

# INOVAÇÃO E CONCORRÊNCIA EM SERVIÇOS DE INFORMAÇÃO ACADÉMICA: DE EUGENE GARFIELD AO *GOOGLE SCHOLAR*

## *INOVATION AND COMPETITION IN SCHOLAR INFORMATION SERVICES: FROM EUGENE GARFIELD TO GOOGLE SCHOLAR*

Eduardo J. C. Beira  
Professor do Programa MIT Portugal  
Escola de Engenharia, Universidade do Minho  
ebeira@dsi.uminho.pt

### RESUMO

Discutem-se os processos de inovação envolvidos, desde a luta empreendedora de Eugene Garfield e do *ISI Institute for Scientific Information* nos últimos 40 anos até meados dos anos 90, e as mudanças que a entrada do *Google Scholar* tem provocado com o seu modelo de negócio baseado num mecanismo de inovação disruptiva. Os processos de inovação envolvidos são discutidos com base em modelos sociais de inovação empresarial. Apresenta-se evidência empírica comparando resultados de procuras no ISI/WoS e no *Google Scholar*, que mostram o efeito de *legacy* da arquitectura de dados do ISI/WoS e que mostram que nenhum deles é perfeito. Ambos subestimam o número real de citações, e, para os casos apresentados, as citações únicas das duas origens são dominantes. Discute-se o significado das actuais “guerras das citações” como uma continuação das anteriores “guerras das ciências” e como uma continuação da permanente procura do significado e fundamento para o conhecimento científico e para a actividade académica. Argumentamos que o modelo aberto e dinâmico do *Google Scholar* é muito mais coerente com a realidade, e permite uma visão mais completa dos múltiplos processos envolvidos na actividade científica e académica – algo que tem importantes implicações sobre as políticas para o sector.

**PALAVRAS-CHAVE:** Inovação empresarial. Inovação disruptiva. Modelos de ciência. Mercados de serviços – informação. Análise de citações.

### 1 INTRODUÇÃO

Foi durante o ano de 2004 que o mundo dos serviços de informação para fins académicos começou a sentir os efeitos de um sistema de que ainda se fazem sentir fortes réplicas e consequências. Em Novembro desse ano a Google lança o serviço *Google Scholar* (GS), um motor de busca especializado em conteúdos académicos, recorrendo a uma adaptação do famoso algoritmo do Google de procura na internet, mas em que agora a ordenação dos resultados da busca se baseia no número de citações da publicação, uma forma alternativa de medir “recomendações”, em vez da estrutura das hiperligações que apontam para um dado site na internet.

Curiosamente (ou talvez não) poucos dias antes do lançamento do *Google Scholar*, o grupo editorial *Elsevier* arrancava com o *Scopus*, um serviço em competição directa e frontal com os produtos do *Information Sciences Institute* (ISI), o serviço de referência e único até aí, e que durante décadas fora desenvolvido pelo *Institute for Scientific Information*, liderado por Eugene Garfield. No ano seguinte (2005) é lançado o serviço *Web of Science* (WoS) integrando cinco bases de dados originalmente do ISI.

Como tem sido reconhecido (Rethlefsen, 2009), no final de 2004 o mercado dos serviços de



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/).

DOI 10.5007/1518-2924.2010v15nesp2p132

Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, n. esp., 2º sem. 2010. ISSN 1518-2924.

informação baseados em índices de citações alterou-se dramaticamente, depois de quase quarenta anos de monopólio do ISI / WoS. Dois anos depois, Vine (2006) confessava que o mundo das bibliotecas, arquivos académicos e dos serviços tinha sido apanhado completamente desprevenido. Cinco anos depois as estatísticas de acesso aos respectivos sites falam por si (ver tabela 1):

		Visitantes únicos mil/mês	Visitantes totais mil/mês	Rank
Google	grátis	148000	3161000	1
Google Scholar	grátis	1304	2624	? (1)
Scirus	grátis	57	63	37415
Scopus	a pagar	29	33	70941
Isiknowledge	a pagar	49	174	43725

(1) Os dados de GS em particular não estão disponíveis, porque é um subdomínio de google.com, o domínio em primeiro lugar no ranking

**Tabela 1:** Estatísticas de acesso a serviços web de informação académica

**Fonte:** Extraído de <http://www.compete.com> (Agosto 2010).

Não só a estrutura competitiva do mercado se alterou de forma irreversível, como os próprios modelos de negócio do incumbente (WoS) foram postos em causa por violentos fenómenos de inovação disruptiva (GS) e de inovação de continuidade (*Scopus*). As implicações destas alterações vão muito para além do próprio mercado de serviços de informação e atingem mesmo o centro dos processos de gestão académica.

Neste trabalho faz-se uma análise dos 40 anos de desenvolvimento que culminaram nos actuais serviços WoS e *Google Scholar*, olhando para essas trajectórias sob o ponto de vista de posicionamento competitivo no sector, e de inovação tecnológica e empresarial, e procurando encontrar aí pistas para os desenvolvimentos futuros dos serviços de informação académicos. Usando dados empíricos, mostram-se limitações de ambos os serviços e refuta-se a ideia muito popular em certos meios académicos de que a base de dados WoS (p. ex. SCI do ISI) seja uma referência credível: na realidade ambos resultam em imagens incompletas da realidade. Mas argumenta-se que o futuro será de modelos disruptivos tipo GS e que os resultados desse modelo dão protagonismo e visibilidade às múltiplas facetas da actividade científica e académica, cuja omissão prejudica na realidade o desenvolvimento da ciência e do conhecimento. As limitações estruturais do WoS, uma *legacy* que não foi ultrapassada, deixa no ar muitas dúvidas sobre o seu futuro a prazo. As “*citation wars*” a que estamos a assistir e cuja continuidade a curto e médio prazo se antecipa, resultam de evoluções tecnológicas associadas à internet e que alteraram o cenário competitivo do sector. Afinal a ciência e o conhecimento são aventuras humanas cuja análise não faz sentido sem os olhares críticos de abordagens como a sociologia e a história da ciência. Finalmente, fazem-se algumas especulações sobre o futuro da concorrência neste mercado.

## 2 EUGENE GARFIELD, O EMPREENDEDOR, E O ISI INSTITUTE FOR SCIENTIFIC INFORMATION

Eugene Garfield (nascido em 1925) reúne as características de empreendedor e de homem de ciência, com uma rara capacidade para as abordagens multidisciplinares. Licenciado em química (U. Columbia, 1948), obteve os graus seguintes já como profissional e empresário: mestrado em “*library science*” (pela mesma Universidade, 1954) e doutoramento em linguística estrutural ligado à algoritmia não numérica (U. Pennsylvania, 1961). Garfield soube tirar partido das então novas tecnologias da informação emergentes, com recurso a computadores digitais e suportes de dados baseados em cartões perfurados, para lançar vários novos produtos de serviços de informação para a comunidade académica, em especial os *Current Contents* e depois os *Science Citation Indexes* (SCI). Em trabalho separado (BEIRA, 2010b) passam-se em revista os passos mais importantes dessa atribulada trajetória pessoal e empresarial<sup>1</sup>, até à integração do ISI nas actividades do grupo Thomson Reuters, já na década de 90. Os desenvolvimentos referidos suscitam vários comentários:

- a) **As raízes do SCI são exteriores ao mundo da ciência e constituem um reaproveitamento de ideais anteriores:** o princípio de um índice de citações não é uma criação do mundo da ciência, mas sim uma reutilização/recuperação de uma prática corrente entre juristas e profissionais do direito americano desde a segunda metade do século XIX (1873): o *Shepard Citator*, cujo modelo hierárquico de indexação serviu de modelo ao SCI do ISI, quase cem anos depois. Garfield escreveu na sua fase inicial em 1954, um texto em que fala de “*Shepardizing the literature of science*” no título e onde começa por analisar o *Shepard Citations*, descreve os esforços para desenvolver um código numérico para as várias publicações e caracteriza um “*citation index to science*” e as necessidades de equipamentos para ajudar o processo, encontrando para isso inspiração em equipamentos electrónicos auxiliares de leitura para cegos e ambliopes, desenvolvidos em 1947 pela RCA (Radio Corporation of America). Struchio e Tackray (1987) contam que Garfield tomou contacto com o *Shepard Index*, na sequência de uma carta de William Adair, em que este sugere o potencial da mesma metodologia de indexação para aplicação à indexação académica em geral. Garfield, que ignorava o *Shepard Index*, conta que correu para a biblioteca mais próxima, a Enoch Pratt Free Library e ficou entusiasmado com o que encontrou – não propriamente com o *Shepard Index* em si, mas sim com o método por trás do produto<sup>2</sup>. Entre as ideias que influenciaram Garfield, tiveram especial importância as do cientista inglês J. Bernal, um dos primeiros a falar de uma “ciência (quantitativa) sobre a ciência”, defendendo o planeamento da actividade científica com vista à resolução dos problemas sociais e uma radical alteração do modelo de publicação de artigos científicos (BERNAL, 1939, p. 269-297)<sup>3</sup>. Bernal fez parte do grupo dos “intelectuais vermelhos” britânicos que nos anos 30 a

---

<sup>1</sup>Para a reconstituição desta cronologia de Garfield socorremo-nos de diversas fontes: TROLLET; O'NEILL (1998), WOUTERS (1999a), GARFIELD (2007b), WILLIAMS (1997), GILAYREVSKII; CHERNY (2009), CHERNYL (2009), BROAD (1978). Na ausência de uma história empresarial compreensiva do ISI, recorreu-se também a diversas fontes, incluindo entrevistas de Garfield para o programa de historiografia oral da Chemical Heritage Foundation (STRUCHIO; TACKRAY, 1987; WILLIAMS, 1997), assim como a tese de Paul Wouters (1999a) e outras publicações (WOUTERS, 1998, 1999, 2003; WOUTERS; DUVAL, 2006). Os múltiplos escritos de Garfield, em grande parte disponíveis no seu site pessoal: <http://www.garfield.library.upenn.edu>, incluem também informação relevante em: <http://www.garfield.library.upenn.edu/pub.html>.

<sup>2</sup> “When I saw the Shepard I found the methodology that I needed for linking all these things, for indexing all these cited references that were sited in the (scientific) review (articles)”. citado por Thackray; Brock (2000).

<sup>3</sup> O sistema funcionaria como: o investigador, fosse qual fosse a sua área de investigação, enviaria o artigo para Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, n. esp., 2º sem. 2010. ISSN 1518-2924.

50 foram influenciados e seduzidos pelos “sucessos” da experiência soviética. O seu livro “*The social function of science*” (BERNAL,1939) teve na altura grande impacto. Uma segunda edição foi publicada pelo The MIT Press em 1967, com um prefácio de Bernal onde avalia positivamente o impacto das ideias do livro nos vinte cinco anos desde a sua publicação – mas na realidade Bernal dificilmente se reconheceria na organização actual da ciência. Nos seus escritos Garfield faz repetidas referências a Bernal, por exemplo, em Garfield, 2007.

- b) **O SCI nasce fora da comunidade científica, promovido e desenvolvido por um outsider do sistema, sem qualquer afiliação com uma instituição académica respeitada** e nos seus primórdios não foi claramente “bem amado” pela comunidade científica - os primeiros pedidos de apoio para a prototipagem e construção de um índice e citações foram inclusive rejeitados por comissões de avaliação da NSF *National Science Foundation*, acabando depois por beneficiar das necessidades da “nova” cultura de racionalização burocrática associada ao aparelho de investimento publico em I&D nos USA a partir do pós guerra, em especial dos anos 60.
- c) **O SCI não foi um produto de sucesso rápido nem fácil** – apenas sobreviveu porque financiado pelos lucros de outros produtos do ISI, em especial os vários *Current Contents*. O posterior sucesso pela viabilização das técnicas da cientometria acabou por se dar de uma forma e com consequências completamente imprevistas pelos seus criadores. Entre as décadas de 60 e 90 o conceito de indexação da literatura académica, o SCI evoluiu de simples sistema de procura de informação (encontrar informação publicada relevante) e de disseminação selectiva de informação para uma ferramenta de avaliação de resultados de investigação. Garfield (1998) viu-se mesmo na obrigação de recordar que o SCI nunca foi inicialmente pensado para estudos quantitativos sobre ciência<sup>4</sup>.

Se definirmos tecnologia como uma combinação de artefactos (computadores e programas), pessoas (utilizadores, em especial) e ideias (como usar os artefactos em novos serviços ou produtos) (HARGADON, 2003, 2006), a trajectória de Garfield e do ISI é um bom exemplo de inovação empresarial como recombinação de redes de pessoas, de artefactos e de ideias com criação de valor acrescentado – o paradigma da inovação tecnológica. Reunindo problemas e soluções de diferentes áreas, Garfield foi um típico “*broker*” de inovação, capaz

---

uma agencia central de publicações, com um certo âmbito territorial (regional, nacional, ), que submeteria o artigo á revisão editorial por um colegas (*peer review*). Uma vez aceite (com ou sem emendas e/ ou alterações) seria reproduzido fotograficamente (e não impresso na forma habitual). Certo numero de cópias seria imediatamente distribuído por todas as bibliotecas e por todos os cientistas que tivessem manifestado interesse (registado) para receber os artigos dessa área. O original do artigo ficaria arquivado (cópia *master* ou filme) para fácil reprodução posterior (a pedido, por exemplo, por cientistas que dele tivessem conhecimento por citação ou pelos abstracts). Bernal (1939) pretendia simplificar o processo de comunicação, reduzir drasticamente os custos e também aumentar muito a rapidez de circulação da informação – viabilizando inclusive a publicação de artigos de grande dimensão ou com características especiais difíceis de imprimir. É claro que no processo desaparecia todo o sector editorial associado (em geral privado). Na agenda de Bernal não deixaria de estar também uma alteração das relações de força e poder na comunidade académica. Parte integrante deste sistema seriam a publicação e distribuição de sumários (abstracts), a cargo da tal agencia central de publicações. Os sumários seriam distribuídos pelas bibliotecas não em livros, mas em cartões (fichas).

