisadores de países em desenvolvimento (GIBBS, 1995).

A função de indexação do SciELO, guiada pelo critério de qualidade, foi concebida para complementar a indexação internacional, especialmente a do *Science Citation Index*, com a visão de ampliar a cobertura dos periódicos com um sistema *online* capaz de medir o desempenho do periódico em número de *downloads* e indicadores bibliométricos baseados em citações, como os proporcionados pelo *Journal Citation Reports* (PACKER, 2014).

Há, assim, uma mudança de alvo, passando-se dos estudos da ciência e seus aspectos sociológicos para a avaliação da pesquisa através da construção de indicadores que permitem avaliar a produtividade e a posição estratégica dos diferentes atores participantes deste processo, dentre outras possibilidades.

Os Estudos Sociais da Ciência são comparados por Velho (1994, p. 309) a um “circo com três arenas”, cada uma com seu próprio discurso e que busca a primazia como um programa instrutivo: a) o “Programa Forte” da sociologia do conhecimento; b) os “Estudos de Laboratório”, que incluem uma variedade de aproximações ao estudo da atividade científica; c) a Cientometria, que inclui todo o tipo de análises quantitativas da ciência, e que se baseia fundamentalmente em arquivos como fontes.

Assim, as dimensões sociais da atividade científica vistas pelas lentes da Sociologia da Ciência e dos Estudos Sociais da Ciência estabelecem relações com o referencial teórico da Bibliometria e da Cientometria. No campo da Ciência da Informação, os estudos métricos da informação – entre eles, a Bibliometria, a Cientometria, a Informetria, a Webometria – constituem-se como campo interdisciplinar dedicado ao estudo quantitativo da ciência e da tecnologia e estão voltados para avaliar a produção científica e tecnológica produzida pela comunidade científica no interior das áreas de conhecimento, representada por artigos, livros, capítulos de livros, trabalhos publicados em anais de eventos, e também patentes (HAYASHI, 2013).

A Bibliometria apesar de utilizar técnicas estatísticas para analisar os textos científicos, o faz a partir do tratamento das referências bibliográficas presentes nesses textos e não no seu conteúdo. Historicamente, as premissas do conceito de Bibliometria apareceram no início do século XIX, e durante o século XX evoluiu em termos de fundamentos, técnicas e aplicações dos métodos bibliométricos. A definição do termo “Bibliometria” foi pleiteada

pelas correntes de autores franceses, que a concedem a Paul Otlet por ter utilizado o termo no seu “Tratado da Documentação”, publicado em 1934 e a dos autores anglo-saxônicos, que atribuem a invenção a Pritchard, por ser este o primeiro autor, em 1969, a cunhar o termo “Bibliometria” no sentido da aplicação dos métodos estatísticos aos livros e outros meios de comunicação (HAYASHI, 2013).

Outras definições dadas à Bibliometria, segundo Spinak (1998) são as seguintes: aplicação de análises estatísticas para estudar as características do uso e criação de documentos; estudo quantitativo da produção de documentos como se reflete nas bibliografias; aplicação de métodos matemáticos e estatísticos ao estudo do uso que se faz dos livros e outros meios dentro e nos sistemas de bibliotecas; estudo quantitativo das unidades físicas publicadas, ou das unidades bibliográficas ou de seus substitutos.

A Cientometria nasce, portanto, a partir dos anos 1960 com o objetivo de estudar a atividade científica como fenômeno social e mediante indicadores e modelos matemáticos.

Nos Estados Unidos, a Cientometria aparece ligada ao nome de Derek de Solla Price, particularmente aos seus livros: *Science since Babylon* (1961) e *Little Science, Big Science* (1963). Com embasamento de trabalhos anteriores, mas defendendo uma *ciência da ciência*. Solla Price ampliou consideravelmente a perspectiva da Bibliometria, e impulsionando este campo de pesquisa nos Estados Unidos. (ROSTAINING, 1997 *apud* HAYASHI, 2013).

Eugene Garfield, em 1960 a partir do *Institute for Scientific Information* (ISI), produtor da base de dados Science Citation Index (SCI) permitiu a divulgação dos trabalhos de Price e o crescimento dos conhecimentos estatísticos dos artigos científicos. Garfield, com o *Science Citation Index* analisou a dimensão coletiva da atividade de pesquisa e o processo dinâmico de construção dos conhecimentos relacionados à pesquisa e a inovação e põe em prática três tipos de crenças: os estudos das ciências e das técnicas se dão pela análise sistemática da produção dos pesquisadores; os estudos quantitativos enriquecem a compreensão e a descrição da dinâmica das tecnociências (conjunto de atividades de pesquisa científica e técnica); a existência de ferramentas sólidas e confiáveis para efetuar estas análises (HAYASHI, 2013).

A bibliometria é baseada “na enumeração e análise estatística da produção científica na forma de artigos, publicações, citações, patentes e outros indicadores mais complexos. É uma ferramenta importante na avaliação de atividades de pesquisa, laboratórios e cientistas, bem como as áreas científicas e desempenho dos países” (OKUBO, 1997, p. 01).

Rostaing<sup>4</sup> (1997, p. 15) citado por Hayashi (2013) assinala que as técnicas bibliométricas se revelaram perfeitamente adaptadas à avaliação da atividade científica ou da atividade de propriedade industrial, enquanto ferramenta de apoio à atividade de vigilância industrial ou tecnológica. De acordo com o autor, esta prática iniciou nos anos 1980, nos Estados Unidos, com os trabalhos de Francis Narin sobre as bases de dados de patentes americanas e posteriormente, o foco se voltou para a aplicação das técnicas bibliométricas em dados de propriedade industrial, facilitada pela existência de bancos de dados que repertoriavam os depósitos de patentes nacionais e internacionais sob a forma de referências bibliográficas.

## 2.2.1 Indicadores de ciência e tecnologia

A reunião do amplo conjunto de teorias sobre a ciência, que se valem da observação do comportamento dos cientistas, da formação e organização de suas comunidades e de sua interação com a sociedade, constituem os Estudos Sociais da Ciência e da Tecnologia, cuja comunicação é considerada como o âmago da questão da ciência. E nela, a produção científica é um recurso indispensável para promover o desenvolvimento da ciência. Nas últimas décadas, os indicadores de produção científica vêm ganhando importância crescente como instrumentos para análise da atividade científica e suas relações com o desenvolvimento econômico e social de todo o mundo (AGRAWAL; BROAD; GUSTON, 2001).

Tais indicadores também são utilizados para formulação de políticas públicas e tomada de decisões estratégicas em C&T. No entanto, a incapacidade de relacionar e interpretar medidas específicas para o desenvolvimento histórico da ciência pode levar a erros na avaliação de acontecimentos passados ao priorizar os acontecimentos futuros. Ou seja, deve-se defender uma utilização mais sistemática das evidências históricas para a formulação de política científica (BELOFF, 1968; CAMPBELL, 2010).

As métricas geralmente dizem respeito a um sistema de medição, indicadores, para uma ampla gama de dados quantitativos de atividades selecionadas. Os indicadores de ciência são instrumentos úteis - e, por vezes, necessários - para a determinação das prioridades nacionais e avaliação de retorno. Contudo, seu uso resulta na dificuldade em medir

---

<sup>4</sup> ROSTAING, H. **La bibliométrie et ses techniques**. Toulouse: Sciences de la Société; Marseille: Centre de Recherche Rétrospective de Marseille, 1997.

diretamente as contribuições da ciência à atividade econômica (DERRICK; PAVONE, 2013; HAGENDIJK, 2004).

Atualmente a bibliometria (cientometria) visa três principais grupos - alvo que determinam claramente temas e subáreas (GLANZEL, 2003; KIERNAN, 2003; KRISHNA; WAAST; GAILLARD, 2000; LARÉDO; MUSTAR, 2004; LEE; BOZEMAN, 2005):

- 1) Bibliometria para bibliometristas (Metodologia) - este é o domínio da pesquisa bibliométrica básica.
- 2) Bibliometria para disciplinas científicas (Informação Científica) - os pesquisadores em disciplinas científicas formam o maior, mas também o mais diversificado grupo de interesse em bibliometria. Devido à sua formação científica, os seus interesses estão fortemente relacionados a sua especialidade. Este domínio pode ser considerado uma extensão da Ciência da Informação. Aqui também encontramos uma fronteira comum com a pesquisa quantitativa em recuperação da informação.
- 3) Bibliometria para a política e gestão (Política Científica) - este é o domínio de avaliação de pesquisa, atualmente o tópico mais importante no campo. Aqui as estruturas nacionais, regionais e institucionais da ciência e da sua apresentação comparativa estão em primeiro plano (GLÄNZEL, 2003, p. 9).

Assim, este trabalho tem seu foco voltado para aplicação do terceiro item – bibliometria para política e gestão – especificamente nas pesquisas sobre a nanotecnologia e, conseqüentemente, quais as contribuições os indicadores bibliométricos de produção científica podem fornecer.

O papel central dos indicadores na política científica é iluminado por Arie Rips na cientometria – um termo genérico usado para descrever múltiplos métodos de coleta, análise e apresentação de indicadores de ciência - pode ser representado por um triângulo equilátero, cujos ângulos são formados pelas perspectivas de cientometristas, cientistas e tomadores de decisões políticas (EVENSEN; CLARKE, 2012; JOHNSON, 2011). Um ângulo não faz um triângulo, ou seja, adotar a perspectiva crítica do decisor político, claramente, não é para rejeitar todos os métodos existentes de medição. Indicadores de ciência são suscetíveis de precisão em pontos onde foi alcançado um consenso. Refinamentos continuados - tais como a visualização de padrões de co-citação e rastreamento da difusão de conceitos entre as disciplinas – pode ser visto como uma promessa para melhorar sua confiabilidade, validade e capacidade de previsão.

Os indicadores científicos podem ser usados de modo macro, médio e micro. Macro indicadores são, por exemplo, a participação de um determinado país na produção global de literatura científica em um período específico, o médio seria a posição de uma instituição, como uma universidade, por exemplo; e para o micro poderia ser o papel de um grupo na produção de artigos em um campo da ciência muito restrito. Combinados a outros indicadores, os estudos bibliométricos podem ajudar tanto na avaliação do estado atual da ciência como na tomada de decisões e no gerenciamento de pesquisas (MACIAS-CHAPULA, 1998).

A abordagem quantitativa possibilita a quantificação e o dimensionamento do universo pesquisado, sendo os dados coletados, analisados e apresentados estatisticamente. Segundo Oliveira (2001), o método quantitativo significa quantificar opiniões, dados, nas formas de coleta de informações, assim como também com o emprego de recursos e técnicas estatísticas desde a mais simples, como percentagem, média, moda, mediana e desvio padrão, até as de uso mais complexo, como coeficiente de correlação, análise de regressão etc. (BOCCATO; FUJITA, 2006).

Verificar a qualidade em C&T assim como em outros contextos das atividades humanas – ingresso em universidades, desempenho dos planos de carreira, elaboração de rankings – culminou com o uso cada vez mais acentuado de indicadores quantitativos, pois permitem além de mensurar, realizar comparações (entre instituições, países, áreas, etc.), o que pode favorecer a tomada de decisão e a gestão de recursos que são quase sempre limitados e escassos.

Por permitirem comparações, os indicadores impulsionam a competitividade, estimulando o crescimento e os avanços do conhecimento gerado em todo o mundo. Entretanto, não se pode desconsiderar as críticas relacionadas a essa forma de avaliar. Elas expressam que, a construção de indicadores trata de processos por demais complexos com procedimentos simplistas, por vezes genéricos e pouco representativos da realidade, por não conseguirem mensurar aspectos subjetivos e intangíveis.

Em contrapartida é preciso reconhecer que os indicadores expressam uma amostra da realidade, e que permitem ressalvas, exceções e considerações. Por outro lado, nenhuma avaliação pode ser realizada sem critérios de comparação. Avaliar é comparar situações e realidades e embora não seja um método perfeito, é até agora o método mais conhecido e o mais objetivo e imparcial.

Do ponto de vista acadêmico, desde os anos 1970 o campo Estudos Sociais da Ciência e da Tecnologia encontra-se consolidado e tem se debruçado a estudar criticamente as dimensões sociais da ciência e da tecnologia no que diz respeito aos condicionantes sociais de mudança científica e tecnológica e aos impactos sociais de tais mudanças. Assim, engloba um amplo espectro de questões – por exemplo: ciência, tecnologia e valores, ciência e ética, ciência e gênero, ciência e arte, ciência e cultura, ciência e cidadania, produção e avaliação do conhecimento científico - que necessitam de uma variedade de ferramentas analíticas utilizadas com base em abordagens interdisciplinares e complementares (HAYASHI, 2013).

Indicadores de contexto: contêm informações sobre algumas dimensões básicas dos países, tais como população, população economicamente ativa (PEA) e produto interno bruto (PIB); indicadores de gastos: refletem o gasto realizado dentro de cada país em atividades científicas e tecnológicas (ACT) e P&D, tanto pelo setor público, como pelo setor privado.[...] relacionados a gastos em C&T, relacionados com o PIB, por habitante, por pesquisador, por setor de financiamento, por setor de execução, por objetivo socioeconômico; indicadores de recursos humanos: refletem os recursos humanos ativos em P&D nos distintos países, incluindo pessoal de C&T e pesquisadores, pesquisadores por gênero, etc.; indicadores de produtos: é utilizado para estimar os resultados das atividades de P&D. Os indicadores relacionados a este conjunto são os de solicitação de patentes, de publicações em base de dados multidisciplinares e de publicações em bases de dados temáticas (HAYASHI et. al., 2006, p. 22).

A partir do processo de construção das bases de dados internacionais, separam-se, por um lado, a literatura *mainstream*, ou seja, aquelas publicações que passam a ser consideradas os canais mais importantes de comunicação científica internacional (GARFIELD, 1985) e que sintetizam o padrão de excelência; e, por outro lado, a periferia científica, ou seja, uma ciência que não apresenta padrão e prestígio internacionais e que, portanto, é vista como não excelente. Nesse contexto, entre as décadas de 1950 e 1960, desenvolve-se o interesse dos organismos públicos na medição das atividades científicas, interesse que se consolida com o desenvolvimento da teoria e metodologia de indicadores de C&T.

A comunicação é traduzida em publicações que é o principal produto da atuação de um cientista, pois é ela quem estabelece o diálogo com o público da comunidade científica, ou seja, entre os pares. Já a divulgação científica visa o público diversificado, fora da comunidade científica (VALERIO; PINHEIRO, 2008).

Assim, a partir do que é publicado pode-se construir indicadores, no Quadro 1 é possível visualizar os principais tipos de indicadores.

Quadro 1 – Indicadores de publicação

<b>Indicadores de publicação</b>	
• Extensão bibliométrica	Número de artigos por país, por disciplina, etc.
• Revista por especialista	Primeira, segunda, terceira e quarta classe de revistas
• Indicador de produção	Número de artigos, livros etc. dividido pelo número de autores
• Índice de atividade	Porcentagem de publicação por X, porcentagem de publicação por Y
• Crescimento da documentação	Taxa de crescimento
• Distribuição de Bradford	Revistas principais
• Distribuição de Lotka	Produtividade dos autores
• Obsolescência	Vida media da publicação
• Distribuição de Zipf	Uso de vocabulário
• Distribuição de Waring	Potencial de publicação
<b>Indicadores de citação</b>	
• Índice de afinidade	• Índice de imediatez
• Índice de atividade	• Fator de impacto
• Relação bibliográfica	• Índice de impacto
• Análise de citação	• Índice de isolamento
• Análise de co-citação	• Índice de abertura
• Fator de consumo	• Fator de popularidade
• Índice de diversidade	• Índice de autocitação
• Fator de eco	
<b>Técnicas de mapeamento</b>	
• Análise de co-ocorrência de palavras	
• Rede de citações	
• Análise de colaboração científica	

Fonte: Spinak (1998)

Os indicadores podem representar, por exemplo, a produção científica de um país através do número de artigos científicos publicados por pesquisadores de certa nacionalidade em periódicos indexados, ou em outro exemplo, representar o percentual de determinada população que considera a C&T como prioridade para investimentos de impostos. Nesse sentido, todo enunciado poderia ser considerado uma representação de uma representação, na medida em que, simultaneamente, representa (ou descreve) um estado de coisas no mundo e representa sua própria enunciação, pela forma que o faz (VOGT; MORALES, 2015)

O enunciado típico dos indicadores de C&T é, portanto, segundo Vogt e Morales (2015), a afirmação. E seu discurso é o de conteúdo, ou seja, da descrição do estado das coisas, da descrição da produção científica, da inovação, da P&D de um país. A linguagem dos indicadores de C&T funciona no registro do dizer, do descrever, do explicitar, do contar.

Os indicadores de C&T têm, no entanto, uma propriedade que é característica própria da afirmação; ou seja, são formas de enunciado que tendem a disfarçar a presença de quem enuncia, e tendem a apresentar-se como que enunciados por si próprios, eludindo os traços de subjetividade que caracterizam as marcas de sua enunciação. Dizer algo sem mostrar-se dizendo-o é uma situação particular de discurso na qual, em nome da objetividade e da verdade, se pretende apagar as origens da enunciação, apresentando o enunciado como se dito por ele próprio. Caso, por exemplo, do discurso demonstrativo da ciência (VOGT; MORALES, 2015).

Com relação ao uso adequado das métricas em avaliação de ciência, um artigo publicado na *Nature* por pesquisadores que integram o Comitê Científico da STI 2014 apresentam os dez princípios do Manifesto de Leiden para o uso adequado dessas métricas, de forma que pesquisadores possam responsabilizar os pareceristas e estes, por sua vez, possam confiar nos indicadores utilizados. A seguir, um resumo dos dez princípios:

- 1. Avaliação quantitativa deve apoiar avaliação qualitativa, conduzida por especialistas.** Métricas quantitativas podem desafiar tendências de viés em avaliação por pares e facilitar o julgamento, fortalecendo a avaliação por pares. Os pareceristas não devem, entretanto, ceder à tentação de substituir a tomada de decisão pelos números. Os indicadores não devem substituir a avaliação pelos pares e todos os envolvidos são responsáveis por suas avaliações.
- 2. Medir o desempenho em relação aos objetivos de pesquisa da instituição, grupo ou pesquisador.** Os objetivos do projeto devem ser indicados no início, e os indicadores utilizados para avaliar o desempenho devem se referir claramente a estas metas. A escolha dos indicadores e as formas como são utilizados deverão considerar os contextos socioeconômicos e culturais mais amplos. Um único modelo de avaliação não se aplica a todos os contextos.
- 3. Proteger a excelência em pesquisa relevante localmente.** Em muitos lugares no mundo, excelência em pesquisa significa publicação em idioma inglês. Vieses são particularmente problemáticos nas ciências sociais e humanidades, na qual a pesquisa é mais regional e de centralidade nacional. Muitas outras áreas tem uma dimensão nacional ou regional – por exemplo, estudos epidemiológicos específicos de determinadas regiões. Este pluralismo e relevância social tendem a ser suprimidos quando se criam artigos de interesse dos *gatekeepers* dos periódicos de alto impacto de língua inglesa. Métricas baseadas em literatura de alta qualidade em língua não inglesa serviriam para identificar e reconhecer excelência em pesquisa de caráter local ou regional.
- 4. Os processos de obtenção de dados e processo de análise devem ser abertos, transparentes e simples.** A construção das bases de dados requeridas para avaliação deve seguir regras predeterminadas e claras. A transparência permite o escrutínio dos resultados. A simplicidade é uma qualidade de um indicador, pois aumenta sua transparência. Entretanto, métricas simplistas podem distorcer os resultados.

5. **Permitir aos avaliados checar dados e análises.** Para assegurar a qualidade dos dados, deve ser facultado aos pesquisadores incluídos em estudos bibliométricos checar se seus resultados foram corretamente identificados, ou submetê-los a auditoria independente. Os sistemas de informação das instituições devem implementar esta verificação, ao mesmo tempo que a transparência deve orientar o processo de seleção dos provedores deste serviço. Leve-se em conta que dados precisos e de alta qualidade consomem tempo e recursos financeiros para coletar e processar.

6. **Levar em conta as variações de áreas na publicação e práticas de citações.** A melhor prática é selecionar uma série de possíveis indicadores e facultar às áreas escolher um dentre eles. Determinadas áreas – notadamente ciências sociais e humanidades – publicam preferencialmente livros ao invés de artigos em periódicos, e outras como ciências da computação tem a maior parte de sua produção científica disseminada em conferências. Esta diversidade de formas de publicação de resultados de pesquisa deve ser levada em conta ao avaliar diferentes áreas. Citações também variam por área, o que requer indicadores normalizados; métodos mais robustos de normalização baseiam-se em percentis.

7. **Embasar avaliação de pesquisadores individuais em um de seus próprios sistemas de avaliação.** É sabido que o índice-*h* aumenta com a idade, mesmo sem novos trabalhos publicados. Além disso, ele depende da base de dados onde é calculado, sendo consideravelmente maior no *Google Scholar* que no *Web of Science*. Ler sobre o trabalho de um pesquisador é muito mais adequado que confiar em um número. Mesmo quando se compara um grande número de pesquisadores, uma abordagem que considere mais informação acerca de sua especialidade, experiência, atividade e influência é sempre melhor.

8. **Evitar objetividade mal colocada e falsa precisão.** Indicadores de ciência e tecnologia tendem à ambiguidade conceitual e incerteza e requerem fortes premissas que não são universalmente aceitas. Assim, boas práticas indicam o uso de indicadores múltiplos para fornecer um quadro mais consistente e próximo do real.

9. **Reconhecer o efeito sistêmico da avaliação e dos indicadores.** Indicadores podem alterar o sistema através dos incentivos que estabelecem, e estes efeitos podem ser antecipados. Isso significa que é sempre preferível usar um conjunto de indicadores. O uso de um único indicador – como número de publicações ou número total de citações – pode levar a erros de interpretação.

10. **Submeter os indicadores a um escrutínio e atualizá-los regularmente.** A missão da pesquisa e os objetivos da avaliação mudam e o sistema de pesquisa também evolui. Quando métricas usuais tornam-se inadequadas, novas emergem. Assim, indicadores devem ser revisados e por vezes modificados (HICKS, 2015, p. 1-2).

A forma como se apresenta os dados é um outro ponto importante quando se estuda e utiliza indicadores. Dessa forma, como exemplo de algumas representações visuais, tem-se: rede de relações, gráfico de dispersão, infográficos, etc. Nas quais é possível ampliar a cognição e ajudar as pessoas a raciocinar sobre a informação apresentada. Além desses exemplos, existe um enorme conjunto de tipos de visualizações possíveis para representar os mais variados conjuntos de dados como afirmam (TAVARES; PIMENTEL; ARAUJO, 2012).

A visualização de informação é compreendida como processos de produção de representações visuais de dados, que visa melhorar a capacidade de realizar uma tarefa através da codificação de informações, muitas vezes altamente abstrata, em uma forma visual. A visualização pode ser estática ou dinâmica e interativa e, através de uma variedade de meios de comunicação - por exemplo, jornal, cartaz, *website* ou *software* (MCINERNEY et. al., 2014, p. 149).

Há os indicadores que tratam especificamente de grafos de redes. Uma rede social é compreendida como um conjunto de dois elementos: atores (pessoas, instituições ou grupos – os nós da rede) e suas conexões (interações ou laços sociais). Redes são metáforas para examinar os padrões de relação de um grupo social, a partir das conexões estabelecidas entre os atores. O foco da rede está na estrutura social, na qual não é possível segregar os atores sociais e nem suas conexões (RECUERO, 2009). Entende-se que “os atores são o primeiro elemento da rede social, representados pelos nós. Trata-se das pessoas envolvidas na rede que se analisa. Como partes do sistema, os atores atuam de forma a moldar as estruturas sociais, através da interação e da constituição de laços sociais” (RECUERO, 2009, p.25).

No passado as redes sociais na ciência eram descritas como campo científico (BOURDIEU, 1983); redes sócio técnicas (LATOUR; WOOLGAR, 1979); arenas transepistêmicas (KNORR-CETINA, 1981); mundo da ciência (NUNES, 1999) entre outras, atualmente apenas redes.

Os atores representam os nós da rede em questão, já as relações de uma rede social podem ser observadas de muitas maneiras. Em síntese, as conexões de uma rede social são constituídas dos laços sociais, que são formados através da interação social entre os atores. Desta maneira pode-se compreender que, são as conexões o principal foco do estudo das redes sociais, uma vez que é a sua variação que altera as estruturas desses grupos (RECUERO, 2009).

### **2.3 Política científica e tecnológica**

Em meados do século XX, Vannevar Bush (BUSH, 1945) elaborou um relatório dirigido ao então presidente dos Estados Unidos, constatando uma aproximação cada vez maior entre ciência e tecnologia e entre elas e a indústria. O documento marcou uma das mais fortes correntes interpretativas sobre a relação entre ciência, tecnologia e desenvolvimento.

Essa abordagem assume que ciência produz conhecimento, o que, automaticamente, leva à produção tecnológica, a qual, por sua vez, produz inovação responsável por melhorar a

competitividade das empresas. Tal perspectiva originou também o debate sobre a ação determinante da oferta ou da demanda no desenvolvimento (é a oferta de novo conhecimento científico que gera e impulsiona novas tecnologias ou é a demanda da indústria por novas tecnologias que estimula a produção de novos conhecimentos), uma discussão economicista que tem sérias implicações nas políticas de ciência e tecnologia (BAUMGARTEN; SANTOS DE LIMA, 2013; MACIEL, 2002).

Nesse sentido, a produção do conhecimento pode ser relativamente simplificada entre uma transição da ciência Modo 1 à ciência Modo 2. Enquanto o Modo 1 refere-se à produção tradicional/disciplinar de conhecimento que é impulsionada pela ciência, o Modo 2 abrange processos basicamente estimulados e influenciado pela demanda e que dependem não apenas de cientistas, mas de um grande número de outros atores que desempenham papéis importantes e reconhecidos (GIBBONS, 1996).

Ao longo dessas sete décadas, mudanças na concepção de como se dá a produção de conhecimento científico e tecnológico, assim como sobre a relação entre ciência e tecnologia têm levado a alterações no desenho da política científica em grande número de países. Junto a mudanças de paradigmas na concepção de ciência e tecnologia (C&T), também ocorreram mudanças nas formas de organização das atividades científicas e tecnológicas, nas instituições que produzem conhecimento, assim como nos instrumentos, processos e procedimentos que têm por objetivo planejar e gerir as atividades relacionadas à C&T. Isso tem gerado um grande grau de incerteza sobre os caminhos mais adequados para planejar, conduzir e avaliar as atividades científicas e tecnológicas (CGEE, 2008).

Do mesmo modo, dada a centralidade do conhecimento científico e tecnológico tanto para a reprodução, quanto para a transformação da relação entre produção, acumulação e distribuição de conhecimento – indissociável da relação entre inclusão social, econômica e política –, a informação sobre ciência, a divulgação científica e o incentivo ao debate informado sobre o tema são condições imprescindíveis para o surgimento de inovações, tendo-se como base sua perspectiva ampliada, o que passa, necessariamente, pela capacidade de integração de saberes (BAUMGARTEN; LIMA, 2013; MACIEL, 2005).

No Brasil, o conjunto de ações federais para a área de C&T é reconhecido como Política Científica e Tecnológica (PCT), ou seja, admite-se a oportunidade de tratá-las juntas, uma vez que a ciência tem se tornado cada vez mais tecnológica e a tecnologia cada vez mais científica. Outra dicotomia pertinente é a que ocorre entre a agenda da ciência e a agenda da

sociedade, de um lado está a comunidade científica e do outro, uma diversidade de atores e interesses, resultando no que se apresenta como Política Científica e Tecnológica.

Como em outros países, no Brasil, somente nos últimos cinquenta anos a “Política Científica e Tecnológica tem sido reconhecida institucionalmente com a criação e manutenção de estruturas formais, mecanismos, procedimentos, apoio burocrático e político especialmente dirigidos às questões a ela pertinentes” (MOREIRA e VELHO, 2008).

Para Dias (2011) há um consenso comum em torno da neutralidade e do determinismo da ciência e da tecnologia, que coincide com uma visão triunfalista e essencialista da C&T, cujos aspectos ideológicos e políticos intrínsecos à Política Científica e Tecnológica são ocultados. Desse modo, as reflexões teóricas comumente realizadas a respeito da política científica e tecnológica tendem a ignorar essas questões, tal fato diferencia a Política Pública em C&T das demais.

Os estudos acerca da política científica e tecnológica nos países latino americanos, incluindo o Brasil, têm sido cada vez mais orientados por conceitos e métodos oriundos dos países ditos “desenvolvidos” (DIAS, 2011). Em busca de uma definição para o que seja Política Pública em âmbito geral, Souza (2006) elenca múltiplas definições uma vez que:

Não existe uma única, nem melhor, definição sobre o que seja política pública. Mead (1995) a define como um campo dentro do estudo da política que analisa o governo à luz de grandes questões públicas e Lynn (1980), como um conjunto de ações do governo que irão produzir efeitos específicos. Peters (1986) segue o mesmo veio: política pública é a soma das atividades dos governos, que agem diretamente ou através de delegação, e que influenciam a vida dos cidadãos. Dye (1984) sintetiza a definição de política pública como “o que o governo escolhe fazer ou não fazer”. A definição mais conhecida continua sendo a de Laswell, ou seja, decisões e análises sobre política pública implicam responder às seguintes questões: quem ganha o quê, por que e que diferença faz (p. 24).

No período de 1950 a início da década de 1960 há a institucionalização da política científica, porém, as ações são datadas de forma descontínua. Inicialmente foram criados institutos que adotaram padrões internacionais e se aperfeiçoaram gradativamente. Era preciso, então, a constituição de um órgão coordenador das atividades científicas e tecnológicas. Foi quando surgiu o CNPq e, em seguida, a CAPES, que passaram a exercer o papel de órgãos facilitadores (MOREL, 1979).

Neste sentido, pode-se dizer que a institucionalização da PCT no Brasil foi marcada pela criação da CAPES e do CNPq. Um processo que se tornou viável apenas ao passo em que a comunidade de pesquisa passou a se articular e a advogar pela criação de instrumentos e

de instituições que pudessem alavancar o avanço científico e tecnológico nacional. Nesse sentido, a sucessão de agendas de diferentes governos configura o “mapa” do Estado, que recentemente tem sido fortemente influenciadas pela Economia da Inovação (DIAS, 2011).

As relações entre Estado e cientistas tiveram, sempre, como espaço privilegiado, as agências de fomento (CNPq e Capes), as quais se caracterizam por longa história de interação com a coletividade científica através dos canais de representação destas, que se constituem nessas instituições – formalmente – segundo critérios embasados na competência técnico-científica. Dessa forma, planejar e implementar as políticas de C&T vem sendo uma atividade compartilhada e, por que não dizer, dirigida pelos próprios cientistas (BAUMGARTEN, 2004).

A evolução dos sistemas de inovação, e o atual conflito sobre qual caminho a ser tomado nas relações universidade-indústria, estão refletidos nos diferentes arranjos institucionais da universidade-indústria e das relações com o governo. Em primeiro lugar, pode-se distinguir uma situação histórica específica que se pode pretender rotular hélice tripla. Nesta configuração o Estado abrange a academia e a indústria e dirige as relações entre eles. Um segundo modelo consiste em separar esferas institucionais com fronteiras fortes, dividindo e mostrando as relações altamente circunscritas entre as esferas. Finalmente, a hélice tripla gera uma infraestrutura de conhecimento em termos de sobreposição, um modelo de intersecção das relações universidade-indústria-governo (ETZKOWITZ; LEYDESDORFF, 2000).

Nas esferas institucionais, cada uma tem o papel de outras e como organizações híbridas os arranjos de hélice tripla geram atualmente interesse normativo. Os autores salientam que quando o conhecimento é cada vez mais utilizado como um recurso para o sistema de produção e distribuição, sua reconstrução pode vir a prevalecer como um modo de "destruição criativa". A hélice tripla denota não só a relação da universidade, indústria e governo, mas também transformação interna dentro de cada uma dessas esferas. A universidade foi transformada a partir de uma instituição de ensino combinada com pesquisa, uma revolução que ainda está em curso, não só nos EUA, mas em muitos outros países (ETZKOWITZ; LEYDESDORFF, 2000).

Costa, Sousa & Mazocco (2010) afirmam que os estudos em Ciência Tecnologia e Sociedade teriam entre seus objetivos, incentivar a alfabetização (divulgação) científica, revelando a ciência como uma atividade de grande importância social. E entre seus públicos, estariam os jornalistas científicos e os receptores da comunicação da ciência por meio da

mídia. A comunicação pública da ciência vem sendo objeto de estudo, devido a sua importância no processo de construção de uma política mais democrática em Ciência e Tecnologia, assegurando a participação da sociedade nessas discussões (COSTA, SOUSA & MAZOCCO, 2010).

O modelo de déficit é descrito como o mais utilizado, principalmente nos países em desenvolvimento, caso do Brasil. Nele a ciência é vista como autônoma em relação ao resto da sociedade e o público é visto como uma massa homogênea e passiva, cuja comunicação por sua vez é unidirecional e linear. Esse modelo assegura a visão de uma ciência autônoma, objetiva e que representa tão somente a “verdade”. Já o modelo de participação pública, modelo da tendência dialógica da comunicação, permite que cientistas e público ocupem o mesmo nível na estrutura, a decisão conjunta nas políticas pública de Ciência e Tecnologia, sobretudo, o diálogo entre Ciência, Tecnologia e Sociedade (CTS). Essas diretrizes estão no escopo dos estudos em CTS, que reconhecem a complexidade da relação entre ciência, tecnologia e sociedade, superando a clássica relação linear entre elas (COSTA, SOUSA & MAZOCCO, 2010).

Dessa forma, nota-se uma relação estreita entre os objetivos e os pressupostos do modelo de participação pública da comunicação pública da ciência e os estudos em CTS. Sendo possível concluir que essa relação deve promover uma maior participação da sociedade nos assuntos relacionados à ciência e à tecnologia, fazendo com que a integração entre elas, tão necessária, seja, de fato, concreta e mais democrática.

Guivant e Macnaghten (2011), reconhecem que há um crescente consenso na literatura de Estudos Sociais da Ciência e da Tecnologia a respeito da importância e da necessidade de participação pública na governança de tecnologias emergentes e controversas. Os autores analisaram três iniciativas relativamente modestas de pesquisas conduzidas no Brasil em relação à percepção de riscos, benefícios e visões da nanotecnologia. Eles não tiveram a pretensão de criar generalizações a partir destes projetos de pesquisa para a situação do Brasil como um todo. Contudo, elas permitiram levantar questões a respeito do que pode acontecer em um contexto no qual nem os cientistas, nem os políticos e muito menos o público estão interessados no debate em torno dos riscos e benefícios das tecnologias emergentes. Para os autores, há necessidade de uma conceitualização diferente de governança tecnológica e sua relação com a globalização, que:

...em vez de meramente contrastar países ocidentais e orientais, leve em conta as especificidades das dinâmicas locais e regionais quando confrontadas pelos mesmos problemas, com alianças específicas e mistas

entre grupos de peritos e leigos, e as articulações entre atores representando várias formas de subpolítica e aqueles representando formas convencionais de política (GUIVANT & MACNAGHTEN, 2011, p. 100).

A respeito disso, é crucial definir mais precisamente o que é entendido por subpolítica e por alternativas à sociedade de risco, evitando abordagens idealizadas da posição de não peritos, e abordagens simplistas de estimativas quantitativas de risco que não identificam algumas transformações que acontecem em relação a tecnologias emergentes, uma vez que a governança tecnológica deve ser analisada em relação à cultura política de um Estado-nação.

### **2.3.1 A nanotecnologia, uma área prioritária?**

Novas intervenções médico-sanitárias serão, provavelmente, cada vez mais demandarão a interseção de três novos campos de saber, que não são derivados apenas das biociências, mas nos quais elas têm uma grande importância: a biotecnologia, a nanotecnologia e as tecnologias de informação (GUIMARÃES, SOUZA e SANTOS, 2012). A nanotecnologia é prioritária, pois é uma área de ponta que pode propiciar o desenvolvimento de diversas indústrias, tornando-as mais competitivas.

A nanotecnologia e a nanociência constituem um campo em ascensão nos últimos anos, campo multidisciplinar do conhecimento científico e tecnológico, que tem como propósito a investigação e a manipulação da matéria na escala de um a cem nanômetros (RATHER; RATHER, 2002).

De acordo com Lisboa Filho e Monteiro (2013) a nanociência e a nanotecnologia consagram-se como uma grande conquista científico-tecnológica do século XXI. Uma vez que ao se manipular a matéria a partir da dimensão nano, acentua-se a relação entre a área superficial e o volume, intensificando algumas propriedades físicas, químicas e mecânicas da matéria, da mesma maneira propicia o aparecimento de outras.

A ciência dos materiais e a nanobiotecnologia podem se beneficiar com as aplicações da nanociência e da nanotecnologia e conseqüentemente uma vasta gama de setores. Tudo isso, encontra-se associado a um fator de lucratividade impactante na economia. O otimismo creditado a nanociência e a nanotecnologia são evidenciados através das políticas desenvolvidas através de programas governamentais, da dotação orçamentária prevista e no quantitativo de pesquisas desenvolvidas nas últimas décadas. Pelo intenso impacto social,

cultural, econômico e ambiental causado pela nanotecnologia, entende-se que se trata de uma revolução tecnocientífica (OLIVÉ, 2009).

Nos países industrializados, a nanotecnologia é um dos focos prioritários das atividades de pesquisa, desenvolvimento e inovação, resultando no contínuo aumento de produtos industrializados. Segundo estimativas da *The Royal Society & The Royal Academy of Engineering*, no período de 2010 a 2015, foi investido um montante na ordem de 5 bilhões de euros no mercado mundial para produtos e processos industriais baseado na nanotecnologia. Desse montante, 2 bilhões de euros vêm da iniciativa privada. Seguindo os rumos dessa tendência, latinoamericanos incluíram a nanotecnologia como área estratégica para fomentar a competitividade (FALADORI, 2012). Das 100.000 patentes existentes no mundo sobre a nanotecnologia, apenas 333 encontram-se na América Latina. Dessas, 33% foram concedidas ao Brasil e ao México. Esse quantitativo e a própria distribuição revelam um descompasso do desenvolvimento da nanotecnologia em relação a algumas localidades (PASTRANA; ÁVILA; MORENO, 2012).

Martins alerta que, devido às pesquisas em nanotecnologia, bem como os processos de produção, além de muito complexos, requererem altos investimentos financeiros. Essas condições inviabilizam a condução dos mencionados processos por pequenos grupos que não detém investimentos financeiros suficientes. Como consequência, a nanotecnologia torna-se susceptível de ser monopolizada por restritas corporações, as quais dispõem de financiamento para o empreendimento (MARTINS, 2006). Essa possibilidade já sinaliza, com justificativas, para a concentração das patentes em alguns países. Ou seja, a quem beneficia o desenvolvimento da nanotecnologia? Haja vista que não gera bem-estar social proporcional ao investimento público.

#### *2.3.1.1 A Nanotecnologia no cenário mundial*

Dados recentes da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) e empresas de consultoria indicam que o mercado de produtos nanotecnológicos movimentava cerca de US\$ 350 bilhões e, em 2020, estima-se que esse valor será superior a US\$ 3 trilhões.

A nanotecnologia, ainda que haja discussão sobre sua definição, pode ser compreendida como o entendimento, o controle, o processamento e a caracterização da matéria em dimensões típicas de 0,1 a 100 nanômetros (nm) (ISO, 2013). A nanotecnologia

tem se tornado cada vez mais relevante no contexto de ciência e engenharia, com um grande potencial para promoção de inovações de impacto para a sociedade (ABDI, 2010; KOSTOFF, KOYTCHOFF e LAU, 2007; MILANEZ, 2011). Os nanomateriais são a base da nanotecnologia e caracterizam-se por apresentar estrutura elementar natural ou engenheirada em escala nanométrica, exibindo comportamentos e propriedades melhorados ou totalmente novos, propiciando a criação de novos dispositivos e materiais (NANOWERK, 2013).

Apesar do grande potencial econômico e da disponibilidade de aplicações comerciais específicas, o desenvolvimento da nanotecnologia e dos nanomateriais é considerado ainda emergente, sujeitos a mudanças técnicas e de regulamentação (ABDI, 2010; NANOWERK, 2013; SALERNO; LANDONI; VERGANTI, 2008). A complexidade e a dinâmica do ambiente tecnológico e de negócios em nanotecnologia e nanomateriais, associadas à crescente globalização da economia mundial e às rápidas mudanças tecnológicas e regulamentares, tornam recomendável atividades de monitoramento, conforme proposto, por exemplo, por Gilad (2003) e por Martino (1993). O intuito é evitar surpresas que comprometam investimentos no desenvolvimento de novos produtos, materiais e processos que podem em casos extremos afetar a sobrevivência de uma organização ou de um Programa Governamental de desenvolvimento na área.

### *2.3.1.2 A nanotecnologia no Brasil*

Estudo realizado por Invernizzi, Korbes e Fuck (2012) analisou dez anos (2000-2010) após as primeiras ações do desenvolvimento inicial da nanotecnologia no Brasil, que se destacou como líder latino-americano na área, em termos de produção científica, recursos humanos, infra-estrutura de investigação, financiamento da investigação, bem como iniciativas na indústria. No final do ano 2000, o MCTI (na época) e o CNPq organizaram o *Workshop* "Tendências em Nanociência e Nanotecnologia". Compareceram 32 pesquisadores de diferentes áreas das ciências físicas, naturais e engenharia, na ocasião chegou-se a um consenso sobre a necessidade de lançar um programa para estimular essa área emergente. Surgindo um grupo de trabalho com a incumbência de mapear a experiência brasileira em nanotecnologia e desenvolver uma agenda (CNPq News, 2000). Este foi o primeiro passo para a formação de uma política brasileira em nanotecnologia. Em 2003 foi submetido a consulta pública um projeto para o desenvolvimento da nanociência e nanotecnologia, este foi o documento base para a inclusão da área no Plurianual 2004-2007 (MCTI, 2004a). O programa

estabeleceu como objetivo central o desenvolvimento de novos produtos baseados em nanotecnologia para aumentar a competitividade dos processos da indústria doméstica.

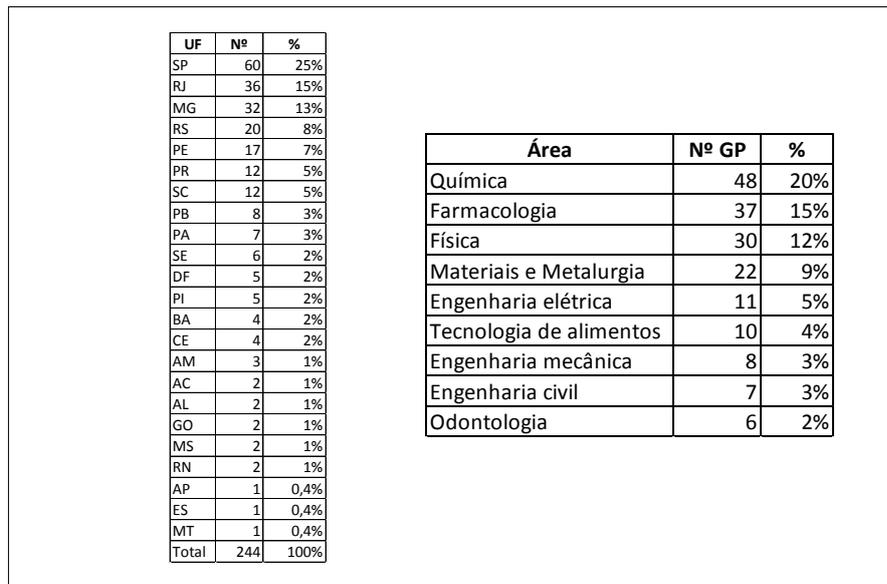
Em 2005 o Governo Federal lança um programa mais amplo, o Plano Nacional de Nanotecnologia (PNN) (MCTI, 2005). O PNN já nasce alinhado com outra política fundamental do então governo: a Política Industrial, Tecnológica e de Comércio Exterior (PITCE), fundada em 2004, que definiu a nanotecnologia como "área portadora de futuro". Por meio de suas agências, o MCTI, investiu desde a formulação do primeiro programa de nanotecnologia em 2004, até 2009, um valor próximo de 314 milhões (MCTI 2008; EMBRAPA NEWS, 2009). Entretanto, não há um número oficial que apresente o total das despesas incorridas no Brasil em nanotecnologia, incorporando todas as ações, programas e fontes.

