

Repensando os *Commons* na Comunicação Científica

Evelyn Cristina Pinto

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO DE MESTRE
EM
CIÊNCIAS

Área de Concentração: Ciência da Computação

Orientador: Prof. Titular Imre Simon

São Paulo, julho de 2006.

Sumário

Resumo.....	1
<i>Abstract</i>	2
Capítulo 1 – Introdução	3
1.1. Contexto	3
1.2. Dificuldades	7
1.3. Objetivos	8
1.4. Metodologia	9
1.4.1. Scripts	10
1.5. Contribuições	11
1.6. Organização	12
Capítulo 2 - <i>Commons</i>	15
2.1. Conceito	15
2.2. <i>Commons</i> de informação.....	17
2.3. <i>Commons</i> digitais	19
2.4. <i>Commons-based peer production</i>	20
Capítulo 3 - <i>Commons</i> Científico	23
3.1. Breve história do <i>commons</i> científico.....	23
3.2. O <i>commons</i> científico hoje	26
3.3. Recursos do <i>commons</i> científico	28
3.4. A comunidade do <i>commons</i> científico	29
3.5. Uso e produção de recursos no <i>commons</i> científico	31
3.6. Regras do <i>commons</i> científico	34
3.7. Crise no acesso ao <i>commons</i> científico	35
Capítulo 4 - Acesso Aberto	37
4.1. Vantagens	37
4.2. O que é?	38
4.3. Breve histórico	40
4.4. Rota 1: <i>Self-archiving</i> em repositórios institucionais e temáticos	40
4.5. Rota 2: Revistas de acesso aberto	43
4.6. O <i>commons</i> do Acesso Aberto	45
4.7. A questão dos incentivos na produção de recursos do <i>commons</i> A. A.	46
4.7.1. O fator de impacto no A. A.	47
4.7.2. A questão da falta de incentivos	48
Capítulo 5 - Transição	51
5.1. A estrutura das revoluções científicas	51
5.2. Rivalidade, controle e liberdade na Rede Mundial	53
5.3. Transição para o acesso aberto	55
5.4. Quatro aspectos da transição	58

5.5. Aspectos legais	60
5.6. Aspectos econômicos	62
5.7. Aspectos sociais	63
Capítulo 6 – Aspectos tecnológicos: Open Archives Initiative	65
6.1. Breve histórico	65
6.2. Algumas definições	66
6.3. O protocolo OAI-PMH	69
6.4. O protocolo na prática	71
6.4.1. Seleção	72
6.4.2. Ferramentas para <i>harvesting</i>	72
6.4.3. Coleta de metadados	73
6.4.4. Heterogeneidade nos metadados e necessidade de normalização	73
6.5. Melhorias no protocolo	75
Capítulo 7 - Evolução do Acesso Aberto	77
7.1. Repositórios	77
7.2. Revistas de acesso aberto	81
7.3. Crescimento social	81
Capítulo 8 - Bibliotecas Temáticas	83
8.1. O porquê das bibliotecas temáticas a partir dos repositórios institucionais	83
8.2. Características desejáveis nas bibliotecas temáticas	85
8.3. Como construir bibliotecas temáticas	87
8.4. O texto completo	87
8.4.1. Disponibilidade do texto completo em repositórios grandes ou importantes	88
8.4.2. Disponibilidade do texto completo nos institucionais	90
8.4.3. Licenças para a coleta do texto completo	91
8.5. Resultados	94
Conclusão	97
Apêndices	
A – Glossário	99
B - Principais Provedores de Dados	102
C - Principais <i>Softwares</i> para Provedores de Dados	106
D - Serviços Importantes e Paralelos ao Acesso Aberto	109
Referência	112

RESUMO

Recentemente estudiosos como Benkler, Lessig, Boyle, Hess e Ostrom retomaram o uso do conceito de commons, mas agora relacionado à informação em geral ou à informação científica. Nesse trabalho, nós lançamos mão desse termo para destacar o caráter cooperativo da pesquisa científica, a importância da transparência e neutralidade no acesso ao commons da Ciência e a natureza anti-rival da informação científica. O conceito de commons nos é muito útil para focar todo o conjunto dos artigos científicos já publicados, quer estejam na forma impressa ou na digital. Ainda permite um estudo através de prismas multidisciplinares e, finalmente, enfatiza a dinâmica das comunidades científicas como um todo.

Em qualquer commons de informação, quanto maior a distribuição do conhecimento, mais dinâmico e eficiente é o processo de evolução do conhecimento. A tecnologia da imprensa tem desempenhado um papel fundamental na divulgação de informação e o seu surgimento marcou uma revolução no conhecimento e na cultura da nossa civilização. A tecnologia digital tem se mostrado mais eficiente ainda, uma vez que a natureza da sua implementação em bits se aproxima mais da natureza anti-rival das idéias do que qualquer outra tecnologia hoje empregada para preservação e distribuição de informação. Em nosso estudo, constatamos que o commons da Ciência pode ser enormemente enriquecido através de práticas cooperativas e de acesso aberto na publicação da academia. Percebemos também que o uso da tecnologia digital no commons científico, especialmente na publicação dos resultados da pesquisa, aumenta grandemente a distribuição do conhecimento acadêmico, suas oportunidades de escrutínio e validação, a dinâmica de amadurecimento das idéias científicas e, conseqüentemente, pode tornar o desenvolvimento da Ciência mais veloz e eficiente.

