

Bibliometria e Cientometria no Brasil: infraestrutura para avaliação da pesquisa científica na Era do Big Data

Bibliometrics and Scientometrics in Brazil: scientific
research assessment infrastructure in the Era of Big Data

Rogério Mugnaini
Asa Fujino
Nair Yumiko Kobashi
(organizadores)

eca 50 ANOS
ESCOLA DE COMUNICAÇÕES E ARTES
UNIVERSIDADE DE SÃO PAULO

USP



ROGÉRIO MUGNAINI
ASA FUJINO
NAIR YUMIKO KOBASHI
(ORGANIZADORES)

BIBLIOMETRIA E CIENTOMETRIA NO BRASIL:
INFRAESTRUTURA PARA AVALIAÇÃO DA PESQUISA
CIENTÍFICA NA ERA DO BIG DATA

**BIBLIOMETRICS AND SCIENTOMETRICS IN BRAZIL: SCIENTIFIC RESEARCH
ASSESSMENT INFRASTRUCTURE IN THE ERA OF BIG DATA**

DOI: 10.11606/9788572051705

SÃO PAULO
ECA – USP
2017

UNIVERSIDADE DE SÃO PAULO

Reitor Prof. Dr. Marco Antonio Zago

Vice-reitor Prof. Dr. Vahan Agopyan

ESCOLA DE COMUNICAÇÕES E ARTES

Diretora Profa. Dra. Margarida Maria Krohling Kunsch

Vice-Diretor Prof. Dr. Eduardo Henrique Soares Monteiro

5º ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA (5º EBBC)

COMISSÃO ORGANIZADORA

Presidente

Prof. Dr. Rogério Mugnaini

Coordenação Científica

Profa. Dra. Nair Yumiko Kobashi

Coordenação Executiva

Profa. Dra. Asa Fujino

Prof. Dr. Rene Faustino Gabriel Junior

Consultores

Prof. Dr. Fábio Mascarenhas e Silva

Profa. Dra. Jacqueline Leta

Prof. Dr. Raimundo Nonato M. dos Santos



MINISTÉRIO DA
EDUCAÇÃO



É permitida a reprodução parcial ou total desta obra, desde que citada a fonte e autoria, proibindo qualquer uso para fins comerciais.

Catálogo na Publicação

Serviço de Biblioteca e Documentação

Escola de Comunicações e Artes da Universidade de São Paulo

B582m Bibliometria e cientometria no Brasil: infraestrutura para avaliação da pesquisa científica na era do Big Data = Bibliometrics and scientometrics in Brazil: scientific research assessment infrastructure in the era of Big Data / Rogério Mugnaini, Asa Fujino, Nair Yumiko Kobashi (organizadores) – São Paulo: ECA/USP, 2017.
218 p.

ISBN: 978-85-7205-170-5

DOI: 10.11606/9788572051705

1. Bibliometria 2. Cientometria 3. Big Data 4. Pesquisa científica – Brasil
I. Mugnaini, Rogério II. Fujino, Asa III. Kobashi, Nair Yumiko

CDD 21.ed. – 025.21

Identidade visual do 5º EBBC

Rene Faustino Gabriel Júnior

Projeto gráfico e editoração eletrônica

Vitor Borysow

Tradução

Carolina de Góes

Versão

Robert Frank Hanson

Normalização

Angélica de Souza Alves de Paula

Bibliometria e Cientometria no Brasil: infraestrutura para avaliação da pesquisa científica na Era do Big Data

Bibliometrics and Scientometrics in Brazil: scientific
research assessment infrastructure in the Era of Big Data

Rogério Mugnaini
Asa Fujino
Nair Yumiko Kobashi
(organizadores)

eca 50 ANOS
ESCOLA DE COMUNICAÇÕES E ARTES
UNIVERSIDADE DE SÃO PAULO

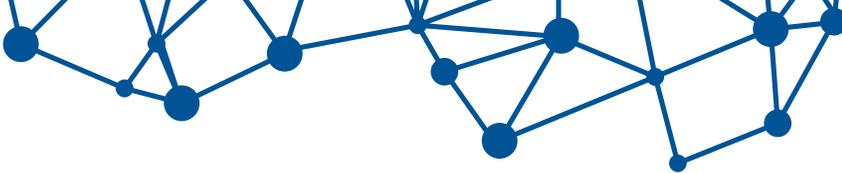
USP



SUMÁRIO

SUMMARY

Apresentação	7
Presentation	13
<i>Rogério Mugnaini, Asa Fujino e Nair Yumiko Kobashi</i>	
Discussões gerais sobre as características mais relevantes de infraestruturas de pesquisa para a cientometria	19
General discussion on the most relevant characteristics of research infrastructures for scientometrics	43
<i>Rodrigo Costas</i>	
Rumo a indicadores para ‘abertura’ de políticas de ciência e tecnologia	67
Towards indicators for ‘opening up’ science and technology policy	79
<i>Ismael Rêfols, Tommaso Ciarli e Andy Stirling</i>	
A pesquisa bibliométrica na era do big data: Desafios e oportunidades	91
Bibliometrics Research in the Era of Big Data: Challenges and Opportunities	101
<i>Dietmar Wolfram</i>	
Avaliação Institucional na USP	111
Institutional Assessment in USP	133
<i>Pedro Vitoriano Oliveira e Vahan Agopyan</i>	
Políticas Públicas em Ciência e Tecnologia no Brasil: desafios e propostas para utilização de indicadores na avaliação	157
Public Policies in Science and Technology in Brazil: challenges and proposals for the use of indicators in evaluation	189
<i>Talita Moreira de Oliveira e Lívio Amaral</i>	



APRESENTAÇÃO

PRESENTATION [*see page 13*] [*go to summary*]

O 5º Encontro Brasileiro de Bibliometria e Cientometria (EBBC) foi realizado de 6 a 8 de julho de 2016, na Universidade de São Paulo, organizado pelo Programa de Pós-Graduação em Ciência da Informação, da Escola de Comunicações e Artes.

O 5º EBBC reuniu especialistas em Bibliometria, Cientometria e áreas afins para discutir as pesquisas em curso e debater as atuais tendências e carências do campo, além de apontar perspectivas futuras para a área, no país.

Os EBBCs são eventos estratégicos para fomentar pesquisas sobre a Bibliometria e a Cientometria. Seus objetivos são: reunir os diferentes grupos de pesquisa que atuam na área; fomentar a interação e a convergência da pesquisa brasileira com o estado da arte internacional; discutir as atuais tendências, necessidades e demandas.

Ganharam relevância as discussões sobre a produção e avaliação de indicadores; o acesso e disponibilização de dados; a importância dos métodos de tratamento e de análise de dados; a qualidade e funcionalidade das ferramentas dos sistemas de informação para armazenamento e recuperação de dados.

Área originalmente circunscrita à Ciência da Informação vem tendo adesão crescente de pesquisadores de diferentes áreas do conhecimento, tanto das ciências humanas e sociais, quanto das ciências exatas e biológicas. Atribui-se tal adesão à percepção da importância dos indicadores bibliométricos e cientométricos para subsidiar políticas de pesquisa e avaliação de Ciência, Tecnologia e Inovação (CT&I), no Brasil e no exterior. Entender e conhecer a diversidade das teorias e aplicações da Bibliometria

e Cientometria são fundamentais para unificar esforços e consolidar a identidade acadêmico-científica desta área no país, ampliar sua massa crítica e conferir maior visibilidade em âmbito internacional.

O evento permitiu estabelecer interações e intercâmbio de ideias entre os pesquisadores brasileiros e estrangeiros. Tais discussões partiram do pressuposto que promover a avaliação da pesquisa científica requer infraestrutura que reúna fontes de dados confiáveis e cultura organizacional que favoreça interações humanas necessárias para analisar métodos, indicadores e resultados obtidos para fundamentar estudos futuros e subsidiar políticas institucionais e governamentais. Neste sentido, um ponto alto do evento foi o diálogo entre as conferências, mesas-redondas, palestras e workshops programados que procuraram mostrar a importância que os estudos bibliométricos e cientométricos assumem para a compreensão mais profunda do papel da ciência e seu planejamento, na contemporaneidade.

Este livro reúne as apresentações dos conferencistas convidados. Os textos, adaptados para apresentação em forma escrita, visam ampliar o acesso às importantes ideias expostas.

A conferência de abertura do 5º EBBC foi proferida por Rodrigo Costas, pesquisador do CWTS (Centre for Science and Technology Studies, Leiden University, Holanda). Seu texto intitula-se “*Discussões gerais sobre as características mais relevantes de infraestruturas de pesquisa para a cientometria*”. É central neste texto, a preocupação com os desafios enfrentados pelo campo da bibliometria/cientometria, alguns deles provocados, paradoxalmente, pela popularização da área. De fato, efeitos preocupantes são a potencial trivialização, os maus usos e mesmo a incompreensão dos indicadores bibliométricos. Nessa medida, Rodrigo chama a atenção para as exigências atuais dos estudos da área. Estas devem contar com: infraestrutura sólida em tamanho e qualidade de recursos humanos, que deve ser multidisciplinar; robustez de recursos tecnológicos de pesquisa; fontes de dados relevantes e consistentemente organizados e classificados tematicamente; domínio de instrumentos e ferramentas específicos de análise de dados. Esse

panorama mostra que as pesquisas bibliométricas e cientométricas voltadas à produção de indicadores são refratárias à improvisação. Envolve custos substanciais; reunião de equipes multidisciplinares, compreensão de que a pesquisa cientométrica não pode estar desconectada dos processos fundamentais de comunicação da ciência.

Ismael Ràfols, do INGENIO (CSIC-UPV), um instituto de política científica na Universitat Politècnica de València, Espanha e também membro do SPRU (Science Policy Research Unit) na University of Sussex, Reino Unido, apresenta o texto *“Rumo a indicadores para ‘abertura’ de políticas de ciência e tecnologia”*, em co-autoria com outros pesquisadores do SPRU – Tommaso Carli e Andy Stirling. O texto trata do uso de indicadores em CT&I, particularmente no que concerne à avaliação da pesquisa. Nessa perspectiva, problematiza o conceito convencional de excelência científica, propondo a necessidade de apresentar perspectivas contrastantes sobre o conceito de excelência. Eles argumentam que pode haver uma necessidade de ampliar os indicadores atuais e fornecer representações mais multidimensionais dos dados. Em resumo, formas mais amplas e plurais de indicadores de CT&I e visualização são necessários para promover políticas de pesquisa mais rigorosas e adequadas para uma variedade de contextos para a ciência prosperar e promover o desenvolvimento humano.

Dietmar Wolfram, da University of Wisconsin-Milwaukee, Estados Unidos, no texto *“A pesquisa bibliométrica na era do big data: desafios e oportunidades”*, discute os benefícios que a disponibilidade dos sistemas denominados *big data* podem proporcionar aos estudos bibliométricos, em face das ferramentas associadas à análise de dados. Pontua, entretanto, os desafios a serem enfrentados. A primeira questão colocada refere-se às barreiras de acesso às fontes de informação, que requerem, via de regra, assinaturas de alto custo. Outro problema refere-se à dispersão e descentralização dessas bases. O autor pergunta-se, ainda, se, de fato, o tamanho descomunal dos *big datas* é efetivamente melhor para produzir dados consistentes. O fato é que o uso adequado desses repositórios exige a realização de limpeza e padronização de dados bastante trabalhosos se o desejo é produzir resultados

confiáveis, significativos para conhecer o estado da arte da CT&I. Conclui que as pesquisas métricas entraram na era dos *big data* tornando possível realizar estudos de grande escala, sendo necessário, no entanto, dispor de melhores métodos que permitam sumarizar e visualizar os dados.

Pedro Vitoriano de Oliveira, em co-autoria com Vahan Agopyan – atual vice-reitor –, ambos da Universidade de São Paulo, apresentam o texto intitulado “*Avaliação Institucional na USP*”. São abordados os mecanismos de Avaliação Institucional da USP, precedidos de antecedentes históricos desse processo nos últimos anos. Foram apresentados os procedimentos adotados para a realização dessa atividade, particularmente daqueles relacionados ao 4º Ciclo da Avaliação Institucional, os caminhos para se chegar aos resultados e alternativas de como colocar em prática as recomendações e críticas que foram apontadas na avaliação. A Avaliação de Instituição de Ensino Superior pode ser entendida como uma ação transformadora, atividade conjunta e contínua, com a finalidade de melhorar a qualidade das várias atividades que são desenvolvidas, como: ensino de graduação e pós-graduação, pesquisa, cultura e extensão, das relações nacionais e internacionais com outras instituições, e da gestão dos recursos humanos e financeiros.

Talita Moreira de Oliveira, da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) apresentou o texto “*Políticas públicas em Ciência e Tecnologia no Brasil: desafios e propostas para utilização de indicadores de avaliação*”, em co-autoria com Livio Amaral. Discutem o processo de avaliação brasileiro e de seus princípios fundamentais em contraposição com o que os manifestos internacionais recomendam – a Declaração de San Francisco sobre Avaliação da Pesquisa (DORA), o Manifesto de Leiden sobre métricas de pesquisa e The Metric Tide (A maré de métricas) –, discutindo as práticas e desafios essenciais para implementar recomendações e indicadores. Na visão dos autores, os indicadores funcionam como instrumentos agregadores e orientadores durante o planejamento e avaliação. No entanto, as métricas não devem ser usadas de forma indiscriminada ou meramente contábil, sem atentar para suas limitações. Desse modo, é central nos processos de avaliação da CAPES a análise por pares, a consulta e

debates constantes com a comunidade científica não apenas para a definição e atualização de critérios, mas também para garantir a transparência ao longo de todo o processo.

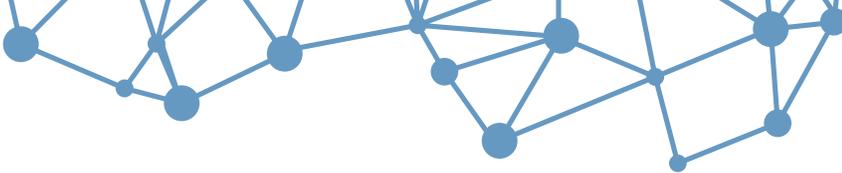
Para finalizar, agradecemos os apoios recebidos da CAPES, bem como do CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), ECA - USP (Escola de Comunicações e Artes da Universidade de São Paulo), FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) e FEA - USP (Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo), que foram fundamentais para a realização do 5º EBBC. À CAPES, especialmente, agradecemos o suporte para publicação deste livro.

São Paulo, 15 de fevereiro de 2017.

*Rogério Mugnaini** – Presidente da Comissão organizadora do 5º EBBC
Asa Fujino – Coordenadora da Comissão executiva
Nair Yumiko Kobashi – Coordenadora da Comissão científica

* mugnaini@usp.br





PRESENTATION

APRESENTAÇÃO [*ver página 7*] [*ir para o sumário*]

The 5th Brazilian Meeting on Bibliometrics and Scientometrics (EBBC) was held at the University of São Paulo, in the period 6th-8th July, 2016 and was organized by the Post-Graduate Program in Information Science, of School of Communication and Arts.

This Meeting brought together specialists in Bibliometrics, Scientometrics and related areas to discuss ongoing research studies and debate current trends and needs in the field, as well as to give future directions to expand the area in the country.

The EBBCs are strategic events that are designed to encourage research into Bibliometrics and Scientometrics. Their aims are as follows: to bring together the different research groups that are involved in the area; to encourage an interaction and alliance of Brazilian research with the state-of-the-art in the international scene; and to discuss current trends, needs and requirements.

Increasing importance has been attached to discussing areas such as creating and assessing indicators, making data available, finding methods for handling and analyzing the data; and improving the quality and functionality of the tools of the information systems needed for the storage and retrieval of data.

An area that was originally confined to Information Science is increasingly being extended to support researchers from different fields of knowledge, both in the human and social sciences and the exact and biological sciences. This support can be attributed to a growing awareness of the

importance of bibliometric and scientometric indicators to assist in policy-making with regard to research and the assessment of Science, Technology and Innovation (ST&I), in Brazil and the outside world. It is essential to understand and know the wide range of theories and application of Bibliometrics and Scientometrics to strengthen and consolidate the academic/scientific identity of this area in the country, as well as to broaden its scope and give it greater prominence on the international stage.

The event made it possible to establish interactions and carry out an exchange of ideas between Brazilian and foreign researchers. These discussions are based on the assumption that making an evaluation of scientific research requires an environment where reliable data sources can be combined. It also allows human reactions to occur which are necessary for the analysis of the methods, indicators and results obtained for future studies and support institutional and governmental policies. This meant that a key feature of the event was the dialogue between the participants of conferences, round-table discussions, lectures and planned workshops which sought to show the importance that is attached to bibliometrics and scientometrics as a means of obtaining a more in-depth understanding of the role of science and its planning in the contemporary world.

This book is a compilation of the presentations of the guest lecturers. The purpose of the papers, which have been adapted for presentation in a written form, is to broaden access to the principal ideas.

The opening conference of the 5th EBBC was delivered by Rodrigo Costas, a researcher from CWTS (Centre for Science and Technology Studies, Leiden University, Holland). The title of his paper is: “*A general discussion on the most relevant characteristics of research infrastructures for scientometrics*”. The central concern of the presentation was with the challenges faced by the field of bibliometrics/scientometrics, some of them caused, paradoxically by the way this area has become popularized. What in fact is disturbing is the potential trivialization, misuse and even misunderstanding of the bibliometric indicators. In the light of this, Rodrigo draws attention to the

current requirements of studies in the area. These should include the following: a sound infrastructure which in terms of size and the quality of its human resources, must be multidisciplinary; robust technological resources for research; key data sources that are collated and classified in an orderly manner; and a mastery of the instruments and tools needed for the data analysis. This panorama shows that the bibliometric and scientometric research studies designed for the creation of indicators are unresponsive to improvisation. The procedure involves substantial costs, and brings together multidisciplinary teams with an understanding that scientometric research cannot be separated from the fundamental processes of communication in science.

Ismael Ràfols, from INGENIO (CSIC-UPV), a science policy institute at the Universitat Politècnica de València, Spain and also a member of SPRU (Science Policy Research Unit) at the University of Sussex, United Kingdom, gave a presentation with the title: *“Toward indicators for opening up science and technology policies”*, in co-authorship with three other researchers from SPRU –Tommaso Carli and Andy Stirling. The paper focuses on the use of indicators in ST&I, in particular regarding research evaluation. From this standpoint, it questions the conventional concept of scientific excellence and stresses the need to show contrasting perspectives of the concept of excellence. They argue that there may be a need to broaden current indicators and provide more multidimensional representations of the data. In summary, broader and more pluralistic forms of ST&I indicators and visualization are necessary to foster research policies more rigorous and appropriate in a variety of contexts for science to flourish and foster human development.

Dietmar Wolfram, of the University of Wisconsin-Milwaukee, United States, gave a talk with the title: *“Bibliometrics research in the era of Big Data: challenges and opportunities”*, in which he discussed the benefits that the availability of what are called *big data* systems, can provide to bibliometric studies, in the light of the tools required for data analysis. However, he underlined the challenges that had to be faced. The first question raised, was regarding

the barriers blocking access to information sources, which, as a general rule, require high-cost subscribers. Another problem concerned the dispersal and decentralization of these bases. The author wonders whether in reality the immense size of the *big data* means they are effectively better for producing consistent data. The fact is that a suitable use of these repositories requires the arduous task of ensuring data cleanliness and standardization, if one wishes to achieve results that are reliable and of enough significance to know the state-of-the-art of the ST&I. It can be concluded that metric research is entering an era of *big data* systems which are making it possible to undertake studies on a large scale, although better methods need to be employed that can allow the data to be summarized and visualized.

Pedro Vitoriano de Oliveira, in co-authorship with Vahan Agopyan – the current Vice-Rector –, both from the Universidade de São Paulo, presented a paper entitled “*Institutional Assessment in the USP*”. It sets out the mechanisms for the Institutional Assessment in USP, preceded by a historical background of this process in recent years. It particularly focuses on the paths, results and alternative methods of putting into practice the recommendations, and taking note of the criticisms, made at the 4th Cycle of Institutional Assessment. The assessment of the Institution of Higher Education can be regarded as a transformative, combined and continuous activity, which is aimed at achieving an improvement in the standards of the various activities that are being carried out. These include the following: teaching – at the level of graduate and post-graduate studies –, research, culture and extension, national and international relations with other analogous institutions and the management of financial and human resources.

Talita Moreira de Oliveira, from the Agency for Support and Evaluation of Graduate Education (CAPES) presented the paper “*Public Policies in Science and Technology in Brazil: challenges and proposals for the use of indicators in evaluation*”, in co-authorship with Livio Amaral. This includes a discussion of the Brazilian evaluation process and its fundamental principles compared to what international manifestos recommend – the San Francisco DORA [Declaration on Research Assessment], the Leiden Manifesto for Research

Metrics and the Metric Tide, discussing the essential practices and challenges for implementing the recommendations and indicators. In the view of the authors, indicators act as auxiliary instruments that provide support for managers and evaluators during planning and assessment. However, metrics should not be used in an indiscriminate or merely numerical manner without attention being paid to their limitations. For this reason, peer review is a fundamental principle for CAPES assessment, together with constant debates with the scientific community, not only for the definition of criteria but also to ensure transparency throughout the whole process.

To conclude, we would like to express out thanks for the support given by CAPES, as well as CNPq (National Council for Scientific and Technological Development), ECA - USP (School of Communications and Arts, University of São Paulo), FAPESP (São Paulo Research Foundation) and FEA - USP (Faculty of Economics, Administration and Accounting, University of São Paulo), that was of fundamental importance for the 5th EBBC undertaking. In addition, we specially acknowledge CAPES for supporting the publication of this book.

São Paulo, February 15, 2017.

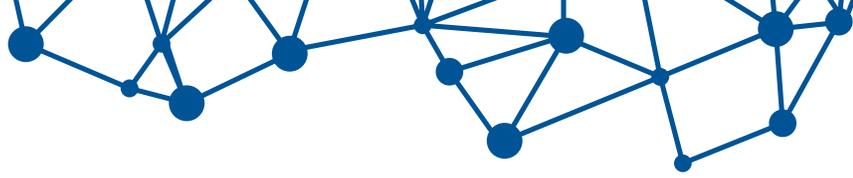
*Rogério Mugnaini** – President of the Organizing Committee of the 5th EBBC

Asa Fujino – Supervisor of the Executive Committee

Nair Yumiko Kobashi – Supervisor of the Scientific Committee

* mugnaini@usp.br





DISCUSSÕES GERAIS SOBRE AS CARACTERÍSTICAS MAIS RELEVANTES DE INFRAESTRUTURAS DE PESQUISA PARA A CIENTOMETRIA

GENERAL DISCUSSION ON THE MOST RELEVANT CHARACTERISTICS OF RESEARCH
INFRASTRUCTURES FOR SCIENTOMETRICS [*see page 43*] [*go to summary*]

*Rodrigo Costas**

INTRODUÇÃO

O uso da cientometria passou por uma importante expansão nos últimos 20 anos. Esse crescente interesse, de alguma forma, ecoa o interesse vindo de muitos diferentes setores, atores e partes interessadas do sistema científico (LEYDESDORFF; WOUTERS; BORNMANN, 2016). Essa expansão também veio acompanhada do desenvolvimento e popularização de programas de computador dedicados à bibliometria¹, assim como o surgimento de fontes de dados novas e mais diversas, com múltiplas possibilidades bibliométricas (por exemplo, Google Scholar, Microsoft Academic

* Centre for Science and Technology Studies (CWTS), Leiden University, Leiden, the Netherlands; rcostas@cwts.leidenuniv.nl

1 Entre esses, podemos mencionar o VOSviewer (<http://www.vosviewer.com>), CitNetExplorer (<http://www.citnetexplorer.nl>), R-packages (<http://www.bibliometrix.org>), Bibexcel (<http://homepage.univie.ac.at/juan.gorraiz/bibexcel/>), CRExplorer (<http://www.crexplorer.net/>), PoP (<http://www.harzing.com/resources/publish-or-perish>), SCI2 (<https://sciz.cns.iu.edu/user/index.php>), etc.

ou Mendeleev, para citar alguns). Além disso, o aumento da capacidade de computação e armazenamento, e o acesso rápido a dados bibliográficos fez com que mais indivíduos tenham acesso a indicadores, ferramentas e aplicações, com base na bibliometria.

Esse cenário está claramente vinculado à ideia, proposta já há 20 anos por (KATZ; HICKS, 1997), da cientometria de computador pessoal, que trazia a promessa da possibilidade de se executar tarefas bibliométricas essenciais (por exemplo, manipulação de dados, unificação, análise ou visualização de informações bibliométricas) a partir de um simples computador, com recursos computacionais comuns. Obviamente a cientometria de computador pessoal pode oferecer muitas vantagens, especialmente um melhor acesso a ferramentas bibliométricas e a popularização e maior interesse pela pesquisa cientométrica. Entretanto, essa popularização também pode trazer novos desafios. Um importante desafio é o que é conhecido como “bibliometria amadora” (RUSHFORTH; DE RIJCKE, 2015) ou abordagens bibliométricas “rápidas e sujas” (BAR-ILAN, 2008), que podem ter o potencial efeito de trivializar e levar ao mau uso e mau entendimento de indicadores bibliométricos.

Para se ter uma melhor ideia do valor da bibliometria mais “profissionalizada” ou avançada (DE RIJCKE; RUSHFORTH, 2015), este capítulo pretende discutir a relevância e as características mais importantes de infraestruturas avançadas para o trabalho bibliométrico hoje existente. O objetivo é oferecer uma reflexão geral dos tipos mais importantes de infraestruturas relevantes à pesquisa cientométrica. Assim, esta descrição poderá ser útil a grupos de pesquisa novos ou emergentes que tenham interesse em desenvolver suas próprias infraestruturas para a pesquisa bibliométrica, assim como a outros usuários de informações bibliométricas (como formuladores de políticas, gestores de pesquisa, etc.) que queiram saber mais sobre as características desse tipo de infraestrutura.

GRUPOS DE PESQUISA CIENTOMÉTRICA EM TODO O MUNDO

Nesta seção, apresentaremos uma breve descrição de alguns dos principais grupos internacionais de pesquisa cientométrica. Esta lista² não tem a intenção de ser completa, mas servirá apenas para dar exemplos de grupos de pesquisa que, até certo ponto, compartilham das características, metodologias e infraestruturas que serão discutidas mais adiante no capítulo.

CWTS – Centro de Estudos em Ciência e Tecnologia, Universidade de Leiden, Holanda

<http://www.cwts.nl>

O Centro de Estudos em Ciência e Tecnologia, Universidade de Leiden, Holanda (CWTS – Centrum voor Wetenschap en Technologie Studies) é um dos maiores e mais bem estabelecidos centros internacionais dedicados à pesquisa cientométrica. O quadro funcional do CWTS agrega indivíduos com experiência em um espectro diverso de disciplinas. Através de seus programas de pesquisa, o escopo de pesquisa do CWTS expandiu a partir de uma pesquisa cientométrica mais tradicional para novos tópicos, como a altmetria e o impacto social da ciência, assim como para abordagens mais qualitativas e de métodos combinados. Um importante *output* do grupo com base em pesquisa é o Leiden Ranking (<http://www.leidenranking.com/>), que é publicado todo ano e apresenta indicadores bibliométricos avançados para mais de 800 universidades em todo o mundo.

-
- 2 A lista tem, claro, viés direcionado ao conhecimento e experiência próprios do autor deste texto, e os grupos mencionados serão apresentados apenas como exemplos de detentores de estruturas cientométricas relativamente grandes. É claro que existem outros grupos, acadêmicos e especialistas que também produzem um excelente trabalho e compartilham de muitas das características e capacidades infraestruturais aqui discutidas.

DZHW – Centro Alemão para a Pesquisa e Estudos Científicos do Ensino Superior (em Hannover, Berlim e Leipzig), Alemanha

<http://www.dzhw.eu/en>

O Centro Alemão para a Pesquisa e Estudos Científicos do Ensino Superior – DZHW (que hoje incorpora o antigo Instituto para a Informação e Garantia de Qualidade da Pesquisa – IFQ) é um instituto que enfoca a prática da pesquisa empírica orientada para a aplicação. Não se trata de um instituto dedicado exclusivamente à pesquisa cientométrica, mas também a variados tópicos relacionados à pesquisa de nível superior. Desde 1º de janeiro de 2016, ele integra, no Departamento 2 “Sistema de Pesquisa e Dinâmica da Ciência”, as linhas de pesquisa e atividades do antigo IFQ, o que inclui a pesquisa bibliométrica, carreiras e avaliações de pesquisas, entre outros assuntos.

Expertisecentrum Onderzoek en Ontwikkelingsmonitoring (ECOOM), Leuven, Bélgica

https://www.ecoom.be/nl/member_details/ku%20leuven

O Centro para o Monitoramento da Pesquisa e Desenvolvimento (Expertisecentrum Onderzoek en Ontwikkelingsmonitoring, ECOOM) é um consórcio interuniversitário que conta com a participação de todas as universidades flamengas (Katholieke Universiteit Leuven, Universiteit Gent, Vrije Universiteit Brussels, Universiteit Antwerpen e Universiteit Hasselt). Sua missão é desenvolver um sistema consistente de indicadores de pesquisa e desenvolvimento e inovação (PD&I) para o governo flamengo, capaz de mapear e monitorar os empenhos em PD&I da região flamenga. Ele combina a experiência científica de uma ampla diversidade de origens, como do campo da cientometria, análise de patentes (tecnometria), avaliações de pesquisa e desenvolvimento e de inovação, entre outros.

Grupo de Pesquisa SCImago, (Granada, Madrid), Espanha

<https://www.facebook.com/SRG-SCImago-Research-Group-189628671202/>³

O Grupo de Pesquisa SCImago dedica-se ao sistema de comunicação acadêmica e ao desenvolvimento de ferramentas para analisar, visualizar e interpretar os dados extraídos de bases de dados de informações científicas. Com uma forte concentração nos dados da Elsevier, o grupo desenvolveu produtos de avaliação como o Scimago Journal & Country rank (<http://www.scimagojr.com/>) e o Scimago Institutions Rankings (<http://www.scimagoir.com/>).

Statistical Cybermetrics Research Group, Universidade de Wolverhampton, Reino Unido

<http://cybermetrics.wlv.ac.uk/>

Esse grupo de pesquisa foca principalmente nas áreas específicas da webometria e da altmetria dentro do universo cientométrico, apesar de que seu trabalho também inclui as pesquisas bibliométricas mais tradicionais. O grupo particularmente visa desenvolver programas de computador e métodos para explorar fontes baseadas na Internet para a pesquisa em webometria, altmetria e cientometria.

Canada Research Chair on the Transformations of Scholarly Communication (CRCTSC), Universidade de Montreal, Montreal, Canadá

<http://crc.ebsi.umontreal.ca/en/>

-
- 3 No momento, o grupo não conta com um site oficial. Desta forma, é difícil avaliar as principais características do grupo, apesar de que a partir de suas publicações e metodologias é possível observar que a maior parte das características de infraestruturas cientométricas expostas abaixo também se aplica ao grupo.

Este é um dos mais novos grupos de pesquisa cientométrica, fora da Europa, que atingiu uma importante capacidade de pesquisa nos últimos anos. Seu principal objetivo é aumentar o entendimento de como o conhecimento está atualmente sendo disseminado através de publicações científicas, e especialmente como os novos tipos de produção de conhecimento, fontes de dados e políticas estão modificando a forma com que os acadêmicos desenvolvem seu trabalho. O grupo consiste de uma importante variedade de tipos de especialidades, que vão desde a bibliometria, altmetria e linguística até as abordagens qualitativas e perspectivas de gênero, entre outros.

*Centre for Research on Evaluation, Science and Technology (CREST),
Universidade de Stellenbosch, Stellenbosch, África do Sul*

<http://www.sun.ac.za/crest/>

Estabelecido em 1995, trata-se de outro centro de pesquisa fora da Europa que cobre os campos da bibliometria, pesquisa em conhecimento do nível superior, recursos humanos em ciência e tecnologia ou avaliação de pesquisa e avaliação de impacto, entre outros tópicos relacionados aos padrões de comunicação entre acadêmicos. Desde 2014, o CREST também é lar para o DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP – <http://www.sun.ac.za/scistip/>), que compreende um amplo espectro de diversos tópicos de pesquisa como a ciência na África, recursos humanos, comunicação científica e indicadores de informação em ciência e tecnologia, entre outros. Assim, o centro também articula uma grande variedade de especialidades, inclusive pesquisas mais quantitativas e voltadas a dados, além de abordagens mais qualitativas.

Como já foi mencionado, essa lista de grupos de pesquisa não é exaustiva. Existem outros grupos, tanto já estabelecidos quanto emergentes, em muitos outros países (como Espanha, Dinamarca, Suécia, Noruega, Finlândia, Estados Unidos, Itália, Áustria, etc.) que compartilham de muitas das características discutidas neste capítulo.

AS PRINCIPAIS CARACTERÍSTICAS DE INFRAESTRUTURAS PARA A PESQUISA CIENTOMÉTRICA

Todos os diferentes grupos de pesquisa acima mencionados possuem suas próprias peculiaridades, focos específicos, missões de pesquisa e objetivos. No entanto, as próximas seções enfocam suas similaridades e características mais importantes, em termos de aspectos infraestruturais relativos às suas pesquisas cientométricas. Desta forma, apontaremos aspectos relacionados aos seus tamanhos e recursos humanos, infraestruturas técnicas e bases de dados, e também em outros desenvolvimentos técnicos necessários para um trabalho mais avançado de pesquisa cientométrica.

1) Recursos humanos

Em geral, a maioria desses grupos dependem de recursos humanos relativamente grandes, com quadros de entre 5 e 10 pessoas, e às vezes mais de 30, consistindo não de apenas pessoal científico, mas também profissionais de gestão, de tecnologia da informação e da comunicação, e de administração. Considerando os recursos humanos científicos, um aspecto importante que caracteriza a maioria desses grupos é a composição multidisciplinar de seus membros. Assim, não é raro encontrar, entre os principais grupos de pesquisas, indivíduos com diferentes especialidades, como cientistas da informação, cientistas da computação, estatísticos, cientistas políticos, especialistas em avaliação de pesquisa e gestão de pesquisa, linguistas, etc.; na maioria dos casos, também combinando perspectivas quantitativas e qualitativas.

Um elemento importante para a sustentabilidade desses grupos é a existência de modelos específicos de negócios que ajudam a manter seus recursos humanos e infraestruturas tecnológicas. Esses modelos de negócios podem variar entre a produção de serviços comerciais e consultorias, e também incluir o desenvolvimento de projetos cientométricos específicos, variando de projetos mais internacionais a projetos mais locais, institucionais ou regionais. A maior parte desses grupos, como a maioria dos pesquisadores

acadêmicos de hoje, dependem muito da aquisição de financiamento externo para sua sustentabilidade.

2) Infraestruturas técnicas

O desenvolvimento de uma infraestrutura técnica significativa para a pesquisa cientométrica é um dos pontos mais importantes para a maioria desses grupos. Desta forma, a disponibilidade de servidores para armazenar os dados e a tecnologia para buscar e acessar os dados (por exemplo, bases de dados SQL, SAS, Access, etc.) são elementos centrais para o trabalho dos grupos⁴. De forma semelhante, garantir uma capacidade suficiente de computação e a otimização dos diferentes tipos de análise é essencial para ganhar eficiência e agilidade na produção de indicadores e estudos.

2.1) Bases de dados

É provável que o principal recurso para se estabelecer um grupo de pesquisa cientométrica seja a disponibilidade de fontes relevantes de dados que permitem o desenvolvimento do trabalho de pesquisa quantitativo. Algumas das mais importantes fontes de dados incluem bases de dados bibliográficos internacionais, que também indexam as relações de citações entre as publicações. Entre as fontes de dados mais importantes para a pesquisa cientométrica, podemos mencionar as seguintes:

1. *Web of Science– Clarivate Analytics*

<http://webofscience.com>

4 Normalmente, os grupos de pesquisa também têm acesso a diferentes bases de dados através de suas interfaces de rede *online*. Porém, esse tipo de acesso tende a ser problemático porque depende das decisões e escolhas metodológicas por parte dos produtores das bases de dados. Os interesses de pesquisa dos diferentes grupos normalmente precisam de acesso a dados e opções de gestão mais avançados. Por isso, muitos grupos optam por possuir versões *in-house* das bases de dados bibliométricas utilizadas em seus trabalhos de pesquisa.

Essa é a base de dados mais utilizada em estudos cientométricos. A fundação da atual Web of Science foi realizada pelo Institute for Scientific Information, fundado por Eugene Garfield na década de 1960. A maior parte dos grupos de pesquisa cientométrica tem acesso (ou possuem versões *in-house*) da Web of Science Core Collection. As principais bases de dados contidas na Core Collection são: Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI) e Arts & Humanities Citation Index (A&HCI), e mais recentemente o Emerging Sources Citation Index (ESCI). Outras bases de dados também ocasionalmente utilizadas por diferentes grupos são o Conference Proceedings Citation Index (CPCI), Book Citation Index (BCI), Data Citation Index (DCI), entre outras. No entanto, a disponibilidade dessas outras fontes varia enormemente entre os grupos de pesquisa, e sua disponibilidade normalmente corresponde a interesses específicos ou ao foco de pesquisa de cada grupo.

A Web of Science Core Collection compreende mais de 12.000 periódicos internacionais, considerados os mais populares periódicos científicos internacionais. As referências das publicações fonte são indexadas, as relações entre documentos citados e citantes são estabelecidas com base em algoritmos internos determinados pela própria Web of Science ou pelos diferentes grupos de pesquisa (OLENSKY; SCHMIDT; VAN ECK, 2016). Uma característica importante dessa base de dados é a grande diversidade de metadados indexados, que incluem os autores das publicações, suas afiliações e países, informações bibliográficas básicas (isto é, título, ano de publicação, nome do periódico, volume, edição, páginas, identificadores do documento, etc) e, mais recentemente, outros elementos paratextuais de informação, como agradecimentos a financiamentos (PAUL-HUS; DESROCHERS; COSTAS, 2016). Essa rica disponibilidade de metadados permite múltiplas possibilidades de abordagens analíticas, que vão desde a análise de colaborações até o estudo textual e linguístico dos títulos e resumos, entre outras.

Sem nos aprofundarmos nas vantagens e limitações da base de dados, podemos dizer que entre as vantagens está a longa trajetória dessa base de dados, sendo uma das mais bem estudadas fontes de dados para a pesquisa

cientométrica. Outra vantagem é o caráter seletivo, que pretende focar os mais importantes periódicos científicos internacionais. Porém, é precisamente essa seletividade que é normalmente apontada como sendo sua mais importante limitação, especificamente pela má cobertura de alguns periódicos locais e de pesquisa em ciências sociais e humanidades (NEUHAUS; DANIEL, 2008).

2. *Scopus – Elsevier*

<https://www.scopus.com/>

Outra das principais fontes de dados bibliométricos é a Scopus. Essa base de dados compreende mais de 20.000 periódicos científicos nacionais e internacionais. Assim como a Web of Science, a Scopus também indexa as referências citadas e as ligações citados/citantes entre as publicações contidas na base de dados. Semelhante à Web of Science, ela também apresenta uma ampla cobertura de uma diversidade de metadados, incluindo autores, afiliações, dados bibliográficos, etc.

Entre as vantagens da Scopus, podemos mencionar a ampla cobertura de periódicos, países e idiomas. Quanto às limitações, podemos mencionar a qualidade questionável de alguns dos metadados nela contidos (FRANCESCHINI; MAISANO; MASTROGIACOMO, 2016), assim como a cobertura temporal mais curta de publicações (a cobertura da base de dados começa em 1996).

3. *Google Scholar*

O Google Scholar é provavelmente a mais popular ferramenta de busca de publicações científicas hoje disponível na Internet. Ela oferece fortes vantagens, como a disponibilidade gratuita (a Scopus e a Web of Science são produtos comerciais). A cobertura do Google Scholar também já foi destacada como a maior de todas as bases de dados existentes (DELGADO; REPISO, 2013), e compreende não apenas artigos científicos, mas também

pré-impressões, livros, capítulos de livros, relatórios, literatura cinzenta, etc., apesar de não estar claro qual é a real cobertura do Google Scholar (ORDUÑA-MALEA et al., 2014). Uma característica importante do Google Scholar relativa à capacidade para a pesquisa cientométrica é que, como a Scopus e a Web of Science, ele também indexa as citações recebidas pelos registros contidos na base de dados.

Apesar dessas importantes vantagens, o Google Scholar já foi criticado por muitos motivos. Nisso, sua natureza de “caixa preta” já foi apontada (WOUTERS; COSTAS, 2012). A falta de informação sobre sua cobertura, a baixa qualidade dos metadados e a falta de informação sobre seu procedimento de correspondência de citações são alguns dos elementos também criticados no Google Scholar (TORRES-SALINAS; RUIZ-PÉREZ; LÓPEZ-CÓZAR, 2009). Além do mais, a falta de uma maneira apropriadamente sistemática para se acessar e coletar dados do Google Scholar preveniu a possibilidade de limpeza e aprimoramento dos dados brutos originais pelos diferentes grupos de pesquisa. Ainda, o Google Scholar não possui a riqueza de metadados oferecida pelas outras bases de dados comerciais; por exemplo, informações cruciais como as afiliações e os países dos autores não constam nos registros.

Frente a essa situação, poucos grupos de pesquisa incorporaram o Google Scholar como sua principal fonte de dados para a pesquisa cientométrica, e a maioria dos estudos baseados nessa fonte concentram-se em conjuntos de dados relativamente pequenos, que ainda são trabalhosos de se obter (PRINS et al., 2016). Talvez o grupo que mais tenha trabalhado com o Google Scholar tenha sido o EC3 em Granada (Espanha) (MARTÍN-MARTÍN et al., 2016), em conjunto com Ann Harzing (HARZING; ALAKAN-GAS, 2016). Entretanto, seus resultados, apesar de promissores, ainda não respaldam uma incorporação apropriada do Google Scholar como uma fonte de grande escala confiável para a pesquisa cientométrica.

4. *Microsoft Academic*

Semelhante ao Google Scholar, a Microsoft Academic também é uma ferramenta de busca de publicações científicas. Diferente do Google Scholar, a Microsoft Academic oferece a possibilidade de busca dentro de uma dada instituição ou entidade (universidades, por exemplo). Estudos preliminares já destacaram a relevância e maior cobertura dessa fonte em comparação às fontes comerciais (SINHA et al., 2015; HARZING, 2016), estimando que a base de dados melhorou significativamente comparada com suas versões anteriores (HARZING; ALAKANGAS, 2016). A Microsoft Academic também oferece alguns cálculos de contagens de citações, apesar do cálculo ser uma estimativa do número de citações recebidas por cada publicação oferecida pela Microsoft Academic⁵.

Quanto às limitações da base de dados, a Microsoft Academic compartilha de muitas das limitações do Google Scholar. Entre essas limitações, podemos mencionar a falta de informações sobre sua real cobertura (apesar de parecer menor do que a do Google Scholar – HARZING, 2016) e a qualidade e exatidão dos metadados fornecidos.

Até onde sabemos, não existem grupos de pesquisa cientométrica que estejam ativamente utilizando pesquisas na Microsoft Academic como fonte primária de dados bibliométricos. Apesar da Microsoft Academic ter sido identificada como uma ferramenta promissora para a pesquisa bibliométrica (WOUTERS; COSTAS, 2012) e apesar de algumas ferramentas, como a Publish or Perish (PoP), também utilizarem essa fonte para cálculos, até hoje não há muitos grupos de pesquisa realizando pesquisas cientométricas baseadas nessa fonte⁶.

5 <https://microsoftacademic.uservoice.com/knowledgebase/articles/838965-microsoft-academic-faq#citations>

6 Um artigo recente por Hug, Ochsner e Brändle (2017) também demonstra o forte potencial da Microsoft Academic como uma fonte de dados para a pesquisa cientométrica.

5. Bases de dados locais e temáticas

Além das bases de dados internacionais, a maioria dos grupos de pesquisa também utiliza fontes de dados de interesse local ou temático. Exemplo disso é que, entre os grupos de pesquisa cientométrica espanhóis, trabalhasse com bases de dados que compreendem a produção relevante espanhola, sendo algumas delas o ICYT (Index of Science and Technology), ISOC (Index of Social Sciences) e IME (Spanish Medical Index)⁷. No caso dos grupos de pesquisa concentradas na produção latino-americana, temos a base de dados SciELO⁸ (que também contém informações de citações), e para os franco-canadenses, é relevante uma base de dados como a Érudit⁹. Em relação à disponibilidade das bases de dados temáticas, a MEDLINE e a PubMed são também bases de dados comuns utilizadas por diferentes grupos de pesquisa cientométrica.

6. Bases de dados altmétricas e de ciência aberta

Nos últimos anos, o surgimento do movimento altmétrico (PRIEM et al., 2010) ofereceu a possibilidade para os diferentes grupos de pesquisa cientométrica estudarem a recepção da ciência em outras fontes. Hoje em dia, é possível estudar como estão sendo mencionadas as publicações científicas em fontes como o Twitter, Facebook, blogs, F1000, Mendeley, etc., abrindo novos caminhos para se analisar a recepção de publicações científicas nas mídias sociais.

Entre essas bases de dados, as mais comuns são a Altmetric.com (<https://www.altmetric.com/>) e a Mendeley.com. A Altmetric.com coletou a recepção em mídias sociais de mais de 5 milhões de publicações de todo o mundo. Essa base de dados oferece estatísticas de quantas vezes uma publicação foi mencionada no Twitter, blogs, Facebook e outras plataformas de mídias sociais. O valor e relevância de todas essas novas fontes de mídias sociais

7 <http://bddoc.csic.es:8080/>

8 <http://www.scielo.org/>

9 <http://www.erudit.org/>

ainda estão sendo debatidos dentro da comunidade científica (WOUTERS; COSTAS, 2012; ROBINSON-GARCÍA et al., 2014).

O Mendeley.com é um gestor de referências online. Os usuários do Mendeley podem salvar publicações e administrar suas bibliotecas individuais para diferentes propósitos (por exemplo, trabalho em pesquisa, estudos, colaborações, etc.). A maior vantagem do Mendeley como fonte altmétrica é que ele informa o número de usuários que salvaram uma dada publicação em suas bibliotecas pessoais, o que é também conhecido como “readership” (público leitor). Além disso, é também possível desagregar a contagem do “readership” por tipologias gerais dos usuários (por exemplo, PhD, professores, alunos, etc. – cf. (ZAHEDI; VAN ECK, 2014)). O Mendeley oferece suas metrias gratuitamente através de sua API (Interface de Programação de Aplicação) aberta (<http://dev.mendeley.com/>), portanto sua fonte pode ser utilizada livremente para se obter metrias do público leitor para praticamente qualquer publicação existente. Porém, sua maior limitação é que a base de dados do Mendeley como um todo não pode ser obtida diretamente, portanto é sempre necessário trabalhar com uma base de dados de fonte original (como a Web of Science, Scopus, MEDLINE, etc.) para realizar buscas na API do Mendeley. Além do mais, é importante observar que os metadados oferecidos pelo Mendeley não são de alta qualidade (ZAHEDI; HAUSTEIN; BOWMAN, 2014), e muitas informações importantes (como afiliações ou países dos autores) não são incluídas nos dados fornecidos pelo Mendeley.