De acordo com o MCTI, há pelo menos 48 universidades ou laboratórios de pesquisa que têm a nanotecnologia como fonte de investigação ativa no país (BAIBICH, 2010). A área da nanotecnologia foi integrado acordos de cooperação em C&T que o país tem com a União Europeia desde 2004. O Brasil mantém cooperação ou está a negociar colaborações em nanotecnologia com vários países, como Japão, China, França, Alemanha, Espanha, Portugal, Estados Unidos, Rússia, Coreia do Sul, Irã, Índia e África do Sul (este último no âmbito da iniciativa da Nanotecnologia dois Índia-Brasil-África do Sul (IBAS)). O Brasil é o país latino-americano que mais investe em P&D em relação ao PIB. No período 2002-2008, a despesa interna bruta em P&D aumentou de 0,98 por cento a 1,09 por cento do PIB. Em 2008, o setor público foi responsável por 55 por cento da despesa (0,59 por cento do PIB), o que representa uma proporção de 15 por cento inferior à média dos países da OCDE, enquanto a despesa bruta em P&D no setor privado atingiu apenas 32 por cento da média dos gastos da iniciativa privada por parte dos países da OCDE (BRITO CRUZ; CHAIMOVICH, 2010).

Atualmente no Brasil há 26 redes temáticas e 16 INCT focados em nanotecnologia com mais de 2.500 pesquisadores e mais de 3.000 estudantes (MCTI, 2014). Todavia, para que a nanotecnologia venha a ser uma nova tecnologia básica, capaz de estimular futuros processos de inovação e impulsionar novas gerações tecnológicas, uma das precondições é que ela interaja de forma transdisciplinar com a eletrônica, a tecnologia de informação, o estudo de materiais, a ótica, a bioquímica, a biotecnologia, a medicina e a micro-mecânica. Em consequência disso, as aplicações da nanotecnologia estendem-se às áreas de materiais sob encomenda e sistemas biotécnicos, apesar da tendência de associá-las com a eletrônica (BONACCORSI, 2008).

Atualmente os grupos de pesquisa atuantes no Brasil são 244 e cerca de 78% deles estão localizados nas regiões Sul, Sudeste e Nordeste. As áreas em que eles cadastrados são as mais diversas, o que revela a multidisciplinaridade da área, contudo nota-se uma forte concentração nas áreas de Química, Farmacologia, Física, Materiais e Metalurgia, como apresentado na Figura 2 abaixo.

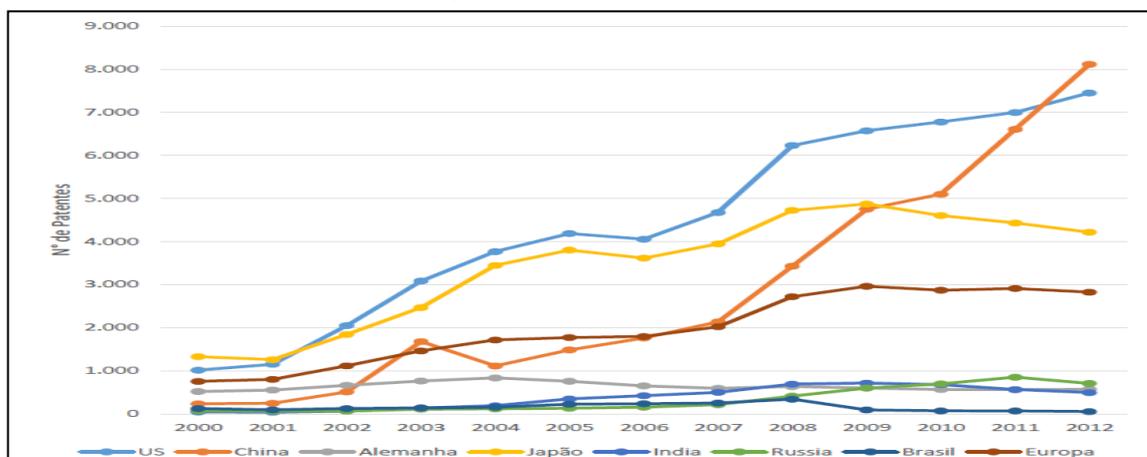
Figura 2 – Composição dos grupos de pesquisa em nanotecnologia



Fonte: MCTI (2015)

O Gráfico 1 apresenta a evolução do número de pedidos de patentes no período de 2000 a 2014 relacionados a área de nanotecnologia. Observa-se queda a partir de 2008 no crescimento do patenteamento referente ao Brasil.

Gráfico 1 – Crescimento do patenteamento em nanotecnologia entre 2000 e 2014

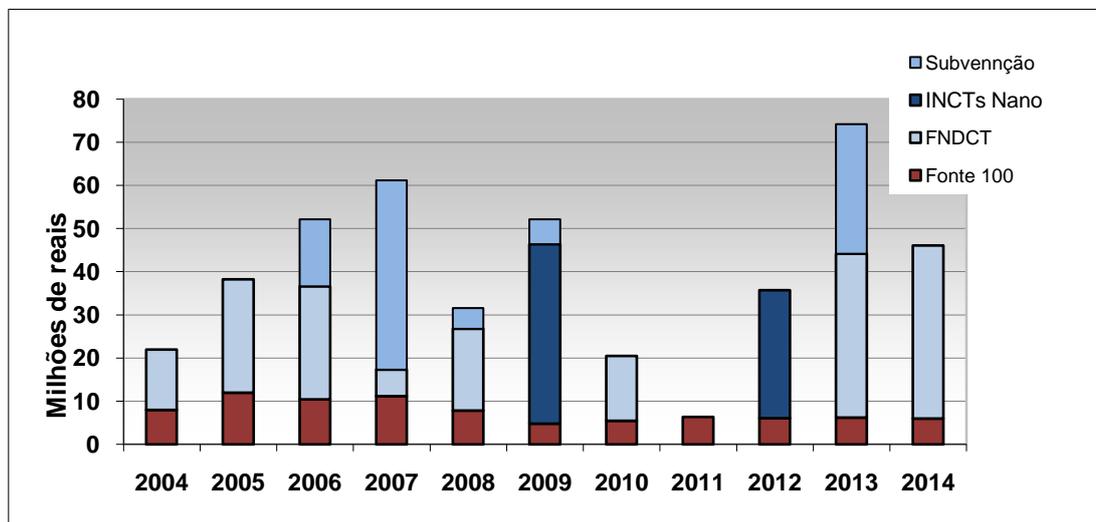


Fonte: Panorama da Nanotecnologia no Brasil UNESCO – (MCTI, 2015)

No Brasil lançou-se em 2013 a Iniciativa Brasileira de Nanotecnologia (IBN). Os fundamentos desta estratégia apresentaram duas frentes: os Institutos Nacional de Ciência, Tecnologia e Inovação (INCT) e o plano “Brasil Maior: Inovar para competir, competir para crescer”. As áreas prioritárias da IBN são: Aeroespacial e Defesa; Agronegócio; Energia; Saúde e Meio Ambiente.

O Gráfico 2 apresenta os investimentos realizados no Brasil em milhões de reais. Em 2009 e 2012 os maiores investimentos (71,30 milhões) foram os relacionados aos incentivos de criação dos INCTs. O ano em que houve o maior investimento foi 2013 com 74,12 milhões. De 2004 a 2014 o total de investimentos somam 440,16 milhões de reais.

Gráfico 2 – Investimentos em nanotecnologia no Brasil de 2004 a 2014



Fonte: MCTI (2015)

## 2.4 Acesso aberto à Informação, Científica e Tecnológica

A Informação em Ciência e Tecnologia (ICT) é constituída de elementos simbólicos utilizados para comunicar o conhecimento científico e técnico, independente de seu caráter (numérico, textual icônico, etc.), dos suportes materiais, da forma de apresentação. Refere-se tanto à substância ou conteúdo dos documentos quanto à sua existência material. Também se emprega este termo ICT para se designar tanto a mensagem (conteúdo e forma) quanto sua comunicação (ação). Quando necessário, distingue-se entre informação bruta (fatos, conceito, representações) e os documentos em que se acha registrada (AGUIAR, 1991).

Para que fosse possível comparar os desempenhos dos desenvolvimentos entre países foi necessário coletar dados e informações para este fim, desde a década de 1960 surgem organizações mundiais com este objetivo.

A Organização das Nações Unidas para a Educação, Ciência e Tecnologia (Unesco) desde 1966 tem-se dedicado à coleta de dados estatísticos em ciência e tecnologia, os quais publica em seu anuário estatístico a partir de 1969. Igualmente a Comunidade Econômica Européia (CEE), através do Comitê para Pesquisa Científica e Técnica (Crest), publica dados sobre os orçamentos nacionais classificados por "objetivos", utilizando a nomenclatura para a análise e comparação dos programas e orçamentos para a ciência. O Conselho Escandinavo para a Pesquisa Aplicada trabalha, desde 1968, na coleta e tratamento de dados sobre as atividades de ciência e tecnologia na região. Essas três últimas instituições subsidiaram o trabalho da Organização para a Cooperação e o Desenvolvimento Econômico (OCDE) para a proposição de um sistema padrão para avaliação da Pesquisa e Desenvolvimento Experimental; dele resultou a publicação Medição de Atividades Científicas e Tecnológicas - Manual Frascati, que o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) traduziu e editou em 1979 (AGUIAR, 1991, p. 7-8).

O termo Informação em Ciência e Tecnologia (ICT) é empregado para englobar as informações que, além de cumprirem as funções relacionadas como específicas da informação científica ou da informação tecnológica, servem ainda para cumprir e apoiar a atividade de planejamento e gestão em ciência e tecnologia: avaliar o resultado do esforço aplicado em atividades científicas e tecnológicas e subsidiar a formulação de políticas, diretrizes, planos e programas de desenvolvimento científico e tecnológico (AGUIAR, 1991, p. 12). Já a Informação Tecnológica é todo tipo de conhecimento sobre tecnologias de fabricação, de projeto e de gestão, que favoreça a melhoria contínua da qualidade e a inovação no setor produtivo (CID/UNB, 2004).

De acordo com Dowbor (2003) a informação é um recurso valioso, e um poderoso racionalizador das atividades sociais. Valiosa também é a limitada capacidade de atenção humana, uma vez que a informação bem organizada e disseminada constitui um elemento essencial da democracia participativa, ao facilitar as opções racionais dos diversos atores sociais. Mas não substitui a iniciativa do Estado e o planejamento estratégico.

De acordo com o detalhamento apresentado por Aguiar (1991) as informações que cumprem a função de subsidiar os desenvolvimentos em Ciência e Tecnologia (C&T) podem ser agrupadas como se apresenta a seguir:

- a) Caracterização da oferta em ciência e tecnologia com informação sobre:
- ✓ entidades atuantes em ciência e tecnologia (áreas de atuação; recursos humanos e laboratoriais; serviços técnico-científicos realizados; atividades de ensino e de pesquisa; publicações regulares etc.);
  - ✓ pesquisas correntes (objetivos; metodologia; equipes e recursos envolvidos; fontes de financiamento etc.);
  - ✓ pesquisas realizadas (objetivos; resultados alcançados; aplicações; possibilidade de transferência de resultados etc.);
  - ✓ pesquisadores (áreas de atuação, de especialização e de formação acadêmica básica; titulação; trabalhos publicados etc.);
  - ✓ relatórios de pesquisa e trabalhos publicados como resultado de pesquisa (título; resumo; equipe executora ou autor/autores; instituição detentora; possibilidades de acesso ao documento etc.) (AGUIAR, 1991, p. 12).
- b) Caracterização da demanda em ciência e tecnologia com informações relacionadas:
- ✓ identificação de problemas locais, regionais, universais cuja solução poderia ser buscada via atividades de pesquisa científica e tecnológica;
  - ✓ planos regionais ou setoriais, geralmente estabelecidos por autoridades governamentais, de natureza econômica, social ou industrial, fixando objetivos e metas de desenvolvimento para cuja consecução a contribuição das atividades científicas e tecnológicas se faz necessária;
  - ✓ perfis tecnológicos de setores produtivos (agropecuário, indústria e setor de serviços), para apontar eventuais carências de desenvolvimento tecnológico que poderiam ser supridas com atividades de pesquisa ou de prestação de serviços técnico-científicos (ensaios, análises e testes; metrologia; normalização; informação científica e tecnológica etc) (AGUIAR, 1991, p. 12).
- c) Estabelecimento de indicadores de desenvolvimento científico e tecnológico com informações representadas por:
- ✓ estatísticas de ciência e tecnologia (número e tipologia de pesquisadores; pesquisadores por área do conhecimento ou por região; recursos aplicados em ciência e tecnologia; cursos de pós-graduação etc.). Análises estatísticas por área do

conhecimento ou por região (distribuição geográfica ou setorial da oferta de ciência e tecnologia);

- ✓ indicadores internacionais (estatísticas internacionais em ciência e tecnologia para fins de comparação entre nações). Análises comparativas, por área de conhecimento ou por região, entre estatísticas nacionais (AGUIAR, 1991, p. 12).

d) Estudos especiais - trabalhos conduzidos com o objetivo de se avaliar:

- ✓ o impacto econômico, de caráter global, correspondente a diferentes níveis de desenvolvimento tecnológico, por setor industrial ou na agropecuária;
- ✓ o impacto social decorrente da mudança de patamares de desenvolvimento tecnológico, por setor ou por região, na renda per capita, na distribuição de renda, no nível de emprego;
- ✓ o impacto ambiental advindo da difusão de tecnologias eventualmente comprometedoras da qualidade do ar, da água e dos solos (AGUIAR, 1991, p. 12).

Sobre a problemática dos sistemas abertos de informação (sobretudo os de ICT), entende-se que somente tornar acessível a produção científica não favorece o conjunto maior da comunicação científica. A comunicação científica não é um meio, mas um processo composto por produtores, usuários e recursos que regem esse conjunto. É preciso ter clareza quanto à função de cada novo recurso informacional que será disponibilizado para a comunidade, o que requer uma definição de suas finalidades na fase de planejamento. Se a função prevista para um determinado sistema for o armazenamento de arquivos eletrônicos, aspectos de organização são secundários. Porém, se houver a expectativa de que seja um SICT dotado de recursos de recuperação da informação e/ou de que sirva como fonte para elaborar indicadores de C&T, é imprescindível se conhecer as diretrizes necessárias próprias à organização da informação para os devidos fins (SILVA & SMIT, 2009, p. 97).

Para lidar com as questões que envolvem a disponibilização e acesso às Informações Científicas e Tecnológicas juntamente com as oportunidades de Acesso Aberto (AA) é mister integrar os avanços da área computacional. Conte e Giardini (2016) ressaltam que nos últimos anos, o estudo dos fenômenos sociais se renovou mediante o seu interesse em Ciências

Sociais Computacionais. Embora este campo não seja novo, já que o primeiro trabalho computacional sobre a evolução dessa cooperação, de *Axelrod*<sup>5</sup>, é do início da década de 80.

De certo modo, a Ciência Social Computacional ressurgiu recentemente, sob a pressão da ciência social quantitativa e da aplicação das análises que envolvem *Big Data* para conjuntos de dados sociais. Entretanto, as autoras reiteram que o *Big Data* não é uma panacéia e o dilúvio de dados que ele fornece levanta mais perguntas do que respostas. Pode-se cogitar que a vivenciada explosão da produção de grandes volumes de dados, juntamente com o desenvolvimento de novas epistemologias, tem e terá consequências de longo alcance, no modo como o conhecimento é produzido, os negócios conduzidos, e a governança promulgada (ANDERSON, 2008; BOLLIER, 2010; FLORIDI, 2012).

IGARASHI et. al (2016) trazem a ideia de que uma ciência dirigida pelos dados é a mãe da ciência, ou seja, é considerada uma ferramenta científica que impulsiona muitos campos dela. Os autores investigam um conceito chamado de três níveis da ciência dirigida pelos dados, que propõe desatar a relação complicada entre muitos campos e diversos métodos. Este conceito se pauta que qualquer problema de análise de dados deve ser discutido em três níveis: teoria computacional, modelagem e algoritmo/ representação<sup>6</sup>.

Neste sentido, o *Big Data* e as novas análises de dados são inovações de rupturas que estão reconfigurando, em muitas instâncias, o modo como a pesquisa científica é conduzida. Há uma necessidade urgente de reflexão ampla e crítica na academia e nas implicações epistemológicas do desdobramento da revolução dos dados, uma tarefa que mal começou a ser abordada, apesar das rápidas mudanças nas práticas de investigação atuais (KITCHIN, 2014).

Para Hey et. al (2009) é possível compreender a ciência dos dados de acordo com quatro paradigmas:

- a) Primeiro – de natureza da ciência experimental surge na forma de empirismo com as descobertas de fenômenos naturais, período pré-Renascença.
- b) Segundo – de natureza da ciência teórica e na forma de modelagem e generalização, período pré-computador;

---

<sup>5</sup> Axelrod, R. *The Evolution of cooperation*. New York, NY: Basic Books, 1984.

<sup>6</sup> Teoria da computação – especifica campos da ciência natural, mensura a ciência. Modelagem – utiliza a física teórica, matemática e estatística. É uma teoria baseada na computação e em sistemas formulados matematicamente. Algoritmo e representação – utiliza estatística, aprendizagem de máquinas, ciência computacional, o algoritmo é desenvolvido para solucionar problemas computacionais oferecidos pela modelagem.

- c) Terceiro – de natureza da ciência computacional e na forma de simulações de fenômenos complexos, período pré-Big Data.
- d) Quarto – de natureza da ciência exploratória na forma de dados intensivos, exploração estatística e mineração de dados, período atual.

Neste ponto, é importante mencionar a *web* profunda, termo utilizado para referir-se ao conjunto de informações disponíveis em bases e *sites* específicos que não estão acessíveis através dos mecanismos de busca tradicionais, como por exemplo, o *Google* (BOUTET; QUONIAM, 2012; BRIN; PAGE, 1998). Entretanto, estima-se que seu volume seja quinhentas vezes maior do que a *web* visível (AGARWAL; DHALL, 2010; LIU; WANG; AGRAWAL, 2012).

Para realizar pesquisas complexas e em diferentes bases de dados da *web* profunda, pesquisadores e profissionais têm desenvolvido ferramentas específicas para a mineração e análise de informações como, por exemplo, agentes inteligentes e mecanismos de *crawler*, *mining* e *scraping* (AGARWAL; DHALL, 2010; FERRARA, 2013; MENA-CHALCO; CESAR-JR, 2009; QUONIAM, 2001; ZHANG, 2013).

E é preciso treinar os cientistas para lidar com o *big data* – conjunto de soluções tecnológicas capaz de lidar com a acumulação contínua de dados pouco estruturados, capturados de diversas fontes e da ordem de petabytes (quatrilhões de bytes) – tanto para realização de projetos científicos, como também para atuarem, eventualmente, em empresas. “O *data scientist* [*cientista capaz de lidar com grandes volumes de dados*] será um requisito imprescindível para o cientista” (ALISSON, 2013, p.01).

A extração de dados de produção científica, identificação de padrões bibliométricos, modelagem e visualização efetiva de redes de interação entre coautores são tópicos relevantes na área de Cientometria. Nos últimos anos, está se dando especial interesse a tais tópicos devido à descoberta de conhecimento que pode ser obtida a partir do tratamento de conjuntos de dados disponíveis nos repositórios de produção científica (por exemplo, banco de dados de produções bibliográficas, de orientação acadêmica, de projetos de pesquisa, e de diretórios de grupos de pesquisa).

#### **2.4.1 Plataforma Lattes**

O CNPq surgiu como Conselho Nacional de Pesquisas e foi criado no início da década de 50, tendo entre suas atribuições “manter relação com instituições nacionais e estrangeiras

para intercâmbio de documentação técnico-científica”(IBICT, 2016). A Plataforma Lattes representa a experiência do CNPq na integração de bases de dados de Currículos, de Grupos de pesquisa, de Instituições e Estatísticas (Painel Lattes) em um único Sistema de Informações. Sua dimensão atual se estende não só às ações de planejamento, gestão e operacionalização de fomento do CNPq, mas também de outras agências de fomento federais e estaduais, das fundações estaduais de apoio à ciência e tecnologia, das instituições de ensino superior e dos institutos de pesquisa. Além disso, se tornou estratégica não só para as atividades de planejamento e gestão, mas também para a formulação das políticas do Ministério de Ciência e Tecnologia e de outros órgãos governamentais da área de Ciência, Tecnologia e Inovação (CNPQ, 2016).

A Plataforma Lattes foi pensada dentro dos moldes da Web 2.0 que como muitos conceitos importantes não tem fronteiras rígidas, mas pelo contrário, um centro gravitacional. Graças a Internet passou a estimular a chamada "arquitetura de participação", algo que não fazia parte da primeira fase dos serviços *online*. Considerada como a segunda geração da Internet, a *Web 2.0* aparece como potencializadora das formas de publicação, compartilhamento e organização das informações *online* (O'REILLY, 2014).

Em linhas gerais, Web 2.0 diz respeito a uma segunda geração de serviços e aplicativos da rede e a recursos, tecnologias e conceitos que permitem um maior grau de interatividade e colaboração na utilização da Internet (BREISSAN, 2007, p.2).

Em resumo, a Plataforma Lattes é uma iniciativa do CNPq em integrar bases de currículos acadêmicos de instituições públicas e privadas em uma única plataforma. Os chamados Currículos Lattes são atualmente considerados um padrão brasileiro de avaliação representando um histórico das atividades científicas, acadêmicas e profissionais de pesquisadores cadastrados, sendo caracterizada pela livre inserção de dados.

Em um contexto em que há uma grande diversidade de fontes de informação, a iniciativa Plataforma Lattes se destaca como um caso de sucesso, pois concentra e integra diversos dados e informações espalhados em ambientes web digitais e analógicos. A estratégia é brilhante, pois ao estimular ou até mesmo exigir o registro da trajetória individual de cada pesquisador, cria-se o mecanismo macro, a completude da base de dados, mesmo com os problemas de confiabilidade ainda presentes. O ponto que chama a atenção é que a somatória de todos esses currículos permite diversas possibilidades de compreender áreas científicas, comportamentos de grupos, análise de redes, internacionalização da ciência, prospecções, trajetórias de cursos/ programas, elaborar estratégias de desenvolvimento das

políticas públicas, estudar a genealogia acadêmica, mensurar gastos públicos com C&T, dentre outras possibilidades.

Para estudos específicos com focos em questões nacionais ou regionais, a PL é sem dúvida a melhor fonte de informação brasileira, pois congrega publicações/ atividades que não estão indexadas nas bases de dados com visibilidade internacional, pois tratam de problemáticas locais que não interessam à comunidade científica, ou aquelas publicadas em periódicos que ainda não atingiram todos os critérios para serem indexados em grandes bases de dados, alguns deles jamais atingirão.

A disponibilização pública dos dados da Plataforma na internet dão maior transparência e mais confiabilidade às atividades de fomento do CNPq e das agências que a utilizam, fortalecem o intercâmbio entre pesquisadores e instituições e é fonte inesgotável de informações para estudos e pesquisas. Na medida em que suas informações são recorrentes e cumulativas, têm também o importante papel de preservar a memória da atividade de pesquisa no país (CNPQ, 2016).

#### *2.4.1.1 A Base de Currículo Lattes*

A intenção de criar um banco de registro dos currículos de pesquisadores brasileiros no CNPq, Base de Currículos Lattes (BCL), ocorreu desde meados dos anos 80, quando foi feita a primeira captação de dados de currículos, ainda em papel, com algumas etapas formalizadas em um sistema informatizado. Ao final dos anos 90, o CNPq desenvolveu um formulário eletrônico que, preenchido pelo pesquisador, deveria ser enviado em disquete para o CNPq, que os carregava na base de dados. Nessa mesma época, o CNPq contratou os grupos universitários Stela, vinculado à Universidade Federal de Santa Catarina, e C.E.S.A.R, da Universidade Federal de Pernambuco, para que, juntamente com profissionais da empresa Multisoft, e técnicos das Superintendências de Informática e Planejamento, desenvolvessem uma única versão de currículo capaz de integrar as já existentes. Assim, em agosto de 1999, o CNPq lançou e padronizou o Currículo Lattes como sendo o formulário de currículo a ser utilizados no âmbito do Ministério da Ciência e Tecnologia e CNPq (CNPQ, 2016a).

Desde então, a Base de Currículo Lattes tornou-se um padrão nacional praticamente compulsório no registro da vida pregressa e atual de estudantes e pesquisadores do país, sendo utilizado pelas principais universidades federais, institutos, centros de pesquisa e fundações de amparo à pesquisa dos estados como instrumento para a avaliação de pesquisadores,

professores e alunos, dessa forma a Plataforma Lattes constitui-se hoje como um grande repositório de dados de pesquisadores de todo o país.

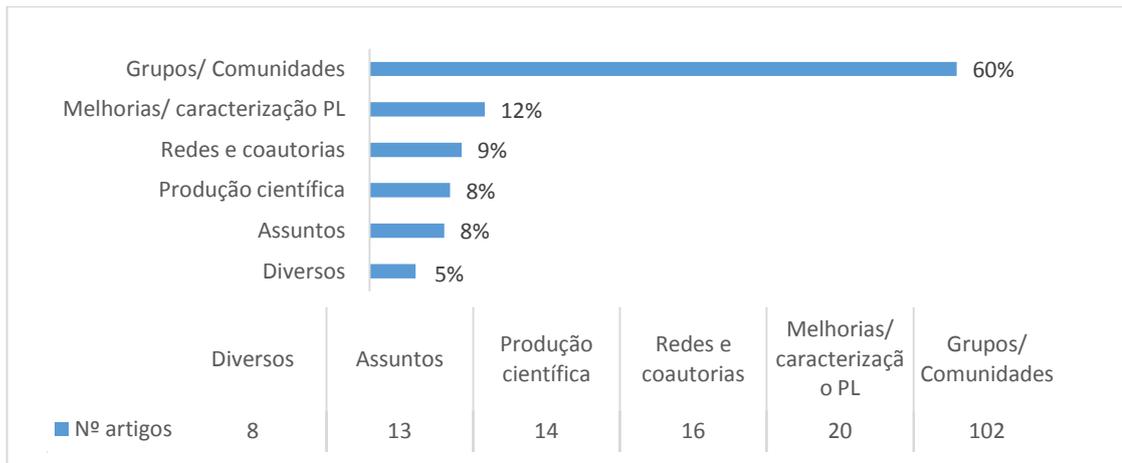
A BCL, de acordo com o CNPq já ultrapassou 3 milhões de currículos, sendo 6,35% de doutores, 10,85% de mestres, 27,69% de graduados, 16,18% de especialistas, 35,54% outros e 3,39% não informado (CNPQ, 2016b).

Uma das funcionalidades da BCL é gerar currículos que se tornam públicos. Esses currículos são documentos que organizam referências a documentos (alguns públicos e outros privados) do arquivo pessoal, ou institucional, dos cientistas. Desta forma, na PL o currículo é um documento que deveria refletir a relação entre os documentos/ atividades dos usuários cadastrados, contudo as informações são distribuídas isoladamente, desconsiderando o contexto na qual foram realizadas (SILVA; SMIT, 2009).

Uma das principais características da BCL é a de que as informações disponíveis estão associadas aos pesquisadores e as instituições que estes se relacionam. A base no todo é um macro dossiê que permite análises em vários níveis: individuais, por grupos, por instituição, por área de formação, por gênero, por área geográfica, por tipo de publicação, por orientações, por projetos financiados, etc. Quando reunidas e organizadas em matrizes, por exemplo, tais possibilidades se multiplicam e revelam as potencialidades das análises permitidas.

Diante desta perspectiva, Brito, Amaral e Quoniam (2016) elaboraram um esquema para compreender as publicações que já se utilizaram da Plataforma Lattes como fonte de informação. Os autores indentificaram seis categorias deste uso nos periódicos analisados. No Gráfico 3 é possível observar que mais da metade das publicações, 60% (102 artigos), estudaram grupos e/ou comunidades específicas, ou seja, partiram de um conjunto de nomes (lista) já identificados e assim realizaram análises dessas atividades registradas na PL.

Gráfico 3 – Distribuição das publicações nas categorias



Fonte: Brito, Amaral e Quoniam (2016)

Neste estudo, Brito, Amaral e Quoniam (2016) afirmam que a PL é uma rica e importante fonte de informação para compreensão de cenários científicos e tecnológicos no Brasil. E que embora tenha sido utilizada como uma fonte de informação para analisar indivíduos ou grupos de indivíduos já conhecidos (60%), há iniciativas que tentam ampliar este uso (8%), realizando o movimento contrário, ou seja, tentam responder quais são os principais pesquisadores de uma determinada área e onde se encontram geograficamente. Entretanto, para este último fim, há grandes dificuldades relacionadas as possibilidades de elaboração de buscas sofisticadas e também para a extração efetiva dos dados.

Sobre o preenchimento dos campos que formam o formulário dos currículos da BCL, eles podem ser classificados em três grupos: autonomia total, autonomia parcial e sem autonomia. Essa especificidade apresenta vieses interessantes, pois quanto mais autonomia for concedida aos usuários, maiores serão as chances das buscas resultarem classes de baixa frequência, resultando em um núcleo reduzido e em alta dispersão. A BCL é baseada num princípio denominado “regras de negócio dos sistemas”, onde cada um dos usuários utiliza e gera a informação que conformará o sistema. No entanto, ao optar por seguir a chamada regra de negócio dos sistemas, o desenvolvimento da BCL priorizou a economia de custos, abrindo mão da sua qualidade (SILVA; SMIT, 2009). O debate sobre “regras de negócio dos sistemas” está relacionado com os novos modelos de serviços da *web*, abertos à participação dos usuários para o compartilhamento de serviços e informações. Na percepção de Catarino e

Baptista (2007), trata-se de um novo paradigma para a organização dos conteúdos de recursos digitais na Web designados, genericamente de folksonomias<sup>7</sup>.

O estudo de Bassoli (2016) exemplifica essas reflexões ao levantar alguns indicadores bibliométricos. Tal estudo trabalhou com o universo de 3.262 currículos de docentes ativos da UFSCar entre 1968 e 2014, conforme a Tabela 1, 30% eram de artigos em periódicos científicos e 70% de artigos em anais de eventos, a coluna “registros” compreende a somatória deles.

Tabela 1 – Artigos publicados encontrados nos Currículos Lattes

Registros	Centros - Universidade Federal de São Carlos	Fora da WoS <sup>8</sup>	WoS <sup>9</sup>	%Fora
11657	Centro de Ciências Exatas e de Tecnologia	3905	6849	33%
5733	Centro de Ciências Biológicas e da Saúde	2420	2724	42%
3970	Centro de Educação e Ciências Humanas	3357	230	85%
1539	Centro de Ciências Agrárias	769	523	50%
930	Centro Ciências Tecnologias para Sustentabilidade	356	503	38%
761	Centro de Ciências Humanas e Biológicas	487	211	64%
430	Centro de Ciências da Natureza	227	197	53%
414	Centro de Ciências em Gestão e Tecnologia	303	89	73%

Fonte: Bassoli (2016)

Na coluna “fora da WoS” está o total de publicações preenchidas nos currículos Lattes, que não estão indexadas na base *Web of Science*, chamando a atenção para o Centro de Educação e Ciências Humanas, que só possui 5,8% do total das publicações indexadas na WoS. Todos os centros possuem publicações não indexadas na WoS, até mesmo centros da

<sup>7</sup> A folksonomia é uma maneira de indexar informações. Esta expressão foi cunhada por Thomas Vander Wal. É uma analogia à taxonomia, mas inclui o prefixo *folks*, palavra da língua inglesa que significa pessoas. O ponto forte da folksonomia é sua construção a partir do linguajar natural da comunidade que a utiliza. Enquanto na taxonomia clássica primeiro são definidas as categorias do índice para depois encaixar as informações em uma delas (e em apenas uma), a folksonomia permite a cada usuário da informação a classificar com uma ou mais palavras-chaves, conhecidas como *tags* (em português, marcadores). Disponível em: <<https://pt.wikipedia.org/wiki/Folksonomia>>. Acesso em: 23 out. 2016.

<sup>8</sup> Algumas publicações não foram consideradas nesta coluna por não terem sido preenchidas corretamente no currículo.

<sup>9</sup> Algumas publicações não foram consideradas nesta coluna por não terem sido preenchidas corretamente no currículo.

área de “Ciências Exatas Tecnologia” e “Ciências Biológicas e da Saúde” que são áreas de interesses mais globalizados e desta base de dados.

#### 2.4.1.2 O Diretório de Grupos de Pesquisa

O Diretório dos Grupos de Pesquisa (DGP) brasileiro é um conjunto de informações em bases de dados sobre os grupos de pesquisa em atividade no país e que faz parte dos sistemas da Plataforma Lattes. Os recursos humanos pertencentes aos grupos, as linhas de pesquisa, os setores de atividade envolvidos, as especialidades do conhecimento, a produção científica, tecnológica e artística e os padrões de interação com o setor produtivo são algumas das informações contidas no Diretório. Os grupos estão localizados em instituições de ensino superior, institutos de pesquisa, etc. As informações individuais dos participantes dos grupos são extraídas dos seus Currículos Lattes. É importante ressaltar que cada grupo é situado no espaço (região, estado e instituição) e no tempo (CNPQ, 2016c).

O Diretório mantém uma base corrente, cujas informações são atualizadas continuamente pelos líderes de grupos, pesquisadores, estudantes e dirigentes de pesquisa das instituições participantes. O DGP pode descrever com precisão os limites e o perfil geral da atividade científico-tecnológica no Brasil. Também é capaz de fornecer uma grande e diversificada gama de informação e detalhes de quem realiza as atividades, como e onde se realizam e sobre o que pesquisam. Tem sido utilizado pela comunidade científica e tecnológica e pelos comitês assessores do CNPq como ferramenta de orientação para suas atividades. De acordo com o último censo do DGP de 2014 são ao todo 35.424<sup>10</sup> grupos de pesquisa, conforme Tabela 2, abaixo. Sendo que a maior parte deles 7.408 (20,9%) são da grande área de Ciências Humanas, seguido pelos das Ciências da Saúde com 5.609 (15,8%).

Tabela 2 – Censo 2014 da distribuição dos grupos de pesquisa

Grande área do conhecimento	Grupos	%	% acumulado
Ciências Humanas	7408	20,9	20,9
Ciências da Saúde	5609	15,8	36,8
Soc. Aplicadas	4841	13,7	50,4
Engs. e Computação	4676	13,2	63,6
Ciências Biológicas	3650	10,3	73,9
C. Exatas e da Terra	3494	9,9	83,8
Ciências Agrárias	3292	9,3	93,1
Ling., Letras e Artes	2454	6,9	100,0
<b>Total</b>	<b>35424</b>	<b>100</b>	<b>-</b>

Fonte: CNPQ (2014)<sup>11</sup>

<sup>10</sup> Houve um aumento de 7.901 grupos de pesquisa. No último Censo eram 27.523.

<sup>11</sup> Disponível em: < <http://lattes.cnpq.br/web/dgp/por-grande-area1>>. Acesso em: 10 out. 2016.

#### **2.4.2 Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)**

O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) é referência em projetos voltados ao movimento do acesso livre ao conhecimento. Exemplo desse compromisso é a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), lançada em 2002, que utiliza as mais modernas tecnologias de arquivos abertos e integra sistemas de informação de teses e dissertações de instituições de ensino e pesquisa brasileiras. A BDTD possui um acervo de mais de 126 mil teses e dissertações de 90 instituições de ensino. O que faz dela a maior biblioteca dessa natureza, no mundo, em número de registros de teses e dissertações de um só país (IBICT, 2016a).

A BDTD tem por objetivo reunir, em um só portal de busca, as teses e dissertações defendidas em todo o País e por brasileiros no exterior. Ela foi concebida e é mantida pelo IBICT com apoio da Financiadora de Estudos e Pesquisas (FINEP), tendo o seu lançamento oficial no final do ano de 2002 (IBICT, 2016b).

A iniciativa de criação de uma base nacional de teses e dissertações, teve, então, as seguintes linhas principais de atuação:

Estudar experiências de desenvolvimento de bibliotecas digitais de teses e dissertações; Desenvolver, em cooperação com membros da comunidade, um modelo para o sistema; Definir padrões de metadados e tecnologias a serem utilizadas pelo sistema; Absorver e adaptar as tecnologias a serem utilizadas na implementação do modelo; Desenvolver um sistema de publicação eletrônica de teses e dissertações para atender instituições que não possuíssem sistemas automatizados para implantar suas bibliotecas digitais; Difundir os padrões e tecnologias adotadas e dar assistência técnica aos parceiros na implantação das mesmas (IBICT, 2016a).

Após dez anos do início do projeto, a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) se apresenta como uma das maiores iniciativas para a disseminação e visibilidade de teses e dissertações, seguindo os preceitos da Iniciativa de Arquivos Abertos (OAI sigla em inglês) e adotando modelo baseado em padrões de interoperabilidade. As instituições de ensino e pesquisa são as provedoras de dados e o IBICT atua como agregador que: coleta os metadados, fornece serviços de informação sobre eles e os expõem para coleta para outros provedores de serviços. Qualquer instituição brasileira de ensino e pesquisa que tenha programa de pós-graduação *stricto sensu* (mestrado e/ou doutorado) pode solicitar a sua participação na BDTD (IBICT, 2016b).

#### **2.4.3 As patentes e a Espacenet**

O capitalismo, como sistema econômico vigente, possui como característica central a produção e a introdução de inovações. Assim como o capitalismo, a economia evolucionista apresenta transformações e cria mecanismos, para que a produção e a inovação industrial estejam sempre envolvidas em um processo de evolução criativa (ALBUQUERQUE, 1998).

As patentes objetivam atribuir direitos e deveres dentro de um sistema que pode e deve se auto ajustar. Na prática dos países tais sistemas de inovação materializam-se no depósito de patentes. No Brasil, é o Instituto Nacional da Propriedade Intelectual (INPI) quem regula as ações que envolvem a proteção industrial. Os aspectos relacionados às legislações patentárias são moldados em cada país, mas seguem orientações da Convenção de Paris desde 1880. É importante destacar que o Brasil está distante da fronteira tecnológica internacional, fazendo muito sentido analisar o tema patentes como uma construção institucional que garante a apropriabilidade dessa mercadoria intangível que é a informação e o conhecimento (ALBUQUERQUE, 1998).

Nesse contexto, a informação é matéria-prima, atua como produto e como insumo na descrição do registro de uma patente, essa valorização da informação está fortalecida em nosso contexto atual, em que vivenciamos a era da informação. Contudo, a informação possui um caráter intangível que dificulta sua posse plena. A informação, nesse contexto econômico, é vista como mercadoria e não como bem público livremente disponível. A apropriação da inovação como parte da dinâmica tecnológica do capitalismo remete-se às oportunidades tecnológicas e às condições de apropriabilidade que são específicas a cada paradigma (ALBUQUERQUE, 1998).

O relatório da Organização Mundial da Propriedade Intelectual (OMPI) publicado em agosto de 2016 apresenta um *ranking* para o índice global de inovação, nele o Brasil ficou em 69º entre os 128 países pesquisados. Essas economias representam 92,8% da população mundial e 97,9% do Produto Interno Bruto (PIB global), a Suíça lidera o ranking das economias mais inovadoras do mundo (OMPI, 2016).

Diante do exposto, as especificidades patentárias em países periféricos, apresentam inovações fundamentalmente incrementais e ainda sofrem com o depósito de patentes estrangeiras que podem restringir o desenvolvimento pleno em determinadas áreas. O simples fato de ter a posse de uma patente não é suficiente para manter uma barreira à entrada, quando o seu proprietário não realiza os investimentos e nem desenvolve as capacitações organizacionais necessárias, quando outro tem vontade e condições de fazê-lo (ALBUQUERQUE, 1998). O sistema de patentes foi criado para avançar a competitividade e

o benefício social, ao incentivar a divulgação pública das invenções em troca de um monopólio estritamente limitado. As invenções, bem como as informações técnicas, podem ser usadas em qualquer lugar, desde que o monopólio da patente não esteja em vigor ou que a proteção não tenha sido solicitada no país em questão.

No INPI são depositados por volta de 23 mil patentes ao ano por estrangeiros enquanto nos EUA são depositadas por volta de 400 mil patentes por ano (OMPI, 2008). Tais números revelam que mais de 350 mil patentes depositadas nos EUA anualmente podem ter seu conteúdo produzido, usado e vendido licitamente no Brasil, uma vez que a proteção patentária tem validade somente nos territórios onde foi concedida. Assim, quando o assunto é patentes é imperioso lembrar que a ciência financiada por fundos públicos é um passo importante no caminho para o desenvolvimento de tecnologias que podem melhorar a sociedade. No entanto, nem sempre a relação entre o progresso da ciência e o seu impacto é muito clara (LENS, 2014).

A nanotecnologia ilustra bem este caso, já que nos últimos anos grandes investimentos foram realizados para o progresso das pesquisas na área. Entretanto ainda não trouxeram os resultados esperados com relação a transferência de tecnologia. O número de publicações aumentou significativamente, mas o número de patentes concedidas ou até mesmo solicitadas ao INPI não pode acompanhar tal crescimento por motivos aparentemente óbvios, pois uma vez que o conhecimento está publicado, já não poderia ser protegido.

Nessa linha de raciocínio pode-se também questionar se patentes em nanotecnologia, protegidas apenas no Brasil, teriam realmente condições de atender aos propósitos dos investimentos realizados, ou seja, tornar a indústria brasileira mais competitiva no mercado internacional, pois raras são as patentes brasileiras com abrangência mundial.

Uma consulta rápida ao INPI, com o termo nano\* nos resumos das patentes, retornou 1.964 patentes em todo o período que compreende os anos 1991 e 2016. Ao se refazer a busca utilizando o filtro “apenas patentes concedidas”, neste mesmo período, encontrou-se 153 patentes, o que representa 7,8%. Ou seja, como há casos em que se leva 12 anos para concessão de um pedido de patente<sup>12</sup>, ainda não é possível se pautar nas informações de concessão. Com a busca nano\* nos títulos na base do INPI é possível perceber que de 2006 até 2016 foram depositadas 1.306 patentes, mas destas, apenas 9 estavam concedidas até a data da busca (11/10/2016). Dentre as diversas razões para este baixo resultado, uma delas é

---

<sup>12</sup> Patente número PI 0215135-9 B1 MÉTODO PARA PREPARAR UMA DISPERSÃO ESTÁVEL DE NANOPARTÍCULAS DE CARBONO EM UM LÍQUIDO. Data do Depósito: 06/12/2002 - Data da Publicação: 04/01/2005 - Data da Concessão: 16/12/2014.

devido ao fato de que muitos pesquisadores publicam artigos científicos com os resultados de suas pesquisas e só depois solicitam patenteamento.

Sobre as características dos documentos de patentes salienta-se a sua classificação, eles são classificados de acordo com diferentes sistemas, dependendo da autoridade responsável pela concessão da patente. O sistema de classificação mais importante é a Classificação Internacional de Patentes (IPC) da WIPO. Introduzido em 1968, o IPC é usado por todos os escritórios de patentes em todo o mundo, alguns dos quais também utilizam um sistema de classificação nacional. O IPC tem uma estrutura hierárquica e subdivide-se em seções, classes, subclasses, grupos e subgrupos. É um dos sistemas de classificação mais preciso disponível, o IPC atualmente divide a tecnologia em cerca de 70.000 sub-áreas (WIPO, 2016).

O sistema de classificação de patentes facilita o arquivamento e recuperação destes documentos. Por exemplo, verificar o estado da arte de uma tecnologia seria praticamente impossível sem classificação, porque a busca com palavras-chave muitas vezes pode produzir resultados imprecisos e incompletos, devido ao idioma no qual os documentos de patentes são escritos e os termos utilizados<sup>13</sup>.

Nas análises de patentes, é importante reconhecer as famílias de patentes para verificar os países onde há proteção para uma mesma invenção. Há várias limitações para buscas efetivas sobre tecnologias, dentre elas tem-se a fase de sigilo (18 meses da data de depósito) dos documentos, pois qualquer base de dados ou ferramenta de busca utilizada vai recuperar apenas documentos que já tenham sido publicados. Outra limitação é a diversidade de fontes de informação que deverão ser consultadas para atestar que não há anterioridade (Escritórios nacionais e internacionais, bases de publicações científicas, dentre outras).

Na base do EPO (Escritório Europeu de Patentes tradução da sigla em inglês) o número do documento da patente pode ser utilizado no campo de busca para encontrar a família da patente. Essa opção é útil quando se quer informações para exploração ou identificação de mercados, por exemplo.

#### **2.4.4 Acesso à Informação no Brasil**

---

<sup>13</sup> Disponível em:

<[https://worldwide.espacenet.com/help?locale=en\\_EP&method=handleHelpTopic&topic=ipc](https://worldwide.espacenet.com/help?locale=en_EP&method=handleHelpTopic&topic=ipc)>. Acesso em: 10 out. 2016.

Acesso aberto ou acesso livre (*open access*) é definido pela disponibilização livre e permanente, na internet, de cópias gratuitas e sem restrições de acesso, de literatura de caráter científico ou acadêmico (BOAI, 2002).