No entanto, o meio digital tem sido utilizado tanto para criar um ambiente de livre circulação de idéias quanto para controlá-las. Por um lado, código computacional tem sido implementado para garantir o acesso apenas aos que pagam pelos altos preços das revistas científicas. Por outro lado, a publicação de revistas on-line de acesso aberto e outras formas alternativas de disseminação de conteúdo científico têm se proliferado. Ainda, o decrescente orçamento das bibliotecas, o crescente preço das assinaturas de revistas científicas e as crescentes restrições aplicadas pelas leis de propriedade intelectual têm minado a natureza livre das idéias científicas e colocado a Comunicação Científica numa crise. Estamos no meio de uma transição de paradigmas quanto à publicação dos resultados de pesquisa científica, onde aspectos legais, tecnológicos e sócio-econômicos estão em renegociação.

À luz das oportunidades da tecnologia digital e da publicação em acesso aberto, as formas de disseminação dos resultados da pesquisa científica presentemente estabelecidas tem sido repensadas. Inserimos essa análise num contexto maior, o paradigma da Comunicação Científica. Isso nos auxilia a fazer um estudo mais abrangente das complexas questões envolvendo nosso tema, analisando os aspectos tecnológicos, legais e sócio-econômicos de uma possível transição para o modelo de publicação de acesso aberto. Tão grandes são as oportunidades desse novo modelo que ele tem agregado em torno de si iniciativas sócio-acadêmicas conhecidas por Movimento de Acesso Aberto à literatura científica.

Atualmente, há muitos testes e modelos de publicação dessa literatura. Em especial, nesse trabalho focamos o modelo de acesso aberto aos resultados científicos, suas vantagens, as dificuldades para seu estabelecimento e como ele tem se desenvolvido. Analisamos a viabilidade de criação de um ecossistema de bibliotecas digitais de acesso aberto, especializadas em cada ramo da Ciência. Nossos modelos de partida baseiam-se em alguns aspectos de serviços como arXiv, CiteSeer e Google Scholar. Entre as muitas conclusões desse estudo, constatamos que bibliotecas desse tipo aumentam sobremaneira a dinâmica de circulação, geração, transformação e renovação do conhecimento científico. Assim, o processo de produção de recursos no commons científico pode se tornar muito mais eficiente.

ABSTRACT

Recent studies done by Benkler, Lessig, Boyle, Hess and Ostrom look at the concept of commons again however, this time in relation to information in general more specifically to scientific information. In this study, we focused on the cooperative character of scientific research, the importance of transparency and neutrality to access the scientific commons. The concept of commons is highly useful to focus on every scientific article that has already been published in print or digitally. This allows studies through several multidisciplines and finally emphasizes the dynamic of scientific communities around the world.

In each commons of information, the higher the distribution of knowledge, the more dynamic and efficient the process of the evolution of this information. Technology of the press has been key in the divulging of information and its expansion marked a revolution in knowledge and culture in our civilization. Digital technology has shown more efficiency. Its implementation into bits is closer to the non-rival nature of the ideas than other technologies used to preserve it and used to distribute information. In our work, we realized that the science of commons should be enriched through cooperative practices and open access to scientific results. We also realized that digital technology in scientific commons improves distribution of scholarly knowledge and the dynamic evolution of scientific ideas so the science development should be even more efficient and faster.

The digital revolution has been used to create a free environment of circulation of ideas and it has also been used to control certain things. On one side, computational code has been implemented to allow access just for people who pay for the service. On the other hand, online journals publishing and other alternative forms of disseminating scientific knowledge have been proliferated. The decreasing budget of libraries, the increasing cost of journal subscriptions and the increasing restrictions applied by intellectual property has enclosed the free nature of scientific ideas and it has put Scholarly Communication into a crisis. We are in the middle of a transitional phase, where legal, technological, social and economic aspects of scientific publishing have being renegotiated.

We inserted our analyses in a larger context, the Scholarly Communication paradigm. This supports a larger study about the complex questions of our subject, analyzing the technological, legal, social and economic aspects of a possible transition to the open access publishing model. This new publishing model is so interesting that some initiatives have started social movements pertaining to it.

Nowadays, there are many tests and publishing models especially in this line of work. We focused on the open access model in scientific results, its advantages, the difficulties of its establishment and how it has been developed. Finally, we propose that the creation of an open access digital libraries ecosystem specialized in every scientific field. Our staring models are services such as: arXiv, CiteSeer and Google Scholar. Among our conclusions, we have realized that following this models stated above, digital libraries can enhance the dynamic of circulation, generation, transformation and renovation of the scientific knowledge.

CAPÍTULO 1 – INTRODUÇÃO

1.1. Contexto

We need to substitute for the book a device that will make it easy to transmit information without transporting material, and that will not only present information to people but also process it for them, following procedures they specify, apply, monitor, and, if necessary, revise and reapply. To provide those services, a meld of library and computer is evidently required.

(Licklider, 1965, p.6).

Economic criteria tend to be dominant in our society. The economic value of information and knowledge is increasing. By the year 2000, information and knowledge may be as important as mobility. We are assuming that the average man of that year may make a capital investment in an "intermedium" or "consol" -- his intellectual Ford or Cadillac -- comparable to the investment he makes now in an automobile, or that he will rent one from a public utility that handles information processing as consolidated Edison handles electric power. In business, government, and education, the concept of "desk" may have changed from passive to active:

A desk may be primarily a display-and-control station in a telecommunication- telecomputation system -- and its most vital part may be the cable ("umbilical cord") that connects it, via wall socket, into the procognitive utility net. Thus our economic assumption is that interaction with information and knowledge will constitute 10 or 20 per cent of the total effort of the society, and the rational economic (or social economic) criterion is that the society be more productive or more effective with procognitive systems than without.