4 “The new generation of scientists and even scientometricians need to be regularly reminded that the *Science Citation Index* (SCI) was not originally created either to conduct quantitative studies, calculate impact factors, nor to facilitate the study of history of science.” (GARFIELD, 1998): “a nova geração de cientistas e mesmo de cientometristas precisam de ser regularmente recordados que o SCI não foi originalmente criado para fazer análises quantitativas, calcular fatores de impacto, ou facilitar o estudo da ciência”

de construir ligações entre “mundos” conceituais e de práticas diferentes, e encontrar e explorar novas combinações tecnológicas capazes de sustentar o crescimento de um negócio. Wouters (1999a, p.77-78) escreveu sobre Garfield:

“Without the drive, perseverance, social capacities, and technical expertise of Eugene Garfield, the immense task of building SCI would probably not even have been thinkable. It's not only a matter of personal traits, but also of being in the right place at right time. Garfield was an outsider on more than one respect, which made it possible for him to think about solutions other people would reject immediately. Garfield was well prepared for information services, Current Contents being the proof of that”<sup>5</sup>.

Os mecanismos sociais de inovação implicam duas etapas cruciais para as quais Garfield se mostrou especialmente dotado e capaz: para além de progressivamente encontrar a tal recombinação, Garfield foi um lutador capaz de construir e consolidar uma crescente comunidade de utilizadores aderente às suas ideias de como usar os serviços (produtos) introduzidos, incluindo uma rede académica cada vez maior interessada nas suas ideias sobre análise de citações e suas aplicações na investigação e também na avaliação da própria investigação académica: os mecanismos sociais de “*bridging and building*” que hoje se reconhecem como mecanismos fundamentais da inovação.

Os processos de inovação falham basicamente porque falha a capacidade de encontrar a recombinação viável entre ideias de “mundos” diferentes ou porque depois falha a capacidade para criar/interessar uma rede crescente de utilizadores nas ideias de uso dessa nova recombinação.

Nesses trinta anos, dos anos 60 aos 90, Garfield foi o explorador das novas oportunidades de negócio e de novas ideias (investigação bibliométrica e análise de citações) que as então novas tecnologias da informação abriram na recolha e processamento de dados sobre literatura científica, começando na década de 60 com cartões perfurados e um IBM1401 com 16K. Wouters (2000) classifica mesmo Garfield como o “alquimista” que encontrou a combinação única do “conceito, computador e literatura”, um belo exemplo de aplicação de um conceito antigo (indexação de citações) num novo contexto (ciência) com uma nova tecnologia (computadores). Mas no processo, como visto, todo o conceito se transformou de modo fundamental. Para isso contribuiu a experiência diversificada de Garfield, mas também as suas qualidades pessoais. Ele próprio se definiu como um vendedor nato:

“I always think that I have been a consummate salesman. I was always very eager to talk to people about our products, and pursue leads. I was a supersalesman. That was something I didn't leave entirely. I still have a temptation to get on the phone and sell somebody our products. And I always tended to engage in almost constant dialogue with the customers”<sup>6</sup> (Sturchio; Thackray, 1987).

---

5 “Sem o entusiasmo, perseverança, capacidades sociais e competência técnica de Eugene Garfield, nunca teria sido possível imaginar a enorme tarefa de construir os índices de citações científicas (SCI). Mas não se trata apenas de traços pessoais, trata-se também de estar no sítio certo no momento certo. Garfield era um *outsider* em mais do que uma dimensão, mas que tornou possível pensar sobre soluções que outras pessoas teriam rejeitado imediatamente. Garfield estava bem preparado para serviços de informação. *Current Contents* é a prova disso”

6 “Sempre me considerei como um vendedor completo. Estava sempre ansioso por falar com as pessoas acerca dos meus produtos e em perseguir pistas de potenciais vendas. Era um super vendedor. Uma coisa que nunca perdi totalmente. Ainda hoje tenho a tentação de pegar no telefone e de vender os nossos produtos a alguém. Sempre me envolvi em quase constante diálogo com os clientes”

### 3 GOOGLE SCHOLAR

O *Google Scholar* (GS) é o resultado de uma pequena equipe de apenas quatro pessoas da Google, num período de alguns escassos meses (tipo período sabático) em que Anurag Acharya e Alex Verstak interromperam o seu trabalho corrente nas equipes do motor de busca na web da Google. Nesse período adaptaram o motor de busca para o caso específico dos conteúdos académicos, melhorando a indexação, a extracção automática de metadados e a ordenação dos resultados das procuras, com o objectivo de melhorar a performance da busca web quando aplicada a conteúdos académicos (BURKE, 2010; HUGHES, 2006).

Note-se que o plano original não era propriamente criar um motor de busca específico para conteúdos académicos, com base numa especificação de modelo de negócio, estruturado e pensado. O GS nasceu simplesmente de uma experiência que resultou numa demo interna, e que se transformou rapidamente num serviço.

Anurag Acharya é a face mais pública da equipe GS. Diplomado em 1987 pelo IIT (Indian Institute of Technology) e doutorado em 1994 pelo Carnegie Mellon, é por vezes citado como o “arquitecto” do GS, brilhante e afável (TENNANT, 2005). A sua passagem pelo serviço académico (Universidade da Califórnia, Santa Barbara) terá ajudado a criar uma visão própria do problema de procura de conteúdos académicos e das suas implicações. Ele próprio fala da sua penosa experiência pessoal para localizar documentação científica durante os seus estudos na Índia. O seu papel relevante no sucesso do GS tornou-o também um ícone na nova Índia com capacidades e ambições nas tecnologias da informação e da comunicação (ASSISI, 2007).

Mais tarde Acharya contaria que a parte técnica mais difícil do projecto foi a extracção da informação relativa aos documentos e respectivas citações. A extracção é basicamente feita por exploração da estrutura das páginas web, e outros formatos, recorrendo a técnicas de *scrapping* (desconstrução e reinterpretação de conteúdos). Indexar, criar os índices, de uma forma rápida e estável, terá sido a parte fácil. A sua experiência anterior de projecto de grandes sistemas distribuídos terá ajudado.

As questões legais de acesso aos conteúdos na web é (será) sempre um factor limitante para serviços de informação deste tipo. A equipe começou a falar com as editoras académicas em 2001. O processo foi lento e difícil, dado o cepticismo das editoras<sup>7</sup>, mas acabou por ter um sucesso suficiente para permitir o lançamento do serviço. Quando foi lançado, nem todas as grandes editoras aderiram ao serviço, facilitando o acesso aos dados nos seus sites para indexação para o GS. O caso mais notório foi a *Elsevier* - que no entanto viria pouco tempo depois, em meados de 2007, a aderir e a abrir o acesso a todos os conteúdos ScienceDirect. Esta atitude de abertura ao GS, depois da recusa inicial, tem um significado e uma importância que não deixa dúvidas. A American Chemical Society será um dos casos ainda não totalmente abertos ao GS, no que respeita aos últimos anos de algumas das suas revistas.

As primeiras impressões da revista *Nature* consideraram o serviço como “impressionante”, quer em escala, quer em funcionalidade (BUTLER, 2004). Garfield (2005) saudou o serviço com expectativa e como uma oportunidade, especialmente para os académicos de artes e

---

7 “Initially the scholarly publishers didn’t believe that scholars used a lowly thing like a search engine. I’m serious. I had to convince people that researchers do use this. It was a mindset that search engines are used for casual things and not for real research. The attitudes really have changed”. BURKE (2010).

humanidades. Escassos dois anos depois Vine (2006, p. 97) comentava que “nada tinha preparado o mundo das bibliotecas para a introdução do GS”.

A inovação dos motores de pesquisa da Google baseia-se em dois grupos de algoritmos proprietários: um de ordenação de resultados e outro de pesquisa e comparação de textos. A tecnologia de ordenação de resultados (*ranking technology*) é famosa, mas parte foi publicada enquanto resultado de investigação na Universidade de Stanford. As tecnologias de comparação de textos (*text matching techniques*) facilitam a desconstrução e análise de conteúdos, e são menos publicitadas.

Mas estas não constituíram a grande alteração introduzida pelo GS. Na realidade pode-se considerar a extensão/adaptação dos algoritmos fundamentais de pesquisa na web ao caso dos conteúdos académicos como uma simples inovação incremental relativamente ao modelo de negócio da Google. Mas para os incumbentes foi uma inovação muito diferente, como se discute adiante.

Um das consequências mais importantes (e mais polémicas) do modelo GS é o seu domínio aberto de procura, uma alteração radical relativamente ao modelo tradicional e fechado do incumbente ISI/WoS. O GS passa a considerar como literatura relevante sob o ponto de vista académico, não só os artigos “*peer reviewed*”, mas também “teses, livros, sumários, e outros artigos”, encontrados pelos *bots* de busca nos sites de “editores de materiais académicos, sociedades profissionais, repositórios “pré-publicação”, universidades e outras organizações académicas, incluindo actas de conferências” (GOOGLE, 2009).

Em vez de tentar definir o que é o espaço académico, por uma definição mais ou menos polémica de quais as publicações cujos conteúdos devem ser indexados, o modelo GS aceita indexar tudo o que encontra em sede académica ou que aparente conteúdo académico. Esta alteração de modelo é radical – não admira que os interesses corporativos (e sem dúvida muitas vezes legítimos) que se criaram à volta do modelo ISI/WoS tenham imediatamente começado a mostrar a sua indignação perante tal apostasia e proclamado mesmo as mais severas excomunhões. Como se argumenta mais à frente, este facto é, no entanto, um retorno a uma realidade que há muito a sociologia e a história da ciência têm evidenciado.

Enquanto que o ISI/WoS criou uma base de dados própria e privada com as referências (entidades) e as transacções associadas (indexação, citações), o GS adopta uma estrutura de dados muito mais dinâmica e flexível e consegue mesmo evidenciar referências citadas algures mas que não estão disponíveis na web. Nasce assim a diferença entre referência propriamente dita e uma *CITATION* (citação), assinalada como tal nos resultados de procura. Uma *CITATION* é citada por outros conteúdos digitais, mas o seu próprio conteúdo (digital) não foi localizado na internet. Por vezes pode acontecer uma duplicação e um mesmo conteúdo ter sido identificado por si e/ou como várias *CITATIONs*, conforme as edições e origens dos conteúdos. Esta heterogeneidade com que uma referência pode aparecer nos resultados de procura do GS tem sido criticada como uma limitação grave – mas na realidade é uma grande inovação relativamente ao modelo ISI/WoS.

O GS rompe com o modelo de sistemas de procura e arquivo baseados em metadados, e estabeleceu um paradigma diferente, capaz de extrair e interpretar os conteúdos de forma alternativa, por técnicas de interpretação textual e engenharia reversa de conteúdos. Como Sadeh (2006) sugeriu, os sistemas de procura com metadados são “*just in time processing*”, enquanto que GS é um sistema federado de busca com “*just in case processing*”. Uma

questão para o futuro é saber se este tipo de sistema de procura vai substituir, ou antes, coexistir com os sistemas baseados em metadados (TENNANT, 2005; WILKINS, 2006). Não é só a tipologia de documentos inventariados que é diferente: é também a largura de banda do espectro de áreas académicas cobertas. A ambição é inventariar toda a literatura académica – em todas as suas áreas, em todas as linguagens, e o mais “para trás” possível no tempo. O próprio Anurag Acharya, citado por Burke (2010), diz que esta ambição é mais facilmente anunciada do que na realidade fácil de atingir. Mas abre a porta a uma maior visibilidade e credibilização do trabalho académico interdisciplinar (sempre marginal nas bases de dados sectoriais) e uma maior visibilidade da produção académica de raiz não anglosaxónica.

Se o ISI/WoS ultrapassou o âmbito sectorial para o âmbito transversal das ciências, o GS aparece a estender a indexação a um âmbito muito mais largo de conteúdos académicos e não apenas relativos às ciências. A longa cauda da produção científica menos citada vem também agora ao de cima – a sua importância pode ser discutida, mas ao menos não é simplesmente ignorada como se fosse não existente.

#### **4 COMPARAÇÃO ENTRE ISI/WoS E GS**

Um grande número de artigos têm comparado os resultados de procuras entre o ISI/WoS e o GS, esgrimindo argumentos pró e contra (HARZING; VAN DER WAL, 2008; HARZING, 2008; NORUZZI, 2005).