A primeira declaração internacional sobre o acesso aberto foi a *Budapest Open Access Initiative* (BOAI) em 2002, ela forneceu a primeira definição de acesso aberto (acesso completamente livre e irrestrito) e definiu as duas vias complementares para o atingir: a via dourada (publicação em revistas em acesso aberto) e a via verde (publicação em repositórios institucionais ou temáticos em acesso aberto). No ano seguinte lançou-se outras duas declarações: a *Bethesda Statement on Open Access Publishing* e a *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* (UNIVERSIDADE DO PORTO, 2012).

No Brasil, o Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica foi lançado em 13 de setembro de 2005, obteve 141 adesões de instituições, nele havia recomendações aos quatro principais grupos de interesse para a informação científica, sendo: comunidade científica; instituições acadêmicas e de pesquisa; e agências de fomento. A partir daí houve a articulação com os diversos segmentos da comunidade científica visando sensibilizá-los quanto à importância das recomendações constantes do manifesto brasileiro. Realizou-se um alinhamento com as diretrizes internacionais consensuadas pela: *Red Federada de Repositorios Institucionales de Publicaciones Científicas (LA Referencia)*; *Open Access Infrastructure for Research in Europe (OpenAIRE)*; *Confederation of Open Access Repositories*, pois precisavam serem compatibilizadas com qualquer sistema que se pretendesse adotar (ALVARES, 2015).

Após tais aproximações iniciais, deu-se a construção do Portal de Repositórios e Publicações de Acesso Livre, Portal OasisBr e também a construção de Repositórios Institucionais Temáticos, as Publicações Eletrônicas<sup>14</sup>, Repositórios Institucionais<sup>15</sup>, a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), Portal do Livro Aberto em Ciência, Tecnologia e Inovação em Publicações Eletrônicas<sup>16</sup>. E para todas as iniciativas

---

<sup>14</sup> BIBLIOTECA DIGITAL BRASILEIRA DE TESES E DISSERTAÇÕES (BDTD); PORTAL DO LIVRO ABERTO EM CIÊNCIA, TECNOLOGIA E INOVAÇÃO; PLATAFORMA SEER/OJS (ALVARES, 2015).

<sup>15</sup> O DSpace tem sido o software mais usado, segundo dados do Registry of Open Access Repository (ROAR) para implementação de repositórios de acesso aberto. O DSpace é uma plataforma que permite o depósito de documentos em qualquer formato (texto, vídeo, áudio e dados) para gestão de documentos e de metadados. Permite a preservação digital de longo prazo (ALVARES, 2015).

<sup>16</sup> KOHA, Solução para gestão integrada de bibliotecas. É composto por vários módulos dando suporte a todas as atividades realizadas no contexto de uma biblioteca: Catálogo Online de Acesso Público (OPAC), catalogação, gestão de autoridades, circulação, aquisições, relatórios, administração e ferramentas. O Koha está traduzido em cerca de 100 idiomas, incluindo o português, e encontra-se implementado em mais de 900 instituições em

houve o estabelecimento de políticas e competência técnica quanto ao uso e desenvolvimento de ferramentas em conformidade com o Open Access (AO) – dados abertos à pesquisa científica (ALVARES, 2015).

Em setembro de 2016 o IBICT lançou um outro manifesto, o "Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã", uma tomada de posição brasileira no que tange o movimento mundial de dar acesso aberto à informação científica. Nele o IBICT divulga sua visão sobre o acesso aberto, e reconhece os dados de pesquisa como um recurso imprescindível para as ações de Ciência Aberta, Ciência para todos, Ciência Cidadã (IBICT, 2016c).

Segundo Costa (2006), o acesso aberto ao conhecimento científico, quando resultante de pesquisas financiadas com recursos públicos, tem sido discutido em diferentes áreas do conhecimento e envolve todos os atores da comunidade científica (composta pelos pesquisadores, editores, bibliotecas, provedores de acesso, agências de fomento, sociedades científicas e outros organismos do contexto acadêmico).

Ortellado (2008) ressalta dois fatores importantes para a surgimento do movimento acesso aberto na literatura científica, sendo: o que ele chamou de crise dos periódicos que tomou as bibliotecas universitárias a partir do final dos anos 1980, que realizaram diversos cancelamentos devido ao alto custo das assinaturas e, o advento da *World Wide Web* que permitiu uma rápida disseminação de ideias e informações em formato digital. Entretanto, a dominação do mercado global de periódicos indexados por editoras transnacionais privadas ainda prevalece.

As políticas nacionais de estímulo às publicações de acesso aberto têm impacto diferente nas diversas áreas do conhecimento, nas ciências sociais essa política tem conseguido liberar o acesso a quase metade dos artigos, nas ciências naturais os resultados não são tão significativos considerando o contexto nacional. Em publicações proprietárias o comportamento é o contrário, há mais representatividade das ciências exatas (ORTELLADO, 2008).

---

todo o mundo. VUFIND - é uma ferramenta de descoberta e entrega com a finalidade de apoiar a criação de portais integradores de busca e recuperação de informação. Possibilita a integração de variadas ferramentas, de forma a criar um portal de busca consolidado. ARQUIVEMATICS - é um sistema de preservação digital utilizado em todo o mundo, projetado para a gestão de objetos digitais, a partir de padrões funcionais internacionais. Esta baseado em uma serie de ferramentas de código aberto integrados que executam as tarefas necessárias. LOCKSS (LOTS OF COPIES KEEP STUFF SAFE) - software de preservação digital para preservar e fornecer acesso ao conteúdo digital, baseado em uma rede de preservação, onde cada um preserva o de todos, garantindo muitas copias de segurança. Ele foi originalmente projetado para revistas acadêmicas, mas agora também é usado para uma grande variedade de materiais (ALVARES, 2015).

Uma série de estudos recentes tem mostrado que embora a maioria das editoras de periódicos científicos seja privada, elas recebem subsídios públicos em pelo menos três pontos da cadeia produtiva: na elaboração do conteúdo dos artigos (com o financiamento da pesquisa científica pelas universidades); no processo de revisão por pares (que não é remunerado pelas editoras, ficando também a cargo das universidades); e, por fim, na própria aquisição dos periódicos pelas bibliotecas universitárias (HOUSE OF COMMONS, 2004<sup>17</sup> apud ORTELLADO, 2008).

Neste sentido, acredita-se que a publicação fechada seja um dificultador do progresso da própria ciência, por criar restrições artificiais à livre circulação e disseminação do conhecimento, além de ser mais um fator a contribuir para a desigualdade entre profissionais e pesquisadores de diferentes países e/ou regiões do mundo (CAMARGO JÚNIOR, 2012).

Ainda nesta discussão sobre o acesso aberto às informações geradas com recursos públicos, a Lei de Acesso a Informação (LAI) tem como comando central, que a informação é a regra e o sigilo é a exceção (BRASIL, 2011).

Sendo assim, se não houver referência quanto ao caráter sigiloso da informação, a mesma deve ser publicizada. Tal dispositivo traz importantes diretrizes, tais como: - Divulgação de informações de interesse público, independentemente de solicitações, o qual tem sido denominado de Transparência Ativa (elencadas no artigo 8º da LAI); - elegeu-se a internet como meio para divulgar as informações; - tornou dever dos Estados garantir o direito dos cidadãos de acesso à informação (BERNARDES; SANTOS; ROVER, 2012, p. 11).

Não se pode esquecer que se vivencia uma Sociedade Informacional, de Governo Aberto, de uma sociedade que a cada dia interage mais na rede e, portanto, demanda por governos mais participativos (BERNARDES, SANTOS, ROVER, 2012). Um exemplo elucidativo desta afirmação no Brasil, é o modo como o acesso aos dados da Plataforma Lattes é disponibilizado, tanto na Base de Currículos Lattes (BCL) como no Diretório de Grupos de Pesquisa (DGP), os dados estão disponíveis de maneira individualizada e estanque, não é possível relacionar e ou visualizar uma integração entre eles. Para este fim, o usuário

---

<sup>17</sup> HOUSE OF COMMONS. Science and Technology Committee. Scientific Publications: Free for All? London: House of Commons, 2004. Disponível em: <http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/399.pdf>. Acesso em 20 de maio de 2008.

necessita baixar os dados um a um e posteriormente trata-los à sua maneira, pois a Plataforma Lattes não disponibiliza mecanismos para extração dos dados.

O estudo de Bernardes, Santos, Rover (2012) constatou que no Brasil o modelo democrático vigente é recente, e ensaia os primeiros passos para sair da cultura do sigilo. Enquanto isso, mecanismos legais tais como a LAI, que obrigam a divulgação de informações na rede, são vistos com repulsa. Todavia, a Suprema Corte brasileira, enquanto guardiã da Constituição Federal e atenta aos compromissos internacionais firmados pelo Estado brasileiro, conjuga esforços para fazer prevalecer a LAI, inclusive com a divulgação de dados dos seus servidores, para servir de exemplo aos demais órgãos (VAZ; RIBEIRO; MATEUS, 2010). Esse é o caminho necessário para que no planejamento e operacionalização de projetos de Governo Aberto, possam efetivamente contribuir para o aprimoramento democrático.

Fornecer dados públicos no formato aberto, oportuniza a reutilização, recombinação e o acesso às informações, também propicia o fortalecimento da democracia e participação cidadã. Isto é, os dados abertos governamentais são os dados produzidos pela administração pública e colocados à disposição das pessoas com o objetivo de tornar possível não apenas sua leitura, mas também a reutilização desses dados em novos projetos, plataformas, *sites* e aplicativos, seu cruzamento com outros dados de diferentes fontes, e sua disposição em visualizações interessantes e esclarecedoras sobre nossa sociedade (VAZ, RIBEIRO, MATEUS, 2010, p. 46).

Para Cruz et al. (2012) o que mais se encontra, ainda, nas páginas de transparência é o ponto de vista dos governos, e não, necessariamente, a transparência que o cidadão deseja. Conhecer a demanda de informação dos cidadãos, entidades externas e entes governamentais, é imperioso para a formatação de bons instrumentos de governo eletrônico, como indica Cruz et al.:

O levantamento de requisitos é uma das etapas do desenvolvimento de *software* que busca levantar junto aos usuários do sistema quais funcionalidades desejadas para o *software*. Pode parecer simples, mas essa fase é considerada extremamente crítica, e quando mal sucedida, torna-se um dos principais fatores do insucesso dos projetos de *software*. Isso ocorre, principalmente, porque o usuário do sistema não consegue explicar ao analista da tecnologia o que ele precisa, ou de outra forma, os analistas podem não estar abertos as ideias dos usuários (CRUZ et al., 2012, p. 262).

Segundo Silva (2007) um grande problema é que a produção e a apropriação dos estoques de informação não são democráticas (SILVA, 2007). Entretanto, a representação e a organização da informação devem dar ao usuário de um Sistema de Recuperação de

Informação (SRI) um acesso fácil a informação de seu interesse, seja para recuperação de informação ou para a navegação entre documentos.

### 3 MÉTODO E DESENVOLVIMENTO DA PESQUISA

Uma das tarefas principais do cientista é delimitar as leis científicas e para isso, deve adotar um método de pesquisa. Desde Francis Bacon, a Ciência tem seguido o princípio de que para relatar uma lei da natureza é preciso testar repetidas vezes, colher e registrar os resultados, aguardar que isso se refaça com vários outros pesquisadores para, então, ser considerada verdadeira. Isto é, uma lei científica é válida quando a comunidade científica, fundada em experiências particulares, colhe resultados semelhantes ou pretensamente iguais repetidas vezes (POPPER, 1968).

Esse é o método da indução, conforme explica Popper (POPPER, 1968). Costuma-se chamar de “indutiva” a uma inferência se ela passa de enunciados particulares, ou experimentos, aos enunciados universais, tais como as hipóteses ou “teorias”. Para testar uma teoria, Popper segue quatro passos, ou espécies de provas: 1º testes internos: buscam a coerência das conclusões extraídas a partir do enunciado; 2º testes da forma: consiste nos testes para se saber se a teoria é, de fato, uma teoria empírica ou científica ou meramente tautologia; 3º testes de inovação: verificação se a teoria realmente é nova ou já está compreendida por outras existentes no sistema; 4º testes empíricos: verificação da aplicabilidade das conclusões extraídas da teoria nova, esses são os mais relevantes entre os testes propostos por Popper, pois a teoria pode passar intacta pelos três testes anteriores e ser falseada pela aplicação empírica de suas conclusões, caso em que a teoria não será considerada válida (POPPER, 1968).

Enfim, ressalta-se que, para Popper, é necessário o enunciado poder ser testado empiricamente, não pela sua verificabilidade, mas pela sua falseabilidade. Pode-se sintetizar os aspectos da epistemologia de Karl Popper abordados neste trabalho em algumas proposições:

- a) A concepção segundo a qual o conhecimento científico é descoberto em conjuntos de dados empíricos (observações/experimentações neutras, livres de pressupostos) - método indutivo - é falsa;
- b) Não existe observação neutra, livre de pressupostos; todo o conhecimento está impregnado de teoria;
- c) O conhecimento científico é criado, inventado, construído com objetivo de descrever, compreender e agir sobre a realidade;
- d) As teorias científicas não podem ser demonstradas como verdadeiras; são conjecturas, virtualmente provisórias, sujeitas a reformulações, a reconstruções;

e) Todo o conhecimento é modificação de algum conhecimento anterior (POPPER, 1968).

Lakatos concorda com Popper quanto a crítica segundo a qual "todas as teorias são igualmente improváveis" (LAKATOS, 1979). Lakatos defende a ideia de uma racionalidade não instantânea, historicamente reconstruída, revelada em seus estudos sobre a metodologia dos programas de pesquisa, a favor de um falseacionismo metodológico sofisticado (LAKATOS, 1987).

Diferentemente de Lakatos, Feyerabend não intenta fornecer uma nova metodologia ou uma nova teoria da racionalidade. Seu objetivo é convencer o leitor de que "todas as metodologias, mesmo as mais óbvias, têm limitações" (FEYERABEND, 1977, p. 43).

Para Feyerabend, o conhecimento não é uma aproximação gradual da verdade. Está mais para um oceano de alternativas mutuamente incompatíveis ou incomensuráveis, cujas teorias, contos de fadas, mitos, que fazem parte do todo, impulsionam as demais partes a manterem articulação maior, fazendo com que todas concorram para o desenvolvimento de nossa consciência (FEYERABEND, 1977).

### **3.1 O delineamento da pesquisa**

De acordo com os objetivos desta pesquisa ela é considerada exploratória e também descritiva.

Na pesquisa exploratória Gil (1999) destaca que ela é desenvolvida no sentido de proporcionar uma compreensão geral sobre determinado fato. Portanto, esse tipo de abordagem é realizada quando o tema escolhido é pouco explorado e torna-se difícil formular hipóteses precisas. Para Andrade (2002) há algumas finalidades primordiais nesse tipo de pesquisa, como: proporcionar maiores informações sobre o assunto investigado; facilitar a delimitação do tema de pesquisa; orientar a fixação dos objetivos e a formulação de hipóteses; ou descobrir um novo tipo de enfoque sobre o assunto.

Já a pesquisa descritiva preocupa-se em observar os fatos, registrá-los, analisá-los, classificá-los e interpretá-los, e o pesquisador não interfere. Assim, os fenômenos do mundo físico e humano são estudados, mas não manipulados pelo pesquisador (ANDRADE, 2002).

De acordo com os objetivos e procedimentos, esta pesquisa se enquadra como um estudo de caso com um levantamento do tipo documental (YIN, 2001). Para tanto, ressalta-se que o estudo de caso e o levantamento são apenas duas das muitas maneiras de se fazer pesquisa em ciências sociais aplicadas. Dentre outras tem-se os experimentos, as pesquisas

históricas, as análises de informações em arquivos, a pesquisa bibliográfica, a documental e a participante são alguns exemplos de outras maneiras de se realizar pesquisa. Em geral, os estudos de caso representam a estratégia adequada quando se colocam questões do tipo “como” e “por que”, quando o pesquisador tem pouco controle sobre os eventos e quando o foco está em fenômenos contemporâneos e inseridos em algum contexto da vida real (YIN, 2001).

Como estratégia de pesquisa, utiliza-se o estudo de caso em muitas situações, nas quais se incluem: política, ciência política e pesquisa em administração pública; sociologia e psicologia comunitária; estudos organizacionais e gerenciais; pesquisa de planejamento regional e municipal, como estudos de plantas, bairros ou instituições públicas; supervisão de dissertações e teses nas ciências sociais – disciplinas acadêmicas e áreas profissionais como administração empresarial, ciência administrativa e trabalho social (YIN, 2001, p. 19-20).

De acordo com GIL (1999) o estudo de caso surge da necessidade de compreensão dos fenômenos sociais complexos.

O Quadro 2 resume os pontos a serem observados ao se escolher as estratégias de pesquisas, a parte sombreada é a que está relacionada com a presente pesquisa.

Quadro 2 – Situações relevantes para diferentes estratégias de pesquisa.

<b>Estratégia</b>	<b>Forma da questão de pesquisa</b>	<b>Exige controle sobre eventos comportamentais?</b>	<b>Focaliza acontecimentos contemporâneos</b>
Experimento	Como, por que	Sim	Sim
Levantamento	Quem, o que, onde, quantos, quanto	Não	Sim
Análise de arquivos	Quem, o que, onde, quantos, quanto	Não	Sim/não
Pesquisa histórica	Como, por que	Não	Não
Estudo de caso	Como, por que	Não	Sim

Fonte: Yin (2001)

Nesse sentido, nota-se que se pode utilizar mais de uma estratégia em qualquer estudo dado, como por exemplo, um levantamento em um estudo de caso ou um estudo de caso em um levantamento, que é o caso do presente estudo. Contudo, identifica-se também algumas situações em que uma estratégia específica possui uma vantagem distinta. No estudo de caso isso acontece quando:

faz-se uma questão do tipo “como” ou “por que” sobre um conjunto contemporâneo de acontecimentos sobre o qual o pesquisador tem pouco ou nenhum controle. Consiste quando o pesquisador através do processo de coletar, analisar e interpretar observações possa fazer inferências relativas às relações causais entre as variáveis sob investigação (YIN, 2001, p. 28).

O estudo de caso como o experimento não permite uma “amostragem”, uma vez que o objetivo do pesquisador é expandir e generalizar teorias (generalização analítica) e não enumerar frequências (generalização estatística) (YIN, 2001). Por esse motivo, a presente pesquisa apoia-se, também, no levantamento. Compreendendo um estudo de caso somente no que diz respeito ao objeto de estudo escolhido, ou unidade de análise, que neste caso é a área prioritária para o governo brasileiro, a Nanotecnologia, área objeto do presente estudo.

A nanotecnologia é uma área de relevante importância econômica, revisões realistas estimam o valor de 49 bilhões de dólares para 2017, excluindo os semicondutores, os nanomateriais são os majoritários do montante e representam 37 bilhões de dólares ((SCHULZ, LEIVA e MILANEZ, 2014).

A nanotecnologia é um campo essencialmente interdisciplinar, pesquisa realizada na *WoS e Derwent Innovations Index*, revelam que as publicações científicas entre 1980 e 1990 registraram um aumento de 1.457% entre as décadas subsequentes. Para as patentes, no mesmo período o número de depósitos de patentes foi de 86,22% e 443,3% respectivamente (SCHULZ; LEIVA; MILANEZ, 2014).

Para a execução da parte empírica, a autora contou com a colaboração do Núcleo de Informação Tecnológica em Materiais (NIT/Materiais), sediado no Departamento de Engenharia de Materiais (DEMa) e parceiro do Programa de Pós Graduação em Ciência, Tecnologia e Sociedade (PPGCTS) da Universidade Federal de São Carlos (UFSCar), e no qual a presente pesquisa se desenvolve. O NIT/Materiais tem contribuído desde 1994 na capacitação de pessoas, na pesquisa e no desenvolvimento de métodos e ferramentas para a análise da informação tecnológica e estratégica na área de materiais e processos, de gestão tecnológica e inovação, diagnóstico tecnológico em propriedade intelectual e indicadores de ciência e tecnologia (C&T), etc.

O NIT/Materiais desenvolve e atua em diversos projetos de pesquisa e extensão realizados junto a órgãos de fomento, instituições de ciência e tecnologia e empresas, tais como a FAPESP (SP), FINEP, SEBRAE (SP), IPT (SP), TECPAR (PR), INPA (AM), EMBRAPA, PETROBRAS, EATON e outras (NIT, 2013). Houve participação, por exemplo, na elaboração de capítulos de indicadores de produção científica, nas edições de 2004 e 2010 dos Indicadores de Ciência, Tecnologia e Inovação em São Paulo (FARIA, et. al., 2011;

GREGOLIN, et. al., 2005), organizadas pela FAPESP. O NIT/Materiais também já desenvolveu outros trabalhos relacionados a nanotecnologia sob outras perspectivas. Eles formam um corpo de conhecimento sólido do grupo e fornecem subsídios que aumentam a efetividade das atividades necessárias à presente pesquisa.

Além do NIT/Materiais, a pesquisa se insere nos estudos recentes do grupo de pesquisa Vlab4u coordenado pelo Prof. Dr. Luc Quoniam. Três membros do grupo, o pesquisador Prof. Dr. Jesús Mena-Chalco, Prof. Dr. André Moraes e Prof. Dr. David Reymond, foram os programadores responsáveis pelos scripts utilizados. ScriptLattes, ScriptExtract, ScriptKeyword pelo Prof<sup>o</sup> Jesús, ScriptsGP pelo Prof<sup>o</sup> André e Patent2Net pelo Prof<sup>o</sup> David.

Um levantamento formal é considerado parte de um estudo de caso, ele inclui tanto os procedimentos de amostragem quanto os instrumentos utilizados em levantamentos habituais. Os dados referentes ao levantamento podem ser extraídos com base em uma amostra retirada de determinada população ou universo que se deseja conhecer (YIN, 2001), no qual considera-se a apresentação de algumas subunidades (apenas o *core* – nesta pesquisa com todos currículos que tenham no mínimo dois artigos e/ou que participam de grupo de pesquisa relacionado a área).

De acordo com Gil (1999) as pesquisas de levantamento se caracterizam pela interrogação direta das pessoas cujo comportamento se deseja conhecer (no caso deste trabalho será uma área da ciência mediante a coleta de dados em bases de dados), procede-se a solicitação de informações a um grupo significativo de pessoas acerca do problema estudado para em seguida, mediante análises quantitativas, obter as conclusões correspondentes aos dados coletados. Quando o levantamento coleta informações de todos os integrantes do universo pesquisado, tem-se um censo. Pelas dificuldades materiais que envolvem sua realização, os censos geralmente são desenvolvidos pelos governos ou instituições de amplos recursos (BEUREN, 2008).

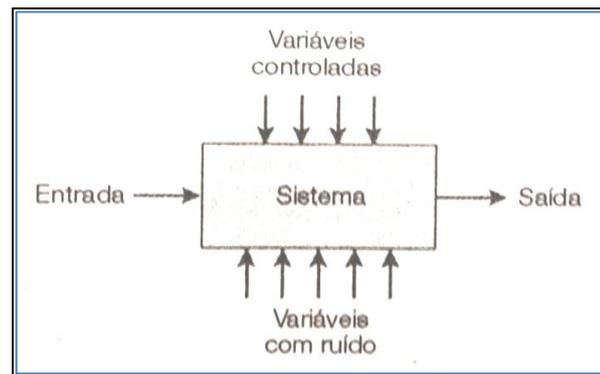
Quanto aos procedimentos utilizados no desenvolvimento da sistemática, é possível classificar esta pesquisa como um levantamento do tipo documental sobre a Nanotecnologia no Brasil, mediante a recuperação de informações junto as bases de dados: Base de Currículos Lattes; Diretório de Grupos de Pesquisa; BDTD e Espacenet.

A presente pesquisa, quanto à abordagem do problema, compreende uma pesquisa quantitativa. De acordo com Richardson (1999, p. 70) pesquisas de natureza quantitativa caracteriza-se pelo emprego de “quantificação tanto nas modalidades de coleta de

informações, quanto no tratamento delas por meio de técnicas estatísticas, desde as mais simples como percentual, média, desvio-padrão, às mais complexas, como coeficiente de correlação, análise de regressão etc.”. É aplicada frequentemente aos estudos descritivos que procuram descobrir e classificar a relação entre variáveis e a relação de causalidade entre fenômenos. Também são comuns quando a tentativa é entender mediante uma amostra o comportamento de uma população (BEUREN, 2008).

As bases de dados utilizadas foram compreendidas como um sistema que pode ser representado de acordo com a Figura 3. As bases incorporam uma alimentação incontrolada (ruído) que combinada com uma alimentação controlada, produz a saída do sistema. Por causa da alimentação incontrolada, os mesmos cenários para a alimentação controlada não resultam saídas idênticas cada vez que o sistema é medido.

Figura 3 – Variáveis controladas e com ruído de um sistema



Fonte: Montgomery e Runger (2009)

Nesse sentido, o experimento desta pesquisa se encaixa no conceito de “experimento aleatório” na estatística – um experimento que pode fornecer diferentes resultados, muito embora seja repetido toda vez da mesma maneira, é chamado de experimento aleatório. Para modelar e analisar um experimento aleatório é necessário entender o conjunto de resultados possíveis de um experimento.

O conjunto de todos os resultados possíveis de um experimento aleatório é chamado de espaço amostral do experimento. O espaço amostral é denotado por  $S$ . Para este caso é objetivo da análise considerar os filtros estabelecidos para que os currículos pertençam ou não ao conjunto escolhido. Nesta pesquisa será:  $S = \{\text{sim}, \text{não}\}$ . Tem-se um espaço amostral “discreto”, pois consiste em um conjunto finito ou infinito contável de resultados.

A população ou a totalidade de elementos sob estudo, podendo ser finita ou infinita são os currículos de pesquisadores da Plataforma Lattes, neste caso os currículos também podem ser compreendidos como elementos ou também como as unidades de análise (BUSSAB; ALBERTO, 2013).

O presente estudo selecionou uma amostra, que é uma parte da população de estudo, ou seja, um grupo finito da população. Mediante a amostragem – processo para obtenção de uma amostra – selecionou em última instância das análises, apenas os currículos de pesquisadores centrais (*core*) (BUSSAB; ALBERTO, 2013).

Dentre as técnicas científicas existentes para a seleção correta de amostras, as mais conhecidas são a amostragem aleatória simples, a amostragem sistemática e a amostragem estratificada. Na amostragem aleatória simples qualquer elemento da população tem a mesma probabilidade de ser escolhido. Na amostragem sistemática os elementos da amostra são escolhidos a partir de uma regra estabelecida. A amostragem estratificada é utilizada quando a população está dividida em estratos ou grupos diferenciados (MONTGOMERY; RUNGER, 2009).

O Quadro 3 apresenta uma síntese da condução do estudo de caso desenvolvido nesta pesquisa.

Quadro 3 – Esquema da condução do estudo de caso

<b>CONDUÇÃO DO ESTUDO DE CASO</b>		
<b>DEFINIR UMA ESTRUTURA CONCEITUAL-TEÓRICA</b>	Mapear a literatura	Realizada para apresentação do referencial teórico e construção da expressão de busca.
	Delinear proposições	Embasadas mediante a utilização das bases de dados de modo temático.
	Delimitar as fronteiras e grau de evolução	Investigação bibliográfica sobre os usos e análises já realizadas nas bases de dados.
<b>PLANEJAR O CASO</b>	Selecionar a(s) unidade(s) de análise e contatos	Selecionar uma dentre as áreas prioritárias para o desenvolvimento científico e tecnológico da ciência brasileira.
	Escolher os meios para coleta e análise dos dados	Seleção e criação de scripts específicos para extração e tratamentos dos dados.
	Desenvolver o protocolo para coleta dos dados	<u>Contexto da pesquisa:</u> delimitado pelas fontes de informação em acesso aberto, expressão de busca utilizada e ferramentas livres. <u>Parte a ser investigada:</u> amostra definida mediante as possibilidades das bases e ferramentas. <u>Variáveis de controle:</u> estabelecidas mediante os filtros e as comparações, visando aumentar a precisão dos dados.
<b>CONDUZIR MELHORIAS (TESTE PILOTO)</b>	Definir meios de controle da pesquisa	Criação dos filtros e algumas comparações
	Testar procedimentos de aplicação	Realização de extrações preliminares.
	Verificar qualidade dos dados	Realização de tratamentos preliminares.
	Fazer os ajustes necessários	Criação de novos scripts e filtros.
<b>COLETAR OS DADOS (FASE EM QUE HÁVERA O LEVANTAMENTO)</b>	Selecionar os casos	Delimitação unidade geral (core em nanotecnologia) subunidades ou amostragens (separar por temas, estados, regiões, etc.).
	Registrar os dados	Extração final e armazenamento dos dados.
	Limitar os efeitos do pesquisador	Utilização dos dados registrados para geração de informações novas.
<b>ANALISAR OS DADOS</b>	Produzir uma narrativa	Explanação e redação dos procedimentos da sistemática.
	Reduzir os dados (indicadores iniciais)	Construção de indicadores gerais.
	Construir painel identificar causalidade	Construção de matrizes para visualizar inter-relacionamento dos dados.
	Construir novos indicadores	Construção de indicadores específicos.
<b>GERAR RELATÓRIO</b>	Desenhar implicações teóricas	Confronto entre os resultados gerados com o referencial teórico apresentado.
	Prover estrutura para replicação	Publicação da sistemática proposta e dos resultados alcançados.

Fonte: baseado em Miguel (2007, p. 221)

### **3.2 Procedimentos específicos à Base de Currículos Lattes**

Para esta subseção cabe ressaltar que a Base de Currículos Lattes foi desenvolvida para armazenar e disponibilizar dados e informações de pessoas, se trata de uma base de dados curricular. Dessa maneira, a proposição desta tese ampliou seu uso, mostrando que é possível analisar áreas, temas e/ou assuntos. Entretanto, para que isto seja realizado são necessários procedimentos específicos que foram levantados de modo exploratório e são apresentados a seguir, sendo: a exploração das especificidades da BCL; a elaboração de uma expressão de busca; a extração dos IDs Lattes; e a criação de filtros para coleta.

#### **3.2.1 Exploração das especificidades**

A *priori* foi realizada uma avaliação da BCL (Base de Currículos Lattes) como um Sistema de Recuperação da Informação, inicialmente elaborou-se uma tabela com os recursos disponíveis em uma base conhecida e com potencialidades técnicas de uso e interface consideradas satisfatórias, a *Web of Science* (WoS). Os itens avaliados foram classificados em: busca; usuário; download e classificação. Entretanto, é preciso destacar que há muitas diferenças entre as duas bases, uma é referencial com arquivos na íntegra e a outra curricular.

A escolha pela WoS deve-se por dois motivos: é uma base referencial muito utilizada em diversos estudos bibliométricos; e a BCL, devido seu caráter inovador, não possui uma base concorrente ou similar. Assim, optou-se em considerar as melhores práticas da WoS em uma tentativa de conhecer suas potencialidades e se apropriar delas. Nesse sentido, foi possível destacar importantes pontos que ocorrem ou que não ocorrem na BCL de maneira satisfatória, ver Quadro 4.

Quadro 4 – Comparação de funcionalidades WoS *versus* Currículo Lattes

	WOS	PL
<b>B U S C A</b>	Os operadores booleanos e caracteres coringa funcionam corretamente	Alguns booleanos não funcionam e não condizem com o <i>help</i> . Não há coringas
	É possível fazer busca combinando todos os tipos de campos. Cada campo tem um código padronizado que facilita a busca	Não é possível escolher campos
	Todos os códigos e operados estão visualmente acessíveis na interface de busca avançada	A busca avançada não possui detalhamento de campos
	A base deixa visíveis exemplos de como realizar busca avançada e disponibiliza tutoriais	Não disponibiliza tutoriais
	Permite elaborar estratégias de busca com inúmeros termos	Permite elaborar estratégias de buscas simples e com vários termos, limitando-se ao tamanho da url de resultados (7386 caracteres)
	Os campos disponíveis para busca são os mesmos contidos no registro do documento	Não é possível escolher campos
	Cada campo do registro também pode ser um filtro de busca	As opções de filtro são precárias. Na opção "preferências" nem um filtro funciona
<b>U S U Á R I O</b>	Disponibiliza uma ferramenta de análise dos resultados que mostra quantidade de registros recuperados por cada campo	Não há qualquer forma de análise dos dados
	O help é bem elaborado	O help é desatualizado e restritivo
	Disponibiliza histórico de busca e permite salvá-lo	Não disponibiliza históricos de buscas
	Permite salvar alertas	Não possui alertas
<b>D O W N L O A D</b>	Permite o download dos registros de documentos	Não há possibilidades de download para conjuntos de documentos
	Exporta em diversos tipos de formatos	Download individual em apenas um formato (xml)
	Permite baixar os registros dos documentos (com os metadados) recuperados com a estratégia de busca	Não permite download de mais de um currículo por vez
	Os registros representam bem o documento indexado	O registro é o documento em si
	Os registros são suficientemente descritivos	Não há registro além do próprio documento
	Permite ranquear mais de 100.000 documentos	Permite ranquear (sem apresentar os critérios para tal) os resultados que gerem uma url inferior a 7.386 caracteres
	A quantidade de documentos recuperados mostrados pela base é consistente	A quantidade de documentos recuperados não é consistente, pois não é revelado os critérios utilizados na busca, intui-se que é pela presença ou ausência do termo utilizado em qualquer parte do documento
	O download dos registros é mais demorado quando a estratégia recupera muitos documentos (500-500)	O download é praticamente impossível para pesquisas quantitativas, pois só permitem recuperações individuais
<b>C L A S S I F.</b>	A base adiciona outras palavras-chave àquelas adicionadas pelo autor, melhorando a representatividade e recuperação do documento	A base não permite considerar campos, como o de palavras-chave
	Todos os documentos ao serem indexados são classificados dentro de áreas e subáreas	Os documentos e seus campos não são indexados e por conseguinte não são classificados em áreas do conhecimento

Fonte: Elaboração própria com base em WoS e Currículo Lattes

As bases de dados referenciais, como a WoS, armazenam documentos (arquivos como artigos, capítulos de livro, dentre outros) e também os metadados, ou seja, dados sobre os

dados que estão presentes nos documentos, ou dados sobre estes documentos. Um conjunto de metadados forma o registro de um documento. A grosso modo é um arquivo que possui campos que descrevem o documento, uma catalogação deles. Um dos campos mais significativos é o que atribui uma classificação ao documento, ou seja, o enquadramento do documento em um determinado assunto, tal processo é chamado de indexação.

Desta maneira, quando uma base dispõe dos documentos e dos registros, o sistema de recuperação normalmente utiliza os registros para buscar e recuperar os documentos, favorecendo uma melhora do nível de precisão das buscas pelo fato de que permite buscas por palavras-chave e também pelos campos disponíveis, melhorando a especificidade das expressões de buscas. Entretanto, na BCL os registros e os documentos são a mesma coisa, ou seja, os currículos são preenchidos mediante campos de um formulário e formam um arquivo XML com todos os dados cadastrados pelo próprio indivíduo. Pode-se considerar que os currículos não possuem registros, pois eles são o próprio registro e também o documento.

Metadados, para Marcondes et al., (2005), são entendidos como elementos de descrição/definição de dados armazenados em sistemas computadorizados, organizado por padrões específicos, de forma estruturada. Sua principal função é documentar, por meio de elementos descritores, qualquer tipo de recurso disponível na Web, de modo a permitir comunicabilidade e interoperabilidade entre sistemas.

É importante salientar que muitos campos do Currículo Lattes são preenchidos em linguagem natural, o que dificulta a existência de padrões de registros no seu acervo. Embora o Currículo Lattes seja direcionado a uma comunidade delimitada (científica e tecnológica, e tenha como usuários os cientistas e pesquisadores brasileiros), a recuperação da informação no seu ambiente depende muito do nível de coincidência entre uma determinada estratégia de busca, como por exemplo, palavras utilizadas pelo usuário que busca a informação e as representações dos registros do acervo da sua base de dados, conforme observou Silva (2007).

### **3.2.2 Elaboração da expressão de busca**

Para a coleta dos dados, o primeiro procedimento consistiu na elaboração de uma expressão de busca específica e adequada, que considerou as limitações principalmente da BCL e dos Grupos de Pesquisa, mas que permitiu recuperações pertinentes.

A BCL está em português, mas oferece ao pesquisador a opção de divulgar seu currículo em inglês, além disso, aceita qualquer idioma para o preenchimento dos campos.

Todavia, publicações em inglês foram crescentes nas áreas de ciências duras e biológicas, mas hoje devido as pressões para a internacionalização da ciência todas as áreas tendem a se comportarem de modo bilíngue, por esta razão as palavras-chave, para a expressão de busca criada, foram utilizadas em inglês e português. Entretanto, tal decisão dependerá sempre das características da área tratada.

Para a realização da busca avançada, a BCL oferece uma interface de ajuda (documentação) online que apresenta os operadores de relação e suas respectivas aplicações em uma expressão de busca, mas nem todos são reconhecidos pelo sistema. Nesta documentação (BOOLEANA, 2014) são apresentados como válidos os operadores de relação: and; or; not; near; (); \_; &; /; !; ~; \*; ?; \$.

No Quadro 5 é possível observar o resultado dos testes verificados, nesta pesquisa, com os operadores mencionados, notando-se que mais da metade deles não são reconhecidos pela Base de Currículos Lattes.

Quadro 5 – Operadores reconhecidos pela Plataforma Lattes

<b>Reconhecidos pela Plataforma</b>	
<b>AND</b>	<i>Liga dois ou mais termos, limitando a busca.</i>
<b>OR</b>	<i>Liga dois termos e reúne todos os documentos que incluam pelo menos um deles.</i>
<b>NOT</b>	<i>Buscará registros que contêm o termo de pesquisa que o precede, mas não o termo que o sucede.</i>
<b>NEAR</b>	<i>Encontra documentos contendo ambas as palavras especificadas ou frases contendo até dez palavras entre elas.</i>
<b>( )</b>	<i>Os parênteses servem para elaborar pesquisas ainda mais complexas, definindo operações menores dentro da expressão inteira. A busca funciona nesse caso considerando os parênteses como se fossem termos isolados, e depois os combina.</i>
<b>Não reconhecidos pela Plataforma</b>	
<b>&amp;</b>	<i>Liga dois ou mais termos, limitando a busca.</i>
<b> </b>	<i>Liga dois termos e reúne todos os documentos que incluam pelo menos um deles.</i>
<b>!</b>	<i>Buscará registros que contêm o termo de pesquisa que o precede, mas não o termo que o sucede.</i>
<b>~</b>	<i>Encontra documentos contendo ambas as palavras especificadas ou frases contendo até dez palavras entre elas.</i>
<b>Não descrito na documentação de busca da Plataforma</b>	
<b>_</b>	<i>Usado para unir bitermos.</i>
<b>*</b>	<i>Funcionaria como coringa com qualquer número de caracteres.</i>
<b>?</b>	<i>Funcionaria como coringa de qualquer caractere isolado.</i>
<b>\$</b>	<i>Funcionaria como coringa localizado em qualquer parte do texto.</i>

**Fonte:** LATTES (2014)

Para uma expressão de busca sofisticada é essencial que a base de dados aceite todos os operadores booleanos, principalmente os que permitam a truncagem das palavras,

normalmente representados pelos caracteres \$, ? ou \*. Mas nenhum deles é descrito como truncador e/ ou válido.

Ao iniciar o estudo, pesquisou-se em diversas fontes de informação as palavras que pudessem descrever a área de nanotecnologia (ARORA *et al.*, 2013; GETERM, 2014; GOGOTSI, 2006; MEYER *et al.*, 2010; NANODICTIONARY, 2014; PORTER, *et al.*, 2008; RUSNANO, 2014; WIKTIONARY, 2014). Após o levantamento inicial a lista com 114 termos (em inglês e português) foi submetida à análise de um especialista na área de nanotecnologia, que aprovou os termos e sugeriu a inserção de mais seis termos, ao todo a lista continha 120 termos, apresentados no Quadro 6.

Quadro 6 – Levantamento dos termos da área

1. bionanotechnology	32. nanomanipulation	63. nanomanufacturing	93. nanorings
2. nanocylinders	33. carbon_nanostructure	64. nanomáquina	94. nanorobótica
3. nanodevice	34. dendrimer	65. nanomaterial	95. nanorod
4. nanodiamond	35. diamondoid	66. nanomechanical	96. nanoscratch
5. nanodiffraction	36. fullerene	67. nanomechanics	97. nanosensor
6. nanodispersions	37. graphene	68. nanomechatronics	98. nanosheets
7. nanodomain	38. molecular_machine	69. nanomedicine	99. nanoshell
8. nanodot	39. nanobalance	70. nanometrology	100. nanosources
9. nanodragster	40. nanobarras	71. nanomontador	101. nanospheres
10. nanoelectromechanical	41. nanobatteries	72. nanomotor	102. nanospring
11. nanoelectronic	42. nanobeads	73. nano-object	103. nanostructured_lipid_carriers
12. nanoemulsion	43. nanobeam	74. nanoparticle	104. nanosurgery
13. nanoespuma	44. nanobiology	75. nanoparticle_tracking_analysis	105. nanosuspension
14. nanoestruturado	45. nanobiotecnologia	76. nanoparticulate	106. nanosystem
15. nanofabrication	46. nanobodies	77. nanopatterning	107. nanotechnologies
16. nanofiber	47. nanocapsule	78. nanopeapods	108. nanotextured
17. nanofilm	48. nanocapturer	79. nanopharmaceuticals	109. nanotige
18. nanofluid	49. nanocarro	80. nanophase	110. nanotitania
19. nanoformulations	50. nanocatalysis	81. nanophotonics	111. nanotomography
20. nanofotônica	51. nanoceramics	82. nanoplasma	112. nanotox
21. nanogel	52. nanochip	83. nanopoliuição	113. nanotoxicologia
22. nanograined	53. nanociência	84. nanopolymers	114. nanotube
23. nanohorn	54. nanoclay	85. nanopore	115. quantum_dot
24. nanohybrids	55. nanocluster	86. nanoporous	116. quantum_dots
25. nanoimprint_lithography	56. nanocomponents	87. nanopowder	117. quantum_well
26. nanoimprinting	57. nanocomposites	88. nanoprecipitation	118. quantum_wells
27. nanoindentation	58. nanocone	89. nanoprobe	119. sol_gel
28. nanolitografia	59. nanocontainers	90. nanoquasicrystalline	120. ultrananocrystalline_diamond
29. nanomachine	60. nanocore	91. nanoreator	
30. nanomachining	61. nanocrystal	92. nanoribbon	
31. nanomagnetism	62. nanocrystalline		

Fonte: elaboração própria

A partir de cada termo presente no Quadro 6, elaborou-se blocos de palavras da seguinte forma, pois como não havia possibilidades de truncagem dos termos, foi necessário buscar de maneira direta o termo no singular/plural e inglês/português:

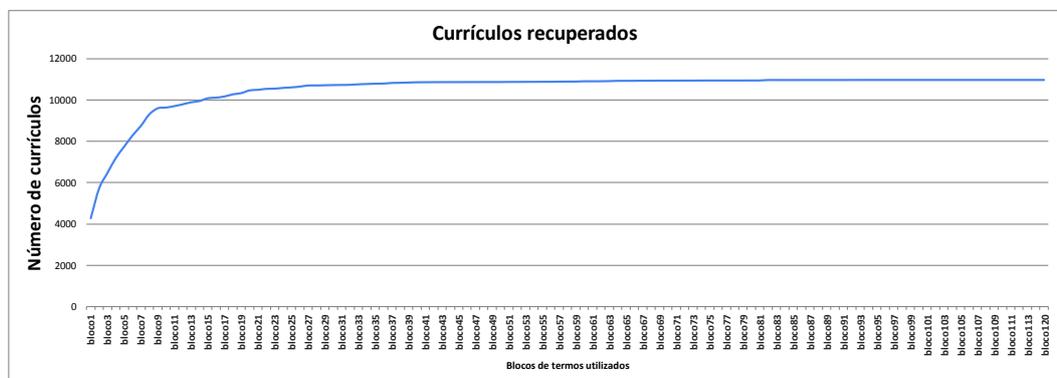
Bloco de palavras 1: (TERMO1 singular inglês OR TERMO1 plural inglês OR TERMO1 singular português OR TERMO1 plural português).

Foram identificados 120 blocos de palavras com 464 termos (no total) ao apresentá-los em inglês/português, singular/plural. Alguns termos foram desconsiderados e ou agrupados, pois não havia equivalente em inglês/português ou por serem cognatos.

Feito isso para todos os termos, buscou-se na base de dados de Currículos Lattes o número de recuperação de cada bloco de palavras individualmente. Após esse levantamento foi possível ordenar em ordem decrescente os blocos que apresentaram os maiores resultados.