If a man wishes to get away from it all and think in peace and quiet, he will have merely to turn off the power. However, it may not be economically feasible for his employer to pay him at full rate for the time he does spend in an unamplified cerebration.

(Licklider, 1965, p.33)

As duas citações acima, ambas proféticas, foram retiradas do livro seminal "Libraries of the Future", um estudo coordenado por J. C. R. Licklider em 1965, mais de 40 anos atrás. Nesses 40 anos, as previsões de Licklider, primeiro diretor da ARPA e avô da Internet, foram basicamente realizadas. Mas, principalmente quanto à primeira dessas previsões, a sua realização é apenas parcial. Ela está tecnologicamente madura para ocorrer, mas de alguma forma não ocorre ou ocorre muito lentamente. Esse fato e o desejo de entendê-lo melhor motivaram esta dissertação.

Na verdade, o nosso estudo partiu de um cenário mais simples do que o cenário apontado por Licklider. Focamos a nossa atenção apenas nos artigos científicos de pesquisa (não incluindo livros, por exemplo). Os nossos pontos de partida são as bibliotecas digitais poderosas como o CiteSeer ou o Google Scholar e os movimentos de Acesso Aberto à Literatura Científica por outro. Esses movimentos procuram tornar disponíveis pela Internet todos os artigos publicados de pesquisa científica, em acesso livre e aberto, para qualquer um que queira consultá-los, seja ele humano ou um até mesmo um robô. Uma vez disponível, o robô poderia incorporar o trabalho científico a uma ou mais bibliotecas digitais. De fato, essas bibliotecas e esses movimentos, sua descrição resumida, a análise do seu progresso, das suas dificuldades e algumas incursões de como eles poderiam ser incrementados é o objetivo básico desta dissertação.

Para realizar o nosso objetivo, embutimos o conjunto dos artigos científicos de pesquisa dentro de um plano bem maior, a Comunicação Científica. E para apoiar a nossa visão e o nosso raciocínio lançamos mão de um conceito pouco explorado nesse contexto, os *commons*. Estudiosos como Benkler, Boyle, Lessig, Hess e Ostrom começaram a modelar a informação científica como se fosse um *commons*. Por *commons*, de forma simplificada e geral, queremos dizer um conjunto de recursos mantidos em comum por certa comunidade de pessoas, onde o acesso é neutro e transparente.

Os *commons* existem em qualquer comunidade. Alguns exemplos de *commons* são o ar, as ruas públicas, os parques, as praias e os rios. Há também *commons* intelectuais, como por exemplo, a teoria do Big Bang e as músicas de Mozart. Podemos distinguir dois tipos de *commons*: os rivais e os não-rivais. Os *commons* não-rivais não podem ser exauridos. Lessig (2001) cita a linguagem como um exemplo de *commons* não-rival e explica: "meu uso dela não inibe o seu uso". As calçadas das ruas também são *commons*, porém rivais: meu uso compete com o seu.

Nesse trabalho, o conceito de *commons* é associado também às idéias e conhecimento científico. Aqui definimos os recursos do *commons* científico como sendo as idéias e o conhecimento científicos, além de artefatos da literatura acadêmica como revistas e artigos científicos, quer estejam na forma digital ou impressa. Nosso estudo focaliza principalmente a publicação de resultados de pesquisa científica. Esses recursos gozam de acesso neutro e transparente e estão disponíveis para uma comunidade ampla e vagamente delimitada. Esses recursos, de propriedade comum, são essenciais para a manutenção vigorosa dos processos produtivos de informação, tão característicos dos processos incrementais da pesquisa científica.

No *commons* científico o cientista deposita seus resultados de pesquisa científica e, futuramente, outro cientista (ou ele próprio) pode retirar esse resultado para utilizar como insumo em outra pesquisa. Esse processo é cíclico e quanto melhor for o acesso a esse *commons*, melhor será a entrada para um novo processo de pesquisa. Mudanças tecnológicas estão reformulando o acesso ao

commons científico e a forma dos cientistas se comunicarem. Essas transformações estão alterando como a informação é gerada, armazenada, distribuída e preservada. O progresso acelerado das tecnologias da informação e comunicação, induzido especialmente pelo advento e desenvolvimento dramático e fortemente transformador da Internet estão na raiz de uma situação de crise na Comunicação Científica.

Muitas revistas já migraram e outras estão migrando para o meio digital, oferecendo diferentes formas de acesso ao leitor. A Internet tem atingido profundamente o modo como os cientistas se comunicam. Apontamos dois principais motivos: a Rede permite maior liberdade na publicação científica e é um meio mais eficiente de implementação de *commons* intelectuais.

Por apresentar características não-rivais, a Rede oferece mecanismos mais baratos e freqüentemente mais eficazes de publicação e distribuição de informação por todo o globo. A Internet permite ainda uma implementação digital, ainda que aproximada, mas mesmo assim bastante fidedigna de um *commons* não-rival, um *commons* onde o seu uso dos recursos não diminui o meu uso desses mesmos recursos. Pelo contrário, quanto mais um recurso é utilizado tanto mais importante, central e significativo ele será para a comunidade. A nosso ver, essa novidade é uma das causas mais profundas da crise focalizada.