Para se compreender a questão, parece importante rever a estrutura de dados de ISI/WoS e os importantes efeitos que daí resultaram. Na secção 4.1 começa por se discutir a arquitectura de ISI/WoS, baseado em estruturas de dados características dos anos 60 e 70, e o seu efeito de herança (*legacy*) sobre o WoS actual. Mostram-se depois esses efeitos através da discussão detalhada do número de citações encontrados no WoS para alguns artigos de Michael Polanyi, publicados nas revistas *Science* e *Naturwissenschaften*. Na secção 4.2 comparam-se os resultados obtidos a partir de WoS e de GS para esses artigos e para mais algumas publicações do mesmo autor, concluindo-se que afinal o número de citações comuns é muito limitado, e que o número corrigido de citações é sempre superior (por vezes mesmo muito superior) aos resultados de cada uma das fontes. Introduce-se o conceito de número de citações “oficial” e “total” (incluindo portanto citações “não oficiais”) no WoS, recorrendo ao “número de citações das citações”, uma métrica pouco usada dadas as dificuldades de acesso aos dados, sugere-se que GS tende a encontrar as citações “importantes” melhor do que WoS – embora isso possa depender das características da referência citada. Finalmente, apresenta-se um quadro de análise das citações WoS conforme a referência citada e a referência citante sejam ou não parte da lista WoS de publicações.

##### **4.1 A estrutura de dados de ISI/WoS e o efeito de “*legacy*”**

A arquitectura de um sistema de gestão de citações é clássica, sob o ponto de vista de sistema de informação, e pode-se resumir num conjunto de entidades (os artigos, referências, documentos, conteúdos) e num conjunto de transacções (as citações: uma citação de uma entidade por outra altera o valor do número de citações da entidade citada).

Um sistema de gestão de citações precisa recolher citações e identificar entidades - uma arquitectura em todo semelhante à arquitectura de outros sistemas de informação clássicos em

gestão, desde a contabilidade à gestão de inventários, etc. Usando a terminologia típica até aos anos 90, o conjunto de entidades cria um *ficheiro mestre* com informação sobre as entidades, que era actualizada (periodicamente) pelos ficheiros de transacções registadas nesse período. Esta arquitectura implicava habitualmente validações prévias ao processamento, sendo esta uma actualização das entidades a partir das transacções. Neste contexto uma transacção só seria completamente processada se envolvesse exclusivamente entidades previamente definidas ou identificadas. Foi com este tipo de arquitectura que Garfield começou a indexação de citações na ISI, e foi com base nela que desenvolveu os produtos e serviços da ISI (as entidades correspondiam aos *Current Contents*, as transacções aos *SCI Science Contents Index*). A unidade de trabalho seria um número de uma revista (publicação, o equivalente a um período de tempo numa aplicação de gestão como a contabilidade), da qual as entidades (artigos publicados) eram extraídas e adicionadas ao conjunto (ficheiro) de entidades. As listas de referências bibliográficas de cada artigo (entidade) eram extraídas e registadas nas transacções (entidade citante, entidade citada). Dadas as limitações de então nos suportes permanentes de informação digital, a codificação das entidades envolvidas na transacção era compactada ao máximo e de acordo com um *layout* fixo. No seu livro publicado em 1979, Garfield discute este tipo de questões em detalhe.

As decisões iniciais tomadas para a recolha das transacções (citações) haveriam de marcar de forma irreversível o sistema SCI do ISI: a identificação da origem (revista, publicação) da entidade citada e citante passava pela prévia codificação abreviada do respectivo título (da publicação) e o registo de um mínimo de dados de cada transacção (citação), de forma a minimizar o espaço ocupado pelo (crescente) volume de transacções arquivadas.

Um ponto fundamental é que por esta razão o título da entidade citada e o título da entidade citante não foram registados nos dados da transacção (citação). Estava aí criado um sério problema futuro para a escalabilidade funcional; uma limitação fundamental da base de dados ISI/WoS para efeitos de procura generalizada é não poder procurar pelo título (ou parte do título) da referência citante ou da referência citada – uma limitação da arquitectura inicial de dados.

Uma simples ordenação das transacções (citações) por entidade citada viabilizou na prática a análise de citações, na medida em que permitia conhecer quem citou quem (e quando e onde) – algo até aí (quase) impraticável. Garfield (1977) conta o espanto que as suas primeiras listas de artigos mais citados provocou nos autores notificados<sup>8</sup>.

O continuado acumular de transacções e entidades num sistema deste tipo criou problemas bem conhecidos de arquivo e acesso aos dados, que as tecnologias das bases de dados (hierárquicas, depois relacionais) e os suportes de arquivo digital de acesso aleatório vieram ultrapassar. Mas o efeito de *legacy* já então estava criado no desenho da arquitectura de dados: o *layout* fortemente estruturado e simplificado dos dados já então existentes marcaria a funcionalidade e escalabilidade futura do sistema.

Na génese desses dados estavam já inscritos os problemas que se viriam a manifestar mais tarde, especialmente quando a funcionalidade e os resultados começaram a ser comparados com outras alternativas:

- a) porque não era viável tratar todas as revistas e jornais pertinentes, rapidamente a ISI começou a construir uma lista de publicações que naturalmente passou a considerar como

---

<sup>8</sup> “The authors of these classics, upon being informed of the status of their papers, expressed varying degrees of surprise and delight” (GARFIELD, 1977).

- de mérito, e que naturalmente começou por ser uma lista das publicações habitualmente reconhecidas como as mais importantes nos meios científicos (em especial da química e das ciências da vida).
- b) mas porque as citações (transacções) podiam envolver entidades (artigos) não registadas (por exemplo, artigos publicados numa revista não objecto de análise pelo ISI) ficou o problema de orfandade parcial de muitas transacções, por a entidade citada não fazer parte das entidades reconhecidas no sistema, e portanto não poder ser validada, apesar da entidade citante fazer parte das entidades registadas (artigo publicado numa revista da lista do ISI).
  - c) eventuais e inevitáveis erros de citação pelos autores (e também alguns de transcrição e recolha) agravam a situação e contribuem para um número significativo de citações órfãs ou soltas, dada a dificuldade histórica de automatizar a resolução das eventuais discrepâncias e erros. Mas toda a lógica da análise de citações baseia-se nas transacções, não nas entidades. Logo o processo de contagem de citações acaba sempre por poder ter que envolver entidades estranhas ao universo do modelo. Daí advém muita confusão sobre o “rigor” e a “qualidade” da informação (número de citações em especial) no ISI/WoS e no GS. Nós argumentamos que a qualidade dos dados não é melhor no WoS do que no GS, sendo certamente imperfeita nos dois (ver exemplos a seguir).

Admitamos que a qualidade dos dados das entidades no WoS (artigos das publicações da lista ISI, obtidos a partir dos respectivos índices) é perfeita. Nesse caso apenas nas transacções que envolvam entidades fichadas (registadas a partir dos índices das publicações) é que a qualidade dos dados da transacção não terá dúvidas (assumindo, o que pode não ser pacífico, que não existem erros de transcrição). Se uma citação (transacção) envolver uma entidade externa ao ficheiro mestre de entidades, a qualidade dos seus dados depende da qualidade com que o autor do artigo fez a citação, e o sistema nada faz para tentar corrigir possíveis discrepâncias. O próprio WoS o reconhece e avisa no site WoS, apesar da manifesta contradição dos termos e do cuidado em escamotear as consequências<sup>9</sup>. No entanto este comentário é enganador: os esforços de unificação, para além da codificação dos títulos das publicações, são quase inexistentes.

Na prática, a principal consequência é que o número de citações que são anunciadas numa pesquisa inicial de artigos de revistas da lista ISI é na realidade apenas o número de citações tipograficamente correctas, publicadas por artigos das publicações da lista ISI, ignorando as citações com gralhas mais ou menos importantes e/ou com informação insuficiente dos artigos da mesma lista. O rigor do número “oficial” de citações anunciadas é, portanto duvidoso e pode deixar de fora um número indeterminado de citações (“não oficiais”) que não estejam formalmente correctas – o que pode ser significativo, em especial para os autores menos citados. Isso para já não falar em todas as citações em publicações fora da lista ISI/WoS.

## 4.2 Alguma evidencia empírica

Michael Polanyi foi um dos cientistas e filósofos mais brilhantes e influentes do século XX

---

<sup>9</sup> “Every cited reference in the citation index contains enough information to uniquely identify the publication that the reference points to. Because only essential bibliographic information is captured, and because author names and source publication titles are unified as much as possible, the same reference cited in two different articles should appear the same way in the database. This unification is what makes possible the Times Cited number on the Full Record page. However, not all references to the same publication can be unified. As a consequence, a cited reference may have variations in Web of Science” ([www.isiknowledge.com](http://www.isiknowledge.com)).  
Enc. Bibli.: R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, n. esp., 2º sem. 2010. ISSN 1518-2924.

(SCOTT; MOLESKI, 2005). Algumas das suas obras têm milhares de citações. Analisaremos alguns casos, resultantes de uma investigação bibliométrica em curso sobre a obra de Polanyi (BEIRA, 2010a).

**Caso 1:** Polanyi contribuiu com três artigos para a revista *Science*, que são facilmente identificados na lista obtida pesquisando pelo nome do autor em WoS, obtendo-se também o respectivo número (“oficial”) de citações.

Consideremos apenas um desses artigos: *Scientific outlook – its sickness and cure* (1957), com 7 citações no WoS. Facilmente se pode obter uma lista com os sete artigos (entidades, referências citantes) que o citam. No entanto, é possível identificar (pelo menos) outras quatro citações adicionais a esse artigo, e que não estão contabilizadas nas referidas sete citações que a ficha da entidade citada regista e anuncia.

Através de uma procura “*cited reference search*” com *Polanyi* como autor e *Science* como nome da publicação, obtém-se uma lista de entidades citadas, gerada directamente a partir das transacções. A lista gerada mostra 22 referências citadas. Três delas incluem a hiperligação “*View record*”, que dá acesso ao registo completo da referência ou entidade citada, e correspondem aos três artigos de Polanyi na *Science*. Pela data (1957) logo se identifica o artigo “*Scientific Outlook – its sickness and cure*”, publicado no volume 125, pag. 480. Mas na mesma lista das 22 referências citadas identificam-se facilmente mais quatro linhas que se referem ao mesmo artigo, todas com uma única citação – todas com o ano (1957) e o volume (125) correctos, mas uma com uma pequena diferença no nome do autor citado (*Polanyi* em vez de *Polanyi M*) e três com diferenças no número da página (468, 482 e 484 em vez de 480). Parece razoável considerar essas quatro referências “*a priori*” como potencialmente legítimas. A causa do erro pode advir do processo de transcrição/digitalização ou um erro de forma da citação pelos respectivos autores (não nos foi possível verificar).

Ou seja: a própria base de dados WoS contém dados facilmente identificáveis e que não são directamente reflectidos nas entidades citadas (de publicações da lista ISI), e que alteram muito (em termos relativos) o número de citações do artigo em questão (mais de 50%: de 7 citações “oficiais” para 11 citações “não oficiais”).

No GS uma procura directa por *polanyi scientific outlook* identifica imediatamente a referência citada, como uma *CITATION*, com 16 citações. Uma análise mais detalhada das listas de citações obtidas pelo ISI/WoS e GS mostra que apenas 3 referências citantes são comuns. Ou seja: ambas as fontes ficam longe das pelo menos 24 citações que se podem reconstituir conjugando as duas listas – 13 citações únicas ISI/WoS, 8 citações únicas GS e 3 comuns.

**Caso 2:** Em 1921, Polanyi publicou dois artigos na conceituada revista alemã *Naturwissenschaften*. Um deles, o *Die chemische konstitution der zellulose*, (v. 9, n. 17, p. 288) não aparece numa procura de entidades no WoS (por nome do autor e da publicação) - o que é estranho, mas mostra que mesmo para artigos de revistas da lista ISI a cobertura apresenta problemas, apesar de WoS reivindicar no seu site uma cobertura do *Science Citation Index Expanded* até 1899.

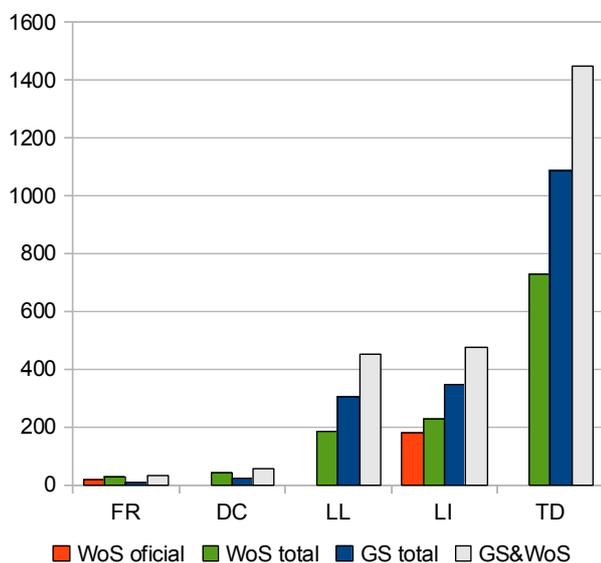
No entanto uma procura sobre as transacções (*cited reference search*) gera uma lista com 14 entidades citadas, das quais apenas uma exhibe a anotação “*View record*” (ou seja, identifica-a como referência “oficial”), mas que se refere ao outro artigo que, nesse ano, Polanyi publicou na referida revista, o *Faserstruktur im Rontgenlicht* (v. 9, n. 17, p. 337).