Posteriormente foi construída uma expressão com todos estes termos ligados pelo operador booleano OR, pois constatou-se discrepâncias nos resultados quando se utiliza mais que um operador booleano. Desse modo, o BLOCO 1 foi adicionado ao BLOCO 2 pelo conectivo *or* e assim sucessivamente. Tal expressão foi utilizada para recuperação dos currículos dos pesquisadores doutores brasileiros e estrangeiros (a tabela que estruturou essa parte do trabalho pode ser consultada no Apêndice A). Entretanto, verificou-se conforme Gráfico 3, que a partir de determinado número de termos visualizava-se uma saturação da recuperação, pois poucos ou nenhum currículo era acrescentado com a adição de novos termos.

Gráfico 3 – Comportamento com o acréscimo de termos



Fonte: Elaboração própria

Lancaster (2004) utiliza como critérios para avaliar a qualidade do processo de recuperação de informações: o coeficiente de precisão e o coeficiente de revocação. O autor define o coeficiente de precisão como sendo a capacidade de evitar documentos inúteis no processo de recuperação de informação. O coeficiente de precisão (Co. Pre.) é medido pela proporção:

$$\text{Co. Pre.} = \frac{\text{No de itens relevantes recuperadas}}{\text{No de itens recuperadas pelo sistema}}$$

Quanto ao coeficiente de revocação (Co. Rev.), o autor define esse critério como a capacidade do sistema em fornecer todas os itens relevantes existentes, medido pela proporção.

$$\text{Co. Rev.} = \frac{\text{No de itens relevantes e recuperadas}}{\text{No de itens relevantes existentes no sistema}}$$

Esse critério é um pouco difícil de aplicar, pois para obter-se o número de itens relevantes existentes no sistema é preciso a realização de uma pesquisa muito genérica sobre um determinado assunto; porém, este não seria necessariamente o intuito normal de uma busca real (BOCCATO; FUJITA, 2006).

A expressão “saturação teórica” foi inicialmente utilizada por Glaser e Strauss (1967). É uma metáfora advinda da área físico-química que nessas ciências serve para descrever o comportamento de solutos junto a um meio solvente. Diz-se que uma solução está saturada quando a concentração do soluto é a máxima possível, considerando as condições físico-químicas em que esta solução se encontra (FONTANELLA; MAGDALENO JUNIOR, 2012).

A expressão de busca gerada com os 120 blocos de palavras, ao ser inserida no campo Busca Avançada na BCL, gerou um resultado inválido, devido à limitação do tamanho da URL (7.386 caracteres é o limite aceitável) de resultados da recuperação, sendo assim foi preciso readequar a expressão de busca, pois gerava uma URL de 24.692 caracteres e desse modo não possibilitava a visualização dos resultados da busca realizada.

Constatou-se que o bloco de palavras associado à nanopartícula e suas variações em inglês e português, plural e singular permitiu obter o maior número de currículos, ao todo 4.267.

Nesse sentido, para reduzir o tamanho da sequência de caracteres, optou-se em aplicar o seguinte filtro: corte de todos os blocos e termos de busca que haviam retornado menos de 3% em relação ao resultado do bloco de maior resultado de recuperação. Como o bloco de maior resultado permitiu obter 4.267 currículos, o limiar considerado para desconsiderar blocos e termos foi de 128 currículos.

A seguir todos os blocos que apresentaram menos que 128 currículos foram retirados

da expressão. No total havia 120 blocos de expressões de busca, foram descartados 86 blocos (aproximadamente 72%)<sup>18</sup>. Foram mantidos 34 blocos (aproximadamente 28%).

Os blocos de palavras que restaram conformaram a expressão final de busca, apresentada no Quadro 7. Esta limiarização sistemática permitiu identificar 24.887 Currículos Lattes de doutores e demais pesquisadores cadastrados, brasileiros e estrangeiros.

Quadro 7 – Expressão de busca

(nanoparticle or nanoparticles or nanopartícula or nanopartículas or nanotechnology or nanotechnologies or nanotecnologia or nanotecnologias or nanostructured or nanostructureds or nanoestruturado or nanoestruturados or nano or nanos or nanocomposite or nanocomposites or nanocompósito or nanocompósitos or sol\_gel or sóis\_géis or nanotube or nanotubes or nanotubo or nanotubos or nanocatalysis or catalyts or nanocatálise or nanocatalisadores or quantum\_dot or quantum\_dots or ponto\_quântico or pontos\_quânticos or nanomaterial or nanomaterials or nanomaterial or nanomateriais or nanocrystal or nanocrystals or nanocrystal or nanocristais or nanocrystalline or nanocrystallines or nanocrystalino or nanocrystalinos or graphene or graphenes or grafeno or grafenos or nanowire or nanowires or nanofio or nanofios or quantum\_well or quantum\_wells or poço\_quântico or poços\_quânticos or nanoscience or nanosciences or nanociência or nanociências or nanofiber or nanofibers or nanofibra or nanofibras or nanocapsule or nanocapsules or nanocapsula or nanocapsulas or nanoparticulate or nanoparticulates or nanoparticulado or nanoparticulados or nanoemulsion or nanoemulsions or nanoemulsão or nanoemulsões or nanosphere or nanospheres or nanoesfera or nanoesferas or nanobiotechnology or nanobiotechnologies or nanobiotecnologia or nanobiotecnologias or nanopowder or nanopowders or nanopó or nanopós or fullerene or fullerenes or fulereno or fulerenos or nanoporous or nanoporoso or nanoporosos or dendrimer or dendrimers or dendrímtero or dendrímteros or nanoindentation or nanoindentations or nanoindentação or nanoindentações or nanoribbon or nanoribbons or nanofita or nanofitas or carbon\_nanostructure or carbon\_nanostructures or nanoestrutura\_de\_carbono or nanoestruturas\_de\_carbono or nanoclay or nanoclays or nanoargila or nanoargilas or nanorod or nanorods or nanohaste or nanohastes or nanomedicine or nanomedicines or nanomedicina or nanomedicinas or nanocluster or nanoclusters or nanocacho or nanocachos)

Fonte: Elaboração própria

Após finalizar tais procedimentos, finalmente, a expressão pode ser utilizada, uma vez que gerou uma *url* (com 7210 caracteres) válida em relação ao número aceitável de caracteres. A coleta final foi realizada dia 27 de março de 2015.

### 3.2.3 Extração dos ID dos Currículos Lattes

Para extrair os dados dos Currículos Lattes, primeiramente foi necessário levantar uma lista de Identificadores (ID). Os identificadores únicos permitem a recuperação dos currículos, evitando a duplicidade e só posteriormente relaciona os dados de cada um deles.

Para este procedimento criou-se dois scripts computacionais programados em *Python* para extração dos IDs, denominados *ScriptExtract1* e *ScriptExtract2*. O primeiro requer a URL da busca avançada realizada na BCL (realizada com a expressão de busca criada no procedimento anterior) e o número total de currículos encontrados na busca da PL. O *ScriptExtract1* extrai todos os 10 dígitos que geram o ID de cada currículo identificado pela

<sup>18</sup> Esses currículos foram descartados, pois do universo final de 9.396 currículos em nanotecnologia, termos que retornaram menos de 3% de 4.267 foram considerados muito específicos, pois a busca não considera campos, mas sim o currículo como um todo.

busca da PL. Entretanto, foi necessário converter esses ID de 10 dígitos para ID com 16 dígitos (adequados como entrada para processamento do *ScriptLattes*) e esta é a função do *ScriptExtract2*, processar a conversão.

Por intermédio do Quadro 8 é possível visualizar o esquema em que a lista, resultado final dos procedimentos, está estruturada. Na primeira coluna encontra-se os números Identificadores de cada currículo e conseqüentemente de cada pesquisador, visto que o cadastro de um currículo está associado ao CPF do pesquisador. Na segunda coluna tem-se apenas uma vírgula por linha, que é o caractere utilizado para separação dos dados entre as colunas 1 e 3. A terceira coluna possui os nomes por extenso de cada pesquisador.

Quadro 8 – Estrutura lista final.

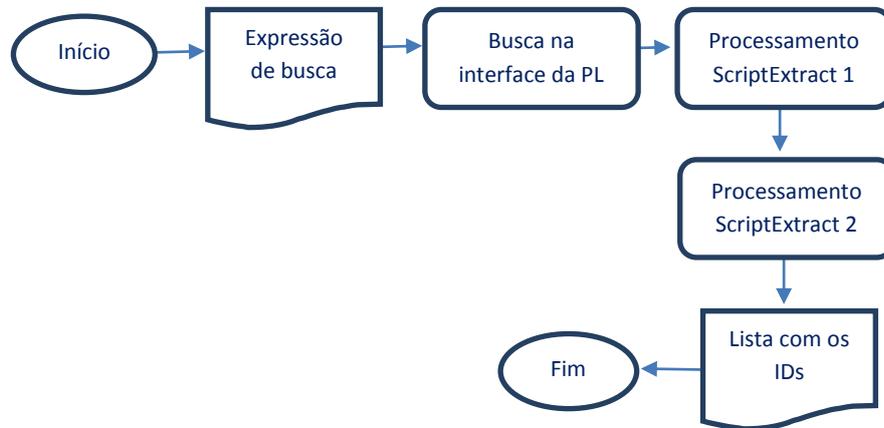
Nº Lattes Identificador		Nome do pesquisador
0000000000000000	,	XXXXXX XXXXXX XXXXXX
0000000000000000	,	XXXXXX XXXXXX XXXXXX
0000000000000000	,	XXXXXX XXXXXX XXXXXX
...		...

Fonte: Elaboração própria

O resultado deste procedimento é uma lista com os Identificadores (IDs) de Currículos Lattes extraídos da BCL. A Figura 4 apresenta os cinco passos realizados no procedimento:

- 1) elaboração de uma expressão de busca adequada para a BCL;
- 2) utilização da Busca Avançada na BCL, ver Anexo 2;
- 3) elaboração e uso de um *script* programado em *Python* (*ScriptExtract1*);
- 4) elaboração e uso do *ScriptExtract2* que converte os IDs de 10 dígitos em IDs de 16 dígitos, que servem como entrada para o *software* livre *ScriptLattes*; e
- 5) a obtenção da “lista final” com todos os IDs dos pesquisadores que possuem a presença de pelo menos um dos termos da expressão de busca em alguma parte de seus currículos.

Figura 4 – Procedimentos para coleta de ID



Fonte: Elaboração própria

### 3.2.4 Criação dos filtros

Para melhorar a precisão e diminuir a revocação dos dados coletados foi necessário a criação de alguns filtros, é importante ressaltar que os filtros dependem das perguntas ou objetivos referentes aos dados, nesse sentido apresenta-se algumas possibilidades. Outros filtros poderão ser criados de acordo com as futuras necessidades, como por exemplo filtros que selecionam apenas os pesquisadores de determinado estado ou região.

Nesse sentido, o procedimento 3 refere-se as possibilidades de filtros em relação ao refinamento dos dados. A primeira possibilidade é utilizando as opções da própria base, como apresentado na Quadro 9:

Quadro 9 – Critérios para busca na PL

<b>Critérios Identificados na Plataforma Lattes</b>	<b>Nº de Currículos</b>
Doutores e demais pesquisadores brasileiros e estrangeiros	24887
Doutores brasileiros e estrangeiros	12773
Doutores brasileiros	11593
<b>Doutores brasileiros e estrangeiros com grupo de pesquisa</b>	<b>9396</b>
Doutores brasileiros e estrangeiros com bolsa produtividade	2543
Doutores brasileiros e estrangeiros com grupo de pesquisa e bolsa produtividade	2386

Fonte: Elaboração própria com base na BCL

O primeiro filtro estabelecido considerou apenas doutores presentes em grupos de pesquisa (opção em negrito na Quadro 9), tal escolha deve-se ao fato de que pesquisadores doutores serem mais atuantes e os que tendem a publicar mais na área, o critério relacionado a pertencer a um grupo de pesquisa revela pesquisadores atuantes no coletivo e que permitem maior colaboração entre os participantes do grupo. E também por que, posteriormente, desejava-se levantar os pesquisadores participantes de Grupos de Pesquisa relacionados com a área de nanotecnologia.

Até este momento a amostra geral da pesquisa estava composta por 9.396 pesquisadores (doutores brasileiros e estrangeiros com grupo de pesquisa) o que extrapola muito a média apontada pelo MCTI (mais de 2.500 pesquisadores, ver Anexo 3)<sup>19</sup>. Isso se explica pelo fato da busca avançada na Base de Currículos Lattes não permitir a escolha do campo onde se deseja realizá-la, sendo assim, as palavras utilizadas na expressão de busca podem ser recuperadas de qualquer parte do currículo como em bancas em que o pesquisador participou, palestras que frequentou, etc.

Para resolver esse problema elaborou-se outro script, o *Scriptkeywords*, que utilizou a lista com os 9.396 IDs extraídos e as palavras-chave da expressão de busca como entrada e, com a sua aplicação foi possível verificar na Base de Currículos Lattes a presença dessas palavras-chave num campo específico. Nesta pesquisa utilizou-se o campo dos títulos dos artigos completos publicados em periódicos.

Com o processamento do *Scriptkeywords* foi possível verificar: o número de artigos completos em periódicos com pelo menos um termo da expressão de busca presente no título; e o número total de artigos completos em periódicos de cada pesquisador. Com esse filtro (utilização da checagem do *Scriptkeywords*) foi possível verificar que dos 9.396 pesquisadores recuperados com os filtros da BCL: 1.264 possuem pelo menos quatro artigos com um ou mais termos da expressão de busca; 443 possuem três artigos com um ou mais termos da expressão de busca; 813 possuem dois artigos com um ou mais termos da expressão de busca; 2.009 possuem um artigo com um ou mais termos da expressão de busca; e 4.866 não possuem nenhum artigo com os termos da expressão de busca.

Ou seja, considerando-se apenas aqueles que possuem no mínimo 2 artigos, chegou-se em um conjunto com 2.520 pesquisadores que é a média apontada pelo MCTI. Percebe-se que pelo fato da BCL não permitir uma busca avançada com delimitação dos campos, há pouca

---

<sup>19</sup> Disponível em: [http://desenvolvimento.gov.br/arquivos/dwl\\_1336677327.pdf](http://desenvolvimento.gov.br/arquivos/dwl_1336677327.pdf). Acesso em: 25 abr 2015.

precisão, pois apenas 27% dos resultados são satisfatórios, recuperou-se cerca de 73% de ruídos (informações inválidas).

Visando maior precisão, criou-se um filtro para os pesquisadores doutores presentes em grupo de pesquisa ligado a área analisada (nanotecnologia). Para tanto, foi necessário programar o quarto *script*, o *ScriptGP*, que permite mediante a interface de busca do Diretório de Grupos de Pesquisa (DGP), levantar todos os grupos de pesquisa dos pesquisadores que compõe uma lista, em seguida levantar todos os pesquisadores pertencentes a cada grupo recuperado no passo anterior.

Há também outro procedimento de filtragem, baseado no critério de participação em INCT da área, mas não foi realizado neste trabalho. Para tanto, seria necessário levantar todos INCT da área de nanotecnologia e seus pesquisadores constituintes, depois comparar com os pesquisadores já recuperados pelo *ScriptLattes*. Contudo, neste procedimento observou-se que as informações disponíveis sobre os INCTs não estão integradas em uma única fonte, os sites individuais dos INCT em nanotecnologia não são padronizados e não fornecem informações completas sobre os membros participantes. Assim, não foi possível concluir este filtro. Observou-se mais uma vez, a importância da BCL no que diz respeito a integralização e completude das informações sobre C&TI no âmbito nacional.

O último procedimento de filtragem utilizou técnicas de análise de rede, por meio do *software* livre *Gephi* e com base nos conceitos de coautoria<sup>20</sup>. O arquivo de entrada para este *software* é um dos arquivos resultantes do *ScriptLattes* (arquivo com extensão GDF), que possibilita análises gráficas e estatísticas baseadas em critérios de coautoria (colaborações ver Figura 12).

### **3.3 Desenvolvimento e aplicação da sistemática para mapeamento de áreas prioritárias**

Os procedimentos utilizados para o desenvolvimento e aplicação da sistemática em áreas prioritárias para o desenvolvimento científico e tecnológico do Brasil, são apresentados em cada subseção a seguir. A sistemática consiste em dois grandes macros procedimentos: 1) a coleta; e 2) o tratamento das informações que são apresentados juntamente com os resultados e análises. A seguir, apresenta-se todos os procedimentos que contemplaram a coleta das informações.

---

<sup>20</sup> A coautoria ou colaboração ocorre quando dois ou mais cientistas, trabalhando juntos em um projeto de pesquisa, compartilham recursos intelectuais, econômicos e/ou físicos. A coautoria é amplamente aceita no mundo científico e até estimulada por colocar diferentes autores, institutos e instituições em contato para enfrentar grandes problemas (GARCIA *et al.*, 2010).

### 3.3.1 A coleta dos dados e informações e as ferramentas utilizadas

Os dados foram coletados de quatro diferentes fontes de informação foram elas: Base de Currículos Lattes, Diretório de Grupos de Pesquisa, BDTD e Espacenet. Essas fontes de informação foram selecionadas com base em dois critérios: 1) oferecer a possibilidade de integrar dados de contextos diferentes (pessoas/ pesquisadores, grupos de pesquisas, teses/ dissertações e patentes) para uma análise com múltiplas dimensões; 2) por permitir de algum modo, o acesso aberto aos dados e informações.

Para Oliveira e Ferreira (2009, p.70) o conceito de fontes de informação diz respeito a “documentos, pessoas ou instituições que fornecem informações pertinentes a determinada área, fatores essenciais para se produzir conhecimento”. A partir daí, infere-se que a função das fontes seria reduzir incertezas e contribuir com a construção de novos saberes individuais e coletivos (VIEIRA & PINHO, 2015).

Para este trabalho ressalta-se as fontes eletrônicas, pois acredita-se que seja apenas uma questão de tempo para que o formato eletrônico prevaleça sobre o impresso (SILVA & TOMAÉL, 2004). Segundo Choo (1998) a internet merece destaque devido a sua importância no processo de disseminação da informação, ou seja, fontes eletrônicas que se utilizam da *Internet* como meio de transmissão de informações, além da própria *Internet* ser considerada, também, uma fonte de informação. Ainda, as bases de dados on-line ou em CD-ROM são citadas como importantes fontes eletrônicas de informação que, possibilitaram a troca de informações e acesso mais rápido entre pesquisadores, no campo da ciência.

De acordo com Souza (2006) a Recuperação da Informação (RI) é uma maneira de obter dados de forma seletiva através do uso de expressões de busca destinadas a identificar os registros que cumprem uma determinada condição, retratada na armazenagem e dispositivos lógicos para confrontar os dados que se deseja recuperar, com os dados armazenados, bem como, métodos e diretrizes para simbolizar o conhecimento e as possíveis estratégias de busca.

Para acessar os dados de um sistema de base de dados há necessidade de um Sistema de Recuperação da Informação (SRI) que suporte e responda adequadamente as pesquisas realizadas pelos usuários. De acordo com Vieira e Corrêa (2011) SRI são programas que respondem as ausências de informação dos usuários e corroboram com a RI pertinente em grandes conjuntos de documentos. Para os autores os entraves centrais da RI são: o excesso de informação oferecido nas buscas realizadas; as expressões inadequadas

utilizadas na pesquisa pelo usuário; e a dificuldade do sistema em absorver a necessidade expressa e apresentar a informação adequada.

Algumas dessas dificuldades se relacionam, segundo Zhang (2008), aos métodos de organização e armazenamento de informações, pois estes afetam e determinam os métodos de recuperação de informação, advindas de bases de dados que armazenam grandes quantidades de dados e que fornecem metadados para o processo de mineração de dados. Daí a importância dos indicadores de ciência e tecnologia que se utilizam das fontes de informação disponível para geração de novas informações e conhecimentos aplicados, culminando no auxílio ao desenvolvimento da ciência e da tecnologia e dos mecanismos de avaliação para os processos de tomada de decisão e planejamento estratégico em PCTI.

#### 3.3.1.1 O ScriptLattes

A base de Currículos Lattes da Plataforma Lattes dispõe de convênios no qual fornece às Instituições públicas de Ciência e Tecnologia um acesso direto, por meio de um *web service*, para recuperação dos currículos em formato XML, que compreendem todas as informações registradas pelos pesquisadores.

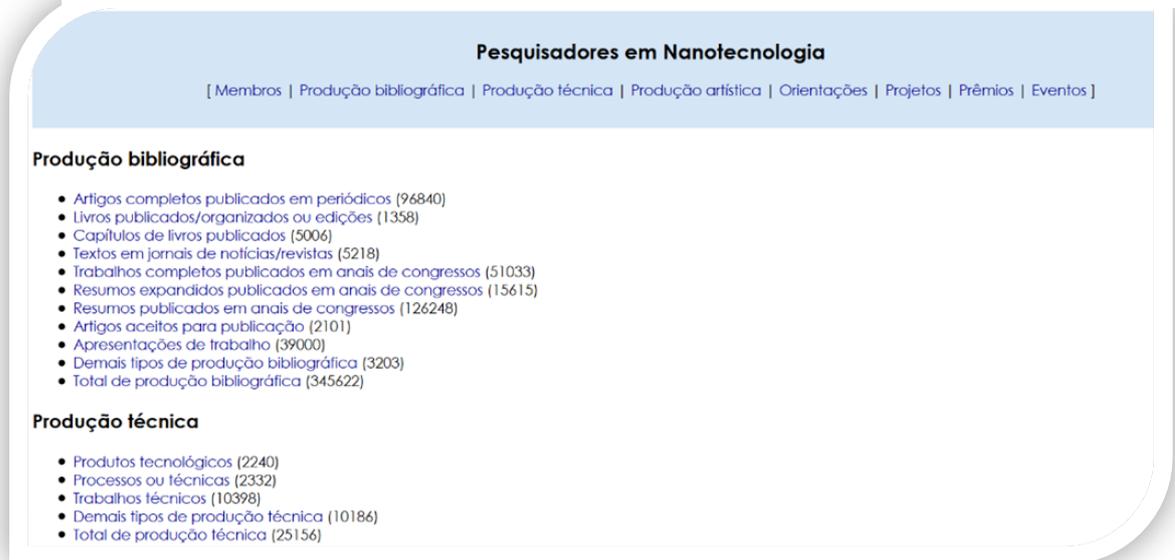
Os dados recuperados pelo *web service* estão em formato bruto, e são disponibilizados de forma oficial com a finalidade de acesso automatizado (MATIAS, 2016). Para isso, ressalta-se a necessidade da configuração de um *proxy* para autenticação e identificação do acesso, que normalmente é permitido, por exemplo, ao reitor da Instituição de Nível Superior, cuja permissão para *download* é recomendada apenas para os dados dos currículos dos pesquisadores da instituição solicitante do convênio. Ao limitar o acesso, tal procedimento consolida-se como mais uma das barreiras para recuperação de informações por assunto e área na referida base. Cabe ressaltar que o acesso disponibilizado para o cidadão comum permite o *download* de um currículo por vez.

As Agências de Fomento ou Fundações Estaduais de Apoio à Pesquisa têm a possibilidade de espelhamento dos dados da BCL para uma base de dados de armazenamento próprio, permitindo a replicação total dos dados. Em resumo, pode-se constatar que as Instituições de Nível Superior não podem analisar áreas temáticas nacionais, uma vez que só tem acesso aos dados referentes a elas mesmas, já as Agências de Fomento, embora possam, requerem duplicação de recursos e esforços operacionais para armazenamento, manuseio e tratamento dos dados espelhados.

Diante do exposto, o presente trabalho utilizou-se do ScriptLattes, a alternativa mais adequada e acessível para a extração dos dados da Base de Currículos Lattes. O *ScriptLattes*<sup>21</sup>, é um *software* que compila as listas de produções, tratando apropriadamente as produções duplicadas e similares. Logo após, são gerados relatórios, em formato HTML, com listas de produções e orientações separadas por tipo e colocadas em ordem cronológica invertida (MENA-CHALCO & CESAR-JR, 2013). Ou seja, tal *software* permite a criação de relatórios acadêmicos de forma automática, considerando apenas as informações cadastradas nos Currículos Lattes. É um programa computacional que proporciona a análise de um conjunto de pesquisadores em cima de um assunto de interesse. Os resultados permitem analisar um conjunto de valiosos indicadores de produção acadêmica, bem como relatos de produção, coautoria, geoposicionamento, etc<sup>22</sup>. O arquivo de configuração com os parâmetros da execução do referido script encontra-se no Anexo 4.

Nessa etapa foram extraídos os dados dos currículos (em formato HTML) de um grupo de pessoas de interesse, utilizou-se a lista com os 2520 pesquisadores, utilizando o ScriptLattes, a Figura 5 apresenta o layout de visualização dos relatórios gerados.

Figura 5 – Visualização dos resultados do ScripLattes



Fonte: VLAB4U (2015)<sup>23</sup>

<sup>21</sup> <http://ScriptLattes.sourceforge.net/> (MENA-CHALCO, 2009)

<sup>22</sup> Para a compilação precisa-se de alguns módulos Python. Para instalar esses módulos deve-se executar como root: # apt-get install python-all python-setuptools python-utidylib python-matplotlib python-levenshtein python-pygraphviz python-numpy tidy python-scipy python-imaging # easy\_install pytidylib

<sup>23</sup> Disponível em: <<http://vlab4u.info/nanocore/>>. Acesso em: 20 ago. 2015.

### 3.3.1.2 O ScriptGP

O Diretório de Grupos de Pesquisa (DGP) reúne as informações de todos os grupos de pesquisa cadastrados no CNPQ. O cadastro desses grupos permite o registro de informações, como: as áreas de atuação; linhas de pesquisa; pesquisadores; estudantes; técnicos; instituições; equipamentos e outros dados sobre um grupo de pesquisa.

O ScriptGP (script que extrai informações do DGP) é na verdade um conjunto de scripts em linguagem de programação *Python* que mediante uma lista de entrada (com os IDs Lattes ou IDs dos Grupos de Pesquisa), processa extrações específicas dos dados do Diretório de Grupos de Pesquisa (DGP) do CNPq, mediante um arquivo de configuração<sup>24</sup>. Ao final o programa gera uma página HTML com todos os dados no formato *PivotTable*, permitindo a análise e cruzamento dos dados. Descreve-se abaixo os *ScriptsGP* utilizados nesta pesquisa:

- O ScriptGP-busca: realiza uma busca por palavras-chave na plataforma de grupos de pesquisa do CNPq e retorna o IDGP de cada grupo encontrado. Assim como nos pesquisadores, o IDGP é um identificador único para cada grupo de pesquisa e permite localizar e recuperar todas as informações do grupo na base do CNPq.
- ScriptGP-cadastro: permite recuperar os principais dados cadastrais dos grupos como, por exemplo, ano de formação, instituição, áreas de conhecimento, cidade, UF, e dados do número de pesquisadores, estudantes e técnicos que pertencem ao grupo. O Script usa como base uma lista de IDGP, obtida por meio da *ScriptGP2-busca* ou *ScriptGP2-lattes* para acessar os grupos e extrair as informações. O programa também gera uma página HTML com todos os dados no formato de uma *PivotTable*, permitindo a análise e cruzamento dos dados.
- *ScriptGP-pesquisadores*: permite recuperar o IDLATTES e nome de todos os pesquisadores pertencentes a um grupo de pesquisa a partir de uma lista de IDGPs. O script também gera uma *PivotTable* com os grupos e pesquisadores.
- *ScriptGP-estudantes*: permite recuperar o IDLATTES e nome de todos os estudantes pertencentes a um grupo de pesquisa a partir de uma lista de IDGPs. O script também gera *PivotTable* com os grupos e pesquisadores.

---

<sup>24</sup> O arquivo de configuração é um arquivo com a extensão “.config” que contém linhas de comando ajustáveis para o processamento do script, um exemplo de seu conteúdo é:

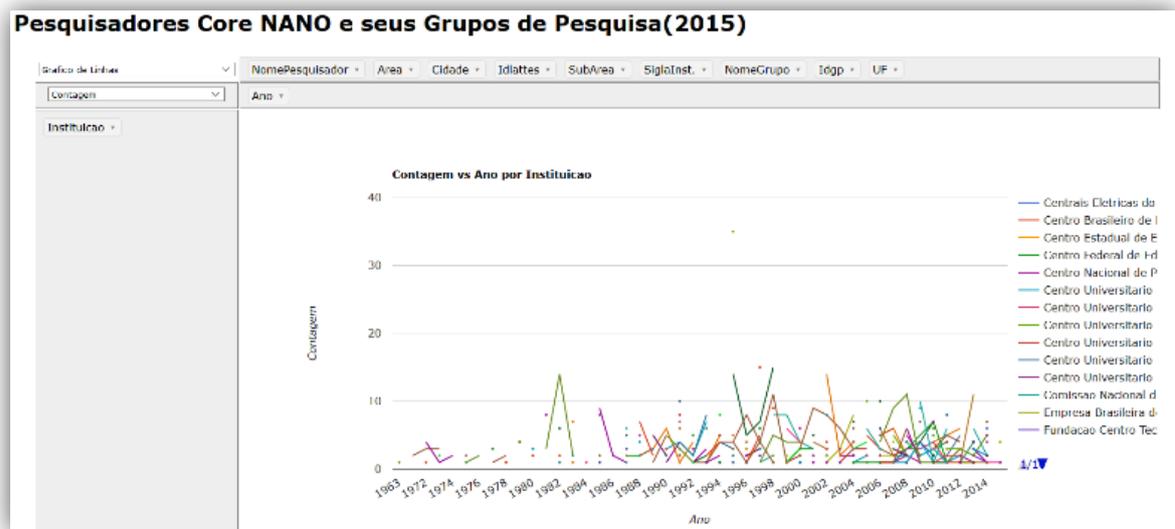
```
sep = \t
arquivolistabusca = C:\Users\Public\Softs\Scripts\ScriptGP20\nanoAline\lista-busca-nano.txt
arquivosaida = C:\Users\Public\Softs\Scripts\ScriptGP20\nanoAline\nano01052015
repercussoes = sim
```

- *ScriptGP-instituições*: permite recuperar as instituições (empresas ou outras organizações) que possuem alguma parceria com os grupos de pesquisa a partir de uma lista de IDGPs. O script recupera o nome, razão social, CNPJ, sigla, endereço, cidade, UF, natureza jurídica e setor de atuação das empresas.

É importante salientar, que além da saída com a extensão TXT um outro resultado dos scripts é um arquivo com a extensão JSON que pode ser processado pelo *ScriptGP-pivottable* e que gera uma visualização HTML automaticamente, como a que pode ser visualizada na Figura 6, a seguir.

O conjunto de *ScriptsGP* foram desenvolvidos especialmente para as pesquisas dos dois grupos da autora (NIT-Materiais e Vlab4u), mas eles são livres e estão disponíveis para download em: <https://bitbucket.org/vlab4u/>.

Figura 6 – Pesquisadores do core em nanotecnologia e seus Grupos de Pesquisa



Fonte: VLABU (2015)

### 3.3.1.3 A extração BDTD

Ao incluir a BDTD o objetivo foi reunir as teses sobre nanotecnologia já defendidas, para um estudo de colaboração entre os membros das bancas e identificação do comportamento das temáticas ao longo do tempo.

Na BDTD (Biblioteca Digital Brasileira de Teses e Dissertações) até o momento da defesa ainda não havia uma possibilidade para extração dos dados. Embora já haja uma importante iniciativa em processo de desenvolvimento por uma equipe de servidores e

pesquisadores do IBICT, um API que permitirá baixar os metadados da BDTD de maneira rápida, eficiente e com qualidade em diversos formatos (como JSON, CSV, Bibtex entre outros).

No momento presente, uma possibilidade acessível foi a de baixar alguns metadados em formato CSV utilizando a opção de URL da busca e substituir o termo Results por Exportcsv<sup>25</sup>. Entretanto, por ser uma alternativa provisória, nesta opção não foi possível baixar as 1.260 teses encontradas, pois para não provocar instabilidade na base há a limitação para o *download* de apenas um mil registros. Desta forma, a alternativa foi solicitar a equipe do IBICT que preparasse um arquivo (extensão CVS) com a seguinte expressão de busca: nano\* - presente nos títulos e palavras-chave de teses e dissertações.

O arquivo CSV foi recebido em 25/10/2016 e continha os seguintes campos: id, oai Identifier, instituição, spelling, language, tipo, acesso, autor, autor2, título, topic, description, data e url.

Entretanto, o arquivo possuía vários problemas, como: falta de padronização no preenchimento dos campos, onde cada instituição compreendeu a informação de preenchimento de modo diferente, por exemplo para o campo “autor 2” algumas preencheram com o nome do orientador, outras preencheram com os nomes de todos os membros da banca; o campo “topic” (assunto) não possuía separador entre as palavras-chave, desta maneira em alguns casos é impossível utilizar a informação, como por exemplo, “TIORIDAZINACLORPROMAZINATRIFLUOPERAZINATHIORIDAZINECHLORPRO MAZINETRIFLUOPERAZINE”. Por estas razões não foi possível elaborar todos os indicadores pretendidos à BDTD, como a rede de colaboração entre os membros de bancas, por exemplo.

#### 3.3.1.4 O Patent2Net

O Patent2Net (P2N) é um software livre com fonte aberta que utiliza o acesso ao *European Patent Office* (EPO) e seus serviços gratuitos *online* para pesquisar patentes. O *European Patent Office* (Escritório Europeu de Patentes) tem uma base de dados de mais de 90 milhões de patentes cujos conteúdos estão descritos em detalhe. É um concentrado de conhecimento tecnológico humano de aproximadamente 90 países. O diferencial do

---

<sup>25</sup> Exemplo: <http://bdtb.ibict.br/vufind/Search/Results?lookfor=&type=AllFields>. Troca Results por Exportcsv: <http://bdtb.ibict.br/vufind/Search/Exportcsv?lookfor=&type=AllFields>

Patent2Net se traduz na facilidade em analisar dados de patentes, nele é possível fazer o *download* e analisar resumo, tabelas e seções de documentos de patentes.

O Patent2Net é um rastreador aberto e gratuito desenvolvido e utilizado por uma equipe internacional em torno de vinte pesquisadores, reunindo professores e estudantes de muitos países: Brasil, Argélia, Senegal, França, Suíça e Portugal. A ferramenta mobiliza desde desenvolvedores até simples usuários, com um objetivo em comum, estudar as demandas de patentes em todo o mundo. Para uma adequada utilização do P2N alguns outros softwares são requeridos, são eles: JAVA, Graphviz, GEPHI, IraMuTeQ; Freeplane; e navegador Mozilla Firefox.

O P2N incorpora diferentes ferramentas para processar e visualizar os dados coletados, são eles:

- *DataTable*: apresenta uma lista de patentes coletadas em formato de tabela com todos seus metadados (país, título, inventor, recorrente, IPC, dentre outros). Tem a possibilidade de excluir ou adicionar campos através do botão "Mostrar / ocultar colunas". Além disso, nele é possível rolar a lista de resultados de patentes pelos botões próximo, anterior, primeiro, último.
- *PivoTable*: mostra os dados bibliográficos de patentes selecionados. É possível escolher as variáveis e um tipo de visualização (mapa de calor, histograma, curva, dentre outros). Também é possível mudar as cores dos gráficos, tamanho da fonte do título, dentre outras formatações.
- *Worldmap*: permite a visualização do país de origem dos inventores e dos requerentes. Cartografias em países de origem dos inventores e requerentes, relativamente a patentes concedidas.
- Grafos de redes (inventores, recorrentes, tecnologia): as redes são apresentadas com grafos do *Gephi*. Nos resultados apresenta-se redes usando apenas uma variável, sendo elas: rede de inventores – mostra a relação entre inventores que trabalharam na mesma patente; rede candidatos – mostra a relação entre os requerentes que trabalharam nas mesmas patentes; rede tecnologias – exhibe os campos tecnológicos de patentes. Apresenta-se também as redes mistas que combinam muitas variáveis, tais como: tecnologias de países – mostra a especialidade (s) ou o campo tecnológico (s) de países; tecnologias de inventores – mostra os especialistas em uma ou várias áreas; tecnologias recorrentes – mostra a sociedade quem são especialistas em uma ou várias

áreas; inventores-recorrentes: mostra as relações entre inventores e requerentes (depositantes).

- Redes de referências, citações, equivalentes: referências - exibe as patentes que citaram outras patentes ou referências externas (jornais, teatro de revista, etc). Existem dois tipos de redes: *Outdegree* – patentes que utilizam uma ou várias outras patentes para a melhoria ou nova versão; *Indegree* – patentes importantes, patentes-chave.
- Redes de citações: exibe as citações sobre patentes: *Indegree* – patentes mais citadas (patentes citadas por outras patentes); *Outdegree* – patentes que citam as mais citadas (patentes que citam outras patentes).
- Patentes de famílias: mostra a lista de famílias de patentes.

O projeto como um todo pode ser acessado em <http://patent2netv2.vlab4u.info/>, onde também é possível conhecer outros casos tratados com P2N. Para acessar e baixar P2N é possível em: <https://github.com/patent2net>. Um exemplo do arquivo de configuração requerido pelo P2N pode ser visualizado no Anexo 5.

As figuras 7, 8 e 9 apresentam o sumário para acessar os dados coletados e tratados com o P2N (dados *on line* e para *download*).

Figura 7 – Resumo das principais informações de uma busca P2N

**Informations:**

- Data directory: Nanofiber
- Request: ta = nanofiber\* and txt = nanotub\*
- Number of patents retrieved: 790
- Family lenght:1614
- Gathering date: 01, Oct 2016
- Abstract: 95 (FR) 5 (DE) 695 (EN) 31 (OL)

Fonte: VLAB4U (2016)

Figura 8 – Análises on-line de ferramentas do P2N

**On-line analysis tools:**

- [Patents datatable, Pivot table](#)
- [Attractivity: Geolocalisation of patent covering \(without EP, WO\), Applicants, Inventors \(when available\)](#)
- [Networks \(Inventor, Applicant, Technology\)](#)
- [Mixed Networks \(Country-Technology, Inventor-Technology, Applicant-Technology, Applicant-Inventor\)](#)
- [Equivalents, Reference \(References to other patents or External references\), Patents citations networks](#)
- [Patent family & Pivot table](#)
- [IPC's Mind-Map \(FreePlane PlugIn\)](#)

Fonte: VLAB4U (2016)

Figura 9 – Dados para tratamentos em outras ferramentas

### Download data

TIP: use "right-click" and "save as" on links!

**Gephi compatible network files (gexf format):**

<a href="#">Applicants</a>	<a href="#">Inventors</a>	<a href="#">Technologies (IPC7)</a>
<a href="#">Citations</a>	<a href="#">References</a>	<a href="#">Equivalents</a>
<a href="#">Countries and technologies</a>	<a href="#">Inventors and applicants</a>	<a href="#">Inventors and technologies</a>
		<a href="#">Applicants and technologies</a>

**FreePlane compatible file**

- [IPCc Mindmap](#)

**Iramuteg tagged abstracts (unicode UTF-8 encoding):**

<a href="#">Abstracts (EN)</a>	<a href="#">Claims (EN)</a>	<a href="#">Descriptions (EN)</a>
<a href="#">Families Abstracts (EN)</a>	<a href="#">Families Claims (EN)</a>	<a href="#">Families Descriptions (EN)</a>
<a href="#">Abstracts (FR)</a>	<a href="#">Claims (FR)</a>	<a href="#">Descriptions (FR)</a>
<a href="#">Families Abstracts (FR)</a>	<a href="#">Families Claims (FR)</a>	<a href="#">Families Descriptions (FR)</a>

**Carrot2 clustering engine XML compatible files:**

- [Abstract XML \(english only\)](#)
- [Abstract XML \(french only\)](#)
- [Abstract XML \(all others language: unknown\)](#)

**Download bibliographic format**

- [Biblatex format, Zotero compatible \(only A,B, C statuses\)](#)

Fonte: VLAB4U (2016)

A base Espacenet recebe dados de cerca de 90 outras bases (escritórios nacionais de cada país participante) e pode haver incompatibilidade de integração dos registros devido aos problemas de interoperabilidade. É importante que haja pelo menos título e resumo em inglês, pois a Espacenet é essencialmente nesta língua, há patentes brasileiras que não atendem esta exigência.

Um depositante (requerente) é uma pessoa ou organização que entrou com um pedido de patente. Pode haver mais do que um requerente por aplicação. O requerente também pode ser o inventor. O formato do requerente para uma pessoa é constituída por o sobrenome seguido do nome (s).

Cabe ressaltar que as palavras utilizadas na expressão de busca do Lattes representam conteúdos genéricos da área como, por exemplo as formas em que se pode encontrar os materiais: argila, pó, fita, tubo, fibra, cristal, dentre outras. Como a base de Currículos Lattes é uma base peculiar, esta estratégia de busca é mais adequada. Ao se pesquisar por patentes a lógica é outra, ou seja, é especificar em níveis menos abrangentes a aplicação da tecnologia em questão. Para a Espacenet o ideal é realizar buscas utilizando o Classificação Internacional de Patentes (CIP).

A expressão de busca realizada na recuperação de patentes deve ser específica, pois a patente é um documento complexo. A busca com palavras-chave não foi uma boa alternativa, pois a Espacenet recebe dados de diversas outras bases dos escritórios nacionais, incluindo o

Brasil, e nem todos enviam os dados dos títulos e *abstract* em inglês, o que poderia causar a não recuperação destes documentos. Neste sentido, optou-se em realizar a busca utilizando o ICP (Classificação Internacional de Patentes). O ICP utilizado está dentro da classe H - Eletricidade, mais precisamente H01G11/36 ligado a nanoestruturas, nanofibras, nanotubos ou fulerenos.

## 4 RESULTADOS E DISCUSSÃO

A presente seção compreende uma apresentação dos resultados dos objetivos propostos e discussão sobre eles. Apresenta também a sistemática desenvolvida neste trabalho para mapeamento de áreas prioritárias, além de algumas análises da área objeto de estudo, a nanotecnologia<sup>26</sup>. Após a coleta dos dados pode-se iniciar o tratamento e elaboração dos indicadores. Para subsidiar a elaboração de alguns indicadores (que não foram gerados de modo automático), utilizou-se o *Excel*, *Iramuteq*<sup>27</sup> e *Gephi*<sup>28</sup> para criar as possibilidades de visualização do corpus criados.

### 4.1 Sistemática para mapeamento de áreas prioritárias

Em meio a uma rede permeada por dados e informações com acesso aberto ou não, ainda há grandes dificuldades para coletá-los e tratá-los adequadamente. Nada obstante, se por um lado as novas tecnologias e os procedimentos de administração de informações e comunicação causaram este aumento no volume de informações disponíveis, é também através deles que provavelmente se encontrará as soluções para dominar e usufruir desta imensidão de dados (DOWBOR, 2003).

Nesta perspectiva, a mineração de dados, aqui representada pelos *scripts* desenvolvidos, surgem para que se tenha acesso e possibilidades de manipulação desses dados. Cada *script (software)* permitiu isoladamente ou associado com os demais, possibilidades de processar a conversão de dados em informação e conhecimento. Atualmente, tais processos de conversão são necessários em todas as esferas do processo de avanço científico e tecnológico, podendo ser mensurados mediante a construção de indicadores que permitem várias formas de visualização e representação das informações coletadas. E ao serem realizados com periodicidade e com fontes de informação de longo

---

<sup>26</sup> Em "<http://vlab4u.info/nano2014/>" é possível acompanhar os dados utilizados e outros resultados potenciais.

<sup>27</sup> Os dados utilizados para o corpus foram os títulos das publicações ou resumos. O Iramuteq é um programa de computador para análise textual baseado em estatísticas (utilizando o software estatístico R) que analisa relações e outras características textuais, o posicionamento, frequência e a estruturação de palavras no texto, de forma que sejam retornados indicadores e visualizações sobre a estrutura e ambientes do texto analisado. Há cinco tipos de análises no Iramuteq: estatísticas textuais, especificidades e AFC, CHD, análise de similitude, nuvem de palavras, cada uma com suas características e especificações. O Iramuteq faz as análises textuais pelas proximidades de relações se baseando na construção textual, mas não com base nos significados das palavras (semântica).

<sup>28</sup> BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi, an open source graph visualization and manipulation software. 2016. [s.l.: s.n.].

alcance (há dados de publicações desde 1964 na BCL, por exemplo), os indicadores podem propiciar uma utilização mais sistemática das evidências históricas para a formulação de política científica (BELOFF, 1968; CAMPBELL, 2010).