A crise é igualmente alimentada por um certo esgotamento dos modelos anteriormente estabelecidos e amplamente praticados. Exemplos: a maioria dos cientistas vai muito menos à biblioteca para ler revistas impressas; o sistema antigo de fichários de cartões bibliográficos quase já não é mais usado; praticamente ninguém mais se comunica através de cartas; o que se faz hoje com motores de busca como o CiteSeer, anteriormente era feito pela mente, principalmente daqueles cientistas famosos por terem um conhecimento completo da bibliografia em certas áreas.

Além dessas transformações, existe uma quantidade vasta de dilemas perplexos na Comunicação Científica hoje. Os preços das assinaturas de revistas científicas estão aumentando. Formas alternativas de publicação livre na Internet estão se proliferando. O orçamento das bibliotecas está diminuindo. Barreiras no acesso à informação científica estão sendo reforçadas através de código computacional de autenticação de usuários nos portais das revistas científicas.

As tendências contraditórias apontam uma crise na Comunicação Científica. Essa crise pode nos levar a uma transição para um novo modelo na comunicação e publicação científicas. Crises assim, que atingem amplas comunidades e colocam a possibilidade ou até mesmo a necessidade de mudanças paradigmáticas referentes a práticas fortemente arraigadas naquelas comunidades podem levar a transições de fase em sistemas complexos, como é o caso da Comunicação Científica. Esse é um dos enfoques pretendidos nessa dissertação. Colocamos os *commons*, propriedade coletiva de todos os interessados, no centro do processo de transição e procuramos ver a transição, que pode

ocorrer ou não, como um processo em que se está "Repensando os *Commons* da Comunicação Científica", título da dissertação.

Antes do advento, desenvolvimento e disseminação da Internet, esse tipo de *commons*, amplamente usado para o compartilhamento e desenvolvimento de idéias, de informação e de intangíveis em geral só podia ser realizado com o uso das mentes humanas. Agora, temos a possibilidade de uma realização aproximada, de base tecnológica, mas de alcance muito mais abrangente do que as possibilidades de realização anterior. Tal fato é um acontecimento de peso. É plausível que leve a transições radicais. A profundidade da novidade e a forma brusca como ela se apresentou estão também contribuindo para as dificuldades enfrentadas pelo processo de renegociação e de repensamento da Comunicação Científica.

É interessante notar que a mesma tecnologia que permite a liberdade na publicação tem sido empregada para reforçar restrições de acesso. Apenas aqueles que estiverem autorizados têm acesso a portais das revistas científicas de acesso fechado. Isso quer dizer, só acessa esse conhecimento científico quem puder ou quiser pagar pelos preços elevados determinados pelas poucas editoras científicas dominantes no mercado. Certamente, esse não é um exemplo de implementação de um *commons* científico, pois o acesso não é neutro.

Nesse processo de transição, as duas formas de publicação (aberta e fechada) estão disputando o mercado e a atenção dos cientistas. Ainda não sabemos qual será o modelo vigente daqui alguns anos. Um possível modelo novo dominante na publicação científica é o Acesso Aberto. Ele é definido basicamente pelo acesso *on-line* e gratuito ao leitor. Segundo estimativas, hoje cerca de 15% dos 2.5 milhões de artigos científicos publicados anualmente já estão em acesso aberto. Há duas principais formas de se implementar o Acesso Aberto à literatura acadêmica: através de repositórios digitais ou das revistas científicas onde não se aplica taxa ao leitor.

Os repositórios digitais podem ser organizados por disciplina, como o arXiv para a Física, ou agregarem a publicação de uma dada instituição, como o eScholarship Repository da Universidade da Califórnia. Para auxiliar a troca de dados entre os repositórios, foi criado o Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Trata-se de uma interface baseada em HTTP (request-response), onde são definidos verbos para facilitar a coleta automática de dados de repositórios digitais distribuídos. Foram definidos dois tipos de serviço: os provedores de dados e os provedores de serviço. Os provedores de dados são os repositórios e bibliotecas digitais que expõem seus dados através do protocolo OAI-PMH. Os provedores de serviço são os serviços implementados sobre a camada de dados importados via protocolo.

Se o modelo de acesso aberto se tornar a forma predominante de publicação científica, emergirá um paradigma radicalmente diferente para a Comunicação Científica. Se toda a literatura científica

estiver aberta e utilizarmos ferramentas computacionais poderosas para classificar, indexar, relacionar e extrair associações entre artigos científicos, o acesso será enormemente aumentado. A pesquisa científica se tornará mais dinâmica e a informação estará muito mais exposta para escrutínio. Como consequência, ocorrerá um desenvolvimento da Ciência mais veloz e eficiente.

Para atingirmos tais avanços, é preciso vencer a crise e enfrentarmos uma árdua negociação e adaptação das regras hoje predominantes no *commons* científico. O modelo de Acesso Aberto hoje enfrenta muitos desafios como a força econômica de algumas editoras comerciais, que procuram impor barreiras ao desenvolvimento desse modelo. Ainda há variadas dificuldades para motivar os cientistas a lerem a literatura de acesso aberto e a mudarem os seus hábitos de publicação.

Neste trabalho, propomos a criação de serviços que, pela sua utilidade, motivarão mais pesquisadores a utilizarem os artigos científicos em acesso aberto e a publicarem mais dessa forma.