Por inspecção, facilmente se verifica que sete das 14 referências da referida lista dizem respeito ao primeiro dos artigos referidos (agregando 43 citações) e seis ao segundo artigo (26 citações agregadas), ficando um caso (com três citações) por resolver numa primeira análise. Ou seja, na base de dados ISI/WoS existe na realidade, informação sobre o primeiro artigo e respectivas citações – embora seja um artigo “oficialmente” não existente. E no caso do segundo artigo, com existência “oficial”, o número de citações “oficiais” (19) é inferior ao número de citações cuja informação realmente consta da base de dados (pelo menos 26).

Acresce um problema de rigor: ambos os artigos foram publicados em alemão, mas o título aparece em inglês no WoS. Noutros casos tivemos oportunidade de verificar a duvidosa qualidade de traduções de títulos do alemão para o inglês.

No GS os dois artigos exibem, respectivamente 22 e 7 citações apenas, com apenas nove e uma citações comuns a ambas as listas - donde se conclui que os dois artigos foram citados pelo menos 56 e 32 vezes.

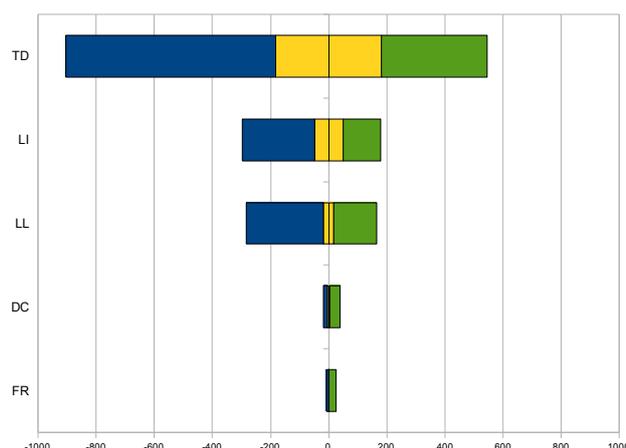
**Caso 3:** Para cinco publicações de M. Polanyi (ver figura 1), em que três são artigos (FR, DC e LI) e dois são livros (LL, TD), comparam-se os resultados a partir de ISI/WoS e GS. Três deles (DC, LL e TD) não têm citações “oficiais” no WoS. As outras duas (FR e LI) têm respectivamente 19 e 178 citações “oficiais”. No entanto, o número total de citações (incluindo as citações “não oficiais”) que é possível encontrar no WoS por *cited research search* é superior. Na figura 1 comparam-se o número de citações “oficiais” e “totais” no WoS, o número de citações no GS e o número total de citações “corrigidas” por conjugação dos resultados das duas listas de citações (WoS totais e GS).



**Figura 1** – Número de citações WoS (oficial, total), GS e “corrigido” para cinco publicações de Michael Polanyi.

**Fonte:** o autor.

A figura 2 mostra o número de citações “corrigidas” obtidas por conjugação dos resultados de procura ISI/WoS (citações “totais”) e GS. À esquerda do eixo vertical, os números de citações únicas GS (com valores negativos). À direita do eixo vertical, o número de citações “totais” únicas no ISI/WoS (com valores positivos). O número de citações comuns a ISI/WoS e GS é igualmente dividido entre os dois lados do eixo vertical.



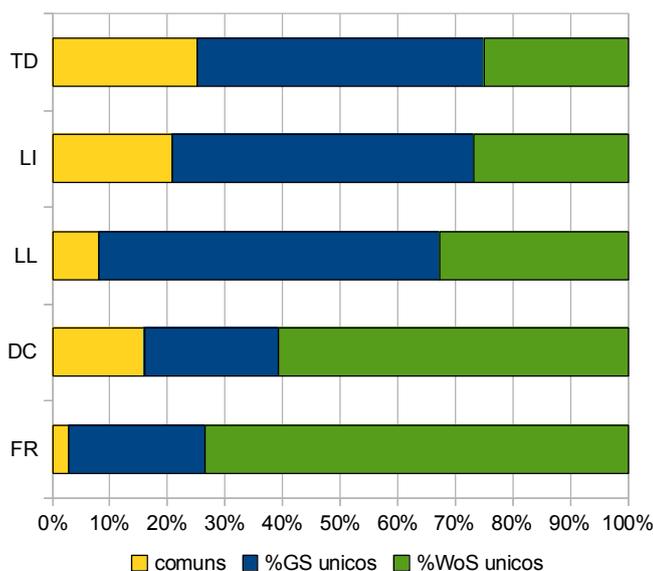
**Figura 2** – Origem das citações na lista “corrigida” de citações de cinco publicações de Michael Polanyi.

**Fonte:** o autor.

Relembremos que na figura 2, à esquerda do eixo vertical aparecem os números de citações únicas GS (com valores negativos). À direita do eixo vertical, o número de citações “totais” únicas no ISI/WoS (com valores positivos). O número de citações comuns a ISI/WoS e GS é igualmente dividido entre os dois lados do eixo vertical.

A estrutura do número de citações corrigidas aparece na figura 3, e da observação da mesma pode-se concluir que

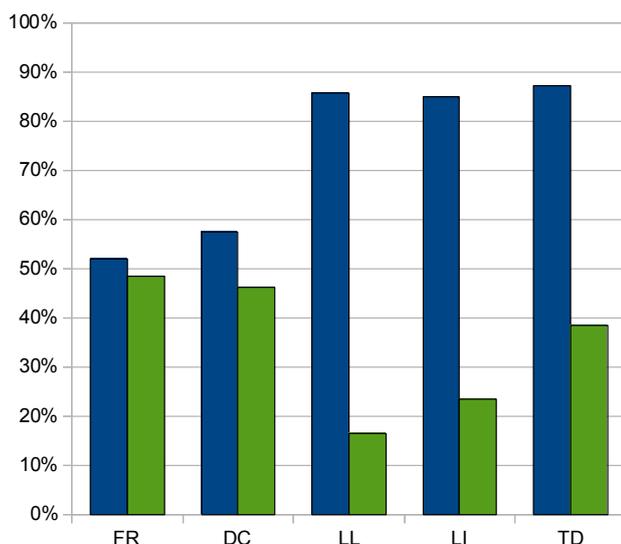
1. o número de citações comuns a GS e ISI/WoS é pequeno nos casos estudados, sendo sempre inferior a 25%, logo a maior parte das citações encontradas são únicas em cada uma das duas fontes de dados
2. nos dois livros (TD, LL) e no artigo LI (muito citado), as referências citantes únicas ao GS dominam largamente
3. nos dois artigos menos populares (DC e FR), as referências citantes únicas de ISI/WoS dominam largamente.



**Figura 3** - Estrutura da lista “corrigida” de citações de cinco publicações de Michael Polanyi, por origem.

**Fonte:** o autor.

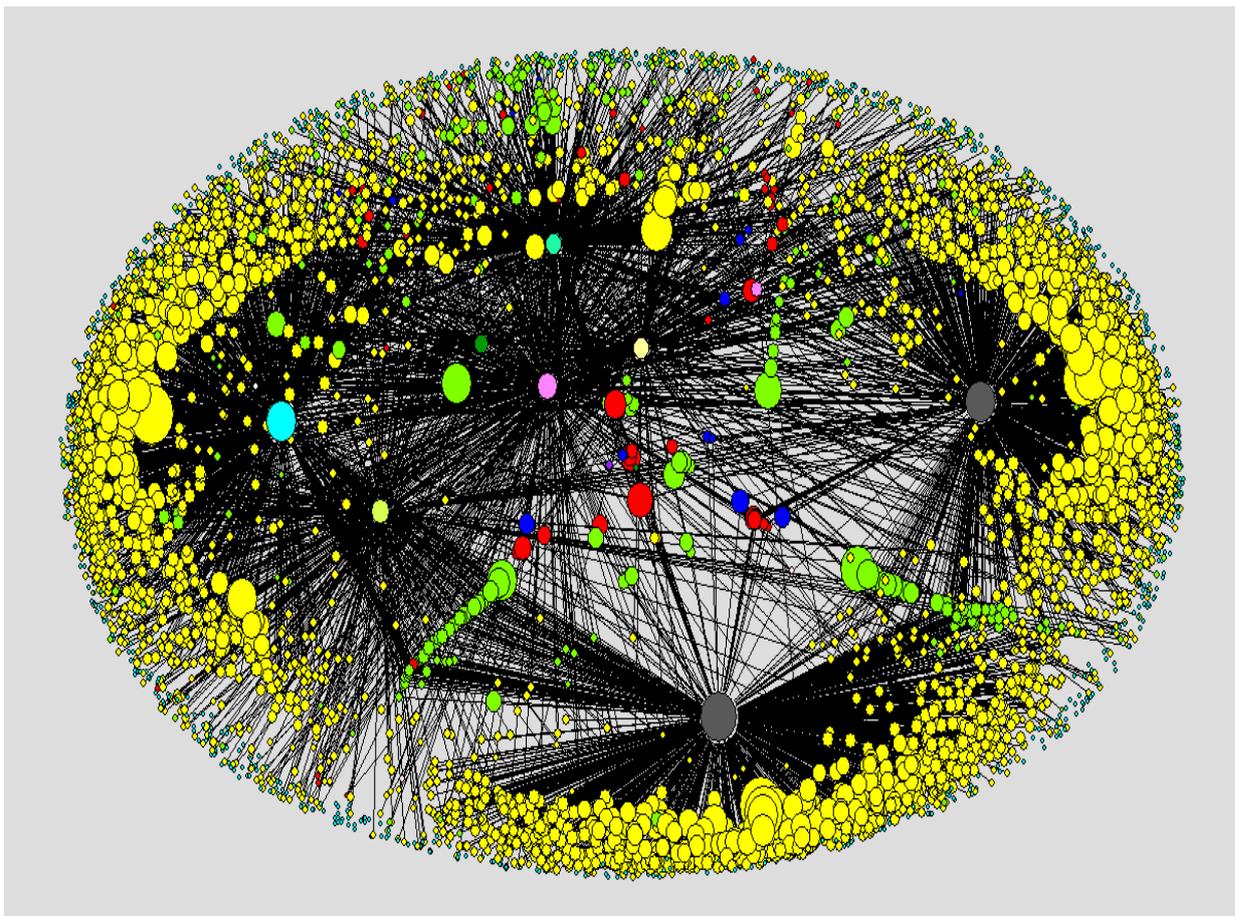
**Caso 4:** Se tomarmos o número de citações de uma referência como uma medida da sua importância (número de recomendações, positivas ou negativas igualmente valorizadas), então o número total de citações das referências citantes, isto é, citações de citações, que abreviadamente designaremos por “coc”, ou citações de segundo nível pode ser tomado como uma medida agregada da importância das referências citantes. Assim o quociente entre o coc das referências citantes captadas pelo GS e o coc da lista total de referências citantes, depois de corrigida pela conjugação das duas listas, representa uma medida do que é que os resultados da procura em GS captam da importância total da lista de citações corrigida. A figura 4 mostra que o GS captou mais de metade da importância das citações (entre 80 e 90% para os três casos mais populares) e que ISI/WoS captou sempre menos de metade.



**Figura 4** – Fracção das citações únicas GS e WoS representam do número de “citações de citações”, para cinco publicações de Michael Polanyi.

**Fonte:** o autor.

O conceito de citações de citações pode ser apresentado através das técnicas de visualização de redes sociais. A figura 5 mostra uma visualização das nove principais referências de Michael Polanyi (cinco livros e cinco artigos) e respectivas citações de primeiro nível, obtidas a partir do GS. As listas com as citações das citações foram obtidas com programas especialmente desenvolvidos para o efeito. A visualização foi construída com *Pajek* (NOOY; MRVAR; BATAGEL, 2005). A rede apresentada foi visualizada com o algoritmo de *Fruchterman Reingold* e tem cerca de 4800 nodos (referências) e cerca de 10000 ligações (arestas, as citações). As citações (referência citante, referência citada) constituem as ligações da rede, com espessura relacionada com o número de citações. A dimensão de cada nodo está relacionada com o respectivo número de citações (mas a relação é não linear). As cores dos nodos (referência) codificam o respectivo grau. A rede permite, por exemplo, identificar e analisar as referências que neste contexto mostram citações cruzadas.



**Figura 5** – Visualização de nove publicações de Michael Polanyi e todas as suas citações.  
**Fonte:** o autor.

**Caso 5:** Na tabela 2, organizam-se algumas conclusões relativas ao ISI/WoS, conforme a referência citante e a referência citada pertençam ou não à lista de publicações ISI/WoS. Para cada uma das quatro situações, assinala-se se:

- há (ou não) informação da referência citada acessível pela opção de procura (*search*) de WoS (ou por *view record* na *cited reference search*);
- a referência citante conta (ou não) para o número de citações “oficiais” WoS da referência citada;
- a referência citante pode (ou não) contar para o número “não oficial” de citações WoS (via *cited reference search*).

		Referência citada	Referência citada
		ISI	não ISI
Referência citante	Não ISI	a) SIM b) NÃO c) NÃO	a) NÃO b) NÃO c) NÃO
Referência citante	ISI	a) SIM b) SIM c) SIM	a) NÃO b) NÃO c) SIM

**Tabela 2-** Quadro de análise da visibilidade “oficial” e “não oficial” das citações em WoS.  
**Fonte:** o autor.

### 4.3 Consequências

Que se pode concluir?