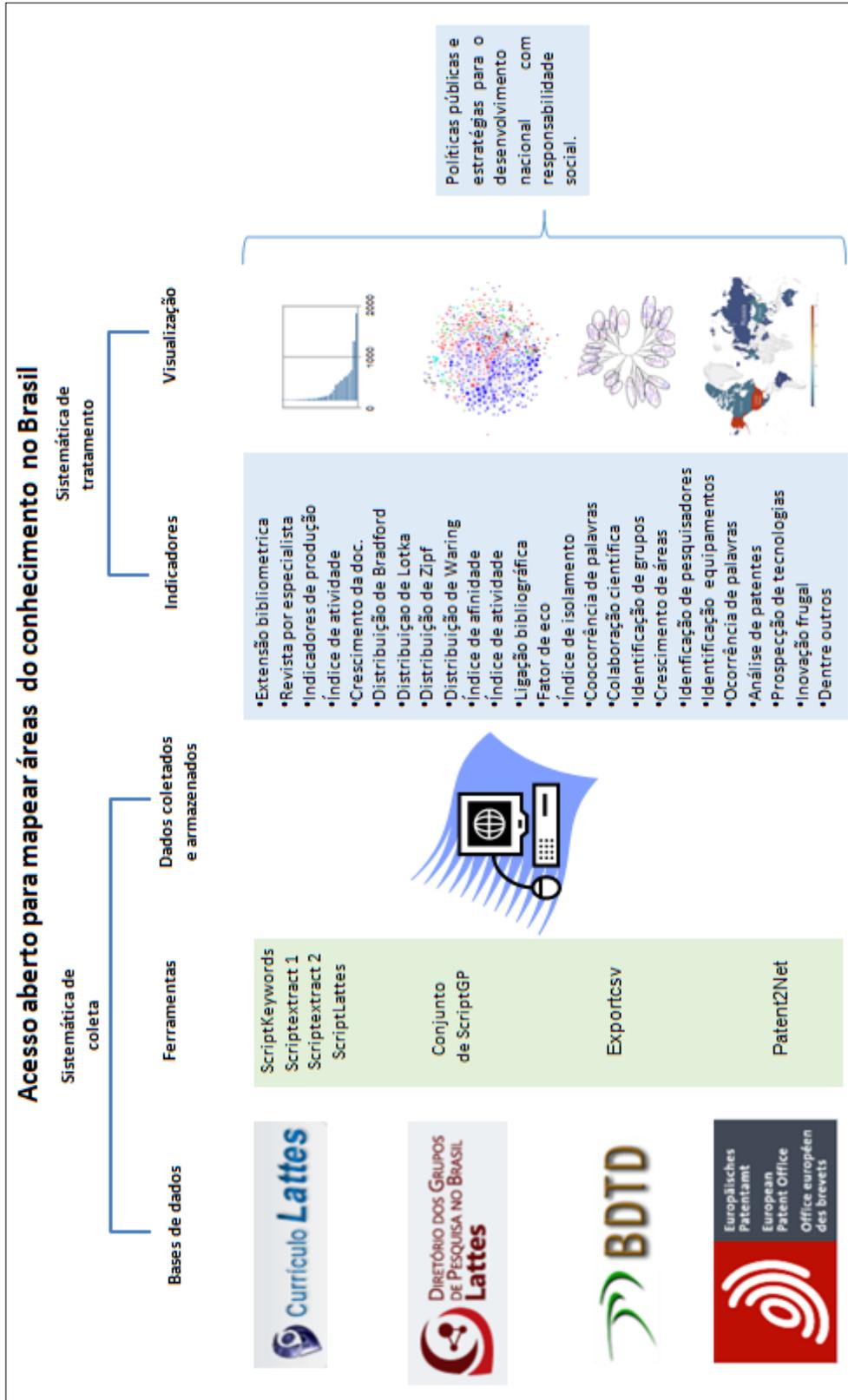
Ainda hoje no Brasil ocorrem mudanças nas formas de organização das atividades científicas e tecnológicas, nas instituições que produzem conhecimento, assim como nos instrumentos, processos e procedimentos que têm por objetivo planejar e gerir as atividades relacionadas à C&T. Todas elas têm gerado um grande grau de incerteza sobre os caminhos mais adequados para planejar, conduzir e avaliar as atividades científicas e tecnológicas (CGEE, 2008).

A Figura 10 sintetiza os principais procedimentos referentes a esta pesquisa na seguinte ordem: os scripts programados, os dados coletados, os potenciais indicadores e as possibilidades de visualização das análises. Nesta representação é possível compreender a dinâmica da principal proposta em questão (elaboração de uma sistemática), ou seja, facilitar a apropriação de informação e conhecimento para sanar necessidades reais de processos como o de tomada de decisão na esfera da C&T.

A análise automatizada é requerida pela necessidade da mineração e seleção de um abrangente conjunto de dados com objetividade (DOWBOR, 2003). É utilizada cada vez mais e principalmente em operações que envolvem grandes capacidades de processamento. O presente trabalho utilizou ferramentas livres para realização da extração de dados e informações de bases de dados de acesso aberto.

A análise humana também foi representada (na Figura 10) e contempla os procedimentos de análise e tratamento realizado por seres humanos, pois é exigida capacidade cognitiva mais sofisticada para criar conexões entre os dados, na qual ressalta-se o empenho de objetividade e foco à resolução de problemas, no caso deste trabalho são aqueles voltados para C&T (TAVARES; PIMENTEL; ARAUJO, 2012).

Figura 10 – Representação dos procedimentos realizados



Fonte: Elaboração própria

Ao passo em que se obtém a junção das análises automatizadas e humanas é possível ressignificar ou até mesmo encontrar novos sentidos para o objeto analisado mediante a construção de indicadores que viabilizam: as representações visuais e suas técnicas de visualização da informação; as análises estatísticas realizadas via softwares específicos em algumas análises (textuais por exemplo); as análises de redes de colaboração mediante a criação de grafos que ilustram conexões e suas frequências; dentre outros. A multiplicidade de indicadores oportunizam uma visão mais holística e integrada do tema abordado, favorecendo análises com maior robustez e completude.

Por fim, indicadores podem alterar o sistema através dos incentivos que estabelecem, e estes efeitos podem ser antecipados. Isso significa que é sempre preferível usar um conjunto de indicadores. O uso de um único indicador pode levar a erros de interpretação. Lembrando que a missão de uma pesquisa e os objetivos da avaliação mudam e o sistema de pesquisa também evolui. Assim, indicadores devem ser revisados e por vezes modificados (HICKS, 2015). Desta maneira todo o detalhamento das informações transformadas em indicadores em C&T apresentado por Aguiar (1991), bem como nesta sistemática, cumprem a função de subsidiar os desenvolvimentos em Ciência e Tecnologia.

Assim, ao se utilizar as fontes de informação e os procedimentos apresentados (Figura 10), descortinam-se oportunidades de apropriação de um conjunto singular de indicadores sobre C&T. É possível manusear dados e informações sobre: as pessoas que compõem a C&T com a Base de Currículos Lattes; o agrupamento dessas pessoas no Diretório de Grupos de Pesquisa; o conhecimento acadêmico pesquisados nas universidades com a Biblioteca Digital Brasileira de Teses e Dissertações; e as aplicações tecnológicas da área com a base de patentes Espacenet. A seguir apresenta-se um exemplo de como isso é possível na prática e para a área objeto deste estudo, a nanotecnologia.

#### **4.2 Mapeamento e análise de uma área prioritária: a nanotecnologia**

Todos os resultados aqui encontrados resultam da sistemática representada na Figura 10. Primeiramente serão abordados os indicadores construídos a partir dos dados e informações da Plataforma Lattes (Base de Currículos Lattes e Diretório de Grupos de Pesquisa), após serão aqueles relacionados com a BDTD (Biblioteca Digital de Teses e Dissertações) e por fim aqueles que utilizaram a base de patentes Espacenet do EPO (Escritório Europeu de Patentes).

#### 4.2.1 Indicadores da Base de Currículos Lattes

São apresentados os indicadores construídos a partir dos dados coletados com os scripts: *ScriptExtract1*, *ScriptExtract2*, *Scriptkeyword* e *ScriptLattes*, que possibilitaram a construção de uma planilha como esta, representada no Quadro 10, dentre suas saídas.

Quadro 10 – Saída do ScriptKeyword

	A	B	C	D	E	F	G
	Nome	Grande área de atuação	Estado	KW	KW	KW	KW
1	Paulo Cesar de Moraes	Ciências exatas e da terra	DF	nanoparti	nanoparti	FALSO	FALSO
2	Edson Roberto Leite	Engenharia de Materiais e Metalurgica	SP	nanoparti	nanoparti	FALSO	FALSO
3	Elson Longo da Silva	Química	SP	nanoparti	nanoparti	FALSO	FALSO
4	Ado Jorio de Vasconcelos	Física	MG	FALSO	FALSO	FALSO	FALSO
5	Antonio Gomes de Souza Filho	Física	CE	nanoparti	nanoparti	FALSO	FALSO
6	Marcos Assuncao Pimenta	Física	MG	FALSO	FALSO	FALSO	FALSO
7	Adriana Raffin Pohlmann	Farmácia	RS	nanoparti	nanoparti	FALSO	FALSO
8	Sílvia Stanisçuaski Guterres	Farmácia	RS	nanoparti	nanoparti	FALSO	FALSO
9	Almir Oliveira Neto	Química	SP	nanoparti	nanoparti	FALSO	FALSO
10	Aldo Felix Craievich	Física	SP	nanoparti	nanoparti	FALSO	FALSO
11	Cid Bartolomeu de Araujo	Física	PE	nanoparti	nanoparti	FALSO	FALSO
12	Estevam Vitorio Spinacé	Química	SP	nanoparti	nanoparti	FALSO	FALSO
13	Oswaldo Luiz Alves	Química	SP	nanoparti	nanoparti	FALSO	FALSO
14	Douglas Soares Galvão	Física	SP	nanoparti	FALSO	FALSO	FALSO

Fonte: Elaboração própria

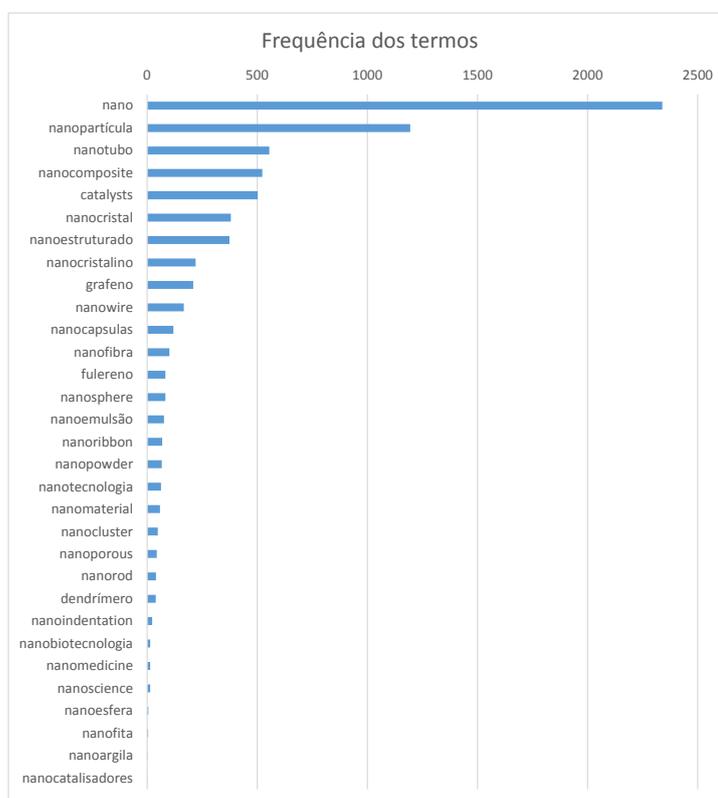
Após ordenar, em ordem decrescente, os pesquisadores que mais publicaram com os termos da expressão de busca utilizada, adicionou-se a coluna estado. Também foi acrescentado, nas células correspondentes, o nome do termo encontrado, quando o termo não foi encontrado estabeleceu-se a palavra “falso” para preencher a célula. Dessa forma, ao selecionar uma linha desta planilha tem-se importantes informações: nome do pesquisador, grande área de atuação, estado, e alguns termos presentes nas publicações dele.

No Gráfico 4 apresenta-se um ranking dos termos mais frequentes da expressão de busca. Foi mantido o radical nano, pois como pode-se perceber foi o termo que apresentou o maior número de frequência, e representa um termo bastante válido com relação ao aumento da revocação (LANCASTER, 2004). O segundo termo mais frequente foi “nanopartícula” e pode ser considerado o primeiro mais específico, iniciando a precisão da recuperação da informação (LANCASTER, 2004), uma vez que apenas o radical “nano” não poderia.

Os três termos subsequentes – nanotubos, nanocompósitos e catalysts – possuem frequência parecida e já elucidam temas mais específicos da área. É possível constatar que a área compreende a aplicação de diversos assuntos, corroborando com a necessidade de uma expressão de busca diversificada e bem elaborada. Segundo Aguiar (1991) este indicador de C&T ilustra por exemplo, os termos mais utilizados nas pesquisas correntes (aplicando os

dados dos projetos em andamento, por exemplo) e as pesquisas realizadas representadas pelos artigos publicados, que foi utilizado para compor o Gráfico 4.

Gráfico 4 – Termos mais frequentes da expressão de busca nos artigos



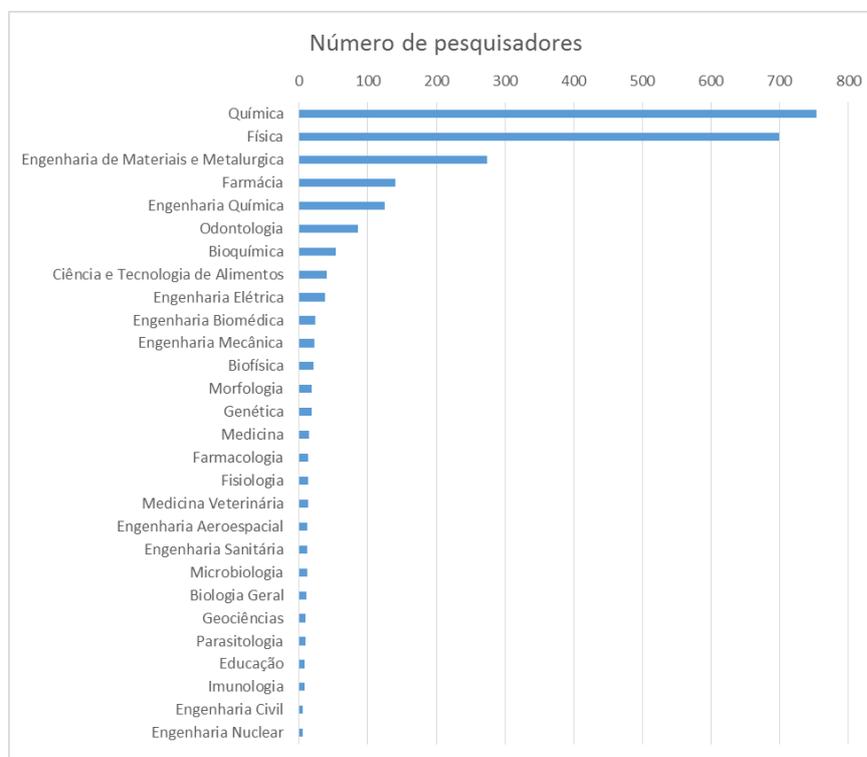
Fonte: Elaboração própria

No Gráfico 5 foi possível mensurar as grandes áreas de atuação dos pesquisadores em nanotecnologia, sendo a Química (com 30%) e a Física (com 27%) as duas grandes principais áreas, somando 1.453 pesquisadores, o que representou 57% da amostra, elucidando uma atuação fortemente relacionada a ciências exatas e naturais.

Para o restante das áreas, com representatividade de 26 % de atuação, nota-se uma aplicação dos conhecimentos advindos de modo bastante pulverizados, pois englobam diversas áreas: Engenharia Química, Odontologia, Bioquímica, Ciência e Tecnologia dos Alimentos, Engenharia Elétrica, Engenharia Biomédica, Engenharia Mecânica, etc. Segundo Ortellado (2008) são justamente as áreas de ciências exatas e naturais as que apresentam mais dificuldades na prática do acesso aberto. Prejudicando a prática das recomendações previstas no *Budapest Open Access Initiative* (BOAI) que trata da disponibilização livre e permanente,

na internet, sem restrições de acesso, de literatura de caráter científico ou acadêmico (BOAI, 2002).

Gráfico 5 – Áreas ligadas a nanotecnologia



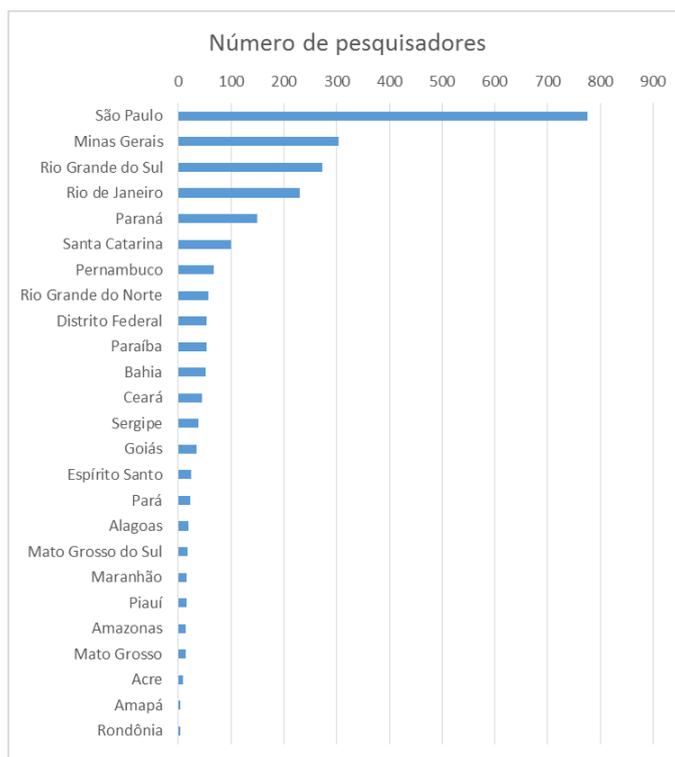
Fonte: Elaboração própria

De acordo com o Gráfico 6 foi possível mensurar as informações sobre a distribuição dos pesquisadores em nanotecnologia por estado brasileiro, ou seja caracterizar geograficamente as entidades atuantes em uma área, neste caso, a localização dos recursos humanos (AGUIAR, 1991).

Considerou-se apenas os brasileiros, dos 2520 pesquisadores da amostra, 138 possuem endereços profissionais estrangeiros e foram desconsiderados no Gráfico 7. O número mantido foi o de 2.382 pesquisadores.

São Paulo possui a maior distribuição com cerca de 32% dos pesquisadores, Minas Gerais com 13%, Rio Grande do Sul com 10%, Rio de Janeiro com 9%, Paraná como 6% e Santa Catarina com 4%. Juntos esses seis estados correspondem a 74% da distribuição total.

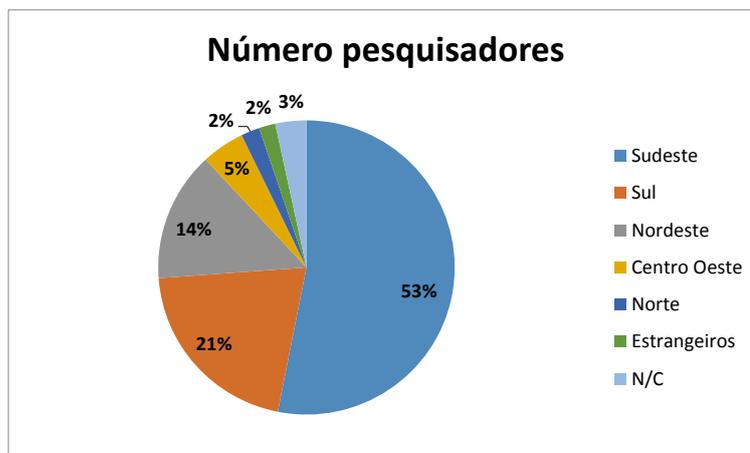
Gráfico 6 – Distribuição dos pesquisadores em nanotecnologia por estados



Fonte: Elaboração própria

Também foi possível consolidar as informações sobre os estados e apresentar um gráfico para as regiões brasileiras. Assim, no Gráfico 7 percebe-se que o Sudeste (com 53%) concentra mais de cinquenta por cento dos pesquisadores em nanotecnologia. A região Sul com 21% de representatividade da distribuição é a segunda. Sul e Sudeste somam juntos 74% dos pesquisadores da área no país. Nordeste com 14% é a terceira maior distribuição. Interessante observar que o número de estrangeiros encontrados é muito parecido com o total de pesquisadores da região norte, confirmando os dados apresentados pelo MCTI (2015).

Gráfico 7 – Distribuição dos pesquisadores em nanotecnologia por regiões



Fonte: Elaboração própria

Na Figura 11 apresenta-se uma visualização da rede de relacionamento entre os termos da expressão de busca encontrados nos títulos dos artigos. Foi construída utilizando o *software Gephi* e um arquivo com a extensão GDF elaborado com base na saída em *Excel* do *Scriptkeyword*. Aplicou-se a distribuição “*Force atlas*” e o cálculo de análise da modularidade e obteve-se a presença de três grupos. Para padronização do método, nesta visualização utilizou-se apenas os termos da expressão de busca, contudo poderiam ter sido utilizadas qualquer lista de palavras, como por exemplo a lista inicial com os 464 termos.

Os termos com maior frequência são representados pelas esferas (nós) maiores, nanopartícula, nanocomposite, catalysts e nanoestruturado são os grandes nós do grupo em azul. Nanotubo, grafeno e nanowire aparecem com os maiores do grupo em verde. As arestas representam a força (grau) da relação entre os termos. Assim, nanocristal e nanocristalino, do grupo em vermelho, apresentam forte ligação com nanopartícula, por exemplo. Nota-se uma pulverização das conexões, pois embora nem todas sejam fortes elas estão presentes em praticamente todos os termos, revelando aspectos da interação entre os atores da rede (RECUERO, 2009).

É possível relacionar, nos grupos encontrados, miniaturas de campos científicos como apontados por Latour e Woolgar (1979) e compreender a dinâmica do relacionamento entre eles.

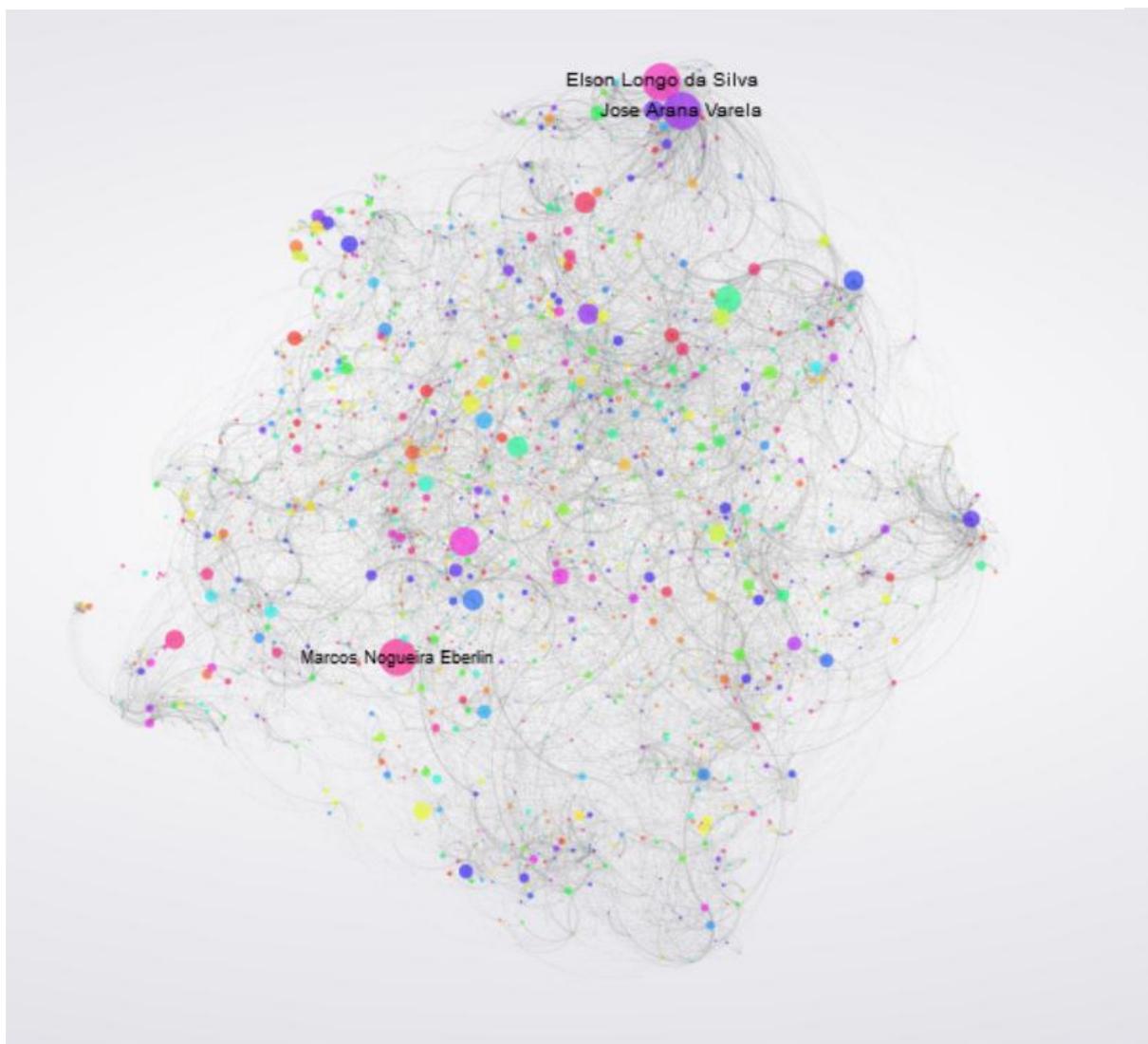


trata de uma rede coesa, cheia de conexões entre as partes envolvidas, justamente por produzirem publicações em conjunto (TARGINO; NEYRA, 2006).

Os conceitos de crédito e credibilidade de Latour e Woolgar (1979) podem ser visualizados na rede apresentada da seguinte forma, os elementos econômicos viabilizam pesquisas melhores e por conseguinte as melhores oportunidades do campo. De acordo com Hochman (1994) os atores mais representativos podem ser os que possuem mais recursos acumulados, são aqueles que convertem dinheiro em equipamentos, prestígios, papers, prêmios, viagens, etc. Elementos que mediante o investimento do capital monetário fornecem o capital social, credibilidade nas relações com os pares e com a sociedade. É importante ressaltar que no esquema de Baumgarten (2004) as avaliações em ciência são baseadas em julgamentos que podem ser por mérito (qualidades intrínsecas) ou por relevância (características externas de seus resultados). Ambos consideram o capital social acumulado por um pesquisador, como em um ciclo, as avaliações tendem a ser realizadas pelos comitês de áreas, constituídos pelos pesquisadores que detêm mais credibilidade e podem direcionar caminhos (FRIGERI & MONTEIRO, 2015).

As conexões neste caso, representam as colaborações em publicações. Os maiores nós representam os atores mais influentes, que podem sê-los por terem maior número de publicações e conexões. As cores representam os agrupamentos que são criados a partir de coeficientes matemáticos do *software* que gerou a rede (Gephi).

Figura 12 – Rede geral em nanotecnologia



Fonte: VLAB4U (2015)

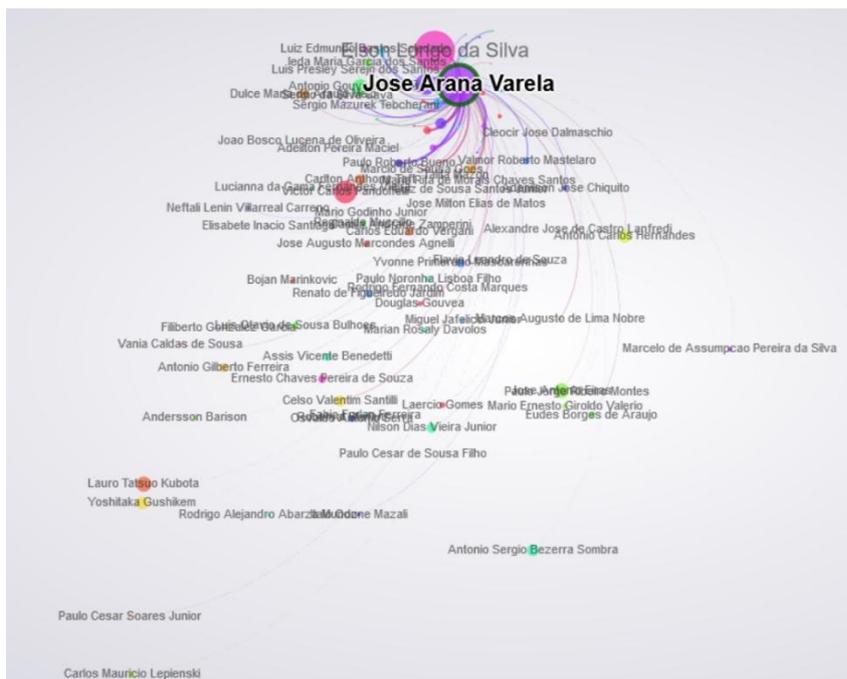
Do mesmo modo que na rede anterior de palavras-chave, as arestas representam o peso da conexão, arestas com espessura maior representam conexões mais frequentes (RECUERO, 2009). As redes são representações gráficas de uma determinada população e não podem ser analisadas com qualidade em representações estáticas e em papel. Por este motivo, as redes devem ser interativas, atualmente a melhor maneira para apresentá-las é com possibilidades de interação em tela<sup>30</sup>.

As Figuras 13 e 14 ilustram a possibilidade de analisar de forma mais particularizada os dois pesquisadores que aparecem na parte superior da rede: Elson Longo da Silva e Jose Arana Varela.

<sup>30</sup> Tais possibilidades podem ser acessadas em: <http://vlab4u.info/Nano-gexf-js/>.

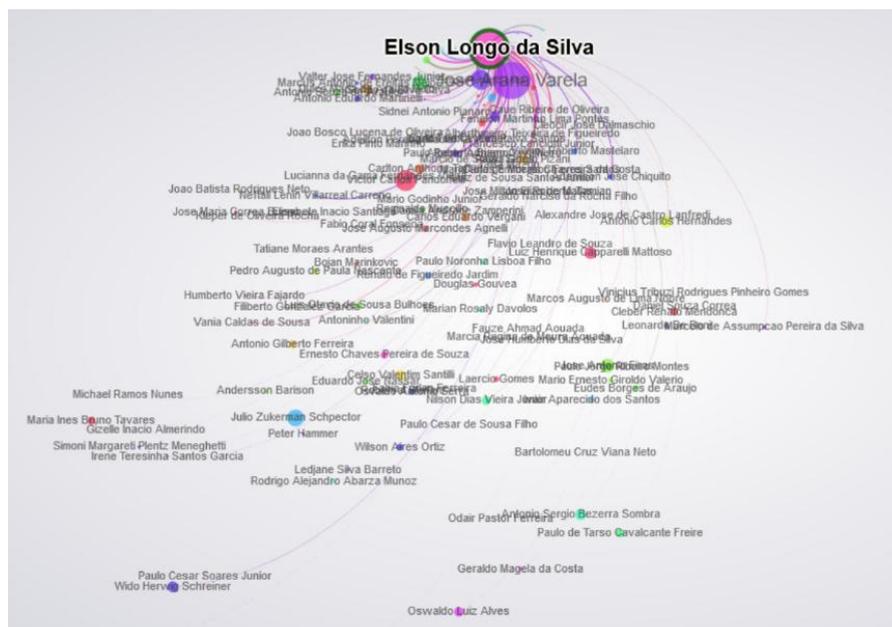
Quando na interface de um computador é possível visualizar apenas um nó (ator da rede) com suas respectivas conexões e deste modo identificar quem são os demais atores que participam dessa relação específica. Para cada nó selecionado é permitido acessar informações detalhadas sobre sua posição na rede, uma parte dessas informações pode ser encontrada na Figura 15.

Figura 13 – Visualização do pesquisador José Arana Varela



Fonte: VLAB4U (2016)

Figura 14 – Visualização do pesquisador Elson Longo da Silva



Fonte: VLAB4U (2016)

Figura 15 – Informações dos nós selecionados

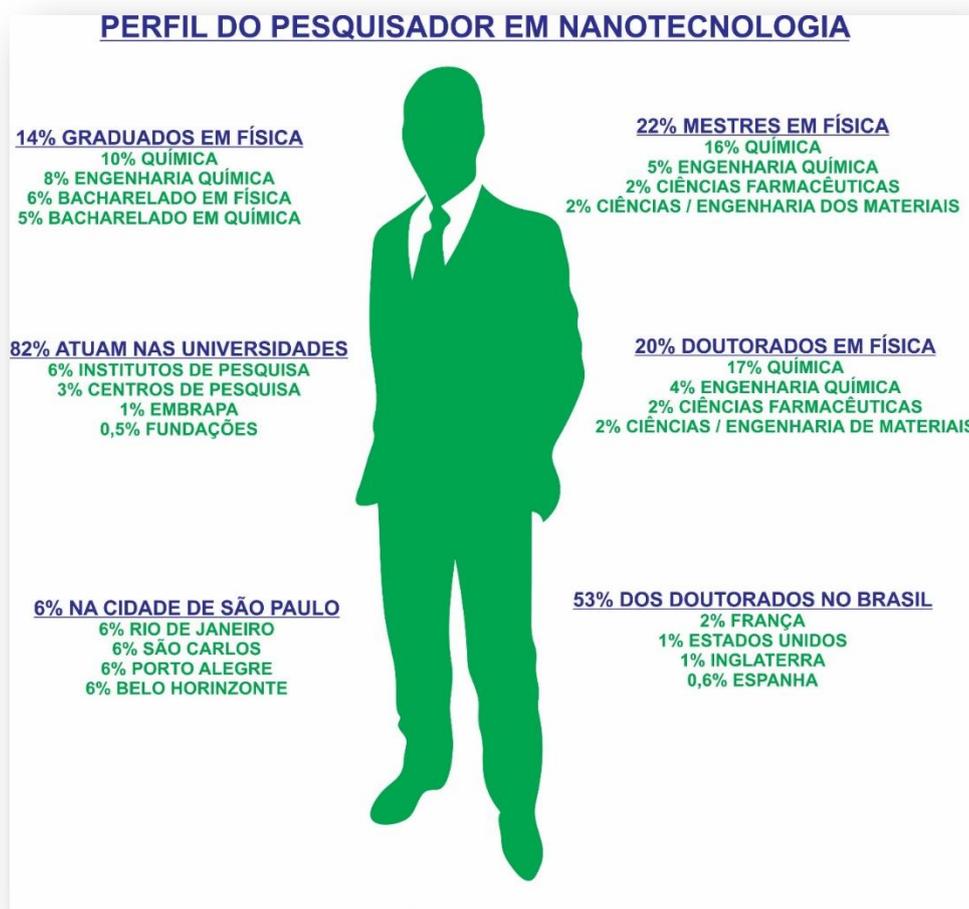


Fonte: VLAB4U (2015)

As informações se referem as características dos nós e sua ligação não direcionada com os outros nós que se relacionam com eles. Uma representação visual no formato de rede pode permitir uma infinidade de análises, neste trabalho o objetivo principal é apresentar mais um indicador da BCL com possibilidades de análises para “n” grupos e áreas.

O infográfico apresentado na Figura 16 é o último indicador da BCL construído para a tese, ele apresenta informações sobre os pesquisadores em nanotecnologia, Aguiar (1991) relaciona este indicador a caracterização da oferta em C&T com as informações sobre áreas de atuação, de formação acadêmica básica, titulação etc. Onde é possível perceber que as áreas de Física e Química têm forte influência em suas formações, como já apresentado no Gráfico 5. Embora estejam espalhados pelas cidades, percebe-se que há maior concentração em cidades que pertence a região sul e sudeste de acordo com os dados apresentados no Gráfico 7.

Figura 16 – Infográfico com o perfil do pesquisador em nanotecnologia



Fonte: Autora com base em VLAB4U (2015)

Outro aspecto interessante é com relação ao local de atuação, cerca de 82% estão atuando ou possuem um vínculo com as universidades, revelando um indicador importante de que a área de nanotecnologia ainda está fortemente ligada a ciência, pois a atividade de pesquisa e desenvolvimento no Brasil é produzida em grande parte nas universidades e com dinheiro público como ressaltado por Costa (2006) ao mencionar as características úteis ao acesso aberto. Em relação ao país onde esses pesquisadores realizaram o doutorado, por erro de preenchimento ou incompatibilidades no arquivo utilizado, não foi possível identificar o país para todos os pesquisadores dessa população, entre os possíveis, tem-se 53% das formações deles realizadas no Brasil.

Embora a nanotecnologia seja uma área com potencial para desenvolver aplicações nas áreas de engenharias e fármacos, por exemplo, as formações específicas nestas áreas ainda não são as mais representativas. Todavia é fato que áreas das ciências básicas, neste caso a química e a física, sempre forneceram contribuições para áreas mais aplicadas.

As métricas apresentadas nos indicadores da BCL dizem respeito a um sistema específico de medição, mediante indicadores de uma ampla gama de dados quantitativos de atividades selecionadas dos currículos Lattes. Úteis, por vezes necessárias, para a determinação das prioridades nacionais e avaliação de retorno (DERRICK; PAVONE, 2013; HAGENDIJK, 2004).

Por fim, com os dados da BCL foi possível apresentar indicadores do comportamento dos termos utilizados na área, no Gráfico 4 e Figura 11, elucidando aspectos temáticos para compreensão da área. Análises das grandes áreas de atuação e formação dos pesquisadores, no Gráfico 5 e Figura 16, fomentando a compreensão dos recursos humanos envolvidos. Distribuição dos pesquisadores por estados da federação e por regiões, no Gráfico 6 e Gráfico 7, localizando no espaço geográfico a real distribuição dos pesquisadores no Brasil. E uma rede de colaboração com foco na produtividade (publicação de artigos), bem como o detalhamento das relações de dois atores de destaque nas Figuras 12, 13 e 14, acarretando possibilidades de compreender a dinâmica dos relacionamentos entre os membros da amostra.

Salienta-se que outros indicadores poderiam ser construídos, já que a BCL é uma fonte de informação valiosa e favorece a recomendação de Hicks (2015) ao ressaltar que as bases de dados internacionais e indexadas não refletem a excelência em pesquisa nas ciências sociais e humanidades, na qual as pesquisas são mais regionais e de centralidade nacional. Muitas outras áreas tem uma dimensão nacional ou regional – por exemplo, estudos epidemiológicos específicos de determinadas regiões. Esta diversidade e relevância social tendem a ser

suprimidas quando o foco está em artigos de interesse dos *gatekeepers* dos periódicos de alto impacto de língua inglesa. Ou seja, métricas baseadas em literatura de alta qualidade em língua não inglesa serviriam para identificar e reconhecer excelência em pesquisa de caráter local ou regional.

O uso da BCL para geração de indicadores corrobora com outra recomendação de Hicks (2015) de que determinadas áreas (notadamente ciências sociais e humanidades) publicam preferencialmente livros ao invés de artigos em periódicos, e outras como ciências da computação tem a maior parte de sua produção científica disseminada em conferências. Esta diversidade de formas de publicação de resultados de pesquisa deve ser levada em conta ao avaliar diferentes áreas. Deste modo, considerar a BCL como fonte de informação permite criar novas e diferentes formas de elaborar indicadores. Infelizmente não há essa iniciativa (BCL) em cada país, o que permitiria comparar os mais diversos indicadores.

## 4.2.2 Indicadores Diretório de Grupos de Pesquisa

Os ScriptGP foram utilizados para compor os indicadores aqui apresentados (evolução anual, distribuição por áreas, por estados e instituições)<sup>31</sup>. A primeira etapa consistiu no processamento de uma lista de termos de busca<sup>32</sup> que identificou 644 Grupos de Pesquisas com pelo menos um dos termos (os mesmos usados na expressão de busca da BCL). Em seguida identificou-se todos os pesquisadores constituintes dos 644 GP. Esse resultado permitiu que a lista de constituintes dos Grupos de Pesquisa fosse comparada com a lista dos 2.520 pesquisadores em nanotecnologia, assim optou-se em manter, para os tratamentos do DGP, apenas os pesquisadores que estavam presentes nas duas listas. Ou seja, aqueles que participam de um grupo de pesquisa ligado a nanotecnologia e tem pelo menos dois artigos publicados (identificados pelo *ScriptKeyword*), resultando em 1.576 pesquisadores.

Estes foram considerados o “core”, ou o núcleo dos pesquisadores da área. A distribuição deles entre as grandes áreas da CAPES é apresentada na Quadro 11.

---

<sup>31</sup> Outras formas de visualização e análises podem ser acessadas em: <http://vlab4u.info/Nano-pivot/>

<sup>32</sup>

busca01,2,nanoparticle nanoparticles nanopartícula nanopartículas nanotechnology nanotechnologies  
 busca02,2,nanotecnologia nanotecnologias nanostructured nanostructureds nanoestruturado  
 nanoestruturados  
 busca03,2,nano nanos nanocomposite nanocomposites nanocompósito nanocompósitos  
 busca04,2,nanotube nanotubes nanotubo nanotubos nanocatalysis catalysts nanocatálise nanocatalisadores  
 busca05,2,nanomaterial nanomaterials nanomaterial nanomateriais nanocrystal nanocrystals nanocristal  
 busca06,2,nanocristais nanocrystalline nanocrystallines nanocristalino nanocristalinos graphene graphenes  
 busca07,2,grafeno grafenos nanowire nanowires nanofio nanofios nanoscience nanosciences nanociência  
 busca08,2,nanociências nanofiber nanofibers nanofibra nanofibras nanocapsule nanocapsules nanocapsula  
 busca09,2,nanocapsulas nanoparticulate nanoparticulates nanoparticulado nanoparticulados nanoemulsion  
 busca10,2,nanoemulsions nanoemulsão nanoemulsões nanosphere nanospheres nanoesfera nanoesferas  
 busca11,2,nanobiotechnology nanobiotechnologies nanobiotecnologia nanobiotecnologias nanopowder  
 busca12,2,nanopowders nanopó nanopós fullerene fullerenes fullereno fullerenos nanoporous nanoporoso  
 busca13,2,nanoporosos dendrimer dendrimers dendrímero dendrímeros nanoindentation nanoindentations  
 busca14,2,nanoindentação nanoindentações nanoribbon nanoribbons nanofita nanofitas  
 busca15,2,nanoclay nanoclays nanoargila nanoargilas nanorod nanorods nanohaste nanohastes nanomedicine  
 busca16,2,nanomedicines nanomedicina nanomedicinas nanocluster nanoclusters nanocacho nanocachos  
 busca17,3,nanoestruturas de carbono  
 busca18,3,poços quânticos  
 busca19,3,quantum dot  
 busca20,3,quantum dots  
 busca21,3,ponto quântico  
 busca22,3,pontos quânticos  
 busca23,3,sol gel

Quadro 11 – Evolução dos grupos de pesquisa dos pesquisadores em nanotecnologia

Ano	Area	Ciencias Agrarias	Ciencias Biologicas	Ciencias da Saude	Ciencias Exatas e da Terra	Ciencias Humanas	Ciencias Sociais Aplicadas	Engenharias	Totais
1963					1				1
1970					2				2
1972					8				8
1973			2		4				6
1974					2				2
1975					4				4
1976					2				2
1977					4				4
1978					3				3
1979					4			4	8
1980					5				5
1981					3			8	11
1982			2		20			1	23
1983					4			8	12
1984					1				1
1985			9		3			8	20
1986	1				1				2
1987					18			5	23
1988				2	13			9	24
1989			3	3	8			3	17
1990	1				21			11	33
1991				4	37			19	60
1992					22			7	29
1993			1		20			11	32
1994		6		4	30			9	49
1995		1	7	2	22			42	74
1996					16			12	28
1997			1	8	40			10	59
1998			1	11	25			19	56
1999		2		1	21	1		1	26
2000		2	3	5	23			14	47
2001			3		24			2	29
2002		5	8	4	38			13	68
2003				3	5			11	19
2004		10		6	46			10	72
2005			1		26			12	39
2006		8	10	15	40			9	82
2007			9	3	39			8	59
2008		3	15	11	42	5		16	89
2009			9	6	46			10	71
2010		5	10	12	46			24	97
2011			1	9	33	1		32	76
2012		1	7	8	24		1	17	58
2013		5	9	7	43			17	81
2014			4	2	42	1		10	59
2015			4	1	1				6
<b>Totais</b>		<b>50</b>	<b>119</b>	<b>127</b>	<b>882</b>	<b>5</b>	<b>1</b>	<b>392</b>	<b>1.576</b>

Fonte: VLAB4U (2015)

Como se pode observar (Quadro 11) são 50 os pesquisadores em Ciências Agrárias, 119 em Ciências Biológicas, 127 em Ciências da Saúde, 882 em Ciências Exatas e da Terra, 5 em Ciências Humanas, 1 em Ciências Sociais Aplicadas e 392 em Engenharias. Em relação a linha do tempo, os anos que mais tiveram adesões, destes pesquisadores nestes GP, foram:

2010 com 97 pesquisadores; 2008 com 89 pesquisadores; 2006 com 82 pesquisadores; e 2013 com 81 pesquisadores.