O tipo de serviço a que nos referimos são as bibliotecas temáticas, criadas a partir de coleta automática de metadados de bibliotecas e repositórios digitais distribuídos. Essas bibliotecas temáticas devem reunir informações sobre artigos científicos e/ou os textos completos dos mesmos, num certo tema, disciplina ou especialidade da Ciência.

Também devem incorporar ferramentas avançadas de indexação e relacionamento de dados bibliográficos e de texto completo semelhantes às hoje encontrados no CiteSeer. Suponhamos que essas bibliotecas possam ser facilmente realizadas, possivelmente com o uso de *softwares* livres, a partir de repositórios digitais que contenham quase a toda a da literatura científica. Então seria possível o desenvolvimento de um ecossistema de bibliotecas temáticas de acesso aberto. A diversidade seria realizada através da relativa facilidade de estabelecimento desses serviços. Essa diversidade, por sua vez, forneceria uma biblioteca temática específica para cada campo ou especialidade que o pesquisador pretendesse investigar. Dessa forma, o acesso ao *commons* científico seria enriquecido grandemente.

1.2. Dificuldades

Aproveitamos essa introdução para realçar ainda três aspectos do trabalho que levam a dificuldades inerentes, sentidas durante a realização da pesquisa. Por um lado, embora o Acesso Aberto seja um tema cada vez mais presente na comunidade científica, a nossa abordagem pode ser considerada inovadora. Existem muito poucos trabalhos para basear o nosso discurso principal. Um segundo aspecto é que a evolução do tema motivador, o Acesso Aberto, é frenética. As novidades surgem com alta frequência e a sua absorção pela comunidade é um processo frequentemente contestado,

lento e gradual. Por último, o tema é fortemente multidisciplinar e é amplamente atingido por uma variedade de aspectos tecnológicos, legais, econômicos e sociais. Mais ainda: esses aspectos são fortemente interdependentes entre si, dificultando ainda mais uma análise global.

Oferecemos dois exemplos muito recentes dessas dificuldades. O primeiro livro dedicado ao Acesso Aberto, "The Access Principle, the case for open access to research and scholarship", de John Willinsky, publicado pela MIT Press em janeiro de 2006, na fase final de redação desta dissertação. Similarmente atropelante, nos últimos dias da redação foi anunciado um simpósio cujo pano de fundo é altamente aderente ao nosso enfoque. Trata-se do simpósio "Cultural Environmentalism at 10"¹ a ser realizado em Stanford para comemorar os 10 anos do livro pioneiro "Shamans, Software and Spleens: Law and the Construction of the Information Society" de James Boyle. Esse autor lançou as sementes do movimento que traz os *commons* intelectuais para o centro da discussão. Esperamos que este trabalho seja uma contribuição, ainda que modesta, para a formação e solidificação da montanha potencial, cuja pedra fundamental foi lançada por Boyle nesse livro.

1.3. Objetivos

Os objetivos principais desse trabalho foram:

1. Analisar o processo de comunicação científica sob a ótica dos *commons*, enfatizando a questão do acesso aberto.
2. Aplicar o estudo de Kuhn sobre paradigmas e transição às transformações recentes da comunicação científica.
3. Entender a falta de sinergia na transição para o modelo de Acesso Aberto; há uma série de dificuldades, principalmente de cunho social, impondo resistência a esse novo modelo de publicação.
4. Enumerar as dificuldades na transição.
5. Sugerir formas de acelerar essa transição.

1- <http://cyberlaw.stanford.edu/conferences/cultural/>

1.4. Metodologia

A metodologia empregada baseou-se nas seguintes técnicas:

- A. Estudar a Comunicação Científica, os *commons*, o Acesso Aberto e o protocolo OAI.
- B. Analisar os problemas do modelo e do protocolo.
- C. Levantar aspectos relevantes da transição.
- D. Elaborar ferramentas para verificar os problemas
- E. Coletar dados estatísticos sobre o movimento de Acesso Aberto.

Como parte do estudo sobre Comunicação Científica, participei da disciplina Comunicação Científica, oferecida pelo Departamento de Biblioteconomia da Escola de Comunicação e Artes da USP. Esse curso se fez muito útil para sedimentar alguns conhecimentos básicos do complexo processo de comunicação científica. O estudo do livro *The Structure of Scientific Revolutions* (1962) de Thomas Kuhn também auxiliou na obtenção de uma visão geral de como se desenrolaram algumas das principais transformações científicas e tecnológicas. Ajudou-me a enxergar o quão profundas e complexas são as atuais mudanças em andamento na Comunicação Científica.

Para poder aplicar adequadamente o conceito de *commons* a recursos como informação científica, foi necessária uma pesquisa sobre esse tema, buscando desde as idéias mais primitivas dele até as mais modernas. Pesquisei desde o conceito medieval, contido na Enciclopédia Britânica até os artigos mais recentes que abordam a idéia de *commons* de informação, como os trabalhos de Benkler, Lessig, Hess e Ostrom.

Em relação aos temas Acesso Aberto e OAI, fiz uma revisão bibliográfica. Foi interessante notar que não havia livros sobre o assunto, apenas artigos científicos, em sua maioria, bastante recentes. Isso me mostrou claramente o quão novo é o assunto. Com o conhecimento obtido através das minhas leituras, elaborei um tutorial em português sobre Acesso Aberto e OAI e o publiquei como um portal na Incubadora da FAPESP, o Clube OAI¹.