- a) Que ISI/WoS está longe de ser perfeito e que não reúne condições de credibilidade para ser usada como referência, muito em especial para questões de avaliação (afinal “o rei vai nu”!);
- b) Que nem o ISI / WoS nem o GS captam a totalidade (ou sequer a quase totalidade) das citações e respectiva importância – afinal ambos têm problemas;
- c) Que a conjugação das duas origens de dados permite estimar de forma (provavelmente) mais correcta as citações de uma referência e respectiva importância;
- d) Que o número “oficial” de citações WoS subavalia, por vezes dramaticamente, as citações “totais” que é possível extrair do próprio WoS, e por isso não merecem credibilidade;
- e) Que WoS poderia facilmente reconhecer as citações “não oficiais” se investisse em ferramentas mais “inteligentes” de análise, tecnologias hoje relativamente acessíveis e que a concorrência já usa nos seus produtos.

A nossa principal conclusão dessa análise e da leitura da literatura publicada é que afinal os resultados do ISI/WoS são insuficientes, e a sua qualidade é duvidosa em muitos casos. Sendo incumbente único até há pouco tempo, a base de dados SCI do ISI/WoS foi, e muitas vezes continua, a ser considerada (e aceite) como a referência. Por exemplo, a convergência que por vezes tem sido encontrada entre WoS e *Scopus* é tomada como um sinal positivo – mas na realidade apenas mostra que ambas estão longe da realidade. Alguns acharão essa convergência positiva e tomam a divergência com o GS como negativa (para o GS). Nós achamos essa divergência positiva e clarificadora, porque evidencia a parcialidade da imagem da realidade que essas fontes (baseadas em domínios fechados de procura) transmitem.

## 5 UMA ANÁLISE DO POSICIONAMENTO COMPETITIVO: O QUE MUDOU EM VINTE ANOS?

Vale a pena fazer uma análise comparativa do sector, contrastando a situação há vinte (ou mesmo apenas dez) anos atrás e agora. Para isso recorreremos ao modelo de análise competitiva de Michael Porter, baseado nas cinco forças condicionantes da estratégia empresarial (PORTER, 1979; 1998a, b), um quadro de análise que avalia o atrito concorrencial dentro de sectores de actividade (que Porter denomina “indústrias”), por cinco forças competitivas: o poder negocial de fornecedores e de clientes, as barreiras à entrada de novos concorrentes, a ameaça de produtos ou serviços substitutos, a intensidade da rivalidade entre concorrentes. A combinação destas forças em cada momento determina como é que o valor económico se distribuí entre as empresas do sector por um lado, e por outro lado, os clientes, os fornecedores, os distribuidores, os substitutos e os novos potenciais actores que entrem no sector.

1. **Fornecedores:** os principais fornecedores a considerar neste caso são as empresas editoras de publicações académicas. Com as casas editoras seleccionadas pelo ISI para cobertura e indexação, o ISI foi tecendo ao longo dos anos uma forte (e legítima) teia de interesses mútuos. Muitos desses fornecedores eram (são) frágeis, com pouca força negocial. Mas outros têm grande força negocial, pelo menos potencial, com uma dimensão empresarial muito superior ao ISI. Mas na prática eram os principais interessados nessa cobertura, e até

mesmo na “defesa do castelo” contra a entrada de novos concorrentes no selecto grupo de publicações aceites para indexação pelo ISI. Logo na prática a posição competitiva do ISI era reforçada pelos grandes fornecedores, dada a posição única do mesmo. A perda dessa posição única transformaria, no entanto a situação, que vinte anos depois está em profunda transformação. Um dos principais fornecedores (*Elsevier*, habitualmente reconhecido como o maior editor académico do mundo) transformou-se num concorrente de peso, com um produto superior (*Scopus*) para o mesmo mercado alvo, com inovações incrementais, melhor integração, melhores facilidades de pesquisa por parte do título de citações, capaz de lidar melhor com as citações órfãs ou formalmente incorrectas, para além de alegar uma cobertura mais vasta de publicações, entre outras. No entanto, o domínio de procura continua fechado a um conjunto de publicações “peer-reviewed”. E um estranho ao sector (Google) lançou um produto concorrente, a *Google Scholar* de uma forma disruptiva, com uma cobertura superior e acima de tudo sem custos de acesso, democratizando de forma até aí impensável a busca, e mesmo o acesso, a conteúdos académicos, abrindo o imenso mercado dos até aí “não clientes” (não clientes por barreiras económicas e de interface no acesso ao WoS). A força dos fornecedores, a quem interessa acima de tudo visibilidade e facilidades de acesso ao ponto de entrada no seu produto, são agora muito maior dado terem mais do que um só incumbente por onde escolher. E foram esses mesmos fornecedores os primeiros a facilitar o acesso via web aos seus conteúdos digitais (GS), apesar da sua descrença inicial. A monumental teia de interesses laboriosamente alinhados está-se a desmoronar.

2. **Cientes:** tal como os fornecedores, a rede de clientes tinha há vinte anos os seus interesses claramente alinhados com os da ISI. Os principais clientes eram (são) as tradicionais instituições de I&D, e as comunidades académicas e de investigadores (Universidades em especial). E ainda mais um novo tipo de instituição para quem as métricas baseadas nos conteúdos das suas bases de dados se tornaram em importantes ferramentas de análise e avaliação de políticas e de projectos: as instituições financiadoras dos sistemas de I&D e as instituições governamentais que regulam e gerem as comunidades académicas e de investigação. Mas tal como com os fornecedores, a perda de posição única do ISI dá-lhes uma maior força negocial, sobre os preços e mesmo sobre a configuração dos produtos. A pressão depressiva sobre os preços tenderá certamente a aumentar. Uma vez mais a teia de interesses mostra rupturas importantes. No cenário actual, com uma oferta de procura gratuita pelo GS, a “guerra” passa pela luta pelo exclusivo da credibilidade académica, e explica muitas das polémicas que atravessam as comparações entre WoS, *Scopus* e GS (e mesmo outros). O simples facto dessa comparação ser tema tão polémico mostra não só a sua importância, mas também evidencia a capacidade corrosiva do GS relativamente ao posicionamento competitivo do WoS.
3. **Barreiras à entrada de novos concorrentes:** eram altíssimas na década de 90, são hoje muito mais baixas. O custo da tecnologia de suporte (computadores, digitalização, bases de dados) diminuiu drasticamente e é cada vez mais acessível, passando a basear-se na web em vez de acessos locais. As profundas mudanças nas tecnologias de procura na web vieram alterar tudo. Reconstruir os conteúdos das bases de dados ISI por outras vias, que não estavam ao alcance do ISI no passado, tornou-se agora viável, desde que o acesso aos conteúdos (rede de fornecedores) seja possível – e é hoje muito mais fácil do que no passado. As facilidades de acesso por metadados (*Scopus* e outros) e sem metadados (*Google Scholar*) abrem uma profunda brecha de fácil entrada na cadeia de valor há vinte

anos monopolizada pelo ISI/WoS. Há vinte anos era muito difícil emular o modelo de negócio ISI, dadas as dificuldades técnicas. Hoje a emulação do modelo é muito mais acessível – e o próprio modelo em si foi profundamente baralhado por um novo actor com uma abordagem profundamente disruptiva (*Google Scholar*). Os novos actores têm agora vantagens de custos e de “*não legacy*” num mercado praticamente sem custos de mudança de um fornecedor para outro, pelo utilizador final (ou pelo menos com custos muito reduzidos). A simplicidade das novas interfaces de procura e análise deixam com valor reduzido os custos de aprendizagem da exploração em ISI/WoS. A capacidade de retaliação do incumbente (WoS) é praticamente nula: reforçar a posição de “*gatekeeper*”, guardião e defensor da excelência académica, será das poucas políticas de retaliação que estarão ao seu alcance, mas cuja eficácia a médio e longo prazo parece pouco relevante.

4. **Substitutos:** em 1992 (e talvez mesmo em 2000) parecia difícil o aparecimento de substitutos. A digitalização de conteúdos académicos estava ainda muito no princípio, os custos eram elevados. Os motores de procura na web não pareciam uma ameaça em curto prazo – até porque os conteúdos (não digitais) estavam ainda muito preservados dentro de “castelos”. Mas entretanto a situação alterou-se radicalmente com a maciça digitalização de conteúdos entretanto ocorrida, e com a passagem para conteúdos nativos sob o ponto de vista digital. O ciclo de produção dos conteúdos começou também a alterar-se de forma desfavorável para o incumbente.
5. **Rivalidade dentro da indústria:** consequência de todas estas alterações, o nível de concorrência aumentou, passando de uma situação de único quase monopolista (ISI/WoS) para um mercado com vários operadores com ofertas concorrentes e, na nossa opinião, a longo prazo convergentes, pelo menos no que diz respeito ao primeiro nível (de topo) de publicações académicas. A realidade é que hoje o WoS tem alternativas e substitutos no mercado, com custos de migração insignificantes. A sua posição competitiva está muito diferente e mais frágil. Na nossa análise a tendência para uma posição competitiva cada vez mais desconfortável e difícil do WoS tenderá a agravar-se nos próximos anos.

Do que se apresentou, conclui-se que as alterações tecnológicas e as inovações introduzidas viraram do avesso os posicionamentos competitivos do sector. O caso do ISI (agora WoS) é exemplar da forma como em vinte ou menos anos se pode passar de uma posição altamente confortável e favorável, mesmo que baseada em vantagens competitivas alicerçadas em recursos tecnológicos e em processos únicos (uma situação típica daquilo que Porter chama de “posicionamento estratégico” (PORTER, 1998). O modelo ISI é agora mais fácil de copiar e com menores custos. O modelo do incumbente WoS não só se viu atacado por mecanismos de inovação incremental, como o *Scopus*, com um posicionamento baseado em vantagens de eficiência operacional, recorrendo uma vez mais à terminologia de Porter, como entretanto foi violentamente sacudido pela inovação disruptiva do modelo *Google Scholar*.

Claro que a internet teve um papel decisivo nessa transformação da arena competitiva em serviços de informação e na alteração do posicionamento competitivo do ISI/WoS. Tal como Porter (2001) previra, o maior impacto da internet seria sobre empresas condicionadas por altos custos de comunicação, de aquisição de informação e custos transaccionais – o cenário ISI/WoS. Neste caso aparecem as três situações em simultâneo: uma maior facilidade de acesso e difusão de resultados, ultrapassadas as bases de dados locais, os CDROMs e o papel impresso; um menor custo de aquisição de informação, com acesso directo ou indirecto a

conteúdos digitais, quer por metadados ou *scrapping* inteligente, e o fim dos *scannings* de versões impressas; e, finalmente um menor custo das transacções, com acessos gratuitos via GS e com interfaces muito mais simples e intuitivas.

## 6 O ATAQUE DISRUPTIVO

Clayton Christensen popularizou o conceito de inovação disruptiva (CHRISTENSEN, 1997; CHRISTENSEN e RAYNOR, 2003; CHRISTENSEN, ANTONY e ROTH, 2004). Uma análise do caso dos serviços de informação à luz do exemplo ISI/WoS versus GS beneficia deste quadro de análise.

Convém clarificar o conceito de inovação disruptiva – algo diferente do conceito tradicional de inovação “radical”. A tipologia de inovações de Clayton Christensen classifica as inovações em dois grandes tipos: inovações disruptivas e inovações de continuidade (*sustaining*).

As inovações de continuidade inovam produtos ou serviços numa linha de continuidade com o estado anterior da arte, mesmo que de forma radical. Melhoram a performance e/ou rentabilidade dos produtos e serviços anteriores de forma progressiva, sem grandes alterações dos mercados consumidores. As duas secções seguintes pretendem clarificar estes conceitos.

### 6.1 Inovações disruptivas

As inovações disruptivas são mais complexas: começam por ser inovações baseadas em paradigmas não alinhados com a oferta prevalecente dos incumbentes, habitualmente com uma baixa performance inicial, mas a baixo custo para os clientes finais, logo uma oferta abaixo do estado corrente da oferta dos incumbentes (razão porque em geral lhes passam despercebidas ou sem interesse na fase inicial). No entanto este modelo de “baixa qualidade” pode atrair duas classes importantes de clientes: os até aí “não clientes”, para quem existiam barreiras importantes ao consumo (económicas, culturais, etc) que se esbatem com a nova oferta, aumentando assim a dimensão do mercado, e também os “clientes sobre-servidos” para quem a oferta incumbente é excessiva (sentir-se-iam satisfeitos com menos funcionalidade e melhor preço) e com facilidade de mudança para um produto menos funcional mas mais acessível. Ambos os grupos de clientes são não clientes ou clientes marginais para os incumbentes, e em geral fora do seu radar. Para os incumbentes é muito difícil responder a este tipo inicial de ataque, não só pela sua falta de visibilidade, mas também porque implicaria uma canibalização da sua oferta para os seus grupos de clientes mais importantes.

Uma inovação disruptiva por um novo concorrente pode ser um ataque brutal aos incumbentes se, por mecanismos agora de inovação de continuidade, a sua performance melhorar rapidamente (e mais rápido do que o crescimento da procura de novas funcionalidades pelo mercado, que em geral tem uma taxa de crescimento muito baixa) e com isso conseguir criar um novo mercado muito mais alargado.

As inovações disruptivas começam por serem mais simples, mais baratas e menos capazes que a oferta dos incumbentes, atraindo novos consumidores na base (não no topo) do

mercado. Se tiver sucesso, podem-se transformar em brutais motores de crescimento empresarial a que os incumbentes podem ter grande dificuldade em responder.