Dentre as possibilidades de análises dos GP, a distribuição dos pesquisadores pelos estados é uma delas. De acordo com a Quadro 12, os cinco estados com maior representatividade são da região Sudeste e Sul do Brasil. Sendo: 486 pesquisadores em São Paulo, 206 em Minas Gerais, 184 no Rio Grande do Sul, 171 no Rio de Janeiro; e 92 no Paraná.

Quadro 12 – Distribuição do core em nano por estado

<b>SP</b>	<b>486</b>
<b>MG</b>	206
<b>RS</b>	184
<b>RJ</b>	171
<b>PR</b>	92
<b>SC</b>	60
<b>PE</b>	51
<b>DF</b>	46
<b>BA</b>	38
<b>PB</b>	31
<b>RN</b>	26
<b>SE</b>	25
<b>CE</b>	23
<b>GO</b>	22
<b>MS</b>	17
<b>PI</b>	16
<b>PA</b>	13
<b>AL</b>	11
<b>AM</b>	11
<b>ES</b>	11
<b>MA</b>	9
<b>MT</b>	7
<b>AP</b>	5
<b>AC</b>	2
<b>RO</b>	1
<b>ñ identificados</b>	12
<b>Totais</b>	<b>1.576</b>

Fonte: VLAB4U (2015)

Os cinco primeiros estados somam juntos 1.139 pesquisadores, o que representa cerca de 73% do total, confirmando a informação apresentada no Gráfico 6 com os dados da BCL.

A região Norte (Acre, Amapá, Amazonas, Pará, Rondonia, Roraima e Tocantins) tem 38 pesquisadores, cerca de 2,5% do total.

Os dados do DGP revelaram mais uma vez a desigualdade na distribuição dos pesquisadores da área, tal constatação pode se relacionar ao que Bourdieu (1975) tratou como luta política pela dominação científica, manifestada pelas relações de poder e pela distribuição desigual de um tipo específico de capital. Também se pode associar os dois pólos científicos da ciência, o dos dominantes, neste caso representado pela concentração e maior número de pesquisadores na região sul e sudeste, e os dominados nas demais regiões por estarem em menor número e menos concentrados, em regiões geograficamente maiores (BOURDIEU, 1975).

Para os estados de São Paulo e Minas Gerais, por serem os mais representativos em número de pesquisadores, apresenta-se as instituições nas quais estão vinculados, ver Quadro 13 e 14.

Em São Paulo, as instituições que mais congregam pesquisadores são a Universidade de São Paulo (USP) com 98, seguida pela Universidade Estadual Paulista Julio de Mesquita Filho (UNESP) com 85, em terceiro lugar ficou a Universidade Estadual de Campinas (UNICAMP) com 55.

Quadro 13 – Instituições do “core nano” em São Paulo

UF	Instituicao	Totais
	Centro Estadual de Educacao Tecnologica Paula Souza	4
	Centro Nacional de Pesquisa em Energia e Materiais	4
	Centro Universitario Central Paulista	1
	Centro Universitario da FEI	1
	Centro Universitario de Araraquara	2
	Comissao Nacional de Energia Nuclear	36
	Empresa Brasileira de Pesquisa Agropecuaria	43
	Instituto Butantan	2
	Instituto de Aeronautica e Espaco	1
	Instituto de Estudos Avancados	3
	Instituto de Pesquisas Tecnologicas do Estado de Sao Paulo	3
	Instituto Israelita de Ensino e Pesquisa Albert Einstein	9
	Instituto Nacional de Pesquisas Espaciais	18
	Instituto Tecnologico de Aeronautica	8
SP	Pontificia Universidade Catolica de Sao Paulo	1
	Universidade Bandeirante de Sao Paulo	1
	Universidade Camilo Castelo Branco	1
	Universidade de Franca	7
	Universidade de Sao Paulo	98
	Universidade de Sorocaba	2
	Universidade de Taubate	2
	Universidade do Sagrado Coracao	3
	Universidade do Vale do Paraiba	9
	Universidade Estadual de Campinas	55
	Universidade Estadual Paulista Julio de Mesquita Filho	85
	Universidade Federal de Sao Carlos	42
	Universidade Federal de Sao Paulo	15
	Universidade Federal do ABC	28
	Universidade Presbiteriana Mackenzie	2
	<b>Totais</b>	<b>486</b>

Fonte: VLAB4U (2015)

Em Minas Gerais, a Universidade Federal de Minas Gerais (UFMG) é a que reúne o maior número de pesquisadores (89). Seguida pela Universidade Federal de Uberlândia com 24 pesquisadores e a Universidade Federal de Viçosa com 19 pesquisadores. Como apresentado na Figura 16, infográfico como o perfil do pesquisador em nanotecnologia, as universidades são as maiores empregadoras dessa área.

Quadro 14 – Instituições do “core nano” em Minas Gerais

UF	Instituicao	Totais
	Centro Federal de Educacao Tecnologica de Minas Gerais	6
	Comissao Nacional de Energia Nuclear	7
	Empresa Brasileira de Pesquisa Agropecuaria	1
	Fundacao Centro Tecnologico de Minas Gerais	1
	Fundacao Ezequiel Dias	1
	Instituto Federal do Sudeste de Minas Gerais	1
	Universidade Federal de Alfenas	5
	Universidade Federal de Itajuba	10
MG	Universidade Federal de Juiz de Fora	11
	Universidade Federal de Lavras	1
	Universidade Federal de Minas Gerais	89
	Universidade Federal de Ouro Preto	9
	Universidade Federal de Sao Joao Del-Rei	8
	Universidade Federal de Uberlandia	24
	Universidade Federal de Vicosa	19
	Universidade Federal do Triangulo Mineiro	5
	Universidade Federal dos Vales do Jequitinhonha e Mucuri	8
	<b>Totais</b>	<b>206</b>

Fonte: VLAB4U (2015)

Seriam possíveis outras combinações entre os dados dos GP utilizando o *Pivotable*, pois além das possibilidades de compor a visualização (gráficos de barras, gráficos de linhas, tabelas com mapa de calor, gráfico de árvore, dentre outras), tem-se também as possibilidades de análises pelos conjuntos de dados, que são: IdGP, IdLattes, área do conhecimento, estado e instituição; nome do pesquisador, ano, cidade, subárea, sigla da instituição.

O Quadro 15 apresenta todos os conjuntos de dados permitidos, o nível de aprofundamento, uma vez que é possível identificar inclusive o nome e IdLattes do pesquisador. Ressalta-se que essa dimensão de integração da informação não é encontrada em outra fonte de informação. É possível visualizar os pesquisadores que formam o núcleo de uma área juntos, podendo traçar análises e comparações rapidamente, chegando no nível do ID Lattes e nome do pesquisador.

No DGP e na BCL encontra-se informações sobre o Grupo de Pesquisa e sobre o pesquisador, mas não se encontra a possibilidade de integrar em uma visualização as informações de dois grupos, por exemplo. Nos indicadores aqui apresentados, é possível integrar grupo de pesquisa de mais de 1.500 pesquisadores. Este nível de detalhamento e integração dos dados, não é encontrado em nenhuma outra fonte até o presente momento, o que reforça a vantagem e o ineditismo dos tratamentos apresentados neste trabalho.

Quadro 15 – Nível máximo de detalhamento dos grupos de pesquisa em nanotecnologia

SiglaInst.	Idlattes	Idgdp	NomeGrupo	Area	UF	Instituicao	Ano	SubArea	NomePesquisador	Cidade	Totais
	0745659049466597	9850240163646282	NANOENDOAMBIENTAL	Ciencias da Saude	RJ	Fundacao Oswaldo Cruz	2003	Saude Coletiva	William Weissmann	Rio de Janeiro	1
	1092649238010894	8598452129413753	Sistemas avancados de liberacao de farmacos	Ciencias da Saude	RJ	Fundacao Oswaldo Cruz	2013	Farmacia	Erlon Henrique Martins Ferreira	Rio de Janeiro	1
	1424782908764033	9850240163646282	NANOENDOAMBIENTAL	Ciencias da Saude	RJ	Fundacao Oswaldo Cruz	2003	Saude Coletiva	Adriano Premebida	Rio de Janeiro	1
	1504062128168961	8598452129413753	Sistemas avancados de liberacao de farmacos	Ciencias da Saude	RJ	Fundacao Oswaldo Cruz	2013	Farmacia	Katty Gyselle de Holanda e Silva	Rio de Janeiro	1
FIOCRUZ	4161702873261271	9548950310720182	Farmacologia Celular, Inflamacao e Inovacao Terapeutica	Ciencias Biologicas	RJ	Fundacao Oswaldo Cruz	1998	Farmacologia	Andressa Bernardi	Rio de Janeiro	1
	4970404116008239	8598452129413753	Sistemas avancados de liberacao de farmacos	Ciencias da Saude	RJ	Fundacao Oswaldo Cruz	2013	Farmacia	Flávia Almada do Carmo	Rio de Janeiro	1
	5692825410779014	3119148054961392	Biologia da interacao patogeno-hospedeiro (BiPaH)	Ciencias Biologicas	BA	Fundacao Oswaldo Cruz	2005	Imunologia	Camila Indiani de Oliveira	Salvador	1
	7143561813892945	9850240163646282	NANOENDOAMBIENTAL	Ciencias da Saude	RJ	Fundacao Oswaldo Cruz	2003	Saude Coletiva	Wilson Engelmann	Rio de Janeiro	1
	7495357939551039	8598452129413753	Sistemas avancados de liberacao de farmacos	Ciencias da Saude	RJ	Fundacao Oswaldo Cruz	2013	Farmacia	Luiz Cláudio Rodrigues Pereira da Silva	Rio de Janeiro	1
										Totais	9

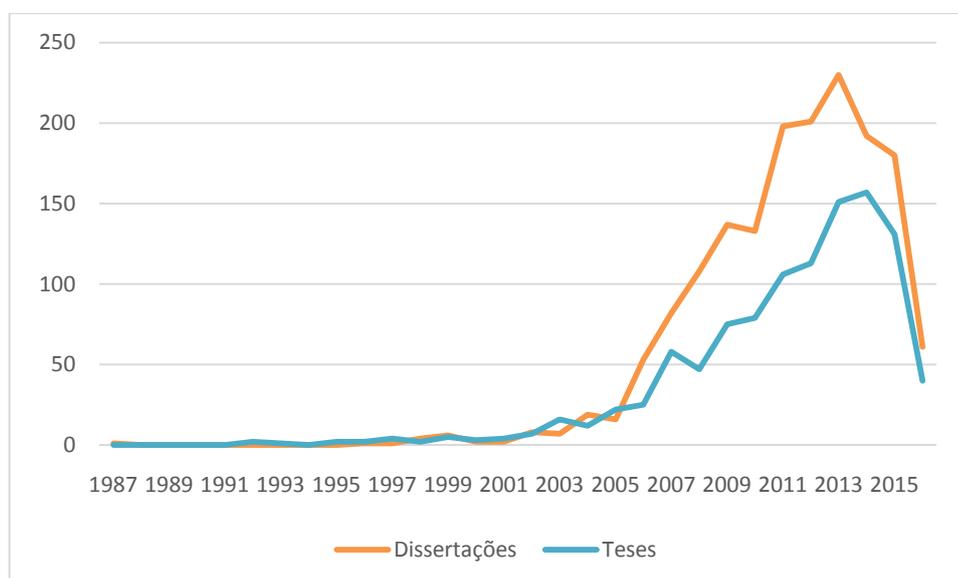
Fonte: Vlab4u (2016)

### 4.2.3 Indicadores BDTD

Com o arquivo recebido pela equipe da BDTD (extensão csv) foi possível elaborar indicadores que retraram o crescimento (teses e dissertações), estatística das palavras-chave, classes dos vocabulários em nanotecnologia, nuvem e rede de palavras.

O primeiro indicador é apresentado no gráfico 8, referente ao número das teses e dissertações ao longo dos anos de 1987 até 09/2016. No total foram coletados os dados de 1.643 dissertações e 1.065 teses com o termo nano\* no título ou nas palavras-chave (campo assunto). Observa-se (Gráfico 8) que só a partir do ano 2003-2004 se iniciou um aumento das pesquisas. Tal crescimento, de acordo com os dados, se manteve até 2014 e começou a diminuir em 2015, como os dados de 2016 ainda não estão completos não se pode afirmar que continuará diminuindo, será necessário acompanhar os próximos depósitos das defesas. Cabe ressaltar que este aumento talvez se deva, em parte, pelo crescimento da base de dados como um todo.

Gráfico 8 – Número das teses e dissertações em nanotecnologia



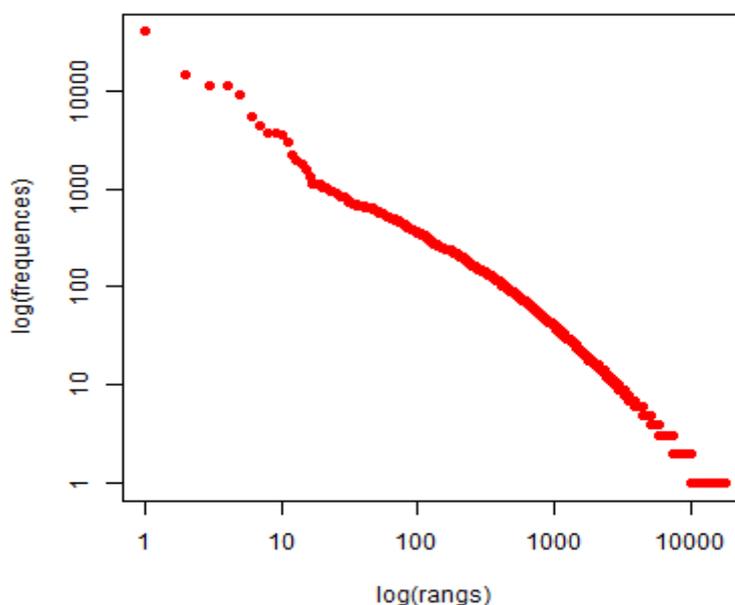
Fonte: Elaboração própria com base na BDTD

Este aumento está relacionado com os investimentos realizados pelo MCTI (2015). De acordo com o Gráfico 1 (MCTI, 2015) os investimentos mais representativos se iniciaram justamente neste mesmo período, ou seja, 2004 com pouco mais de 20 milhões de reais. Os pontos máximos de investimento, 2007 e 2013 do Gráfico 1, são justamente os anos que

antecedem os anos com mais dissertações e teses defendidas, provavelmente um reflexo do investimento de 440,16 milhões de reais realizados entre 2004 e 2014 (por subvenção, INCT's Nano, FNDCT, Fonte 100).

O gráfico 9 apresenta o resultado da análise da estatística textual do corpus utilizado, composto por 922 resumos de teses (algumas teses não possuíam resumos), que utilizou os verbos e substantivos. Nota-se neste gráfico, que quanto maior é a frequência de uma forma/palavra em cada resumo, menor é a sua quantidade no corpus.

Gráfico 9 – Estatística das frequências das palavras dos resumos em nananotecnologia



Fonte: Elaboração própria com base em BDTD (2016)

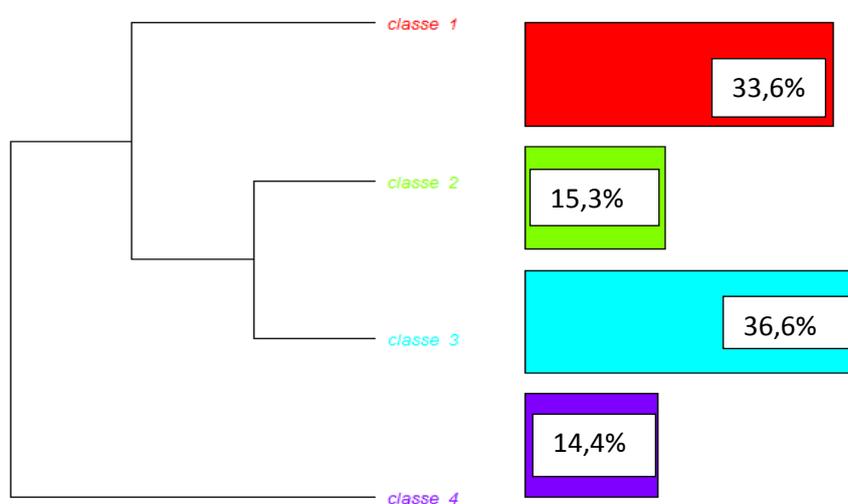
Neste caso, as palavras com frequência 1 existem em grande quantidade no texto, e palavras com frequência 100 aparecem em menor quantidade.

No Iramuteq, é possível realizar também o tratamento com a Classificação Hierárquica Descendente (CHD) permite uma classificação pelo método Reinert. É uma análise que permite o uso da lógica de correlação, onde utiliza as segmentações do corpus textual, juntamente com a lista de formas reduzidas e o dicionário embutido para apresentar um esquema hierárquico de classes. O Iramuteq identifica as classes de vocabulário, possibilitando deduzir quais ideias o corpus textual deseja transmitir. As classes são separadas

por diferentes cores, dessa forma ilustram quantos vocabulários formam o corpus textual, cuja porcentagem indica a abrangência do vocabulário.

Para a BDTD utilizou-se como corpus os resumos de 922 teses. Cabe ressaltar que em algumas teses o resumo não foi cadastrado, estas foram eliminadas do arquivo que compôs o corpus textual. Na classificação (CHD) realizada pelo Iramuteq, identificou-se quatro classes de vocabulários para o corpus com os resumos das teses em nanotecnologia, ver Figura 17.

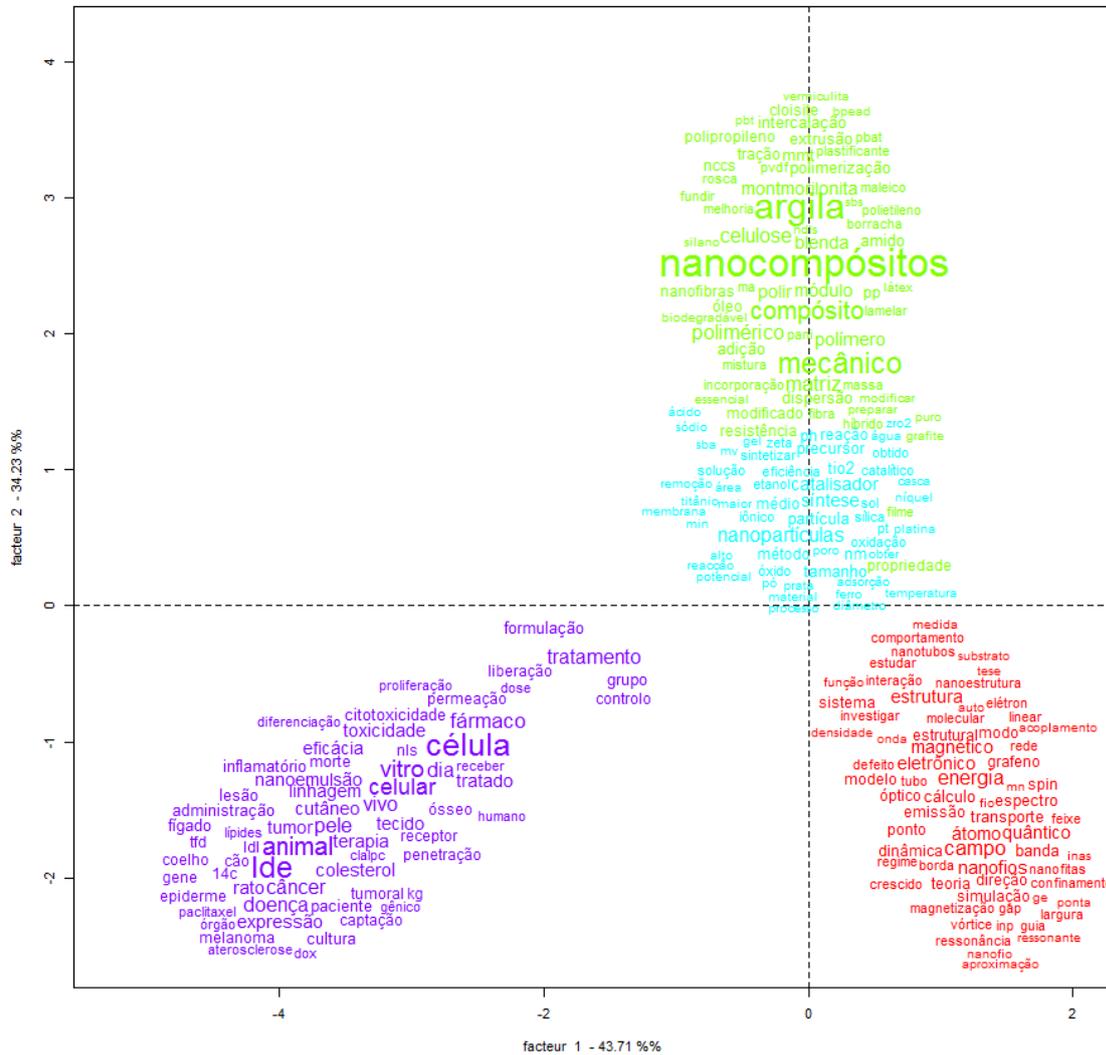
Figura 17 – Classes dos vocabulários em nanotecnologia



Fonte: Elaboração própria com base em BDTD

O Iramuteq permite gerar formas de visualização dessas quatro classes identificadas. A Figura 18 ilustra, com as palavras que compõem os vocabulários, o posicionamento dessas classes. Elas representam o posicionamento das classes de vocabulários no corpus textual, pode-se ver quais classes se complementam e concentram o corpus, e quais se distanciam do centro e mostram certa especificidade.

Figura 18 – Visualização dos vocabulários em classes



Fonte: Elaboração própria com base em DBTD (2016)

Na Figura 18, é possível afirmar que a classe em roxo que apresenta as palavras: animal, célula, vitro, cutâneo, inflamatório, pele, tumor, doença, dentre outras, está ligada com as pesquisas em nanotecnologia que tratam de aspectos relacionados a saúde, como em fármacos, por exemplo, com 14,4% de representatividade. As palavras da classe em vermelho, foram: magnético, eletrônico, grafeno, nanotubos, confinamentos, campo, dentre outras, revelam uma classe composta pelas pesquisas em nanotecnologia que abordam a área da elétrica e eletrônica com 33,6%.

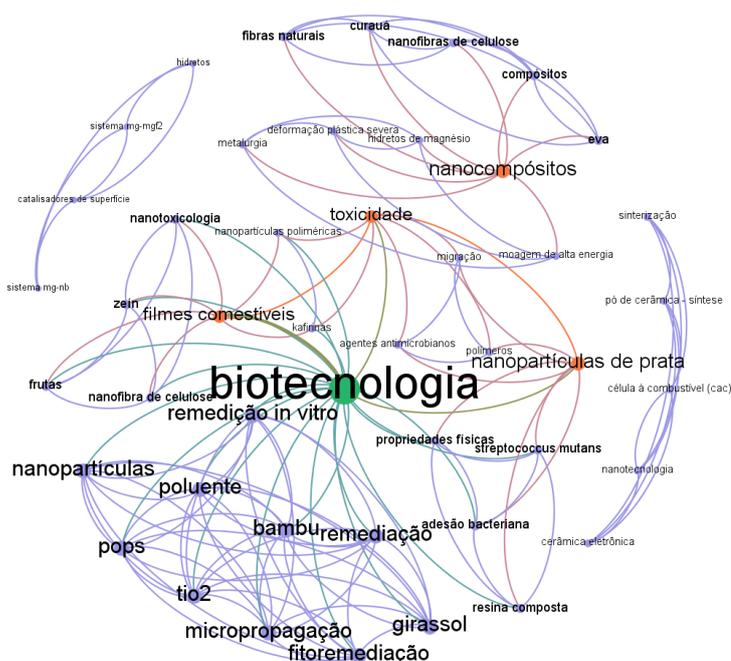
As palavras da classe em azul elucidam aspectos da área de nanotecnologia que estuda aspectos dos materiais naturais, no qual encontra-se as palavras: síntese, casca, membrana, etanol, água, dentre outras, com 36,6 de representatividade. A última classe, em verde,



A Figura 19 apresenta palavras bastante amplas como: material, propriedade, sistema, estudo, efeito, resultado, amostra, método. Ou seja, palavras que revelam características de procedimentos ou algo relacionado a eles. A palavra nanopartículas é a palavra específica maior e mais central, confirmando o que foi identificado com os dados da Base de Currículos Lattes (Figura 11). Outras palavras que aparecem na nuvem e que também estiveram presentes na expressão de busca da BCL, foram: nanotubos, nanocompósitos, nanoestrutura, nanoestruturados, nanofibras, carbono, nanocristais, dentre outras. Há também a presença de palavras relacionadas às técnicas realizadas nesses estudos, como: espectroscopia e microscopia.

Com o propósito de apresentar que é possível criar uma rede, bem como demonstrar o potencial de análises com as palavras-chave, foi construída uma rede de palavras-chave demonstrativa com apenas 10 teses da UFSCar. O número reduzido da amostra se justifica devido a necessidade de correção do campo palavras-chave no arquivo de origem (enviado pela equipe BDTD), uma vez que ainda não há padronização nos marcadores de separação entre elas. A Figura 20 apresenta essa rede demonstrativa com as palavras-chave de apenas dez teses (escolhidas aleatoriamente) da UFSCar.

Figura 20 – Demonstrativo de uma rede de palavras-chave de teses da UFSCar



Fonte: Autora com base em BDTD (2016)

Foram identificados três grupos (verde, laranja e roxo) onde a palavra “biotecnologia” obteve maior destaque. Algumas palavras-chave não estão conectadas com as mais representativas e se apresentam separadas do grupo principal (a maioria do grupo roxo), mas conectadas entre elas, sendo: sistema Mg-Nb, catalisadores de superfície, hidretos, sistema Mg Mg<sub>2</sub>, etc. O grupo laranja trata das teses ligadas a nanopartículas de prata, nanocompósitos, toxicidade, fibras naturais, nanofibras de celulose, etc. O grupo verde apresenta o relacionamento entre as palavras: biotecnologia, bambu, remediação, girassol, nanopartículas, resina composta, filmes comestíveis, dentre outras.

Ao interpretar as ligações entre as palavras-chave, nota-se a intersecção entre subáreas da nanotecnologia, seria possível levantar essas ligações não só entre as palavras, mas também entre diferentes instituições de pesquisa, o que poderia revelar as áreas mais proeminentes de um determinado conjunto de programas de pós-graduação, evidenciando possíveis parcerias e/ou outros interesses em comum. Outra análise interessante seria a dos membros que formam a banca de defesa, proporcionaria verificar as colaborações entre instituições.

Por fim, é fato que a BDTD é uma excelente fonte de informação em acesso aberto e que representa um modelo de iniciativas deste contexto, mas devido as dificuldades para atual extração dos metadados, não foi possível demonstrar mais possibilidades de elaboração de indicadores.

#### **4.2.2 Indicadores Espacenet**

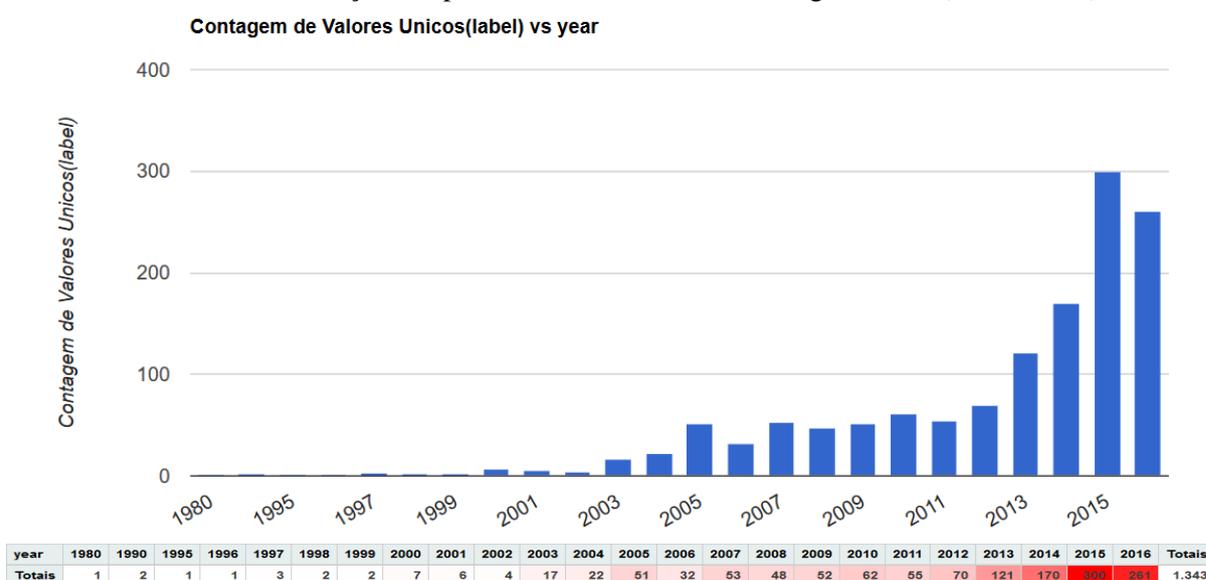
Para coletar dados da Espacenet, utilizou-se o código “H01G11/36” da Classificação Internacional de Patentes. Este código é da seção H que trata sobre Eletricidade: H01G 11/00 capacitores híbridos, isto é, capacitores tendo eletrodos positivo e negativo diferentes; capacitores de dupla camada elétrica (supercapacitores); Processos para fabricação destes ou de suas partes. A classificação “H01G11/36” se refere a aplicação de tecnologia ligada a: nanoestruturas, por exemplo: nanofibras, nanotubos ou fulerenos (WIPO, 2016). Deste modo, tratou-se de uma parte das tecnologias que envolvem nanotubos ligados a Eletricidade.

A busca por essa tecnologia foi realizada em 14 de outubro de 2016, foi utilizada dessa forma por que a busca com palavras-chave muitas vezes pode produzir resultados imprecisos e incompletos, devido ao idioma no qual os documentos de patentes são escritos e os termos

utilizados. Utilizando a base Espacenet, o P2N extraiu dados de 1.344 patentes com 2616 famílias (patentes relacionadas).

No Gráfico 10 é possível acompanhar o aumento das patentes relacionadas a esta tecnologia. Há patentes relacionadas desde 1980 até 2016, só no ano de 2015 foram identificadas 300 patentes. É uma tecnologia em expansão e crescimento, de 2011 até o momento da coleta, havia 977 patentes, cerca de 73% do total de patentes da análise. Tal fato também se deve, em parte, pelo crescimento da base de dados como um todo.

Gráfico 10 – Evolução das patentes relacionadas a tecnologia tratada (H01G11/36)



Fonte: VLAB4U (2016)

No Quadro 16 identifica-se em quais países as patentes estão protegidas. A primeira patente é de 1980 e pertence ao Japão que é o segundo país com mais patentes protegidas em seu território.

A China lidera o *ranking* desta tecnologia com 421 patentes, só em 2016 já foram 144. Com exceção do Japão, que sempre teve um número razoavelmente constante de patentes nesta tecnologia, os demais países começaram a patentear-la com mais ênfase a partir de 2012. Pode-se dizer que os líderes desta tecnologia de acordo com a Espacenet são: China (421); Japão (291); e Estados Unidos (260).

Quadro 16 – Países onde as patentes estão protegidas

country	year	1980	1990	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Totais
AT																		1	1	1						3
AU						1			1			1	1	1								1			1	7
BR								1																	1	2
CA						1						1		1			1		1	1		1	1	2	1	11
CN								1		2			1	5		5	6	7	6	7	8	42	53	134	144	421
DE															1	1				1	1				3	8
EP									1	1		1	2	2	1	2	4	5	3	4	6	6	14	18	6	76
FR														1		1	1				1		1	1		6
GB																			1		1	2	1			5
HK																								1		1
IL										1																1
JP		1	2	1	1		2		2	1		6	12	25	21	27	19	24	16	14	17	14	18	32	36	291
KR									1	1	2	2	4	6	2	7	3	4	4	8	5	4	11	30	17	111
RO																										1
RU									1												3			1	2	8
SG																						1	1			2
TW												1		1	1				2		1	1	1	5	1	14
US									1		2	2	1	5	1	2	9	6	18	8	23	38	50	55	39	260
WO						1						3	1	4	5	8	5	5	9	9	8	9	18	20	10	115
<b>Totais</b>		<b>1</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>3</b>	<b>2</b>	<b>2</b>	<b>7</b>	<b>6</b>	<b>4</b>	<b>17</b>	<b>22</b>	<b>51</b>	<b>32</b>	<b>53</b>	<b>48</b>	<b>52</b>	<b>62</b>	<b>55</b>	<b>70</b>	<b>121</b>	<b>170</b>	<b>300</b>	<b>261</b>	<b>1.343</b>

Fonte: VLAB4U (2016)

Com o objetivo de observar a representatividade do Brasil nesta tecnologia, filtraram-se os resultados da coluna “country” por "BR", utilizando agora o *DataTable* e encontrou-se duas patentes ligadas a esta tecnologia (H01G11/36), apresentada no Quadro 17.

Quadro 17 – Patentes sobre H01G11/36 protegidas no Brasil

Country	Title	Applicant	CPC	Pub year
br	Search	Search	Search	Search
BR	Materiais nanoporosos, fabricao de materiais nanoporosos e aplicaes de materiais nanoporosos	Cambridge Entrpr Ltd	B01D67/003, B01D69/02, B01D71/28, B01D71/80, C08J5/2231, C08J9/26, C08J9/40, H01G11/24, H01G11/26, H01G11/36, H01G11/46, H01G11/48, H01G11/50, H01G11/86, Y02E60/13, B01D2323/30, B01D2323/345, B01D2325/02, C08J2201/026, C08J2201/038, C08J2201/046, C08J2205/042, C08J2353/00, Y10T428/24975, Y10T428/249981, B32B7/02, B29C67/202, B32B3/26, H01G9/0032	2016
BR	Manofibras graf ticas em capacitores eletroqu micos	Hyperion Catalysis Int	B82Y30/00, H01G9/155, H01G9/22, H01G11/02, H01G11/24, H01G11/36, H01G11/48, H01G4/005, Y02T10/7022, Y02E60/13, H01M4/583, Y10S977/788, Y10S977/734	1999

Fonte: VLAB4U (2016)

Os depositantes foram Cambridge Enterprise (empresa inglesa) e Hyperion Catalysis (empresa norte americana<sup>33</sup>). A patente norte americana já está com o prazo de licença terminando (1999 até 2019). Desta maneira, considerando que as patentes WO<sup>34</sup> sejam 115 e que as depositadas e protegidas especificamente no Brasil sejam duas, há oportunidade para o Brasil explorar 1.226 patentes relacionadas a tecnologia “H01G11/36”.

Neste sentido, tal quantitativo e a própria distribuição revelam um descompasso do desenvolvimento da nanotecnologia em relação a algumas localidades (PASTRANA; ÁVILA; MORENO, 2012). Ou seja, devido às pesquisas em nanotecnologia, bem como os processos de produção, além de muito complexos, requererem altos investimentos financeiros. Essas condições inviabilizam a condução dos mencionados processos por pequenos grupos que não detém investimentos financeiros suficientes. Como consequência, a nanotecnologia torna-se susceptível de ser monopolizada por restritas corporações, as quais dispõem de financiamento para o empreendimento (MARTINS, 2006). Por esta razão talvez haja concentração das patentes em alguns países como apresentado na Figura 21, abaixo.

Sendo assim, arrisca-se dizer que todo o investimento em nanotecnologia já realizado no Brasil não gera bem-estar social proporcional ao investimento público (LENS, 2014; MARTINS 2006). Neste caso é possível observar que avaliação de retorno mediante a transferência das tecnologias é uma métrica possível para tal verificação, corroborando com a afirmação de Derrick & Pavone (2013) sobre a dificuldade em medir diretamente as contribuições da ciência à atividade econômica.

Outra informação interessante e facilmente visualizada com os tratamentos do P2N, são os mapas, gerados com o propósito de visualizar a atratividade das patentes. Neste indicador de geolocalização apresentado na Figura 21, é possível visualizar, de forma simples e rápida, três distribuições diferentes: a primeira apresenta a distribuição dos países depositantes; a segunda os países que detém a proteção; e a terceira os países dos inventores.

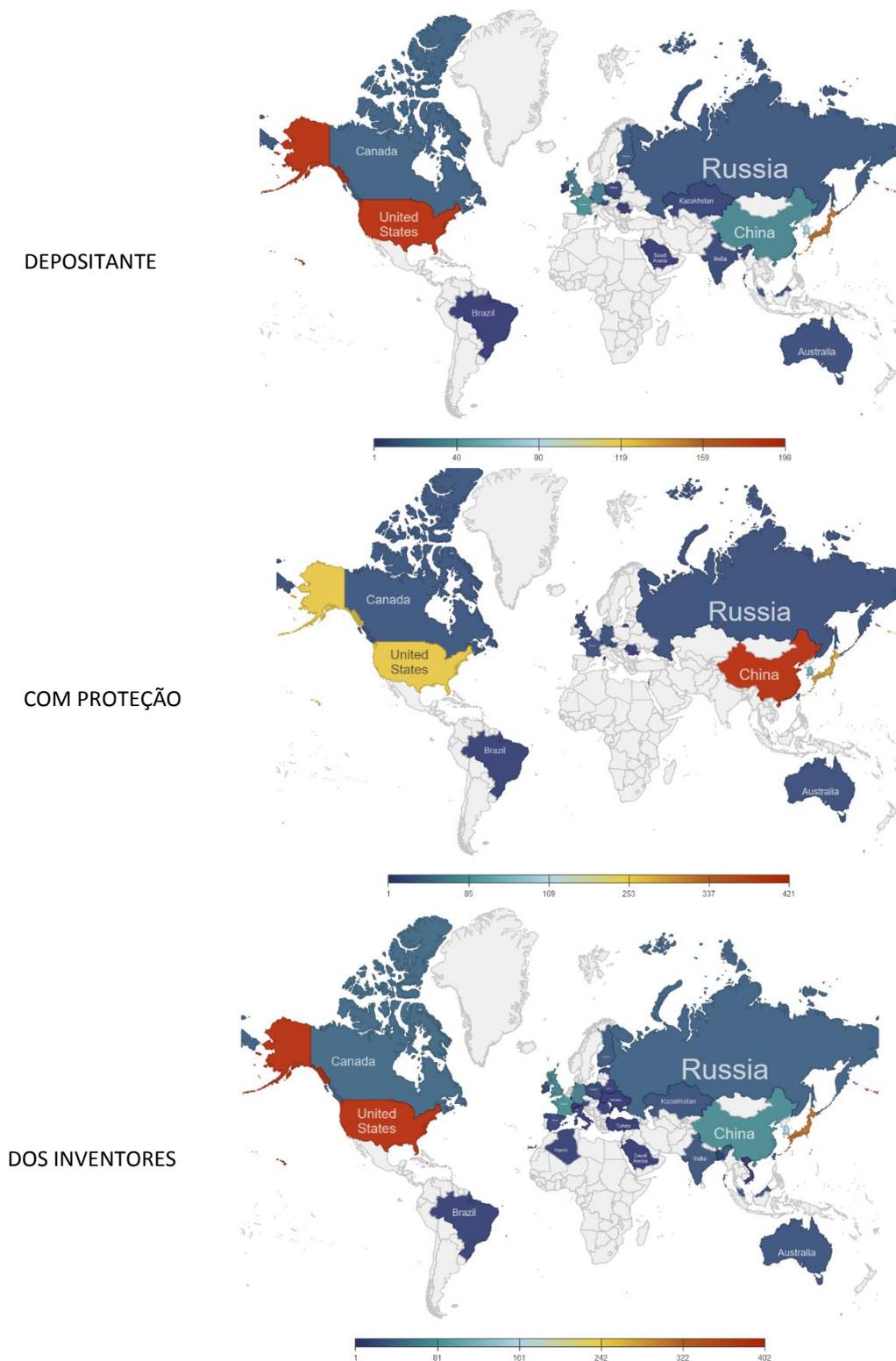
Essa visualização é importante para verificar presença não simultânea nestas atuações, pois como observa-se o país onde há proteção nem sempre é o país do inventor (muitos pesquisadores atuam em empresas estrangeiras) e o país depositante nem sempre protege a patente em seu próprio território (interesses particulares nos mercados potenciais).

---

<sup>33</sup> Quando a busca é feita considerando as famílias das patentes, encontrou-se mais uma patente norte americana estendida no Brasil, com o título “Dispositivos elétricos contendo fibras infundidas com nanotubos de carbono e métodos para sua produção, nº US2011304964.

<sup>34</sup> São patentes que depositadas no *Patent Cooperation Treaty* (PCT), um tratado internacional que oficializa um sistema para o depósito de pedidos de patente e que permite que se obtenham patentes em diversos países a partir de um único pedido.

Figura 21 – Mapa dos países depositantes, com proteção e do inventor

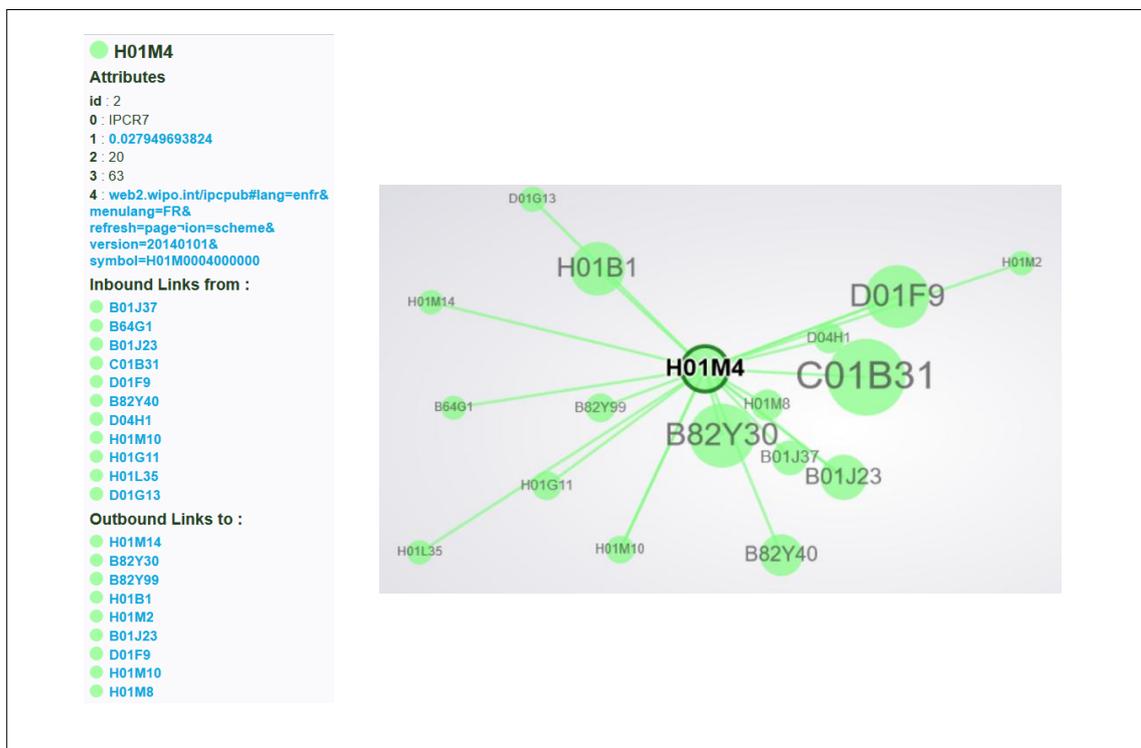


Fonte: VLAB4U (2016)

Nas imagens, uma escala representada por cores e valores indicou a posição de cada país, onde o azul mais escuro estaria próximo de 1 e o vermelho mais escuro próximo do valor máximo do conjunto de dados. Por exemplo, com relação aos países com patentes protegidas, o Brasil como já visto, tem apenas 2 patentes protegidas, os Estados Unidos têm 260, a China 421, o Japão 292 e a Coreia do Sul 111. Sobre os países depositantes, o Brasil tem apenas uma patente, os Estados Unidos 198, a China 26, o Japão 155, e a Coreia do Sul 75. É importante ressaltar que para gerar os mapas o P2N não considerou as patentes com abrangência mundial (WO) e européia (EP).

O P2N também gera tratamento das patentes mediante redes de colaboração utilizando o Gephi, de modo automático. Há a geração de 7 diferentes redes inter-relacionando depositante, inventor, tecnologias, dentre outros. Na interface da tela é possível separar um único nó e visualizar suas conexões. A Figura 22 apresenta um exemplo, análise da tecnologia “H01M4” e suas conexões.