1- <http://clube-oai.incubadora.fapesp.br>

O acompanhamento de fóruns de discussão na Internet bastante dinâmicos como o *American Scientist Open Access* mostraram-me o caráter polêmico e profundo de diversas questões relacionadas à implementação do Acesso Aberto. Serviços de notícia como o *FOS News* mostraram-me o quão dinâmica é a temática do meu trabalho.

Para analisar os problemas do modelo de Acesso Aberto e do protocolo OAI: i) realizei *harvesting* de mais de 120 repositórios; ii) estudei como os metadados e protocolo têm sido implementados; iii) instalei e estudei a ferramenta *EPrints*, um *software* livre para implantação de repositórios institucionais (esse serviço está hospedado no portal Arca do IME-USP e é conhecido por Coruja).

Já para levantamento de aspectos importantes da transição, primeiramente segmentei a análise da transição em aspectos tecnológicos, econômicos, legais e sociais. Acompanhei as negociações, novidades e formas de implementação do Acesso Aberto através de fóruns de discussão como o *American Scientist Open Access* e o do *FOS News*. Apliquei conceitos do livro de Kuhn às transformações recentes na Comunicação Científica, focando especialmente o novo modelo de Acesso Aberto.

Finalmente, para coletar dados estatísticos sobre a evolução do Acesso Aberto, encontrei informações recentes e úteis no verbete Open Access da Wikipedia, nos serviços de registro de repositórios Registry of Open Access Repositories (ROAR) e Directory of Open Access Repositories (DOAR). Ainda, utilizando minha coleção de metadados obtidos via protocolo OAI-PMH, gerei minhas próprias estatísticas, através de programas que implementei na linguagem de *scripts* Python.

1.4.1. Os scripts

O *script* *oaister-pages.py* realiza duas funções: contabilizar a distribuição de registros nos repositórios listados pelo OAIster e mostrar quem são os maiores repositórios contidos neste portal. O programa acessa uma a uma as páginas de listagem de repositórios do OAIster, que contém informações sobre a quantidade de registros armazenados em cada repositório. Essas páginas estão localizadas em URLs compostas por uma URL base¹ mais uma letra do alfabeto² seguida da extensão ".html".

1- <http://oaister.umdl.umich.edu/o/oaister/browse>

2- Por exemplo, todos os repositórios cujos metadados estão agregados no OAIster e cuja letra inicial do seu nome é A, eram listados na página HTML de URL <http://oaister.umdl.umich.edu/o/oaister/browsea.html>. O mesmo se deu para todas as letras do alfabeto.

Depois de abrir uma dessas páginas, foi extraída da página a quantidade de registros de cada um dos repositórios listados e as respectivas URLs dos repositórios. Esses dados foram armazenados em listas. Em seguida, a lista de tamanhos de repositórios da página foi varrida para classificar cada repositório, de acordo com as seguintes categorias, conforme sua quantidade de registros: repositórios de mais de 100 mil registros; entre 100 mil e 50 mil; entre 50 mil e 10 mil; entre 10 mil e 5 mil; entre 5 mil e mil; entre mil e 100 registros. Ao final, foi gerado um arquivo com as URLs de todos os repositórios com mais de 100 mil registros e foi listada a quantidade de repositórios em cada categoria relacionada.

O outro *script* implementado por mim foi o `full-texts-counter.py`. A partir de metadados coletados de um determinado repositório via protocolo OAI, o programa selecionava a URL fornecida no campo de metadado *identifier*. Esse endereço normalmente se referia a uma página HTML que descrevia um documento digital ou se referia diretamente ao objeto digital (no nosso caso, artigos científicos). Para as URLs obtidas através desse campo de metadado, o programa tentava descobrir se ele se refere a um documento aberto ou fechado. Ao final do processamento, foi exibido um resumo da quantidade de textos completos abertos do repositório. Para ver o código desses programas e obter mais informações, acesse a seção¹ de bibliotecas temáticas no Clube OAI.

1.5. Contribuições

As principais contribuições dos esforços desse trabalho foram:

a) Apontamentos de algumas dificuldades da transição para o acesso aberto, tais como:

- a falta de padrões na implementação do protocolo OAI e conseqüente dificuldade em se construir serviços úteis sobre os metadados coletados;
- a pouca quantidade de artigos científicos realmente em texto completo aberto dentro de repositórios supostamente de Acesso Aberto;
- a resistência de editoras predominantes no mercado;
- a falta de motivação da comunidade científica com o modelo de acesso aberto.

1- <http://clube-oai.incubadora.fapesp.br/tematicas/documentacao>.

b) Análise de algumas transformações na Comunicação Científica sob o ponto de vista:

- de um *commons*;
- dos paradigmas e transição de Kuhn, enfatizando o novo modelo de acesso aberto;
- da segmentação em aspectos tecnológicos, econômicos, legais e sociais.

c) Investigação:

- da implementação de licenças para coleta do texto completo dentro dos repositórios institucionais;
- do crescimento dos repositórios institucionais – tanto ao analisar gráficos e artigos já existentes quanto gerando nossos próprios gráficos e tabelas através de programas codificados por mim;
- da quantidade de textos completos realmente disponíveis em acesso aberto.

d) Apoio ao movimento de acesso aberto através:

- da disponibilização de um servidor EPrints para o Instituto de Matemática e Estatística da USP;
- da construção do portal Clube OAI.