Por sua vez um inovador disruptivo tenderá, uma vez instalado, a subir (por inovações de continuidade) na escala de valor e a atacar o mercado principal dos incumbentes. Assimetrias de motivação e de recursos/competências tornam em geral difíceis a reacção de um incumbente sujeito a um ataque disruptivo. Mudanças de políticas governamentais podem ter um papel relevante na abertura de espaços de oportunidade para novos paradigmas de oferta disruptiva.

## **6.2 Inovações de continuidade**

Por sua vez as inovações de continuidade podem ser por deslocação (na cadeia de valor) ou não, ocupando estas ultimas um continuo entre inovações incrementais e inovações radicais.

As inovações por deslocação ocorrem em pontos modulares da cadeia de valor. Um novo actor especializado que entra na cadeia de valor e oferece produtos ou serviços com um novo nível de integração, é um caso típico de deslocação, em que o valor passa de um nível da cadeia de valor para outro nível da cadeia. As inovações por deslocação tendem a ter os grupos mais importantes de clientes por alvo. Os incumbentes facilmente podem co-optar este tipo de inovações. A criação de valor pode ser significativa, mas é pouco sustentável em longo prazo.

As inovações por não deslocação, em que há mais do que uma simples mudança modular na cadeia de valor podem ser algo entre os limites da “inovação radical” e da “inovação incremental”, conforme a dimensão ou dramatismo da alteração do produto ou serviço. As inovações incrementais melhoram de forma marginal uma oferta anterior do incumbente e são a forma mais comum de concorrência empresarial por inovação. Mas é em geral fácil para o incumbente reagir e neutralizar eventuais vantagens de novos actores neste tipo incremental de inovação. No outro extremo, a inovação radical implica alterações complexas de natureza sistémica, com melhorias substanciais para os principais mercados do produto ou serviço. Os incumbentes estão geralmente melhor posicionados para integrar este tipo de inovação de continuidade, dados os recursos necessários e a necessidade de integrar de forma nova as múltiplas interdependências na cadeia de valor necessárias para a sua implementação. Incumbentes com sucesso em inovações radicais de continuidade podem alterar de forma permanente a estrutura de um sector – mas é sempre um campo de batalha preferencial entre incumbentes e muito ingrato para novos agentes.

## **6.3 A disrupção (GS) de uma posição competitiva confortável (ISI/WoS)**

Como é que este quadro de análise se aplica aos serviços de informação académicos, que estamos a discutir?

O ISI/WoS é o incumbente natural, com uma posição invejável até final do século passado. Os seus produtos eram topo de gama, caros e laboriosamente alinhados com os interesses dos seus principais clientes e fornecedores. A informação (estruturada) nas suas bases de dados parecia constituir um forte inexpugnável: replicar todo o investimento feito pelo ISI ao longo de décadas parecia uma loucura a um novo concorrente, que estaria sempre numa posição de desvantagem em relação ao incumbente, se o terreno concorrencial se desenrolasse no

mesmo plano tecnológico.

Os motores de busca da internet, e o Google em especial, são simples de usar, são gratuitos e com uma alta performance. Procurar informação na web era no entanto algo mundano e trivial no final do século passado. O volume de conteúdos académicos (artigos em especial) disponíveis em páginas web era ainda muito limitado. O centro do negócio do ISI até nem era a internet, mas sim as publicações, bases de dados em CDROMs e acessos (ao SCI) por redes privadas. E mesmo que os motores de busca localizassem um artigo de uma revista, onde estava a indexação, e as citações desse artigo? Na realidade o ataque disruptivo começa com o Google e outros motores de busca, mesmo ainda antes do *Google Scholar*: com um motor de busca na web gratuito começou a ser possível localizar informação sobre conteúdos académicos já disponíveis na web, sem necessidade de acesso ao WoS. Claro que os resultados não tinham a extensão (e a credibilidade) dos obtidos via WoS, para além de não incluírem informação sobre as citações e não terem facilidades de acesso a estas. Mas isso é precisamente uma das características iniciais da inovação disruptiva: a oferta inicial tem menos qualidade do que a do incumbente, mas essa baixa qualidade é aceitável para alguns utilizadores, e acima de tudo, para novos utilizadores.

Em 2000 já o ISI estava na internet – uma típica inovação de continuidade, muito mais incremental do que radical: os mesmos produtos e as mesmas bases de dados, a mesma lógica e o mesmo modelo de negócio, mas agora (também) a partir da web. A posição parecia forte. Os motores de busca (Google em especial) eram outro negócio, outro mundo, que nada tinha a ver com o mercado de topo do incumbente (a “excelência” académica).

Em 2004 tudo começou a ruir: o lançamento do *Scopus* é uma inovação de continuidade para os mesmos clientes centrais do mercado ISI/WoS. Como inovação de continuidade será mais radical do que incremental: baseia-se em acessos a conteúdos digitais (não locais) via metadados, por um novo concorrente, até aí fornecedor. Com funcionalidades melhoradas e mais inteligência embutida nos processos de procura e análise, passa a concorrer directamente com a oferta incumbente WoS.

O lançamento do GS introduziu uma forma alternativa de construir a indexação de conteúdos académicos, baseados nas citações das publicações, identificadas a partir das suas versões digitais por um processo automático. A recolha e interpretação de dados foi automatizada no mundo da web. De um momento para o outro um motor de busca especializado, um novo actor que nada tinha a ver até aí com os mercados dos serviços de informação para fins académicos, passa a competir com o incumbente (e com o *Scopus* também), e de forma disruptiva: gratuito, rápido, simples de usar, capaz inclusive de mostrar o número de citações de cada publicação (e a sua listagem e acesso). E ainda por cima com uma escalabilidade impressionante, com custos de escala marginais, e uma definição aberta de conteúdo académico. Rapidamente captou grande número de utilizadores académicos que até ai não tinham acesso aos serviços WoS – os tais não clientes do incumbente, ou os tais clientes capazes de se sentirem satisfeitos com um produto talvez não tão perfeito, ou com uma aura não tão académica ou institucional.

Ao longo dos primeiros anos a dimensão (escala) dos conteúdos recenseados foi aumentando, assim como a inteligência dos processos de procura foi melhorando. Começam a aparecer as

inquietantes comparações, mesmo que inicialmente muito reverenciais ao incumbente. Primeiro é a onda de ataque dos “guardiões do templo”: como seria de esperar, o incumbente fala acima de tudo de “falta de qualidade” do novo serviço, sob múltiplos argumentos. Depois começam a aparecer as análises empíricas que mostram que afinal o serviço incumbente afinal não é assim tão bom como se pregava: afinal o carácter compreensivo do incumbente é discutível (até aí não havia na prática termos de comparação), a base de dados do incumbente tem *também* imensos problemas e lacunas, afinal o padrão assumido (o incumbente) não é bem o que assumia ser.

Poucos anos depois o agente disruptivo vai ganhando uma credibilidade crescente. O incumbente tem (e terá) dificuldade em reagir: afinal como pode concorrer com uma oferta gratuita, cada vez mais abrangente, simples e eficiente, aberta, e que usa as tecnologias web mais sofisticadas, um mundo onde as competências do incumbente são claramente limitadas?. Adoptar o novo modelo, mesmo assumindo que o incumbente tem as motivações e as competências necessárias para tal, seria largamente incompatível com a manutenção dos serviços anteriores e respectivos preços.

A sua maior esperança estará porventura fora do próprio mercado – em convencer os agentes políticos do aparelho da ciência e da academia de que são o “padrão” ou referencial legítimo onde se devem basear as métricas que alimentam o aparelho burocrático de avaliação dos sistemas de I&D e de avaliação do ensino superior. É também a essa luz que muita da polémica actual deve ser lida. Estamos a assistir a uma batalha social provocada por um ataque de uma inovação disruptiva a um incumbente que tinha construído a sua posição com a ajuda e o interesse institucional dos próprios produtores e editores dos conteúdos académicos. Pelo menos com os editores, esse alinhamento está a desaparecer. E antecipamos que também com os produtores (académicos) isso tenderá a acontecer.

## 7 CITAÇÕES, CIÊNCIA E ACADEMIA

A “batalha das citações” é também uma batalha sobre o que é a ciência e a actividade académica entre dois pólos que têm dominado a polémica das últimas décadas:

1. a *ciência elite*, que ignora tudo o que não é publicado no topo da pirâmide e para quem apenas isso é relevante e tudo o resto simplesmente não existe. O acesso a publicar em certas revistas aproxima-se de um ritual de iniciação. A visão predominante da academia actual é largamente influenciada pelo conceito incremental de “ciência normal” de Thomas Kuhn (KUHN, 1962). Steve Fuller, o truculento filósofo americano (mas radicado no Reino Unido) que lançou a epistemologia social, escreveu que a visão khuniana de ciência “normal” combinava todas as qualidades da Máfia, de uma dinastia monárquica e de uma ordem religiosa, sem as salvaguardas constitucionais assumidas pelas democracias modernas (FULLER, 2003, p. 46).
2. a *ciência comunidade*, uma actividade emergente de um complexo e permanente borbulhar de diferentes contribuições em diferentes níveis e via diferentes canais de comunicação. Lewenstein (1999), por exemplo, dá uma boa imagem dessa visão de ciência e da forma como os múltiplos canais de comunicação interagem, a propósito do conhecido episódio da “fusão a frio” (comparar as respectivas figuras 1 e 4, p.18. 23).

As fronteiras abertas do GS permitem evidenciar que a actividade académica se faz por múltiplas facetas e tipos de conteúdos, e permite a sua valorização, e que a vida de um cientista não é exclusivamente uma vocação redutora para escrever artigos num certo conjunto fechado de publicações. Pelo contrário é uma via feita com múltiplas contribuições académicas ao longo de um contínuo processo de aprendizagem e de interacção em redes sociais.

No início deste século a OCDE (2000, p. 165) popularizou um modelo de desenvolvimento da inovação por parcerias entre a ciência e o mundo empresarial (indústrias) baseado na analogia do *iceberg*, com uma pequena parte visível, mas sustentado por uma enorme massa gelada invisível (BEIRA, 2002). A ideia de *iceberg* é interessante, mas demasiado frígida. Nós preferimos antes a analogia do vulcão – mas não essencial, a estrutura do conceito é a mesma.

Segundo esta versão do modelo as erupções imprevisíveis de inovação, *spin offs* tecnológicos, patentes, instituições conjuntas de I&D (“*research labs*”), e contratos de investigação entre a academia e as empresas, seriam o resultado visível de um permanente borbulhar menos visível (ou mesmo invisível) de iniciativas, como a mobilidade de investigadores e profissionais entre os dois mundos, as co-publicações (entre membros do mundo académico e empresarial, ou de publicações de membros de um desses mundos em publicações vocacionadas para o outro mundo), as conferências conjuntas, as exposições e feiras, os media especializados, os contactos informais entre as redes de redes profissionais envolvendo académicos e profissionais empresariais, e as ligações informais resultantes do fluxo de graduados para a indústria.

Neste modelo, as incubadoras, os parques de ciência de tecnologia, os *clusters* empresariais e as redes de intermediários na comercialização da ciência e da tecnologia têm o papel de “*brokers*” que ajudam na aceleração dos mecanismos de erupção.

As políticas governamentais devem ter um papel catalizador (acelerador) não só na parte visível, mas também e acima de tudo, na parte invisível do vulcão, através das políticas de educação, do mercado de trabalho, das políticas de compras, das políticas financeiras e de financiamento, das políticas regionais e de planeamento urbano e (especialmente) das políticas de concorrência e de regulamentação sectoriais e de I&D.

Para uma discussão mais detalhada do papel dos factores “não mercado” na inovação disruptiva recomenda-se a leitura do que é dito por Christensen; Anthony e Roth (2004, cap.4, p. 73 a 96). Este modelo é especialmente inspirador. Significa que concentrar a atenção nas erupções (mais ou menos imprevisíveis) e ignorar o permanente borbulhar invisível do processo é simplesmente retardar ou mesmo matar o processo. Promover as erupções implica alimentar a caldeira menos visível do processo, “por baixo” e não “por cima”. Ou seja: reconhecer a importância das actividades e experiências empresariais na progressão académica (e vice versa), promover e valorizar publicações de académicos na literatura empresarial (e vice versa), promover oportunidades para intercâmbio e contacto entre académicos e profissionais das empresas, etc. É claro que as regras da vida académica tendem em geral a desvalorizar estes aspectos. Se o modelo é realista, então significa que tendem a retardar as erupções do vulcão, sob o pretexto de seleccionar as erupções genuínas ou de “qualidade certificada” (pelos critérios de visibilidade).

Uma das consequências deste modelo é que publicações mesmo de menor nível ou qualidade académica (seja qual for o critério adoptado) são importantes. Fazem parte do tal borbulhar menos visível do vulcão. Logo, incentivar as erupções visíveis do vulcão, passa por incentivar em primeiro lugar o tal borbulhar invisível – é um processo de aprendizagem tácita para publicar em níveis superiores. E nesse processo muitos diamantes por lapidar (ou mesmo já lapidados) são produzidos. Impor um “*números clausus*” nas publicações científicas (uma implicação directa do modelo de negócio ISI/WoS) é arrefecer a parte menos visível do vulcão. Nesse sentido a abordagem do GS é renovadora, na medida em que valoriza e dá protagonismo a uma parte da produção até aqui quase invisível do interior do vulcão.