Da perspectiva do estudo de movimentos como o Open Access ou o Open Data, bases de dados como a OpenAIRE (<https://www.openaire.eu/>) ou a DataCite (<https://www.datacite.org/>) são fontes importantes que os grupos de pesquisa cientométrica estão começando a estudar. Essas fontes de dados estão disponíveis através de APIs públicas (<http://api.openaire.eu/>; <https://mds.datacite.org/static/apidoc>), o que permite que qualquer usuário interessado possa baixar e analisar os dados.

Um importante elemento conceitual que precisa ser levado em conta quando se trabalha com essas novas fontes altmétricas e de dados/ciência

aberta é que elas ainda são novas e mais pesquisa se faz necessária para se entender seu valor e relevância para a pesquisa cientométrica (HAUSTEIN; BOWMAN; COSTAS, 2016). Uma limitação prática ainda mais importante dessas fontes de dados é que elas são bastante dependentes de identificadores como DOIs (Digital Object Identifier – identificador digital de objetos) ou identidades PubMed, e ainda assim a qualidade de seus metadados é altamente problemática (ZAHEDI; HAUSTEIN; BOWMAN, 2014).

É importante destacar a crescente disponibilidade de bases de dados de textos integrais, como os periódicos do PLOS ou o PubMed Central. Além disso, algumas grandes editoras comerciais, como a Elsevier, a Springer e a Wiley, também disponibilizam (sob certas condições) as bases de dados de textos integrais de suas bases de dados. Essas possibilidades permitem o desenvolvimento de mais tipos de análises de mineração de texto, inseridas no contexto geral das perspectivas do *big data* que estão atualmente ganhando importância na pesquisa científica (EKBIA et al., 2015).

2.2) Outros desenvolvimentos tecnológicos

Além da disponibilidade de bases de dados e fontes de dados específicas, o trabalho cientométrico normalmente também requer ferramentas e instrumentos adicionais, necessários para a análise adequada de dados e indicadores derivados. Algumas dessas infraestruturas necessárias estão descritas abaixo:

1. *Limpeza de dados e desenvolvimento de tesouros e algoritmos específicos*

Muito frequentemente, os dados fornecidos pelas diferentes fontes de dados apresentam severos problemas de padronização e qualidade de dados (HOOD; WILSON, 2003). Portanto, por exemplo, os nomes das afiliações de instituições na Web of Science ou na Scopus normalmente requerem limpeza e harmonização extensivas (WALTMAN et al., 2012). De forma similar, outras informações, como dados de agradecimentos a financiamentos, também requerem limpeza e harmonização. A solução mais comum é o

desenvolvimento de tesouros especializados de organizações e/ou financiadores, nos quais as variantes dos nomes de diferentes unidades são homogeneizadas sob forma canônica, assim como as relações entre elas (por exemplo, de dependência ou equivalência) são estabelecidas (FERNÁNDEZ et al., 1993; SIRTES, 2013).

Outro elemento importante é a harmonização dos diferentes elementos bibliográficos (como títulos, ano de publicação, nomes de periódicos, etc.). Isso é menos crítico para bases de dados comerciais mais estabelecidas (WoS or Scopus), já que seus metadados são normalmente bem padronizados, mas é mais relevante para fontes de dados que agregam dados de diferentes fontes (como a DataCite ou OpenAIRE) ou dados da Internet (como o Google Scholar ou Microsoft Academic).

Um caso especial que requer atenção quando se trabalha com informações cientométricas é a importância da desambiguação de nomes de autores (SMALHEISER; TORVIK, 2009). Esse aspecto é crítico em trabalho que envolva informações em nível de indivíduo-pesquisador, e a falta do conhecimento de problemas como homonímia ou sinonímia pode distorcer os resultados das análises de forma significativa. Portanto, os grupos cientométricos tendem a investir importantes recursos na desambiguação de nomes de atores em suas bases de dados de forma manual, automatizada ou semi-automatizada (D'ANGELO; GIUFFRIDA; ABRAMO, 2011; CARON; VAN ECK, 2014).

2. Classificações temáticas e esquemas disciplinares

Outro elemento central da pesquisa cientométrica é a disponibilidade de instrumentos que permitam a possibilidade da análise temática. Assim, a disponibilidade de classificações temáticas torna-se relevante no desenvolvimento da pesquisa no trabalho cientométrico (WALTMAN; VAN ECK, 2012).

As Web of Science Subject Categories são classificações bastante comuns utilizadas na pesquisa cientométrica. Essa classificação é composta de cerca de 250 áreas de pesquisa. É uma classificação muito conveniente, como já é

incorporada na base de dados da Web of Science, assim tornando o uso da classificação muito prático. Uma classificação semelhante é também apresentada pela base de dados da Scopus da Elsevier. Porém, esse tipo de classificação não está livre de limitações. Uma das mais importantes limitações é que elas são baseadas na classificação de periódicos científicos, em vez de publicações científicas individuais, exibindo diferentes graus de exatidão (cf., WANG; WALTMAN, 2016). Desta forma, todas as publicações de um dado periódico são classificadas na mesma área de pesquisa que a do periódico, em vez de receber sua própria classificação individual. Isso é especialmente problemático para grandes periódicos multidisciplinares, como o *Nature*, *Science*, *PNAS*, e, especialmente, o *PLOS ONE* e *Scientific Reports*.

Alternativamente, outras classificações são também utilizadas por diferentes grupos de pesquisa, muitas vezes relacionadas ao seu respectivo ambiente local ou outros interesses. Portanto, exemplos de classificações incluem a da National Science Foundation (NSF); a classificação da OECD; classificações mais específicas a um país, como a NOWT na Holanda; e outras classificações multifuncionais, como a da ScienceMetrix Classification (<http://www.science-metrix.com/en/classification>). De forma parecida, classificações disciplinares são ocasionalmente utilizadas. O Medical Subject Headings (MeSH) é um bom exemplo de uma classificação disciplinar relevante.

Mais recentemente, novas classificações foram desenvolvidas com base nas relações de citações entre publicações científicas. Um exemplo desses esquemas classificatórios mais avançados é o que foi desenvolvido pelo CWTS (WALTMAN; VAN ECK, 2012). Essas classificações têm a vantagem de serem criadas com base nas citações de uma publicação a outra, gerando agrupamentos de publicações ligadas por relações de citações, e são, portanto, mais exatas do que sistemas de classificação baseados em periódicos e outros sistemas construídos manualmente (KLAVANS; BOYACK, 2016). O alto nível de detalhamento dessas classificações oferece a possibilidade de análises disciplinares com uma maior resolução.

3. Citações e cálculos métricos baseados em citações

Um elemento fundamental para o desenvolvimento de pesquisa de qualquer grupo de pesquisa cientométrica é a análise do impacto baseado em citações de periódicos científicos. A forma com que essas citações são calculadas não é um elemento irrelevante para o desenvolvimento de infraestrutura dos grupos. Normalmente, a identificação de relações entre documentos, através das citações, é proporcionada pelos fornecedores originais dos dados (por exemplo, Web of Science, Scopus ou Google Scholar) mas, em alguns casos, essa identificação é realizada pelos próprios grupos (OLENSKY; SCHIMIDT, VAN ECK, 2016).

Uma importante razão para se trabalhar tanto com esquemas de classificações adequados quanto com procedimentos de identificação de citações é o cálculo de indicadores de citações normalizados por área (WALTMAN et al., 2011), apesar de haver também procedimentos de normalização que não dependem de tais classificações (WALTMAN; VAN ECK, 2013). O cálculo desses indicadores é avançado e sofisticado. Portanto, não se fazem necessárias apenas a classificação e a identificação de citações, mas também um entendimento apropriado dos efeitos e problemas relacionados aos múltiplos elementos que precisam ser levados em conta no processo de normalização; por exemplo, tipos de documentos, ano de publicação, determinação da publicação, janelas de citação, etc.

CONSIDERAÇÕES FINAIS

Durante os últimos anos, houve um importante crescimento e popularização de ferramentas e dados cientométricos. Esse crescimento pode estar ligado ao surgimento do “big data”, e também à crescente atenção à informação cientométrica por diferentes grupos de atores (LEYDESDORFF; WOUTERS; BORNMANN, 2016). Entre esses atores estão os produtores dos dados originais, gestores em ciência, cientistas e bibliometristas. Segundo

Leydesdorff, Wouters e Bornmann (2016), gestores em ciência e cientistas são usuários ‘comuns’, enquanto produtores e bibliometristas são usuários profissionais. Neste capítulo, contribuimos para mostrar algumas das mais importantes características das infraestruturas que distinguem o trabalho de grupos de pesquisa profissionais dos usuários comuns de cientometria.

Em primeiro lugar, é importante observar que o desenvolvimento de infraestruturas substanciais para a pesquisa cientométrica tem alto custo (em termos de recursos econômicos, pessoal, bases de dados, tecnologias, etc.) e a decisão de iniciar um grupo de pesquisa cientométrica avançada deve contemplar esse custo.

Em segundo lugar, algumas das principais características relevantes (em diferentes graus) para a maioria dos grupos de pesquisa no campo da cientometria, que foram ilustradas neste capítulo, basicamente apontam para questões como: a necessidade de montar grupos com pessoas de diversas especialidades; a importância de poder contar com infraestruturas técnicas e de dados (como servidores, interfaces de consulta, etc.) adequadas; a importância de monitorar problemas e mudanças nas fontes de dados; a atenção ao surgimento de novas fontes de dados; e o desenvolvimento de instrumentos analíticos específicos (por exemplo, tesouros, classificações, metodologias de desambiguação, técnicas de normalização, algoritmos de identificação de citações, etc.) para vencer as limitações existentes em dados brutos.

Por fim, o desenvolvimento de infraestruturas avançadas de pesquisa cientométrica não deve ser desconectado dos desenvolvimentos mais fundamentais no contexto do sistema de comunicação da ciência. Já disse Wouters (2014): “podemos estar à beira da evolução de uma complexidade cada vez maior das infraestruturas do conhecimento, que podem, ou frustrar ou fazer progredir o desenvolvimento do conhecimento científico e acadêmico”. Portanto, é essencial entender como os “traços” deixados no sistema de comunicação da ciência são adequadamente capturados e estudados através dessas infraestruturas de pesquisa, e como as infraestruturas podem ser aprimoradas e adaptadas para lidar com as mudanças contínuas nas novas formas de produção de conhecimento.

REFERÊNCIAS

- BAR-ILAN, J. Informetrics at the beginning of the 21st century-A review. **Journal of Informetrics**, v. 2, n. 1, p. 1-52, 2008. doi: 10.1016/j.joi.2007.11.001.
- CARON, E.; VAN ECK, N. J. Large scale author name disambiguation using rule-based scoring and clustering. In: INTERNATIONAL CONFERENCE ON SCIENCE AND TECHNOLOGY INDICATORS, 19., 2014, Leiden. **Proceedings...** Leiden: CWTS-Leiden University, 2014. p. 79-86.
- D'ANGELO, C. A.; GIUFFRIDA, C.; ABRAMO, G. A heuristic approach to author name disambiguation in bibliometrics databases for large-scale research assessments. **Journal of the Association for Information Science and Technology**, v. 62, n. 2, p. 257-269, 2011. doi: 10.1002/asi.21460.
- DE RIJCKE, S.; RUSHFORTH, A. To intervene or not to intervene; is that the question? On the role of scientometrics in research evaluation. **Journal of the Association for Information Science and Technology**, v. 66, n. 9, p. 1954-1958, 2015. doi: 10.1002/asi.23382.
- DELGADO, E.; REPISO, R. The impact of scientific journals of communication: comparing Google Scholar Metrics, Web of Science and Scopus. **Comunicar**, v. 21, n. 41, p. 45-52, 2013. doi: 10.3916/C41-2013-04.
- EKBBIA, H.; MATTIOLI, M.; KOUPEL, I.; ARAVE, G.; GHAZINEJAD, A.; BOWMAN, T.; SURI, V. R.; TSOU, A.; WEINGART, S.; SUGIMOTO, C. R. Big Data, bigger dilemmas: a critical review. **Journal of the Association for Information Science and Technology**, v. 66, n. 8, p. 1523-1545, 2015. doi: 10.1002/asi.23294.
- FERNÁNDEZ, M. T.; CABRERO, A.; ZULUETA, M. A.; GÓMEZ, I. Constructing a relational database for bibliometric analysis. **Research Evaluation**, v. 3, n. 1, p. 56-62, 1993. doi: 10.1093/rev/3.1.55.
- FRANCESCHINI, F.; MAISANO, D.; MASTROGIACOMO, L. The museum of errors/horrors in Scopus. **Journal of Informetrics**, v. 10, n. 1, p. 174-182, 2016. doi: 10.1016/j.joi.2015.11.006.
- HARZING, A. W. Microsoft Academic (Search): a Phoenix arisen from the ashes? **Scientometrics**, v. 108, n. 3, p. 1637-1647, 2016. doi: 10.1007/s11192-016-2026-y.

- HARZING, A. W.; ALAKANGAS, S. Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, v. 106, n. 2, p. 787-804, 2016. doi: 10.1007/s11192-015-1798-9.
- HARZING, A. W.; ALAKANGAS, S. Microsoft Academic: is the phoenix getting wings? *Scientometrics*, v. 110, n. 1, p. 371-383, 2017. doi: 10.1007/s11192-016-2185-x.
- HAUSTEIN, S.; BOWMAN, T. D.; COSTAS, R. Interpreting “altmetrics”: viewing acts on social media through the lens of citation and social theories. In: SUGIMOTO, C. R. (Ed.). *Theories of informetrics: a festschrift in honor of blaise cronin*. Berlin: De Gruyter Mouton, 2016. p. 372-405.
- HOOD, W. W.; WILSON, C. S. Informetric studies using databases: opportunities and challenges. *Scientometrics*, v. 58, n. 3, p. 587-608, 2003. doi: 10.1023/B:SCIE.0000006882.47115.c6.
- HUG, S. E.; OCHSNER, M.; BRÄNDLE, M. P. Citation analysis with microsoft academic. *Scientometrics*, p. 1-8, 2017. doi: 10.1007/s11192-017-2247-8.
- KATZ, J. S.; HICKS, D. Desktop scientometrics. *Scientometrics*, v. 38, n. 1, p. 141-153, 1997. doi: 10.1007/BF02461128.
- KLAVANS, R.; BOYACK, K. W. Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *Journal of the Association for Information Science and Technology*, p. 1-15, 2016. doi: 10.1002/asi.23734.
- LEYDESDORFF, L.; WOUTERS, P.; BORNMANN, L. Professional and citizen bibliometrics: complementarities and ambivalences in the development and use of indicators – a state-of-the-art report. *Scientometrics*, v. 109, n. 3, p. 2129-2150, 2016. doi: 10.1007/s11192-016-2150-8.
- MARTÍN-MARTÍN, A.; ORDUNA-MALEA, E.; AYLLÓN, J. M.; LÓPEZ-CÓZAR, E. D. Back to the past: on the shoulders of an academic search engine giant. *Scientometrics*, v. 107, n. 3, p. 1477-1487, 2016. doi: 10.1007/s11192-016-1917-2.
- NEUHAUS, C.; DANIEL, H. D. Data sources for performing citation analysis: an overview. *Journal of Documentation*, v. 64, n. 2, p. 193-210, 2008. doi: 10.1108/00220410810858010.
- OLENSKY, M.; SCHMIDT, M.; VAN ECK, N. J. Evaluation of the citation matching algorithms of CWTS and iFQ in comparison to the Web of Science. *Journal of the*

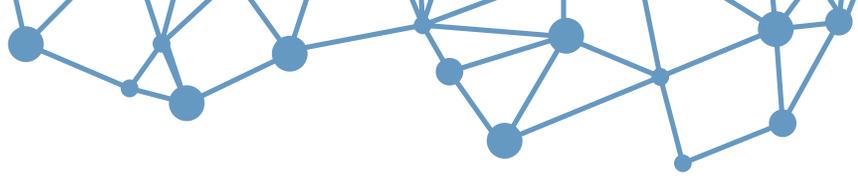
- Association for Information Science and Technology, v. 67, n. 10, p. 2550-2564, 2016. doi: 10.1002/asi.23590.
- ORDUÑA-MALEA, E.; AYLLÓN, J. M.; MARTÍN-MARTÍN, A.; LÓPEZ-CÓZAR, E. D. About the size of Google Scholar: playing the numbers. *EC3 Working Papers*, v. 18, p. 1-42, 2014. doi: 10.13140/RG.2.2.16418.43207.
- PAUL-HUS, A.; DESROCHERS, N.; COSTAS, R. Characterization, description, and considerations for the use of funding acknowledgement data in Web of Science. *Scientometrics*, v. 108, n. 1, p. 167-182, 2016. doi: 10.1007/s11192-016-1953-y.
- PRIEM, J.; TARABORELLI, D.; GROTH, P.; NEYLON, C. Altmetrics: a manifesto. 2010. Disponível em: <<http://altmetrics.org/manifesto>>. Acesso em: 14 fev. 2017.
- PRINS, A. A. M.; COSTAS, R.; LEEUWEN, T. N. VAN; WOUTERS, P. F. Using Google Scholar in research evaluation of humanities and social science programs: a comparison with Web of Science data. *Research Evaluation*, v. 25, n. 3, p. 264-270, 2016. doi: 10.1093/reseval/rvvo49.
- ROBINSON-GARCÍA, N.; TORRES-SALINAS, D.; ZAHEDI, Z.; COSTAS, R. New data, new possibilities: exploring the insides of Altmetric.com. *El Profesional de la Información*, v. 23, n. 4, p. 359-366, 2014. doi: 10.3145/epi.2014.jul.03.
- RUSHFORTH, A.; DE RIJCKE, S. Accounting for Impact? The Journal Impact Factor and the making of biomedical research in the Netherlands. *Minerva*, v. 53, n. 2, p. 117-139, 2015. doi: 10.1007/s11024-015-9274-5.
- SINHA, A.; SHEN, Z.; SONG, Y.; MA, H.; EIDE, D.; HSU, B. J.; WANG, K. An overview of Microsoft Academic Service (MAS) and applications. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 24., 2015, Florence. *Proceedings...* New York: ACM, 2015. p. 243-246. doi: 10.1145/2740908.2742839.
- SIRTES, D. Funding acknowledgments for the German Research Foundation (DFG). The dirty data of the Web of Science database and how to clean it up. In: INTERNATIONAL SOCIETY OF SCIENTOMETRICS AND INFORMETRICS CONFERENCE, 14., 2013, Vienna. *Proceedings...* Vienna: AIT GmbH, 2013. p. 784-795.
- SMALHEISER, N. R.; TORVIK, V. I. Author name disambiguation. *Annual Review of Information Science and Technology*, v. 43, n. 1, p. 1-43, 2009. doi: 10.1002/aris.2009.1440430113.

- TORRES-SALINAS, D.; RUIZ-PÉREZ, R.; LÓPEZ-CÓZAR, E. D. Google Scholar as a tool for research assessment. *El Profesional de la Información*, v. 18, n. 5, p. 501-510, 2009. doi: 10.3145/epi.2009.sep.03.
- WALTMAN, L.; VAN ECK, N. J. Source normalized indicators of citation impact: an overview of different approaches and an empirical comparison. *Scientometrics*, v. 96, n. 3, p. 699-716, 2013. doi: 10.1007/s11192-012-0913-4.
- WALTMAN, L.; CALERO-MEDINA, C.; KOSTEN, J.; NOYONS, E. C. M.; TIJSSEN, R. J. W.; VAN ECK, N. J.; VAN LEEUWEN, T. N.; VAN RAAN, A. F. J.; VISSER, M. S.; WOUTERS, P. The Leiden ranking 2011/2012: data collection, indicators, and interpretation. *Journal of the Association for Information Science and Technology*, v. 63, n. 12, p. 2419-2432, 2012.
- WALTMAN, L.; VAN ECK, N. J. A new methodology for constructing a publication-level classification system of science. *Journal of the Association for Information Science and Technology*, v. 63, n. 12, p. 2378-2392, 2012. doi: 10.1002/asi.22748.
- WALTMAN, L.; VAN ECK, N. J.; VAN LEEUWEN, T. N.; VISSER, M. S.; VAN RAAN, A. F. J. Towards a new crown indicator: an empirical analysis. *Scientometrics*, v. 87, n. 3, p. 467-481, 2011. doi: 10.1007/s11192-011-0354-5.
- WANG, Q.; WALTMAN, L. Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics*, v. 10, n. 2, p. 347-364, 2016. doi: 10.1016/j.joi.2016.02.003.
- WOUTERS, P. The citation: from culture to infrastructure. In: CRONIN, B.; SUGIMOTO, C. R. (Eds.). *Next Generation metrics: harnessing multidimensional indicators of scholarly performance*. Cambridge: MIT Press, 2014. p. 47-66.
- WOUTERS, P.; COSTAS, R. *Users, narcissism and control - tracking the impact of scholarly publications in the 21 st century*. Utrecht: SURF Foundation, 2012. 50 p.
- ZAHEDI, Z.; HAUSTEIN, S.; BOWMAN, T. D. Exploring data quality and retrieval strategies for Mendeley reader counts. In: WORKSHOP ON INFORMETRIC AND SCIENTOMETRIC RESEARCH, 14., 2014, Seattle. *Proceedings...* Seattle: ASIS&T, 2014.
- ZAHEDI, Z.; VAN ECK, N. J. Visualizing readership activity of Mendeley users using VOSviewer. In: ACM WEB SCIENCE CONFERENCE, 14., 2014,

Indiana. *Proceedings...* Bloomington: Indiana University, 2014. doi: 10.6084/m9.figshare.1041819.v2.

AGRADECIMENTOS

Agradeço a Rogério Mugnaini da Universidade de São Paulo pelo convite para participar do 5º Encontro Brasileiro de Bibliometria e Cientometria (5º EBBC – <http://www.ebbc.inf.br/ebbc5/index.php/ebbc5#>) e pelo apoio no desenvolvimento deste texto. Agradeço aos participantes do Encontro pelos pareceres e comentários durante o evento. Agradeço também a Ludo Waltman do CWTS – Universidade de Leiden por seus comentários e discussões sobre uma versão inicial deste texto.



GENERAL DISCUSSION ON THE MOST RELEVANT CHARACTERISTICS OF RESEARCH INFRASTRUCTURES FOR SCIENTOMETRICS

DISCUSSÕES GERAIS SOBRE AS CARACTERÍSTICAS MAIS RELEVANTES DE INFRAESTRUTURAS DE PESQUISA PARA A CIENTOMETRIA [ver página 19] [ir para o sumário]

*Rodrigo Costas**

INTRODUCTION

The use of scientometrics has experienced an important expansion during the last 20 years. This growing interest somehow echoes the interest coming from many different sectors, actors and stakeholders of the scientific system (LEYDESDORFF; WOUTERS; BORNMANN, 2016). This expansion has also been accompanied by the development and popularization of dedicated bibliometric software¹, as well as the emergence of new and more diverse data sources with multiple bibliometric possibilities (e.g. Google Scholar, Microsoft Academic or Mendeley, just to name a few). Moreover, the increase in computing capacity, storage and rapid access to bibliographic

* Centre for Science and Technology Studies (CWTS), Leiden University, Leiden, the Netherlands; rcostas@cwts.leidenuniv.nl

1 Among these we can mention VOSviewer (<http://www.vosviewer.com>), CitNetExplorer (<http://www.citnetexplorer.nl>), R-packages (<http://www.bibliometrix.org>), Bibexcel (<http://homepage.univie.ac.at/juan.gorraiz/bibexcel/>), CRExplorer (<http://www.crexplorer.net/>), PoP (<http://www.harzing.com/resources/publish-or-perish>), SCI2 (<https://sciz.cns.iu.edu/user/index.php>), etc.

data has provoked that more individuals have access to indicators, tools and applications, based on bibliometric information.

This landscape clearly links with the idea, proposed already 20 years ago by (KATZ; HICKS, 1997) of “desktop scientometrics”, promising the possibility of performing essential bibliometric tasks (e.g. data manipulation, unification, analysis or display of bibliometric information) from a simple PC and using desktop computing. Obviously “desktop scientometrics” have many advantages, particularly the better access to bibliometric tools and the popularization and broader interest for scientometric research. However, this popularization may also bring new challenges. An important challenge are the so-called “amateur bibliometrics” (RUSHFORTH; DE RIJCKE, 2015) or “quick and dirty” bibliometric approaches (BAR-ILAN, 2008), which may have the potential effect of trivializing, misusing and misunderstanding bibliometric indicators.

In order to provide a better idea of the value of the more “professionalized” or advanced bibliometrics (DE RIJCKE; RUSHFORTH, 2015), this chapter aims at discussing the relevance and most important characteristics of advanced infrastructures for scientometric work existing nowadays. The objective is to provide a general reflection on the most important types of infrastructures that are relevant for scientometric research. Thus, this description may be of help for new or emerging research groups interested in developing their own infrastructures for scientometric research, as well as for other users of bibliometric information (e.g. policy makers, research managers, etc.) that want to know more about the characteristics of this type of infrastructures.

RESEARCH TEAMS IN SCIENTOMETRICS WORLDWIDE

In this section we present a short description of some of the major international scientometric research groups. This list² is by no means intended to be exhaustive, but it is just to provide examples of research teams that, to some extent, share the characteristics, methodologies and infrastructures that will be discussed later on in the chapter.

CWTS – Centre for Science and Technology Studies, Leiden University, the Netherlands

<http://www.cwts.nl>

The Centre for Science and Technology Studies (CWTS – Centrum voor Wetenschap en Technologie Studies) is one of the biggest and well established international centers devoted to research in scientometrics. CWTS staff combines individuals with expertise from a diverse spectrum of disciplines. Through its research programs, the research scope of CWTS has expanded from more traditional scientometrics research to new topics such as altmetrics, societal impact of science, as well as more qualitative and mixed methods approaches. An important research-based output of the team is the Leiden Ranking (<http://www.leidenranking.com/>), which is published each year and provides advanced bibliometric indicators for more than 800 Universities worldwide.

DZHW – German Centre for Higher Education Research and Science Studies (in Hannover, Berlin and Leipzig), Germany.

<http://www.dzhw.eu/en>

-
- 2 The list is of course biased towards the own knowledge and experience of the author of this text, and the teams mentioned are presented just as examples of having relatively large scientometric infrastructures. There are of course other teams, groups of scholars and specialists that also do excellent work and share many of the characteristics and infrastructural capacities here discussed.

The German Centre for Higher Education Research and Science Studies – DZHW (now incorporating the former Institute for Research Information and Quality Assurance – IFQ) is an institute focused on carrying out application-oriented empirical research. It is not an institute solely devoted to scientometric research but to several Higher Education Research topics. From January 1st (2016) it integrates as Department 2 “Research System & Science Dynamics” the research lines and activities of the former IFQ, which includes research on bibliometrics, careers or research evaluation, among other topics.

Expertisecentrum O&O Monitoring (ECOOM), Leuven, Belgium

https://www.ecoom.be/nl/member_details/ku%20leuven

The Centre for Research & Development Monitoring (Expertisecentrum Onderzoek en Ontwikkelingsmonitoring, ECOOM) is an interuniversity consortium with participation of all Flemish universities (Katholieke Universiteit Leuven, Universiteit Gent, Vrije Universiteit Brussels, Universiteit Antwerpen and Universiteit Hasselt). Its mission is to develop a consistent system of R&D and Innovation (RD&I) indicators for the Flemish government, able to map and monitor the RD&I efforts in the Flemish region. It combines scientific expertise from a large diversity of backgrounds, going from the field of scientometrics, patent analysis (Technometrics), R&D Surveys or innovation surveys, among others.

SCImago Research Group, (Granada, Madrid), Spain.

<https://www.facebook.com/SRG-SCImago-Research-Group-189628671202/3>

-
- 3 Currently there is no official website of the group itself. Therefore it is difficult to assess the group's main characteristics, although through their publications and methodologies it is possible to guess that most of the characteristics of scientometric infrastructures explained below also apply to this group.

The SCImago Research Group is devoted to the study of the scholarly communication system and the development of tools to analyze, visualize and interpret the data extracted from scientific information databases. With a strong focus on Elsevier data, the group has developed evaluation products such as the Scimago Journal & Country rank (<http://www.scimagojr.com/>) and the Scimago Institutions Rankings (<http://www.scimagoir.com/>).

Statistical Cybermetrics Research Group, University of Wolverhampton, United Kingdom

<http://cybermetrics.wlv.ac.uk/>

This research group focuses mainly on the specific areas of webometrics and altmetrics within the scientometric realm, although their work also includes the more traditional bibliometric research. The team is particularly aimed at developing software and methods to exploit Web-based sources for webometrics, altmetrics and scientometrics research.

Canada Research Chair on the Transformations of Scholarly Communication (CRCTSC), Université de Montréal, Montréal, Canada

<http://crc.ebsi.umontreal.ca/en/>

This is one of the youngest research teams in scientometric research, outside Europe, that has achieved an important research capacity in the last years. Its main goal is to increase the understanding of how knowledge is currently being disseminated through scientific publications, and particularly how new types of knowledge production, sources of data and policies are changing the way academics develop their work. The team includes an important range of types of expertise, ranging from bibliometrics, altmetrics, linguistics to qualitative approaches and gender perspectives, among others.

Centre for Research on Evaluation, Science and Technology (CREST), Stellenbosch University, Stellenbosch, South Africa.

<http://www.sun.ac.za/crest/>

Another research center from outside Europe established in 1995, covering the fields of bibliometrics, research on higher education knowledge, human resources in science and technology or research evaluation and impact assessment among other topics related with the communication patterns of scholars. Since 2014 CREST also hosts the DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP – <http://www.sun.ac.za/scistip/>) which covers a large spectrum of diverse research topics such as Science in Africa, Human resources, Science communication or STI indicators among others. Thus, the center combines also a large variate of expertise, including more quantitative and data driven research, as well as more qualitative approaches.

As mentioned before, the list of research groups is not exhaustive. There are already established, as well as new emerging research teams in multiple countries (e.g. Spain, Denmark, Sweden, Norway, Finland, USA, Italy, Austria, etc.) that share many of the characteristics that are discussed in this chapter.

MAIN CHARACTERISTICS OF INFRASTRUCTURES FOR SCIENTOMETRIC RESEARCH

All the different research groups mentioned above have their own peculiarities, specific foci, research missions and objectives. However, in the following sections the focus is on their most important similarities and characteristics in terms of infrastructural aspects related with their scientometric research. Thus, we will focus on aspects related with their size and human

resources, technical infrastructures, databases, as well as other technical developments necessary for more advanced scientometric research work.

1) Human resources

In general, most of these teams depend on relatively large human resources, with staff composed of between 5 to 10, and sometimes up to more than 30 people, combining not only scientific staff, but also management, ICT and administration personnel. Regarding the scientific human resources, an important aspect that characterizes most of these teams is the multidisciplinary composition of their members. Thus, it is not rare to find among the major research teams individuals with different expertise, going from information scientists, computer scientists, statisticians, economists, political scientists, experts in research evaluation and research management, linguistics, etc.; in most of the cases also combining quantitative and qualitative perspectives.

An important element regarding the sustainability of these teams is the existence of specific business models that help to maintain their human resources and technological infrastructures. These business models may vary from producing commercial services and consultancies, as well as the development of specific scientometric projects, ranging from more international projects to more local, institutional or regional projects. Most of these teams, as most research scholars nowadays, have a strong dependence on the acquisition of external funding for their sustainability.

2) Technical infrastructures

The development of a significant technical infrastructure for scientometric research is one of the cornerstones of most of these teams. Thus, the availability of servers to store the data, the technology to query and access the data (e.g. SQL, SAS, Access databases, etc.) are central elements for the

work of the teams⁴. Similarly, ensuring enough computing capacity and the optimization of the different types of analysis is essential in order to gain in efficiency and agility in the production of indicators and studies.

2.1) Databases

Probably the main asset for establishing a research team in scientometrics is the availability of relevant data sources that will allow the development of the quantitative research work. Some of the most important data sources include international bibliographic databases that also index the citations linkages across the publications. Among the most important data sources for scientometric research we can mention the following:

1. *Web of Science– Clarivate Analytics*

<http://webofscience.com>

This is the most used database for scientometric studies. The foundation of today's Web of Science was the Institute for Scientific Information, founded by Eugene Garfield in the 60s. Most of the major scientometric research teams have access to (or have in-house versions of) the Web of Science Core Collection. The main databases included in the Core Collection are Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI) and Arts & Humanities Citation Index (A&HCI) and more recently the Emerging Sources Citation Index (ESCI). Other databases, are also occasionally used by different teams, including here for example the Conference Proceedings Citation Index (CPCI), Book Citation Index (BCI), Data Citation Index (DCI), , etc. However, the availability of these other sources varies greatly

4 Usually, the research teams also have access to the different database through their online web interfaces. However, this type of access tends to be problematic as it relies on the decisions and methodological choices made by the producers of the databases. The research interests of the different teams usually need of more advanced data access and management options. For this reason, many teams opt for having in-house versions of the bibliometric databases used for their research work.

among research teams, and their availability normally responds to specific interests or research foci of the different teams.

The Web of Science Core collection covers more than 12,000 international journals, considered as the international “mainstream” journals in science. The references of the source publications are indexed and the linkages between cited and citing publications are established based on internal algorithms determined either by the Web of Science itself or by the different research teams (OLENSKY; SCHMIDT; VAN ECK, 2016). An important characteristic of this database is the great diversity of metadata indexed, including the authors of the publications, their affiliations and countries, the basic bibliographic information (i.e. title, publication year, journal name, volume, issue, pages, identifiers of the documents, etc.), and more recently other para-textual elements of information such as funding acknowledgements (PAUL-HUS; DESROCHERS; COSTAS, 2016). This rich availability of metadata allows for multiple possibilities of analytical approaches, ranging from collaboration analysis to the textual and linguistic study of titles or abstracts, among others.

Without deeply entering into the advantages and limitations of the database, we can mention that among the advantages stands the long trajectory of this database, being one of the best studied data sources for scientometric research. Another advantage is its selective character, which intends to focus on the most important international scientific journals. However, it is precisely this selectivity that has been usually highlighted as its most important limitation, namely the bad coverage of some local journals as well as social sciences and humanities research (NEUHAUS; DANIEL, 2008).

2. Scopus – Elsevier

<https://www.scopus.com/>

Another major bibliometric data source is Scopus. This database covers more than 20,000 international and national scientific journals. Like Web of Science, Scopus also indexes the cited references and the linkages cited/citing

between the publications within the database. Similar to Web of Science it also has a large coverage of very diverse of metadata, including authors, affiliations, bibliographic data, etc.

Among the advantages of Scopus we can mention the large coverage of journals, countries and languages. Regarding the limitations we can mention the questionable quality of some of the metadata covered by Scopus (FRANCESCHINI; MAISANO; MASTROGIACOMO, 2016) as well as the shorter temporal coverage of publications (the database coverage starts in 1996).

3. Google Scholar

Google scholar is probably the most popular search engine of scientific publications currently available online. It has strong advantages such as its free availability (both Scopus and Web of Science are commercial products). The coverage of Google Scholar has been also highlighted as the largest of all existing databases (DELGADO; REPISO, 2013), including not only scientific articles, but also pre-prints, books, book chapters, reports, grey literature, etc., although it is unclear what the actual coverage of Google Scholar is (ORDUÑA-MALEA et al., 2014). An important feature of Google Scholar regarding its capacity for scientometric research is that, like Web of Science and Scopus, it also indexes the citations received by the records included in the database.

In spite of these important advantages, Google Scholar has also been criticized for several reasons. Thus, its “black box” nature has been pointed out (WOUTERS; COSTAS, 2012). The lack of information on its actual coverage, the low quality of the metadata, and the lack of information about its citation matching procedure are some of the elements also criticized about Google Scholar (TORRES-SALINAS; RUIZ-PÉREZ; LÓPEZ-CÓZAR, 2009). Moreover, the lack of a proper systematic manner to access and collect Google Scholar data has also prevented the possibility of cleaning and enhancing the original raw data by the different research

groups. Furthermore, Google Scholar does not have the richness in metadata as provided by the other commercial databases, and for example critical information such as the affiliations or the countries of the authors are missing from the records.

Given this situation, not many research teams have incorporated Google Scholar as their main data source for scientometric research, and most of the studies based on this source focused on relatively small datasets, which still are very labor-intensive to obtain (PRINS et al., 2016). Perhaps the team that has worked the most with Google Scholar is the EC₃ team in Granada (Spain) (MARTÍN-MARTÍN et al., 2016), together with Ann Harzing (HARZING; ALAKANGAS, 2016). However, their results, although promising, still do not support a proper incorporation of Google Scholar as a reliable large-scale source for scientometric research.

4. Microsoft Academic

Similar to Google Scholar, Microsoft Academic is also a search engine of scientific publications. Unlike Google Scholar, Microsoft Academic has the possibility of searching for a given institution or entity (e.g. universities). Preliminary studies have highlighted the relevance and larger coverage of this source in contrast to the commercial sources (SINHA et al., 2015; HARZING, 2016), estimating that the database has significantly improved with respect to its previous versions (HARZING; ALAKANGAS, 2016). Microsoft Academic also offers some calculation of citation counts, although this calculation is an estimation of the number of citations received by each publication provided by Microsoft Academic⁵.

Regarding the limitations of this database, Microsoft Academic shares many of the limitations of Google Scholar. Among these limitations we can mention the lack of information about its actual coverage (although it seems

5 <https://microsoftacademic.uservoice.com/knowledgebase/articles/838965-microsoft-academic-faq#citations>

to be lower than Google Scholar – HARZING, 2016), the quality and accuracy of the metadata provided.

To the best of our knowledge, there are no research teams in scientometrics that are actively using Microsoft Academic research as a primary source of bibliometric data. Although Microsoft Academic has been identified as a promising tool for scientometric research (WOUTERS; COSTAS, 2012) and although some tools such as Publish or Perish (PoP) also include this source for calculations, so far there are not many research teams that are actively doing scientometric research based on this source⁶.

5. Local and thematic databases

In addition to the more international databases, most of the research teams also use data sources with more local or thematic interest. Thus, for example, among the Spanish scientometric research teams they work with databases covering the Spanish-relevant production. For example databases such as ICYT (Index of Science and Technology), ISOC (Index of Social Sciences) or IME (Spanish Medical Index)⁷. In the case of research teams with a focus on Latin American production they have the database Scielo⁸ (which also contains citation information), and for the French Canadian related information, a database such as Érudit⁹ is of strong relevance. Regarding the availability of thematic databases, MEDLINE or PubMed are also common databases used by the different research teams in scientometrics.

6. Altmetric and open science databases

During the last years, the emergence of the altmetrics movement (PRIEM et al., 2010) has also opened the possibility for the different

6 A recent paper by Hug, Ochsner and Brändle (2017) also demonstrates the strong potential of Microsoft Academic as a data source for scientometric research.

7 <http://bddoc.csic.es:8080/>

8 <http://www.scielo.org>

9 <http://www.erudit.org/>

research teams in scientometrics to study the reception of science in other sources. Nowadays it is possible to study how scientific publications are being mentioned in sources such as Twitter, Facebook, blogs, F1000, Mendeley, etc., opening new venues of analyzing the social media reception of scientific publications.

Among these databases, the most common ones include Altmetric.com (<https://www.altmetric.com/>) and Mendeley.com. Altmetric.com has collected the social media reception of more than 5 million publications from all over the world. This database provides statistics on the number of times a publication has been mentioned in Twitter, blogs, Facebook and some other social media platforms. The value and relevance of all these new social media sources is still under discussion within the scientific community (WOUTERS; COSTAS, 2012; ROBINSON-GARCÍA et al., 2014).

Mendeley.com is an online reference manager. Users of Mendeley can save publications and manage their individual libraries for different purposes (e.g. research work, study, collaboration, etc.). The main advantage of Mendeley as an altmetric source is that it provides the number of users that have saved a given publication in their personal libraries, also known as “readership”. Additionally, it is also possible to disaggregate the counts of “readership” by general typologies of users (e.g. PhD, Professors, Students, etc. – cf. (ZAHEDI; VAN ECK, 2014)). Mendeley offers its metrics for free through its open API (<http://dev.mendeley.com/>), thus this source can be freely used to obtain readership metrics for nearly any existing publication. The main limitation however is that the Mendeley database as a whole cannot be directly obtained, and therefore it is always necessary to work with an original source database (e.g. Web of Science, Scopus, MEDLINE, etc.) to query the Mendeley API. Furthermore, it is important to keep in mind that the metadata provided by Mendeley is not of high quality (ZAHEDI; HAUSTEIN; BOWMAN, 2014) and many important pieces of information (such as affiliations or countries of the authors) are not included in the data provided by Mendeley.

From the perspective of studying movements such as Open Access or Open Data, databases like OpenAIRE (<https://www.openaire.eu/>) or

DataCite (<https://www.datacite.org/>) are important sources that research teams in scientometrics are starting to study. These data sources are available through public APIs (<http://api.openaire.eu/>; <https://mds.datacite.org/static/apidoc>), allowing any interested user to download and analyze their data.

An important conceptual element that needs to be taken into account when working with these new altmetric and open science/data sources is that they are still new and more research is necessary in order to understand their value and relevance for scientometric research (HAUSTEIN; BOWMAN; COSTAS, 2016). A more important practical limitation of these data sources is their strong reliance on publication identifiers as DOIs or PubMed ids, and still the quality of their metadata is highly problematic (ZAHEDI; HAUSTEIN; BOWMAN, 2014).

It is also important to highlight here the increasing availability of full-text databases, such as PLOS journals or PubMed Central. Also some large commercial publishers such as Elsevier, Springer or Wiley also allow (under certain conditions) the full-text databases of their databases. These possibilities also allow for the development of more text mining types of analysis, framed in the general context of the big data perspectives currently gaining importance in scientific research (EKBI et al., 2015)

2.2) Other technological developments

In addition to the availability of databases and specific data sources, usually scientometric work also needs additional tools and instruments that are necessary for the proper analysis of the data and derived indicators. Some of these necessary infrastructures are described below:

1. Cleaning of data and development of thesauri and specific algorithms desenvolvimento

Very frequently the data provided by the different data sources have severe problems of standardization and data quality (HOOD; WILSON,

2003). Thus, for example, the affiliation names of institutions in Web of Science or Scopus usually need extensive cleaning and harmonization (WALTMAN et al., 2012). Similarly, other pieces of information such as funding acknowledgment data also need extensive cleaning and harmonization. The most common solution is the development of specialized thesauri of research organization and/or funders, in which the different name variants of the different units are homogenized under a canonical form, and also the relationships (e.g. of dependence or equivalence) among them are established (FERNÁNDEZ et al., 1993; SIRTES, 2013).

Another important element is the harmonization of the different bibliographic elements (e.g. titles, publication year, journal names, etc.). This is less critical for the more established commercial databases (WoS or Scopus) as their metadata are usually quite standardized, but it is more relevant for data sources that aggregate data from different sources (e.g. DataCite or OpenAIRE) or online data (e.g. Google Scholar, Microsoft Academic).

A special case that requires attention when working with scientometric information is the importance of author-name disambiguation (SMALHEISER; TORVIK, 2009). This aspect is critical when working with information at the individual-researcher level, and the lack of knowledge of problems like homonymy or synonymy can significantly distort the results of the analyses. Therefore, scientometric teams also tend to invest important resources in manual, automated or semi-automated disambiguation of author names in their databases (D'ANGELO; GIUFFRIDA; ABRAMO, 2011; CARON; VAN ECK, 2014).

2. Thematic classifications and disciplinary schemes

Another central element in scientometric research is the availability of instruments that allow the possibility of thematic analysis. Thus, the availability of thematic classifications becomes relevant in the development of research in scientometrics work (WALTMAN; VAN ECK, 2012).

A very common classification used in scientometric research are the Web of Science Subject Categories. This classification consists of about 250 research areas. It is a very convenient classification as it is already incorporated in the Web of Science database, thus making the use of this classification very straightforward. A similar classification is also provided by Elsevier's Scopus database. However, this type of classifications are not free of limitations. One of the most important limitations is that they are based on the classification of scientific journals, instead of individual scientific publications, exhibiting different degrees of accuracy (cf., WANG; WALTMAN, 2016). Thus, all publications in a given journal are classified in the same research area of the journal, instead of having their own individual classification. This is especially problematic for large multidisciplinary journals, such as *Nature*, *Science*, *PNAS*, and especially *PLOS ONE* and *Scientific Reports*.

Alternatively other classifications are also used by different research teams, many times related to their own local environment or to other different interests. Thus, examples of classifications include the National Science Foundation (NSF), the classification of the OECD, more country-specific classification like the NOWT classification in the Netherlands, or other multiple-purpose classifications like the ScienceMetrix classification (<http://www.science-metrix.com/en/classification>). Similarly, disciplinary classifications are also occasionally used. For example the Medical Subject Headings (MeSH) is a good example of a relevant disciplinary classification.

More recently, new classifications have been developed based on the citation relationships among scientific publications. An example of this more advanced classificatory schemes is the one developed at CWTS (WALTMAN; VAN ECK, 2012). These classifications have the advantage that they are created on the basis of citations from one publication to another, creating clusters of publications linked by citation relationships and therefore more accurate than journal-based classification systems and other manually constructed systems (KLAVANS; BOYACK, 2016). The high level of detail of these classifications, offer the possibility of disciplinary analysis with a higher resolution.

3. Citations and citation-based metrics calculation

A pivotal element in the research development of any research team in scientometrics is the analysis of the citation impact of scientific publications. How these citations are calculated is not a minor element in the infrastructural developments of the teams. Usually, the citation matching is provided by the original data providers (e.g. Web of Science, Scopus or Google scholar), however in some cases this citation matching is developed by the teams themselves (OLENSKY; SCHIMIDT, VAN ECK, 2016).

One important reason for working with both adequate classification schemes and citation matching procedures is the calculation of field-normalized citation indicators (WALTMAN et al., 2011), although there are also normalization procedures that are not dependent on such classifications (WALTMAN; VAN ECK, 2013). The calculation of these indicators is quite advanced and sophisticated. Thus, it is not only the classification and the identification of citation linkages that are necessary, but also a proper understanding of the effects and problems related with the multiple elements that need to be accounted for in the normalization process; for example document types, publication years, determination of publication and citation windows, etc.