Além disso, este trabalho oferece uma pauta para discussão de assuntos essenciais e polêmicos sobre a qualidade e a eficiência na Comunicação Científica. Ainda agrega em um só documento diversos assuntos relevantes a todo cientista interessado em aumentar a divulgação do seu trabalho e tornar mais eficiente o processo de comunicação científica.

1.6. Organização

Este trabalho está organizado em oito capítulos e mais quatro apêndices. No capítulo um, nós apresentamos a introdução, o contexto, as dificuldades, os objetivos, a metodologia e as contribuições deste trabalho.

O capítulo dois apresenta o conceito de *commons*, desde suas idéias mais primitivas até as mais modernas, associando esse termo à informação em geral e à informação em forma digital. Discutimos ainda o papel da rivalidade de recursos dentro de um *commons*. Nesse capítulo apresentamos também um conceito relacionado, o *commons-based peer production*, um novo modo de produção, descrito por Benkler, e que possui algumas características de *commons*.

No capítulo três, definimos um *commons* científico, basicamente um *commons* de informação científica. Apresentamos um pouco de contexto histórico, os recursos atuais de um *commons* científico, a comunidade agregada por esses recursos e as suas regras de uso. Ainda explicamos a crise atual dentro desse *commons*.

O capítulo quatro é um dos mais fundamentais da dissertação, onde descrevemos o modelo de acesso aberto à literatura científica, um modo de publicação onde a informação científica permanece livremente disponível na Internet, semelhantemente a um *commons*. Esse capítulo mostra as formas básicas de emprego desse modelo, suas características, suas vantagens e suas dificuldades.

A transição para o modelo de Acesso Aberto é explicada no capítulo cinco, através do conceito de transição de Thomas Kuhn, abordado na obra clássica *A Estrutura das Revoluções Científicas*. Nessa parte da dissertação também destacamos os aspectos legais, sociais e econômicos da transição. Os aspectos tecnológicos são abordados e explicados no capítulo seguinte, onde detalhamos o protocolo OAI-PMH e levantamos alguns dos seus problemas.

O capítulo sete explica, através de gráficos e dados estatísticos, a evolução do Acesso Aberto, abordando tanto o crescimento dos repositórios quanto o das revistas científicas baseadas nesse modelo. Ainda apresenta alguns indicadores de crescimento social do movimento.

No capítulo oito, propomos um serviço que pode aumentar o uso do Acesso Aberto e acelerar o processo de transição. Apresentamos idéias de como uma biblioteca digital poderia atrair mais leitores e autores para a forma de publicação de Acesso Aberto. Também mostramos as dificuldades para a construção do serviço que propomos e algumas possíveis soluções.

No final desse trabalho, apresentamos a conclusão e alguns apêndices contendo glossário, lista e descrição dos principais provedores de dados, dos principais *softwares* para implementação desses provedores e também alguns serviços importantes, paralelos aos serviços de Acesso Aberto.

CAPÍTULO 2 - *COMMONS*

2.1. Conceito



Fig 1.1

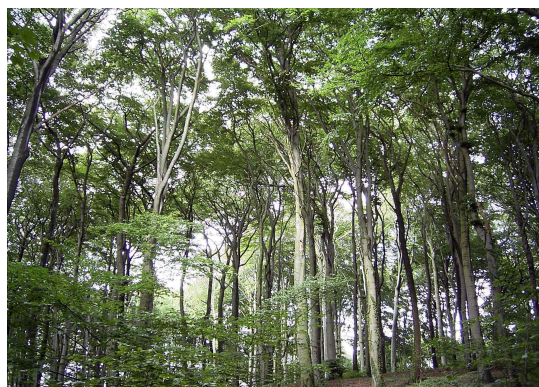


Fig. 1.2

Os *commons*¹ existem em qualquer comunidade e são regidos por algumas normas ou regras. Estas podem ser muito explícitas como leis ou contratos ou podem ser menos formais como, por exemplo, a ética ou o bom senso ao utilizar um recurso. No *Oxford English Dictionary*, os *commons* são recursos mantidos em comum, por uso ou posse conjunta. Na maior parte dos casos, trata-se de recursos aos quais qualquer membro da comunidade possui um direito de uso sem precisar de permissão de ninguém. Em alguns casos, a permissão é necessária, mas ela é garantida de forma neutra. Os parques e as ruas públicas são bons exemplos.

Nas figuras acima estão ilustrados outros dois exemplos do conceito de *commons*. A primeira representa a idéia inicial sobre *commons*, que se originou no sistema medieval de cultivo de terras na Inglaterra. Cada vilarejo era circundado por terra arável cultivada em faixas, por campinas, pastos e terras excedentes. Destas últimas, cada chefe de família tinha o direito de extrair madeira para combustível e para reparos. Quando os campos continham as plantações e as campinas descansavam (para reabastecer de feno), os cavalos, os bois, as ovelhas e as vacas ficavam nessas terras excedentes e contribuía com adubo. Cada chefe de família tinha o direito de trazer suas bestas para pastar nessas terras. Os *commons* não eram exatamente públicos, mas da comunidade local (Enciclopédia Britannica, verbete *commons*).

A segunda imagem nos sugere um conceito um pouco mais moderno sobre *commons*, aliado à ecologia e à preservação do meio-ambiente (Hardin, 1968). Nesse caso, são bons exemplos o ar, os rios, o mar, as florestas, etc. Recentemente, Benkler, Boyle, Lessig, Hess e Ostrom², entre outros,

1- O termo *commons* não é traduzido aqui nesse trabalho por não haver uma tradução apropriada para o mesmo.