Com ligações fracas na parte não visível do vulcão será impossível criar fortes erupções. A polinização cruzada de ideias faz-se largamente nessa parte. A maior parte das políticas governamentais para ciência e tecnologia preocupam-se especialmente com a parte visível do vulcão (as erupções, e geralmente só depois de acontecerem) e com os agentes facilitadores, e ignoram (e por vezes desprezam mesmo) os incentivos ao borbulhar permanente e pouco visível do interior do processo vulcânico. Basta ver a subvalorização (ou simplesmente não valorização) da experiência e ligações empresariais de académicos, um facto que acontece em muitos países (BEIRA, 1999).

Este mesmo modelo vulcânico é também uma boa analogia para a actividade da ciência e para a actividade académica, com consequências semelhantes, mesmo sem entrar na discussão das “*science wars*” entre os post-modernos do construtivismo social e os defensores da tradição objectivista de ciência baseada em fundamentos metodológicos - não vamos aqui discutir se a realidade existe ou não por si, ou se é uma construção mais ou menos social. Mas é razoável considerar as “*citation wars*” como mais uma faceta, ou uma nova onda temporal, das “*science wars*” (ROSS, 1996; ASHMAN; BARINGER, 2001) e do que elas representam sobre a permanente procura dos fundamentos e das raízes epistemológicas do conhecimento. Como os sociólogos da ciência têm mostrado, por exemplo Latour; Woolgar (1979) e Latour (1987), a ciência é uma aventura humana sujeita a fortes pressões sociais e grandes paixões e dramas pessoais, que se exprime e constrói de múltiplas formas e interacções em variadas redes sociais. Reduzi-la a um só tipo e a um só universo fechado (modelo WoS) é castrador e limitante. Por isso o modelo aberto do GS promete progressiva visibilidade a todo o ecossistema académico e social que participa na actividade científica e académica.

Michael Polanyi abalou a epistemologia ao argumentar que a validade do conhecimento científico resulta do consolidar do confronto de ideias na “*républica da ciência*”, expressão de Polanyi que ficou conhecida na literatura como designando um modelo de ciência hierárquica, mas simultaneamente uma democracia de mestres e uma escola de aprendizes, com fórmulas de governação autónomas (POLANYI, 1962). Mas, na realidade, a visão de ciência de Polanyi é mais a de uma “*sociedade de exploradores*” na incerta e dolorosa procura da descoberta. “*Uma sociedade de exploradores*” é precisamente o título do ultimo capítulo do livro “*The tacit dimension*” editado em 1966, livro onde Polanyi explora pela última vez as consequências da sua filosofia baseada no conhecimento pessoal (POLANYI, 1958) sobre a actividade de conhecer em ciência. Nesse modelo é bem reconhecido o papel dos conflitos e das paixões na ciência, mesmo o carácter de permanente polémica e confronto que caracteriza a vida académica. E a chave da objectividade da ciência resulta de um processo emergente (no sentido da moderna teoria da complexidade): a aceitação (ou não) de factos e teorias através do turbilhão da polémica e do confronto na república da ciência. Na epistemologia de Polanyi o conhecimento científico tem os mesmos fundamentos de todas as outras actividades de

conhecer, baseado que é em componentes tácitos do conhecimento pessoal, e não em fundamentos próprios e autónomos.

## 8 CONCLUSÕES

O que é que podemos dizer sobre o futuro dos serviços de informação académica, tendo identificado as alterações que o sector conheceu nos últimos anos e o efeito do ataque disruptivo?

Quer a teoria, quer a prática, sugerem que a capacidade do WoS pode responder ao ataque são escassas. Parece fora de questão que o WoS co-optar o modelo GS. Poderia fazê-lo comprando o GS (ou a empresa mãe!), um cenário difícil de antever nas circunstâncias actuais. Ou poderia fazê-lo desenvolvendo de raiz um motor de busca do tipo GS – algo que se afigura pouco provável, por várias razões: pela assimetria de motivações e de recursos/competências, e de trajetórias e experiências empresariais (já atrás referido), mas também pelo problema financeiro posterior: depois de existir o GS, alguém estará disposto a pagar pelo acesso a um serviço semelhante?

A médio prazo é de esperar uma constante actualização do GS, quer para trás – na medida em que mais conteúdos antigos aparecem digitalizados na internet, quer para frente – novos conteúdos entretanto publicados começam em geral nativos pela internet. O processo tenderá a dar-lhe uma cobertura que incluirá a dos serviços incumbentes e ainda outros conteúdos académicos adicionais.

Sob o ponto de vista de serviço de procura e localização de informação, uma alternativa ao modelo GS, a existir, será certamente um serviço gratuito, o que implica uma entidade que possa sobreviver sem que esse serviço por si contribua directamente para os resultados empresariais. Recorda-se que a Google não cobra directamente pelo serviço GS aos seus utilizadores, mas cobra sim aos anunciantes de outros serviços – mas até agora não inclui anúncios nas páginas do GS.

Mas sob o ponto de vista de acesso à informação (conteúdos completos), a posição neutra do GS (não detém direitos sobre conteúdos, não é proprietária de publicações académicas) reforça a sua força disruptiva. A adesão dos conteúdos Science Direct à indexação pelo GS é bem elucidativa da força do modelo GS: uma casa editorial, mesmo com a importância e dimensão da *Elsevier*, não se pode dar ao luxo de ignorar a crescente popularidade do GS como uma porta de acesso aos seus conteúdos. Mas se a neutralidade editorial do GS é uma vantagem, já essa neutralidade joga a desfavor do WoS num modelo não gratuito, que certamente desmotiva a necessária autorização para indexação de conteúdos dos concorrentes. E sem conteúdos próprios associados (como a *Elsevier*) cujo acesso seja possível monetizar, fica apenas o negócio da procura de informação – uma luta em que não terá grandes hipóteses contra o modelo disruptivo do GS.

Logo, antecipamos que:

1. A WoS terá crescentes dificuldades em responder às alterações do mercado. O serviço de

localização de informação indexada não poderá continuar a gerar fundos no futuro, tal como aconteceu no passado. Restringir-se apenas ao mercado de conteúdos “*peer-reviewed*” será lutar ombro a ombro com *Scopus*, sem a força e o peso editorial da *Elsevier*. Mas tal especialização pode ser vulnerável: bastará o GS passar a adicionar uma marca que assinale se cada referência ou citação é ou não *peer-reviewed*, algo simples de fazer com base na publicação onde aparece.

2. A co-optação entre os dois principais incumbentes não é impossível. Quer a Thomson Reuters quer a *Elsevier* podem ser capazes de tal operação. A tabela 4, apresentada a seguir, compara as três principais empresas discutidas, e ainda a Microsoft e a IBM, sob o ponto de vista financeiro. A capitalização bolsista do Google é muito superior à do grupo Thomson Reuters e da *Elsevier*. Em qualquer dos casos o negócio não poderá estar na procura de informação, mas poderá estar antes em serviços especializados de bibliometria e cientometria – aspirar a ser o fornecedor “oficial” de dados (com base em literatura apenas *peer reviewed*) para efeitos de avaliação dentro do ecossistema académico. A fusão entre as duas empresas é um cenário muito razoável, que permitiria um espaço de sobrevivência num mercado tendencialmente decrescente, não só nos serviços de informação académicos, mas especialmente nos serviços de informação jurídicos e médicos.
3. De acordo com a teoria, um disruptor tenderá a subir na cadeia de valor pela adição de funcionalidades adicionais aos seus produtos/serviços, aumentando o seu valor acrescentado e entrando progressivamente nos segmentos mais altos do mercado. Neste caso significaria, por exemplo, o GS passar a oferecer serviços semelhantes aos que acabamos de referir (bibliometria, cientometria). Mas aqui, a teoria das motivações e recursos/competências dá vantagem aos incumbentes: esse não é o massificado mundo web da Google, mas o mundo restrito das complexas relações de política institucional dentro do sistema académico, onde *Scopus* e WoS têm maiores competências, experiência e motivação. Por isso não antecipamos que a Google entre por aí. Mas pode fazer algo que teria efeito semelhante: disponibilizar o acesso aos seus dados a um operador terceiro que se especializasse nesse tipo de serviços. Pode-se argumentar, com base nesta lógica, que um cenário interessante para a WoS ou *Scopus* seria acordar com o Google ser precisamente esse operador especializado, embora qualquer serviço pago sobre os dados GS levantará potenciais questões sobre as facilidades de acesso aos conteúdos da concorrência e a distribuição da receita. Mas o GS parece, antes mais, interessado em aumentar o universo de cobertura do que em especializar serviços. Em Novembro de 2009 o *Google Scholar* passou também a integrar informação legal americana, relativa a casos ao nível federal e ao nível estadual, incluindo acesso integral aos textos e citações dos casos e incluindo citações em documentos legais e não legais. Uma vez mais, quer a Reed *Elsevier* (serviço LexisNexis), quer a Thomson Reuters (serviço Westlaw) se vêm na mira do modelo disruptivo do GS, desta vez nos serviços de informação jurídicos e legais. Um acesso que tinha grandes barreiras (económicas) de acesso passa a estar disponível de forma simples, rápida, eficiente e gratuita, e com uma grande cobertura. Tal como o desenvolvimento inicial do GS, a componente legal terá sido desenvolvida como actividade marginal por técnicos da Google (ARNOLD, 2009). Os resultados de 2009 e posteriores de ambos os grupos parecem estar a reflectir o aumento de atrito concorrencial. Não são só os serviços de informação jurídicos que são (serão) afectados – a própria prática jurídica e dos tribunais também o será. De modo semelhante, as facilidades de acesso ao Google e ao *Google Scholar* estão a mudar a prática de serviços médicos. Nesse sentido, recomenda-se a leitura do editorial do *BMJ British Medical*

*Journal* (GIUSTINI, 2005), e ainda o livro de Miah e Rich (2008).

4. Métricas baseadas no GS tenderão a ter uma importância crescente. Mas aqui aparecem problemas associados ao uso do GS para além de motor de busca de informação, que precisariam de ser resolvidos. Em primeiro lugar não é possível fazer a descarga (*download*) dos resultados de uma procura no GS, o que significa que o seu processamento posterior é difícil. Esta “não funcionalidade” parece ser uma questão de política de produto, pela Google. Os actuais termos de acesso ao serviço GS impedem na prática uma oferta de ferramentas que o façam, pois o seu uso violaria esses termos de acesso. Ainda por cima, para as referências com mais de mil citações, apenas as primeiras mil citações são visíveis no GS. Estas limitações podem sugerir que o GS não pretende vir a oferecer facilidades mais avançadas de processamento de resultados de procura, como por exemplo as oferecidas pelo concorrente Microsoft, ou que então procura reservar esse tipo de especialização para o futuro. A Microsoft lançou em 2006 um serviço experimental: Windows Live Academic Search, disponível em <http://academic.research.microsoft.com>, da Microsoft Research Asia, sobre literatura académica *peer reviewed*, centrada em ciências da computação. A funcionalidade da interface de *output* é excelente, permitindo a identificação e visualização de redes de autores, e gráficos sobre a evolução temporal do número de publicações e citações. Permite também a identificação de nomes parecidos em formas avançadas de refinamento dos resultados da procura.
  
5. A curto e médio prazo, a comunidade académica precisará de conjugar resultados das diferentes fontes para ter resultados mais completos e métricas mais realistas para análise de citações. Esse poderá ser um interessante espaço para operadores especializados que usem os resultados de várias origens para obter listas corrigidas de citações. As técnicas de “*mash-up*” (FISCHER; BAKALOV; NUAREZ, 2009) começam já a gerar protótipos com alguma operacionalidade. Estas técnicas são centrais para a construção da chamada “*web semantica*”. Protótipos de um novo OCS (*Online Citation Service*) baseado neste tipo de abordagem têm sido desenvolvidos, por exemplo, pelo Abteilung Datenbanken, da Universidade de Leipzig (ver <http://dbs.uni-leipzig.de/en>).

	Capitalização	Empregados	Volume de negócios	Resultados operacionais	Resultados líquidos
	Agosto 2010	2009	2009	2009	2009
	bilhões USD	mil	bilhões USD	% vol. Neg.	% vol. Neg.
Reed Elsevier	9	17	5	13	7
Thomson Reuters	29	55	13	9	7
Google	145	22	24	35	28
Microsoft	209	89	62	37	30
IBM	157	399	96	18	14

Fonte: Google Finance (Agosto 2010)

**Tabela 4** – Dados financeiros das empresas Reed Elsevier, Thomson Reuters, Google, Microsoft e IBM.

**Fonte:** Google Finance (<http://www.google.com/finance>, consultada em Agosto 2010).

Uma última previsão parece fácil: as *citation wars* estão para durar, pois o que está em jogo não são só as bases de dados propriamente ditas e a sua qualidade, mas concepções de fundo

sobre a ciência e a academia. Por trás das *citation wars* escondem-se muitas *hidden agendas*, como seria de esperar. É a vida, a ciência e a academia.

## REFERÊNCIAS:

Todos os acessos a sites, em Agosto 2010.

ARNOLD, S.. *Cicumvallation: reed Elseviers and Thomson as vercingetorix*. *Beyond Search* blog, 27 Nov. 2009. Disponível em: <http://arnoldit.com/wordpress/2009/11/27/cicumvallation-reed-Elsevier-and-thomson-as-vercingetorix/>.

ASHMAN, K.; BARINGER, P. *After the science wars*. Routledge, 2001.