FINAL REMARKS

During the last years there has been an important growth and popularization of scientometric tools and data. This growth can also be linked to the rise of “big data” as well as to the growing attention to scientometric information by different groups of actors (LEYDESDORFF; WOUTERS; BORNMANN, 2016). These actors include the producers of the original data, science managers, scientists and bibliometricians. According to Leydesdorff, Wouters and Bornmann (2016), science managers and scientists are a kind of ‘citizen’ users, while producers and bibliometricians are more

professional types of users. In this chapter we contribute to this idea by depicting some of the most important infrastructural characteristics that would distinguish the work of the more professional research teams from the more ‘citizen’ types of users of scientometrics.

In the first place, it is important to highlight that the development of substantial scientometric research infrastructures is costly (in terms of economic resources, people, databases, technologies, etc.), and the decision of starting an advanced research team in scientometrics must contemplate these costs.

Secondly, some of the most important characteristics that are (with different degrees) important for most of the major research teams in the field of scientometrics that have been depicted in this chapter, basically point to questions such as the need of assembling teams of people with diverse backgrounds, the importance of counting with suitable technical and data infrastructures (e.g. servers, querying interfaces, etc.), the importance of monitoring issues and changes in the data sources, the attention to the emergence of new data sources, or the development of specific analytical instruments (e.g. thesauri, classifications, disambiguation methodologies, normalization techniques, citation matching algorithms, etc.) to overcome the limitations existing in the raw data.

Finally, the development of advanced scientometric research infrastructures cannot be disconnected from the more fundamental developments in the context of the communication system of science. As stated by Wouters (2014) “[w]e may be on the verge of the evolution of an increasing complexity of knowledge infrastructures, which may either frustrate or bring forward the development of scientific and scholarly knowledge”. Therefore it is essential to understand how the “traces” left in the communication system of science are properly captured and studied through these research infrastructures and how the infrastructures can be improved and adapted to cope with the ongoing changes in the novel forms of knowledge production.

REFERENCES

- BAR-ILAN, J. Informetrics at the beginning of the 21st century-A review. **Journal of Informetrics**, v. 2, n. 1, p. 1-52, 2008. doi: 10.1016/j.joi.2007.11.001.
- CARON, E.; VAN ECK, N. J. Large scale author name disambiguation using rule-based scoring and clustering. In: INTERNATIONAL CONFERENCE ON SCIENCE AND TECHNOLOGY INDICATORS, 19., 2014, Leiden. **Proceedings...** Leiden: CWTS-Leiden University, 2014. p. 79-86.
- D'ANGELO, C. A.; GIUFFRIDA, C.; ABRAMO, G. A heuristic approach to author name disambiguation in bibliometrics databases for large-scale research assessments. **Journal of the Association for Information Science and Technology**, v. 62, n. 2, p. 257-269, 2011. doi: 10.1002/asi.21460.
- DE RIJCKE, S.; RUSHFORTH, A. To intervene or not to intervene; is that the question? On the role of scientometrics in research evaluation. **Journal of the Association for Information Science and Technology**, v. 66, n. 9, p. 1954-1958, 2015. doi: 10.1002/asi.23382.
- DELGADO, E.; REPISO, R. The impact of scientific journals of communication: comparing Google Scholar Metrics, Web of Science and Scopus. **Comunicar**, v. 21, n. 41, p. 45-52, 2013. doi: 10.3916/C41-2013-04.
- EKBIA, H.; MATTIOLI, M.; KOUPEL, I.; ARAVE, G.; GHAZINEJAD, A.; BOWMAN, T.; SURI, V. R.; TSOU, A.; WEINGART, S.; SUGIMOTO, C. R. Big Data, bigger dilemmas: a critical review. **Journal of the Association for Information Science and Technology**, v. 66, n. 8, p. 1523-1545, 2015. doi: 10.1002/asi.23294.
- FERNÁNDEZ, M. T.; CABRERO, A.; ZULUETA, M. A.; GÓMEZ, I. Constructing a relational database for bibliometric analysis. **Research Evaluation**, v. 3, n. 1, p. 56-62, 1993. doi: 10.1093/rev/3.1.55.
- FRANCESCHINI, F.; MAISANO, D.; MASTROGIACOMO, L. The museum of errors/horrors in Scopus. **Journal of Informetrics**, v. 10, n. 1, p. 174-182, 2016. doi: 10.1016/j.joi.2015.11.006.
- HARZING, A. W. Microsoft Academic (Search): a Phoenix arisen from the ashes? **Scientometrics**, v. 108, n. 3, p. 1637-1647, 2016. doi: 10.1007/s11192-016-2026-y.

- HARZING, A. W.; ALAKANGAS, S. Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, v. 106, n. 2, p. 787-804, 2016. doi: 10.1007/s11192-015-1798-9.
- HARZING, A. W.; ALAKANGAS, S. Microsoft Academic: is the phoenix getting wings? *Scientometrics*, v. 110, n. 1, p. 371-383, 2017. doi: 10.1007/s11192-016-2185-x.
- HAUSTEIN, S.; BOWMAN, T. D.; COSTAS, R. Interpreting “altmetrics”: viewing acts on social media through the lens of citation and social theories. In: SUGIMOTO, C. R. (Ed.). *Theories of informetrics: a festschrift in honor of blaise cronin*. Berlin: De Gruyter Mouton, 2016. p. 372-405.
- HOOD, W. W.; WILSON, C. S. Informetric studies using databases: opportunities and challenges. *Scientometrics*, v. 58, n. 3, p. 587-608, 2003. doi: 10.1023/B:SCIE.0000006882.47115.c6.
- HUG, S. E.; OCHSNER, M.; BRÄNDLE, M. P. Citation analysis with microsoft academic. *Scientometrics*, p. 1-8, 2017. doi: 10.1007/s11192-017-2247-8.
- KATZ, J. S.; HICKS, D. Desktop scientometrics. *Scientometrics*, v. 38, n. 1, p. 141-153, 1997. doi: 10.1007/BF02461128.
- KLAVANS, R.; BOYACK, K. W. Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *Journal of the Association for Information Science and Technology*, p. 1-15, 2016. doi: 10.1002/asi.23734.
- LEYDESDORFF, L.; WOUTERS, P.; BORNMANN, L. Professional and citizen bibliometrics: complementarities and ambivalences in the development and use of indicators – a state-of-the-art report. *Scientometrics*, v. 109, n. 3, p. 2129-2150, 2016. doi: 10.1007/s11192-016-2150-8.
- MARTÍN-MARTÍN, A.; ORDUNA-MALEA, E.; AYLLÓN, J. M.; LÓPEZ-CÓZAR, E. D. Back to the past: on the shoulders of an academic search engine giant. *Scientometrics*, v. 107, n. 3, p. 1477-1487, 2016. doi: 10.1007/s11192-016-1917-2.
- NEUHAUS, C.; DANIEL, H. D. Data sources for performing citation analysis: an overview. *Journal of Documentation*, v. 64, n. 2, p. 193-210, 2008. doi: 10.1108/00220410810858010.
- OLENSKY, M.; SCHMIDT, M.; VAN ECK, N. J. Evaluation of the citation matching algorithms of CWTS and iFQ in comparison to the Web of Science. *Journal of the*

- Association for Information Science and Technology, v. 67, n. 10, p. 2550-2564, 2016. doi: 10.1002/asi.23590.
- ORDUÑA-MALEA, E.; AYLLÓN, J. M.; MARTÍN-MARTÍN, A.; LÓPEZ-CÓZAR, E. D. About the size of Google Scholar: playing the numbers. *EC3 Working Papers*, v. 18, p. 1-42, 2014. doi: 10.13140/RG.2.2.16418.43207.
- PAUL-HUS, A.; DESROCHERS, N.; COSTAS, R. Characterization, description, and considerations for the use of funding acknowledgement data in Web of Science. *Scientometrics*, v. 108, n. 1, p. 167-182, 2016. doi: 10.1007/s11192-016-1953-y.
- PRIEM, J.; TARABORELLI, D.; GROTH, P.; NEYLON, C. Altmetrics: a manifesto. 2010. Disponível em: <<http://altmetrics.org/manifesto>>. Acesso em: 14 fev. 2017.
- PRINS, A. A. M.; COSTAS, R.; LEEUWEN, T. N. VAN; WOUTERS, P. F. Using Google Scholar in research evaluation of humanities and social science programs: a comparison with Web of Science data. *Research Evaluation*, v. 25, n. 3, p. 264-270, 2016. doi: 10.1093/reseval/rvvo49.
- ROBINSON-GARCÍA, N.; TORRES-SALINAS, D.; ZAHEDI, Z.; COSTAS, R. New data, new possibilities: exploring the insides of Altmetric.com. *El Profesional de la Información*, v. 23, n. 4, p. 359-366, 2014. doi: 10.3145/epi.2014.jul.03.
- RUSHFORTH, A.; DE RIJCKE, S. Accounting for Impact? The Journal Impact Factor and the making of biomedical research in the Netherlands. *Minerva*, v. 53, n. 2, p. 117-139, 2015. doi: 10.1007/s11024-015-9274-5.
- SINHA, A.; SHEN, Z.; SONG, Y.; MA, H.; EIDE, D.; HSU, B. J.; WANG, K. An overview of Microsoft Academic Service (MAS) and applications. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 24., 2015, Florence. *Proceedings...* New York: ACM, 2015. p. 243-246. doi: 10.1145/2740908.2742839.
- SIRTES, D. Funding acknowledgments for the German Research Foundation (DFG). The dirty data of the Web of Science database and how to clean it up. In: INTERNATIONAL SOCIETY OF SCIENTOMETRICS AND INFORMETRICS CONFERENCE, 14., 2013, Vienna. *Proceedings...* Vienna: AIT GmbH, 2013. p. 784-795.
- SMALHEISER, N. R.; TORVIK, V. I. Author name disambiguation. *Annual Review of Information Science and Technology*, v. 43, n. 1, p. 1-43, 2009. doi: 10.1002/aris.2009.1440430113.

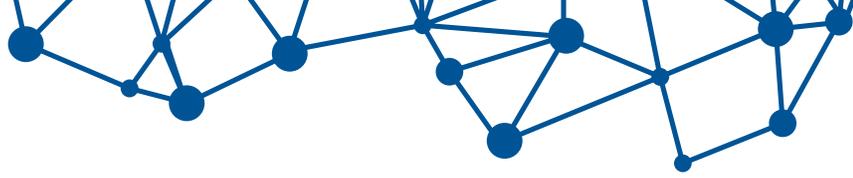
- TORRES-SALINAS, D.; RUIZ-PÉREZ, R.; LÓPEZ-CÓZAR, E. D. Google Scholar as a tool for research assessment. *El Profesional de la Información*, v. 18, n. 5, p. 501-510, 2009. doi: 10.3145/epi.2009.sep.03.
- WALTMAN, L.; VAN ECK, N. J. Source normalized indicators of citation impact: an overview of different approaches and an empirical comparison. *Scientometrics*, v. 96, n. 3, p. 699-716, 2013. doi: 10.1007/s11192-012-0913-4.
- WALTMAN, L.; CALERO-MEDINA, C.; KOSTEN, J.; NOYONS, E. C. M.; TIJSSEN, R. J. W.; VAN ECK, N. J.; VAN LEEUWEN, T. N.; VAN RAAN, A. F. J.; VISSER, M. S.; WOUTERS, P. The Leiden ranking 2011/2012: data collection, indicators, and interpretation. *Journal of the Association for Information Science and Technology*, v. 63, n. 12, p. 2419-2432, 2012.
- WALTMAN, L.; VAN ECK, N. J. A new methodology for constructing a publication-level classification system of science. *Journal of the Association for Information Science and Technology*, v. 63, n. 12, p. 2378-2392, 2012. doi: 10.1002/asi.22748.
- WALTMAN, L.; VAN ECK, N. J.; VAN LEEUWEN, T. N.; VISSER, M. S.; VAN RAAN, A. F. J. Towards a new crown indicator: an empirical analysis. *Scientometrics*, v. 87, n. 3, p. 467-481, 2011. doi: 10.1007/s11192-011-0354-5.
- WANG, Q.; WALTMAN, L. Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics*, v. 10, n. 2, p. 347-364, 2016. doi: 10.1016/j.joi.2016.02.003.
- WOUTERS, P. The citation: from culture to infrastructure. In: CRONIN, B.; SUGIMOTO, C. R. (Eds.). *Next Generation metrics: harnessing multidimensional indicators of scholarly performance*. Cambridge: MIT Press, 2014. p. 47-66.
- WOUTERS, P.; COSTAS, R. *Users, narcissism and control - tracking the impact of scholarly publications in the 21 st century*. Utrecht: SURF Foundation, 2012. 50 p.
- ZAHEDI, Z.; HAUSTEIN, S.; BOWMAN, T. D. Exploring data quality and retrieval strategies for Mendeley reader counts. In: WORKSHOP ON INFORMETRIC AND SCIENTOMETRIC RESEARCH, 14., 2014, Seattle. *Proceedings...* Seattle: ASIS&T, 2014.
- ZAHEDI, Z.; VAN ECK, N. J. Visualizing readership activity of Mendeley users using VOSviewer. In: ACM WEB SCIENCE CONFERENCE, 14., 2014,

Indiana. *Proceedings...* Bloomington: Indiana University, 2014. doi: 10.6084/m9.figshare.1041819.v2.

ACKNOWLEDGMENTS

Rogério Mugnaini from the Universidade de São Paulo is thanked for the invitation to participate in the 50 Encontro Brasileiro de Bibliometria e Cien-tometria (50 EBBC – <http://www.ebbc.inf.br/ebbc5/index.php/ebbc5#>) and for his support in the development of this text. The participants of the Encontro are also thanked for their feedback and comments during the event. Ludo Waltman from CWTS–Leiden University is also kindly thanked for his comments and discussions on an early draft of this text.





RUMO A INDICADORES PARA ‘ABERTURA’ DE POLÍTICAS DE CIÊNCIA E TECNOLOGIA¹

TOWARDS INDICATORS FOR ‘OPENING UP’ SCIENCE AND TECHNOLOGY POLICY [see page 79]

[go to summary]

Ismael Rafols^{a,b}, Tommaso Ciarli^b e Andy Stirling^b

INTRODUÇÃO

Ao longo dos últimos anos, tem-se observado muitos debates críticos acerca do uso simplista de ferramentas cientométricas para a avaliação formal ou informal de organizações (por exemplo, rankings universitários) ou indivíduos (por exemplo, o índice h) atuantes na ciência e tecnologia (C&T) (ROESSNER, 2000; VAN RAAN, 2005; WEINGART, 2005). Como reação a essas críticas, houve um empenho para aprimorar a robustez das medições ao se ampliar a variedade dos *inputs* considerados nas avaliações cientométricas. Alguns exemplos envolvem a inclusão de periódicos nacionais ou regionais e livros (MARTIN et al., 2010) ou, mais recentemente, a ‘altmetria’ (isto é, a métrica baseada em fontes de dados alternativas, vide Priem et al., 2010). Com isso, as comunidades de indicadores e de políticas da C&T voltaram a acreditar que a cientometria deve contar com múltiplas

¹ Este artigo foi publicado previamente em 2012, nos anais da 17th *International Conference on Science and Technology Indicators (STI)*, realizada em Montreal.

^a *Ingenio* (CSIC-UPV), Universitat Politècnica de València, València, Espanha, i.rafols@ingenio.upv.es

^b SPRU (Science Policy Research Unit), University of Sussex, Brighton, Reino Unido

fontes de dados que podem proporcionar ‘indicadores parciais convergentes’ (MARTIN; IRVINE, 1983).

Conquanto essa ‘ampliação’ da variedade de dados utilizados como ‘*inputs*’ à avaliação cientométrica seja, sob nossa perspectiva, louvável (STIRLING, 2003), propomos neste artigo que é necessário considerar também uma segunda dimensão. Isso se aplica na medida em que os ‘*outputs*’ apreciativos se ‘abrem’ a conceitualizações contrastantes do fenômeno que está sob investigação, assim permitindo uma observação mais ponderada e rigorosa às opções de políticas alternativas, tanto pelos tomadores de decisão quanto em um debate mais abrangente de políticas (STIRLING, 2005; STIRLING et al., 2007; LEACH; SCOONES; STIRLING, 2010). Usamos um estudo comparativo recente sobre o desempenho e interdisciplinaridade de seis unidades organizacionais (RAFOLS et al., 2012) para ilustrar a diferença entre o aumento da variedade de *inputs* (‘amplitude’) e o aprimoramento da diversidade de *outputs* no contexto das tomadas de decisões políticas (‘abertura’). Assim, a avaliação de políticas poderá subsidiar as tomadas de decisões de uma forma ‘plural e condicionada’ mais rigorosa – reconhecendo a maneira com que suposições e métricas normativas divergentes podem render entendimentos contrastantes do fenômeno sob investigação e das respostas apropriadas às políticas (STIRLING, 2008).

ESTRUTURA CONCEITUAL: ‘ABERTURA’ E ‘AMPLITUDE’ NA AVALIAÇÃO DE POLÍTICAS

Muitos indicadores de C&T foram desenvolvidos nos últimos 50 anos como meio de revelar os ‘pontos fortes’ e ‘pontos fracos’ da ‘capacidade’ e ‘desempenho’ de um dado país em ciência e tecnologia (GODIN, 2003). Desenvolvimentos da OECD (Organização para a Cooperação e Desenvolvimento Econômico) e da US National Science Board (NSB – Conselho Nacional da Ciência dos Estados Unidos) derivaram de uma ‘estrutura pura de contabilidade baseada nos benefícios econômicos esperados da ciência’

(GODIN, 2007) e, portanto, têm tendência a assumir um entendimento básico da produção e excelência científicas, influenciados por conceitos econômicos como ‘eficiência’ e ‘efetividade’ (NARIN, 1987). Os primeiros estudos cientométricos tomaram o cuidado de expor limitações metodológicas, por exemplo, alegando explicitamente que citações eram representações e medidas ‘parciais e imperfeitas’ de impacto, em vez de qualidade (MARTIN; IRVINE, 1983). No entanto, seja por cautela ou não, a ênfase dos estudos cientométricos tradicionalmente prioriza a produção de uma ‘boa’ medida de um dado conceito, como ‘excelência científica’, em vez de oferecer perspectivas contrastantes do significado de ‘excelência’.

Nos últimos anos, variados desenvolvimentos paralelos começaram a desafiar esse *status quo* cientométrico. Primeiro, a sutil difusão de medidas cientométricas simplistas (e possivelmente prejudiciais), como o índice *h* em vários níveis de gestão, renovou o debate sobre o abuso e mau uso de indicadores (WEINGART, 2005). Segundo, a cientometria tradicional é desafiada por fontes de dados alternativas, como bases de dados de países até agora ignoradas (por exemplo, o SciELO do Brasil), e novos indicadores baseados na Internet, como a frequência de *downloads* ou popularidade de sites 2.0 tal qual o *academia.eu* (PRIEM et al., 2010). Terceiro, houve o surgimento de novas ferramentas para a visualização de dados (por exemplo, o *Gapminder* de Hans Rosling), para grandes análises de redes (por exemplo, ROSVALL; BERGSTROM, 2008) e para o mapeamento científico (BÖRNER, 2010), que estão facilitando radicalmente a apresentação de informações quantitativas multidimensionais a leigos.

Cada uma dessas tendências está empurrando as políticas de C&T em direção a indicadores baseados em *inputs* com maior diversidade de dados. Esses portfólios mais amplos de *inputs* podem, a princípio, tornar as análises cientométricas em algo mais robusto. Porém, argumentamos aqui que essa ‘amplitude’ aprimorada de *inputs* não necessariamente se traduzirá em um processo político mais plural e condicional. A ‘abertura’ não envolve apenas ‘mais’ indicadores, e nem é uma questão de ‘posicionamento’ ou de contextualização (LEPORI, 2006). Trata-se do planejamento e uso de indicadores voltados

explicitamente para proporcionar entendimentos e opções plurais das políticas. Para que haja uma ‘abertura’ de políticas, os indicadores utilizados na avaliação precisam ser concebidos novamente como ‘dispositivos discutíveis, que permitam um aprendizado coletivo’ (BARRÉ, 2010, p. 227).

Dessa forma, distinguimos duas dimensões em qualquer processo de avaliação de políticas, como ilustrado na Figura 1. A primeira dimensão, ‘amplitude’, se refere à profundidade, extensão e escopo, com os quais a avaliação inclui diferentes tipos de conhecimento que podem descrever o fenômeno sob investigação (LEACH; SCOONES; STIRLING, 2010). A segunda dimensão, ‘abertura’, se refere ao grau em que os *outputs* apreciativos proporcionam interpretações plurais e condicionais do fenômeno – assim permitindo opções de políticas contrastantes para serem debatidas de forma rigorosa. Diferente das ferramentas analíticas que ‘restringem’ a avaliação, pelo estabelecimento de uma classificação absoluta de ‘melhores’ escolhas, as ferramentas de ‘abertura’ permitem que tomadores de decisões comparem a forma com que, sob diferentes suposições, as análises possam resultar em diversas classificações de opções.

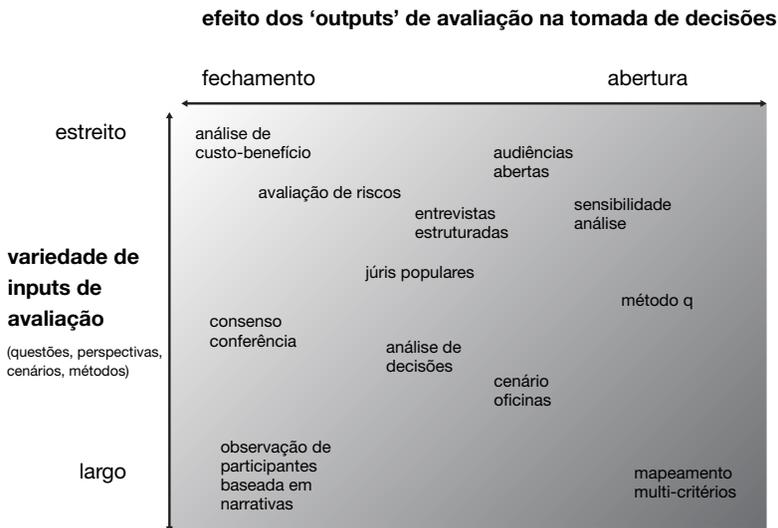


Figura 1. Características dos métodos de avaliação.

Fonte: Stirling et al., 2007, p. 57.



Figura 2. Diferença entre ‘ampliação’ da variedade de *inputs* utilizados em indicadores (esquerda) e ‘abertura’ das tomadas de decisão (direita).

A avaliação cientométrica convencional é bastante estreita; tanto na amplitude de *inputs* quanto na abertura de *outputs* (como ilustrado na Figura 2). Assim como ocorre com a análise de custo-benefício, essa limitação resulta da medição de desempenho em apenas uma ou duas dimensões (por exemplo, produção e eficiência, ou quantidade de publicações e citações) e do foco desproporcional em seleções artificialmente singulares de escolhas metodológicas identificadas como ‘melhores possíveis’ com as quais lidar com dados empíricos (como rotinas de normalização ou procedimentos de agregação) – mesmo onde alternativas igualmente razoáveis proporcionam classificações de *outputs* díspares.

Algumas das ferramentas analíticas dos indicadores de C&T podem ser relativamente amplas, em termos da variedade de *inputs*. Por exemplo, a classificação de universidades de Xangai leva em conta seis diferentes *inputs*, e o Painel Europeu de Avaliação da Inovação inclui um total de 25 indicadores. Entretanto, ambas as ferramentas geram um índice composto que utiliza ponderações simples para agregar múltiplas dimensões em um único escalar. Elas são amplas em *inputs*, mas estreitas em *outputs* (como ilustrado na Figura 2 à esquerda). Tais pontos escalares ‘limitam’ os debates sobre desempenho ao estabelecer, de forma unívoca, qual universidade é a ‘melhor’, ou qual país é o ‘mais’ inovador. Esses indicadores compostos já se mostraram potencialmente enganosos, já que ‘é grande o escopo para a manipulação de painéis através de seleção, ponderação e agregação’ (GRUPP; MOGGEE, 2004).

Uma maneira óbvia de lidar com dimensões plurais de *inputs* é a utilização de representações multidimensionais, como gráficos de ‘aranha’ (GRUPP; SCHUBERT, 2010) – preferencialmente após a redução de dimensões com embasamento conceitual e matemático. Porém, na análise cientométrica (e mais ainda, na bibliométrica), a variedade de *inputs* de uma dada propriedade (impacto de produtividade ou citações) é normalmente limitada pela natureza das fontes de dados. Em casos assim, podem os estudos quantitativos capturar e transmitir resultados diversos sob diferentes suposições analíticas? Nossa resposta é sim. Mesmo quando as fontes dos dados são relativamente limitadas, ainda há escopo para abertura (Figura 2 à direita). Mesmo com *inputs* limitados, é possível desenvolver ferramentas que auxiliem os tomadores de decisões a investigar como diferentes conceitualizações e operacionalizações matemáticas podem produzir resultados contrastantes (mesmo com exatamente os mesmos dados). Ao investigar a forma com que diferentes suposições levaram a diferentes métodos e classificações, o analista poderá oferecer recomendações ‘plurais e condicionais’ – e os decisores políticos poderão ser mais reflexivos e explícitos sobre os aspectos normativos de suas escolhas.

ABRINDO MEDIDAS DE INTERDISCIPLINARIDADE E DESEMPENHO

Aqui, exploraremos e ilustraremos o processo de ‘abertura’ através da análise de uma comparação bibliométrica recente de desempenho e interdisciplinaridade em seis organizações acadêmicas (RAFOLS et al., 2012). Tanto o ‘desempenho’ quanto a ‘interdisciplinaridade’ são conceitos complexos que podem ser apenas parcialmente capturados por indicadores bibliométricos. Os indicadores em questão derivaram de apenas duas fontes de dados: atributos genéricos de periódicos e as referências contidas em cada publicação.²

- 2 Esses dados são tratados com o uso de informações contextuais complementares, como a classificação dos periódicos em categorias de disciplina (matéria), e os padrões gerais de citações através dos periódicos em toda a Web of Science.

Ainda assim, apesar da limitação de *inputs*, mostramos que é possível conceber diferentes conceitualizações de interdisciplinaridade e desempenho, e produzir múltiplas operacionalizações de algumas delas.

A Figura 3 mostra duas conceitualizações de interdisciplinaridades utilizando mapas de sobreposição de la ciência (RAFOLS et al., 2010). Por um lado, podemos entender a interdisciplinaridade como *diversidade* disciplinar. Assim, as medidas da diversidade da distribuição das publicações (ou referências) de uma unidade através de diferentes categorias temáticas (como ilustrado pela propagação de nós no mapa da ciência) capturam o grau em que uma unidade cobre diferentes abordagens disciplinares. Por outro lado, podemos conceitualizar a interdisciplinaridade como o grau de *coerência* nas redes de categorias das publicações. O objetivo disso é capturar o grau de fertilização cruzada entre as disciplinas, que seria demonstrado pela medida em que as referências das publicações atravessam o mapa da ciência (como ilustrado pelas linhas verdes, que mostram casos de citações cruzadas cinco vezes ou mais do que esperado). Na análise, observou-se que a unidade mais interdisciplinar quanto à diversidade não foi a mais coerente – por isso, há um bom motivo para se diferenciar essas conceitualizações. Ainda assim, uma comparação entre três unidades de Estudos de Inovação (EI) e três unidades de Negócios e Administração (NA) mostra que, sob qualquer uma das várias conceitualizações e operacionalizações, as unidades de EI se mostraram mais interdisciplinares que as de NA. Assim, nesta escala maior, a contribuição do empenho de abertura foi proporcionar evidências mais robustas da diferença entre EI e NA.

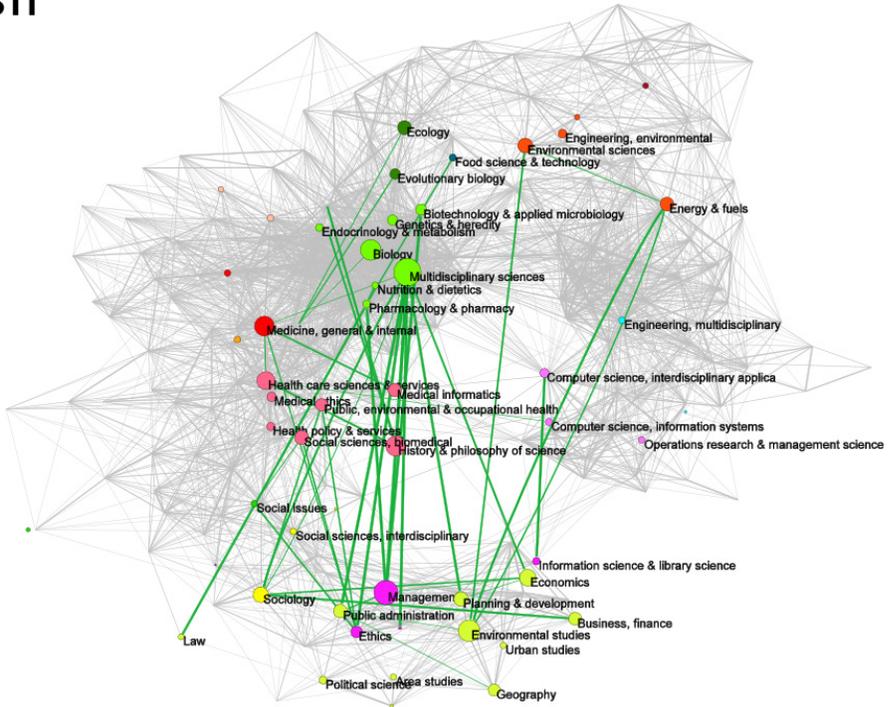


Figura 3. Sobreposição do número de referências nas categorias da Web of Science (fonte) pelo ISSTI (Institute for the Studies of Science Technology and Innovation – Instituto de Estudos da Tecnologia e Inovação da Ciência, Universidade de Edimburgo) no mapa global da ciência. Cada nó representa uma subdisciplina (Categoria Temática), e o tamanho do nó representa o número de referências. As ligações verdes indicam a ocorrência de referências (ou citações) cinco vezes ou mais do que esperado entre Categorias Temáticas, pelo ISSTI. As linhas cinzas indicam um certo nível de similaridade entre as Categorias Temáticas. O grau de sobreposição no fundo cinza ilustra o grau de similaridade entre diferentes áreas da ciência por todos os dados da Web of Science de 2009. A diversidade de referências (como refletido na propagação de nós pelo mapa) e as citações entre diferentes Categorias Temáticas (a quantidade de ligações cruzadas) são interpretadas como sinais de interdisciplinaridade.

Fonte: Ràfols et al., 2012.

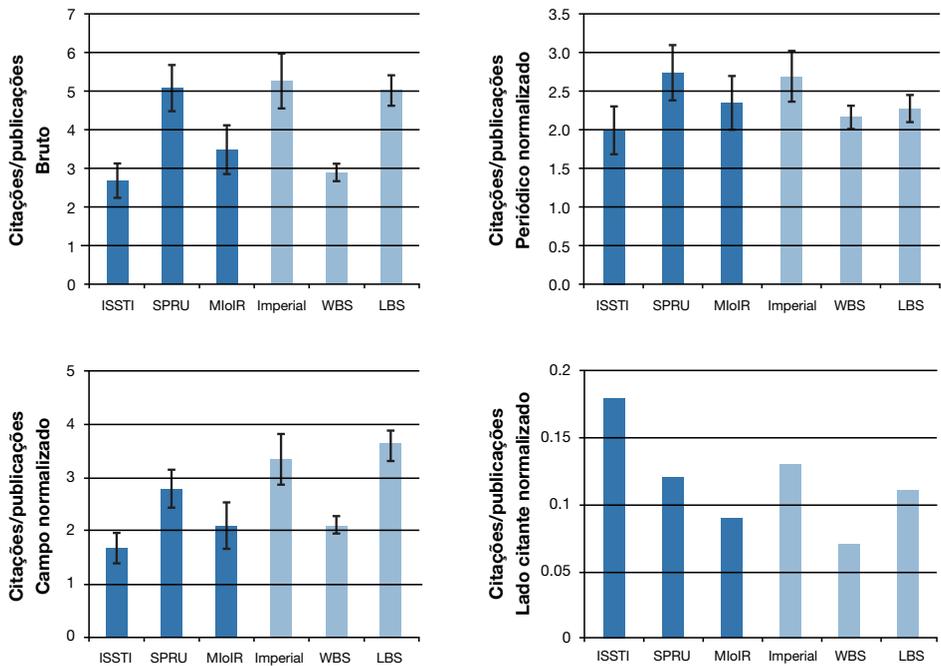


Figura 4. Exemplo de abertura usando diferentes normalizações em uma medida do número médio de citações por publicação em uma dada organização.

Fonte: Ràfols et al., 2012.

CONCLUSÕES E IMPLICAÇÕES DAS POLÍTICAS

O objetivo deste artigo é ilustrar que mesmo ferramentas analíticas limitadas e aparentemente rígidas, como os indicadores cientométricos, deixam espaço para uso de política mais explícita em relação à dependência dos *outputs* analíticos de pressupostos normativos. Argumentamos que essa ‘abertura’ é distinta (e complementar) à ‘ampliação’ da variedade de *inputs* de dados.

Os indicadores de políticas e gestão da C&T (assim como em outras esferas sociais) não apenas se tornaram conhecidos como ferramentas de medição, como também constituem óbvias ‘tecnologias de governança’ (DAVIS; KINGSBURY; MERRY, 2012). Os indicadores têm um papel performativo, incentivando e assim ‘orientando’ os cientistas para uma compreensão

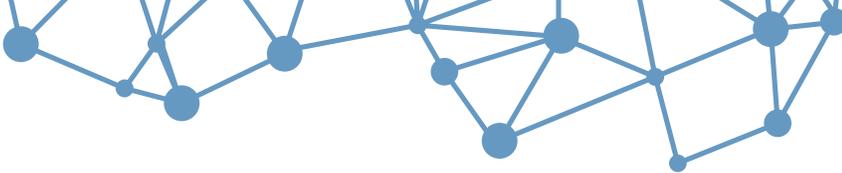
específica do ‘bom’ desempenho. ‘Medidas estatísticas tendem a substituir o debate político pelo conhecimento técnico’ (MERRY, 2011). Sob essas circunstâncias, torna-se imperativo que hajam debates mais abertos envolvendo as escolhas normativas cruciais subjacentes aos indicadores (BARRÉ, 2010). Em suma, são necessárias formas mais amplas e plurais de indicadores de C&T e de ferramentas de visualização, a fim de facilitar a ‘abertura’ de uma avaliação política mais rigorosa e confiável.

REFERÊNCIAS

- BARRÉ, R. Towards socially robust S&T indicators: indicators as debatable devices, enabling collective learning. *Research Evaluation*, v. 19, n. 3, p. 227-231, 2010. doi: 10.3152/095820210X512069.
- BÖRNER, K. *Atlas of Science: visualizing what we know*. Cambridge, MA: MIT Press, 2010. 272 p.
- DAVIS, K. E.; KINGSBURY, B.; MERRY, S. E. Indicators as a technology of global governance. *Law and Society Review*, v. 46, n. 1, p. 71-104, 2012. doi: 10.1111/j.1540-5893.2012.00473.x.
- GODIN, B. Science, accounting and statistics: the input-output framework. *Research Policy*, v. 36, n. 9, p. 1388-1403, 2007. doi: 10.1016/j.respol.2007.06.002.
- GODIN, B. The emergence of S&T indicators: why did governments supplement statistics with indicators? *Research Policy*, v. 32, n. 4, p. 679-691, 2003. doi: 10.1016/S0048-7333(02)00032-X.
- GRUPP, H.; MOGEE, M. E. Indicators for national science and technology policy: how robust are composite indicators? *Research Policy*, v. 33, n. 9, p. 1373-1384, 2004. doi: 10.1016/j.respol.2004.09.007.
- GRUPP, H.; SCHUBERT, T. Review and new evidence on composite innovation indicators for evaluating national performance. *Research Policy*, v. 39, n. 1, p. 67-78, 2010. doi: 10.1016/j.respol.2009.10.002.

- LEACH, M.; SCOONES, I.; STIRLING, A. **Dynamic sustainabilities: technology, environment, social justice**. London: Earthscan, 2010. 224 p. doi: 10.4324/9781849775069.
- LEPORI, B. Methodologies for the analysis of research funding and expenditure: from input to positioning indicators. **Research Evaluation**, v. 15, n. 2, p. 133-143, 2006. doi: 10.3152/147154406781775995.
- MARTIN, B. R.; IRVINE, J. Assessing basic research: some partial indicators of scientific progress in radio astronomy. **Research Policy**, v. 12, n. 2, p. 61-90, 1983. doi: 10.1016/0048-7333(83)90005-7.
- MARTIN, B. R.; TANG, P.; MORGAN, M.; GLANZEL, W.; HORNBOSTEL, S.; LAUER, G.; LENCLUD, G.; LIMA, L.; OPPENHEIM, C.; VAN DEN BESSELAAR, P.; ZIC-FUCHS, M. **Towards a bibliometric database for the social sciences and humanities – a European scoping project**. Brighton, UK: SPRU, 2010. 55 p.
- MERRY, S. E. Measuring the World: indicators, human rights, and global governance. **Current Anthropology**, v. 52, n. 3, p. S83-S95, 2011. Supplement.
- NARIN, F. Bibliometric techniques in the evaluation of research programs. **Science and Public Policy**, v. 14, n. 2, p. 99-106, 1987. doi: 10.1093/spp/14.2.99.
- PRIEM, J.; TARABORELLI, D.; GROTH, P.; NEYLON, C. **Altmetrics: a manifesto**. 2010. Disponível em: <<http://altmetrics.org/manifesto>>. Acesso em: 28 fev. 2017.
- RAFOLS, I.; PORTER, A. L.; LEYDESDORFF, L. Science overlay maps: a new tool for research policy and library management. **Journal of the Association for Information Science and Technology**, v. 61, n. 9, p. 1871-1887, 2010. doi: 10.1002/asi.21368.
- RAFOLS, I.; LEYDESDORFF, L.; O'HARE, A.; NIGHTINGALE, P.; STIRLING, A. How journal rankings can suppress interdisciplinary research: a comparison of Innovation Studies and Business & Management. **Research Policy**, v. 41, n. 7, p. 1262-1282, 2012. doi: 10.1016/j.respol.2012.03.015.
- ROSVALL, M.; BERGSTROM, C. T. Maps of random walks on complex networks reveal community structure. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 4, p. 1118-1123, 2008. doi: 10.1073/pnas.0706851105.
- ROESSNER, D. Quantitative and qualitative methods and measures in the evaluation of research. **Research Evaluation**, v. 9, n. 2, p. 125-132, 2000. doi: 10.3152/147154400781777296.

- STIRLING, A. Opening up or closing down? Analysis, participation and power in the social appraisal of technology. In: LEACH, M.; SCOONES, I.; WYNNE, B. (Eds.). **Science and citizens: globalization and the challenge of engagement**. London: Zed Books, 2005. p. 218-231.
- STIRLING, A. "Opening Up" and "Closing Down": power, participation, and pluralism in the social appraisal of technology. **Science, Technology, and Human Values**, v. 33, n. 2, p. 262-294, 2008. doi: 10.1177/0162243907311265.
- STIRLING, A. Risk, uncertainty and precaution: some instrumental implications from the social sciences. In: BERKHOUT, F.; LEACH, M.; SCOONES, I. (Eds.). **Negotiating environmental change: new perspectives from social science**. Cheltenham: Edward Elgar, 2003. p. 33-76.
- STIRLING, A.; LEACH, M.; MEHTA, L.; SCOONES, I.; SMITH, A.; STAGL, S.; THOMPSON, J. **Empowering designs: steps towards more progressive social appraisal of sustainability**. Brighton: STEPS Centre, 2007. 72 p.
- VAN RAAN, A. F. J. Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric methods. **Scientometrics**, v. 62, n. 1, p. 133-143, 2005. doi:10.1007/s11192-005-0008-6.
- WEINGART, P. Impact of bibliometrics upon the science system: inadvertent consequences? **Scientometrics**, v. 62, n. 1, p. 117-131, 2005. doi: 10.1007/s11192-005-0007-7.



TOWARDS INDICATORS FOR ‘OPENING UP’ SCIENCE AND TECHNOLOGY POLICY¹

RUMO A INDICADORES PARA ‘ABERTURA’ DE POLÍTICAS DE CIÊNCIA E TECNOLOGIA [ver página 67]

[ir para o sumário]

Ismael Ràfols^{a,b}, Tommaso Ciarli^b e Andy Stirling^b

INTRODUCTION

Recent years have seen much critical debate over the simplistic use of scientometric tools for formal or informal appraisal of science and technology (S&T) organisations (e.g. in university rankings) or individuals (e.g. the h-index) (ROESSNER, 2000; VAN RAAN, 2005; WEINGART, 2005). As a reaction to these critiques, efforts have been made to improve the robustness of measurements by broadening the range of inputs considered in scientometric evaluations. Examples include the inclusion of books and national or regional journals (MARTIN et al., 2010), or more recently ‘altmetrics’ (i.e. metrics based on alternative data sources, see Priem et al., 2010). In doing so, the S&T indicator and policy communities have reverted to an early conventional wisdom that scientometrics should rely on multiple sources of data that may provide ‘converging partial indicators’ (MARTIN; IRVINE, 1983).

While this ‘broadening out’ of the range of data used as ‘inputs’ in scientometric appraisal is, in our view, commendable (STIRLING, 2003), we propose

1 This article was first published online in 2012, as part of the proceedings of the 17th International Conference on Science and Technology Indicators (STI), held at Montréal.

^a *Ingenio* (CSIC-UPV), Universitat Politècnica de València, València, Espanha, i.rafols@ingenio.upv.es

^b SPRU (Science Policy Research Unit), University of Sussex, Brighton, Reino Unido

in this paper that a second dimension also needs to be considered. This relates to the extent to which the ‘outputs’ of appraisal ‘open up’ contrasting conceptualisations of the phenomena under scrutiny and consequently allow for more considered and rigorous attention to alternative policy options, both by decision makers and within wider policy debate (STIRLING, 2005; STIRLING et al., 2007; LEACH; SCOONES; STIRLING, 2010). We use a recent comparative study on the performance and interdisciplinarity of six organisational units (RAFOLS et al., 2012) to illustrate the difference between increasing the range of inputs (‘broadening out’) and enhancing the diversity of outputs to policy decision making (‘opening out’). In this way, policy appraisal can inform decision making in a more rigorous ‘plural and conditional’ fashion – acknowledging the way in which divergent normative assumptions and metrics can yield contrasting understandings of both the phenomena under scrutiny, and of appropriate policy responses (STIRLING, 2008).

CONCEPTUAL FRAMEWORK: ‘OPENING UP’ VERSUS ‘BROADENING OUT’ IN POLICY APPRAISAL

Many S&T indicators have been developed over the past 50 years as means to reveal the ‘strengths’ and ‘weaknesses’ of a given country’s ‘capacity’ and ‘performance’ in science and technology (GODIN, 2003). Developments by the OECD and US National Science Board (NSB), were derived from ‘a pure accounting framework based on the anticipated economic benefits of science’ (GODIN, 2007) and hence with a tendency to take an essentialist understanding of scientific excellence and production, influenced by economic concepts such as ‘efficiency’ and ‘effectiveness’ (NARIN, 1987). Initial scientometric studies were careful to declare methodological limitations, for example stating explicitly that citations were proxies and ‘partial and imperfect’ measures of impact rather than quality (MARTIN; IRVINE, 1983). But whether cautious or not, the emphasis of scientometric studies has traditionally lain in producing a ‘good’ measure of a given

concept such as ‘scientific excellence’, rather than in providing contrasting perspectives on what the meaning of ‘excellence’ is.

In recent years, various parallel developments have begun to challenge this scientometric *status-quo*. First, the pervasive diffusion of simplistic (and very possibly damaging) scientometric measures such as the h-index at various levels of management has renewed the debate over abuse and misuse of indicators (WEINGART, 2005). Second, traditional scientometrics is challenged by alternative data sources, like databases from hitherto excluded countries (e.g. Brazil’s Scielo), and new web-based indicators such as publication download frequency or popularity in 2.0 websites like *academia.eu* (PRIEM et al., 2010). Third, new tools have emerged for data visualisation (e.g. Hans Rosling’s *Gapminder*), for large network analysis (e.g. ROSVALL; BERGSTROM, 2008) and, for science mapping (BÖRNER, 2010), which are radically easing the presentation of complex multidimensional quantitative information to non-experts.

Each of these trends is pushing S&T policy towards use of indicators based on more diverse data inputs. These broader portfolios of inputs can, in principle, make scientometric analyses more robust. However, we contend here that this improved ‘breadth’ of inputs need not necessarily translate into a more plural and conditional policy process. ‘Opening up’ is not just about ‘more’ indicators, nor is it only a matter of ‘positioning’ or contextualising (LEPORI, 2006). It is about the design and use of indicators aimed explicitly at providing plural policy understandings and options. For S&T policy to be ‘opened up’, indicators used in appraisal need to be re-conceived as ‘debatable devices, enabling collective learning’ (BARRÉ, 2010).

In this way, we distinguish two dimensions in any process of policy appraisal, as illustrated in Figure 1. The first dimension, ‘breadth’ refers to the depth, extent and scope with which appraisal includes different types of knowledge that can describe the phenomena under scrutiny (LEACH; SCOONES; STIRLING, 2010). The second dimension, ‘openness’, refers to the degree to which the outputs of appraisal provide plural and conditional interpretations of the phenomena – and thus allow contrasting policy

options to be rigorously debated. Unlike analytical tools that ‘close down’ appraisal by establishing an absolute ranking of ‘best’ choices, ‘opening up’ tools allow decision-makers to contrast how under different assumptions the analysis may result in different rankings of options.

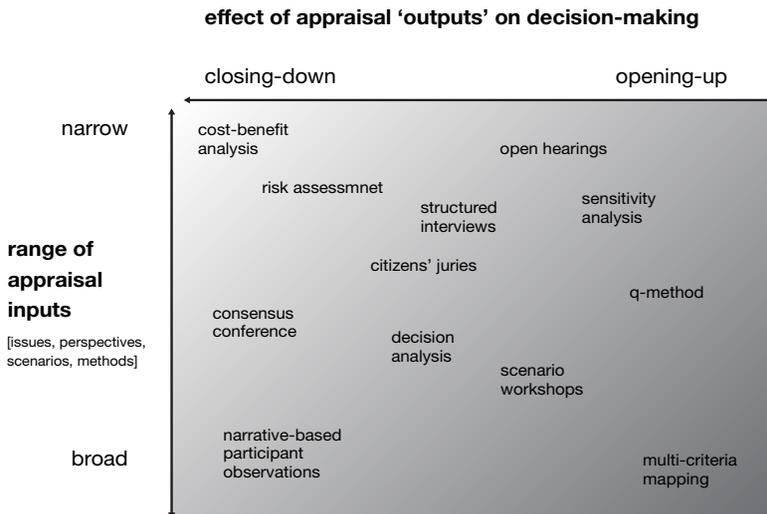


Figure 1. Characteristics of appraisal methods.
Source: Stirling et al., 2007, p. 57.