Figura 22 – Conexões da tecnologia H01M4

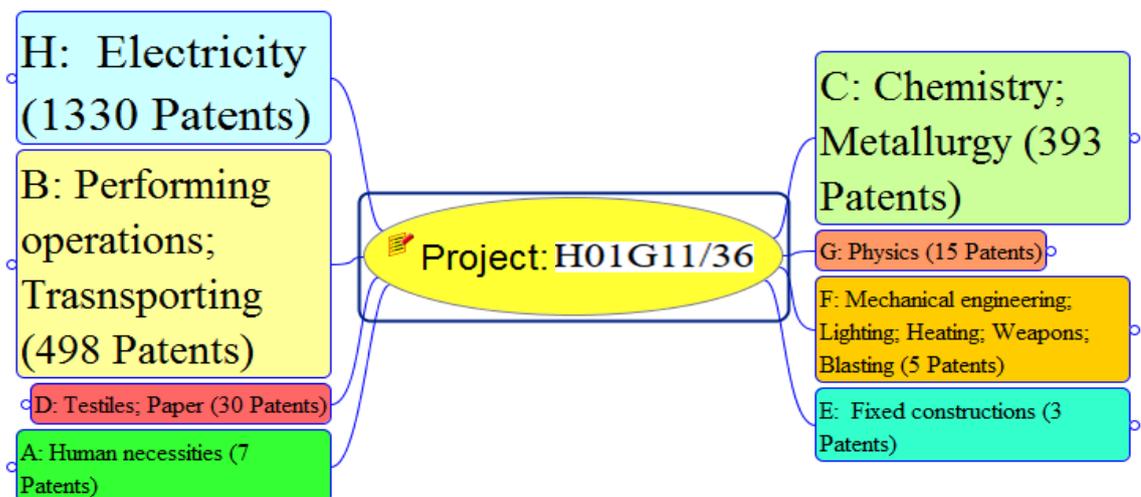


Fonte: VLAB4U (2016)

Este nó “H01M4” trata sobre eletrodos, a conexão se dá com tecnologia relacionadas com carbono e seus compostos (C01B31), filamentos manufaturados ou similares de outras substâncias e aparelhos especialmente adaptados para manufatura de filamentos de carbono

(D01F9), nanotecnologia para ciência de materiais ou de superfícies – por exemplo, nanocompósitos (B82Y30), dentre outras. Ao lado da rede do nó selecionado é possível identificar os principais atributos do nó (H01M4). Outro resultado disponível de forma automática é o do *Mind Map*, que cria um mapa conceitual com as patentes de acordo com suas classificações. Na Figura 23 é possível visualizar a opção mais simplificada do tratamento “H01G11/36”.

Figura 23 – Mapa conceitual por classificação das patentes analisadas

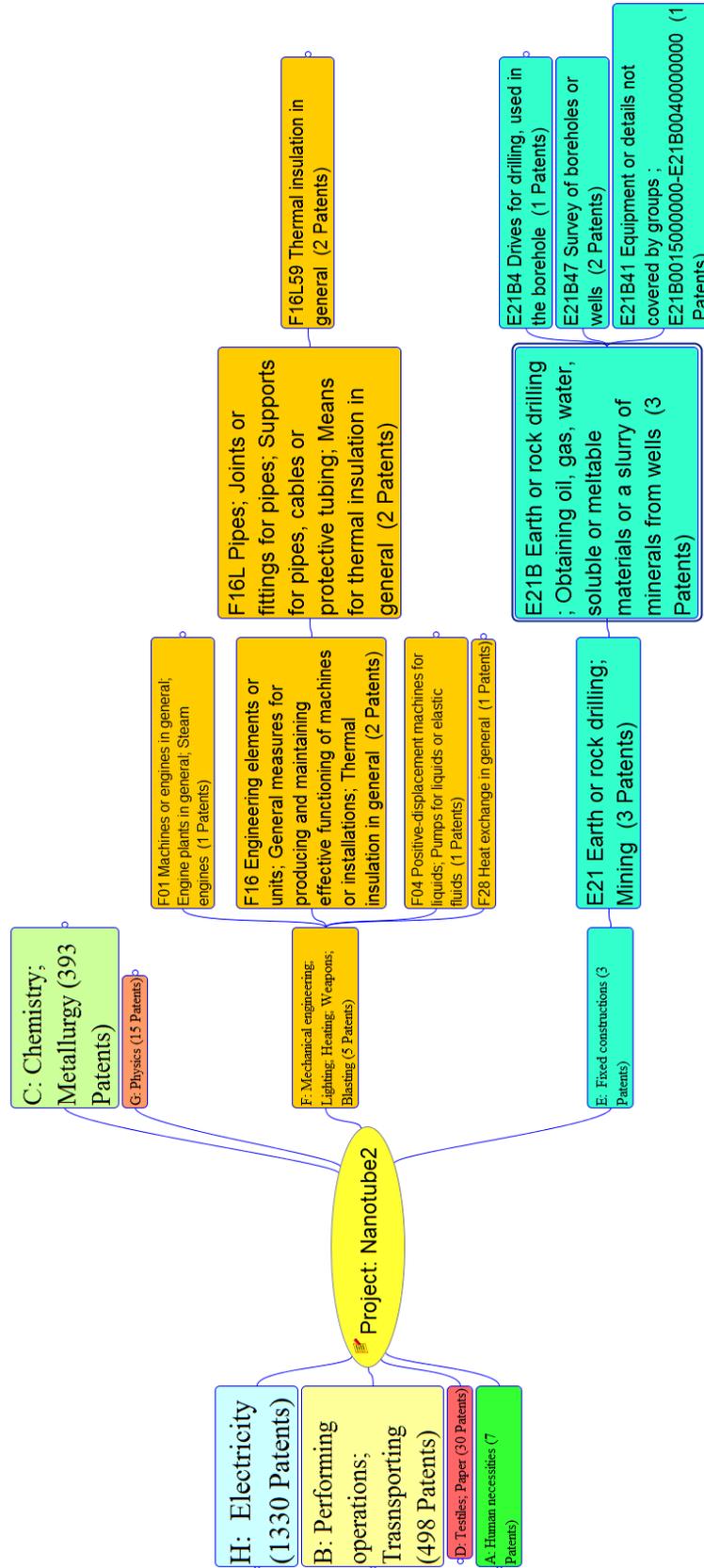


Fonte: VLAB4U (2016)

Não é possível apresentar no formato análogo (este texto) todos os níveis de detalhamento do *MindMap* criado com o P2N por estar bem completo e extenso, deste modo, selecionou-se a representação das áreas E e F, por serem menores, para demonstrar o nível do detalhamento, ver Figura 24.

Ao clicar no botão a direita da caixa E e F, abrem-se novas caixas com os detalhes baseados nos níveis de classificação das patentes, desta maneira, ao passo que se torne interessante ou necessário se aprofundar em uma classe, por exemplo a E que trata de perfuração em solo, rapidamente se visualiza o número de patentes relacionadas e suas classificações na Classificação Internacional de Patentes.

Figura 24 – Detalhamento das áreas E e F



Fonte: VLAB4U (2016)



### *Pré conclusão da presente seção*

Em síntese considera-se que os indicadores da Base de Currículos Lattes integraram dados e informações que antes se encontravam isolados. Diversas palavras presentes nos títulos dos artigos analisados no Gráfico 4, também estão presentes na nuvem de palavras dos resumos das teses (Figura 23), confirmando de certo modo, a similitude dos dados de duas diferentes fontes, demonstrando também que os currículos Lattes alimentam o DGP.

O Gráfico 5 que analisou as áreas de atuação utilizando dados da BCL também corrobora com as classes de palavras identificadas pelo Iramuteq na Figura 18, com os dados da BDTD e com as áreas de formação apresentadas no infográfico da Figura 16. Da mesma maneira, os estados com maior incidência de pesquisadores em nanotecnologia, são os mesmos apresentados no DGP e na BCL, justamente pelo fato de que uma base (BCL) alimenta as informações dos indivíduos da outra (DGP). Notou-se também um declínio no crescimento de teses que estudam a nanotecnologia nos últimos dois anos (desde 2014), entretanto ao se analisar as patentes (H01G11/36), na Espacenet verificou-se um aumento do patenteamento. Os indicadores apresentados representam apenas uma amostra das possibilidades de análises da área aqui abordada. O procedimento de criação de corpus e sua visualização pode ser realizado em outras análises, como por exemplo, análises referentes a prospecção de uma área.

Cabe ressaltar que tanto as fontes de informação utilizadas como os indicadores apresentados são canais de comunicação da ciência e devem representar a democratização em C&T, seja para usos particulares ou coletivos. De acordo com Hicks (2015) o próximo passo da análise é submeter os indicadores à avaliação por pares e facilitar o julgamento. Uma vez que os indicadores não devem substituir a avaliação pelos pares e todos os envolvidos são responsáveis por suas avaliações, ressaltando que um único modelo de avaliação não se aplica a todos os contextos.

Com vistas aos consideráveis avanços que o acesso aberto aos dados de pesquisa pode proporcionar e de acordo com o "Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã" lançado este ano pelo IBICT, os dados coletados e utilizados para a elaboração dos resultados apresentados podem ser consultados na *web*. Desta maneira, a reunião do conjunto de procedimentos elaborados e descritos, oportunizou a formalização de procedimentos para coleta, tratamento e análise de áreas temáticas da ciência e da tecnologia – favorecendo por exemplo, a compreensão de como e onde uma área se desenvolve no país,

subsidiando a criação de indicadores sobre o desenvolvimento nacional, além de possibilitar a identificação de pesquisadores atuantes, dentre outras possibilidades.



## 5 CONCLUSÃO

A sistemática empregada no presente trabalho serviu de instrumento de integração e aquisição de informações a nível micro, médio e macro dos dados científicos nacionais. Além de poder ser replicada para o mapeamento de qualquer outra área estudada no Brasil. Neste contexto, há movimentos que impulsionam o avanço do conhecimento a partir de demandas sociais, mas também há pesquisas que preveem as demandas. Como tais: Conhecer o que se produz? Como produz? Quando produz? Com quem produz? Onde produz? Quanto custa para produzir? – dentre outras, são perguntas que podem ser respondidas pelo conjunto de informações disponíveis nas bases utilizadas. Entretanto, da forma como estão disponibilizadas hoje não proporcionam plenos ganhos de desenvolvimento das pesquisas em C&T.

As principais dificuldades encontradas para o desenvolvimento desta pesquisa estiveram relacionadas aos equipamentos, pois algumas etapas exigiram grandes capacidades de processamentos computacionais. Esta dificuldade foi contornada, utilizando-se um servidor do orientador da pesquisa, localizado na França e também um servidor do NIT-Materiais e toda a estrutura de rede de internet desses dois ambientes. Outra dificuldade que merece ser destacada é o ineditismo operacional de grande parte dos procedimentos, o que exigiu, além da fase de desenvolvimento, muitos testes e ajustes junto aos desenvolvedores. É importante ressaltar que todos os softwares e ferramentas utilizadas são de utilização e código fonte livres.

O principal resultado referente a BCL e os seus indicadores foi apresentar uma possibilidade de identificar quem são os 2.500 pesquisadores apontados pelo MCTI (2015), além de compreender suas formações de base e na pós graduação, bem como a distribuição geográfica de cada um deles (não apenas a somatória) e o relacionamento das palavras-chave utilizadas atualmente na área. Um resultado factível e bastante lastimável foi a implementação do captcha como barreira ao uso do ScriptLattes, representando um retrocesso ao acesso aberto a informação C&T de uma importante iniciativa brasileira (Base de Currículos Lattes).

O principal resultado referente ao DGP foi a apresentação do núcleo (core competence) da área, cada um dos 1.576 pesquisadores, que participam de Grupos de Pesquisa em nanotecnologia e que também publicaram dois artigos com as palavras-chave levantadas. Além da integração, dos dados desses pesquisadores, a um nível de detalhamento ainda não possibilitado de outra forma, ver <http://vlab4u.info/Nano-pivot/>.

O principal resultado da BDTD foi verificar a relação direta entre os investimentos e o aumento das pesquisas na área. Constatavam-se mecanismos de visualização dos relacionamentos das subáreas estudadas em nanotecnologia, além de integrar indicadores que facilitam a compreensão da dimensão ensino da pesquisa em nanotecnologia no Brasil. Entretanto é necessário ressaltar a urgente necessidade de uma alternativa à extração efetiva e oficial dos metadados da BDTD.

O principal resultado da Espacenet foi a constatação da facilidade de gerar diversos indicadores de informação tecnológica em uma área do conhecimento, constando também a falta de representatividade do Brasil no cenário mundial de uma determinada tecnologia. O que trouxe à tona a questão da dificuldade brasileira em transferir o conhecimento gerado na academia para o mercado. Afirma-se que uso do P2N permite a democratização do acesso, tratamento e uso das informações em patentes da base Espacenet.

Com relação aos objetivos específicos, estes foram alcançados de modo satisfatório. Foi possível levantar as bases de dados de acesso aberto (BCL, DGP, BDTD e Espacenet) e as ferramentas computacionais de uso livre (ScriptExtract1, ScriptExtract2; ScriptKeyword; Scriptlattes – Conjunto de ScriptGP – CSV para BDTD – P2N – Gephi e Iramuteq), que se mostraram adequadas para a coleta de dados e informações das bases de dados utilizadas.

Foram elaboradas formas válidas de recuperação da informação para coletar dados e informações dos Currículos Lattes (apresentada na subseção 3.2), Diretório de Grupos de Pesquisa (apresentada na subseção 3.3.1.2), Biblioteca Digital Brasileira de Teses e Dissertações (apresentada na subseção 3.3.1.3) e Espacenet (apresentada na subseção 3.3.1.4), todas com o foco no objeto de estudo, a nanotecnologia.

A sistemática (apresentada na Figura 10), ilustrou os procedimentos sugeridos para o processo de coleta, tratamento e análise de dados e informações, da Base de Currículos Lattes, Diretório de Grupos de Pesquisa, Biblioteca Digital Brasileira de Teses e Dissertações e Espacenet relacionadas a Nanotecnologia. Também revelou um caminho interessante para integrar e gerar informações úteis, mas até o momento não poderia ser replicada em outros países, pois a Base de Currículo Lattes e o Diretório de Grupos de Pesquisa são iniciativas brasileiras bem particulares. Neste sentido, ainda não há possibilidade de comparar todos os indicadores gerados neste trabalho, se houvesse, poderia por exemplo, comparar os investimentos em C&T e o seu retorno. Sendo assim, a aplicação da sistemática mediante a construção de indicadores sobre a nanotecnologia foi realizada de forma satisfatória e revelou indicadores sólidos, pertinentes e integrados sobre a área objeto de estudo deste trabalho.

Atualmente não seria possível acessar os indicadores apresentados de outra maneira, pois o grande desafio que ainda se apresenta, é o da organização da informação segundo as necessidades práticas dos atores sociais que intervêm no processo de desenvolvimento social, neste caso das instituições em C&T. Sendo assim, uma alternativa emergente é a publicação da sistemática em um periódico de alcance internacional, como o *Scientometrics*, para que se possa divulgar as iniciativas brasileiras em gestão da informação em C&T, bem como a própria sistemática, efetivando uma divulgação de alcance globalizado da iniciativa.

É importante ressaltar que quando o contexto envolve as pesquisas em C&T financiadas com recursos públicos o acesso aberto é altamente recomendado e necessário. Entretanto, há necessidade de compreensão das exceções, como por exemplo, quando estiverem envolvidas informações sigilosas como aquelas passíveis de patenteamento.

As análises dos dados revelaram que a PL (com a BCL e DGP), BDTD e Espacenet são ricas e importante fonte de informação para compreensão de cenários científicos e tecnológicos do Brasil. Nesta perspectiva, o presente trabalho proporciona uma nova alternativa para utilização da PL ao fornecer uma sistemática composta por passo a passo e um caso ilustrativo de como realizar os procedimentos, bem como baixar e ter acesso as ferramentas livres.

Dentre as sugestões para possíveis desdobramentos e trabalhos futuros, tem-se os seguintes pontos: prospectar os assuntos futuros sobre a nanotecnologia, levantando os assuntos que possivelmente serão tratados em médio prazo, utilizando os dados dos projetos e orientações em andamento cadastrados nos Currículos Lattes; mensurar a Responsabilidade Social da Pesquisa em nanotecnologia, mediante os dados dos “Produtos Tecnológicos” cadastrados nos Currículos Lattes; mapear os pesquisadores, os grupos de pesquisa e os INCT sobre nanotecnologia, representando em grafos de rede e também em um mapa do Brasil onde estão os pesquisadores, os grupos de pesquisa e os INCT que pesquisam sobre a nanotecnologia; analisar os assuntos utilizados nas teses sobre nanotecnologia, construindo um mapa temático para análise e representação dos temas das publicações sobre nanotecnologia nos últimos anos; analisar outras aplicações de tecnologias utilizando a CIP e a Espacenet, etc.

As contribuições mais significativas desta pesquisa estão relacionadas com o levantamento das informações específicas da área de nanotecnologia, bem como com uma sistemática para coleta, tratamento e análise de dados e informações em bases de dados de acesso aberto, utilizando ferramentas livres que podem ser replicadas para tratamento de

qualquer outra área prioritária ou não do Brasil. Para ampliar as análises referentes a nanotecnologia será interessante colher as interpretações de especialistas, o que será realizado em uma próxima publicação em periódico avaliado nesta área de concentração, como um dos desdobramentos da tese.

É correto afirmar que o objetivo principal da pesquisa de desenvolver e aplicar uma sistemática para coleta, tratamento e análise de informações sobre a nanotecnologia no Brasil, utilizando essencialmente dados de acesso aberto e ferramentas livres, foi realizado satisfatoriamente e revelou uma significativa e importante sistemática para mapeamento de qualquer outra área prioritária do Brasil. Espera-se que o uso das ferramentas e da sistemática desenvolvida possibilite maior quantidade e qualidade de informações para apoiar a pesquisa e desenvolvimento de ações públicas para compreensão do comportamento da área de nanotecnologia.

Foi constatado a PL (com a BCL e DGP) e a BDTD contribuem para o entendimento da dinâmica do desenvolvimento científico e tecnológico brasileiro como fonte de informação, apesar das dificuldades de extração, tratamento, análise e representação visual da informação. Com relação ao conhecimento adquirido nas bases de patentes, ele se faz a cada dia mais essencial quando o assunto aborda C&T. Dentre suas diversas contribuições uma situação real é que as pessoas podem inadvertidamente infringir patentes que não conhecem ou evitar áreas de inovação em que têm direito de atuar com criatividade, ou até mesmo realizar investimentos desastrosos com base em informações incompletas sobre as patentes. Assim, pessoas físicas que queiram obter mais informações sobre determinada tecnologia ou processo tecnológico, são usuários potenciais da sistemática aqui desenvolvida.

É importante ressaltar que dias antes da defesa desta tese o CNPq realizou modificações na interface de busca da Base de Currículos Lattes, que vão em deslencontro do que foi aqui discutido e apresentado, pois na nova apresentação não é mais possível realizar buscas com os operadores booleanos. Entretanto, tal atitude reforça todas as críticas aqui apresentadas, de que a Base de Currículos Lattes estava e continua sendo subutilizada como fonte de informação da ciência brasileira.

Conclui-se que as Informações em C&T, com acesso aberto, permitem uma democrática geração de conhecimentos para este contexto. As fontes de informação e o conjunto de ferramentas levantadas possibilitaram a elaboração de indicadores capazes de elucidar múltiplas dimensões da produção científica e tecnológica – de pesquisadores e grupos de pesquisa – além do ensino e pesquisa na pós graduação.

Também foi possível concluir que o conjunto de resultados aqui obtidos formam uma amostra de como órgãos nacionais de fomento à pesquisa e de formação de recursos humanos de alto nível, como a Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), podem direcionar seus investimentos em áreas do conhecimento consideradas relevantes, por intermédio do conhecimento descoberto em bases de dados. Os indicadores obtidos e sua aplicação reforçam o objetivo do processo de data mining, no sentido de transformar dados em informação a ser empregada no processo decisório das organizações relacionadas à preservação e a inovação do conhecimento.

## Referências

- ABDI. **Panorama de nanotecnologia**. Disponível em: <<http://www.abdi.com.br/Paginas/estudo.aspx?f=Nanotecnologia>>. Acesso em: 30 set. 2013.
- AGARWAL, B. B.; DHALL, S. Web mining: information and pattern discovery on the world wide web. **IJSTM**, 2010.
- AGRAWALA, S. BROAD, K.; GUSTON, D. H. Integrating climate forecasts and societal decision making: Challenges to an emergent boundary organization. **Science Technology and Human Values**, v. 26, n. 4, p. 454-477, 2001.
- AGUIAR, A. C. Informação e atividades de desenvolvimento científico, tecnológico e industrial: tipologia proposta com base em análise funcional. **Ci. Inf.**, Brasília, 20(1):7-15, jan./jun. 1991.
- ALBORNOZ, M. (Eds.). **Indicadores de ciencia y tecnología: estado del arte y perspectivas**. Caracas: Nueva Sociedad, 1994. p. 307-348.
- ALBUQUERQUE, E. M. Patentes segundo a abordagem neo-schumpeteriana: uma discussão introdutória. **Revista de Economia Política**. São Paulo, v. 18, n. 4, p. 65-83, out./dez., 1998.
- ALISSON, E. **E-Science revoluciona a forma como se faz ciência**. 2013. Disponível em: <http://exame.abril.com.br/ciencia/escience-revoluciona-a-forma-como-se-faz-ciencia/>. Acesso em: 20 abr 2016.
- ALVARES, L. Ações do Ibiict em Direção ao Acesso Aberto à Informação Científica. Brasília. 16 nov. 2015. 25 slides. Apresentação em Power-point. Disponível em: <http://lillian.alvarestech.com/Apresentacoes/AcessoAbertoIBICT1617112015.pdf>. Acesso em: 20 set. 2016.
- ANDERSON, C. **The end of theory: the data deluge makes the scientific method obsolete**. Wired, 2008. Disponível em: [http://www.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://www.wired.com/science/discoveries/magazine/16-07/pb_theory). Acesso em: 12 ago 2016.
- ANDRADE, M. M. **Como preparar trabalhos para cursos de pós-graduação**. 5. ed. São Paulo: Atlas, 2002.
- ARAÚJO, C. A. Bibliometria: evolução histórica e questões atuais. **Em Questão**, v. 12, n. 1, 2007.
- ARORA, S. K. ET AL. Capturing new developments in an emerging technology: An updated search strategy for identifying nanotechnology research outputs. **Scientometrics**, v.95, n.1, p.351-370, 2013.
- BAIBICH, M. **Marco regulatório para nanotecnologia**. 2010. Disponível em: <https://jornalggm.com.br/blog/luisnassif/marco-regulatorio-para-a-nanotecnologia#content>. Acesso em: 10 mar. 2016.

BASSOLI, M. **Avaliação do Currículo Lattes como fonte de informação para construção de indicadores de produção científica: o caso da UFSCar.** 2016. 78 p. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) – Universidade Federal de São Carlos, São Carlos, 2016.

BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi, an open source graph visualization and manipulation software. 2016. [s.l: s.n.].

BAUMGARTEN, M. Avaliação e gestão de ciência e tecnologia: Estado e coletividade científica. **Revista Crítica de Ciências Sociais**, n. 70, p. 33–56, 2004.

BAUMGARTEN, M.; SANTOS DE LIMA, L. Divulgação e comunicação em C&T: mediações para a apropriação social do conhecimento. In: XV CONGRESSO BRASILEIRO DE SOCIOLOGIA. Salvador: 10 set. 2013

BERNARDES, M. B.; SANTOS, P. M.; ROVER, A. J. **Teoria e prática de governo aberto: lei de acesso à informação nos executivos municipais da região sul.** Florianópolis: Funjab, 2012.

BEUREN, I. M. **Metodologia da pesquisa aplicável às ciências sociais.** São Paulo: Atlas, 2008.

BOCCATO, V. R. C.; FUJITA, M. S. L. Estudos de avaliação quantitativa e qualitativa de linguagens documentárias: uma síntese bibliográfica. **Perspect. ciênc. inf**, v. 11, n. 2, p. 267–281, ago. 2006.

BOLLIER, D. **The promise and peril of big-data.** Washington: The Aspen Institute, 2010.

BONACCORSI, A. Search regimes and the industrial dynamics of science. **Minerva: a review of science, learning and policy.** The Netherlands, [s.d.]. Forthcoming. Disponível em: <<http://www.springer.com/education/journal/11024>>. Acesso em: 05 jun. 2014

BOURDIEU. O campo científico. In: ORTIZ, R. (Ed.). **Pierre Bourdieu: sociologia.** São Paulo: Ática, 1983.

BOURDIEU, P. The Specificity of the Scientific Field and the Social Conditions of the Progress of Reason. **Social Science Information**, v. 14, n. 6, p. 304–317, 1975.

BOUTET, C. V.; QUONIAM, L. Towards active SEO (search engine optimization) 2.0. **Journal of Information System and Technology Management**, v. 9, n. 3, p. 443–458, 2012.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações. DF, v. 134, n. 248, 18 nov. 2011. Seção 1, p. 27834-27841.

BRESSAN, R. T. Dilemas da rede: web 2.0, conceitos, tecnologias e modificações. **Anagrama**, v. 1, n. 2, p. 1-13, mar. 2009. Disponível em: <<http://www.revistas.usp.br/anagrama/article/view/35306/38026>>. Acesso em: 05 jan. 2016.

BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. *Internacional Word Wide Web Conference. Anais...* In: WWW 1998. Brisbane, Australia: 14 abr. 1998. Acesso em: 27 set. 2013

BRITO, A. G. C.; AMARAL, R. M.; QUONIAM, L. A plataforma lattes como fonte de informação. In: III Encontro Internacional – Dados, Tecnologia e Informação (III DTI), 2016. Marília:-SP. **Anais...** Marília: UNESP, 2016.

BRITO, A. G. C.; QUONIAM, L. Análise das publicações sobre a plataforma. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO: além das nuvens, expandindo as fronteiras da Ciência da Informação, 15, 2014, Belo Horizonte. **Anais...** Belo Horizonte: ECI/UFMG, 2014.

BRITO CRUZ, C. H.; CHAIMOCH, H. Brazil. In: **The Current Status of Science around the World**. Unesco Science Report 2010. Paris, Unesco, 2010.

BUDAPEST OPEN ACCESS INITIATIVE. READ THE ORIGINAL (BOAI) declaration. Budapest, 2002. Disponível em: <<http://www.budapestopenaccessinitiative.org/read>>. Acesso em: 18 set 2016.

BUSH, V. Science the endless frontier: a report to the president by Vannevar Bush, director of the office of scientific research and development. Washington: United States Government Printing Office, 1945.

BUSSAB, W. O.; ALBERTO, M. P. **Estatística básica**. São Paulo: Saraiva, 2013.

CAMARGO JR, K. R. A indústria de publicação contra o acesso aberto . **Revista de Saúde Pública**, São Paulo, v. 46, n. 6, 2012, p. 1090-1094. Disponível em: <<http://www.revistas.usp.br/rsp/article/view/53464>>. Acesso em: 05 nov 2016.

CASTELLS, M. **A sociedade em rede**. São Paulo: Paz e Terra, 2010. v. v. 1

CATARINO, M. E.; BAPTISTA, A. A. Folksonomia: um novo conceito para a organização dos recursos digitais na Web. **Data Grama Zero**. v. 8, n. 3, jun. 2007.

CENTRO DE DOCUMENTAÇÃO E INFORMAÇÃO (CID). **Glossário Geral de Ciência da Informação**. Brasília: CID/UNB, 2004.

CGEE. **Avaliação de políticas de ciência, tecnologia e inovação**: diálogo entre experiências internacionais e brasileiras. Brasília: Centro de Gestão e Estudos Estratégicos, 2008.

CHOO, C. W. The Art of Scanning the Environment. **Bulletin of the American Society for Information Science and Technology**, v. 25, n. 3, p. 21–24, 2005.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTIFICO E TECNOLÓGICO (CNPQ). Lattes. 2016. Disponível em: < <http://lattes.cnpq.br/>>. Acesso em: 1 out. 2016.

\_\_\_\_\_. Sobre a plataforma Lattes, 2016a. Disponível em: <http://lattes.cnpq.br/>. Acesso em: 10 out. 2016.

\_\_\_\_\_. Estatísticas da base de currículos da Plataforma Lattes, 2016b. Disponível em: <<http://estatico.cnpq.br/painelLattes/>>. Acesso em: 1 out. 2016.

\_\_\_\_\_. Diretorio de grupos de pesquisa, 2016c. Disponível em: <http://lattes.cnpq.br/web/dgp>. Acesso em: 07 ago. 2016.

\_\_\_\_\_. Aquarius. Plataforma Aquarius, 2014a. Disponível em: <<http://aquarius.mcti.gov.br/app/artigo/>>. Acesso em: 17 set. 2014.

\_\_\_\_\_. Booleana. Wiki CNPq ajuda. 2014b. Disponível em: <<http://ajuda.cnpq.br/index.php/Booleana>>. Acesso em: 21 set. 2014.

CONTE, R.; GIARDINI, F. Towards Computational and behavioral social science. **European Psychologist**. Hogrefe Publishing, 21, 2016. 131-140 p. DOI: 10.1027 / 1016-9040 / a000257.

COSTA, S. Filosofia aberta, modelos de negócios e agências de fomento: elementos essenciais a uma discussão sobre o Acesso Livre à informação científica. **Ciência da Informação**, v. 35, n.2, 2006, p. 39-50.

COSTA, A. R. F.; SOUSA, C. M.; MAZOCCO, F. J. Modelo de comunicação pública da ciência: agenda para um debate teórico-prático. **Conexão – Comunicação e Cultura**, UCS, Caxias do Sul, v. 9, n. 18, jul./dez. 2010.

CRUZ, C. F., et al. Transparência da gestão pública municipal: um estudo a partir dos portais eletrônicos dos maiores municípios brasileiros. **Revista de Administração Pública**, v. 46, n. 1, 2012, pp. 153-176.

DAVYT, A.; VELHO, L. A. A avaliação da ciência e a revisão por pares: passado e presente. Como será o futuro? **História, Ciências, Saúde - Manguinhos**, v.7, n. 1, p. 93- 116, 2000.

DERRICK G. E.; PAVONE, V. Democratising research evaluation: Achieving greater public engagement with bibliometrics-informed peer review. **Science and Public Policy**, v. 40, n. 5, p. 563-575, 2013.

DIAS, R. B. O que é a política científica e tecnológica? **Sociologias**, v. 13, n. 28, p. 316- 344, dez. 2011.

DIGIAMPIETRI, L. A. et. al. Minerando e caracterizando dados dos currículos lattes. Brazilian Workshop on Social Network Analysis and Mining. **Anais...** In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING. Curitiba, PR, Brasil: 2012

DOWBOR, L. **Informação para a Cidadania e o Desenvolvimento Sustentável**. 2003.

EPSTEIN, I. Thomas S. Kuhn: a cientificidade entendida com vigência de um paradigma. In: In: Oliva, Alberto (org.). **Epistemologia: a cientificidade em questão**. Campinas: Papirus, 1990. p. 103-129.

ETZKOWITZ, H.; LEYDESDORFF, L. The dynamics of innovation: from national systems and “mode 2” to a triple helix of university – industry – government relations. **Research Policy**, n. 29, p. 109-123, 2000.

EVENSEN, D. T.; CLARKE, C. E. Efficacy Information in Media Coverage of Infectious Disease Risks. **Science Communication** . v. 34, n. 3, 2011. p. 392 – 418.

FALADORI, G. et al. Características distintivas del desarrollo de las nanotecnologías em América Latina. **Sociologias**, v. 14, n. 30, mai-ago , 2012. pp 330-363.

FARIA, L. I. L. et al. **Análise da produção científica a partir de publicações em periódicos especializados**: indicadores de ciência, tecnologia e inovação em São Paulo. São Paulo: FAPESP, 2011.

FERRARA, E. et. al. Web data extraction, applications and techniques: a survey. **arXiv**, 2013.

FEYERABEND, P. **Contra o método**. 3ªed. Rio de Janeiro: Francisco Alves, 1989.

\_\_\_\_\_. **Contra o método**. Rio de Janeiro: Livraria Francisco Alves Editora, 1977.

FLORIDI, L. Big data and their epistemological challenge. **Philosophy and Technology**, v. 25, n. 4: 2012. pp.435–437.

FONTANELLA, B. J. B.; MAGDALENO JUNIOR, R. Saturação teórica em pesquisas qualitativas: contribuições psicanalíticas. **Psicologia em Estudo**, v. 17, n. 1, p. 63–71, mar. 2012.

FRIGERI, M.; MONTEIRO, M. O Qualis e a rotina editorial dos periódicos científicos. **Revista eletrônica de jornalismo científico**, Com Ciência SBPC/ Labjor, 15 mar. 2015.

GALILEU. Eu não sou um robô. *Galileu*, mar/2016 p. 8-9.

GARFIELD, E. In tribute to Derek John de Solla Price: a citation analysis of Little Science, Big Science. **Scientometrics**, v. 7, n. 3, p. 487–503, 1985.

GIBBONS, M. et al. **The new production of knowledge**: dynamics of science and research in contemporary societies. London: Sage Publications, 1996.

GIBBS, W. W. Lost science in the Third World. **Science**, v. 2, n. 273, p. 76–83, 1995.

GIL, A. C. **Métodos e técnicas de pesquisa social**. São Paulo: Atlas, 1999.

GILAD, B. **Early warning**: using competitive intelligence to anticipate market shifts, control risk, and create powerful strategies. New York: AMACOM, 2003.

GLÄNZEL, W. **Bibliometrics as a research field**: a course on theory and application of bibliometric indicators. Course handouts: 2003. 115 p. Disponível em: <[http://nsdl.niscair.res.in/bitstream/123456789/968/1/Bib\\_Module\\_KUL.pdf](http://nsdl.niscair.res.in/bitstream/123456789/968/1/Bib_Module_KUL.pdf)>. Acesso em: 01 jan. 2014.

GLASER, B.; STRAUSS, A. L. **The discovery of grounded theory**: strategies for qualitative research. Chicago: Aldine, 1967.

GOGOTSI, Y. (Ed.). **Nanomaterials handbook**. New York: Taylor & Francis, 2006.

GREGOLIN, J. A. R. ET AL. **Análise da produção científica a partir de indicadores bibliométricos**: indicadores de ciência, tecnologia e inovação em São Paulo 2004. São Paulo: FAPESP, 2005.

GUIMARÃES, R.; SOUZA, L. E. P. F.; SANTOS, L. M. P. Ciência, tecnologia e pesquisa em saúde. In: **Políticas e sistemas de saúde no Brasil**. Giovanella L. et AL (orgs.). 2 ed., Rio de Janeiro: Editora Fiocruz, 2012. p. 239-58.

GUIMARÃES, R. **Avaliação e fomento de C&T no Brasil**: propostas para os anos 90. Brasília: MCTI/CNPq, 1994.

GUIVANT, J. S.; MACNAGHTEN, P. O mito do consenso: uma perspectiva comparativa sobre governança tecnológica. **Ambientes & Sociedade**. v. XIV, n. 2, p. 89-104, jul./dez. 2011.

HAGENDIJK, R. P. The Public Understanding of Science and Public Participation in Regulated Worlds. **Minerva**, v. 42, n. 1, p. 41–59, 1 mar. 2004.

HAYASHI, M. C. P. I. Afinidades eletivas entre a cientometria e os estudos sociais da ciência Maria Cristina Piumbato Innocentini Hayashi. **Filosofia e Educação (Online)**, v. 5, n. 2, p. 57–88, out. 2013.

HAYASHI, M. C. P. I.; et. al. Indicadores de CT&I no pólo tecnológico de São Carlos: primeiras aproximações. **Revista Digital de Biblioteconomia e Ciência da Informação**, v. 3, n. 2, p. 17–30, 2006.

HEY, T, ET. AL. On eScience: a transformed scientific method. In: Hey T, Tansley S, Tolle K (eds). **The fourth paradigm**: data-intensive scientific discovery. Redmond: Microsoft Research, pp. xvii–xxxii. 2009.

HICKS, D. Bibliometrics: The Leiden Manifesto for research metrics. **Nature**, v. 520, n. 7548, p. 429–431, 22 abr. 2015.

HOCHMAN, G. A. HOCHMAN, G. A ciência entre a comunidade e o mercado: leituras de Kuhn, Bourdieu, Latour e Knorr-Cetina. In: PORTOCARRERO, V. (Org.). In: **Filosofia, história e sociologia das ciências**: abordagens contemporâneas. Rio de Janeiro: Fiocruz, 1994.

IGARASHI, Y. et al. Three levels of data-driven science. **International Meeting on High-Dimensional Data-Driven Science**. Journal of Physics: Conference Series 699. 2016.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). Histórico. 2016a. Disponível em: <http://www.ibict.br/sobre-o-ibict/historico-1>. Acesso em: 20 ago. 2016.

\_\_\_\_\_. O que é?. 2016b. Disponível em:  
<http://bdtd.ibict.br/vufind/Contents/Home?section=what>. Acesso em: 20 ago. 2016.

\_\_\_\_\_. Manifesto. 2016c. Disponível em: [http://www.ibict.br/Sala-de-  
 Imprensa/noticias/2016/ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-  
 brasileira-para-ciencia-cidada](http://www.ibict.br/Sala-de-Imprensa/noticias/2016/ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-brasileira-para-ciencia-cidada). Acesso em: 02 out. 2016.

INVERNIZZI, N.; KÖRBES, C.; FUCK, M. P. Política de nanotecnología en Brasil: a 10 años de las primeras redes. In: **Perspectivas sobre el desarrollo de las nanotecnologías en América Latina**. FOLADORI, G.; INVERNIZZI, N.; LAU, E. Z. México, Miguel Ángel Porrúa, 2012.

INTERNACIONAL ORGANIZATION STANDARDIZATION (ISO). Technical committees -ISO/TC 229 -Nanotechnologies. Disponível em:  
 <[http://www.iso.org/iso/standards\\_development/technical\\_committees/other\\_bodies/iso\\_technical\\_committee.htm?commid=381983](http://www.iso.org/iso/standards_development/technical_committees/other_bodies/iso_technical_committee.htm?commid=381983)>. Acesso em: 30 set. 2013.

JOHNSON, I. M. Bibliometrics and the brain dead. **Information Development**, v. 27, n. 2, p. 92–93, 5 jan. 2011.

KIERNAN, V. Diffusion of News about Research. **Science Communication**, v. 25, n. 1, p. 3–13, 9 jan. 2003.

KITCHIN, R. Big Data, new epistemologies and paradigm shifts. **Big Data & Society**. April–June 2014: 1–12.

KNORR-CETINA, K. **The Manufacture of Knowledge**: an essay on the Constructivist and Contextual Nature of Science. Oxford: Pergamon Press, 1981.

KOSTOFF, R. N.; KOYTCHEFF, R. G.; LAU, C. G. Y. Global nanotechnology research literature overview. **Technological Forecasting and Social Change**, v. 74, n. 9, p. 1733–1747, 2007.

KRISHNA, V. V.; WAAST, R.; GAILLARD, J. The Changing Structure of Science in Developing Countries. **Science Technology & Society**, v. 5, n. 2, p. 209–224, 1 mar. 2000.

LAKATOS, I. O Falseamento e a Metodologia dos Programas de Pesquisa. In: Lakatos, I.; Musgrave, A. (Eds.). **A Crítica e o desenvolvimento do conhecimento**. São Paulo: Cultrix, 1979.

\_\_\_\_\_. **História de las ciências y sus reconstrucciones racionales**. Madrid: Tecnos, 1987.

LARÉDO, P.; MUSTAR, P. Public Sector Research: A Growing Role in Innovation Systems. **Minerva**, v. 42, n. 1, p. 11–27, 1 mar. 2004.

LATOURETTE, B.; WOOLGAR, S. **Laboratory life**: the social construction of scientific facts. Beverly Hills: Sage, 1979.

LEE, S.; BOZEMAN, B. The Impact of Research Collaboration on Scientific Productivity. **Social Studies of Science**, v. 35, n. 5, p. 673–702, 10 jan. 2005.

LEWENSTEIN, B. V. The meaning of public understanding of science in the United States after World War II. **Public Understanding of Science**, v. 1, p. 45-68, 1992.

LIU, T.; WANG, F.; AGRAWAL, G. Stratified sampling for data mining on the deep web. **Frontiers of Computer Science**, v. 6, n. 2, p. 179–196, 2012.

LISBOA FILHO, P. N; MONTEIRO, M. A. Análises da nanociência e nanotecnologia na perspectiva de ciência, tecnologia e sociedade CTS. **MOMENTO - Revista de Física**, v. 46, n. E, p. 79-91, 2013. Disponível em: <<http://hdl.handle.net/11449/135290>>. Acesso em: 10 out 2016.

MACIAS-CHAPULA, C. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ci. Inf.**, v. 27, n. 2, p. 134–140, ago. 1998.

MACIEL, M. L. Ciência, tecnologia e inovação: a relação entre conhecimento e desenvolvimento. **BIB**, n. 54, p. 67–80, sem 2002.

\_\_\_\_\_. Estímulos e desestímulos à divulgação do conhecimento científico. In: **Conhecimentos e redes: sociedade, política e inovação**. Baumgarten, M. (Org.). Porto Alegre: UFRGS, 2005.

MARCONDES, C.H. et al. **Bibliotecas digitais saberes e práticas**. Salvador, BA : EDUFBA; Brasília; IBICT. 2005.

MARTINO, J. P. **Technological forecasting for decision making**. New York: Mcgraw-Hill, 1993.

MARTINS, P. R. Nanotecnologia e Meio Ambiente Para Uma Sociedade Sustentável. In: (org.). **Nanotecnologia, Sociedade e Meio Ambiente**. São Paulo: SP, Xamã, 2006. pp 114-132.

MATIAS, M. S. O. **Base referencial para o povoamento de repositórios institucionais: coleta automatizada de metadados da Plataforma Lattes**. 2016. 94 p. Dissertação (Mestrado em Gestão de Organizações e Sistemas Públicos) – Universidade Federal de São Carlos, São Carlos, 2016.

MEYER, M.; DEBACKERE, K.; GLANZEL, W. Can applied science be good science? Exploring the relationship between patent citations and citation impact in nanoscience. **Scientometrics**, v.85, n.2, p.527-539, 2010.

MCINERNY, G. J.; et. al. Information visualisation for science and policy: engaging users and avoiding bias. **Trends in ecology & evolution**, v. 29, n. 3, p. 148–57, 2014.

MINISTÉRIO DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO (MCTI). A nanotecnologia no Brasil. Iniciativa Brasileira de Nanotecnologia. 2014. Disponível em: <<http://nano.mct.gov.br/a-nanotecnologia-no-brasil/>>. Acesso em: 12 set. 2014.

\_\_\_\_\_. Estr egia Nacional de Ci ncia, Tecnologia e Inova o 2012 -2015. Balan o das atividades estruturantes, 2012. Dispon vel em: <<http://livroaberto.ibict.br/docs/218981.pdf>>. Acesso em: 30 set. 2014.

MENA-CHALCO, J. P. ScriptLattes. S o Paulo, SP: [s.n.].

MENA-CHALCO, J. P.; CESAR-JR, R. M. ScriptLattes: an open-source knowledge extraction sytem from the lattes plataform. v. 15, n. 4, p. 31–39, 2009.

MERTON, R. K. **Sociologia: teoria e estrutura**. S o Paulo: Editora Mestre Jou: 1970.

MIGUEL, P. A. C. Estudo de caso na engenharia de produ o: estrutura o e recomenda es para sua condu o. **Produ o**, v. 17, n. 1, p. 216–229, abr. 2007.

MILANEZ, D. H. **Nanotecnologia**: indicadores tecnol gicos sobre os avan os em materiais a partir da an lise de documentos de patentes. Disserta o (Mestrada em Ci ncias dos Materiais). Universidade Federal de S o Carlos, S o Carlos, 2011.

MONTGOMERY, D. C.; RUNGER, G. C. **Estat stica aplicada e probabilidade para engenheiros**. 4. ed. Rio de Janeiro: LTC, 2009.

MOREIRA, M. L.; VELHO, L. P s-gradua o no Brasil: da concep o “ofertista linear” para “novos modos de produ o do conhecimento” implica es para avalia o. **Avalia o**, v. 13, n. 3, p. 625–645, 2008.

MOREL, R. L. M. **Ci ncia e Estado**: a pol tica cient fica no Brasil. S o Paulo: T. A. Queiroz, 1979. S o Paulo: T. A. Queiroz, 1979.