ASSISI, F.. *Anurag acharya helps Geogle's Scholarly leap*. *INDOLink*, 11 Jul. 2007. Disponível em: <http://timesofindia.indiatimes.com/world/us/Now-Google-goes-scholarly/articleshow/929365.cms>.

BEIRA, E. Universidade, industria e sociedade: desafios. In: Conceição, P. et al. (eds.). *Novas ideias para a Universidade*. IST Press, Lisboa, 1999.

BEIRA, E. *Academic and business views on partnerships for innovation*. . Disponível como Working Paper Mercados e empresas: dinâmicas e estratégias, WP18 (2002), U. Minho, em: [www.dsi.uminho.pt/ebeira](http://www.dsi.uminho.pt/ebeira).

BEIRA, E. *Michael Polanyi, fifty years later: A bibliometric approach, contrasting WoS and Google Scholar*. Disponível como Working Paper Mercados e empresas: dinâmicas e estratégias, WP102 (2010a), U. Minho, em: [www.dsi.uminho.pt/ebeira](http://www.dsi.uminho.pt/ebeira).

BEIRA, E. *Eugene Garfield, from ISI to Thomson Reuters: a timeline*. Disponível como Working Paper Mercados e empresas: dinâmicas e estratégias, WP104 (2010b), U. Minho, em: [www.dsi.uminho.pt/ebeira](http://www.dsi.uminho.pt/ebeira).

BERNAL, J. *The social function of science*. G. Routledge & Sons, 1939.

BROAD, W. Librarian turned entrepreneur makes millions off mere footnotes. *Science*, n. 202, 1978, p. 853-857.

BROWN, A. J. D. *Bernal: the sage of science*. Oxford University Press, 2007.

BURKE, A. Science 2.0 Pioneers. *SEED*, 20 Maio 2010. Disponível em: [http://seedmagazine.com/content/article/science\\_2.0\\_pioneers/](http://seedmagazine.com/content/article/science_2.0_pioneers/).

BUTLER, D. Science searches shift up a gear as Google starts Scholar engine. *Nature*, n. 432, 2004, p. 423

CHRISTENSEN, C. *The Innovators dilemma: when new technologies cause great firms to fail*. Harvard Business Press, 1997.

CHRISTENSEN, C.; RAYNOR, M. *The innovator's solution: creating and sustaining succesful growth*. Harvard Business Press, 2003.

CHRISTENSEN, C.; ANTHONY, S.; ROTH, E. *Seeing what's next: using the theories of innovation to predict industry change*. Harvard Business Press, 2004.

CHERNYL, A. The ISI Web of Knowledge, a modern system for the information support of scientific research: a review. *Scientific and Technical Information Processing*, n. 36, 2009, p. 351-358.

FISCHER, T.; BAKALOV, F.; NAUERZ, A. An overview of current approaches to mashup generation. *Proceedings of the 5th Conference on Professional Knowledge Management*, Solothurn, Switzerland, Março 2009

FULLER, S. *Khun vs. Popper. The struggle for the soul of science*. Icon Books (UK), 2003.

GARFIELD, E. *Association-of-ideas techniques in documentation: Shepardizing the literature of science*. 1954. Disponível em: <http://www.garfield.library.upenn.edu>.

GARFIELD, E. Citation indexes for science: a new dimension in documentation through association of ideas. *Science*, n. 122, 1955, p. 108-111.

GARFIELD, E. Introducing citation classics: the human side of scientific reports. *Current Contents*, n. 3, Jan. 1977, p. 1.

GARFIELD, E. *Citation indexing – its theory and application in science, technology and humanities*. J. Wiley & Sons, 1979.

GARFIELD, E. Essays of an information scientist. (1977-78). *ISI Institute for Scientific Information*, v. 3, 1979, p. 1-2.

GARFIELD, E. From citation indexes to informetrics: is the tail now wagging the dog? *Libri*, n. 48, 1998, p. 67–80.

GARFIELD, E. A prospective view of citation indexing and information retrieval in the 21st century. *Contributions to Information Science & Technology Award*. ASIS&T LA, Chapter, 27, Jan. 2005.

GARFIELD, E. Tracing the influence of JD Bernal on the world of science through citation analysis. *Bernal Symposium on Protein Crystallisation*. Dublin, n. 3-4, Set. 2007a.

GARFIELD, E. On the shoulders of giants – tracing the impacts of information retrieval systems on science policy. *60th anniversary celebration, Department of Energy, Office of Scientific and Technical Information*. Oak Ridge (USA), 18 Set. 2007b.

GARFIELD, E. From Information Retrieval to Scientometrics - is the dog still wagging its tail? *Keynote Address at WIS & COLLNET*, Dalian, China, 13 Setembro 2009.

GIUSTINI, D. How Google is changing medicine. *BMJ*, n. 331, 2005, p. 1487.

GILAYREVSKIL, R.; CHERNYL, A. Doctor Eugene Garfield: scientific-information activities. *Scientific and Technical Information Processing*, n. 36, 2009, p. 156-159.

GOOGLE, *Annual report*, 2009

HARGADON, A. *How breakthroughs happen: the surprising truth about how companies innovate*. Harvard Business Press. 2003.

HARGADON, A. Bridging old worlds and building new ones: towards a microsociology of creativity. In: THOMPSON, L. e H. CHOI (eds.). *Creativity and innovation in organizational teams*. Lawrence Erlbaum Associates, Inc., 2006, p. 199-215.

HARZING, A. *Google Scholar – a new data source for citation analysis*. 2008. Disponível em: [http://www.harzing.com/pop\\_gs.htm](http://www.harzing.com/pop_gs.htm).

HARZING, A.; VAN DER WAL, R. *Google Scholar as a new source for citation analysis*. *Ethics in Science and Environmental Politics*, n. 8, 2008, p. 61-73.

HAYES, R. History review: the development of information science in the United States. In: BOWDEN, M.; HALIN, T.; WILLIAM, R. *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*. Information Today Inc., 1999, p. 223-236.

HUGHES, T. An interview with Anurag Acharya, *Google Scholar* leading engineer. *Google Librarian Central Newsletter*, Dez. 2006. Disponível em: [http://www.google.com/librariancenter/articles/0612\\_01.html](http://www.google.com/librariancenter/articles/0612_01.html) .

IDCH. *International Directory of Company Histories*. St James Press, 2005.

KHUN, T. *The structure of scientific revolutions*. The University of Chicago Press, 1962.

LEDERBERG, J. How the Science Citation Index got started. In: *The Web of Knowledge: a Festschrift in honour of Eugene Garfield*. CRONIN, B.; ATKINS, H. ASIS American Society, Information Today Inc., 2000, p. 25-63.

LATOUR, B. *Science in action: how to follow scientists and engineers through society*. Harvard University Press, 1987.

LATOUR, B.; WOOLGAR, S. *Laboratory life: the construction of scientific facts*. Sage Publications, 1979.

LEWENSTEIN, B. Fax to facts: cold fusion and the history of science information. In: BOWDEN, M. ;HALIN, T.; WILLIAM, R. *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*. Information Today Inc., 1999, p. 14-26.

MIAH, A.; RICH, E. *The medicalization of cyberspace*. Routledge, 2008.

MIKKI, S. *Google Scholar* compared to Web of Science: a literature review. *Nordic Journal of Information Literacy in Higher Education*, n. 1, 2009, p. 41-51.

NOOY, W.; MRVAR, A.; BATAGEL, V. *Exploratory social network analysis with Pajek*. Cambridge University Press, 2005. Disponível em: <http://pajek.imfm.si/doku.php>.

NORUZZI, A. *Google Scholar: the new generation of citation indexes*, *Libri*, 2005, p. 170-180

OCDE. *OCDE science, technology and industry outlook, 2000*. OCDE, 2000.

POLANYI, M. *Personal knowledge: towards a post-critical philosophy*. University of Chicago Press, 1958.

POLANYI, M. The republic of science: its political and economic theory. *Minerva*, n. 1, 1962, p. 54-74.

POLANYI, M. *The tacit dimension*. Doubleday & Co., New York, 1966.

PORTER, M. How Competitive Forces Shape Strategy. *Harvard Business Review*, Mar./Apr.,

1979. Reimpresso com o título: The five forces that shape strategy. *Harvard Business Review*, Jan. 2008.

PORTER, M. *Competitive strategy: techniques for analyzing industries and competitors*. Simon and Schuster, 1998.

PORTER, M. *On competition*. Harvard Business Press, 1998.

PORTER, M. What is strategy? *Harvard Business Review*, Dez., 1996, p. 61-78.

PORTER, M. Strategy and the internet. *Harvard Business Review*, Mar., 2001, p. 1-18.

RAIGHATTA, C. Now, Google goes scholarly! *The Times of India*, 20 Nov. 2004, Disponível em: <http://timesofindia.indiatimes.com/world/us/Now-Google-goes-scholarly/articleshow/929365.cms>.

RETHLEFSEN, M. Citation indexes: more than just Web of Science. *e-knowledge.net: Quarterly Newsletter of the Mayo Clinic Libraries*, v. Winter, n.21, Mar. 2006.

ROSS, A. (ed.) *Science wars*. Duke University Press, 1996.

SADEH, T. *Google Scholar* versus metasearch systems. *HEP Libraries Webzine*, 11/01/2006. Disponível em: <http://library.web.cern.ch/library/Webzine/12/papers/1/>.

SCOTT, W.; MOLESKY, M. *Michael Polanyi: scientist and philosopher*. Oxford University Press, 2005.

STURCHIO, J.; THACKRAY, A. Eugene Garfield. Transcript of an interview. *Oral History Program*, Chemical Heritage Foundation, 16 Novembro 1987. Disponível em: <http://www.garfield.library.upenn.edu/oralhistory/interview.html>

SWANN, B.; APRAHAMIAN, F. (eds.) *J. D. Bernal: a life in science and politics*. Verso, 1999.

TENNANT, R. Digital Libraries: is metasearching dead? *Library Journal*, 07/15/2005. Disponível em: <http://www.libraryjournal.com/article/CA622685.html>.

THACKRAY, A.; BROCK, D. Eugene Garfield: history, scientific information and chemical endeavor. *The Web of Knowledge: a Festschrift in honour of Eugene Garfield*. CRONIN, B.; ATKINS, H. ASIS American Society, Information Today Inc., cap. 3, 2000, p. 11-24.

THOMSON REUTERS. *2010 Fact Book*. Disponível em: [http://thomsonreuters.com/investor\\_relations](http://thomsonreuters.com/investor_relations)

TROLLEY, J.; O'NEILL, J. The evolution of citation indexing – from computer printout to the Web of Science. In: BOWDEN, M.; HALIN, T.; WILLIAM, R. *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*. Information Today Inc., 1999, p. 124-126.

WILKINS, J. Metasearch vs. *Google Scholar*, 2006, disponível em <http://scholarlypublishing.org/jpwilkin/archives/6>

WILLIAMS, R. Eugene Garfield. Transcript of an interview. *Oral History Program*, Chemical Heritage Foundation, 29 Julho 1997. Disponível em: <http://www.chemheritage.org/discover/collections/oral-histories/details/garfield-eugene->

1997.aspx.

WOUTERS, P. The signs of science. *Scientometrics*, n. 41, 1998, p. 225-241.

WOUTERS, P. *The citation culture. Doctoral thesis.* University of Amsterdam, 9 Mar. 1999a.

WOUTERS, P. The creation of the Science Citation Index. In: BOWDEN, M.; HALIN, T.; WILLIAM, R. *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems.* Information Today Inc., 1999b, p. 127-136.

WOUTERS P. Garfield as alchemist. *The Web of Knowledge: a Festschrift in honour of Eugene Garfield.* CRONIN, B.; ATKINS, H. ASIS American Society, Information Today Inc., cap. 3, 2000, p. 65-72.

WOUTERS, P. *The citation culture.* Stanford University Press, 2003.

WOUTERS, P.; DUVAL, J. Aux origines de la scientometrie. La naissance du Science Citation Index. *Actes de la recherche en sciences sociales*, n. 164, 2006, p. 11-22.

VINE, R. *Google Scholar.* *J. Med. Libr. Assoc.*, v. 94, n. 1, Jan. 2006, p. 97.

## **ABSTRACT**

During last forty years scholarly information services developed from near non existence to a significant business with deep implications in the management of science and academic institutions. We discuss the innovative business struggle of Eugene Garfield and ISI Institute for Scientific Information until mid 90's, and the changes after the *Google Scholar* entered the market with an highly disruptive business model. The processes are discussed in the context of social frameworks for business innovation. Changes in the market during last decade are analysed and the implications for future are explored. Empirical data is presented comparing search results from ISI/WoS and *Google Scholar* that show the legacy effect of ISI / WoS data architecture. Results also show limitations from both services. The real number of citations is underestimated by both sources, and unique citations from each source are the majority for the cases discussed. The actual "citations wars" are discussed in the context of the previous "science wars", and the permanent search for the meaning and justification of science and academic activities. We argue that the open and dynamic model of *Google Scholar* is much more coherent with the reality, allowing a much more complete view of all the burgeoning and varied processes involved in the continuous struggle of science and academic work – something with important policy implications.

**KEYWORDS:** Business innovation. Disruptive innovation. Information services markets. Science models. Citation analysis.

*Originalis recebidos em: 30/06/2010*

*Aceito para publicação em: 08/10/2010*