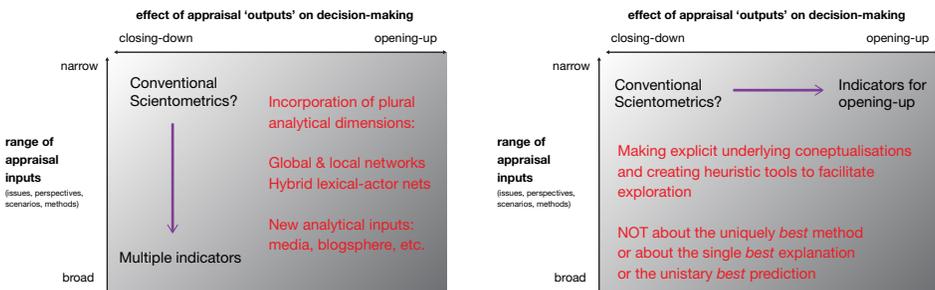


Figure 2. Difference between ‘broadening out’ the range of inputs used in indicators (left) and ‘opening up’ decision making (right).

Conventional scientometric appraisal is rather narrow: both in the breadth of inputs and the openness of outputs (as illustrated in Figure 2). As with cost-benefit analysis, this narrowness results from measuring performance only in one or two dimensions (e.g. production and efficiency, or number of publications and citations) and focusing disproportionately on artificially singular selections of allegedly ‘best possible’ methodological choices with which to handle empirical data (like normalisation routines or aggregation procedures) – even where equally reasonable alternatives yield disparate output rankings.

Some of the analytical tools in S&T indicators can be relatively broad in terms of the range of inputs. For example, the Shanghai ranking of universities takes into account six different inputs, and the European Innovation Scoreboard includes a total of 25 indicators. However, both tools create a composite index that uses simple weightings to aggregate multiple dimensions into a single scalar. These are broad in inputs but narrow in outputs (as illustrated in the left side of Figure 2). Such scalar scores ‘close down’ debates on performance by univocally establishing which university is ‘best’ or which country is ‘most’ innovative. Such composite indicators have been shown to be potentially misleading as ‘the scope for manipulation of scoreboards by selection, weighing and aggregation is great’ (GRUPP; MOGGEE, 2004).

An obvious way to handle plural input dimensions is to use multidimensional representations, such as ‘spider’ charts (GRUPP; SCHUBERT, 2010) – preferably after conceptually and mathematically grounded reduction of dimensions. But in scientometric (and even more so, in bibliometric) analysis, the range of inputs on a given property (productivity or citation impact) is often limited by the nature of data sources. In such cases, can quantitative studies capture and convey diverse outcomes under different analytical assumptions? Our answer is yes. Even when data sources are relatively narrow, there is still scope for opening up (on the right hand side of Figure 2). Even with narrow inputs, tools can be developed that help decision makers scrutinize how different conceptualisations and associated mathematical operationalisations may yield contrasting results (even of exactly the same data). By investigating

how different assumptions lead to different methods and rankings, the analyst can provide ‘plural and conditional’ advice – and policy makers can be more reflective and explicit about the normative aspects of their choices.

OPENING UP MEASURES OF INTERDISCIPLINARITY AND PERFORMANCE

Here we will explore and illustrate the process of ‘opening up’, by reviewing a recent bibliometric comparison of performance and interdisciplinarity in six academic organisations (RAFOLS et al., 2012). Both ‘performance’ and ‘interdisciplinarity’ are complex concepts that can only partially be captured by bibliometric indicators. Indicators in question were derived from only two data sources: generic journal attributes and the references contained in each publication.² Yet in spite of this narrowness of inputs, we show it is possible to conceive of different conceptualisations of interdisciplinarity and performance, and make multiple operationalisations of some of them.

Two conceptualisations of interdisciplinarity are shown in Figure 3 using overlay maps of science (RAFOLS et al., 2010). On the one hand, we can understand interdisciplinarity as disciplinary *diversity*. Thus diversity measures of the distribution of publications (or references) of a unit across disparate subject categories (as illustrated by the spread of nodes over the map of science) captures the degree to which a unit covers different disciplinary approaches. On the other hand, we can conceptualise interdisciplinarity as the degree of *coherence* in their network of categories where they publish. This aims to capture the degree of cross-fertilisation between disciplines, which would be shown by the extent to which the references of publications criss-cross the map of science (as illustrated by the green lines, which show cases of cross-citation 5-fold above expectation). In the analysis it was found that the most interdisciplinary unit in terms of diversity was not the

2 These data are treated using complementary contextual information such as the classification of journals into disciplinary subject category, and the overall citation patterns across journals in all the web of science.

most coherent – hence there is good reason to differentiate these conceptualisations. Nevertheless, a comparison between three Innovation Studies (IS) units and three Business and Management units (BM) units showed that under any of the various conceptualisations and operationalisations IS units were more interdisciplinary than BM units. Thus, at this larger scale, the contribution of the opening-up effort was to provide more robust evidence of the difference between IS and BM.

ISSTI

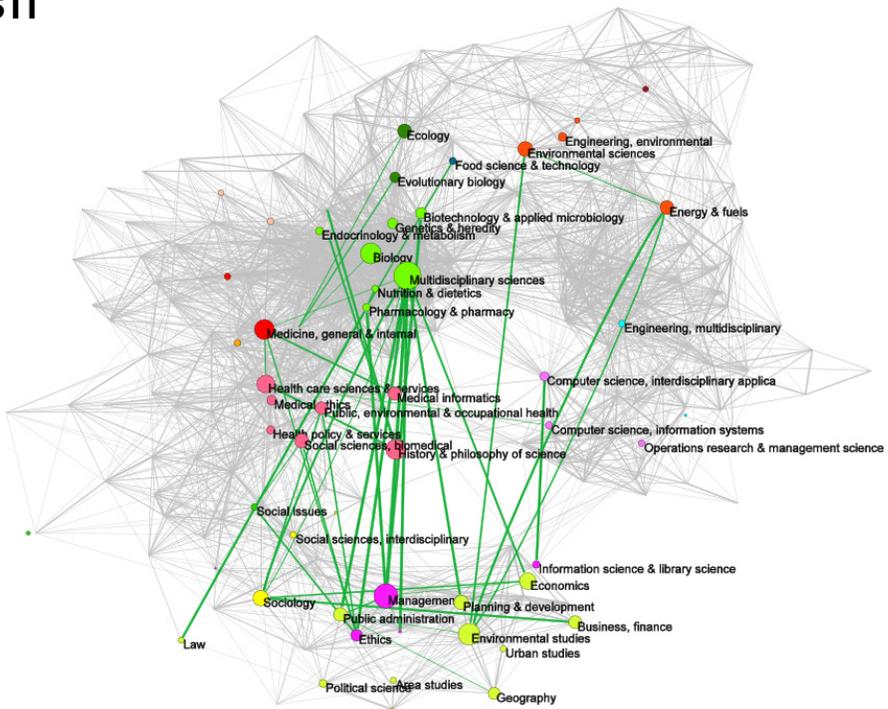


Figure 3. Overlay of number of references on Web of Science Categories (source) by of the Institute for the Studies of Science Technology and Innovation (ISSTI, University of Edinburgh) on the global map of science. Each node represents a sub-discipline (Subject Category), and node size the number of references. Green links indicate 5-fold above expectation referencing (or citing) between Subject Categories by ISSTI. Grey lines indicate a certain level of similarity between Subject Categories. The degree of superposition in the grey background illustrates the degree of similarity between different areas of science for

all 2009 Web of Science data. Diversity of references (as reflected in the spread of nodes over map) and referencing across disparate Subject Categories (the amount of cross-linking) are interpreted as signs of interdisciplinarity.

Source: Ràfols et al., 2012.

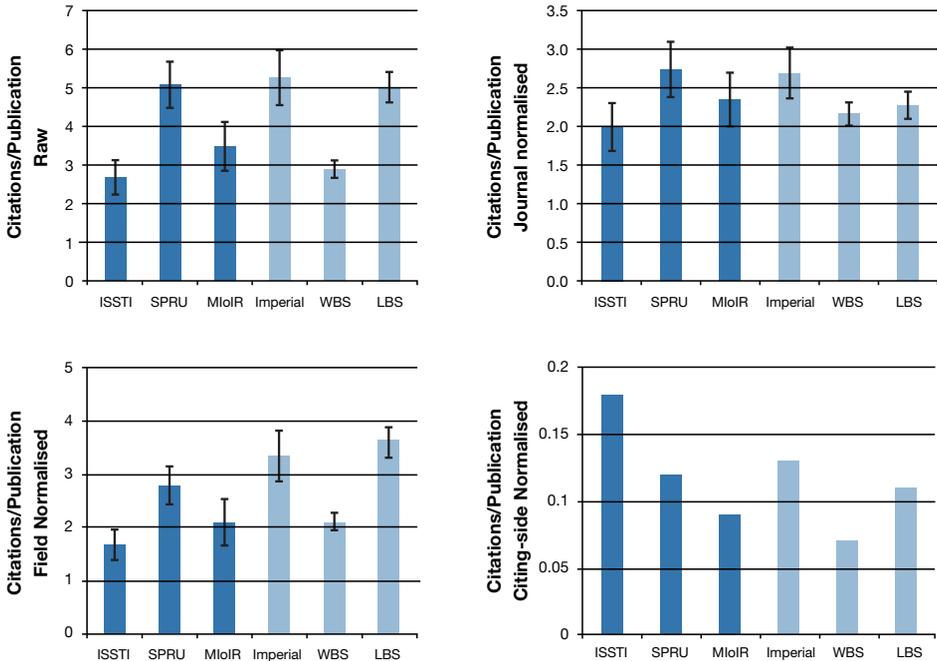


Figure 4. Example of opening-up by using different normalisations to a measure of the average number of citations per publication in a given organisation.

Source: Ràfols et al., 2012.

CONCLUSIONS AND POLICY IMPLICATIONS

This paper aims to illustrate that even analytical tools as narrow and apparently rigid as scientometric indicators leave room for policy usage that is more explicit about the dependence of analytic outputs on normative

assumptions. We have argued that this ‘opening up’ is distinct (and complementary) to the ‘broadening out’ of the range of data inputs.

Indicators in S&T policy and management (as well as in other social spheres) have not only become pervasive as measurement tools, but constitute obvious ‘technologies for governance’ (DAVIS; KINGSBURY; MERRY, 2012). Indicators play a performative role, incentivising and thus ‘guiding’ scientists towards particular understandings of ‘good’ performance. ‘Statistical measures tend to replace political debate with technical expertise’ (MERRY, 2011). Under these circumstances, it becomes imperative to bring out into more open debate the crucial normative choices underlying indicators (BARRÉ, 2010). In short, both broader and more plural forms of S&T indicators and visualisation tools are needed, in order to facilitate the ‘opening up’ of more rigorous and accountable policy appraisal.

REFERENCES

- BARRÉ, R. Towards socially robust S&T indicators: indicators as debatable devices, enabling collective learning. *Research Evaluation*, v. 19, n. 3, p. 227-231, 2010. doi: 10.3152/095820210X512069.
- BÖRNER, K. *Atlas of Science: visualizing what we know*. Cambridge, MA: MIT Press, 2010. 272 p.
- DAVIS, K. E.; KINGSBURY, B.; MERRY, S. E. Indicators as a technology of global governance. *Law and Society Review*, v. 46, n. 1, p. 71-104, 2012. doi: 10.1111/j.1540-5893.2012.00473.x.
- GODIN, B. Science, accounting and statistics: the input-output framework. *Research Policy*, v. 36, n. 9, p. 1388-1403, 2007. doi: 10.1016/j.respol.2007.06.002.
- GODIN, B. The emergence of S&T indicators: why did governments supplement statistics with indicators? *Research Policy*, v. 32, n. 4, p. 679-691, 2003. doi: 10.1016/S0048-7333(02)00032-X.

- GRUPP, H.; MOGEE, M. E. Indicators for national science and technology policy: how robust are composite indicators? *Research Policy*, v. 33, n. 9, p. 1373-1384, 2004. doi: 10.1016/j.respol.2004.09.007.
- GRUPP, H.; SCHUBERT, T. Review and new evidence on composite innovation indicators for evaluating national performance. *Research Policy*, v. 39, n. 1, p. 67-78, 2010. doi: 10.1016/j.respol.2009.10.002.
- LEACH, M.; SCOONES, I.; STIRLING, A. *Dynamic sustainabilities: technology, environment, social justice*. London: Earthscan, 2010. 224 p. doi: 10.4324/9781849775069.
- LEPORI, B. Methodologies for the analysis of research funding and expenditure: from input to positioning indicators. *Research Evaluation*, v. 15, n. 2, p. 133-143, 2006. doi: 10.3152/147154406781775995.
- MARTIN, B. R.; IRVINE, J. Assessing basic research: some partial indicators of scientific progress in radio astronomy. *Research Policy*, v. 12, n. 2, p. 61-90, 1983. doi: 10.1016/0048-7333(83)90005-7.
- MARTIN, B. R.; TANG, P.; MORGAN, M.; GLANZEL, W.; HORNBOSTEL, S.; LAUER, G.; LENCLUD, G.; LIMA, L.; OPPENHEIM, C.; VAN DEN BESSELAAR, P.; ZIC-FUCHS, M. *Towards a bibliometric database for the social sciences and humanities* – a European scoping project. Brighton, UK: SPRU, 2010. 55 p.
- MERRY, S. E. Measuring the World: indicators, human rights, and global governance. *Current Anthropology*, v. 52, n. 3, p. S83-S95, 2011. Supplement.
- NARIN, F. Bibliometric techniques in the evaluation of research programs. *Science and Public Policy*, v. 14, n. 2, p. 99-106, 1987. doi: 10.1093/spp/14.2.99.
- PRIEM, J.; TARABORELLI, D.; GROTH, P.; NEYLON, C. *Altmetrics: a manifesto*. 2010. Available from: <<http://altmetrics.org/manifesto>>. Viewed: 28 Feb. 2017.
- RAFOLS, I.; PORTER, A. L.; LEYDESDORFF, L. Science overlay maps: a new tool for research policy and library management. *Journal of the Association for Information Science and Technology*, v. 61, n. 9, p. 1871-1887, 2010. doi: 10.1002/asi.21368.
- RAFOLS, I.; LEYDESDORFF, L.; O'HARE, A.; NIGHTINGALE, P.; STIRLING, A. How journal rankings can suppress interdisciplinary research: a comparison of Innovation Studies and Business & Management. *Research Policy*, v. 41, n. 7, p. 1262-1282, 2012. doi: 10.1016/j.respol.2012.03.015.

- ROSVALL, M.; BERGSTROM, C. T. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*, v. 105, n. 4, p. 1118-1123, 2008. doi: 10.1073/pnas.0706851105.
- ROESSNER, D. Quantitative and qualitative methods and measures in the evaluation of research. *Research Evaluation*, v. 9, n. 2, p. 125-132, 2000. doi: 10.3152/147154400781777296.
- STIRLING, A. Opening up or closing down? Analysis, participation and power in the social appraisal of technology. In: LEACH, M.; SCOONES, I.; WYNNE, B. (Eds.). *Science and citizens: globalization and the challenge of engagement*. London: Zed Books, 2005. p. 218-231.
- STIRLING, A. "Opening Up" and "Closing Down": power, participation, and pluralism in the social appraisal of technology. *Science, Technology, and Human Values*, v. 33, n. 2, p. 262-294, 2008. doi: 10.1177/0162243907311265.
- STIRLING, A. Risk, uncertainty and precaution: some instrumental implications from the social sciences. In: BERKHOUT, F.; LEACH, M.; SCOONES, I. (Eds.). *Negotiating environmental change: new perspectives from social science*. Cheltenham: Edward Elgar, 2003. p. 33-76.
- STIRLING, A.; LEACH, M.; MEHTA, L.; SCOONES, I.; SMITH, A.; STAGL, S.; THOMPSON, J. *Empowering designs: steps towards more progressive social appraisal of sustainability*. Brighton: STEPS Centre, 2007. 72 p.
- VAN RAAN, A. F. J. Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, v. 62, n. 1, p. 133-143, 2005. doi:10.1007/s11192-005-0008-6.
- WEINGART, P. Impact of bibliometrics upon the science system: inadvertent consequences? *Scientometrics*, v. 62, n. 1, p. 117-131, 2005. doi: 10.1007/s11192-005-0007-7.





A PESQUISA BIBLIOMÉTRICA NA ERA DO *BIG DATA*: DESAFIOS E OPORTUNIDADES

BIBLIOMETRICS RESEARCH IN THE ERA OF BIG DATA: CHALLENGES AND OPPORTUNITIES

[see page 101] [go to summary]

*Dietmar Wolfram**

INTRODUÇÃO

O tema *big data* tem o potencial de afetar quaisquer disciplinas em que grandes conjuntos de dados são utilizados. Neste ensaio, farei uma reflexão sobre os desafios e oportunidades que o *big data* tem a oferecer à bibliometria, cientometria, e infometria (aqui referidas como “métricas”). Nos últimos anos, os pesquisadores de métricas começaram a lidar com a questão (por exemplo, MOED, 2012; ROUSSEAU, 2012).

Em primeiro lugar, como se define o termo *big data*? No momento, não há uma definição universalmente aceita. Trata-se simplesmente de tamanho, ou é também relacionado ao que investigadores podem fazer com os volumes massivos de dados de pesquisas hoje disponíveis? Uma crescente variedade de projetos de pesquisas das ciências físicas e biomédicas dependem de coletas de dados que teriam sido inconcebíveis há uma geração. Por exemplo, o Grande Colisor de Hádrons do CERN (GCH), que é empregado no estudo de física de partículas, já gerou cerca de 75 petabytes de dados durante o período de três anos (HOFFMAN, 2013). Isso requer uma

* School of Information Studies, University of Wisconsin-Milwaukee; dwolfram@uwm.edu

quantidade enorme de espaço de armazenamento, assim como métodos eficientes para acessar qualquer expressão desses dados. Além disso, iniciativas da *big science*, como o GCH, exigem a participação de grandes equipes de pesquisa, em alguns casos envolvendo milhares de participantes, e têm seus próprios desafios de gestão de projeto.

Na pesquisa de métricas, não trabalhamos com conjuntos de dados ou equipes de pesquisa tão grandes quanto os das disciplinas envolvidas na *big science*. Com isso em mente, o que significa o conceito de *big data* no contexto bibliométrico? Como em outras áreas científicas, o tamanho dos conjuntos de dados disponíveis utilizados na pesquisa de métricas também aumentou. Ao mesmo tempo, é esperado dos pesquisadores de métricas que empreguem grandes conjuntos de dados para aumentar a confiabilidade e a capacidade de generalização de suas descobertas.

Quando olhamos para os conjuntos de dados utilizados em alguns dos primeiros estudos métricos de conjuntos de dados bibliográficos e linguísticos, podemos considerá-los modestos pelos padrões de hoje. A coleta e análise manuais de dados tornou o processamento de grandes conjuntos de dados computacionalmente laborioso. Isso é evidente em muitos estudos bibliométricos clássicos:

- Alfred Lotka utilizou 6891 e 1325 autores em seus estudos de produtividade científica para o Chemical Abstracts e o Auerbach's Geschichtstafeln der Physik, respectivamente (LOTKA, 1926). Lotka limitou os dados do Chemical Abstracts a autores cujos sobrenomes começavam com as letras A e B.
- Samuel C. Bradford utilizou 326 periódicos com 1332 referências para seu estudo inicial sobre a concentração e difusão de literatura sobre um dado tema presente em diferentes periódicos (BRADFORD, 1934).
- George K. Zipf, que estudou regularidades no uso da linguagem, contou com conjuntos de dados que continham normalmente menos que 10.000 tipos de palavras. Seu comentário sobre o trabalho de um pesquisador anterior: “O volume total de quase 11 milhões de palavras corridas

de Kaeding até o momento excede uma amostra de tamanho ideal que nos é de pouca utilidade prática.” (ZIPF, 1949, Section 3.IV) reflete sua preocupação em ter dados em excesso.

Os tempos mudaram. Hoje, há muito mais dados disponíveis em forma legível por máquinas, o que simplifica a coleta de grandes conjuntos de dados. Os pesquisadores de métricas trabalham com conjuntos de dados medidos em milhões de pontos de dados. A habilidade computacional continua a melhorar, e novas ferramentas analíticas de análise numérica e de texto permitem que grandes conjuntos de dados sejam processados rapidamente e visualizados de forma que padrões de pequena e grande escalas possam ser descobertos.

O QUE É “GRANDE” EM BIBLIOMETRIA?

Considerando os tamanhos dos conjuntos de dados associados com o *big data*, muitas vezes medidos em terabytes e petabytes, pode-se qualificar como *big data* os conjuntos de dados coletados e analisados por pesquisadores de métricas? O tamanho do conjunto de dados é certamente um critério, mas outro aspecto importante que caracteriza o *big data* é relacionado às exigências de desempenho. Magoulas e Lorica (2009) alegam que dados se tornam *big data* “... quando as exigências de tamanho e desempenho para a gestão de dados se tornam fatores significantes no projeto e na decisão de se implementar um sistema de gestão e análise de dados” (p. 2). Isso descreve adequadamente o ambiente atual em que operam os pesquisadores de métricas. Idealmente, os pesquisadores gostariam de trabalhar com toda uma população de dados, e não apenas uma amostra. As bases de dados bibliográficos e de citações de hoje armazenam dezenas de milhões de registros, que por sua vez representam um subconjunto de todos os trabalhos publicados. Hoje é possível analisar e visualizar relações entre dezenas de milhões de documentos (vide, por exemplo, <http://www.mapofscience.com/>). Mesmo com a disponibilidade da população de todos os trabalhos publicados, essas fontes de bases de dados

podem ser consideradas como mais limitadas que conjuntos de dados coletados nas ciências físicas. Com o acesso adequado, é possível que pesquisadores de bibliometria processem o conteúdo integral de múltiplas bases de dados, em que as unidades de análise são autores, publicações, periódicos ou citações. Apesar do crescimento anual da quantidade de literatura científica, o volume total é ainda pelo menos uma ordem de grandeza menor do que as populações de dados coletadas em certas áreas das ciências biomédicas e naturais. Em termos de análise métrica dos produtos da comunicação acadêmica, os pesquisadores apenas podem processar o que foi publicado e o que está disponível. Isso não implica que o processamento de conjuntos de dados bibliométricos não tenha os mesmos tipos de desafios encontrados em outras disciplinas. O tamanho desses conjuntos de dados é ainda considerável, e pode aumentar drasticamente quando pesquisadores trabalham com textos integrais de documentos, e não com apenas substitutos de documentos compostos de referências bibliográficas ou citações. O Hathitrust Research Center, por exemplo, oferece acesso aos textos integrais de mais de três milhões de obras de domínio público (<https://analytics.hathitrust.org/>). Há ainda outras formas de discurso registrado com potencial de maiores volumes de dados que atraíram a atenção de pesquisadores de métricas nas últimas décadas. Dados originários da Internet para estudos métrico, sejam mídias sócias ou registros de buscas de usuários e padrões de navegação na Internet, proporcionam um potencial para conjuntos de dados ainda maiores. Nesses casos, a velocidade com que novos dados tornam-se disponíveis pode dificultar a coleta e análise da população de dados, de uma perspectiva computacional e legal. Esses dados também levantam questões éticas para a pesquisa métrica, em que técnicas de mineração de dados e de texto em grandes conjuntos de dados podem revelar informações pessoalmente identificáveis, como ficou evidente em 2006 com a divulgação de um conjunto de dados de registro de pesquisa pela AOL (NAKASHIMA, 2006).

DESAFIOS DO *BIG DATA* PARA A BIBLIOMETRIA

Diversos desafios continuados relativos ao acesso a dados e sua análise se apresentam aos pesquisadores de métricas. Primeiramente, como em muitas outras áreas de investigação acadêmica, a acessibilidade aos dados é uma barreira em potencial. Os dados desejados podem estar disponíveis apenas através de fornecedores que exigem assinaturas. Restrições de acesso em nível de usuário final podem limitar a quantidade de dados que podem ser baixados de uma só vez. A compra de conjuntos de dados completos é possível, mas os preços podem ser proibitivos. Cada vez mais, sites de mídias sociais que geram dados publicamente disponíveis estão oferecendo Interfaces de Programação de Aplicação (APIs) para acessar esses dados, mas elas podem limitar a quantidade de dados que podem ser baixados dentro de um dado período, ou negar acesso a dados pelos motivos de privacidade acima mencionados. Em geral, dados publicamente disponíveis são muitas vezes descentralizados, o que dificulta a identificação e coleta dos dados. Pelo lado positivo, um desafio mais antigo, que tornou-se mais brando nos últimos anos, é o aumento de velocidade de transferência de dados e a diminuição do custo para o armazenamento de *big data*.

Com o *big data*, tem havido a crença de que “mais é melhor”. No entanto, pode-se dizer que esse é sempre o caso? Mais dados é algo positivo, mas os dados são “bons” dados? A habilidade de acesso à população de dados permite uma representação completa. Porém, é possível que alguns tipos de dados ou dados de diferentes fontes exijam extensas limpezas e padronizações antes de se tornarem utilizáveis para análise. Isso é especialmente verdadeiro para dados baseados na Internet, por exemplo, em que registros de servidores de rede muitas vezes abrangem conteúdo supérfluo que não é relacionado diretamente a ações iniciadas por usuários (HAN; WOLFRAM, 2016). Além do mais, um pesquisador de métricas notificou que “... a computação de força bruta com o *big data* pode levar a falsas descobertas e correlações espúrias...” (PRATHAP, 2014). Independente da completude do conjunto de dados, as conclusões tiradas acerca do fenômeno de interesse,

aos pesquisadores de métricas, podem aplicar-se apenas às amostras utilizadas. Um exemplo disso é que as características gerais dos conjuntos de dados podem mudar com as diferentes quantidades de dados utilizados. Distribuições de frequência e estatísticas descritivas ligadas a conjuntos de dados de métricas podem variar imensamente com o tamanho do conjunto de dados, enquanto outros aspectos dos dados podem permanecer relativamente estáveis. (AJIFERUKE; WOLFRAM; FAMOYE, 2006).

Os grandes conjuntos de dados e as relações complexas entre os dados levam a uma representação de dados de altas dimensões, que por sua vez está vinculada a uma crescente sobrecarga computacional. Um exemplo disso é que a representação de dados bibliométricos de interesse (como autores, publicações e periódicos) como espaços vetoriais, tal qual utilizada na pesquisa de recuperação de informação, torna possível acessar as relações entre as entidades de interesse. Infelizmente, armazenar dados nessa forma pode resultar em matrizes compostas de milhões de linhas e colunas. Técnicas de redução de dimensionalidade se fazem necessárias para diminuir a sobrecarga de processamento, enquanto preservando a essência das relações obsevasdas. Técnicas estatísticas, como a análise de agrupamento e a análise fatorial para dados quantitativos, juntamente com métodos baseados em linguagem, empregadas onde apropriadas, podem revelar essas relações ou simplificar a representação de dados.

A aplicação de técnicas apropriadas de análise e resumo de dados apresenta ainda outro desafio em potencial. Estão disponíveis uma variedade de técnicas de análise de redes e ligações, assim como de mineração de texto e de dados, para oferecer vislumbres de relações evidentes ou ocultas dentro de conjuntos de dados. De forma parecida, as técnicas estatísticas exploratórias que resumem e visualizam conjuntos de dados permitem que pesquisadores identifiquem padrões que seriam, sob outras circunstâncias, ocultos, devido ao volume de dados. O que os pesquisadores descobriram com essas diferentes ferramentas e técnicas é que elas podem produzir diferentes resultados mesmo a partir dos mesmos dados. Quais técnicas são melhores e proporcionam a maior validade? Não há apenas uma forma correta de

analisar ou visualizar os dados, então os pesquisadores devem justificar seus casos de forma convincente quando apresentam e interpretam resultados.

OPORTUNIDADES PARA A BIBLIOMETRIA

Apesar dos crescentes volumes de dados e opções para a análise e apresentação de dados oferecerem desafios para a identificação das melhores abordagens, essas técnicas e avanços da computação fazem com que seja possível processar grandes volumes de dados de formas que não eram possíveis até mesmo há poucos anos.

A pesquisa bibliométrica tradicionalmente enfocou os aspectos bibliográficos e baseados em citações do discurso registrado. A aplicação de técnicas de análise de redes, junto aos programas relacionados, tornou possível estudar grandes conjuntos de dados através de múltiplas disciplinas. Abordagens de mineração de dados (THELWALL, 2001), mineração de texto (SONG; CHAMBERS, 2014), métodos empregados para agrupamento e classificação (GLÄNZEL; SCHUBERT, 2003) e métodos para a detecção de comunidades (BOHLIN et al., 2014) podem identificar padrões e relações ocultos em dados. Programas de análise de redes publicamente disponíveis, como o Pajek (<http://mrvar.fdv.uni-lj.si/pajek/>) e o Gephi (<https://gephi.org/>), assim como programas desenvolvidos especificamente para suportar pesquisas métricas, como o CiteSpace (<http://cluster.cis.drexel.edu/~cchen/citespace/>), o Sci2 (<https://sci2.cns.iu.edu/user/index.php>) e o VOSviewer (<http://www.vosviewer.com/>) oferecem rotinas para o resumo e visualização de grandes conjuntos de dados.

A análise baseada em ligações com base em citações (citação direta, cocitação, acoplamento bibliográfico) ou coautoria tem sido um ponto focal para a pesquisa métrica durante décadas. Uma desvantagem relacionada ao uso desses tipos de dados é que se não há ligação, não há relação. Pesquisadores de diferentes áreas ou de diferentes partes do mundo que não estão cientes uns dos outros podem até compartilhar interesses em comum, mas

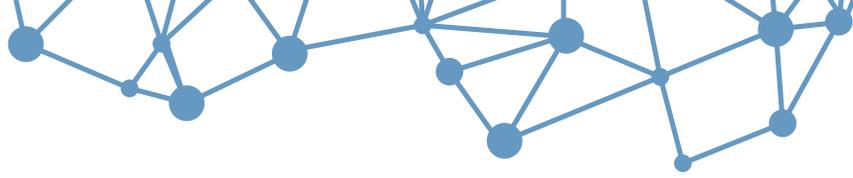
suas similaridades não seriam identificadas sem colaborações formais ou ligações explícitas baseadas em citações.

A análise baseada em linguagem apresenta outro caminho por onde se pode estudar as relações entre entidades de interesse. Os primeiros empenhos com foco em linguagem dependiam da análise de coocorrência de palavras em títulos ou palavras-chave do resumo, ou de cabeçalhos de assuntos (por exemplo, CALLON; COURTIAL; LAVILLE, 1991; COURTIAL, 1994). Os desenvolvimentos no processamento de linguagem natural (PLN), aprendizado de máquinas, mineração de texto e modelagem de tópicos agora permitem que o foco se volte a um corpus maior de textos integrais. Esses métodos podem escalar para acomodar conjuntos de dados maiores, mas em consequência disso, requerem conjuntos de dados grandes o suficiente para treinar modelos de forma efetiva, o que por sua vez apresentam resultados mais confiáveis. Apesar de haver uma alta sobrecarga computacional para o treinamento de modelos para algumas dessas técnicas, a vantagem para os pesquisadores é a redução em dimensionalidade resultante para a análise de dados. No caso de comparações baseadas em texto utilizando modelagem de tópicos (WALLACH, 2006), centenas de milhares de palavras-chave candidatas que são usadas para indexar publicações podem ser reduzidas a poucas centenas de assuntos para representar as publicações. As aplicações de métodos baseados em texto se estendem ao estudo do impacto acadêmico (SONG & DING, 2014) e a abordagens complementares à análise de citações para o estudo de similaridades de perfis de pesquisas de autores (LU; WOLFRAM, 2012). Lu e Wolfram, por exemplo, estudaram similaridades de pesquisas de autores com base em modelos de tópicos e observaram que produziram as relações mais coerentes entre autores quando os compararam a abordagens baseadas em citações. Métodos baseados em texto também podem ser combinados a métodos baseados em citações para firmar os pontos fortes de cada abordagem (GLENISSON et al., 2005).

REFERÊNCIAS

- AJIFERUKE, I.; WOLFRAM, D.; FAMOYE, F. Sample size and informetric model goodness-of-fit outcomes: A search engine log case study. **Journal of Information Science**, v. 32, n. 3, p. 212-222, 2006.
- BOHLIN, L.; EDLER, D.; LANCICHINETTI, A.; ROSVALL, M. Community detection and visualization of networks with the map equation framework. In: DING, Y.; ROUSSEAU, R.; WOLFRAM, D. (Eds.). **Measuring scholarly impact: methods and practice**. Cham: Springer International Publishing, 2014. p. 3-34. doi: 10.1007/978-3-319-10377-8.
- BRADFORD, S. C. Sources of information on specific subjects. **Engineering**, v. 137, n. 3550, p. 85-96, 1934.
- CALLON, M.; COURTIAL, J. P.; LAVILLE, F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. **Scientometrics**, v. 22, n. 1, p. 155-205, 1991. doi: 10.1007/BF02019280.
- COURTIAL, J. P. A cword analysis of scientometrics. **Scientometrics**, v. 31, n. 3, p. 251-260, 1994. doi: 10.1007/BF02016875.
- GLÄNZEL, W.; SCHUBERT, A. A new classification scheme of science fields and sub-fields designed for scientometric evaluation purposes. **Scientometrics**, v. 56, n. 3, p. 357-367, 2003. doi: 10.1023/A:1022378804087.
- GLENISSON, P.; GLÄNZEL, W.; JANSSENS, F.; DE MOOR, B. Combining full text and bibliometric information in mapping scientific disciplines. **Information Processing and Management**, v. 41, n. 6, p. 1548-1572, 2005. doi: 10.1016/j.ipm.2005.03.021.
- HAN, H.; WOLFRAM, D. An exploration of search session patterns in an image-based digital library. **Journal of Information Science**, v. 42, n. 4, p. 477-491, 2016. doi: 10.1177/0165551515598952.
- HOFFMAN, M. LHC collider recorded over 75 petabyte until today's shutdown. **Science World Report**, 14 Feb. 2013. Disponível em: <<http://www.scienceworldreport.com/articles/4973/20130214/lhc-collider-cern-datacenter-100-petabyte-physics-shutdown.htm>>. Acesso em: 28 fev. 2017.
- LOTKA, A. J. The frequency distribution of scientific productivity. **Journal of the Washington Academy of Science**, v. 16, n. 12, p. 317-323, 1926.

- LU, K.; WOLFRAM, D. Measuring author research relatedness: a comparison of word-based, topic-based, and author cocitation approaches. *Journal of the Association for Information Science and Technology*, v. 63, n. 10, p. 1973-1986, 2012. doi: 10.1002/asi.22628.
- MAGOULAS, R.; LORICA, B. Introduction to big data. *Release 2.0*, n. 11, p. 1-39, 2009.
- MOED, H. F. The use of big datasets in bibliometric research. *Research Trends*, n. 30, p.31-33, 2012.
- NAKASHIMA, E. AOL takes down site with users' search data. *The Washington Post*, 8 Aug. 2006. Disponível em: <<http://www.washingtonpost.com/wp-dyn/content/article/2006/08/07/AR2006080701150.html>>. Acesso em: 28 fev. 2017.
- PRATHAP, G. Big data and false discovery: analyses of bibliometric indicators from large data sets. *Scientometrics*, v. 98, n. 2, p. 1421-1422, 2014. doi: 10.1007/s11192-013-1063-z.
- ROUSSEAU, R. A view on big data and its relation to Informetrics. *Chinese Journal of Library and Information Science*, v. 5, n. 3, p. 12-26, 2012.
- SONG, M.; CHAMBERS, T. Text mining with the Stanford CoreNLP. In: DING, Y.; ROUSSEAU, R.; WOLFRAM, D. (Eds.). *Measuring scholarly impact: methods and practice*. Cham: Springer International Publishing, 2014. p. 215-234. doi: 10.1007/978-3-319-10377-8.
- SONG, M.; DING, Y. Topic modeling: Measuring scholarly impact using a topical lens In: DING, Y.; ROUSSEAU, R.; WOLFRAM, D. (Eds.). *Measuring scholarly impact: methods and practice*. Cham: Springer International Publishing, 2014. p. 235-257. doi: 10.1007/978-3-319-10377-8.
- THELWALL, M. A web crawler design for data mining. *Journal of Information Science*, v. 27, n. 5, p. 319-325, 2001. doi: 10.1177/016555150102700503.
- WALLACH, H. M. Topic modeling: beyond bag-of-words. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 23., 2006, Pittsburgh. *Proceedings...* New York: ACM, 2006. p. 977-984.
- ZIPE, G. K. *Human behavior and the principle of least effort: an introduction to human ecology*. Cambridge: Addison-Wesley, 1949. 573 p.



BIBLIOMETRICS RESEARCH IN THE ERA OF BIG DATA: CHALLENGES AND OPPORTUNITIES

A PESQUISA BIBLIOMÉTRICA NA ERA DO *BIG DATA*: DESAFIOS E OPORTUNIDADES [ver página 91]

[ir para o sumário]

*Dietmar Wolfram**

INTRODUCTION

The topic of big data has the potential to affect any discipline where large datasets are used. In this essay, I reflect on challenges and opportunities big data has to offer bibliometrics, scientometrics and informetrics (referred to here as “metrics”). In recent years, metrics researchers have begun to grapple with this issue (e.g., MOED, 2012; ROUSSEAU, 2012).

First, how does one define the term big data? Currently, there is no universally accepted definition. Is it simply about size or does it also relate to what investigators are able to do with the massive volumes of research data now available? A growing array of research projects in the physical and biomedical sciences rely on data collections that would have been inconceivable a generation ago. For example, CERN’s Large Hadron Collider (LHC), which is used to study particle physics, has generated about 75 petabytes of data over a three-year period (HOFFMAN, 2013). This requires an enormous amount of storage space as well as efficient methods to access any aspect of the data. Furthermore, big science initiatives like the LHC require

* School of Information Studies, University of Wisconsin-Milwaukee; dwolfram@uwm.edu

the participation of large research teams, in some cases involving thousands of participants, which has its own project management challenges.

In metrics research, we do not work with datasets or research teams that are nearly as large as disciplines involved in big science. With this in mind, what does the concept of big data mean in a bibliometrics context? Like other scientific areas, the size of available datasets used for metrics research has increased. At the same time, metrics researchers are expected to employ large datasets to increase the reliability and generalizability of their findings.

When we look back to the datasets used for some of the earliest metrics studies of bibliographic and language datasets, we would consider them quite modest by today's standards. The manual collection and analysis of data made the processing of large datasets computationally demanding. This is evident in several classic bibliometric studies:

- Alfred Lotka used 6891 & 1325 authors for his studies of scientific productivity for Chemical Abstracts and Auerbach's *Geschichtstafeln der Physik*, respectively (LOTKA, 1926). Lotka had limited the data from Chemical Abstracts to authors whose last names began with the letters A and B.
- Samuel C. Bradford relied on 326 journals with 1332 references for his initial study on the concentration and scatter of literature on a topic across different journals (BRADFORD, 1934).
- George K. Zipf, who studied regularities in language use, relied on datasets that contained usually fewer than 10,000 word types. His comment on an earlier researcher's work: "Kaeding's total bulk of nearly 11 million running words so far overshoots a sample of optimum size that it is of little practical use to us." (ZIPE, 1949, Section 3.IV) reflected his concern with having too much data.

Times have changed. Today, much more data is available in machine-readable form, which simplifies the collection of large datasets. Metrics researchers rely on datasets that are measure in millions of data points. Computational

ability continues to improve and new analytical tools for numeric and textual analysis make it possible for large datasets to be processed rapidly and visualized so that small and large-scale patterns may be discovered.

WHAT IS “BIG” IN BIBLIOMETRICS?

Given the sizes of datasets associated with big data, often measures in terabytes and petabytes, do the datasets collected and analyzed by metrics researchers qualify as big data? Dataset size is certainly a criterion, but another important aspect that characterizes big data relates to performance requirements. Magoulas and Lorica (2009) stated that data become big data “... when the size and performance requirements for data management become significant design and decision factors for implementing a data management and analysis system” (p. 2). This aptly describes the current environment in which metrics researchers operate. Ideally, researchers would like to work with the entire population of data and not just a sample. Current bibliographic and citation databases store tens of millions of records, which in turn represent a subset of all published works. It is now possible to analyze and visualize relationships among tens of millions of documents (see, for example, <http://www.mapofscience.com/>). Even with the availability of the population of all published works, these database sources can be regarded as more bounded than datasets collected in the physical sciences. With adequate access, it is possible for bibliometrics researchers to process the entire contents of multiple databases, where the units of analysis are authors, publications, journals, or citations. Although the amount of scientific literature continues to grow annually, the total volume is still at least an order of magnitude smaller than populations of data collected in some areas of the natural and biomedical sciences. When it comes to the metric analysis of products of scholarly communication, researchers can only process what has been published and what is available. This does not imply that the processing of bibliometric datasets is not without the types

of challenges found in other disciplines. The size of these datasets are still considerable and can increase dramatically when researchers work with the full text of documents and not just document surrogates consisting of bibliographic reference data or citations. The HathiTrust Research Center, for instance, provides access to the full text of more than three million public domain works (<https://analytics.hathitrust.org/>). Other forms of recorded discourse that have the potential for higher data volume have attracted the attention of metrics researchers in recent decades. Internet-based data for metrics study, whether records of user search and browsing patterns of the Internet or social media provide the potential for much larger datasets. In these cases, the speed with which new data becomes available may make the collection and analysis of the population of data challenging from a computational and legal perspective. These data also raise ethical issues for metrics research, where data and textual mining techniques on large datasets may reveal personally identifiable information, as was evident in 2006 with the release of a search log dataset by AOL (NAKASHIMA, 2006).

BIG DATA CHALLENGES FOR BIBLIOMETRICS

Several ongoing challenges related to data access and analysis of data face metrics researchers. First, as with many areas of scholarly inquiry, data accessibility is a potential barrier. Desired data may only be available through vendors for which subscriptions are needed. Restrictions for end user-level access may limit the amount of data that can be downloaded at one time. Purchase of complete datasets is possible, but can be cost prohibitive. Increasingly, social media websites that generate publicly available data are providing Application Program Interfaces (APIs) to access these data, but they may limit the amount of data that can be downloaded within given periods of time, or deny data access for the privacy reasons cited above. In general, publicly available data is often decentralized, making identification and collection of the data difficult. On the positive side, one earlier

challenge that has become less of an issue in recent years is the increase in speed for data transfer and the decline in cost for big data storage.

With big data, there has been the belief that “more is better.” However, can we say this is always the case? More data can be good, but is it “good” data? The ability to access the population of data makes for complete representation. However, some types of data or data from different sources may require extensive cleaning and standardization before becoming usable for analysis. This is particularly true for Internet-based data, for instance, where Web server logs often contain superfluous content that does not relate directly to user-initiated actions (HAN; WOLFRAM, 2016). Furthermore, one metrics researcher has warned “... brute force computation with big data may lead to false discoveries and spurious correlations ...” (PRATHAP, 2014). Regardless of the completeness of the dataset, conclusions drawn about phenomena of interest to metrics researchers may only apply to the samples used. As one example of this, overall characteristics of datasets can change with different amounts of data used. Frequency distributions and descriptive statistics associated with metrics datasets can vary greatly with the size of the dataset, while other aspects of the data can remain relatively stable. (AJIFERUKE, WOLFRAM; FAMOYE, 2006).

Large datasets and complex relationships within the data lead to high dimensional data representation, which in turn is associated with increased computational overhead. As an example, the representation of bibliometric data of interest (e.g., authors, publications, journals) as vector spaces, as used in information retrieval research, makes it possible to assess the relationships between the entities of interest. Unfortunately, storing data in this form can result in matrices consisting of millions of rows and columns. Dimensionality reduction techniques are needed to reduce processing overhead while preserving the essence of the observed relationships. Statistical techniques such as cluster analysis and factor analysis for quantitative data, along with language-based methods, where appropriate, can reveal these relationships or simplify the data representation.

The application of appropriate data analysis and summarization techniques presents another potential challenge. A plethora of network and link analysis as well as text and data mining techniques are available to provide insights into overt or hidden relationships within datasets. Similarly, exploratory statistical techniques that summarize and visualize datasets allow researchers to identify patterns that would be otherwise hidden due to the volume of the data. What researchers have found with these different tools and techniques is that they can produce different outcomes even for the same data. Which techniques are best and provide the greatest validity? There is no one correct way to analyze or visualize the data, so researchers must state their case convincingly when presenting and interpreting outcomes.

OPPORTUNITIES FOR BIBLIOMETRICS

Although the growing volumes of data and options for data analysis and presentation provide challenges for the identification of best approaches, these techniques and advances in computation make it possible to process large volumes of data in ways that were not possible even a few years ago.

Bibliometrics research has traditionally focused on bibliographic and citation-based aspects of recorded discourse. The application of network analysis techniques and associated software has made it possible to study large datasets across multiple disciplines. Data mining approaches (THELWALL, 2001, 2001), text mining (SONG; CHAMBERS, 2014), methods used for clustering and classification (GLÄNZEL; SCHUBERT, 2003) and methods for community detection (BOHLIN et al., 2014) can identify hidden relationships and patterns in data. Publicly available network analysis software such as Pajek (<http://mrvar.fdv.uni-lj.si/pajek/>) and Gephi (<https://gephi.org/>), as well as software specifically developed to support metrics research such as CiteSpace (<http://cluster.cis.drexel.edu/~cchen/citespace/>), Sci2 (<https://sci2.cns.iu.edu/user/index.php>) and VOSviewer

(<http://www.vosviewer.com/>) provide routines for the summarization and visualization of large datasets.

Link-based analysis based on citations (direct citation, co-citation, bibliographic coupling) or co-authorship has been a focal point of metrics research for decades. One downside associated with the use of these data types is that if there is no linkage, there is no relationship. Researchers in different fields or different parts of the world who are not aware of each other may share common interests, but their similarities would not be identified without explicit citation-based linkages or formal collaborations.