MUGNAINI, R. **M tricas da produ o cient fica**. 25 mar. 2015. Dispon vel em: [http://cvirtual-ccs.bvsalud.org/tiki-read\\_article.php?articleId=636](http://cvirtual-ccs.bvsalud.org/tiki-read_article.php?articleId=636). Acesso em: 10 abr. 2015.

NANOWERK. **Roadmaps at 2015 on nanotechnology application in the sectors of: materials, health & medical systems, energy**. Dispon vel em: [http://www.nanowerk.com/nanotechnology/reports/nano\\_reports.php?page=2&cat=General](http://www.nanowerk.com/nanotechnology/reports/nano_reports.php?page=2&cat=General) . Acesso em: 30 set. 2013.

NUCLEO DE INFORMA O TECNOL GICA (NIT). 2013. Dispon vel em: <<http://www.nit.ufscar.br/>>. Acesso em: 30 set. 2013.

NUNES, J. A. Para al m das duas culturas: tecnoci ncias, tecnoculturas e teoria cr tica. **Revista Cr tica de Ci ncias Sociais**, v. 52, p. 15–59, 1999.

ORGANIZACAO MUNDIAL DA PROPRIEDADE INTELECTUAL (OMPI). ** ndice global de inova o**. 2016. Dispon vel em:<<https://www.globalinnovationindex.org/gii-2016-report>>. Acesso em: 02 nov. 2016.

O'REILLY. **O que   Web 2.0?** Dispon vel em: <[www.oreilly.com](http://www.oreilly.com)>.

OKUBO, Y. **Bibliometric indicators and analysis of research systems**: methods and examples. Paris, OECD, 1997.

OLIVÉ, L. Problemas axiológicos y éticos de la nanotecnología. **Mundo Nano, Revista Interdisciplinaria em Nanociencias y Nanotecnología**, v.1, n.2, jun. 2009, pp 48-60.

OLIVEIRA, E. F.; FERREIRA, K. E. Fontes de informação on line em arquivologia: uma avaliação métrica. **Biblos**, Rio Grande, v. 23, n. 2, p. 69-76, 2009.

ORTELLADO, P. As políticas nacionais de acesso à informação científica. **Liinc em Revista**, v.4, n.2, 2008, p. 186-195.

PACKER, A. L. **SciELO -15 Anos de Acesso Aberto [livro eletrônico]**: um estudo analítico sobre acesso aberto e comunicação científica. Paris: Unesco, 2014.

PASTRANA, H. F; ÁVILA, A.; MORENO, G. Nanotecnología, patentes y la situación em America Latina. **Mundo Nano**, v. 5. n. 9, jul-dic, 2012. pp 57-67.

POPPER, K. **The logic of scientific discovery**. New York: Harper & Row, 1968.

PORTOCARRERO, V. **Filosofia, história e sociologia das ciências**: abordagens contemporâneas. Rio de Janeiro: Fiocruz, 1994.

QUONIAM, L. et. al. Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil. **Ci. Inf.**, v. 30, n. 2, p. 20–28, ago. 2001.

RATHER; M; RATHER, D. **Nanotechnology**: a gentle introduction to the big idea. Prentice Hall, 2002.

REGALADO, A. Brazilian science: riding a gusher. **Science**, n. 330, p. 306–312, 2010.

RECUERO, R. **Redes sociais na internet**. Porto Alegre: Sulina, 2009.

RICHARDSON, R. J. **Pesquisa social**: métodos e técnicas. 3. ed. São Paulo: Atlas, 1999.

SAGAN, C. **O mundo assombrado pelos demônios**: a ciência vista como uma vela no escuro. São Paulo: Companhia das Letras, 2006.

SALERNO, M.; LANDONI, P.; VERGANTI, R. Designing foresight studies for Nanoscience and Nanotechnology (NST) future developments. **Technological Forecasting and Social Change**, v. 75, n. 8, p. 1202–1223, out. 2008.

SCHULZ, P. A. B.; LEIVA, D. R.; MILANEZ, D. H. Introdução aos nanomateriais. **In: Engenharia de materiais para todos**. 2. ed. São Carlos: Edufscar, 2014. p. 220.

SILVA, F. B. M. E. Ciência da informação: uma ciência para a informação científica e tecnológica?. **Informação & Sociedade**: Estudos, v. 17, n. 3, p. 93-96, 2007. Disponível em: <<http://www.brapci.ufpr.br/brapci/v/a/7627>>. Acesso em: 21 Out. 2016.

SILVA, F. M.; SMIT, J. W. Organização da informação em sistemas eletrônicos abertos de informação científica e tecnológica: análise da plataforma lattes. **Perspect. ciênc. inf. [online]**, v. 14, n. 1, p. 77–98, 2009.

- SILVA, T. E.; TOMAÉL, M. I. Fontes de informação na Internet: a literatura em evidência. In: TOMAÉL, M. I.; VALENTIM, M. L. P. (Orgs.). **Avaliação de fontes de informação na Internet**. Londrina: Eduel, 2004. p. 01-17.
- SOUZA, C. Políticas públicas: uma revisão da literatura. **Sociologias**, v. 8, n. 16, p. 20–45, 2006.
- SPINAK, E. Scientometric indicators. **Ci. Inf.**, v. 27, n. 2, p. 141–148, 1998.
- TARGINO, M. G.; NEYRA, O. N. B. Ciência, divulgação científica e eventos técnicos–científicos. CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO. **Anais...**Brasília: 2006
- TAVARES, R. L.; PIMENTEL, M.; ARAUJO, R. M. Information Visualization in Political Discussions. **Brazilian Symposium on Collaborative Systems**, p. 75–83, out. 2012.
- UNIVERSIDADE DO PORTO. História do acesso aberto. 2012. Disponível em: [https://sigarra.up.pt/up/pt/WEB\\_GESSI\\_DOCS.download\\_file?p\\_name=F-1080500796/4\\_Historia\\_do\\_acesso\\_aberto.pdf](https://sigarra.up.pt/up/pt/WEB_GESSI_DOCS.download_file?p_name=F-1080500796/4_Historia_do_acesso_aberto.pdf). Acesso em: 17 out 2016.
- VALERIO, P. M.; PINHEIRO, L. V. R. Da comunicação científica à divulgação. **TransInformação**, v. 20, n. 2, p. 159–169, ago. 2008.
- VANTI, N. A. P. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da Informação**, Brasília, v. 31, n. 2, p. 152-162, maio/ago. 2002.
- VAZ, J. C.; RIBEIRO, M. M.; MATHEUS, R. Dados governamentais abertos e seus impactos sobre os conceitos e práticas de transparência no Brasil. **Cadernos PPGAU – UFBA**, v. 9, edição especial (2010) -Democracia e Interfaces Digitais para a Participação Pública. Salvador, 2010, v. 9, p. 45-62: 20 set. 2016.
- VELHO, L. **Indicadores científicos: aspectos teóricos y metodológicos**. 1994.
- VESSURI, H. M. C. Perspectivas recientes em el estudio social de la ciência. **Interciência**, n. 2, p. 60-68, mar-apr, 1991.
- VIEIRA, J. M. L.; CORREA, R. F. Visualização da Informação na construção de interfaces amigáveis para sistemas de recuperação de informação. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 16, n. 32, p. 73-93, 2011.
- VIEIRA, J. M. L.; PINHO, F. A. A contribuição da organização e da visualização da informação para os sistemas de recuperação de informação. **Informação & Informação**, Londrina, v. 20, n. 1, p. 110-136, jun. 2015.
- VLAB4U. Nano 2014. 2014. Disponível em: <<http://vlab4u.info/nano2014>>. Acesso em: 20 jun. 2015.
- \_\_\_\_\_. Nano 2014. 2015. Disponível em: <<http://vlab4u.info/nano2014>>. Acesso em: 20 jun. 2015.

VOGT, C.; MORALES, A. P. O discurso dos indicadores de C&T e de sua percepção. **Revista eletrônica de jornalismo científico**, 10 mar. 2015.

WAGNER, C. S. **The new invisible college**. Washington, DC: Brookings Institution Press, 2008.

WIKTIONARY. Category: English words prefixed with nano. 2013. Disponível em: <[http://en.wiktionary.org/wiki/Category: English\\_wor](http://en.wiktionary.org/wiki/Category:English_wor)>. Acesso em: Oct. 14, 2013.

YIN, R. K. **Estudo de caso: planejamento e métodos**. 2. ed. Porto Alegre: Bookman, 2001.

ZHANG, C. et. al. Research collaboration in health management research communities. **BMC Medical Informatics and Decision Making**, v. 13, n. 52, 2013.

ZIMAN, J. M. **Real science: what it is and what it means**. Cambridge: Cambridge University Press, 2000.

\_\_\_\_\_. Percepción publica y cultura científica: ciência y sociedade civil. **Revista Iberoamericana de Ciencia, Tec Soc**. Buenos Aires, v.1, n.1, set. 2003.

## APÊNDICE A – Esquema para a construção da expressão de busca

Conjuntos de palavras	Recuperação individual	Expressão incrementada	Recuperação	Acrescimos
(nanoparticle or nanoparticles or nanopartícula or nanopartículas)	4267	=C2	4267	
(nanotechnology or nanotechnologies or nanotecnologia or nanotecnologias)	2338	=(C2&" or ")&C3	5726	1459
(nanostructured or nanostructureds or nanoestruturado or nanoestruturados)	2101	=(E3&" or ")&C4	6476	750
(nano or nanos)	1829	=(E4&" or ")&C5	7201	725
(nanocomposite or nanocomposites or nanocompósito or nanocompósitos)	1822	=(E5&" or ")&C6	7751	550
(sol_gel or sóis_géis)	1586	=(E6&" or ")&C7	8293	542
(nanotube or nanotubes or nanotubo or nanotubos)	1579	=(E7&" or ")&C8	8757	464
(nanocatalysis or catalysts or nanocatálise or nanocatalisadores)	1381	=(E8&" or ")&C9	9311	554
(quantum_dot or quantum_dots or ponto_quântico or pontos_quânticos)	741	=(E9&" or ")&C10	9603	292
(nanomaterial or nanomaterials or nanomaterial or nanomateriais)	716	=(E10&" or ")&C11	9640	37
(nanocrystal or nanocrystals or nanocristal or nanocristais)	660	=(E11&" or ")&C12	9711	71
(nanocrystalline or nanocrystallines or nanocristalino or nanocristalinos)	640	=(E12&" or ")&C13	9798	87
(graphene or graphenes or grafeno or grafenos)	529	=(E13&" or ")&C14	9897	99
(nanowire or nanowires or nanofio or nanofios)	477	=(E14&" or ")&C15	9956	59
(quantum_well or quantum_wells or poço_quântico or poços_quânticos)	460	=(E15&" or ")&C16	10092	136
(nanoscience or nanosciences or nanociência or nanociências)	459	=(E16&" or ")&C17	10120	28
(nanofiber or nanofibers or nanofibra or nanofibras)	404	=(E17&" or ")&C18	10178	58
(nanocapsule or nanocapsules or nanocapsula or nanocapsulas)	396	=(E18&" or ")&C19	10281	103
(nanoparticulate or nanoparticulates or nanoparticulado or nanoparticulados)	359	=(E19&" or ")&C20	10340	59
(nanoemulsion or nanoemulsions or nanoemulsão or nanoemulsões)	354	=(E20&" or ")&C21	10472	132
(nanosphere or nanospheres or nanoesfera or nanoesferas)	275	=(E21&" or ")&C22	10499	27
(nanobiotechnology or nanobiotechnologies or nanobiotecnologia or nanobiotecnologias)	254	=(E22&" or ")&C23	10546	47
(nanopowder or nanopowders or nanopó or nanopós)	239	=(E23&" or ")&C24	10555	9
(fullerene or fullerenes or fullereno or fullerenos)	320	=(E24&" or ")&C25	10592	37
(nanoporous or nanoporoso or nanoporosos)	178	=(E25&" or ")&C26	10617	25
(dendrimer or dendrimers or dendrímero or dendrímeros)	169	=(E26&" or ")&C27	10655	38
(nanoindentation or nanoindentations or nanoindentação or nanoindentações)	166	=(E27&" or ")&C28	10708	53
(nanoribbon or nanoribbons or nanofita or nanofitas)	160	=(E28&" or ")&C29	10711	3
(carbon_nanostructure or carbon_nanostructures or nanoestrutura_de_carbono or nanoestruturas_de_carbono)	153	=(E29&" or ")&C30	10720	9
(nanoclay or nanoclays or nanoargila or nanoargilas)	135	=(E30&" or ")&C31	10730	10
(nanorod or nanorods or nanohaste or nanohastes)	134	=(E31&" or ")&C32	10735	5
(nanomedicine or nanomedicines or nanomedicina or nanomedicinas)	131	=(E32&" or ")&C33	10742	7

(nanocluster or nanoclusters or nanocacho or nanocachos)	130	=(E33&" or ")&C34	10766	24
<b>PARA A EXPRESSÃO DE BUSCA UTILIZOU-SE ATÉ AQUI</b>				
(nanosystem or nanosystems or nanosistema or nanosistemas)	123	=(E34&" or ")&C35	10782	16
(nanodevice or nanodevices or nanodispositivo or nanodispositivos)	116	=(E35&" or ")&C36	10795	13
(nanofilm or nanofilms or nanofilme or nanofilmes)	104	inserido pelo especialista		
(nanohybrid or nanohybrids or nanohíbrido or nanohíbridos)	99	=(E36&" or ")&C38	10802	7
(nanofluid or nanofluids or nanofluido or nanofluidos)	89	=(E38&" or ")&C39	10832	30
(nanotoxicology or nanotoxicologies or nanotoxicologia or nanotoxicologias)	89	=(E39&" or ")&C40	10840	8
(nanomechanical or nanomechanicals or nanomecânico or nanomecânicos)	84	=(E40&" or ")&C41	10852	12
(nanopore or nanopores or nanoporo or nanoporos)	83	=(E41&" or ")&C42	10864	12
(nanomagnetism or nanomagnetisms or nanomagnetismo or nanomagnetismos)	74	=(E42&" or ")&C43	10867	3
(nanosensor or nanosensors or nanosensores)	72	=(E43&" or ")&C44	10870	3
(nanolithography or nanolithographies or nanolitografia or nanolitografias)	63	=(E44&" or ")&C45	10872	2
(nanomechanic or nanomechanics or nanomecânica or nanomecânicas)	60	=(E45&" or ")&C46	10873	1
(nanoelectronic or nanoelectronics or nanoeletrônico or nanoeletrônicos)	56	=(E46&" or ")&C47	10874	1
(nanofabrication or nanofabrications or nanofabricação or nanofabricações)	52	=(E47&" or ")&C48	10875	1
(nanodiamond or nanodiamonds or nanodiamante or nanodiamantes)	47	inserido pelo especialista		
(nanoceramic or nanoceramics or nanocerâmica or nanocerâmicas)	41	=(E48&" or ")&C50	10876	1
(nanogel or nanogels or nanogéis)	39	=(E50&" or ")&C51	10877	1
(nanomanufacturing or nanomanufacturings or nanofabricação or nanofabricações)	34	=(E51&" or ")&C52	10877	0
(nanoprecipitation or nanoprecipitations or nanoprecipitação or nanoprecipitações)	34	=(E52&" or ")&C53	10877	0
(nanoreactor or nanoreactors or nanoreator or nanoreatores)	34	=(E53&" or ")&C54	10882	5
(nanoparticle_composite or nanoparticles_composite or compósito_de_nanopartícula or compósito_de_nanopartículas)	33	=(E54&" or ")&C55	10882	0
(nanocone or nanocones)	30	=(E55&" or ")&C56	10885	3
(nanodot or nanodots or nanoponto or nanopontos)	27	=(E56&" or ")&C57	10886	1
(nanophotonic or nanophotonics or nanofotônico or nanofotônicos)	27	=(E57&" or ")&C58	10889	3
(nanostructured_lipid_carrier or nanostructured_lipid_carriers or portador_lipídico_nanoestruturado or portadores_lipídicos_nanoestruturados)	27	=(E58&" or ")&C59	10889	0
(bionanotechnology or bionanotechnologies or bionanotecnologia or bionanotecnologias)	26	=(E59&" or ")&C60	10894	5
(nanodispersion or nanodispersions or nanodispersão or nanodispersões)	26	=(E60&" or ")&C61	10898	4
(nanoscratch or nanoscratches or nanorisco or nanoriscos)	26	=(E61&" or ")&C62	10900	2
(diamondoid or diamondoids or diamantoíde or diamantoídes)	25	=(E62&" or ")&C63	10912	12
(nanoring or nanorings or nanoanel or nanoanéis)	25	=(E63&" or ")&C64	10912	0
(nanophase or nanophases or nanofase or nanofases)	23	=(E64&" or ")&C65	10915	3
(nanoformulation or nanoformulations or nanoformulação or nanoformulações)	21	=(E65&" or ")&C65	10921	6

		)&C66		
(nanomachine or nanomachines or nanomáquina or nanomáquinas)	21	=(E66&" or )&C67	10933	12
(nanoprobe or nanoprobos or nanosonda or nanosondas)	21	=(E67&" or )&C68	10935	2
(nanoshell or nanoshells or nanoconcha or nanoconchas)	20	=(E68&" or )&C69	10936	1
(nanopatterning or nanomicrousinagem or nanomicrousinagens)	20	inserido pelo especialista		
(nanometrology or nanometrologies or nanometrologia or nanometrologias)	19	=(E69&" or )&C71	10940	4
(nanotribology or nanotribologies or nanotribologia or nanotribologias)	17	=(E71&" or )&C72	10940	0
(nanobalance or nanobalances or nanobalança or nanobalanças)	15	=(E72&" or )&C73	10943	3
(ultrananocrystalline_diamond or ultrananocrystallines_diamonds or diamante_ultrananocristalino or diamantes_ultrananocristalinos)	13	inserido pelo especialista		
(nanopolymer or nanopolymers or nanopolímero or nanopolímeros)	12	=(E73&" or )&C75	10943	0
(nanopharmaceutical or nanopharmaceuticals or nanofarmacêutico or nanofarmacêuticos)	11	=(E75&" or )&C76	10944	1
(nanodomain or nanodomains or nanodomínio or nanodomínios)	9	=(E76&" or )&C77	10946	2
(nanosuspension or nanosuspensions or nanosuspensão or nanosuspensões)	9	=(E77&" or )&C78	10946	0
(nanoelectromechanical or nanoelectromechanicals or nanoeletromecânico or nanoeletromecânicos)	8	=(E78&" or )&C80	10949	3
(nanorobotic or nanorobotics or nanorobótica or nanorobóticas)	7	=(E80&" or )&C81	10950	1
(nanobar or nanobars or nanobarra or nanobarras)	6	=(E81&" or )&C82	10950	0
(nanocomponent or nanocomponents or nanocomponente or nanocomponentes)	6	=(E82&" or )&C83	10950	0
(nanocontainer or nanocontainers)	6	=(E83&" or )&C84	10950	0
(nanograined or nanograineds or nanogranulação or nanogranulações)	6	=(E84&" or )&C85	10950	0
(nanotextured or nanotextureds or nanotexturizado or nanotexturizados)	6	=(E85&" or )&C86	10952	2
(nanoquasicrystalline or nanoquasicrystallines or nanoquasicristalino or nanoquasicristalinos)	5	inserido pelo especialista		
(nanobiology or nanobiologies or nanobiologia or nanobiologias)	5	=(E86&" or )&C88	10953	1
(nanocore or nanocores or nanonúcleo or nanonúcleos)	5	=(E88&" or )&C89	10973	20
(nanocylinder or nanocylinders or nanocilindro or nanocilindros)	5	=(E89&" or )&C90	10973	0
(nano-object or nano-objects or nano-objeto or nanoobjetos)	5	=(E90&" or )&C91	10973	0
(nanobeam or nanobeams or nanoviga or nanovigas)	4	=(E91&" or )&C92	10973	0
(nanohorn or nanohorns or nanochifre or nanochifres)	4	=(E92&" or )&C93	10974	1
(nanoimprint_lithography or nanoimprint_lithographies or litografia_de_nanoimpressão or litografias_de_nanoimpressão)	4	=(E93&" or )&C94	10974	0
(nanomanipulator or nanomanipulators or nanomanipulador or nanomanipuladores)	4	=(E94&" or )&C95	10974	0
(nanomotor or nanomotors or nanomotores)	4	=(E95&" or )&C96	10974	0
(nanoplasm or nanoplasmis or nanoplasma or nanoplasmas)	4	=(E96&" or )&C97	10974	0
(nanospring or nanosprings or nanomola or nanomolas)	4	=(E97&" or )&C98	10974	0
(nanobody or nanobodies or nanocorpo or nanocorpos)	3	=(E98&" or )&C99	10976	2

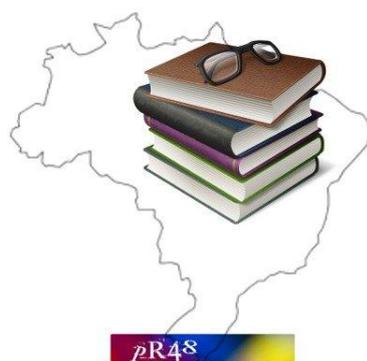
(nanocar or nanocars or nanocarro or nanocarros)	3	=(E99&" or ")&C100	10976	0
(nanotomography or nanotomographies or nanotomografia or nanotomografias)	3	=(E100&" or ")&C101	10977	1
(nanotox or nanotoxs)	3	=(E101&" or ")&C102	10977	0
(nanobattery or nanobatteries or nanobateria or nanobaterias)	2	=(E102&" or ")&C103	10977	0
(nanobead or nanobeads or nanopérola or nanopérolas)	2	=(E103&" or ")&C104	10977	0
(nanochip or nanochips)	2	=(E104&" or ")&C105	10977	0
(nanodiffraction or nanodiffractions or nanodifração or nanodifrações)	2	=(E105&" or ")&C106	10977	0
(nanofoam or nanofoams or nanoespuma or nanoespumas)	2	=(E106&" or ")&C107	10977	0
(nanomechatronic or nanomechanics or nanomecatrônico or nanomecatrônicos)	2	=(E107&" or ")&C108	10977	0
(nanopeapods)	2	=(E108&" or ")&C109	10977	0
(nanosurgery or nanosurgeries or nanocirurgia or nanocirurgias)	2	=(E109&" or ")&C110	10977	0
(nanoimprinting or nanoimprintings or nanoimpressão or nanoimpressões)	1	=(E110&" or ")&C111	10977	0
(nanomachining or nanomachinings or nanousinagem or nanousinagens)	1	=(E111&" or ")&C112	10977	0
(nanoparticle_tracking_analysis or nanoparticles_tracking_analysis or análise_de_monitoramento_de_nanopartícula or análise_de_monitoramento_de_nanopartículas)	1	=(E112&" or ")&C113	10977	0
(nanosource or nanosources or nanofonte or nanofontes)	1	=(E113&" or ")&C114	10977	0
(nanotitania or nanotitanio or nanotitanias)	1	inserido pelo especialista		
(bionanotechnology or bionanotechnologies or bionanotecnologia or bionanotecnologias)	0	=(E114&" or ")&C116	10977	0
(nanocapturer or nanocapturers or nanocapturador or nanocapturadores)	0	=(E116&" or ")&C117	10977	0
(nanocrystalline_ceramic or nanocrystallines_ceramics or cerâmica_nanocristalina or cerâmicas_nanocristalinas)	0	=(E117&" or ")&C118	10977	0
(nanodragster or nanodragsters)	0	=(E118&" or ")&C119	10977	0
(nanogranulated or nanogranulateds or nanogranulado or nanogranulados)	0	=(E119&" or ")&C120	10977	0
(nanofitter or nanofitters or nanomontador or nanomontadores)	0	=(E120&" or ")&C121	10977	0
(nanopollution or nanopollutions or nanopoluição or nanopoluições)	0	=(E121&" or ")&C122	10977	0
(nanotige or nanotiges)	0	=(E122&" or ")&C123	10977	0

## ANEXO 1 – Petição para a retirada dos captchas dos currículos da Plataforma Lattes



Fonte: CNPQ (2016)

[Jesús Mena-Chalco - Santo Andre, Brasil<sup>35</sup>](#)



O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), órgão do Ministério da Ciência, Tecnologia e Inovação (MCTI) para o incentivo à pesquisa no Brasil, teve a excelente iniciativa de criação de um banco de registros nacional de currículos de pesquisadores nos meados dos anos 1980. A primeira captação de dados foi realizada em papel, mas com algumas etapas

apoiadas por um sistema computacional. Ao final dos anos 1990 o CNPq, conjuntamente com universidades e empresas, desenvolveu a Plataforma Lattes visando à padronização dos dados já coletados, e também a uma forma mais ágil de registro dos currículos de pesquisadores no âmbito do MCTI.

Hoje, é inegável que o currículo Lattes é um padrão nacional e internacional que mantém um registro da vida pregressa e atual de estudantes, pesquisadores e acadêmicos no Brasil. A Plataforma Lattes, destaque no contexto internacional como um repositório curricular ímpar, contém atualmente mais de 4 milhões de currículos cadastrados e mais de 19 milhões de publicações registradas. Esta Plataforma transformou-se em um repositório único e precioso para a coleta de dados e para a análise de informações relacionadas à ciência brasileira, consolidando-se em um modelo que assegura ao CNPq confiabilidade nas evoluções implementadas nessa ferramenta.

<sup>35</sup> Disponível em: <https://www.change.org/p/peti%C3%A7%C3%A3o-para-a-retirada-dos-captchas-dos-curr%C3%ADculos-da-plataforma-lattes>. Acesso em: 21 out 2016.

Até 28 de abril de 2015 os currículos Lattes eram livremente disponibilizados para consulta e coleta de dados. Após essa data o CNPq adotou um sistema de verificação/validação que emprega o recurso do captcha para o acesso aos currículos. A adoção desse sistema, segundo o CNPq, deve-se à demanda de usuários para evitar a extração de informações e a sua publicação indevida por sites não autorizados (entenda-se como sites comerciais, já relatado na seguinte página).

Os dados nos currículos Lattes são informações públicas (a plataforma preserva as informações individuais) que muitas universidades, centros e institutos de pesquisa, professores e alunos têm utilizado para fins de pesquisa não-comercial. Em particular, todos os programas brasileiros de pós-graduação têm utilizado a Plataforma Lattes para coleta de dados e envio de relatórios de produção acadêmica à Plataforma Sucupira - nova ferramenta integradora dos dados e informações dos programas de pós-graduação do país, lançada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) em março de 2014 para ser a fonte de referência do Sistema Nacional de Pós-Graduação. Para exemplificar, em 2012, foram avaliados pela CAPES 3.342 programas de pós-graduação (entre doutorados e mestrados, acadêmicos ou profissionalizantes). Ao todo, foram consideradas informações sobre a produção bibliográfica de aproximadamente 72.802 professores doutores (permanentes, colaboradores ou visitantes). Com a adoção dos captchas, serão necessárias mais de 101 horas-homem somente para digitar os caracteres a fim de acessar uma única vez os mesmos 72.802 currículos (assumindo 5 segundos para inserir os caracteres exigidos). A leitura e coleta de dados certamente demanda ainda mais tempo. Se considerarmos 1 minuto para ler e extrair a informação bibliográfica de cada currículo, serão necessárias mais de 1.213 horas.

Os currículos Lattes são extensamente utilizados pela comunidade acadêmica. Se apenas 1% (i.e., 40 mil) dos currículos forem consultados diariamente, serão necessárias mais de 55 horas-homem para simplesmente digitar os caracteres para validar o acesso pelo captcha. O custo de energia e recursos que devem ser gastos por essas horas-extra têm um grande impacto que merece ser reavaliado pelo CNPq.

É importante frisar que o CNPq tem como política fornecer, para cada instituição, uma única conta para acesso aos dados consolidados de seu corpo de alunos, pesquisadores e professores, mas tal conta não é facilmente franqueada (como indicado no seguinte artigo da Revista FAPESP). Dentre as regras para acesso (pelo serviço web de extração de currículos) determinadas pelo CNPq, a instituição deve informar um endereço IP de onde partirão as requisições e somente esse IP tem permissão de extrair os currículos, além de que é indicado utilizar os dados relacionados apenas à própria instituição. Assim, trabalhos transversais, que exploram mais de uma instituição, não podem ser realizados.

Os captchas e a falta da livre disponibilização dos dados dos currículos Lattes, nos termos da Lei de Acesso à Informação (Nº 12.527), são um freio à promoção da Ciência Brasileira (já discutido na seguinte página). As informações curriculares são bloqueadas para acesso a motores de busca que indexam produções bibliográficas (como o Google Acadêmico e do Microsoft Academic Search). Nenhum currículo é acessado sem a validação pelo captcha. A promoção dos trabalhos dos pesquisadores registrados na Plataforma Lattes é restringida, perdendo-se assim uma oportunidade de incentivar sua visibilidade e aumentar as citações de trabalhos científicos brasileiros. Adicionalmente, possíveis avaliadores estrangeiros terão que submeter-se ao mesmo mecanismo de validação para acessar os currículos, restringindo também o impacto da ciência nacional no exterior. Os captchas desencorajam qualquer tipo de uso que considera dados da Plataforma Lattes, tais como avaliações, identificação de especialistas, agrupamentos de pesquisadores pelas coautorias, cruzamentos de informações, entre outras diversas análises propiciadas pela riqueza de dados dessa

valiosa fonte. Outra questão importante a ser mencionada diz respeito à acessibilidade para pessoas com necessidades especiais. Os captchas são inconvenientes para pessoas com deficiência visual.

Muitos trabalhos acadêmicos têm se beneficiado da Plataforma Lattes por meio do uso de ferramentas de software-livre para a coleta de dados. O uso dos captchas nos currículos inviabiliza pesquisas acadêmicas que visam explorar e descobrir conhecimento científico nos dados nela registrados. Os pesquisadores de dados, como bibliometristas e cientometristas, ficam impossibilitados de efetivar trabalhos caracterizando e/ou evidenciando o crescimento da ciência nacional. Há muita ciência que pode ser feita da análise dos currículos, e essa ciência é fundamental para o melhor planejamento das instituições de ensino e pesquisa.

Esperamos que uma solução para a retirada dos captchas seja encontrada. Uma alternativa para a análise de dados da Plataforma Lattes pode ser oferecida por meio de um servidor computacional a partir do qual o conteúdo da plataforma, atualizado, por exemplo, uma vez por mês, permitiria o acesso de forma sistemática. Plataformas internacionais, como o Digital Bibliography & Library Project (DBPL), da Universidade de Trier, Alemanha, oferecem serviços livres para todos seus dados. Essa abordagem facilitaria o desenvolvimento de softwares que demandam consulta de grandes volumes de dados como ocorre na Plataforma Lattes.

Os dados da Plataforma Lattes são um patrimônio da Ciência Brasileira. O CNPq tem realizado um extraordinário papel ao implementar a Plataforma.

**Resumindo, a adoção de captcha nos currículos Lattes:**

- Gera um custo elevado, oneroso e desnecessário para a coleta de dados;
- Constitui um freio à promoção da Ciência Brasileira;
- Desencoraja a consulta de dados da Plataforma Lattes;
- Diminui a visibilidade e o impacto da produção científica brasileira nos âmbitos nacional e internacional; e
- Impossibilita estudos acadêmicos envolvendo mais de uma instituição, gerando prejuízo à Ciência Brasileira.

**Com estes argumentos, os peticionários apelam ao CNPq para que retire os captchas dos currículos da Plataforma Lattes.**

**Este abaixo-assinado foi entregue para:**

- **Presidente do CNPq**  
Hernan Chaimovich Guralnik
- **Diretor de Gestão e TI do CNPq**  
Luiz Alberto de Freitas Brandão

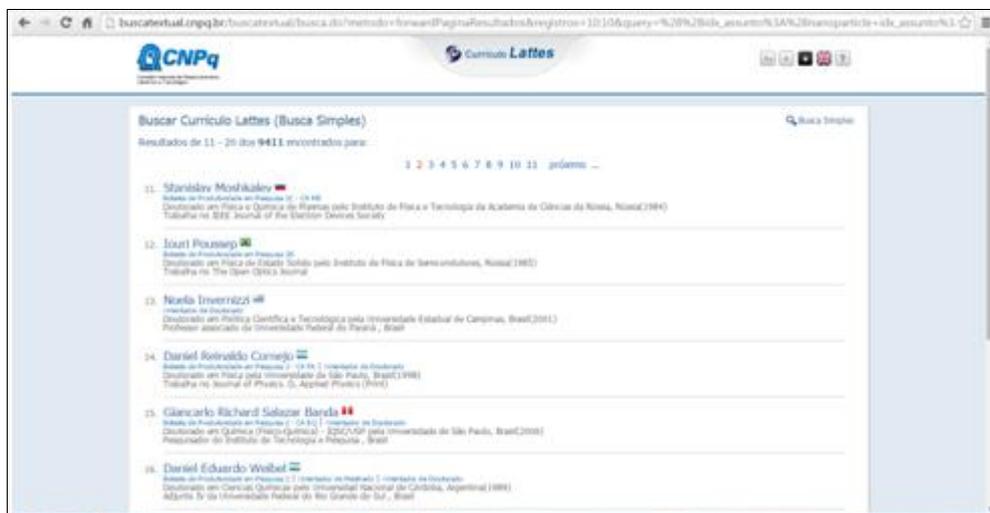
ANEXO 2 – Interface da busca avançada e recuperação dos Currículos Lattes

Interface da busca avançada



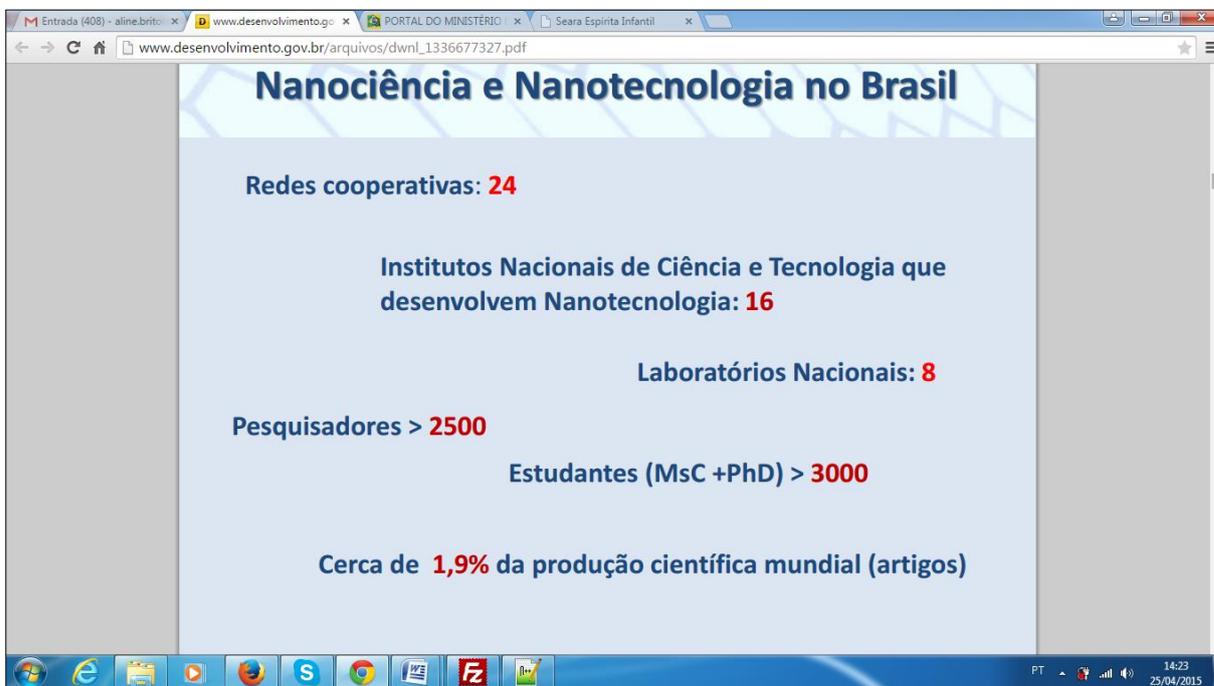
Fonte: CNPQ (2016)

Interface da recuperação dos resultados



Fonte: CNPQ (2016)

## ANEXO 3 – Apresentação do MCTI sobre nanotecnologia



Fonte: CNPQ (2016)

## ANEXO 4 – Arquivo de configuração para o ScriptLattes

```

# ----- #
# INFORMAÇÕES GERAIS #
# ----- #
global-nome_do_grupo = Teste 01 do ScriptLattes (período 2006-2014)
global-arquivo_de_entrada = ./exemplo/teste-01.list
global-diretorio_de_saida = ./exemplo/teste-01/
global-email_do_admin = jesus.mena@ufabc.edu.br
global-idioma = PT # ainda não implementada na versão Python

global-itens_desde_o_ano = 2006 # se não for indicado o ano então serão consideradas as publicações
global-itens_ate_o_ano = 2014 # identificadas em todos os CVs Lattes
global-itens_por_pagina = 1000
global-criar_paginas_jsp = nao
global-google_analytics_key = UA-5266248-1 # usado com contas do google analytics
global-prefixo = teste-01 # usado nos arquivos temporários (úteis para futuras análises)

global-salvar_informacoes_em_formato_xml = sim # armazenado no arquivo 'database.xml'

# Atribuição dos Qualis
global-identificar_publicacoes_com_qualis = nao
global-arquivo_qualis_de_periodicos = # ./exemplo/qualis_computacao_periodicos_2012.csv
global-arquivo_qualis_de_congressos = # ./exemplo/qualis_computacao_conferencias_2012.csv

# cache de CVs Lattes (geralmente utilizado para grandes grupos de CVs).
# Se não for indicado então serão utilizadas as últimas versões dos CVs.
global-diretorio_de_armazenamento_de_cvs = ./exemplo/cache
global-diretorio_de_armazenamento_de_doi = # ./exemplo/doi

# ----- #
# RELATÓRIOS DE PRODUÇÃO EM C,T & A #
# Serão criadas listas de publicações apenas para os tipos = sim #
# ----- #
relatorio-salvar_publicacoes_em_formato_ris = sim # armazenado no arquivo 'publicacoes.ris'

relatorio-incluir_artigo_em_periodico = sim
relatorio-incluir_livro_publicado = sim
relatorio-incluir_capitulo_de_livro_publicado = sim
relatorio-incluir_texto_em_jornal_de_noticia = sim
relatorio-incluir_trabalho_completo_em_congresso = sim
relatorio-incluir_resumo_expandido_em_congresso = sim
relatorio-incluir_resumo_em_congresso = sim
relatorio-incluir_artigo_aceito_para_publicacao = sim
relatorio-incluir_apresentacao_de_trabalho = sim
relatorio-incluir_outro_tipo_de_producao_bibliografica = sim

relatorio-incluir_software_com_patente = sim
relatorio-incluir_software_sem_patente = sim
relatorio-incluir_produto_tecnologico = sim
relatorio-incluir_processo_ou_tecnica = sim
relatorio-incluir_trabalho_tecnico = sim
relatorio-incluir_outro_tipo_de_producao_tecnica = sim

relatorio-incluir_producao_artistica = sim

```

```

# ----- #
# RELATÓRIOS DE ORIENTAÇÕES                                     #
# ----- #
relatorio-mostrar_orientacoes                                = sim

relatorio-incluir_orientacao_em_andamento_pos_doutorado     = sim
relatorio-incluir_orientacao_em_andamento_doutorado        = sim
relatorio-incluir_orientacao_em_andamento_mestrado         = sim
relatorio-incluir_orientacao_em_andamento_monografia_de_especializacao = sim
relatorio-incluir_orientacao_em_andamento_tcc              = sim
relatorio-incluir_orientacao_em_andamento_iniciacao_cientifica = sim
relatorio-incluir_orientacao_em_andamento_outro_tipo       = sim

relatorio-incluir_orientacao_concluida_pos_doutorado        = sim
relatorio-incluir_orientacao_concluida_doutorado           = sim
relatorio-incluir_orientacao_concluida_mestrado            = sim
relatorio-incluir_orientacao_concluida_monografia_de_especializacao = sim
relatorio-incluir_orientacao_concluida_tcc                 = sim
relatorio-incluir_orientacao_concluida_iniciacao_cientifica = sim
relatorio-incluir_orientacao_concluida_outro_tipo          = sim

# ----- #
# RELATÓRIOS ADICIONAIS                                       #
# ----- #
relatorio-incluir_projeto                                   = sim
relatorio-incluir_premio                                   = sim
relatorio-incluir_participacao_em_evento                   = sim
relatorio-incluir_organizacao_de_evento                   = sim
relatorio-incluir_internacionalizacao                      = nao

# ----- #
# GRAFO DE COLABORAÇÕES                                       #
# ----- #
grafo-mostrar_grafo_de_colaboracoes                       = sim
grafo-mostrar_todos_os_nos_do_grafo                       = sim
grafo-considerar_rotulos_dos_membros_do_grupo            = nao # informação dada no arquivo de entrada
.list (quarta coluna)
grafo-mostrar_aresta_proporcional_ao_numero_de_colaboracoes = sim

grafo-incluir_artigo_em_periodico                          = sim
grafo-incluir_livro_publicado                              = sim
grafo-incluir_capitulo_de_livro_publicado                 = sim
grafo-incluir_texto_em_jornal_de_noticia                  = sim
grafo-incluir_trabalho_completo_em_congresso              = sim
grafo-incluir_resumo_expandido_em_congresso               = sim
grafo-incluir_resumo_em_congresso                         = sim
grafo-incluir_artigo_aceito_para_publicacao               = sim
grafo-incluir_apresentacao_de_trabalho                   = sim
grafo-incluir_outro_tipo_de_producao_bibliografica       = sim

grafo-incluir_software_com_patente                        = sim
grafo-incluir_software_sem_patente                        = sim
grafo-incluir_produto_tecnologico                          = sim
grafo-incluir_processo_ou_tecnica                         = sim
grafo-incluir_trabalho_tecnico                            = sim
grafo-incluir_outro_tipo_de_producao_tecnica              = sim

```

```
grafo-incluir_producao_artistica          = sim
grafo-incluir_grau_de_colaboracao        = sim # medida similar ao PageRank

# ----- #
# MAPA DE GEOLOCALIZAÇÃO                  #
# ----- #
mapa-mostrar_mapa_de_geolocalizacao = nao
mapa-incluir_membros_do_grupo          = sim
mapa-incluir_alunos_de_pos_doutorado = nao
mapa-incluir_alunos_de_doutorado      = nao
mapa-incluir_alunos_de_mestrado       = nao # NÃO sugerido pois pode demorar muito baixar todos os CVs
Lattes (Use apenas para grupos menores)

# ----- #
# FIM                                     #
# ----- #
```

## ANEXO 5 – Arquivo de configuração para o Patent2Net

```

#insert below your request in cql format as done in the example
#insert also a compliant name (no space or special characters) for
#the "patents universe" corresponding to the request
request: ta = nanofiber* and ta = nanotub*
#####
# Name of files OR directories
DataDirectory: Nanofiber

#####
#Patent2Net options : Set to *True* or False
#####
# Collecting
# patents list corresponding to the request. e.g the "Patent Universe" (PU)
GatherPatent: True

# patents bibliographic data corresponding to patent list
GatherBiblio: True

# patent content (description, abstract, claims) completing patents bibliographic data
GatherContent: True
OPSGatherContentsv2-Iramuteq: True

#patents families extending the PU to families
GatherFamily: True

#####
# Networks. Produce both online and for download and local use of Gephi
# Inventors' network
InventorNetwork: True

# Applicants' network
ApplicantNetwork: True

# IPCs' network
CrossTechNetwork: True

# Applicants and Inventors network
ApplicantInventorNet: True

# Applicants and IPCs' network
ApplicantCrossTechNetwork: True

# Inventors and IPCs' network
InventorCrossTechNetwork: True

# Countries and IPCs' network
CountryCrossTechNetwork: True

# Network of references for the PU (each reference cited by each patent within the PU, includes external bibliographic refs)
ReferencesNetwork: True

# Network of citation for the PU (each patent cited within the PU cited in the rest of the patents world)

```

CitationsTechNetwork: True

# Network of equivalent patents for each patent within the PU

EquivalentsTechNetwork: True

# The total network

CompleteNetwork: True

# Complete families network

FamiliesNetwork: True

FamiliesHierarchicNetwork: True

#####

# Mindmaps of IPCs both online and for download and local use of freeplane

P2N-FreePlane: True

#####

# export patents as a bibliographic file in bibTex format

FormateExportBiblio: True

#####

# Produce the world maps of inventors, applicants, deposit country

FormateExportAttractivityCartography: True

FormateExportCountryCartography: True

#####

# Patents as a table

FormateExportDataTable: True

FormateExportDataTableFamilies: True

#####

# Prepare interactive pivottable

FormateExportPivotTable: True

#####

# prepare Abstracts, descriptions, claims for text analysis

# for download and local use of Carrot2 and Iramuteq

FusionCarrot2: True

FusionIramuteq2: True