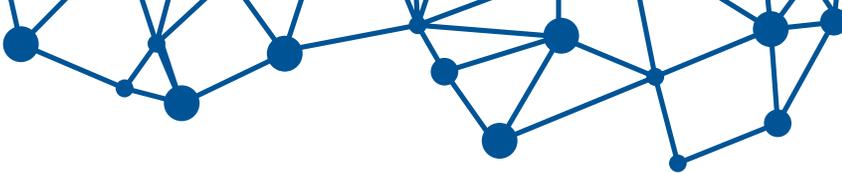
Language-based analysis presents another avenue by which relationships among entities of interest may be studied. Early efforts focusing on language relied on co-word analysis of title or abstract keywords or subject headings (e.g., CALLON; COURTIAL; LAVILLE, 1991; COURTIAL, 1994). Developments in natural language processing (NLP), machine learning, text mining and topic modeling now make it possible to focus on larger corpora of full text. These methods can scale to accommodate larger datasets, but in turn require large enough datasets to effectively train models, which will then result in outcomes that are more reliable. Although there is high computational overhead for the training of models for some of these techniques, the payoff for researchers is the resulting dimensionality reduction for the analysis of the data. In the case of text-based comparisons using topic modeling (WALLACH, 2006), hundreds of thousands of candidate keywords that are used to index publications may be reduced to a few hundred topics to represent the publications. Applications of text-based methods extend to the study of scholarly impact (SONG; DING, 2014) and complementary approaches to citation analysis for the study of author research profile similarity (LU; WOLFRAM, 2012). Lu and Wolfram, for example, studied author research similarity based on topic models and found they produced the most coherent relationships among authors when compared to citation-based approaches. Text-based methods also can be combined with citation-based methods to build on the strengths of each approach (GLENISSON et al., 2005).

REFERENCES

- AJIFERUKE, I.; WOLFRAM, D.; FAMOYE, F. Sample size and informetric model goodness-of-fit outcomes: A search engine log case study. *Journal of Information Science*, v. 32, n. 3, p. 212-222, 2006.
- BOHLIN, L.; EDLER, D.; LANCICHINETTI, A.; ROSVALL, M. Community detection and visualization of networks with the map equation framework. In: DING, Y.; ROUSSEAU, R.; WOLFRAM, D. (Eds.). *Measuring scholarly impact: methods and practice*. Cham: Springer International Publishing, 2014. p. 3-34. doi: 10.1007/978-3-319-10377-8.
- BRADFORD, S. C. Sources of information on specific subjects. *Engineering*, v. 137, n. 3550, p. 85-96, 1934.
- CALLON, M.; COURTIAL, J. P.; LAVILLE, F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, v. 22, n. 1, p. 155-205, 1991. doi: 10.1007/BF02019280.
- COURTIAL, J. P. A cword analysis of scientometrics. *Scientometrics*, v. 31, n. 3, p. 251-260, 1994. doi: 10.1007/BF02016875.
- GLÄNZEL, W.; SCHUBERT, A. A new classification scheme of science fields and sub-fields designed for scientometric evaluation purposes. *Scientometrics*, v. 56, n. 3, p. 357-367, 2003. doi: 10.1023/A:1022378804087.
- GLENISSON, P.; GLÄNZEL, W.; JANSSENS, F.; DE MOOR, B. Combining full text and bibliometric information in mapping scientific disciplines. *Information Processing and Management*, v. 41, n. 6, p. 1548-1572, 2005. doi: 10.1016/j.ipm.2005.03.021.
- HAN, H.; WOLFRAM, D. An exploration of search session patterns in an image-based digital library. *Journal of Information Science*, v. 42, n. 4, p. 477-491, 2016. doi: 10.1177/0165551515598952.
- HOFFMAN, M. LHC collider recorded over 75 petabyte until today's shutdown. *Science World Report*, 14 Feb. 2013. Available from: <<http://www.scienceworldreport.com/articles/4973/20130214/lhc-collider-cern-datacenter-100-petabyte-physics-shutdown.htm>>. Viewed: 28 Feb. 2017.

- LOTKA, A. J. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science*, v. 16, n. 12, p. 317-323, 1926.
- LU, K.; WOLFRAM, D. Measuring author research relatedness: a comparison of word-based, topic-based, and author cocitation approaches. *Journal of the Association for Information Science and Technology*, v. 63, n. 10, p. 1973-1986, 2012. doi: 10.1002/asi.22628.
- MAGOULAS, R.; LORICA, B. Introduction to big data. *Release 2.0*, n. 11, p. 1-39, 2009.
- MOED, H. F. The use of big datasets in bibliometric research. *Research Trends*, n. 30, p.31-33, 2012.
- NAKASHIMA, E. AOL takes down site with users' search data. *The Washington Post*, 8 Aug. 2006. Available from: <<http://www.washingtonpost.com/wp-dyn/content/article/2006/08/07/AR2006080701150.html>>. Viewed: 28 Feb. 2017.
- PRATHAP, G. Big data and false discovery: analyses of bibliometric indicators from large data sets. *Scientometrics*, v. 98, n. 2, p. 1421-1422, 2014. doi: 10.1007/s11192-013-1063-z.
- ROUSSEAU, R. A view on big data and its relation to Informetrics. *Chinese Journal of Library and Information Science*, v. 5, n. 3, p. 12-26, 2012.
- SONG, M.; CHAMBERS, T. Text mining with the Stanford CoreNLP. In: DING, Y.; ROUSSEAU, R.; WOLFRAM, D. (Eds.). *Measuring scholarly impact: methods and practice*. Cham: Springer International Publishing, 2014. p. 215-234. doi: 10.1007/978-3-319-10377-8.
- SONG, M.; DING, Y. Topic modeling: Measuring scholarly impact using a topical lens In: DING, Y.; ROUSSEAU, R.; WOLFRAM, D. (Eds.). *Measuring scholarly impact: methods and practice*. Cham: Springer International Publishing, 2014. p. 235-257. doi: 10.1007/978-3-319-10377-8.
- THELWALL, M. A web crawler design for data mining. *Journal of Information Science*, v. 27, n. 5, p. 319-325, 2001. doi: 10.1177/016555150102700503.
- WALLACH, H. M. Topic modeling: beyond bag-of-words. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 23., 2006, Pittsburgh. *Proceedings...* New York: ACM, 2006. p. 977-984.
- ZIPE, G. K. *Human behavior and the principle of least effort: an introduction to human ecology*. Cambridge: Addison-Wesley, 1949. 573 p.





AVALIAÇÃO INSTITUCIONAL NA USP

INSTITUTIONAL ASSESSMENT IN USP [see page 133] [go to summary]

*Pedro Vitoriano Oliveira** e *Vahan Agopyan***

I. INTRODUÇÃO

Avaliar consiste em empreender um diagnóstico a partir da comparação entre as metas que se pretendem alcançar, confrontando-se os objetivos estabelecidos previamente com os efeitos obtidos, em um decurso de tempo, a fim de perceber as potencialidades e os pontos críticos. É uma atividade que possibilita a formação de um vínculo de caráter histórico, pelo qual se pretende a análise do momento presente da entidade – o qual é reflexo das escolhas e práticas passadas – a fim de possibilitar a programação relativamente a momentos posteriores e a possíveis mudanças.

Avalia-se para verificar as características existentes em um dado objeto ou situação, porém este processo não se exaure na verificação destes atributos. Normalmente, ele é utilizado como fundamento para a realização de outras atividades, isoladas ou reunidas, as quais são destinadas à consecução de determinados objetivos.

Compreende-se avaliação de Instituição de Ensino Superior (IES) como uma ação transformadora, conjunta e contínua, com o objetivo essencial de colaborar para a melhoria da qualidade das várias atividades que nela se desenvolvem como ensino – nos níveis de graduação e pós-graduação

* Instituto de Química, Universidade de São Paulo; pvolivei@iq.usp.br

** Escola Politécnica, Universidade de São Paulo

– pesquisa, cultura e extensão, das relações internacionais e nacionais com outras instituições análogas, a gestão dos recursos financeiros, humanos e de todas as responsabilidades a eles imputados, visando o melhor funcionamento institucional. Trata-se de procedimento autocrítico e responsável para atender ao bem comum e à missão institucional, devendo ser parte integrante do planejamento da instituição. Ao final, a contribuição de maior relevância que a avaliação trará é a criação e o fortalecimento de uma cultura de qualidade no contexto do plano institucional da Universidade.

Nas IES, a avaliação institucional pode ocorrer utilizando-se de quatro etapas distintas (GONZÁLES-GONZÁLES, 2012):

- Autoavaliação – que consiste em um processo com vistas a promover tanto o autoconhecimento como o reconhecimento, de modo que as figuras do avaliado e do avaliador são as mesmas, não havendo a atuação de sujeitos que não façam parte da entidade que ora se avalia.
- Avaliação externa – desempenhada por especialistas ora de uma área do conhecimento, ora de uma disciplina específica, com fundamento na autoavaliação previamente realizada, aplicando-se apenas no âmbito dimensional de programas acadêmicos, devendo os resultados alcançados em sua efetivação serem utilizados, exclusivamente, por parte do programa que se avalia.
- Acreditação – modalidade avaliativa pela qual o organismo acreditador – no âmbito das instituições de ensino superior, representados pelo Ministério da Cultura (MEC) ou Conselho Estadual de Educação (CEE) – concede o certificado de avaliação, baseando-se na autoavaliação realizada, de modo prévio, pela Instituição, além de conferir fé pública em relação ao quilate acadêmico inerente a esta.
- Certificação profissional – em que um organismo de certificação profissional – como a Ordem dos Advogados do Brasil (OAB), ou o Conselho Federal de Medicina (CFM), entre outros – são responsáveis pela emissão do certificado de avaliação do principal produto da IES, o graduado, utilizando como parâmetros para a análise o nível de informação

que possibilita conferir fé pública às qualificações de ordem acadêmica e profissional e, neste último caso, é avaliado o resultado da atuação educacional sendo, por isso, desenvolvida em turmas de egressos ou nestes considerados isoladamente.

No que diz respeito à Universidade de São Paulo (USP), relativamente à aplicação de avaliações institucionais no decurso do tempo, há uma prevalência da utilização das modalidades da Autoavaliação e da Avaliação Externa. No entanto, as certificações profissionais, requeridas pela OAB e CFM, tem a participação de egressos de cursos da USP.

No que concerne ao planejamento nas IES, este constitui uma atividade extremamente necessária tanto para o processo avaliativo como ao desenvolvimento adequado das atribuições organizacionais. Para tanto, não se deve planejar de modo mecânico ou baseado em metas com pouca probabilidade de serem alcançadas: é imprescindível a ocorrência de reflexões ou discussões, desde o momento em que se traçam as diretrizes a serem observadas (contemplando distintos lapsos temporais para o alcance de objetivos), a fim de sistematizar as atividades e adotar as ações com vistas ao fortalecimento dos aspectos favoráveis (percebidos por meio da avaliação prévia), bem como para que possam ser solucionados possíveis problemas que tenham sido percebidos por meio do processo de avaliação. Nesse, as duas ações – avaliar e planejar – ocorrem de forma agregada, com a finalidade de transformação, ocorrendo de modo contínuo e conjuntamente. Assim articuladas (autoavaliação e planejamento), visam promover a excelência da educação superior. Por intermédio dessas ações, pode-se fomentar e robustecer os parâmetros para a inserção, no âmbito institucional universitário, de uma cultura baseada na valorização da qualidade. Trata-se, ainda, de um procedimento que se realiza com o objetivo de assegurar a autonomia universitária, destinado a fomentar a autocrítica, de maneira responsável, em prol dos benefícios de todos os participantes deste meio e da missão escolhida pela instituição, sendo imprescindível que esteja contemplado como etapa do planejamento desta (GONZÁLES-GONZÁLES, 2012).

Para o Sistema Nacional da Educação Superior – Sinaes – em atendimento a diretrizes emanadas pela Comissão Nacional de Avaliação da Educação Superior – CONAES – bem como pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep – considerados conjuntamente, os processos avaliativos precisam formar um conjunto direcionado a integrar as várias instâncias da realidade que se avalia, garantindo a coerência em relação aos aspectos conceituais, epistemológicos e práticos, bem como possibilitando a consecução das metas pretendidas por meio de inúmeros expedientes ou estratégias que se adotam.

2. HISTÓRICO

2.1. Contexto da Avaliação Institucional: Brasil, Estado de São Paulo e USP

A Constituição Federal de 1988 (BRASIL, 1988), em seu Art. 209, dispõe sobre a necessidade do procedimento de avaliação nas instituições de ensino pátrias em todos os níveis:

Art. 209: o ensino é livre à iniciativa privada, mediante avaliação de qualidade pelo poder público.

Neste sentido, cabe ao Estado proceder à verificação sobre o atendimento dos padrões de qualidade por parte das IES privadas que atuam no âmbito educacional. Segundo o Art. 209, não é incumbência do Estado a avaliação de IES públicas, sendo, no entanto, prevista essa avaliação na Lei das Diretrizes e Bases da Educação Nacional (LDB).

A Lei Federal nº 9.394, de 20/12/1996 (BRASIL, 1996), conhecida como Lei das Diretrizes e Bases da Educação Nacional (LDB), no seu Art. 9, apresenta duas regras dirigidas à avaliação, nos incisos VI e IX:

VI – assegurar processo nacional de avaliação do rendimento escolar no ensino fundamental, médio e superior;

IX – cabe ao governo federal autorizar, reconhecêr, credenciar, supervisionar e avaliar cursos e instituições de educação superior.

No mesmo diploma legal, o Art. 46 dispõe que:

a autorização e o reconhecimento de cursos, bem como o credenciamento de instituições de educação superior, terão prazos limitados, sendo renovados, periodicamente, após processo regular de avaliação.

A avaliação das Instituições de Ensino Superior (IES), no Estado de São Paulo, foi instituída pelo Conselho Estadual de Educação a partir da Deliberação CEE 04/2000.

Realizando-se uma comparação entre os regramentos estabelecidos nas instâncias federal e estadual, não se verificam divergências entre eles. De acordo com o Sinaes e Inep, a Avaliação Institucional está associada aos seguintes aspectos:

- à melhoria da qualidade da educação superior
- à orientação da expansão de sua oferta
- ao aumento permanente da sua eficácia institucional e efetividade acadêmica e social
- ao aprofundamento dos compromissos e responsabilidades sociais das instituições de educação superior, por meio da valorização de sua missão pública, da promoção dos valores democráticos, do respeito à diferença e à diversidade, da afirmação da autonomia e da identidade institucional.

No Estado de São Paulo, as regras foram determinadas pelo Conselho Estadual de Educação. Este apresenta os objetivos a serem alcançados com a Avaliação Institucional:

1. Orientar, acompanhar e fiscalizar as universidades e centros universitários do Estado de São Paulo.
2. Analisar periodicamente o desempenho e atualização institucional no que se refere a:
 - eficácia e eficiência do ensino
 - importância da pesquisa
 - relevância da produção cultural e científica
 - eficácia da formação profissional
 - importância das ações comunitárias
 - condições da graduação e pós-graduação
 - qualidade da gestão.

Cabe destacar que, por haver reconhecimento da necessidade e relevância dos procedimentos avaliativos para a melhor atuação das Instituições de Ensino Superior em suas atividades, a Avaliação Institucional foi instituída no âmbito da USP anteriormente à regulamentação em nível estadual, a CEE 04/2000.

2.2. Avaliação Institucional na USP

A Universidade de São Paulo deu início aos procedimentos relacionados à sua avaliação institucional, de maneira sistematizada, por meio de ciclos de Avaliação Departamental. Esta atividade teve início quando foi constituída a Comissão Permanente de Avaliação (CPA), à data de 7/4/1992, (Resolução: 3920/92), em uma proposta apresentada pela Comissão de Assuntos Acadêmicos (CAA) e Comissão Especial de Regimes de Trabalho (CERT) da USP.

Desde então, verifica-se que houve a realização de quatro ciclos de avaliação (USP, 2005, 2010, 2016a) sendo que, desde 2000, correspondem a ciclos de cinco anos seguidos (Figura 1), em conformidade com a determinação emanada pelo Conselho Estadual de Educação do Estado de São Paulo, segundo o disposto no Art. 3º da CEE 04/2000.

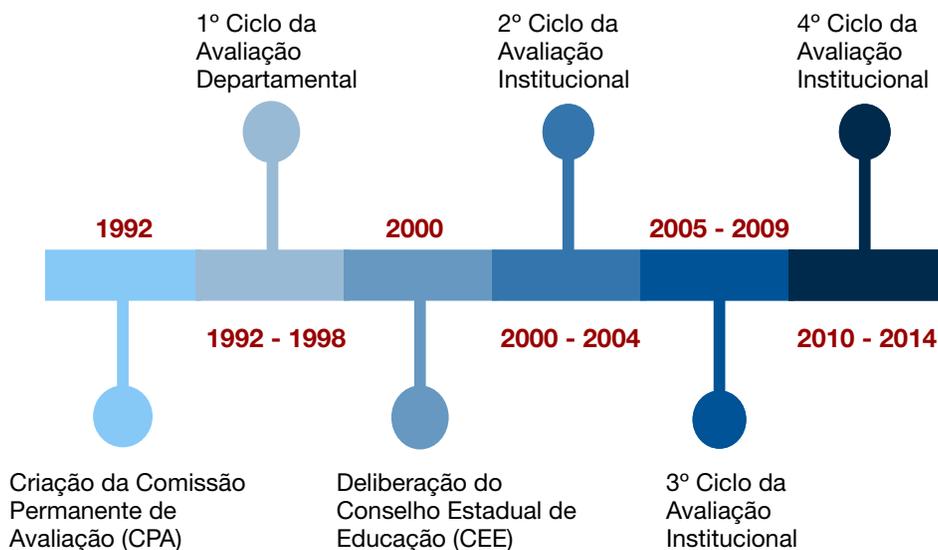


Figura 1. Perspectiva histórica da Avaliação Institucional na USP desde 1992.

No 4º Ciclo de Avaliação Institucional USP (USP, 2016a), correspondente ao quinquênio 2010 – 2014 decidiu-se pela continuidade do mesmo padrão dos ciclos anteriores, o que evidencia a importância relativa à existência do procedimento entre as práticas, exigida no âmbito da Universidade de São Paulo. Neste contexto, foi utilizado um instrumento de avaliação aperfeiçoado, denominado Formulário de Avaliação, sendo um com questões específicas para os Departamentos e outro para as Unidades, os quais serão detalhados no item 3.1. Todas as Unidades de Ensino e seus Departamentos, Institutos, Centros Especializados, Museus e Hospitais fizeram uso deste mecanismo, direcionado para a autoavaliação.

Visando integrar o interesse dos outros processos avaliativos da USP, tais como da Comissão Especial de Regime de Trabalho (CERT), Comissão Central de Avaliação para Progressão de Nível na Carreira Docente (CCAD), Agência USP de Cooperação Acadêmica Nacional e Internacional (AUCANI), Agência USP de Inovação (AUSPIN) e das Pró-Reitorias (Graduação, Pós-Graduação, Pesquisa e Cultura e Extensão) foram acrescentadas, nos formulários, questões elaboradas por esses setores, refletindo a

preocupação em se construir uma cultura em prol da qualidade baseada no engajamento e atuação sistêmica de todos os membros da comunidade acadêmica e administrativa da USP.

Participaram da construção do 4º Ciclo de Avaliação Institucional USP (USP, 2016a), além dos 10 componentes que constituem a CPA, responsáveis pelos setores acadêmicos e administrativos, três membros da Vice-Reitoria (assessoria técnica e apoio administrativo) e a Superintendência de Tecnologia da Informação (STI) da USP. A contribuição desses participantes será detalhada abaixo, no item 3, procedimentos adotados. Também foram envolvidos 189 assessores externos aos quadros da USP, que fizeram a avaliação de 42 Unidades de Ensino, 6 Centros e Institutos Especializados, 4 Museus Universitários e 2 Hospitais, totalizando 54 Unidades, além dos 216 Departamentos das Unidades de Ensino.

Uma lista de sugestões de nomes de pesquisadores, nacionais e internacionais, que poderiam fazer parte da Comissão de Assessores Externos foi elaborada pela Unidade e submetida à CPA para avaliação. Nesta etapa, foram considerados o perfil técnico e o grau de relacionamento do(a) indicado(a) com a Unidade. Pesquisadores com envolvimento em projetos de pesquisa com pesquisadores das Unidades não foram recomendados. Ao final, após avaliação das indicações das Unidades, a CPA indicou a relação de nomes a serem considerados como titulares e suplentes para comporem a Comissão de Assessores Externos. A Tabela 1 mostra o panorama numérico geral da evolução das Unidades e Departamentos que participaram dos diferentes ciclos avaliativos da USP, bem como dos assessores externos envolvidos nos processos. As variações nos números de Departamentos e Unidades se devem a eventuais junções e incorporações de novos Centros à Universidade de São Paulo. Particularmente, nesse 4º Ciclo de Avaliação Institucional da USP foram consideradas como Unidades independentes os Centros e Institutos Especializados, Museus e Hospitais.

Tabela 1. Avaliação Institucional da USP em Números (1992 – 2014)

AVALIAÇÃO INSTITUCIONAL	1º CICLO 1992-1998	2º CICLO 2000-2004	3º CICLO 2004-2009	4º CICLO 2010-2014
DEPARTAMENTOS	210	199	207	216
UNIDADES	35	36	39	54
TOTAL DE ASSESSORES EXTERNOS	421	287	152	183*
a) Assessores Externos Nacionais	289	156	107	122
b) Assessores Externos Internacionais	132	131	45	61

* 7 Unidades foram visitadas por duas comissões

3. PROCEDIMENTOS ADOTADOS NO 4º CICLO DE AVALIAÇÃO INSTITUCIONAL USP

O 4º Ciclo da Avaliação Institucional (USP, 2016a) foi realizado em 5 etapas distintas:

- Primeira – consistiu de visitas coordenadas aos Campi USP para exploração e orientação sobre o Processo da Avaliação Institucional, incluindo informações sobre o sistema operacional para a autoavaliação das Unidades, Departamentos, Institutos e Centros Especializados, Museus e Hospitais. As visitas ocorreram em 3 datas diferentes, no final do ano de 2014, imediatamente antes do início do encaminhamento dos formulários para preenchimento pelas Unidades e Departamentos. Participaram dessas discussões dirigentes, chefes de departamentos, presidentes de comissões e integrantes dos setores administrativos.

- Segunda – ocorreu o preenchimento dos formulários pelos Departamentos e Unidades. O processo foi instruído de forma que os formulários fossem preenchidos de forma coordenada e convergente, primeiro os Departamentos para que os dirigentes pudessem sintetizar, no formulário da sua Unidade, os principais aspectos relacionados aos seus departamentos, bem como da gestão e infraestrutura. Os Departamentos cadastraram os seus formulários das autoavaliações, eletronicamente, para que os dirigentes pudessem acessá-los e organizar a autoavaliação da Unidade, que também foi cadastrada eletronicamente. Essas atividades foram finalizadas até 30 dias antes da visita da Comissão de Assessores Externos, para que os seus integrantes tivessem tempo de analisar todo o material, antes da visita à Unidade.
- Terceira – foram as visitas das Comissões de Assessores Externos às Unidades da USP, que ocorreram de forma coordenada entre junho e novembro de 2015. Os integrantes das comissões, formadas por professores/pesquisadores externos aos quadros da USP, tinham certa aderência às atividades desenvolvidas na Unidade e/ou Departamentos.
- Quarta – compreendeu a avaliação das atividades-fim, feita de forma global pela Comissão de Assessores Seniores. Nessa etapa, seis pesquisadores foram convidados a darem os seus pareceres sobre a graduação, pós-graduação, pesquisa, cultura e extensão, internacionalização e gestão, com base nos formulários das autoavaliações das Unidades e Departamentos, dos pareceres emitidos pelas Comissões de Assessores Externos, e nas entrevistas com integrantes das Unidades (diretores, chefes, professores, pesquisadores, alunos de graduação e pós-graduação e funcionários).
- Quinta – Elaboração do relatório final.

3.1. Sobre o formulário de autoavaliação

Os formulários adotados para as autoavaliações dos Departamentos e Unidades, no 4º Ciclo de Avaliação Institucional USP (USP, 2016a), foram fundamentados em três eixos:

- Conjunto de intenções (missão, visão e proposta educacional);
- Autoavaliação (gestão, articulação, infraestrutura, serviços técnicos administrativos, docentes, processos de ensino aprendizagem, graduação, pós-graduação, pesquisa, cultura e extensão, e internacionalização);
- Plano institucional com previsão de metas e ações.

A missão é a finalidade, expressa em tempo presente, que orienta todas as decisões e ações educativas de uma organização universitária. Na educação superior descreve o dever e ser da instituição, os valores que lhe dão vida, as necessidades que irá satisfazer e as atividades-fim que desenvolve. É o propósito declarado. É aquilo que a Unidade é e a razão pela qual existe. Compõe-se de enunciados sintéticos que capturam a essência da instituição. A missão deve compor-se de três elementos importantes:

- Declaração de princípios e valores da Instituição – A Unidade internacionaliza princípios e valores que obedecem às motivações associadas a sua fundação ou ao contexto que origina a instituição. Este tipo de declaração de princípios gerais não se operacionaliza em propósitos a serem realizados, mas é importante para que os membros da comunidade acadêmica, os avaliadores, entendam o contexto de valores mais amplo no qual opera a instituição.
- Caracterização acadêmica da Instituição – A referência é a identidade acadêmica e educativa da instituição, que deve estar explícita na missão. Na Unidade convergem programas, metodologias de ensino, níveis de estudo, relações com a sociedade e finalidades de pesquisa, entre outras. Nenhuma delas é neutra e a forma como cada instituição se posiciona determina a sua missão. Estes elementos compõem a parte central da missão para uma entidade educativa. Alguns exemplos de caracterização acadêmica são: i) Instituições com programas marcadamente profissionais, em contraste com outras que dão importância à educação em competências gerais; ii) Instituições que privilegiam mecanismos de aprendizagem autônomos e ativos, enquanto outras são mais conservadoras em

aspectos pedagógicos; iii) Há Instituições em que os cursos de graduação e pós-graduação têm papel equivalentes, enquanto que outras conservam interesse em pós-graduação; iv) Há Instituições em que a produção em investigação é parte da essência da instituição, enquanto que em outras somente se contempla o ensino.

- Relação da Instituição com o seu contexto social – Uma missão baseada estritamente no ajuste dos propósitos declarados não pode deixar de considerar o contexto do qual ela faz parte e no qual seus egressos operam. A educação é um serviço público e, como tal, tem o dever de pertinência à sociedade. A maneira como cada instituição responde a esse dever é própria e deve refletir na missão. Alguns exemplos de relação e seu contexto social são: i) A titulação de estudantes e a identificação de egressos bem-sucedidos no mercado de trabalho, sinaliza retorno social importante. ii) Uma exigência distintiva que caracteriza uma instituição que, mediante seus cursos, pesquisas e serviços, se coloca no dever de contribuir ativamente na análise e solução de problemas de sua região ou país. Uma pode ser mais seletiva em razão dos programas que oferece, enquanto outra pode privilegiar a inclusão.

Partindo da missão, a visão descreve a situação que a organização universitária deseja ter em um futuro próximo. É um exercício que busca pensar, a partir das condições atuais até o futuro, qual é a melhor interpretação que a Unidade deve fazer de sua própria missão. Sua construção é um exercício estratégico e não simplesmente uma atividade operacional. Requer conhecimento dos desafios que a Unidade enfrenta (TORO, 2012). Mudanças de visão podem estar relacionadas a diferentes circunstâncias, tais como:

- Tendência de massificação da educação e de necessidade de instituições mais inclusivas que atendam a essa demanda;
- Condições socioeconômicas locais e globais, nas quais opera a instituição, e suas tendências;
- Transformação das áreas do conhecimento e suas relações crescentes;

- Tendências em educação, como evolução das competências esperadas dos graduados, evolução das metodologias pedagógicas e uso de recursos tecnológicos para sua implementação;
- Necessidades locais de pesquisa versus tendências globais.

Ao insistir no adjetivo estratégico procura-se projetar a instituição no cenário futuro no qual operará. Isso a leva a fazer escolhas de caminhos de longo prazo, o que lhe permitirá ser bem-sucedida – considerando sua missão e ambiente. A visão implica em decisões complexas e mudanças institucionais, significa assumir riscos e implica comprometer recursos e operações. Tudo isso não coincide com o cotidiano e a operação atual, daí sua natureza estratégica (TORO, 2012).

Em síntese, pode-se dizer que as atividades de planejamento e gestão institucional e avaliação e geração de planos de melhoria têm por objetivo levar a instituição, de um tempo presente com sua missão estabelecida, a uma visão proposta em médio e longo prazos.

No eixo autoavaliação, as Unidades e Departamentos foram orientados, na elaboração das respostas, a considerarem as metas estabelecidas pelas Unidades e Departamentos por ocasião do 3º Ciclo de Avaliação (USP, 2010), bem como as manifestações dos assessores externos naquela ocasião.

Com o objetivo de se obter relativa uniformidade de informações, foram organizados roteiros específicos para as autoavaliações das Unidades e Departamentos, constituindo-se em elementos de base para o processo de avaliação, contemplando principalmente os aspectos qualitativos. Essas informações faziam parte do Guia da Avaliação Institucional USP 2010-2014, distribuída a todos os Departamentos e Unidades.

Além disso, o 4º Ciclo de Avaliação Institucional (USP, 2016a) foi caracterizado pela inserção de alguns diferenciais em relação aos anteriormente realizados, como o preenchimento dos formulários em redação bilíngue (versões em português e inglês), de forma a possibilitar aos assessores internacionais acessar, de modo mais efetivo, as informações prestadas pelos

Departamentos e Unidades, valorizando a transparência e a necessária ampliação do acesso às informações relativas a este tema.

3.2. Organização do processo da Avaliação Institucional

Para subsidiar o processo da Avaliação Institucional foi desenvolvido pela Superintendência de Tecnologia da Informação – STI um sistema de informatização para o preenchimento *on-line* dos formulários, para o cadastro dos pareceres dos avaliadores externos, além de toda a comunicação, submissão e análise das informações. Nesse sistema foi aberto um campo para cada Unidade, onde puderam ser cadastrados os relatórios de ciclos de avaliações anteriores, tabelas, planilhas e gráficos, retirados dos bancos de dados dos sistemas corporativos da USP, com informações de atividades das Unidades e seus Departamentos, para que pudessem ser consultadas e comparadas com outras autoavaliações dos ciclos anteriores. Sendo assim, os dados das últimas avaliações da USP, registrados no Sistema USP, constituirão, além de registro histórico para futuras Comissões de Avaliação, de uma rica fonte de consulta para dirigentes da Universidade e das Unidades, que poderão nortear decisões e diretrizes a serem tomadas.

As Comissões de Assessores Externos foram constituídas por professores e pesquisadores externos aos quadros da USP e, com o intuito de sistematizar o trabalho deles, foi elaborado um roteiro de sugestões, que ficou disponível no site para que fosse utilizado pelos Assessores Externos, antes e durante as visitas às Unidades. Este roteiro continha as seguintes sugestões:

- Sumarizar o desenvolvimento acadêmico da Unidade, com base no relatório de autoavaliação e nas demais informações existentes, indicando pontos que: i) sejam considerados adequados; ii) que mereçam destaque; e, iii) que necessitem aprimoramento e sugestões de como a Unidade poderá investir para estimular a qualidade das atividades acadêmicas.
- Explicitar a situação da Graduação e da Pós-Graduação, com base em indicadores objetivos disponíveis e nos projetos em desenvolvimento,

bem como na articulação entre as atividades-fim, a interdisciplinaridade, o estágio de internacionalização das atividades-fim, a articulação interna e externa da Unidade e as assimetrias (desigualdades) internas na realização das atividades-fim.

- Analisar a objetividade das metas acadêmicas propostas e a consonância com a missão da Unidade/Departamentos e com o estágio de desenvolvimento acadêmico da Unidade.
- Comentar as manifestações obtidas durante as entrevistas com o corpo discente (Graduação/Pós-Graduação) sobre o desenvolvimento das atividades acadêmicas na Unidade/Departamento/Programa de Pós-Graduação.
- Analisar o planejamento e a gestão acadêmica e administrativa da Unidade e dos Departamentos.
- Analisar a situação da infraestrutura da Unidade/Departamentos em relação a sua organização administrativa, abrangendo também Recursos Humanos, materiais, biblioteca e tecnologia da informação e inovação.
- Comentar a interação da Unidade com outras instituições congêneres e sua relação com outros setores da sociedade civil e organizada (indústria, governo, etc.), nos âmbitos nacionais ou internacionais, e seu impacto acadêmico, econômico e/ou social.
- Apresentar, se pertinente, sugestões e/ou recomendações à administração das Unidades, dos Departamentos, da Universidade e à Comissão Permanente de Avaliação da CPA.

Ao final das visitas, cada Comissão elaborou um parecer com comentários sobre pontos positivos, críticas e sugestões que foram cadastradas no site da avaliação.

A Comissão de Assessores Seniores, formada por 6 integrantes teve como objetivo avaliar as atividades-fim de forma global, analisando o desempenho de todas as Unidades para as respectivas atividades em que foram incumbidos. Aos integrantes dessa Comissão de Seniores foram solicitados considerarem nas suas avaliações, aspectos: i) considerados adequados; ii)

que mereçam destaque; e, iii) que necessitem aprimoramento e sugestões de como a USP poderá investir para estimular a qualidade. Além desses, também foram orientados, de forma específica, a considerar outros aspectos na avaliação das atividades a eles incumbidas, com base nos indicadores solicitados nos formulários das autoavaliações dos Departamentos e Unidades (disponíveis no site <https://gvr.uspdigital.usp.br/cpa>):

- Graduação e pós-graduação – avaliar a graduação e pós-graduação na USP, com base em indicadores objetivos disponíveis nos projetos em desenvolvimento, bem como na articulação com as outras atividades-fim, a interdisciplinaridade, o estágio de internacionalização, a articulação interna e externa da Unidade e as assimetrias internas na realização da atividade.
- Pesquisa e Cultura-Extensão – avaliar aspectos gerais da pesquisa e da cultura e extensão na USP, com base nas interações das Unidades com outras instituições congêneres e sua relação com outros setores da sociedade civil e organizada (indústria, governo, etc.), nos âmbitos nacionais ou internacionais, e seu impacto acadêmico, econômico e/ou social.
- Cooperação Internacional – analisar a internacionalização na USP com base nas informações e dados estatísticos fornecidos; mobilidade da graduação e pós-graduação, participação em redes acadêmicas internacionais, projetos de cooperação internacional, indicadores de desempenho, etc. Críticas e sugestões para a USP frente às ações de internacionalização foram solicitadas.
- Administração – analisar o planejamento e a gestão acadêmica e administrativa das Unidades, a situação da infraestrutura das Unidades em relação a sua organização administrativa, abrangendo também recursos humanos, materiais, biblioteca e tecnologia da informação e inovação.

4. RESULTADOS DA AVALIAÇÃO INSTITUCIONAL USP

O conceito utilizado na USP é o da avaliação como um processo contínuo, envolvendo avaliação-planejamento. O planejamento segue uma avaliação dos sucessos e fracassos em relação a objetivos previamente definidos e dos pontos fortes e fracos dos seus Departamentos e Unidades. Na busca pela excelência, o planejamento estratégico constitui importante instrumento que pode indicar a melhor direção a ser seguida. Assim, entende-se que o processo de avaliação institucional da Universidade de São Paulo pode ajudar nessa tarefa, indicando, a partir do conhecimento de sua realidade, caminhos para compreender e melhor utilizar seus pontos fortes; conhecer e eliminar ou adequar seus pontos fracos; usufruir das oportunidades externas; e, conseguir elaborar um efetivo plano de trabalho.

Ao longo dos últimos anos, a Avaliação Institucional tem sido um importante instrumento de gestão para os dirigentes. Apoia a tomada de decisões e favorece o acompanhamento e planejamento das Unidades e da Universidade, com o compromisso da Reitoria e das Unidades para implementarem as ações voltadas às metas institucionais. O processo de avaliação, com certeza, tem influência importante na busca da excelência da Universidade de São Paulo.

4.1. Discussões dos Resultados do 4º Ciclo da Avaliação Institucional USP

O Relatório foi o documento gerado e entregue ao Conselho Estadual de Educação do Estado de São Paulo (CEE) como parte dos requisitos exigidos para a avaliação quinquenal das Instituições de Ensino Superior. Seu conteúdo abordou, além da íntegra dos pareceres emitidos pelos Assessores Seniores, um diagnóstico feito pela Comissão Permanente de Avaliação (CPA), fundamentados nos pareceres elaborados pelas Comissões de Assessores Externos que visitaram as Unidades da USP. Esse Relatório bem como todos os documentos gerados com o 4º Ciclo da Avaliação Institucional (USP, 2016a) estão registrados no site <https://uspdigital.usp.br/wsusuario/>.

O desafio que se apresentou logo após o encerramento das atividades do 4º Ciclo da Avaliação Institucional da USP (USP, 2016a) foi o de transformar essa atividade, mais que uma tarefa a ser cumprida, exclusivamente por exigência legal, em uma oportunidade para se reunir importantes indicadores das diversas frentes de trabalho da USP, de modo a se tornar em elemento de gestão e aprimoramento desta Instituição. Sob esta perspectiva, foi instituído um Grupo de Trabalho – Análise da Avaliação Institucional (GT) que, juntamente com colaboradores dos diferentes Campi da USP, teve o objetivo de promover discussões, no âmbito das Unidades de Ensino e seus Departamentos, Institutos, Centros Especializados, Museus e Hospitais, sobre os resultados gerados com a Avaliação Institucional USP 2010-2014, visando ao aperfeiçoamento e operacionalização do processo, bem como permitir ampla reflexão sobre os resultados da avaliação, buscando o alcance da qualidade e da excelência almejados pela Instituição.

O roteiro de orientações de trabalho para as Comissões de Assessores Externos previa que fossem identificadas ações proativas, em andamento nas Unidades e Departamentos, bem como de aspectos críticos e sugestões de ações para vencer esses desafios. As comissões seguiram as recomendações e identificaram inúmeras ações importantes que estão em curso ou em vias de serem colocadas em prática nas Unidades e destacaram nos seus pareceres. Do mesmo modo, também identificaram pontos que podem dificultar o desenvolvimento das atividades-fim nas Unidades e, em alguns casos, sugeriram alternativas para contornar esses problemas, como reestruturação curricular, apoio às atividades extracurriculares, controle de evasão, entre outros.

Às Unidades foi solicitado que identificassem esses aspectos relatados nos pareceres dos assessores externos e que preparassem documento sucinto, discriminando as ações proativas que mereceram destaques, pontos críticos e ações destacadas para melhoria das atividades de gestão, graduação, pós-graduação, pesquisa, cultura e extensão, e internacionalização. Afora os aspectos apontados pelas Comissões de Assessores Externos, as Unidades, em um exercício complementar ao da autoavaliação, identificaram aspectos positivos bem como aspectos críticos e propuseram alternativas de gestão tanto para a

Unidade como para a USP de como essas dificuldades poderiam ser contornadas. As sugestões foram específicas, visando particularmente as necessidades das Unidades ou de seus Departamentos; em alguns casos, definições de metas e acompanhamento do seu cumprimento foram destaques.

Adicionalmente, foi solicitado às Unidades que apresentassem críticas e sugestões para o aperfeiçoamento e operacionalização do processo da Avaliação, sugerindo: i) indicações dos melhores e piores indicadores do formulário; ii) comentar e dar sugestões sobre a estrutura do formulário; e iii) comentar e dar sugestões sobre o Processo da Avaliação Institucional (preenchimento do formulário, visita da Comissão de Assessores Externos, composição das Comissões, Comissão de Assessores Seniores, entre outras).

As reuniões ocorreram no âmbito das Congregações ou CTAs das Unidades, contando com a presença de representantes do GT, e as discussões basearam-se nos formulários das autoavaliações dos Departamentos e Unidades, nos pareceres emitidos pelas Comissões de Assessores Externos, de Assessores Seniores bem como no conteúdo do Relatório Final da Avaliação.

As atividades relacionadas à análise e discussão dos resultados do 4º Ciclo da Avaliação Institucional (USP, 2016a) ocorreram entre abril e julho de 2016, primeiramente, nas Congregações ou CTAs das 54 Unidades, com a presença de representantes do GT. Como resultados dessas reuniões, as Unidades prepararam documentos sintetizando as solicitações acima mencionadas e enviaram para o GT-AAI. Em seguida, as Unidades nomearam representantes para que as propostas fossem discutidas em 7 Workshops, que ocorreram entre agosto e setembro de 2016.

Numa ação convergente, todas as contribuições enviadas pelas 54 Unidades da USP deram origem a dois documentos sínteses: um relacionado às contribuições para gestão e o outro relacionado às contribuições para o processo da Avaliação Institucional que será enviado à nova CPA (ver item 5).

Nos documentos com as contribuições para gestão foram relacionadas ações proativas em curso, aspectos críticos apontados e algumas propostas de ações para enfrentá-los. Nas contribuições para o processo da avaliação, foram relacionadas sugestões indicando quais questões do formulário

permitiram gerar importantes indicadores de qualidade, bem como críticas e sugestões sobre o formulário e sobre o processo de Avaliação Institucional da USP, que podem servir de contribuições para a futura Comissão Permanente de Avaliação. No geral, nos dois documentos, as sugestões foram organizadas de forma específica para gestão, graduação, pós-graduação, pesquisa, cultura e extensão e internacionalização.

Ao longo do desenvolvimento do 4º Ciclo de Avaliação Institucional, constatou-se a obtenção de contribuições muito relevantes para o processo, haja vista que permitiu uma ampla oportunidade para se discutir e refletir a respeito dos resultados obtidos, contemplados no Relatório Final da avaliação no item “Análise dos resultados da avaliação institucional baseado nos pareceres das comissões de assessores externos”, feito pela CPA, além de fornecer substrato para a ocorrência de processos avaliativos posteriores, valorizando a participação conjunta dos membros da comunidade acadêmica, chamados a participar, democraticamente, dos procedimentos de avaliação.

Em decorrência do exposto, foi possível perceber, a imprescindibilidade de que sejam proporcionadas condições necessárias para a ocorrência de um aperfeiçoamento contínuo, o qual depende da análise criteriosa e constante no âmbito das práticas institucionais.

5. PERSPECTIVAS FUTURAS

Pela importância que a USP dá para avaliação como uma ferramenta para a busca constante da melhoria da qualidade, e pela repercussão dos resultados dos processos de avaliação institucional na comunidade acadêmica para o planejamento de suas atividades-fim, o Conselho Universitário, na reunião de novembro de 2016, aprovou um novo regimento para a Comissão Permanente de Avaliação, formalizado pela Resolução 7272 de 23 de novembro de 2016 (USP, 2016b).

Pela nova abordagem da CPA, a avaliação docente será integrada à avaliação institucional. Dessa forma, os docentes serão apreciados num

panorama mais amplo, levando em conta os planejamentos dos seus respectivos Departamentos e Unidades. Além disso, o novo regimento consolida a dimensão institucional da avaliação.

A estrutura da Comissão ficou constituída por uma Comissão Plenária, presidida pelo Vice-Reitor da Universidade e duas Câmaras, uma de Avaliação Institucional e a outra de Atividades Docentes. Os membros das Câmaras estão sendo selecionados e no começo do segundo trimestre de 2017, as atividades da nova CPA serão iniciadas.

6. REFERÊNCIAS

BRASIL. Constituição (1988). *Constituição da República Federativa do Brasil*. Brasília, DF: Senado Federal: Centro Gráfico, 1988. 292 p.

BRASIL. Lei nº 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. *Diário Oficial [da República Federativa do Brasil]*, Brasília, DF, v. 134, n. 248, 23 dez. 1996. Seção I, p. 27834-27841.

GONZÁLES-GONZÁLEZ, J. Evaluación – planeación como instrumento de mejoramiento permanente del educación superior. In: ENCONTRO DE AVALIAÇÃO INSTITUCIONAL DA USP, 7., 2012, São Paulo.

TORO, J. R. Acreditación y aseguramiento de calidad. Revisión y algunos desafíos. In: REUNIÓN DE LA JUNTA DIRECTIVA DE CINDA, 45., 2012, Buenos Aires.

UNIVERSIDADE DE SÃO PAULO (USP). 2º Relatório da Avaliação Institucional USP 2000-2004. São Paulo, 2005.

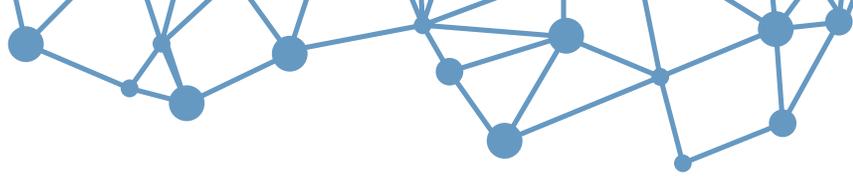
UNIVERSIDADE DE SÃO PAULO (USP). 3º Relatório da Avaliação Institucional USP 2005-2009. São Paulo, 2010.

UNIVERSIDADE DE SÃO PAULO (USP). 4º Relatório da Avaliação Institucional USP 2010-2014. São Paulo, 2016a.

UNIVERSIDADE DE SÃO PAULO (USP). Resolução 7272 de 23/11/2016 – Regimento da Comissão Permanente de Avaliação. São Paulo, 2016b.

AGRADECIMENTOS

Desejamos expressar os nossos agradecimentos às pessoas que fizeram parte do 4º Ciclo da Avaliação Institucional, inicialmente aos integrantes da Comissão Permanente de Avaliação que vem colaborando, de forma expressiva, ao longo dos últimos anos nessa tarefa, as Professoras Sonia Teresinha de Souza Penin (FE-USP) e Emma Otta (IP-USP) e os Professores Álvaro de Vita (FFLCH-USP), Fernando Luís Medina Mantelatto (FFCLRP-USP), Geraldo Duarte (FMRP-USP), José Alberto Cuminato (ICMC-USP), Marco Antonio Saidel (EP-USP), Rodney Garcia Rocha (FO-USP) e Rui Curi (ICB-USP); aos colaboradores que participaram na etapa final da análise dos resultados, as Professoras Eucia Beatriz Lopes Petean (FFCLRP-USP), Júlia Maria Matera (FMVZ-USP), Maria Aparecida de Andrade Moreira Machado (FOB-USP), Maria Cristina Motta de Toledo (EACH-USP), Maria Vitória Lopes Badra Bentley (FCFRP-USP), Silvana Mishima (EERP-USP), Wanda Maria Risso Gunther (FSP-USP) e os Professores Aluísio Augusto Cotrin Segurado (FM-USP), Luiz Gustavo Nussio (ESALQ-USP), Paulo José do Amaral Sobral (FZEA-USP), Tito José Bonagamba (IFSC-USP), e Valmor Alberto Augusto Tricoli (EEFE-USP); a Profa. Ângela Maria Magosso Takayanagui (EERP-USP) que foi responsável pela elaboração do material inicial, aos colaboradores que participaram da organização Cláudia Regina Pires, Edna Maria Brazolim e Mônica Jimenez (gabinete da vice-reitoria), Gisele Lopes Batista Pinto, Marino Hilário Catarino e Rafael Germano Rossi e Silvio Fernandes de Paula (técnicos do STI); e a todos os dirigentes USP, diretores de Unidades, chefes de departamentos, presidentes de comissões, professores(as), assistentes administrativos, secretárias(os), alunos(as) de graduação e pós-graduação.



INSTITUTIONAL ASSESSMENT IN USP

AVALIAÇÃO INSTITUCIONAL NA USP [ver página 111] [ir para o sumário]

*Pedro Vitoriano Oliveira** and *Vahan Agopyan***

I. INTRODUCTION

Assessment entails conducting a diagnosis based on the comparison between the goals that are to be achieved, confronting the previously established objectives with the effects obtained over a period of time, in order to recognize the potentialities and the critical points. It is an activity that allows a historical link, in which an attempt can be made to analyze the present time of the phenomenon – which is a reflection of choices and past practices – with a view to setting out a program that can be adapted to subsequent periods and possible changes.

Assessment is carried out to determine the existing features of a given object or situation, although this process does not end with the determination of these attributes. It is generally employed as the basis of other activities, whether in isolation or collectively, which are designed to achieve particular objectives.

It is considered as assessment of the Institute of Higher Education (IES) as being a transformative, shared and continuous activity with the basic aim of seeking to improve the quality of the different activities in which it is involved such as the following: teaching – at undergraduate and graduate levels

* Chemistry Institute, University of São Paulo; pvolivei@iq.usp.br

** Politécnica School, University of São Paulo

– research, culture and extension, national and international relations of institutions with their counterparts, and in the management of financial and human resources and all the responsibilities they take on, with a view to ensuring they can operate more effectively. It is a question of a self-critical and responsible method of proceeding to meet common needs and fulfil an institutional mission and should be an integral part of the planning of an institution. Finally, the intrinsic value of assessment is that it will create and strengthen a culture of quality within the institutional planning of the university.

In the IES, institutional assessment can take place by following four distinct stages (GONZÁLES-GONZÁLES, 2012):

- Self-assessment – which involves fostering both self-knowledge and recognition in so far as the assessed and the assessor are the same person and no subjects are included who do not form a part of the phenomenon that is being evaluated at that time.
- External assessment – this is carried out by specialists either from an area of knowledge or a particular subject-area that is based on the self-assessment that has been made previously. It is only applied within the sphere of academic programs and the results achieved from putting it into effect, should be solely used for the part of the program that is being evaluated.
- Accreditation – an evaluative modality through which the crediting body – in the sphere of higher education institutions represented by the Ministry of Culture (MEC) or the State Council of Education (CEE) – awards a certificate in assessment and quality assurance which is based on the self-assessment carried out by the Institution previously, as well as bestowing public trust on the academic excellence that is inherent in it.
- Professional certification – in which a professional body – such as the Order of Attorneys in Brazil (OAB), or the Federal Council of Medicine (CFM), among others, – is responsible for issuing assessment certificates for the main group of people from the IES – graduates. These are employed as parameters for analyzing the degree of information that

can bestow public trust on the academic and professional qualifications, and, in the case of the latter, the result of the educational performance is assessed and in this way applied to the classes of the graduates or those considered in isolation.

In the case of the University of São Paulo (USP) – and with regard to the way institutional assessment has been conducted in the course of time – there has been a greater tendency to make use of the Self-Assessment and External Assessment modalities. However, the professional certification required by the OAB and CFM, relies on the students USP from the undergraduate courses.

As regards the planning in the IES, this is essential both for the evaluative process and for an appropriate exercise of organizational powers. However, this does not mean that planning should be undertaken in a mechanical way or based on goals that have little prospect of being attained: it is essential to have periods of reflection and discussion from the time the guidelines that must be followed are laid down, (and to take account of the intervals of time needed to achieve the objectives). The purpose of this is to arrange the activities in a systematic way and take measures that can strengthen any positive factors (which are found out by means of the previous assessment), as well as to tackle any problems that may have been highlighted by the evaluative process. The two activities – assessment and planning – take place in an aggregated way with the aim of bringing about change, and run together in a continuous manner. When self-assessment and planning are interwoven like this, they are able to foster excellence in higher education. By means of these measures, it is possible to instigate and strengthen the parameters for embedding a culture based on an evaluation of quality within the institutional framework of universities. It is thus a question of adopting a procedure which seeks to ensure autonomy for universities. It also aims at encouraging self-criticism in a responsible way, for the benefit of all the participants in this environment and those who share the aspirations of the institution, since it is essential for this to be covered in the planning stage (GONZÁLES-GONZÁLES, 2012).

In the case of the National System of Higher Education – SINAES – when seeking to comply with the guidelines laid down by the National Committee for the Assessment of Higher Education – CONAES – (together with the National Institute of Research Studies “Anísio Teixeira”) – INEP – the evaluative processes must be combined to integrate the various powers involved in the real situation being assessed. This is to ensure consistency with regard to conceptual, epistemological and practical factors, as well as to make possible the achievement of excellence in the goals sought by means of the various expedients and strategies that are adopted.

2. HISTORICAL BACKGROUND

2.1. Context of Institutional Assessment: Brazil, State of São Paulo and USP

Article 209 of the Federal Constitution (BRASIL, 1988), dwells on the need for assessment procedures in the country’s learning institutions at all levels:

Art. 209: teaching is open to private enterprise, provided that the following conditions are met:

- 1. Compliance with the general rules of education*
- 2. Authorization and evaluation of quality by the government*

Thus it is the responsibility of the State to proceed with determining if the standards of quality are being met by the private IES which operate in the educational sphere. According to Art. 209, it is not incumbent on the State to assess the public IES, since this assessment is envisaged by the “Lei das Diretrizes e Bases da Educação Nacional” [Law setting out the Guidelines for the Foundation of National Education] (LDB).

Article 9 of Federal Law nº 9.394, of 20/12/1996 (BRASIL, 1996), known as the “Lei das Diretrizes e Bases da Educação Nacional” (LDB), stipulates two rules for assessment in Clauses VI and IX:

VI – it must ensure a national assessment of educational achievement at basic education, secondary school and higher education levels;
IX – it is the responsibility of the Federal Government to authorize, recognize, accredit, supervises and assesses courses and higher education institutions.

In the same act of legislation, Art. 46 states that:

The authorization and recognition of courses as well as the accreditation of higher education institutions must meet strict deadlines since they will be periodically renewed after the regular proceedings of assessment.

The assessment of Higher Education Institutions in the State of São Paulo was established by the State Council of Education on the basis of Resolution CEE 04/2000.

When a comparison was made between the regulations enacted by the federal and State authorities, no divergences were found between them. In accordance with SINAES and INEP, the Institutional Assessment was concerned with the following factors:

- Improving the quality of higher education
- Ensuring that its offer was expanded
- Constantly increasing its institutional efficacy and academic and social effectiveness
- Broadening the commitments and social responsibilities of the higher education institutions by recognizing the value of its public mission, fostering democratic values, respecting differences and diversity, and affirming autonomy and an institutional identity.

In the State of São Paulo, the rules were laid down by the State Council of Education. These sets out the objectives that will be achieved by means of Institutional Assessment:

1. To guide, monitor and inspect the universities and university centers in the State of São Paulo.
2. To periodically analyze the performance and institutional activities with regard to:
 - The effectiveness and efficiency of the teaching
 - The importance of research
 - The value of cultural and scientific output
 - The effectiveness of professional training
 - The importance of community action
 - The conditions for graduate and post-graduate studies
 - The quality of the management.

It should be stressed that to ensure that the need for evaluative procedures for a better performance in Higher Education Institutions was fully recognized, the Institutional Assessment was established in the sphere of USP before being standardized at a State level, (CEE 04/2000).

2.2. Institutional Assessment in USP

The University of São Paulo set in motion the procedures related to its institutional assessment in a systematic manner, by means of the cycles of the Departmental Assessment. This activity began when the Permanent Assessment Committee (CPA) was set up on 7/4/1992, (Resolution: 3920/92), in a proposal presented to the Committee of Academic Subjects (CAA) and the Special Committee of Work Regimes (CERT) of the USP.

Since that time, it was found that four assessment cycles (USP, 2005, 2010, 2016a) have been carried out since 2000, which correspond to the cycles of five years that followed (Figure 1), in compliance with the determination issued by the State Council of Education of the State of São Paulo, as stipulated in Art. 3^o of CEE 04/2000.

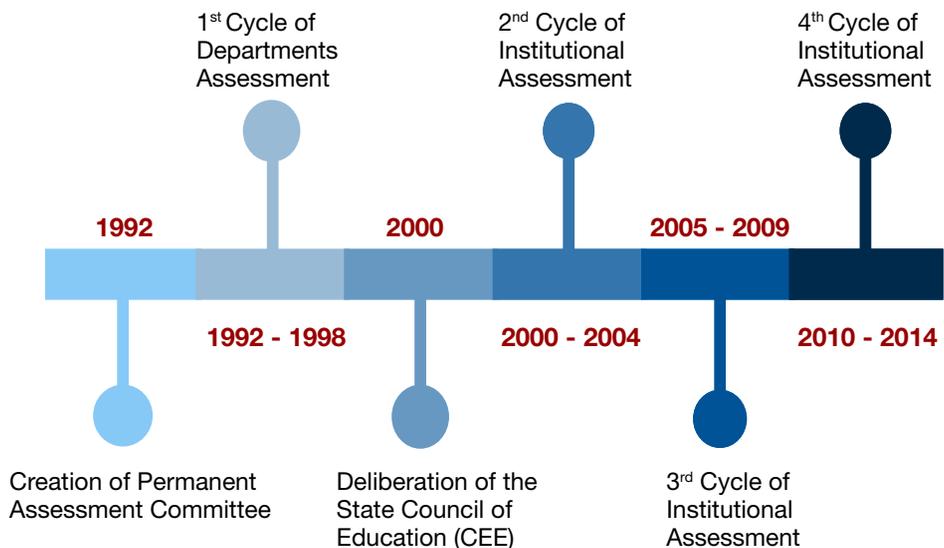


Figure 1. Historical background of the Institutional Assessment in USP since 1992.

In the 4th Cycle of Institutional Assessment in USP (USP, 2016a), which corresponds to the five-year period (2010 – 2014), it was decided to continue with the same pattern of past cycles, which is evidence of the importance attached to the existence of the practical procedure, as required within the sphere of the University of São Paulo. This involved employing an improved evaluative instrument called *Formulário de Avaliação* [the Assessment Form], which is one system for including specific questions for the Departments and another for the Faculties, which will be detailed in item 3.1. All the teaching faculties and their departments, institutes, specialist centers, museums and hospitals made use of this mechanism which was geared towards self-assessment.

There was an increase in the number of questions prepared in the assessment forms with the aim of integrating them with the requirements of other evaluative processes of USP, such as the Special Committee of Work Practices (CERT), the Central Assessment Committee for Advances in Teaching Careers (CCAD), the USP Agency for National and International

Cooperation (AUCANI), the USP Agency of Innovation (AUSPIN) and the Rectors (Graduate and Post-Graduate Studies, Research, Culture and Extension). These reflect the concern to establish a culture in support of quality based on the full and systematic involvement of all the members of the academic and administrative community of USP.

These members took part in forming the Cycle of Institutional Assessment (USP), as well as the 10 components that constitute the CPA and include those that are responsible for the academic and administrative sector, together with three members of the Vice-Rector's committee (for technical assessment and administrative support) and the Supervision of Information Technology (STI) of USP. The contributions made by these participants will be listed below in the item on adopted procedures. 189 external assessors were involved in the USP framework, who carried out the evaluation of 42 teaching faculties, 6 specialist centers and institutes, 4 university museums and 2 hospitals, making a total of 54 Faculties, as well as the 216 departments in the teaching faculties.

A list of suggested national and international researcher names that could be on the Committee of External Advisors was prepared by the Faculty and submitted to the CPA for evaluation. In this stage, the technical profile and the degree of relationship of the person with the Faculty/Departments were considered. Researchers with involvement in research projects with researchers from the Faculty/Departments were not recommended. At the end, after evaluating the suggestions of the Faculty, the CPA indicated the list of names to be considered as holders and substitutes to compose the Committee of External Advisors

Table 1 shows the evolving pattern of the number of faculties and departments that took part in the different USP evaluative cycles, as well as the external assessors who were also involved.

The variations in the numbers of Departments and Faculties are due to the joints and incorporations of new Centers to the University of São Paulo. In particular, in this 4th Cycle of Institutional Assessment (USP, 2016a), the

Centers and Specialized Institutes, Museums and Hospitals were considered as independent Faculties.

Table 1. Institutional Assessment of USP in Numbers (1992 – 2014)

INSTITUTIONAL ASSESSMENT	1 ST CYCLE 1992-1998	2 ND CYCLE 2000-2004	3 RD CYCLE 2004-2009	4 TH CYCLE 2010-2014
DEPARTMENTS	210	199	207	216
FACULTIES	35	36	39	54
TOTAL NUMBER OF EXTERNAL ASSESSORS	421	287	152	183*
a) National External Assessors	289	156	107	122
b) International External Assessors	132	131	45	61

* 7 faculties were visited by two committees

3. PROCEDURES FOLLOWED IN THE 4TH CYCLE OF INSTITUTIONAL ASSESSMENT IN USP

The 4th Cycle of Institutional Assessment (USP, 2016a) was carried out in 5 distinct stages:

- First – this consisted of coordinated visits to the USP campuses to examine and supervise the Institutional Assessment process and involved providing information about the operational system for the self-assessment of the faculties, departments, specialist institutes and centers, museums and hospitals. The visits took place on three different dates at the end of 2014, immediately before the preparation of the assessment forms that had to be completed by the faculties and departments. The directors of

faculties, heads of departments, chairmen of the committees and members of the administrative sectors, took part in the discussions.

- Second – the assessment forms were completed by the departments and faculties. The procedure was arranged in a way that allowed the forms to be completed in a coordinated and synchronized way. They were first filled by the departments so that the directors of faculties could summarize the main factors related to their departments (as well as to management and infrastructure) and include them in the forms of their faculties. The departments registered the self-assessment forms electronically so that the directors could have access to them and then organize the self-assessment of their Faculty, which was also registered electronically. These activities were completed up to 30 days before the visit of the External Assessment Committee which meant that the members had time to analyze all the material before their visit to the Faculty.
- Third – visits were paid by the External Assessment Committee to the USP Faculties and these took place in a coordinated manner between June and November 2015. The members of the committees which comprised teachers/researchers outside the USP framework, had some links with the activities carried out in the Faculty and/or Departments.
- Fourth – this encompassed the assessment of core subjects and was carried out by the Senior Assessment Committee on a global scale. In this stage, six researchers were invited to give their opinions about undergraduate and graduate in USP, research, culture and extension, together with internationalization and management. This was undertaken on the basis of the self-assessment forms of the Faculties and Departments, the opinions expressed by the External Assessment Committees based on the self-assessment evaluation and the interviews with the members of the Faculties (directors of studies, heads of departments, teachers, researchers, undergraduates and graduates students and the administrative staff).
- Fifth – preparation of the final report.

3.1. On the self-assessment form

The forms used for the self-assessment of the Departments and Faculties in the 4th Cycle of Institutional Assessment USP were grounded on three key areas:

- A set of goals (mission, vision and educational planning);
- Self-assessment (management, coordination, infrastructure, technical/administrative services, the teaching staff, teaching/learning processes, undergraduate and graduate, research, culture and extension, and internationalization);
- Institutional planning with predicted goals and future activities.

The mission is the final objective and is expressed in the present tense and determines all the decisions and educational activities of a university organization. In higher education it describes the duties and nature of the institution, the values that endow it with life, the needs it seeks to satisfy and the goals it establishes. This is its declared purpose and is why the Faculty is its *raison d'être*. It comprises a set of declarations that capture the essence of the institution. The mission should comprise three key features:

- A declaration of the principals and values of the Institution – the Faculty embodies the principles and values which are in accordance with the reasons for its foundation or the circumstances that gave rise to its origin. This kind of declaration of general principles is not put into effect by fulfilling a set of objectives but it is important for the members of the academic community to be aware that the assessors understand the broader context of the values underlying the work of the institution.
- Academic characterization of the Institution – the reference-point is the academic/educational identity of the institution that must be made clear in the mission. A number of factors including programs, teaching methodologies, levels of study, relations with society and the goals of research converge in the Faculty. None of these is neutral and the way

each institution positions itself, determines its mission. These features form the central part of the mission of an educational body. The following are some examples of what characterizes an academic body: i) Institutions with distinctly professional programs in contrast with others that attach importance to general skills; ii) Institutions that give prominence to mechanisms of active and autonomous learning, while others are more conservative in pedagogical areas; iii) There are Institutions in which the graduate and post-graduate courses play an equivalent role whereas others are restricted to post-graduate studies; iv) Investigative research is an essential part of some Institutions, while others are only concerned with teaching.

- The relation of the Institution with its social context – a mission that is strictly based on its ability to make adjustments to its declared objectives cannot ignore the context in which it is embedded and within which its graduates operate. Education is a public service and as such, has the responsibility to address the needs of society. An appropriate way must be found for each institution to fulfil this obligation and this must be reflected in its mission. Some examples of how it can be related to its social context are as follows: i) If an institution provides qualifications to its students and finds out which of the graduates are successful in the job market, this is an important social exchange. ii) There is a clear requirement on the part of an institution to take on the responsibility (through its courses, research activities and services) to play an active role in analyzing and tackling problems in its particular region or country. As a result of the programs on offer, it is possible to be more selective, whereas other institutions may only encourage inclusion.

Setting out from the mission, the “vision” describes the situation that the university organization seeks to achieve in the near future. It is an exercise that, within the conditions that prevail now and are expected in the future, requires thought about how the Faculty can best interpret what it should do to fulfil its own mission. This interpretation is a strategic exercise

and not simply an operational activity. It requires a knowledge of the challenges facing the Faculty (TORO, 2012). Changes of vision can be caused by different circumstances such as:

- A tendency towards mass education and the need for more inclusive institutions that can meet this increasing demand;
- Local and global socioeconomic conditions within which the Institution operates and its tendencies;
- Changes in areas of knowledge and their growing relationships;
- Current trends in education such as the evolving pattern of new skills expected of graduates and new forms of pedagogical methodology, as well as the use of technological resources needed to implement them;
- The local needs of research versus global trends.

By insisting on the adjective 'strategic', an attempt is made to make plans for the Institution in the future setting within which it will operate. This leads to making choices about the paths to follow in the long term to ensure it will be successful – while taking account of its mission and environment. The vision which entails making complex decisions and institutional changes, means taking risks, pledging resources and undertaking new operations. All this does not correspond with what currently takes place in the everyday world and for this reason is of a strategic nature (TORO, 2012).

In summary, it can be said that the purpose of the planning, institutional management, assessment and introduction of improvements, is to take the institution from a time in the present where it has an established mission, to a vision that is planned for the medium and long term.

In the self-assessment axis, when preparing their answers, the Faculties and Departments took account of the goals set out by the Faculties and Departments on the occasion of the 3rd Assessment Cycle (USP, 2010), as well as by the declarations of the external assessors involved.

Specific schedules were arranged for the self-assessment of the Faculties and Departments with the aim of ensuring there was a relative degree of

uniformity with regard to the information. These formed the basis of the evaluative process and mainly covered qualitative factors. This information was part of the Guide of Institutional Assessment in USP 2010-2014, distributed to all Departments and Faculties.

In addition, the 4th Cycle of Institutional Assessment (USP, 2016a) was characterized by the inclusion of some differentials with regard to what had previously been carried out. These included the completion of the forms in a bilingual publication (with versions in Portuguese and English), in a way that allowed international assessors to have access to the information provided by the Faculties and Departments in the most effective manner and to benefit from the transparency and necessary broadening of access to information about this issue.

3.2. Organization of the Institutional Assessment procedure

A computerization system to assist in the Institutional Assessment was designed by the Supervision of Information Technology – STI. This was to enable the forms to be completed online and the opinions of the external assessors to be registered, as well as to handle all the communications, submissions and analysis of the information. A field was opened up in each Faculty where it was possible to register the following: the reports of the cycles of previous assessments, tables, spreadsheets and graphs retrieved from the databases of the corporate systems of USP, together with information about the activities in the Faculties and their Departments. This was so they could be consulted and compared with other self-assessments in previous cycles. In view of this, the data from the last USP assessments registered in the USP system (together with the historic register for future Assessment Committees) will constitute a valuable source of material for consultation for the directors of the University and faculties and make it possible to influence them and offer guidelines for their decision-making.

The External Assessment Committees were established by external teachers and researchers in the USP frameworks; a schedule of suggestions

was planned with the aim of putting their work in systematic order and this was made available in the site so that it could be used by the External Assessors before and during their visits to the Faculties. This schedule contained the following suggestions:

- To summarize the academic progress of the Faculty on the basis of the self-assessment report and other existing information and indicate points that: i) could be regarded as appropriate; ii) that are worth highlighting; and, iii) that need improvement and suggestions about how the Faculty can take measures to raise the standard of its academic activities.
- To explain the situation of Graduate and Post-Graduate Studies on the basis of the objective indicators available and in the ongoing projects. As well as this to establish links between core subjects, interdisciplinarity, the stage when core subjects are at an international level, the internal and external links with the Faculty and the inner asymmetries (inequalities) that are revealed when undertaking the core subjects.
- To analyze the objectivity of the academic goals put forward and their agreement with the mission of the Faculty/Department and with the stage of academic progress made by the Faculty.
- To comment on the observations made during the interviews with the group of students (Undergraduate and Graduate level) about the academic activities carried out in the Faculty/Department/Graduate Program.
- To analyze the planning and academic and administrative management of the Faculty and the Departments.
- To analyze the situation regarding the facilities in the Faculty/Department with regard to administrative organization, which also encompasses human resources, materials, a library and information and innovation technology.
- To comment on the interaction of the Faculty with other similar institutions and their relations with other sectors of civil and organized society (industry, government, etc.), in the national or international sphere and their academic, economic and/or social impact.

- If appropriate, to make suggestions and/or recommendations to the administrative section of the Faculties, and Departments of the University and the Permanent Assessment Committee (CPA).

At the end of the visits, each Committee set out its opinion which included positive points, criticisms and suggestions – and these were registered in the evaluation site.

The objective of the Committee of Senior Assessors, which comprised 6 members, was to assess the core subjects in a global way by analyzing the performance of all the Faculties and the respective activities that were within their area of responsibility. The members of this Committee of Seniors were requested to take account of the following factors in their assessments: i) could they be regarded as suitable; ii) do they deserve to stand out; and, iii) what needs improving or what suggestions should be made about how USP can take steps to enhance their quality. In addition, they were encouraged in a specific way to take account of other factors in evaluating the activities and those in charge of them, based on the indicators requested in the self-assessment forms of the Departments and Units (available at <https://gvr.uspdigital.usp.br/cpa>):

- Graduate and Post-Graduate Students – to assess graduate and post-graduate studies in USP, on the basis of the available objective indicators and ongoing projects, as well as combining them with other goals – interdisciplinarity, the stage of internationalization, the internal and external linking of the Faculty and the inner asymmetries involved when undertaking the activity.
- Research, Culture and Extension – to assess general aspects of research, culture and extension in USP, on the basis of interactions between the Faculties and other similar institutions and their relations with other sectors of civil and organized society (industry, government, etc.), in national and international spheres; and to measure their academic, economic and/or social impact.

- International Cooperation – to analyze the internationalization of USP on the basis of the information and statistical data supplied; the mobility of graduate and post-graduate studies, participation in international academic networks, projects involving international cooperation, performance indicators, etc. Criticisms and suggestions are welcome on the question of how the USP is carrying out activities for greater internationalization.
- Administration – to analyze the planning and academic/administrative management of the Faculties, the infrastructural facilities of the Faculties with regard to administrative organization, by encompassing all the human resources, materials, library, information technology and innovation.

4. RESULTS OF THE USP INSTITUTIONAL ASSESSMENT

The concept adopted by the USP is that assessment is a continuous process involving a planned evaluation. The planning follows an assessment of successes and failures with regard to previously defined goals and the strong and weak points of the Departments and Faculties. In the search for excellence, strategic planning constitutes an important instrument that can show the best path to follow. Thus, it can be understood that the institutional assessment procedures of the University of São Paulo can assist in this task and on the basis of a knowledge of its situation, can: i) point out the best way to understand and make better use of its strong points; ii) learn or get rid of or adapt to its weak points; iii) enjoy the benefits of outside opportunities and iv) devise an effective work plan.

In recent years, Institutional Assessment has been an important management tool for the directors. It supports decision-making and assists in monitoring and planning the Faculty and University with the commitment of the Principal and the Faculties, and enables them to take measures aimed at achieving institutional goals. There is a tradition behind the assessment process which has undoubtedly led to excellence at the University of São Paulo.

4.1. Discussions of the results of the 4th Cycle of Institutional Assessment in USP

The Report was a paper that was drawn up and sent to the State Council of Education of the State of São Paulo (CEE) as a part of the requirements for the 5-year assessment of the Higher Education Institutions. As well as bringing together the opinions expressed by the Senior Assessors, it addressed the investigative inquiry carried out by the Permanent Assessment Committee which was based on the opinions formed by the External Assessment Committees who visited the USP Faculties. This Report, together with all the material produced during the 4th Cycle of Institutional Assessment (USP, 2016a), is registered in the following site: <https://uspdigital.usp.br/wsusuario/>.

After the activities of the 4th Cycle of Institutional Assessment (USP, 2016a) had been completed, the challenge that was posed was to make changes in them rather than regarding them as a task that had to be fulfilled solely to meet legal requirements. This provided an opportunity to bring together key indicators for various aspects of the work of USP, in so far as it became a feature in the management and improvement of this Institution. From this standpoint, a study group was set up for the Analysis of Institutional Assessment (GT) which, together with the participants from different USP campuses, sought to encourage discussion (within the domain of the teaching faculties and their departments, institutes, specialist centers, museums and hospitals) on the results obtained from the 4th Cycle Institutional Assessment of USP (2010-2014). Its ultimate aim was to improve the process and put it into effect, as well as to allow a full period of reflection on the results of the assessment, in an attempt to achieve the degree of quality and excellence desired by the Institution.

In the schedules for overseeing the work for the External Assessment Committees, it was envisaged that the ongoing proactive activities in the Faculties and Departments would be identified, as well as critical comments and suggestions for taking steps to overcome these challenges. The

committees followed the recommendations and discovered many important activities that are being planned (or on the way to being put into practice) in the Faculties and which are highlighted in their opinions. In the same way, certain points were found that may make it difficult to improve the core subjects in the Faculties and in some cases, alternatives were suggested to circumvent these problems, such as curricular restructuring, support for extracurricular activities, evasion control, among others.

The Faculties were requested to locate these factors recorded in the opinions of the external assessors and draw up a concise paper to discriminate between the proactive activities that are worth highlighting, from the critical points and the activities given prominence for the improvement of management, undergraduate and post-graduate students, research, culture and extension and internationalization. Apart from the factors pointed out by the External Assessment Committees, in a complementary exercise of self-evaluation, the Faculties identified some positive features as well as the criticized factors, and advocated an alternative form of management, both for the Faculty and for the USP as a whole, as a means of getting round these difficulties. The suggestions were specific, aiming in particular at the needs of the Faculties or their Departments. In some cases, definitions of goals and follow-up of their fulfillment were highlights.

In addition, the Faculties were requested to make criticisms and suggestions for improving the assessment process and putting it into effect by suggesting: i) a means of determining the best and worst indicators of the assessment form; ii) making comments and recommendations about the framework of the form; and iii) making comments and recommendations about the Institutional Assessment procedure (filling in the form, visits from the External Assessment Committee, the composition of the Committees and the Senior Assessment Committee, among other factors).

The meetings were held within the sphere of the Assemblies or CTAs of the Faculties and were attended by the GT representatives. The discussions were based on a) the self-evaluation forms of the Departments and Faculties,

b) the opinions expressed by the External Assessment Committees and Senior Assessors and c) the contents of the Final Assessment Report.

The activities related to the analysis and discussion of the results of the 4th Cycle of Institutional Assessment (USP, 2016a), took place between April and June 2016, – at first in the Assemblies or CTAs of the 54 Faculties together with the GT representatives. In giving their results, the Faculties prepared papers that summarized the requests mentioned above and sent them to the GT. Following this, the Faculties appointed representatives so that the ideas could be discussed in 7 Workshops, and these were held between August and September 2016.

Running in parallel with this, all the contributions sent to the 54 Faculties of USP gave rise to two summarized documents one regarding the assistance to management and the other for the Institutional Assessment which was sent on to the new CPA (see Item 5).

In the documents for management, there were reports about ongoing proactive activities, critical comments, and some suggestions of measures for confronting them. With regard to the evaluative process, suggestions were made with regard to which questions in the form could lead to key indicators of quality, as well as criticisms and suggestions about the form and about the USP Institutional Assessment which could act as guidelines for a future Permanent Assessment Committee. In general, the suggestions in the two documents were arranged in a specific way for management, graduate and post-graduate studies, research, culture and extension and internationalization.

In the course of the 4th Cycle of Institutional Assessment, it was confirmed that very useful contributions had been made to the process, since it allowed ample opportunity for discussion and reflection on the results obtained, all registered in the Final Report of the evaluation in the item “Analysis of the results of the institutional evaluation based on the opinions of the committees of external advisors”, made by the CPA. It also provided a firm basis for subsequent evaluative processes to take place and recognized the

value of including members of the academic community who were invited to take part in the assessment procedures in a democratic way.

As a result of this overview, it can be seen that it is essential that the right conditions should be established for a continuous improvement and that this depends on a rigorous and constant analysis within the realm of institutional practices.

5. FUTURE PERSPECTIVES

At a meeting held in November 2016, the University Council approved a new set of regulations for the Permanent Assessment Committee of 23rd November 2016 (USP, 2016b). This was to underline the importance attached by the USP to assessment as a tool for the constant search for an improvement of quality and reflects the repercussions felt by the results of the institutional assessment among the academic community when planning its courses for core subjects.

In adopting a new approach to CPA, the evaluation of teaching will be integrated with institutional assessment. In this way, the value of teachers will be recognized within a wider sphere and account will be taken of the planning of their respective Departments and Faculties. As well as this, the new regulations will consolidate the institutional dimension of assessment.

The framework of the Committee will be established by a Plenary Session presided over by the Vice-Rector of the University and two corporations – one of Institutional Assessment and the other of teaching activities. The members of the Corporations are currently being selected and the activities of the new CPA will get underway at the beginning of the second quarter of 2017.

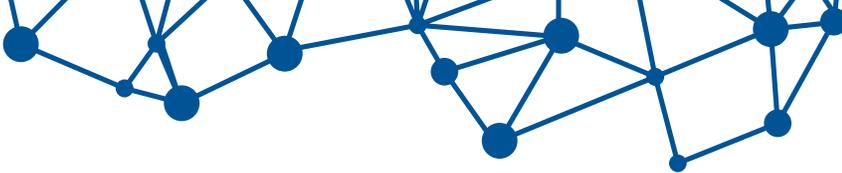
6. REFERENCES

- BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**. Brasília, DF: Senado Federal: Centro Gráfico, 1988. 292 p. [Constitution of the Federal Republic of Brazil]
- BRASIL. Lei nº 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. **Diário Oficial [da República Federativa do Brasil]**, Brasília, DF, v. 134, n. 248, 23 dez. 1996. Seção I, p. 27834-27841. [Law setting out the Guidelines for the Foundation of National Education]
- GONZÁLES-GONZÁLEZ, J. Evaluación – planeación como instrumento de mejoramiento permanente del educación superior . In: ENCONTRO DE AVALIAÇÃO INSTITUCIONAL DA USP, 7., 2012, São Paulo. [Assessment - planning as a means of making a permanent improvement in Higher Education; 7th Meeting of Institutional Assessment at University of São Paulo]
- TORO, J. R. Acreditación y aseguramiento de calidad. Revisión y algunos desafíos. In: REUNIÓN DE LA JUNTA DIRECTIVA DE CINDA, 45., 2012, Buenos Aires. [Accreditation and ensuring quality. A review and some challenges]
- UNIVERSIDADE DE SÃO PAULO (USP). 2º Relatório da Avaliação Institucional USP 2000-2004. São Paulo, 2005. [2nd Institutional Assessment Report]
- UNIVERSIDADE DE SÃO PAULO (USP). 3º Relatório da Avaliação Institucional USP 2005-2009. São Paulo, 2010.
- UNIVERSIDADE DE SÃO PAULO (USP). 4º Relatório da Avaliação Institucional USP 2010-2014. São Paulo, 2016a.
- UNIVERSIDADE DE SÃO PAULO (USP). Resolução 7272 de 23/11/2016 – Regimento da Comissão Permanente de Avaliação. São Paulo, 2016b. [Resolution 7272, 23/11/2016 – Regulations of the Permanent Assessment Committee]

ACKNOWLEDGMENT

We would like to express our gratitude to the personnel of the 4th Cycle of Institutional assessment, and first mention the members of the Permanent Assessment Committee who have done significant work on this task in recent years. These include the following Professors: Sonia Teresinha de Souza Penin (FE-USP) and Professor Emma Otta (IP-USP), Álvaro de Vita (FFLCH-USP), Fernando Luís Medina Mantelatto (FFCLRP-USP), Geraldo Duarte (FMRP-USP), José Alberto Cuminato (ICMC-USP), Marco Antonio Saidel (EP-USP), Rodney Garcia Rocha (FO-USP) and Rui Curi (ICB-USP). We would also like to thank their female colleagues who took part in the final analysis of the results: Eucia Beatriz Lopes Petean (FFCLRP-USP), Júlia Maria Matera (FMVZ-USP), Maria Aparecida de Andrade Moreira Machado (FOB-USP), Maria Cristina Motta de Toledo (EACH-USP), Maria Vitória Lopes Badra Bentley (FCFRP-USP), Silvana Mishima (EERP-USP), Wanda Maria Risso Gunther (FSP-USP); and their male colleagues Professors Aluísio Augusto Cotrin Segurado (FM-USP), Luiz Gustavo Nussio (ESALQ-USP), Paulo José do Amaral Sobral (FZEA-USP), Tito José Bonagamba (IFSC-USP), and Valmor Alberto Augusto Tricoli (EEFE-USP); Prof. Ângela Maria Magosso Takayanagui (EERP-USP) who was responsible for preparing the initial material, and other members of staff who were involved in the organization: Cláudia Regina Pires, Edna Maria Brazolim and Mônica Jimenez (from the Office of the Vice-Rector), Gisele Lopes Batista Pinto, Marino Hilário Catarino and Rafael Germano Rossi and Silvio Fernandes de Paula (STI technical staff); and all the USP directors, directors of faculties, heads of departments, chairmen of committees, teachers, assistant administrators, secretaries and graduate or post-graduate students.





POLÍTICAS PÚBLICAS EM CIÊNCIA E TECNOLOGIA NO BRASIL: DESAFIOS E PROPOSTAS PARA UTILIZAÇÃO DE INDICADORES NA AVALIAÇÃO

PUBLIC POLICIES IN SCIENCE AND TECHNOLOGY IN BRAZIL: CHALLENGES AND PROPOSALS FOR
THE USE OF INDICATORS IN EVALUATION [*see page 189*] [*go to summary*]

**Talita Moreira de Oliveira e **Livio Amaral*

INTRODUÇÃO¹

O sistema de avaliação da pós-graduação brasileira foi implantado pela CAPES em 1976. A agência é responsável pela recomendação de novos programas de pós-graduação e pela avaliação periódica de desempenho dos mesmos, o que, conseqüentemente, garante a manutenção dos programas no Sistema Nacional de Pós-Graduação (SNPG) com rigorosos padrões de qualidade. Esse processo, desde o início, sempre foi realizado com significativa participação da comunidade acadêmica, científica e tecnológica por meio de comissões de avaliação, atualmente divididas em 49 áreas. O processo de avaliação tem sido a base para o fomento e incentivo à

* CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; talita.oliveira@capes.gov.br

** Departamento de Física, Universidade Federal do Rio Grande do Sul

1 Os autores agradecem ao Dr. Jorge Almeida Guimarães pela leitura crítica do manuscrito.

pós-graduação e pesquisa, não apenas pela CAPES, mas também como um indicador para outras agências governamentais.

Os principais fundamentos da avaliação da CAPES residem em três principais pontos, que são: i) a qualidade assegurada pela análise dos pares, ii) a consulta e o debate constantes com a comunidade acadêmica, científica e tecnológica para definição e atualização dos critérios de avaliação e iii) a ampla transparência dos procedimentos e dos resultados. É um sistema que exige constante atualização e flexibilidade a ponto de abarcar a crescente expansão do número de programas e a evolução de suas características.

A expansão e consolidação da pós-graduação *stricto sensu* (mestrado e doutorado) são fundamentais para garantir a formação de pessoal qualificado para atuar no setor produtivo e nas universidades e promover o desenvolvimento socioeconômico do país. As políticas nacionais de educação e de ciência e tecnologia permitiram o avanço da pós-graduação e da produção científica brasileira.

Os indicadores funcionam como instrumento agregador e orientador, permitindo simulações e elaboração de cenários de forma facilitada, frente a um universo complexo que envolve grande quantidade de dados, fontes externas de informações e diversas variáveis envolvidas.

Atualmente, as ações da CAPES envolvem, além da avaliação da pós-graduação *stricto sensu*, o fomento a formação de recursos humanos de alto nível no país e no exterior, a promoção da cooperação científica internacional, o acesso e divulgação da produção científica e a indução e fomento a formação inicial e continuada de professores para a educação básica nos formatos presencial e a distância.

A AVALIAÇÃO DA PÓS-GRADUAÇÃO BRASILEIRA

A avaliação da pós-graduação *stricto sensu* tem como objetivo analisar detalhadamente o panorama e as atividades deste nível de ensino no Brasil, produzindo estudos e indicadores que fundamentam políticas governamentais

de apoio e crescimento da pós-graduação, sendo atualmente o mais importante instrumento para o fomento, tanto pela própria CAPES quanto por outras agências. Por exemplo, subsidia na identificação de áreas estratégicas para a ciência e tecnologia do país, assim como na indução de redução de assimetrias regionais e na expansão de programas de pós-graduação.

O sistema de avaliação se divide em dois macro-processos, o de entrada de novos cursos, por meio da avaliação de propostas submetidas por instituições de ensino e pesquisa e o de permanência, por meio da avaliação periódica dos programas (Figura 1).

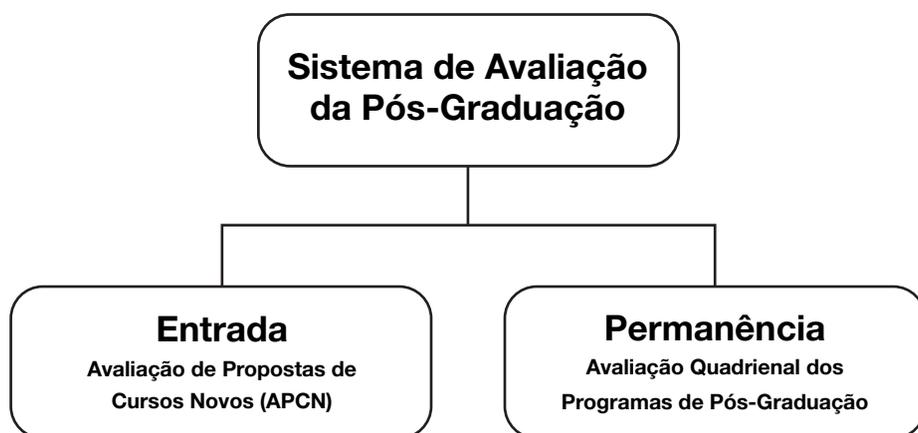


Figura 1. Macroprocessos do sistema de avaliação da pós-graduação

Fonte: CAPES

Todo programa de pós-graduação precisa, para início e continuidade de funcionamento, obrigatoriamente passar pela avaliação realizada pela CAPES. O sistema de avaliação da pós-graduação brasileira é baseado na análise feita pelos pares. Cada área de avaliação é liderada por um coordenador, um coordenador adjunto de programas acadêmicos e um coordenador adjunto de programas profissionais e composta por consultores *ad-hoc* com destacada experiência de ensino e pesquisa em sua área de especialidade. Os consultores que participam das comissões são renovados a cada atividade

avaliativa em pelo menos 50%, o que contribui também com a minimização dos vícios de avaliação.

As comissões de área são formadas de acordo com o tipo de atividade avaliativa a ser desempenhada, que podem ser:

- i. **Análise de propostas de cursos novos:** As instituições interessadas em oferecer um curso de pós-graduação submetem uma proposta à CAPES, que passa pela análise de mérito das comissões de área e pela aprovação do Conselho Técnico-Científico da Educação Superior (CTC-ES), que por sua vez podem ou não recomendá-la para fins de reconhecimento pelo Ministério da Educação.
- ii. **Classificação de produtos técnicos e científicos:** Os produtos resultantes das atividades dos programas passam por um processo de avaliação para fins de aferir sua qualidade. São feitas classificações de periódicos, livros, de eventos educacionais, científicos e tecnológicos e de produtos técnicos. Cada um destes produtos tem sua própria e independente escala de classificação, e não são estabelecidas regras de correspondência ou equivalência entre elas.
- iii. **Seminários de Acompanhamento:** Encontros anuais regulares para discussão de critérios e indicadores da área e acompanhamento do desempenho dos programas em conjunto com a comunidade acadêmica.
- iv. **Avaliação Periódica:** Os programas passam por ciclos avaliativos, atualmente em intervalos de quatro anos, com fins de avaliar seu desempenho em termos de atividades de ensino, pesquisa e extensão e de formação de alunos. São atribuídas notas no intervalo de 1 a 7, sendo que aqueles que recebem 1 ou 2 são descredenciados do sistema. Nota 3 corresponde ao mínimo padrão de qualidade exigido. Notas 6 e 7 representam programas de excelência, comparáveis a nível internacional.

A dimensão do SNPG dá uma ideia do desafio que se tem em avaliar um universo crescente de cursos. Segundo Ferreira e Moreira (2003), em 1976, o número de doutores titulados era de apenas 188. Daquele ano até

os dias atuais, a pós-graduação mantém um ritmo elevado de crescimento. Entre 1998 e 2014 (Figura 2), o número de cursos cresceu 172%, possuindo neste último ano 2016 doutorados, 3130 mestrados e 540 mestrados profissionais, contendo cerca de 230 mil matriculados em pós-graduação e mais de 69 mil titulados.

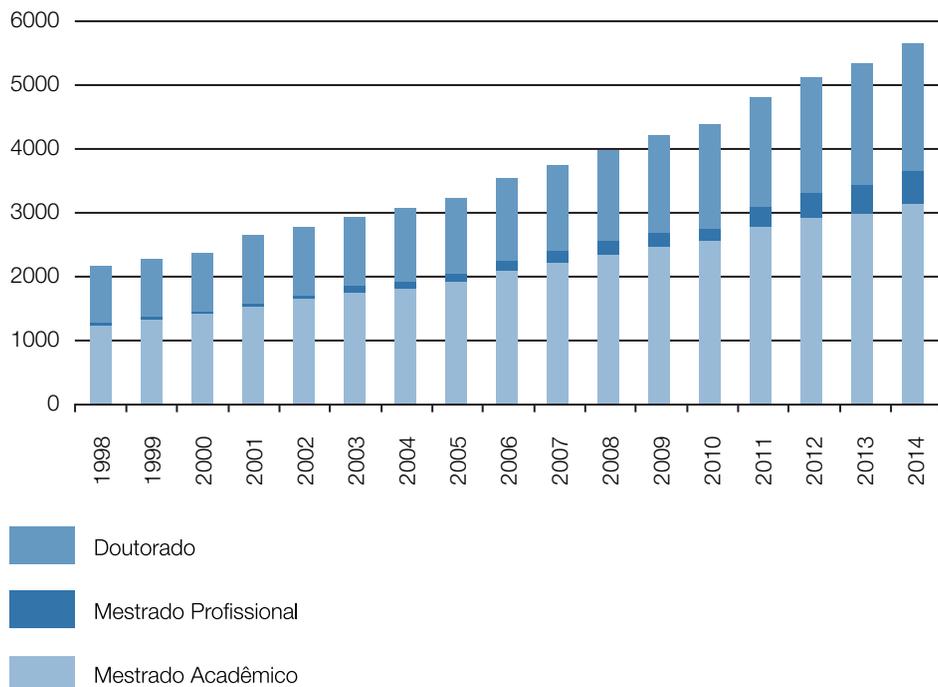


Figura 2. Evolução do número de cursos de pós-graduação

Fonte: Geocapes. Elaboração CAPES

CRITÉRIOS E INDICADORES DE AVALIAÇÃO

Os Manifestos Internacionais

O uso de indicadores de desempenho e métricas para avaliação da pesquisa e produção científica tem sido objeto de intensa discussão e crescente manifestação da comunidade científica mundial. Recentemente, três importantes declarações internacionais foram feitas a respeito desse assunto.

Em 2012, um grupo de editores e divulgadores de periódicos científicos, durante a reunião anual da Sociedade Americana de Biologia Celular publicou a *Declaração de San Francisco sobre Avaliação da Pesquisa* (DORA, 2012) <http://www.ascb.org/dora/>, contendo uma série de recomendações para a melhoria da forma de avaliação da produção científica. Vários outros interessados, como agências financiadoras, instituições e pesquisadores, também assinaram a declaração, mostrando seu compromisso em realizar uma avaliação mais apurada, focada primordialmente na qualidade e no impacto do que meramente em quantidade e em métricas indiretas dos periódicos de publicação.

No ano de 2015, foi divulgado o *Manifesto de Leiden* sobre métricas de pesquisa, como resultado das discussões ocorridas durante a Conferência Internacional de Indicadores em Ciência e Tecnologia (STI 2014), em Leiden. Propôs dez princípios de boas práticas para a medição do desempenho da pesquisa. Uma das preocupações colocadas no manifesto é que as avaliações têm se baseado cada vez mais em métricas e menos em uma avaliação mais criteriosa feita por especialistas (HICKS et al., 2015).

No mesmo ano de 2015, houve a publicação do *The Metric Tide (A maré de métricas)*, relatório a respeito do papel das métricas na avaliação e gestão da pesquisa, encomendado pelo Higher Education Funding Council for England (Hefce), agência financiadora da Inglaterra (WILSDON et al., 2015). O documento aborda os usos e limitações de métricas de pesquisa e o seu potencial de contribuição para o desenvolvimento e o impacto da pesquisa de excelência. Aponta cinco princípios do que seriam as “métricas responsáveis”, que pressupõem o uso adequado de indicadores quantitativos.

Em síntese, a grande ênfase dessas declarações reside no fato de que as métricas não devem ser usadas de forma indiscriminada ou meramente contábil, sem se atentar para suas limitações. Além disso, elas não devem ser indicadores únicos para avaliação, sobrepondo a análise de especialistas.

A seguir, foi feita uma análise comparativa dos princípios para o bom uso das métricas declarados pelos manifestos internacionais e as práticas de avaliação utilizadas pela CAPES no processo de avaliação dos programas de pós-graduação.

Os princípios de avaliação da CAPES à luz dos manifestos internacionais

1. Indicadores: importantes, mas não soberanos

O princípio fundamental da avaliação é que ela deve contar com a análise criteriosa de grupos de especialistas que tenham propriedade técnica, qualificação científico-acadêmica e vivência experimental para analisar o trabalho feito por seus pares cientistas e pesquisadores.

Porém, considerando o grande e crescente volume de informações disponíveis, não há como realizar uma análise minuciosa e individual de cada produto da pós-graduação. Os indicadores aparecem como ferramentas auxiliares no processo de avaliação de forma a sinalizar tendências, apontar grupos medianos ou que se destacam de forma negativa ou positiva.

Os manifestos mostram a importância de se praticar um cenário de complementariedade entre uma avaliação quantitativa baseada em indicadores e uma avaliação fundada na qualidade assegurada pela análise dos pares.

O QUE DIZEM OS MANIFESTOS

MANIFESTO DE LEIDEN	<i>Princípio 1: A avaliação quantitativa deve apoiar a análise qualitativa de especialistas</i>
THE METRIC TIDE	<i>Princípio da Humildade: a avaliação quantitativa deve apoiar, mas não substituir a avaliação qualitativa feita por especialistas.</i>
DORA	<i>A avaliação deve ser feita não somente com métricas de publicações, mas preferencialmente baseada no conteúdo científico.</i>

O processo de avaliação brasileiro tem como princípio fundamental a avaliação por pares, por meio da criação de comissões que conduzem as diversas atividades avaliativas. Os coordenadores de área administram as atividades em conjunto com consultores *ad hoc*, indicados dentre os membros da comunidade acadêmico-científico-tecnológica e renovados constantemente. Os indicadores são usados como ferramenta auxiliar a análise de mérito, não sendo o único e soberano parâmetro. A análise qualitativa de conteúdo é considerada fundamental para a avaliação e só é possível de ser feita pelos consultores.

2. A importância da combinação de indicadores e da percepção de suas limitações

Todo tipo de indicador possui suas vantagens, mas também limitações. O Manual do Ministério do Planejamento, baseado na OCDE, coloca as principais propriedades e elementos que caracterizam um bom indicador. Deve-se ter em mente que o mau uso dos indicadores leva a distorções em qualquer tomada de decisão. Gestores e avaliadores precisam dispor de indicadores de medição factível, de simples obtenção a partir de fontes confiáveis e com uma forma de obtenção transparente e auditável. Devem também ser estáveis ao longo do tempo e com uma periodicidade regular (BRASIL, 2012).

Os manifestos dizem da importância em não se basear a avaliação em um único indicador, tendo em vista que ele pode não representar toda a complexidade de um sistema de pesquisa e conduzir para uma análise enviesada.

O QUE DIZEM OS MANIFESTOS

MANIFESTO DE LEIDEN	<i>Princípio 9: Reconhecer os efeitos sistêmicos da avaliação e dos indicadores</i>
THE METRIC TIDE	<i>Princípio da Diversidade: reconhecer a variabilidade entre áreas, e usar uma série de indicadores para refletir e apoiar a pluralidade de trajetórias de pesquisa e carreira de pesquisador em todo o sistema.</i>
DORA	<i>Utilize uma série de métricas de artigos e indicadores pessoais, como evidência do impacto de artigos publicados individualmente e outros resultados da pesquisa.</i>

O Manifesto de Leiden alerta para o efeito influenciador dos indicadores usados na avaliação na forma como os pesquisadores se comportam, em busca de atender critérios e garantir, em contrapartida, o apoio de agências de fomento.

A avaliação da pós-graduação se baseia em várias dimensões e conjunto de indicadores que representam todo o universo que envolve as atividades de um programa de pós-graduação, no seu papel de formação de pessoal e de geração e disseminação de conhecimento.

A Ficha de Avaliação é composta de cinco quesitos a serem avaliados para todos os programas, conforme mostrado na **Tabela 1**. Cada quesito possui itens que detalham mais o que deve ser analisado, além de possuir indicadores próprios. No Anexo há uma descrição detalhada dos itens e pesos de avaliação.

Tabela 1. Quesitos da Ficha de Avaliação e descrição

QUESITO	O QUE É AVALIADO?	DESCRIÇÃO E EXEMPLOS DE INDICADORES
PROPOSTA DO PROGRAMA	Estrutura curricular, planejamento e infraestrutura	Análise qualitativa baseada em textos informados pelos programas, que retratam de forma descritiva o escopo do programa em termos de sua proposta curricular, áreas de concentração, disciplinas, projetos de pesquisa, o planejamento para formação discente, aprimoramentos e desenvolvimento futuro e a infraestrutura para ensino, pesquisa e extensão.
CORPO DOCENTE	Perfil docente, formação, experiência e atuação no programa	<p>Perfil do corpo docente no que se refere a sua titulação e experiências e sua dedicação às atividades de ensino e pesquisa do programa.</p> <p><i>% de docentes permanentes² em relação ao total de docentes</i></p> <p><i>% Docentes com titulação de doutor</i></p> <p><i>% Docentes com orientações concluídas de mestrado e de doutorado</i></p> <p><i>% Docentes Permanentes com projetos de pesquisa com financiamento ou com bolsa produtividade</i></p> <p><i>% Docentes com atividade de docência</i></p> <p><i>% Docentes Permanentes com alguma atividade na graduação (docência, orientação de trabalho de conclusão de curso ou iniciação científica)</i></p>
CORPO DISCENTE, TESES E DISSERTAÇÕES	Qualidade dos trabalhos de conclusão, eficiência na formação dos discentes	<p>Perfil do corpo discente, sua participação e desempenho no programa, escopo e qualidade das teses e dissertações produzidas.</p> <p><i>Análise qualitativa do conteúdo das teses e dissertações</i></p> <p><i>Número de Alunos de mestrado/doutorado matriculados e titulados</i></p> <p><i>Tempo médio de titulação</i></p> <p><i>Número de discentes autores de itens de produção intelectual</i></p> <p><i>% de discentes matriculados/titulados em relação ao número de docentes</i></p>

2 Docente permanente é aquele que possui atividades de ensino, orientação e pesquisa no programa e que tenha vínculo com a instituição (Portaria CAPES 174/2014).

QUESITO	O QUE É AVALIADO?	DESCRIÇÃO E EXEMPLOS DE INDICADORES
PRODUÇÃO INTELECTUAL	Produção qualificada do programa	<p>A produção intelectual é quantificada e qualificada. A qualificação é feita por meio das classificações de produtos (livros, eventos, produtos técnicos e artísticos) e de periódicos, o chamado Qualis.</p> <p><i>Distribuição de produções em relação ao corpo docente</i></p> <p><i>Participação de discentes e egressos como autores de produções</i></p> <p><i>Produção média dos docentes</i></p> <p><i>Nº de artigos por estrato do Qualis por ano ou Nº de produtos por estrato por ano (total do programa e por docente, discente e egresso)</i></p> <p><i>Número de produções vinculadas às teses ou dissertações</i></p>
INSERÇÃO SOCIAL	Impacto, colaboração, egressos	<p>Inserção e impacto regional, nacional e/ou internacional do programa, colaboração com outros programas, instituições e centros de pesquisa, visibilidade das ações do programa à comunidade.</p> <p>Análise qualitativa de:</p> <p><i>Atividades de extensão ou equivalentes junto às comunidades locais.</i></p> <p><i>Políticas afirmativas visando o acesso e permanência de professores da educação básica da rede pública e grupos sociais historicamente excluídos.</i></p> <p><i>Atividades na educação básica e ensino médio, com produção de material didático e de formação.</i></p> <p><i>Atividades acadêmicas destacadas, como prêmios, participação em sociedades científicas, divulgação científica, assessorias especiais, participação em conselhos.</i></p> <p><i>Cooperação com setor público e privado</i></p> <p><i>Transferência de conhecimento novo para setores sociais que dele necessitam e qualificação de profissionais para lidar com questões socialmente relevantes.</i></p> <p><i>Avaliação dos impactos sociais dos projetos de pesquisa.</i></p>

QUESITO	O QUE É AVALIADO?	DESCRIÇÃO E EXEMPLOS DE INDICADORES
ATRIBUIÇÃO DE EXCELÊNCIA	Nível de desempenho com padrão internacional	Os programas que recebem notas de excelência devem demonstrar nível de desempenho comparado a padrões internacionais de excelência Análise qualitativa de: <i>Comissões e convênios de cooperação internacional;</i> <i>Mobilidade internacional de discentes e docentes;</i> <i>Medidas informativas (divulgação e comunicação em idiomas estrangeiros);</i> <i>Financiamento estrangeiro;</i> <i>Linhas de pesquisa e centros de referência;</i> <i>Bibliotecas de alto padrão;</i> <i>Publicações com visibilidade internacional;</i> <i>Participação em eventos, cursos internacionais, assessorias, consultorias, editorias, visitas, etc;</i> <i>Prêmios, reconhecimento ou destaque internacional;</i> <i>Redes de pesquisa;</i> <i>Dupla titulação.</i>

Fonte: Fichas de Avaliação/CAPES

3. Indicadores não são imutáveis

Indicadores, uma vez definidos, não precisam se tornar algo imutável, ou seja, eles sempre vão exigir atualização periódica e naturalmente precisam ter capacidade de refletir e se adequar às mudanças. De acordo com Brasil (2012), modelos e teorias são continuamente aperfeiçoados e, portanto, a pertinência dos indicadores deve sempre passar por avaliações críticas, com o cuidado de que a atualização deve ser feita a partir de pesquisas e embasamento metodologicamente confiáveis.

Os manifestos explicitam a necessidade de revisão e atualização regular dos indicadores.

O QUE DIZEM OS MANIFESTOS

MANIFESTO DE LEIDEN

Princípio 10: Examinar regularmente os indicadores e atualizá-los

THE METRIC TIDE

Reflexividade: reconhecer e antecipar os efeitos sistêmicos e potenciais dos indicadores e atualizá-los em resposta.

No que concerne ao processo de avaliação, os Documentos de Área, que contêm as orientações e critérios estabelecidos pelas áreas, passam por revisão e atualização contínua quanto a suas métricas de avaliação. Geralmente, são atualizados e publicados a cada período avaliativo, ou seja, a cada 4 anos. Mas, as atividades realizadas anualmente geralmente possuem atualização de critérios de forma mais frequente, por exemplo, as orientações para avaliação de APCN e classificações da produção intelectual.

4. O limite da precisão

Um dos princípios de um bom indicador, segundo Brasil (2012), é o de Mensurabilidade, que representa a capacidade de alcance e mensuração com maior precisão possível e sem ambiguidade. Os manifestos defendem que o indicador precisa ter um escopo abrangente e confiabilidade assegurada, mas que não se deve basear em uma precisão ilusória. O que defende o manifesto de Leiden é que há incertezas e imprecisão em indicadores de Ciência e Tecnologia e, portanto, não deve ser levada tão a fundo a exatidão dos números. Ou seja, muitos indicadores são representados por números, inteiros ou decimais, mas isso deve ser interpretado com cautela, às vezes o mais correto é considerar uma faixa de valores.

O QUE DIZEM OS MANIFESTOS

MANIFESTO DE LEIDEN

Princípio 8: Evite solidez mal colocada e falsa precisão.

THE METRIC TIDE

Robustez: as métricas devem se basear nos melhores dados possíveis em termos de precisão e escopo.

Quando se considera a escala de avaliação da CAPES, não é feita apenas uma soma regular de números. Para atingir a nota final, os avaliadores fazem uma combinação balanceada de critérios quantitativos e qualitativos para análise das diversas atividades anuais do programa, como mostrado na Tabela 1. Cada item e quesito da Ficha de Avaliação recebe um conceito na escala: “Muito Bom”, “Bom”, “Regular”, “Fraco”, “Insuficiente”, o que, de forma ponderada (Conceito x Peso do item), resulta em uma nota final de 1 a 7.

5. Transparência das ações

Um dos principais pontos tratados por todas as declarações internacionais é o de publicidade e acessibilidade às fontes de informação, às formas de cálculo e aos resultados da avaliação, para que os interessados saibam como serão e foram efetivamente avaliados.

O QUE DIZEM OS MANIFESTOS	
MANIFESTO DE LEIDEN	<p><i>Princípio 4: Manter a coleta de dados e processos analíticos abertos, transparentes e simples.</i></p> <p><i>Princípio 5: Permitir que os avaliados possam verificar dados e análises.</i></p>
THE METRIC TIDE	<p><i>Princípio da Transparência: manter a coleta de dados e processos analíticos abertos e transparentes, de modo que aqueles que estão sendo avaliadas possam testar e verificar os resultados;</i></p>
DORA	<p><i>Seja explícito a respeito dos critérios usados para avaliação da produtividade científica e claramente destaque que o conteúdo científico de um artigo é muito mais importante do que as métricas da publicação ou de identidade do jornal onde foi publicado.</i></p> <p><i>Seja aberto e transparente ao fornecer dados e métodos usados para calcular todas as métricas.</i></p>

Todo o processo de avaliação é baseado no Coleta de Dados, sistema que recebe informações anuais dos programas, preenchido pelo coordenador do programa e validado pela instituição de ensino, na figura do pró-reitor de pós-graduação. Em 2013, foi lançada a Plataforma Sucupira, que é uma ferramenta online para a Coleta de Informações, submissão de propostas

APCN e consultas diversas. A grande mudança de cenário foi que, com a nova Plataforma, os dados passaram a ser públicos e de acesso aberto. Assim, qualquer interessado pode consultar as informações de um programa à medida que vão sendo preenchidas, o que garante transparência e consequentemente um ciclo maior de confiança dos dados, já que um discente ou docente pode verificar se os dados que o programa informou estão corretos e, em caso negativo, solicitar correção.

Assim, o mesmo conjunto de dados utilizado pelas comissões no momento da avaliação pode ser obtido por qualquer interessado, permitindo fazer simulações ou comparações de forma autônoma, o que é fundamental para a transparência e isonomia do processo.

6. Reconhecer a diversidade entre áreas e o contexto do impacto dos programas

Cada área do conhecimento possui heterogeneidade de seus padrões, seja quanto a características do ensino e da pesquisa ou quanto ao escopo de publicações e citações, assim como a ponderação dada a elas. O impacto da pesquisa possui também abrangência diferenciada, que pode ser considerado a nível local, regional, nacional ou internacional.

O QUE DIZEM OS MANIFESTOS

MANIFESTO DE LEIDEN

Princípio 3: Proteger a excelência da pesquisa localmente relevante.

Princípio 6: Considerar variações por área em práticas de publicação e citação.

THE METRIC TIDE

Princípio da Diversidade: reconhecer a variabilidade entre áreas, e usar uma série de indicadores para refletir e apoiar a pluralidade de trajetórias de pesquisa e carreira de pesquisador em todo o sistema.

DORA

Para efeitos de avaliação da pesquisa, considere o valor e o impacto de todos os resultados da investigação (incluindo conjuntos de dados e software) para além das publicações e considere uma vasta gama de medidas de impacto incluindo indicadores qualitativos de impacto, como por exemplo, influência na política e na prática.

Considere variação nos tipos de artigos (por exemplo, revisões versus artigos de pesquisa) e em diferentes áreas do conhecimento quando as métricas são usadas, agregadas ou comparadas

A CAPES estabelece os princípios gerais da avaliação, como o padrão da Ficha, seus quesitos e itens gerais que devem obrigatoriamente constar em todas as áreas. Porém, cada uma das 49 áreas pode customizar seus critérios e indicadores, desde que siga o mínimo requerido nas normatizações.

Assim, as áreas podem dar importância distinta para os produtos da pós-graduação. As áreas pertencentes a grande área de Humanidades geralmente dão um peso elevado para livros. A área de Ciência da Computação é uma das que pontuam fortemente eventos. A Biotecnologia valoriza a produção de patentes, tendo em vista o seu caráter inovador. Os indicadores para cada um dos itens também variam. De qualquer forma, toda a produção do programa é considerada para fins de avaliação, não somente artigos científicos, mas também livros, artigos em conferências, produção técnica e artística.

No caso dos programas profissionais há diferenciação nos itens de avaliação. Produtos técnicos são mais valorizados e há maior variedade dos tipos de trabalhos de conclusão, que podem ser um desenvolvimento de software, relatório técnico, produção de material didático ou instrucional (Portaria CAPES 17/2009).

Não há preferência ou discernimento com o idioma da publicação. Algumas áreas recomendam que as publicações sejam feitas em inglês, visando maior internacionalização do programa, mas isso não é mandatório.

Aquelas publicações indexadas em bases de dados internacionais, como a Web of Science e a Scopus são avaliadas com base em métricas bibliométricas disponíveis. Porém os comitês também consideram as bases de dados com maior cobertura regional, como a Scielo. Aquelas não indexadas são avaliadas por seu impacto local ou regional, considerando a importância do conteúdo desenvolvido e objetivos da pesquisa.

Por exemplo, dentro da área de Ciências Agrárias, o desenvolvimento de uma técnica agrícola em região de seca ou o plantio de um cultivar próprio para determinada região podem ter impacto expressivo local e regionalmente, mas não a nível nacional. A produção (um artigo ou livro) resultante desta pesquisa provavelmente não terá grande número de citações, mas isso não significa que não tenha tido resultado relevante. Assim como acontece com a valorização de programas em Medicina Tropical, em Literatura Brasileira, Ensino de História local, em que a menor disseminação não significa falta de qualidade ou de prestígio.

7. Medir a qualidade e o impacto preferencialmente que a quantidade

O ponto fundamental aqui é primar pela qualidade dos trabalhos e dos pesquisadores ao invés de meramente adotar uma contagem numérica. O impacto que a pesquisa tem em transformar algo ao seu redor, em contribuir com a solução de um problema da sociedade também deve ser considerado.

O QUE DIZEM OS MANIFESTOS

MANIFESTO DE LEIDEN

Princípio 2: Meça o desempenho em relação às missões de pesquisa da instituição, do grupo ou do pesquisador.

Princípio 7: Utilize como base de avaliação de pesquisadores um julgamento qualitativo de seu portfólio

THE METRIC TIDE

Quando os indicadores padrão são inadequados, os pesquisadores individuais devem procurar uma variedade de fontes de dados para documentar e apoiar reivindicações sobre o impacto de seu trabalho.

DORA

Avalie a pesquisa baseado em seus próprios méritos

Os docentes são avaliados por todas as suas atividades no programa, incluindo não apenas artigos publicados, mas também a sua participação em projetos de pesquisa, orientação de alunos, disciplinas ministradas e experiência profissional. O foco de avaliação é todo o conjunto de atividades do programa e a qualidade de formação dos discentes, mas a produtividade docente também é individualmente analisada em coerência com o escopo do programa.

Um dos itens de avaliação é a Inclusão Social, detalhado na **Tabela 1**. Parte da nota é atribuída a questões de cooperação, transferência e divulgação do conhecimento.

Considerando a disponibilidade das informações continuamente ao longo do ano a partir da Plataforma Sucupira, os consultores estão tendo a oportunidade de preparar e validar os indicadores previamente, para que durante a semana presencial de avaliação na CAPES, apenas as discussões finais e a decisão sobre o mérito possam ser deliberadas em plenária.

O PNPG 2011-2020 coloca a importância de se basear a avaliação na “qualidade e excelência dos resultados, na especificidade das áreas de conhecimento e no impacto dos resultados na comunidade acadêmica e empresarial e na sociedade” Os indicadores não podem se ater apenas ao mero produtivismo, mas a ênfase maior deve ser em refletir a “relevância do conhecimento novo, sua importância no contexto social e o impacto da inovação tecnológica no mundo globalizado e competitivo”. Colocou-se como

orientação para os critérios de avaliação a definição de indicadores que reflitam a inovação, a geração e transferência de conhecimento, a formação qualificada de recursos humanos, por meio do acompanhamento de egressos e a indução da pesquisa em problemas relevantes e estratégicos para o desenvolvimento do país (BRASIL, 2010).

8. Bons indicadores necessitam de dados estruturados e confiáveis

Para qualquer cálculo de indicador, a qualidade da informação é essencial. Para isso, os sistemas de registro de dados precisam adotar padrões para estruturação e descrição da informação. Atualmente, alguns padrões vêm sendo adotados internacionalmente, além de identificadores únicos que garantem a unicidade e a confiabilidade do registro.

Cita-se, por exemplo, os identificadores únicos de pessoas (ex. ORCID, ResearcherID, Scopus Author Identifier), organizações (ex. ISNI, GRID), periódicos (ex. ISSN), artigos (ex. DOI), livros (ex. ISBN). Além disso, a adoção de padrões de modelos de dados (ex. CERIF) e dicionário de dados (ex. CASRAI), permitem a interoperabilidade (troca) de informações entre sistemas. Essas iniciativas de padronização, registro, geração e divulgação de conhecimento têm sido feitas na comunidade europeia, por exemplo, por meio do EUROCRIS (<http://eurocris.org/>) e nos Estados Unidos, pela iniciativa da rede VIVO (<http://vivoweb.org/>).

O Metric Tide deixa explícita a importância de se melhorar a infraestrutura de dados que suporta o gerenciamento de informações de pesquisa. O relatório cita especificamente o uso do ORCID, que será obrigatório para todos os pesquisadores no próximo ciclo avaliativo das instituições no Reino Unido, o DOI para as publicações e o ISNI, para mapeamento de instituições. Também coloca a necessidade de se aperfeiçoar a interoperabilidade entre sistemas de gerenciamento de pesquisa, para fins de troca de informações entre agências financiadoras, editoras, universidades, etc.

Ações em consonância com estas orientações estão sendo discutidas na CAPES para incorporação no planejamento estratégico da Plataforma

Sucupira para os próximos anos. A certificação das informações atualmente é confiada aos coordenadores de programa, que ainda passa pela validação pelos pró-reitores das instituições. Porém, ações para auditoria dos dados em bases internacionais e em outras fontes de informações ainda precisam ser mais amplamente adotadas para assegurar a melhor qualidade das informações.

OUTRAS CONSIDERAÇÕES E DETALHAMENTOS

Classificação da Produção Intelectual

A produção intelectual dos programas de pós-graduação não é considerada apenas de forma quantitativa, mas também é avaliada por meio de um processo de qualificação, baseado em uma escala de classificação.

A qualificação de artigos científicos é feita indiretamente pela classificação dos jornais ou periódicos de divulgação. Essa sistemática é denominada Qualis Periódicos e é a única requerida para todas as áreas de avaliação. Recomenda-se a leitura do artigo intitulado “Dez coisas que você deveria saber sobre o Qualis”, para uma discussão mais aprofundada a este respeito (BARATA, 2016).

Os critérios de classificação da produção intelectual também não se baseiam em apenas um indicador. A classificação Qualis de Periódicos leva em conta principalmente indicadores bibliométricos dos periódicos fornecidos pelas bases indexadoras. Os principais são o *Fator de Impacto*, do Journal of Citation Reports, base da Web of Science, o *SJR* e *Cites per Doc*, da Scopus e o *índice h* do Google Scholar.

Cada ciclo de avaliação analisa as publicações dos últimos quatro anos, portanto, os indicadores baseados em citação não podem ser considerados absolutos porque os artigos mais recentes deixarão de ter valores significativos.

No caso da avaliação da CAPES existe ainda um importante aspecto. Uma parte considerável dos artigos produzidos pelos programas não é publicada em periódicos indexados nessas bases internacionais de dados, tais

como Web of Science e Scopus e, portanto, não há como obter todo ou parte destes indicadores bibliométricos. Além disso, o fator de impacto dos artigos brasileiros é relativamente baixo devido ao baixo grau de cooperação internacional. Por causa disso, o Qualis não se baseia apenas em métricas, são considerados também critérios mais qualitativos, como por exemplo, a presença em bases indexadoras, a linha editorial, as normas de submissão e avaliação dos artigos, os meios de divulgação, a periodicidade e a importância para a área.

Há uma crítica ao uso indiscriminado do Fator de Impacto como o primeiro e muitas vezes único parâmetro para avaliar e comparar a produção científica de pesquisadores e instituições. A Declaração DORA e o Metric Tide citam especificamente as limitações deste indicador e sugerem a adoção de outras métricas para representar o desempenho de jornais, como por exemplo, o Fator de Impacto de 5 anos, o Eigen Factor, o SCImago e o índice h.

A busca por novas métricas com objetivo de superar algumas limitações dos indicadores disponíveis aumentou. É o caso da nova métrica proposta pelo NIH (National Institute of Health), *Relative Citation Ratio*, uma medida da influência de um artigo com base na taxa relativa de citações (NAIK, 2016).

A forma de se estratificar os artigos de forma indireta como é feito no Qualis é muitas vezes criticada, mas por outro lado há o desafio de se avaliar uma quantidade elevada de produções informadas anualmente, que chegam a ordem de 70 mil livros, 215 mil trabalhos em anais, 6 mil produções artísticas e 367 mil produções técnicas em média por ano (Tabela 2). Uma das opções seria utilizar o número de citações de artigos (o que não garante a qualidade, mas permite ter uma ideia do impacto) e combiná-lo com a classificação do periódico dada pelo Qualis. Além disso, poder-se-ia optar pela avaliação qualitativa individual apenas das melhores publicações, como preconiza o próprio PNPG 2011-2020 (BRASIL, 2010).

Tabela 2. Evolução do quantitativo de produção intelectual registrada pelos programas nos anos de 2013 a 2015

PRODUÇÃO INTELECTUAL			
	2013	2014	2015
ARTÍSTICA	5.001	6.434	6.853
BIBLIOGRÁFICA	521.540	512.584	508.442
Artigo	196.629	205.340	202.163
Livro	66.406	70.143	73.335
Trabalho em anais	233.148	208.330	204.678
Outros	25.357	28.771	28.266
TÉCNICA	360.645	366.011	374.623
Total	887.186	885.029	889.918

Fonte: CAPES

Ao se avaliar os documentos de todas as áreas de avaliação, referentes aos parâmetros adotados na Avaliação de 2013, observa-se que existem três critérios gerais para avaliação de periódicos científicos:

1. Indicadores bibliométricos fornecidos por bases de dados nacionais e internacionais;
2. Presença em bases indexadoras, principalmente aquelas mais relevantes para a área;
3. Critérios qualitativos, que incluem: existência de editor responsável e conselho editorial, presença de ISSN, características da linha editorial, normas de submissão, avaliação por pares, afiliação institucional de autores, idioma de publicação, formas de divulgação, periodicidade e outros critérios mais qualitativos que envolvem a valoração de periódicos em cada área.

Com relação ao primeiro critério, observa-se que 81% das áreas utilizam o Fator de Impacto (Journal of Citation Reports da Web of Science) e, destas, 74% o consideram como principal definidor de classificação. 56% das áreas utilizam o SJR (Scientific Journal Rankings – Scimago Journal & Country Rank, base da SCOPUS) e, destas, 89% o consideram apenas como auxiliar para definir classificação. Outros indicadores utilizados são o índice h, o Cites per Doc e alguns mais específicos para determinadas áreas. Sete áreas não relataram usar nenhum indicador bibliométrico (Figura 3).

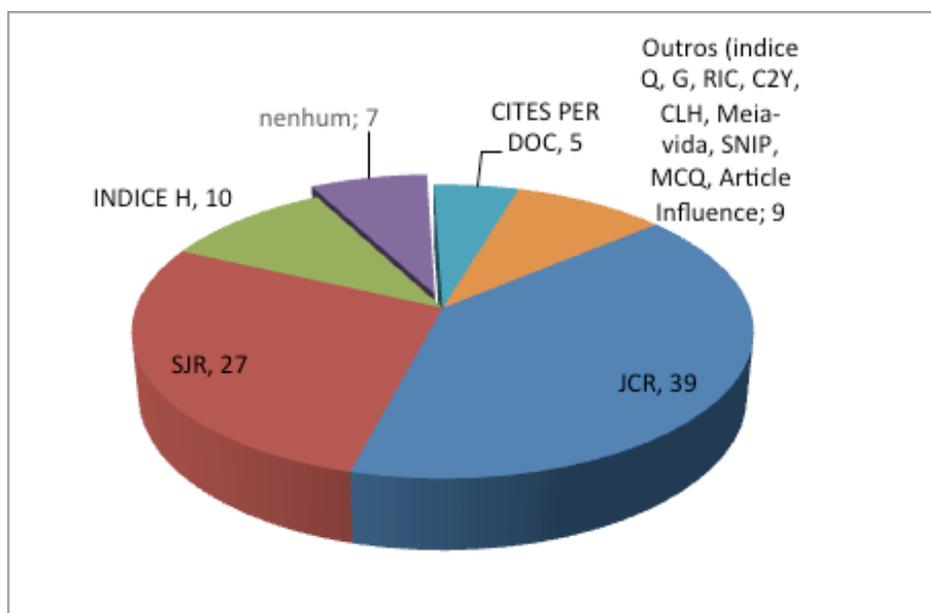


Figura 3. Número de áreas e o indicador bibliométrico usado para classificação Qualis

A diversidade de indicadores utilizados pelas áreas na avaliação da pós-graduação está em consonância com a recomendação de não se usar apenas um parâmetro.

Quanto ao segundo critério, 61 bases indexadoras foram relatadas por todas as áreas. 79% das áreas consideram, como parâmetro, o periódico estar indexado na Scielo, 63% na Scopus e 60% na Web of Science, que são as principais por indexarem periódicos de todas as áreas. As demais bases são

mais específicas, por exemplo, a PubMed e MedLine para as Medicinas e áreas da Saúde, a CAB Abstracts, AGRIS (Agricultural Science) e FSTA (Food Science and Technology Abstracts) para as áreas de Ciências Agrárias, a PSYCINFO para as Ciências Sociais e BIOSIS para as Biológicas (Figura 4).

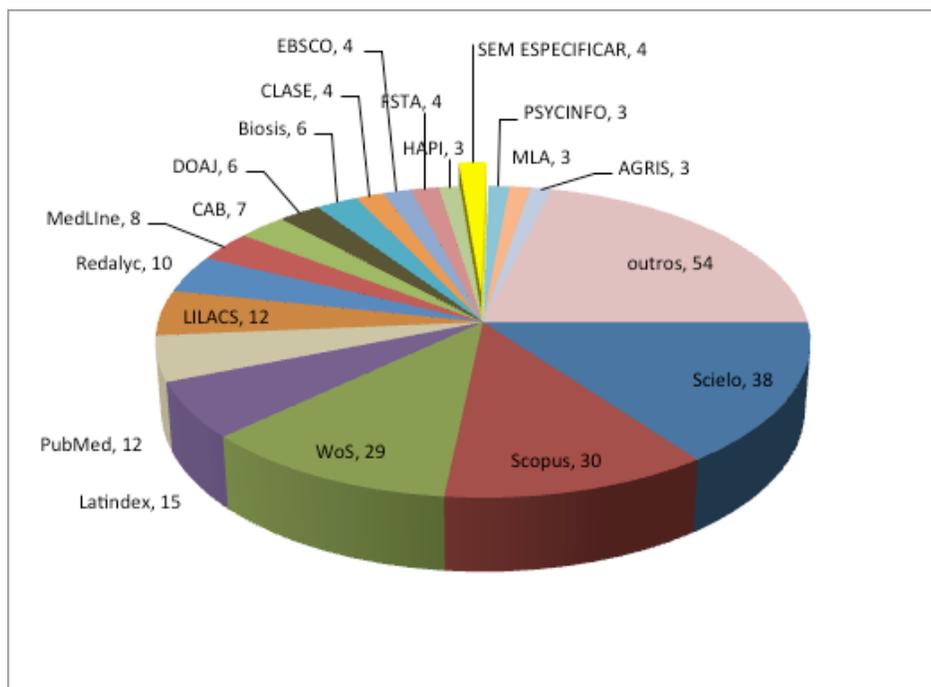


Figura 4. Bases Indexadoras utilizadas pelas áreas de avaliação como critérios de classificação Qualis

Critérios para classificação de livros

Além da classificação de periódicos, há também a avaliação de livros. Considera-se, como critérios de classificação, o cadastro no ISBN, número de páginas, características de editoria e ficha catalográfica, o ineditismo e a relevância de seu conteúdo. As obras são classificadas em escala que vai de L1 (menor qualidade) a L4.

Algumas áreas também realizam a classificação de eventos e produtos técnicos, porém devido a grande diversidade de produções não há ainda sistematização de sua realização. Cabe enfatizar que não existe correlação ou regras de correspondências entre escala de livros e escala Qualis para os periódicos.

APERFEIÇOAMENTO DE TECNOLOGIAS DE INFORMAÇÃO NECESSÁRIAS AO DESENVOLVIMENTO DO SNPG

Qualidade da Informação: Perspectivas

A necessidade de aperfeiçoamento constante do processo de avaliação e de suas ferramentas para apoiar a análise dos pares requer fases evolutivas para a Coleta de Dados, que vem se transformando de um simples registro administrativo para uma ferramenta de gerenciamento de informações. A Plataforma Sucupira permitiu alcançar um estágio de maior transparência para a comunidade, mas ainda há ações a serem tomadas para proporcionar mais qualidade da informação. Isso inclui a integração com outros sistemas de Ciência e Tecnologia e a colaboração com outros atores envolvidos, o que permite alcançar a perspectiva de gestão do conhecimento (Figura 5).

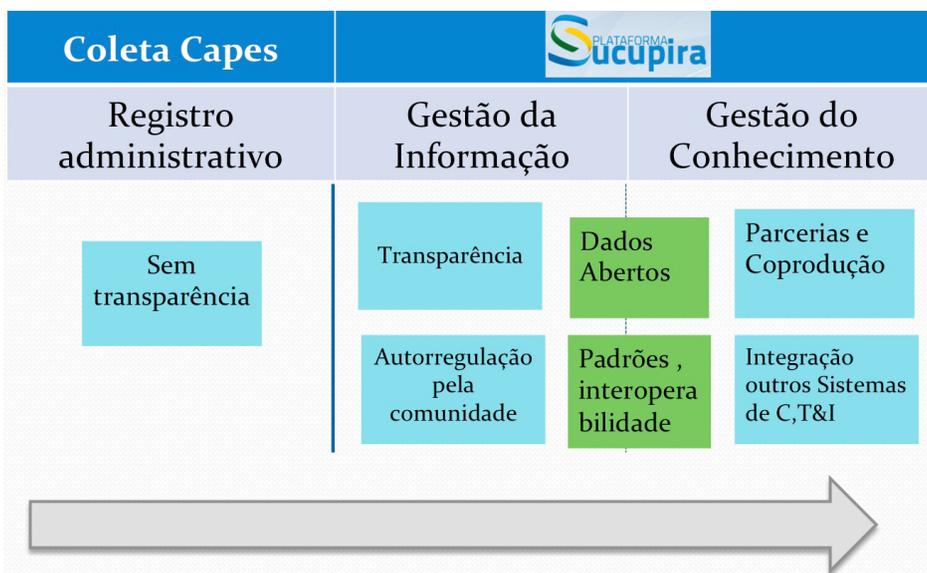


Figura 5. Evolução da ferramenta de Coleta de Dados

Ferramentas de bancos de dados e visualizações

Em 2015, a CAPES lançou o projeto do Acervo de Dados Digitais (ADD), que vem fazendo um trabalho de organização, padronização e descrição dos dados. No caso da avaliação, a Plataforma Sucupira funciona como a porta de entrada de dados a partir dos programas de pós-graduação, permitindo consultas diversas. O ADD cria o acervo de dados, com a documentação dos metadados (“dados que descrevem os dados”, ou seja, são informações descritivas para facilitar a compreensão dos dados) e a disponibilização de microdados (informação detalhada na sua forma mais bruta). Com isso, é possível disponibilizar informação organizada para a sociedade, o que facilita também a criação de planilhas e visualização gráfica (Figura 6).

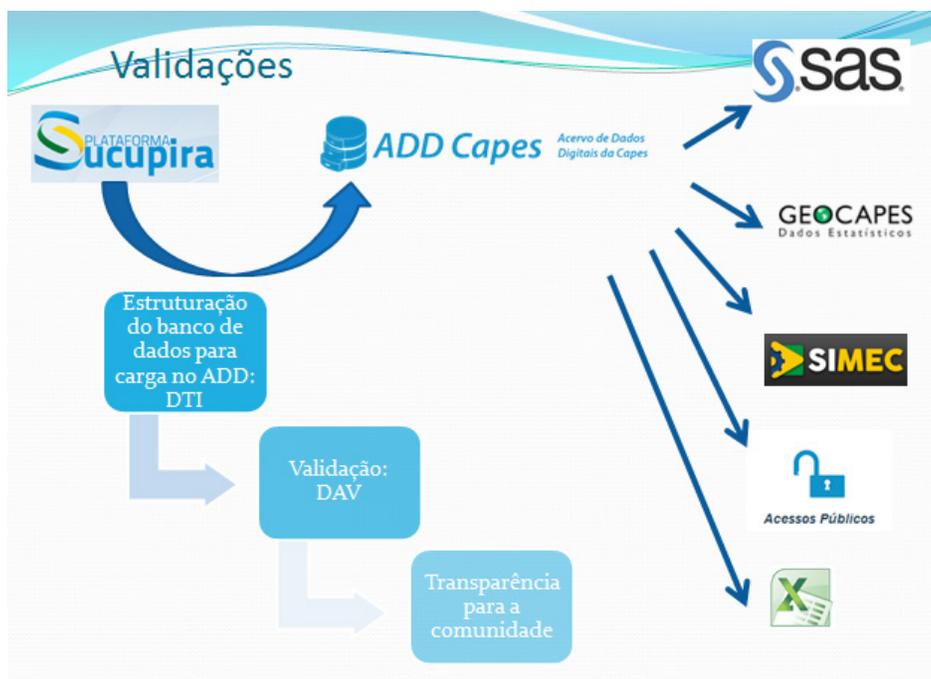


Figura 6. Estrutura do Acervo de Dados Digitais da CAPES

Os acessos públicos às informações da CAPES estão disponíveis por meio do Sistema de Disseminação de Informações, em <http://sdi.capes.gov.br/>. Neste espaço, é possível ter acesso ao Portal de Dados Abertos, ao Banco de Teses e Dissertações e ao Portal Transparência CAPES, com a relação dos pagamentos realizados individualmente e por projeto.

CONSIDERAÇÕES FINAIS

No mundo de *big data*, ou seja, “grande volume de dados estruturados ou não”, as agências de fomento e avaliação precisam se munir de ferramentas que permitam a geração de conhecimento e apoio à tomada de decisão. Os indicadores obtidos – com os cada vez mais poderosos algoritmos de extração –, são ferramentas auxiliares neste processo, que municiam avaliadores e

gestores. No entanto, cabe enfatizar que eles devem ser usados com permanente crítica e moderação, tendo em vista suas naturais limitações.

No que tange o modelo de avaliação brasileiro, os processos e práticas da CAPES atendem em grande medida os princípios dos manifestos internacionais. Naturalmente, o modelo precisa de discussão e aperfeiçoamentos constantes, o que já é feito sempre em diálogo aberto com a comunidade. A avaliação por pares é um dos fundamentos primários das atividades avaliativas conduzidas pela CAPES, desde quando foram iniciadas, tornando o processo confiável e certificado pela qualidade.

Um dos pontos importantes de aperfeiçoamento é a adoção de iniciativas a fim de refinar as tecnologias de informação necessárias ao desenvolvimento do Sistema Nacional de Pós-Graduação, visando menor trabalho operacional de preenchimento de dados, maior confiabilidade das informações e, em consequência, garantindo maior abrangência e representatividade dos indicadores.

As iniciativas dos manifestos internacionais são bastante válidas para provocar a discussão em diferentes esferas de avaliação e financiamento e promover a sua evolução.

Todo tipo de indicador possui suas vantagens, mas também limitações. Gestores e avaliadores precisam dispor de indicadores adequados, mas é necessário assegurar que seu uso não traga distorções em qualquer tipo de tomada de decisão.

ANEXO

Ficha de Avaliação para Programas Acadêmicos

QUESITOS / ITENS	PESO
1 – PROPOSTA DO PROGRAMA	0
1.1. Coerência, consistência, abrangência e atualização das áreas de concentração, linhas de pesquisa, projetos em andamento e proposta curricular.	1.1 + 1.2 ≥ 60%
1.2. Planejamento do programa com vistas a seu desenvolvimento futuro, contemplando os desafios internacionais da área na produção do conhecimento, seus propósitos na melhor formação de seus alunos, suas metas quanto à inserção social mais rica dos seus egressos, conforme os parâmetros da área.	1.1 + 1.2 ≥ 60%
1.3. Infraestrutura para ensino, pesquisa e, se for o caso, extensão.	1.3 ≥ 5%
2 – CORPO DOCENTE	15 ou 20%
2.1. Perfil do corpo docente, consideradas titulação, diversificação na origem de formação, aprimoramento e experiência, e sua compatibilidade e adequação à Proposta do Programa.	2.1 ≥ 10%
2.2. Adequação e dedicação dos docentes permanentes em relação às atividades de pesquisa e de formação do programa.	2.2 ≥ 20% (2.2 + 2.3 ≥ 60%)
2.3. Distribuição das atividades de pesquisa e de formação entre os docentes do programa.	2.3 ≥ 30% (2.2 + 2.3 ≥ 60%)
2.4. Contribuição dos docentes para atividades de ensino e/ou de pesquisa na graduação, com atenção tanto à repercussão que este item pode ter na formação de futuros ingressantes na PG, quanto (conforme a área) na formação de profissionais mais capacitados no plano da graduação. Obs: este item só vale quando o PPG estiver ligado a curso de graduação; se não o estiver, seu peso será redistribuído proporcionalmente entre os demais itens do quesito.	2.4 ≥ 10%
3 – CORPO DISCENTE, TESES E DISSERTAÇÕES	30 ou 35%
3.1. Quantidade de teses e dissertações defendidas no período de avaliação, em relação ao corpo docente permanente e à dimensão do corpo discente.	3.1 + 3.2 + 3.4 ≥ 40% (3.1 ≥ 10%)
3.2. Distribuição das orientações das teses e dissertações defendidas no período de avaliação em relação aos docentes do programa.	3.1 + 3.2 + 3.4 ≥ 40% (3.2 ≥ 10%)

- 3.3. Qualidade das Teses e Dissertações e da produção de discentes autores da pós-graduação e da graduação (no caso de IES com curso de graduação na área) na produção científica do programa, aferida por publicações e outros indicadores pertinentes à área. $\geq 30\%$
- 3.4. Eficiência do Programa na formação de mestres e doutores bolsistas: Tempo de formação de mestres e doutores e percentual de bolsistas titulados. $3.1 + 3.2 + 3.4 \geq 40\%$

4 – PRODUÇÃO INTELECTUAL

35 ou 40%

- 4.1. Publicações qualificadas do Programa por docente permanente. $4.1 + 4.4 \geq 40$
- 4.2. Distribuição de publicações qualificadas em relação ao corpo docente permanente do Programa. $4.2 \geq 30$
- 4.3. Produção técnica, patentes e outras produções consideradas relevantes. $4.3 \geq 5$
- 4.4. Produção artística, nas áreas em que tal tipo de produção for pertinente. $4.1 + 4.4 \geq 40$
($4.1 \geq 4.4$)

5 – INSERÇÃO SOCIAL

10 ou 15%

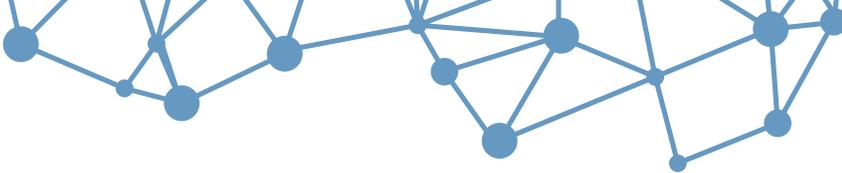
- 5.1. Inserção e impacto regional e (ou) nacional do programa. $5.1 \geq 15\%$
- 5.2. Integração e cooperação com outros programas e centros de pesquisa e desenvolvimento profissional relacionados à área de conhecimento do programa, com vistas ao desenvolvimento da pesquisa e da pós-graduação. $5.2 \geq 20\%$
- 5.3. Visibilidade ou transparência dada pelo programa à sua atuação. $15 \geq 20\%$

REFERÊNCIAS

- BARATA, R. C. B. Dez coisas que você deveria saber sobre o Qualis. *Revista Brasileira de Pós-Graduação*, v.13, n. 30, p. 13-40, 2016. doi: 10.21713/2358-2332.2016.v13.947.
- BRASIL. Ministério do Planejamento, Orçamento e Gestão. Indicadores. *Orientações Básicas Aplicadas à Gestão Pública*. 1ª edição. Brasília, DF: MP, 2012. 64 p.
- BRASIL. *Plano Nacional de Pós-Graduação 2011-2020*. Brasília, DF: CAPES, 2010. 608 p.
- DECLARATION ON RESEARCH ASSESSMENT (DORA). San Francisco Declaration on Research Assessment. 2012. Disponível em: <<http://www.ascb.org/files/SFDeclarationFINAL.pdf?x30490>>. Acesso em: 28 fev. 2017.

- FERREIRA, M. M.; MOREIRA, R. L. (Eds.). *CAPES 50 anos: depoimentos ao CPDOC/ FGV*. Brasília, DF: FGV/CPDOC/CAPES, 2003.
- HICKS, D.; WOUTERS, P.; WALTMAN, L.; DE RIJCKE, S.; RAFOLS, I. The Leiden Manifesto for research metrics. *Nature*, v. 520, n. 7548, p. 429-431, 2015. doi: 10.1038/520429a.
- NAIK, G. The quiet rise of the NIH's hot new metric. *Nature*, v. 539, n. 7628, p. 150, 2016. doi: 10.1038/539150a.
- WILSDON, J.; ALLEN, L.; BELFIORE, E.; CAMPBELL, P.; CURRY, S.; HILL, S.; JONES, R.; KAIN, R.; KERRIDGE, S.; THELWALL, M.; TINKLER, J.; VINEY, I.; WOUTERS, P.; HILL, J.; JOHNSON, B. *The Metric Tide: report of the independent review of the role of metrics in research assessment and management*. Bristol: HEFCE, 2015. 163 p. doi: 10.13140/RG.2.1.4929.1363.





PUBLIC POLICIES IN SCIENCE AND TECHNOLOGY IN BRAZIL: CHALLENGES AND PROPOSALS FOR THE USE OF INDICATORS IN EVALUATION

POLÍTICAS PÚBLICAS EM CIÊNCIA E TECNOLOGIA NO BRASIL: DESAFIOS E PROPOSTAS PARA UTILIZAÇÃO DE INDICADORES NA AVALIAÇÃO [ver página 157] [ir para o sumário]

**Talita Moreira de Oliveira and **Livio Amaral*

INTRODUÇÃO¹

The Brazilian graduate evaluation system was implemented by the Federal Agency for Support and Evaluation of Graduate Education (CAPES) in 1976. The agency is responsible for recommending new graduate programs and for the periodic evaluation of their performance which, as a result, ensures quality standards within the National Graduate Studies System (SNPG). This process is carried out with significant participation of the academic-scientific-technological community through evaluation committees, currently divided into 49 areas. The evaluation process has been the basis for fostering graduate education and research, not only by CAPES itself, but also as an indicator for other government agencies.

* CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; talita.oliveira@capes.gov.br

** Departamento de Física, Universidade Federal do Rio Grande do Sul

¹ The authors acknowledge Dr. Jorge Almeida Guimaraes for critically reading this manuscript

The main principles of CAPES evaluation relies on quality based on peer review; criteria debated and updated by the Brazilian academic-scientific-technological community in each evaluation period and transparency of criteria, procedures and results. It is a system that needs constantly improvement and requires enough flexibility to be ready to deal with the growing number of courses and the evolving pattern of their different features.

The expansion and consolidation of graduate courses (i.e Master's and Doctoral) are essential to ensure properly qualified personnel who can take part in industry and universities, and therefore fosters the socio-economic growth of the country. The national education policies for science and technology helped to develop graduate studies and scientific output in Brazil.

The indicators act as instruments for simulations and planning strategies which make it easier to confront a complex world which includes large amounts of data, external sources of information and a wide range of variables.

Currently, as well as assessing graduate studies, the activities of CAPES are designed to elaborate policies and support training of human resources, both in the country and abroad. They also involve improving international scientific cooperation and encouraging both the initial and continuous training of teachers for basic education.

ASSESSMENT OF GRADUATE STUDIES IN BRAZIL

The evaluation of graduate studies aims to analyze in detail the panorama and activities of this level of education in Brazil, producing studies and indicators that base government policies and the continuous growth of graduate studies being currently the most important instrument for funding by CAPES itself and other agencies. For example, it subsidizes the identification of strategic areas for science and technology in the country, as well as inducing reduction of regional asymmetries and the expansion of graduate programs.

The assessment system can be divided into two macro-processes – a) the admission of new courses (Master's Degree, Doctoral or Professional

Master) presented by universities and research institutions to CAPES and b) the periodical assessment of approved programs (Figure 1).

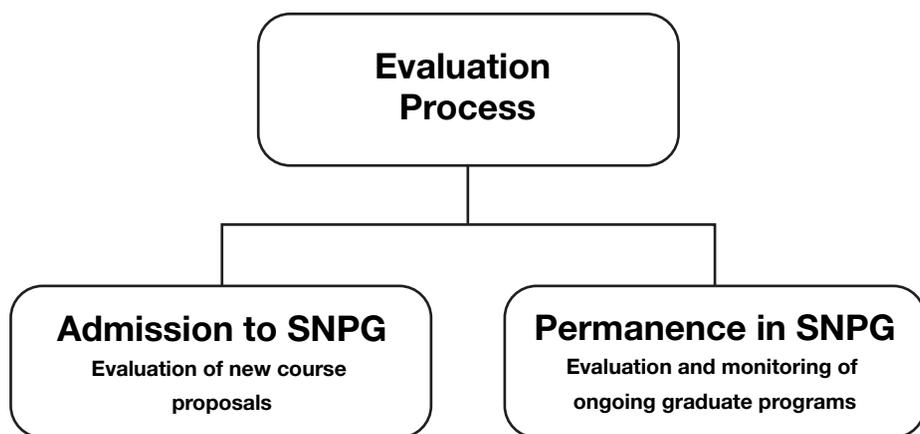


Figure 1. Macroprocesses of the system for assessing graduate studies

Source: CAPES

Every graduate program has to go through CAPES assessment both before entering in operation and, once in operation, in order to continue its activities. The assessment system is carried out by peer review where panels are divided into 49 areas. Each of them is led by a coordinator with outstanding experience in teaching and research in his/her area of expertise on a three-year term of office. Committees are formed by consultants chosen by the academic community among its own members. The committees are not permanent; the consultants are renewed in at least 50% for each evaluation activity.

Area committees are formed according to the type of evaluation activity to be performed, which can be:

- i. **Analysis of new courses proposals:** Institutions interested in offering a graduate course submit a proposal to CAPES, which goes through the merit analysis of the area committees and the approval of the Technical and Scientific Higher Education Board (CTC-ES), which in turn may or may not recommend it for formal purposes by the Ministry of Education.

- ii. **Classification of technical and scientific products:** The products resulting from programs activities undergo an evaluation process in order to assess their academic quality and real operational procedures. Classifications are made for scientific journals, books, educational, scientific and technological conferences, technical products, infrastructure aspects, administrative routines and daily activities. Each of these products has its own independent classification scale.
- iii. **Follow-up seminars:** Regular annual meetings to discuss criteria and indicators of the area and follow-up of program performance together with the academic community.
- iv. **Periodic Evaluation:** The programs go through evaluative cycles, currently at intervals of four years, in order to assess their performance in terms of teaching, research, extension activities and student training. Programs receive grades from 1 to 7, and those who receive 1 or 2 are not authorized to continue their activities. Grades 6 and 7 indicate outstanding programs, comparable to those of recognized international research centers.

The dimension of the SNPG gives an idea of the challenge to assess the growing number of courses. According to Ferreira and Moreira (2003), in 1976, there were only 188 people with doctoral degrees obtained in Brazilian courses. From that year up to now, graduate studies have witnessed a regular and steady growth. Between 1998 and 2014 (Figure 2), the number of courses grew by 172% and SNPG displayed 2,016 PhD, 3,130 Master and 540 Professional Master's courses that enrolled around 230,000 students and graduated more than 69,000 students.

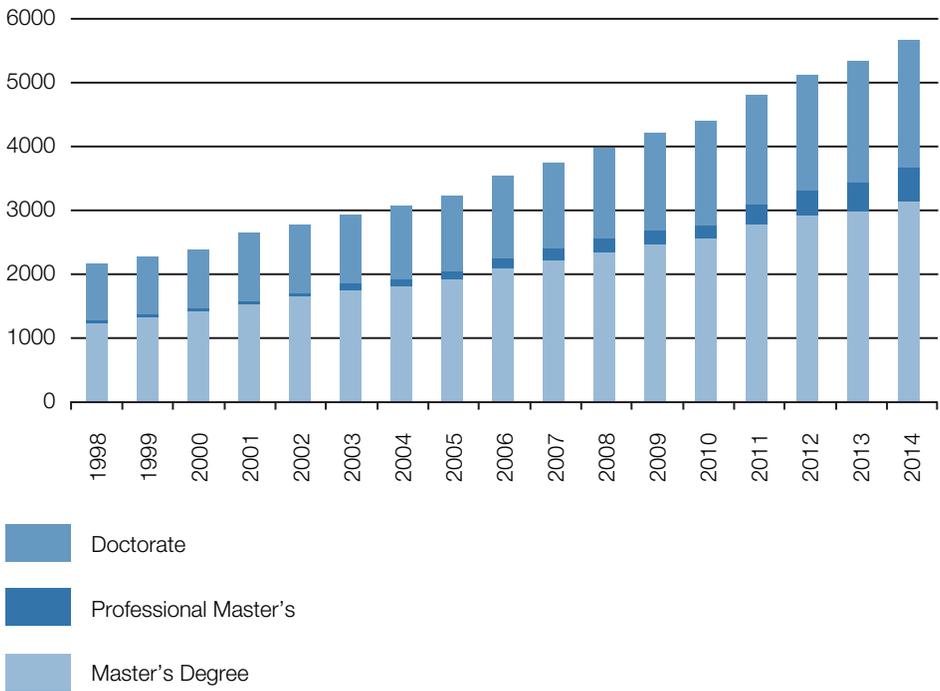


Figure 2. Growth of graduate courses
Source: CAPES

CRITERIA AND ASESSMENT INDICATORS

The International Manifestos

The use of metrics to evaluate scientific research performance has been the subject of intense discussion and increasing manifestation of the world scientific community. Recently, three important international manifestos have been made on this topic.

In 2012, a group of editors and publishers of scientific journals, during the annual meeting of the American Society for Cell Biology published the San Francisco Declaration on Research Assessment (DORA, 2012) (<http://www.ascb.org/dora/>), containing a set of recommendations for improving the way of evaluation of scientific research. Many other parties, such as funding agencies, institutions and researchers, have also supported DORA's

declaration, showing their commitment to a more accurately assessment focused primarily on the quality and impact of research outputs.

In 2015, the Leiden Manifesto for research metrics was released as a result of discussions during the International Conference on Science and Technology Indicators (STI 2014) in Leiden. It was proposed ten principles of good practice for measuring research performance (HICKS et al., 2015). One of the concerns raised by the Leiden is that evaluations have been increasingly based on metrics rather than on experts' judgment.

During the same year, The Metric Tide was published, which is a *report on the role of metrics in research assessment and management*, commissioned by the Higher Education Funding Council for England (HEFCE) (WILSDON et al., 2015). The document addresses the uses and limitations of research metrics and their potential contribution for improving research excellence. It points out five principles of what was called "responsible metrics", which indicate the proper use of quantitative indicators.

We drew on the concept of the principles for the good use of metrics stated by the international manifestos and matched with the evaluation practices used by the Brazilian agency CAPES into the process of graduate programs assessment. It was aggregated similar principles of the Leiden, Metric Tide and Dora and discoursed comparably to well related CAPES practices.

Comparison of the International Manifestos and CAPES procedures

1. Indicators: important but not unconditional

The fundamental principle of evaluation is that it must rely on the careful analysis of groups of experts who have technical property, scientific and academic qualification and truly experimental experience to analyze the work done by their peers.

However, considering the large and growing volume of incoming information and size of the scientific academic and technological communities, there is no way to conduct a thorough and individual analysis of each

graduate process and product. Indicators appear as auxiliary tools in the evaluation process in order to show trends and point out groups that stand out in a negative or positive way.

The manifestos show the importance of practicing a scenario of complementarity between a quantitative evaluation based on metrics and an evaluation based on the quality assured by peer review.

WHAT SAID THE MANIFESTOS	
LEIDEN MANIFESTO	<i>Principle 1: Quantitative evaluation should support qualitative, expert assessment</i>
THE METRIC TIDE	<i>Humility: recognising that quantitative evaluation should support – but not supplant – qualitative, expert assessment</i>
DORA	<i>When involved in committees making decisions about funding, hiring, tenure, or promotion, make assessments based on scientific content rather than publication metrics.</i>

The Brazilian evaluation process has peer evaluation as fundamental principle, through the formation of national committees that conduct the various evaluation activities. The area coordinators manage the activities together with ad-hoc consultants. Indicators are used as an auxiliary tool to analyze merit, not being the absolute and sovereign parameter. The qualitative analysis of content is considered fundamental for the evaluation and it is only possible to be done by the consultants.

2. The importance to combine indicators and the perception of its limitations

Every type of indicator has its advantages, but also limitations. OECD states the main properties and elements that characterize a good indicator. It alerts that the misuse of the indicators leads to distortions in any decision making. Managers and evaluators need to have feasible measurement indicators that are simple to obtain from reliable sources and at regular

intervals. They should also be stable over time and with a transparent and auditable way of obtaining them (OECD, 2003).

The manifestos say of the importance of not basing the evaluation on a single indicator, since it may not represent all the complexity of a research system and lead to a biased analysis.

WHAT SAID THE MANIFESTOS	
LEIDEN MANIFESTO	<i>Principle 9: Recognize the systemic effects of assessment and indicators</i>
THE METRIC TIDE	<i>Diversity: accounting for variation by field, and using a range of indicators to reflect and support a plurality of research and researcher career paths across the system.</i>
DORA	<i>Use a range of article metrics and indicators on personal/supporting statements, as evidence of the impact of individual published articles and other research outputs</i>

The Leiden Manifesto alerts about the influencing effect of the indicators used in the evaluation on how researchers behave in order to meet criteria and ensure, in return, the support of evaluation and funding agencies.

The graduate evaluation in Brazil is based on several dimensions and a set of indicators that represent the entire universe that involves the activities of a graduate program in its role of people training and generation and dissemination of knowledge.

The Evaluation Form is composed by five topics to be evaluated for all programs, as shown in Table 1. Each topic has items that detail what should be analyzed, besides its own indicators (See Appendix for a more detailed description of itens and weights).

Table 1. Topics of Evaluation Form and description

TOPIC	WHAT IS EVALUATED?	DESCRIPTION AND EXAMPLES OF INDICATORS
PROGRAM PROPOSAL	Curricular structure, planning and infrastructure	Qualitative analysis based on texts informed by the programs, which describe its scope in terms of curricular proposal, areas of expertise, disciplines, research projects, planning for student training, improvements and future development and infrastructure for teaching, research and extension.
PROFESSORS	Teaching profile, training, experience and performance in the program	<p>Profile of faculty considering academic degree, experience and suitability for program proposal, adequacy and dedication to research and training activities and the distribution of research and educational activities among professors.</p> <p><i>% of professors with doctor's degree</i></p> <p><i>Number of master's and doctoral degrees advised by each professor</i></p> <p><i>% of professors participating on funded research projects and providing scholarships</i></p> <p><i>% of professors with some undergraduate activity (teaching, student advising)</i></p>
STUDENTS, THESIS AND DISSERTATIONS	Quality of thesis and dissertations and efficiency in student training	<p>Student's profile, its participation and performance in the program, scope and quality of the thesis and dissertations produced.</p> <p><i>Qualitative analysis of the content of thesis and dissertations</i></p> <p><i>Number of master's/ doctoral students enrolled and graduated</i></p> <p><i>Average time for graduation</i></p> <p><i>Number of student authors of intellectual production</i></p> <p><i>% of students enrolled/ graduated in relation to the number of professors</i></p>

TOPIC	WHAT IS EVALUATED?	DESCRIPTION AND EXAMPLES OF INDICATORS
INTELLECTUAL PRODUCTION	Qualified output	<p>Intellectual production is quantified and qualified. The qualification is made through the products classification (books, conferences, technical and artistic products) and journals, called Qualis.</p> <p><i>Balanced distribution of productions in relation to the faculty</i></p> <p><i>Participation of students as authors of productions</i></p> <p><i>Average teacher output</i></p> <p><i>Number of articles per Qualis stratum per year or Number of products per stratum per year</i></p> <p><i>Number of productions linked to thesis or dissertations</i></p>
SOCIAL INCLUSION	Impact, collaboration, graduated follow-up	<p>Regional, national and / or international impact of the program, collaboration with other programs, institutions and research centers, joining activities with the community.</p> <p>Qualitative analysis of:</p> <p><i>Extension or equivalent activities with local communities.</i></p> <p><i>Policies aiming at the access and permanence of teachers at public basic education.</i></p> <p><i>Activities in basic education and high school, with production of didactic material and training</i></p> <p><i>Prominent academic activities, such as awards, participation in scientific societies, scientific divulgation, special consultations, participation in councils.</i></p> <p><i>Cooperation with public and private sector</i></p> <p><i>Transfer of new knowledge to social sectors and qualification of professionals to deal with socially relevant issues.</i></p> <p><i>Evaluation of the social impacts of research projects.</i></p>

TOPIC	WHAT IS EVALUATED?	DESCRIPTION AND EXAMPLES OF INDICATORS
EXCELLENCE ATTRIBUTE	Outstanding performance compliant with international standards	<p>Programs that receive excellence grades (6 or 7) must demonstrate a level of performance compared to international standards of excellence</p> <p>Qualitative analysis of:</p> <p><i>International cooperation;</i></p> <p><i>International mobility of students and professors;</i></p> <p><i>Dissemination and communication in foreign languages;</i></p> <p><i>Foreign financing;</i></p> <p><i>Creation of reference centers;</i></p> <p><i>High-standard libraries;</i></p> <p><i>Publications with international visibility;</i></p> <p><i>Participation in conferences, international courses, consulting, publishing, visits;</i></p> <p><i>Awards, recognition or international prominence;</i></p> <p><i>Research networks;</i></p> <p><i>Joint degrees.</i></p>

Source: CAPES Evaluation Forms

3. Indicators are not unchangeable

Once defined, indicators do not need to become something unchangeable, meaning they will always require periodic updating and of course need to be able to reflect and adjust to changes. According to Brasil (2012), models and theories are continually improved and, therefore, the pertinence of the indicators should always pass through critical evaluations, with the care that the updating should be done from methodologically reliable research.

The manifestos explain the need to review and regularly update the indicators.

WHAT SAID THE MANIFESTOS	
LEIDEN MANIFESTO	<i>Principle 10: Scrutinize indicators regularly and update them.</i>
THE METRIC TIDE	<i>Reflexivity: recognising and anticipating the systemic and potential effects of indicators, and updating them in response</i>

Regarding the evaluation process, the Area Documents, which contain the guidelines and criteria established by the areas, are reviewed and updated on their evaluation metrics. Generally, they are updated and published every evaluation period, that is, every 4 years. However, criteria for activities carried out annually usually are updated more often, for example, new course proposal assessment guidelines and intellectual production classifications.

4. The limit of accuracy

One of the principles of a good indicator, according to OECD (2003), is Measurability, which represents the ability to reach and measure with the greatest possible precision and without ambiguity. The manifestos argue that the indicator needs to have a comprehensive scope and assured reliability, but should not be based on an illusory precision. What Leiden manifesto advocates is that there are uncertainties and inaccuracies in Science and Technology indicators, and therefore the accuracy of numbers should not be taken too far. That is, many indicators are represented by numbers, integers or decimals, but this must be interpreted with caution, sometimes the most correct is to consider a range of values.

WHAT SAID THE MANIFESTOS	
LEIDEN MANIFESTO	<i>Principle 8: Avoid misplaced concreteness and false precision.</i>
THE METRIC TIDE	<i>Robustness: basing metrics on the best possible data in terms of accuracy and scope</i>

Considering the CAPES evaluation scale, it is not just a regular sum of numbers. In order to reach the final grade, the evaluators make a balanced

combination of quantitative and qualitative criteria for the analysis of the various annual activities of the program, as shown in Table 1. Each topic and item in the Evaluation Form receives a scaled score: “Very Good”, “Good”, “Fair”, “Poor”, “Insufficient”, which, in a weighted way (Score x Weight of item), results in an integer numerical grade, without any decimals, in a scale from 1 to 7.

5. Publicity and transparency

One of the main points stated by all international manifestos is the publicity and transparency of sources of information, calculation methods and evaluation results, to be clear for everyone how the evaluation will be and had been conducted.

WHAT SAID THE MANIFESTOS	
LEIDEN MANIFESTO	<p><i>Principle 4: Keep data collection and analytical processes open, transparent and simple.</i></p> <p><i>Principle 5: Allow those evaluated to verify data and analysis.</i></p>
THE METRIC TIDE	<p><i>Transparency: keeping data collection and analytical processes open and transparent, so that those being evaluated can test and verify the results ;</i></p>
DORA	<p><i>Be explicit about the criteria used in evaluating the scientific productivity of grant applicants and clearly highlight, especially for early-stage investigators, that the scientific content of a paper is much more important than publication metrics or the identity of the journal in which it was published</i></p> <p><i>Be open and transparent by providing data and methods used to calculate all metrics</i></p> <p><i>Provide the data under a licence that allows unrestricted reuse, and provide computational access to data, where possible</i></p>

The entire graduate evaluation process in Brazil is based on the Data Collection application. It is filled by the program’s coordinator and validated by the educational institution, usually through the postgraduate dean. In 2013, the Sucupira Platform was launched, an online tool for Data Collection, submission of new courses proposals and various sources of queries.

The big change of scenery was that with the new Platform, the data became, as definitely as possible, public and open access. Thus, any interested party can consult the information of a program as they were filled, which guarantees transparency and consequently a greater cycle of trustable information, since a student or teacher can verify if the data informed by the program are correct and, if not, it is possible to request for correction.

Thus, the same set of data used by the committees at the time of evaluation can be obtained by any interested party, allowing making simulations or studies in an autonomous way, which is fundamental for the transparency and isonomy of the process.

6. Recognize the diversity between areas and the impact of research

Each evaluation area has its own characteristics, whether related to teaching and research or to the scope of publications and citations. The impact of the research also has differentiated coverage, which can be considered at local, regional, national or international level.

WHAT SAID THE MANIFESTOS	
LEIDEN MANIFESTO	<i>Principle 3: Protect excellence in locally relevant research.</i> <i>Principle 6: Account for variation by field in publication and citation practices.</i>
THE METRIC TIDE	<i>Diversity: accounting for variation by field, and using a range of indicators to reflect and support a plurality of research and researcher career paths across the system;</i>
DORA	<i>For the purposes of research assessment, consider the value and impact of all research outputs (including datasets and software) in addition to research publications, and consider a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice.</i> <i>Account for the variation in article types (e.g., reviews versus research articles), and in different subject areas when metrics are used, aggregated, or compared.</i>

CAPES establishes the general principles for evaluation, such as the standard of the Evaluation Form, its requirements and general items that must be

included in all areas. However, each of the 49 areas can customize its criteria and indicators, as long as it follows the minimum required in the regulations.

Thus, areas may give distinctive importance to intellectual products. Areas belonging to the Humanities generally give higher weight to books. Computer Science is one area that punctuates strongly scientific conferences. Biotechnology values heavily the production of patents, considering their innovative character. Indicators for each of the items also vary. In any case, all the production of the program is considered for evaluation purposes, not only scientific articles, but also books, conferences papers, technical and artistic production.

In the case of master's professional programs there is differentiation in the evaluation items. Technical products are valued most and there is a greater variety of types of graduate work, which can be a software development, technical report, protocol, production of didactic or instructional material.

There is no a priori differentiation on weight assessment of intellectual products as a function of language of the publication. Some areas recommend that publications be written in English, aiming for greater internationalization of the program, but this is not mandatory.

Those publications indexed in international databases such as Web of Science and Scopus are evaluated based on available bibliometric metrics. However, the committees also consider databases with greater regional coverage, such as Scielo (Scientific Electronic Library Online). Non-indexed ones are evaluated for their local or regional impact, considering the importance of the content developed and the objectives of the research.

For example, within the area of Agrarian Sciences, the development of an agricultural technique in a drought region or the planting of a specific cultivar for a particular region can have a significant impact locally and regionally, but not at a national level. The output (an article or book) from this research will probably not have a large number of citations, but that does not mean that it has not had relevant results. The same is valid for the valorization of programs in Tropical Medicine, in Brazilian public health

literature, teaching of local history among others, in which the less dissemination does not mean lack of quality or prestige.

7. Measure the quality and impact preferably than quantity

The key point here is to prevail the work and outputs' quality over merely taking a numerical count. The impact that research has on transforming the environment around it, in contributing to the solution of a society problem must also be considered.

WHAT SAID THE MANIFESTOS	
LEIDEN MANIFESTO	<i>Principle 2: Measure performance against the research missions of the institution, group or researcher.</i> <i>Principle 7: Base assessment of individual researchers on a qualitative judgement of their portfolio.</i>
THE METRIC TIDE	<i>When standard indicators are inadequate, individual researchers should look for a range of data sources to document and support claims about the impact of their work.</i>
DORA	<i>The need to assess research on its own merits</i>

Teachers and researchers are assessed for all their activities in the program, including not only published articles, but also their participation in research projects, belonging to scientific societies, student guidance, courses taught and professional experience. The focus of evaluation is the entire set of program activities and the quality of student training, but professor productivity is also individually analyzed in line with the scope of the program.

One of the evaluation items is Social Inclusion, detailed in Table 1. Part of the grade is attributed to collaborative networks, public non-academic essay, transfer and dissemination of knowledge.

Considering the persistent availability of information from the Sucupira Platform, the area committees are having the opportunity to prepare

and validate the indicators in advance, so, during the assessment itself at CAPES, only the final discussions and decision may be taken in plenary.

The Brazilian National Plan for Graduate Studies 2011-2020 stresses the importance of centring evaluation on “quality and excellence of results, the specificity of the areas of knowledge and the impact of results on the academic community and on society”. Indicators cannot be confined to mere productivism but the emphasis should be on reflecting “the relevance of new knowledge, its importance in the social context and the impact of technological innovation in the globalized and competitive world” (BRASIL, 2010).

Evaluation criteria takes into account characteristics of indicators that reflect innovation, the generation and transfer of knowledge, the qualification of students graduated, and the induction of research into relevant and strategic problems aiming at the development of the country.

8. Good indicators need structured and reliable data

For any calculation of indicator, the quality of the information is essential. Therefore data recording systems need to adopt standards for structuring and describing information. Currently, some standards have been adopted internationally, as well as single identifiers that guarantee the uniqueness and reliability of the registry.

For example, the unique identifiers of people (eg. ORCID, Researcher-ID, Scopus Author Identifier), organizations (eg. ISNI, GRID), journals (eg. ISSN), articles (eg. DOI), books (eg. ISBN). In addition, the adoption of data model standards (eg. CERIF) and data dictionary (eg. CASRAI), to allow interoperability of information between systems. These initiatives of standardization, registration, generation and dissemination of knowledge have been made in the European community, through EUROCRIS (<http://eurocris.org/>) and in the United States, through the VIVO initiative (for example <http://vivoweb.org/>), for example.

The Metric Tide makes explicit the importance of improving the data infrastructure that supports the management of information. The report

specifically cites the use of ORCID, which will be mandatory for all researchers in the next Research Excellence Framework (*REF*), DOI for publications and ISNI, for mapping institutions. It also states the need to improve the interoperability between research management systems for the purpose of exchanging information between funding agencies, publishers, universities, etc.

Actions in line with these guidelines are being discussed at CAPES for incorporation into the strategic planning of the Sucupira Platform for the coming years. The certification of the information is currently entrusted to the program coordinators, which is still validated by deans of the institutions. However, actions to audit data on international databases and other sources of information still need to be more widely adopted to ensure the best quality of information.

OTHER CONSIDERATIONS AND REMARKS

Intellectual Production and Classification

The intellectual output of the programs is not considered only quantitatively in the evaluation. It also goes through a qualification process, which is called production classification.

The qualification of scientific articles is done indirectly by the classification of scientific journals. This activity is called Qualis and is the only one required for all areas of evaluation. The detailed description and in-depth discussion of the Qualis process is described in the article untitled “Ten things you should know about the Qualis” *text in Portuguese* (BARATA, 2016).

The criteria for classifying intellectual production are not based on just one indicator. Qualis classification takes into account mainly journals’ bibliometric indicators provided by databases, specially Impact Factor from Journal of Citation Reports (Web of Science database), SJR and Cites per Doc from Scimago Journal and Country Rank (Scopus database), and h-index from Google Scholar.

Every cycle of evaluation analyses publications from the preceding four years, therefore indicators based on citation cannot be taken as absolute because the more recent articles will no longer have significant values.

In the case of CAPES evaluation there is still important aspects. A considerable part of the articles produced by graduate programs is not published in indexed journals at Web of Science or Scopus and, therefore, there is no way to obtain all or part of these bibliometric indicators. Also, Brazilian articles' impact factor is relatively low due to the low rate of international cooperation. Because of that, Qualis is not only based on metrics, but also on more qualitative criteria, such as the presence in large databases, the editorial line, rules for submission and evaluation of the articles, publication policies, periodicity and importance for the area.

There is a critique on the indiscriminate use of the Impact Factor as the first and often single parameter to evaluate and compare the scientific output of researchers and institutions. The DORA Declaration and the Metric Tide specifically cite the limitations of this indicator and suggest the adoption of other metrics to represent journals performance, such as the 5-year Impact Factor, Eigen Factor, SCImago, and h-index.

The search for new metrics seeking to overcome some limitations of the available indicators has risen. It is the case of the new metric proposed by NIH, the Relative Citation Ratio, a measure of an article's influence based on a relative field's citation rate (NAIK, 2016).

The way in which articles are indirectly classified as done in Qualis is often criticized, but on the other hand there is the challenge of evaluating a huge amount of productions that are registered each year. The volume reaches the order of 70 thousand books, 215 thousand conference papers, 6 thousand artistic productions and 367 thousand technical productions in average per year (Table 2). One of the options would be to use the number of articles citations (which does not guarantee the quality, but allows to have an idea of the impact) and to combine it with the journal classification given by Qualis. In addition, the evaluation should be focused only on the best publications, as encouraged by the PNPG 2011-2020 itself (BRASIL, 2010).

Table 2. Some data of the Intellectual Production during 3 years period (2013-2015)

INTELLECTUAL PRODUCTION			
	2013	2014	2015
ARTISTIC	5.001	6.434	6.853
BIBLIOGRAPHIC	521.540	512.584	508.442
Articles	196.629	205.340	202.163
Books	66.406	70.143	73.335
Conference Papers	233.148	208.330	204.678
Others	25.357	28.771	28.266
THECNICAL	360.645	366.011	374.623
Total	887.186	885.029	889.918

Source: CAPES database

When evaluating the documents of all evaluation areas, with regard to the parameters adopted for assessment in 2013, it was found that there were three general criteria for assessing scientific journals:

1. Bibliometric indicators supplied by national and international databases;
2. Their presence in indexed databases, particular those that are most relevant to the area;
3. Qualitative criteria which include: the existence of the editor responsible and an advisory editorial board and their respective authorities while academic leaderships, the clear identification of ISSN, the type of editorial line, the standards for submission, peer review assessment, the institutional affiliation of the authors, kinds of publications, periodicity, and other more qualitative criteria which involve the evaluation of journals in each area.

With regard to the first criterion, it was observed that 81% of the areas use the Impact Factor (Journal Citation Reports from the Web of Science) and of these, 74% regard it as the main definer of the classification. 56% of the areas make use of SJR (Scientific Journal Rankings – Scimago Journal & Country Rank, database of SCOPUS) and of these, 89% view it only as secondary for the classifications. Other indicators that are used are H-Index, Cites per Doc and some of a more specific kind for determined areas. Seven areas did not report using any bibliometric indicator (Figure 3).

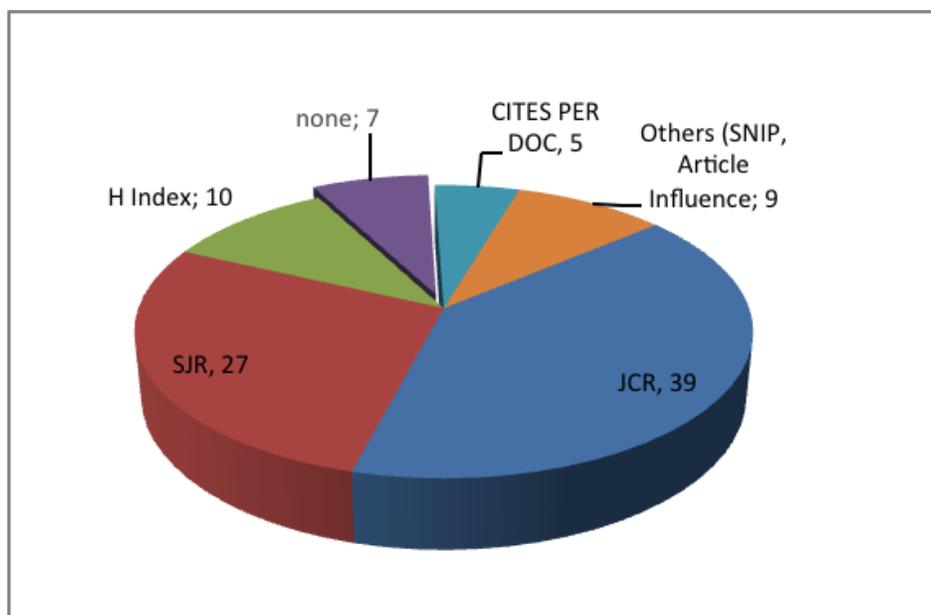


Figure 3. Number of areas and the bibliometric indicator used for Qualis classification

The wide diversity of indicators used by evaluation areas complies with the recommendation to avoid only using a single parameter.

With regard to the second criterion, 61 indexed databases were reported by all areas. 79% of the areas consider, as a parameter, the journal be indexed in Scielo, 63% in Scopus and 60% in the Web of Science, which are the main databases for indexing journals from all areas. The other databases are more specific – for example, PubMed and MedLine for Medicine and Health

areas, CAB Abstracts, AGRIS (Agricultural Science) and FSTA (Food Science and Technology Abstracts) for the Agrarian Sciences, PSYCINFO for the Social Sciences and BIOSIS for the Biological Sciences (Figure 4).

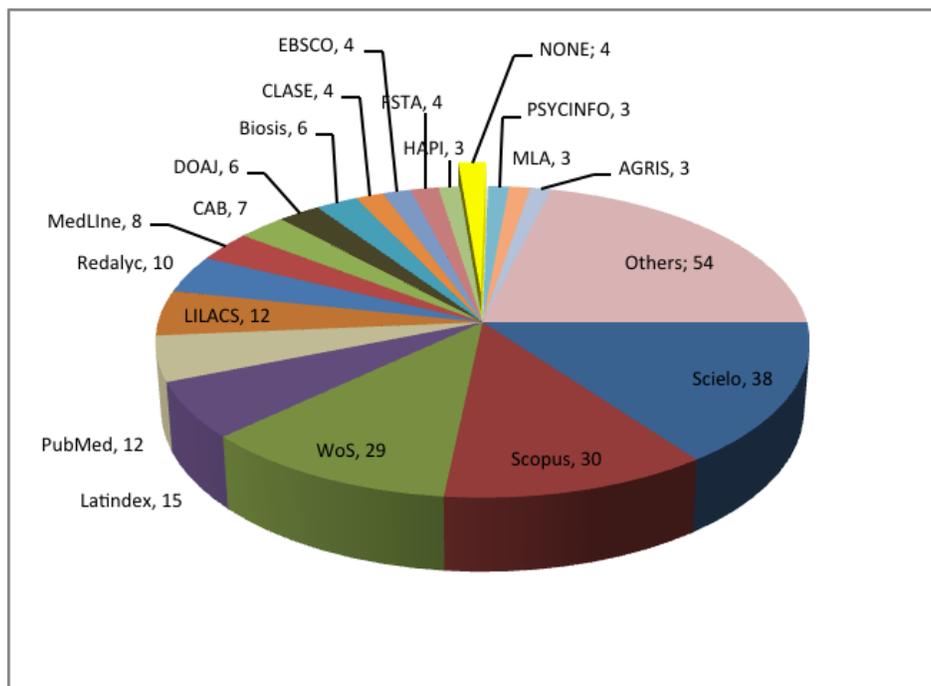


Figure 4. Indexed databases used by evaluation areas as criteria for Qualis classification.

Criteria for the classification of books and other productions

The evaluation of books – still clearly incipient – considers, as classification criteria, the registration in the ISBN, number of pages, characteristics of the editor and catalog, existence of well-prescribed editorial practices by the publishers, prevalence of editorial staff, the novelty and relevance of its content. They are classified in a scale from L1 (the lowest quality) to L4.

Some areas also perform the classification of conferences and technical products, however due to the great diversity of productions there is still

no broad consensus on how to conduct it in a systematic way. It should be stressed that there is no correlation at all or numerical rules for making the scale of books correspond to the Qualis scale for journals.

IMPROVEMENT OF INFORMATION TECHNOLOGY NEEDED FOR THE DEVELOPMENT OF THE SNPG

Quality of Information: Perspectives

The need to constant improvement of the process of evaluation and its tools to support peer analysis requires evolving stages for Data Collection that has been transforming from a merely administrative register to a tool for information management. Sucupira Platform permitted achieves a stage of more transparency for community but there are still actions to be taken in order to provide more quality of information. This includes integration with other systems of Science and Technology and collaboration with other actors involved, which allows reaching the knowledge management perspective (Figure 5).

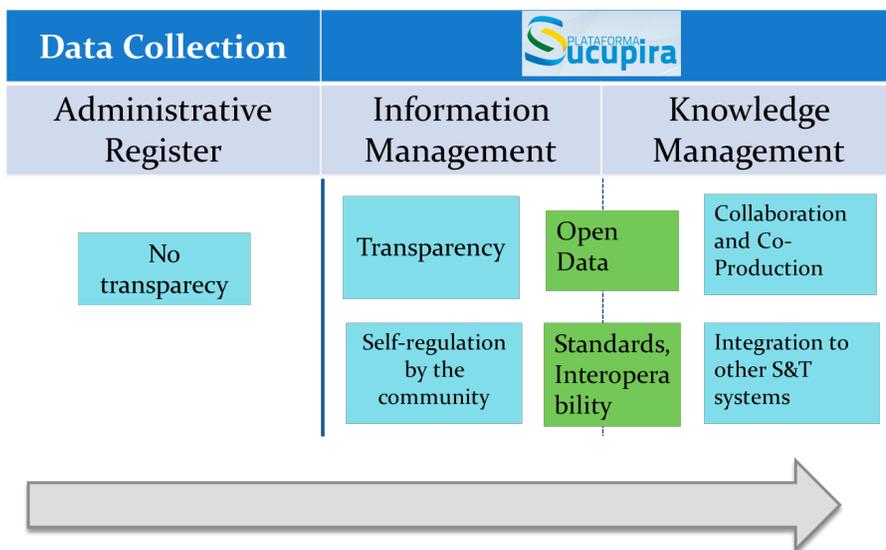


Figure 5. Evolving tool for Data Collection

DATABASE TOOLS AND VISUALIZATIONS

In 2015, CAPES introduced its “Acervo de Dados Digitais” (ADD) [Digital Data Repository] project, which carried out a task that involved organizing, standardizing and describing the data. In the case of the assessment, the Sucupira Platform operates as the gateway to data obtained from graduate programs that allow several kinds of consultation. ADD creates the repository of data, containing metadata documentation (“data which describe the data”, or in other words, information that is descriptive and can make it easier to understand the data) and microdata availability (detailed information in its most raw form). This provides information for society and can also enable the creation of spreadsheets and graphic visualizations (Figure 6).

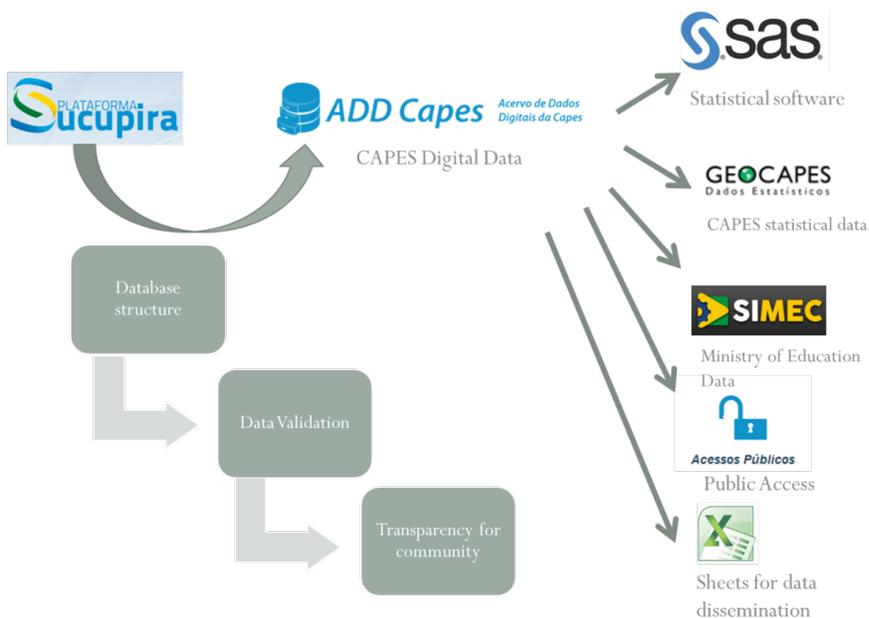


Figure 6. Framework for CAPES Digital Data Collection

Public access to the information from CAPES is made available by means of the Information Dissemination System (<http://sdi.capes.gov.br/>). In this space, it is possible to have access to the Open Data Portal, the Thesis and Dissertations Repository and the CAPES Transparency Portal, containing information about scholarships and projects funding.

FINAL CONSIDERATIONS

In the world of *big data* in which there is availability of large volume of data, funding and assessment agencies must be supplied with tools that can support decision-making. The indicators obtained – with increasingly more powerful algorithms for extraction – are the auxiliary tools for peers and managers. However, their use implies in being aware about their natural limitations.

With regard to the Brazilian model for assessment, the procedures and practices of CAPES, to a great extent, comply with the principles of the

international manifestos. Naturally, the model must be constantly discussed and improved, which is already being done in the form of an open dialogue with the community. Peer review assessment has been one of the leading principles of the evaluative activities carried out by CAPES, ever since it was first established and this has made the process reliable, as well as ensuring quality certification.

One key aspect of improvement is the adoption of measures that are aimed at refining the information technology that is needed for the development of the National System of Graduate Studies. This involves less operational work to filling in the data and greater reliability of the information and, consequently, ensuring greater representativeness of the indicators.

The international manifestos have enough validity to give rise to a discussion in different spheres of assessment and funding, thus encouraging improvement in procedures.

Every kind of indicator has its benefits but also its drawbacks. Managers and peers need to have feasible measurement indicators but it is necessary to assure their proper use, avoiding distortions in any kind of decision-making.

APPENDIX

Evaluation Form for Academic Programs

QUESTIONS/ ITENS	WEIGHT
1 – PROGRAM’S PLAN	0
1.1. Coherence, consistency, wide-ranging and updated in areas of expertise, lines of research, ongoing projects and curricular planning.	1.1 + 1.2 ≥ 60%
1.2. Planning of the program with a view to future development, taking note of the international challenges, its aims to improve teaching and its objectives with regard to enhancing social inclusion in accordance with the parameters of the area.	1.1 + 1.2 ≥ 60%
1.3. Infrastructure for teaching, research, and if applicable, extension.	1.3 ≥ 5%
2 – TEACHING STAFF	15 ou 20%
2.1. Profile of the teaching staff, considering their educational and professional experience and their adequacy to the Program Proposal.	2.1 ≥ 10%
2.2. Suitability and commitment of teachers to research and teaching activities	2.2 ≥ 20% (2.2 + 2.3 ≥ 60%)
2.3. Distribution of research and educational activities among teachers	2.3 ≥ 30% (2.2 + 2.3 ≥ 60%)
2.4. Contribution made by the teachers to teaching activities and/or research at undergraduate level	2.4 ≥ 10%
3 – STUDENTS – THESIS AND DISSERTATIONS	30 ou 35%
3.1. Number of thesis and dissertations concluded in the evaluation period averaged by teachers and students enrolled.	3.1 + 3.2 + 3.4 ≥ 40% (3.1 ≥ 10%)
3.2. Number of thesis and dissertations distributed into teachers advising	3.1 + 3.2 + 3.4 ≥ 40% (3.2 ≥ 10%)
3.3. Qualification of thesis and dissertations in the scientific productions of the program, evaluated through publications and other appropriate indicators in the area.	≥ 30 %
3.4. Efficiency of the training program for Master’s and PhD students with scholarships : time for Master’s and PhD conclusion and percentage of students with scholarships	3.1 + 3.2 + 3.4 ≥ 40%

4 – INTELLECTUAL OUTPUT	35 ou 40%
4.1. Qualified publications of the Program by teaching staff.	4.1 + 4.4 ≥ 40
4.2. Distribution of qualified publications into teachers	4.2 ≥ 30
4.3. Technical production, patents and other works regarded as relevant.	4.3 ≥ 5
4.4. Artistic work, in the areas in which this is relevant.	4.1 + 4.4 ≥ 40 (4.1 ≥ 4.4)
5 – SOCIAL INCLUSION	10 ou 15%
5.1. Inclusion and impact of the program at regional and/or national level.	5.1 ≥ 15%
5.2. Integration and cooperation with other programs and research centers	5.2 ≥ 20%
5.3 – Visibility or transparency given by the program to its performance.	15 a 20%

REFERENCES

- BARATA, R. C. B. Dez coisas que você deveria saber sobre o Qualis. [Ten things you should know about Qualis] *Revista Brasileira de Pós-Graduação*, v.13, n. 30, p. 13-40, 2016. doi: 10.21713/2358-2332.2016.v13.947.
- BRASIL. Ministério do Planejamento, Orçamento e Gestão. Indicadores. **Orientações Básicas Aplicadas à Gestão Pública**. 1ª edição. Brasília, DF: MP, 2012. 64 p. [Ministry of Planning, Budgetary Control and Management. Indicators – Basic Guidelines for Public Management]
- BRASIL. **Plano Nacional de Pós-Graduação 2011-2020**. Brasília, DF: CAPES, 2010. 608 p.
- DECLARATION ON RESEARCH ASSESSMENT (DORA). San Francisco Declaration on Research Assessment. 2012. Available from: <<http://www.ascb.org/files/SF-DeclarationFINAL.pdf?x30490>>. Viewed: 28 Feb. 2017.
- FERREIRA, M. M.; MOREIRA, R. L. (Eds.). **CAPES 50 anos: depoimentos ao CPDOC/ FGV**. Brasília, DF: FGV/CPDOC/CAPES, 2003. [50 Years of CAPES. Testimonies made to CPDOC/GVF - Getúlio Vargas Foundation]

- HICKS, D.; WOUTERS, P.; WALTMAN, L.; DE RIJCKE, S.; RAFOLS, I. The Leiden Manifesto for research metrics. *Nature*, v. 520, n. 7548, p. 429-431, 2015. doi: 10.1038/520429a.
- NAIK, G. The quiet rise of the NIH's hot new metric. *Nature*, v. 539, n. 7628, p. 150, 2016. doi: 10.1038/539150a.
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). OECD Environmental Indicators: Development, measurement and use. Paris: OECD, 2003. 37 p.
- WILSDON, J.; ALLEN, L.; BELFIORE, E.; CAMPBELL, P.; CURRY, S.; HILL, S.; JONES, R.; KAIN, R.; KERRIDGE, S.; THELWALL, M.; TINKLER, J.; VINEY, I.; WOUTERS, P.; HILL, J.; JOHNSON, B. **The Metric Tide**: report of the independent review of the role of metrics in research assessment and management. Bristol: HEFCE, 2015. 163 p. doi: 10.13140/RG.2.1.4929.1363.

eca50 ANOS
ESCOLA DE COMUNICAÇÕES E ARTES
UNIVERSIDADE DE SÃO PAULO

USP