2- Benkler (2004b); Boyle (2003); Lessig (2004); Hess, Ostrom(2003).

associaram recursos intelectuais ao conceito de *commons*. Destes estudos emerge o conceito de *commons* intelectuais, cujas propriedades discutiremos na seção 2.2.

Podemos distinguir dois tipos de *commons*: os rivais e os não-rivais. Os *commons* não-rivais não podem ser exauridos. Lessig (2001) cita a linguagem como um exemplo de *commons* não-rival e explica: "meu uso dela não inibe o seu uso". As equações matemáticas de Maxwell também são um recurso não-rival. O uso que Albert Einstein (ou qualquer outra pessoa) fez delas não diminui as minhas oportunidades de uso.

Já as calçadas das ruas também são *commons*, porém rivais: meu uso compete com o seu. Um livro da biblioteca também é um bem rival, pois somente um usuário pode lê-lo por vez. Se um recurso é não-rival então existe basicamente um problema para sua sobrevivência: se há incentivo suficiente para produzi-lo. Já o recurso rival apresenta este e ainda outro problema: se o consumo efetuado por alguns deixará o suficiente para os outros.

Hardin (1968) mostrou claramente como a exploração indiscriminada de recursos rivais pode levar à tragédia. O usuário não enxerga o custo total de suas ações. Para ele, a melhor estratégia de curto prazo é explorar mais do que a sua parte no recurso. Todavia, se um certo número de indivíduos segue essa mesma estratégia, o recurso é excessivamente explorado, tornando-se inútil.

Garret Hardin ilustra esse fato fazendo alusão ao conceito primitivo, onde os *commons* são uma porção de pasto compartilhada por animais da vila. Cada fazendeiro adiciona mais animais para pastar no *commons*. Em poucos anos, o solo torna-se enfraquecido pelo excesso de uso do pasto, os *commons* tornam-se inúteis e a vila perece. A tragédia dos *commons* reaparece nos problemas relacionados à preservação ecológica, como a questão do desmatamento das florestas e da poluição de *commons* como o ar e a água.

Entretanto, é interessante observar que as duas imagens no início do capítulo, utilizadas para explicar o conceito de *commons*, foram retiradas de um outro *commons*, não-rival e mais moderno, de mídia digital e oferecido pelo portal-enciclopédia Wikipedia¹, o Wikimedia Commons².

1- <http://en.wikipedia.org>

2- http://commons.wikimedia.org/wiki/Main_Page

2.2. Commons de Informação

Numa carta a Isaac McPherson em 1813 sobre o caráter do poder da patente, Jefferson ilustrou a natureza de uma idéia:

If nature has made one thing less susceptible than all others of exclusive property, it is the action of the thinking power called an idea, which an individual may exclusively possess as long as he keeps it to himself; but the moment it is divulged, it forces itself into the possession of everyone, and the receiver cannot dispossess himself of it. Its peculiar character, too, is that no one possesses the less, because every other possesses the whole of it. He who receives an idea from me, receives instruction himself without lessening mine; as he who lites his taper at mine, receives light without darkening me. That ideas should freely spread from one to another over the globe, for the moral and mutual instruction of man, and improvement of his condition, seems to have been peculiarly and benevolently designed by nature, when she made them, like fire, expansible over all space, without lessening their density at any point, and like the air in which we breathe, move, and have our physical existence, incapable of confinement, or exclusive appropriation. Inventions then cannot, in nature, be a subject of property.

Note como Jefferson apropriadamente ressaltou o caráter não-rival da informação. Essa propriedade tem um impacto profundo na produção de informação. Ajuda a facilitar a manutenção de um *commons* para informação, já que o recurso não é esgotável. Se fosse rival, haveria necessidade de um sistema de controle para garantir a produção e evitar o super uso. Mas como é não-rival, então um sistema de controle é necessário simplesmente para garantir que o recurso seja criado. As leis de propriedade intelectual nasceram para desempenhar esse tipo de controle, para garantir incentivo à criação artística e intelectual.

Por outro lado, o controle deve ser moderado. Ele não deve dificultar muito o acesso. Se o acesso tiver muitas barreiras, o incentivo à criação também pode diminuir. Além disso, a informação é tanto a entrada quanto a saída no processo de produção de informação. Então, quanto maior e melhor for o acesso à informação já existente, maior e melhor serão as oportunidades de escrutínio dela e de geração de nova informação. A respeito disso, Lessig (2001, p. 105) comenta:

The content an author must draw upon varies with the writing. Some part is new – this is the part we think as “creative”. But as many have argued, we’ve come to exaggerate the new

and forget that a great deal in the creative is actually old. The new builds on the old, and hence depends, to a degree, on access to the old. Academics writing textbooks about poetry need to be able to criticize and hence, to some degree, use the poetry they write about. Playwrights often base their plays upon novels by others. Novelists use familiar plots to tell their story. Historians use facts about the history they retell. Filmmakers retell stories from our culture. Musicians write within a genre that itself determines how much of the past content it needs to be within that genre (...) All of this creativity depends in part on access to, and use of, the already created.

Segundo Benkler (2004a) para se introduzir *commons* em ambientes de informação, é necessário que:

- a informação e o meio de comunicação estejam livres de controle exclusivo de uma instituição governamental ou não-governamental;
- empregue-se um desenho de sistema que possibilite que as entradas (informações) sejam utilizadas como num *commons*, ou seja, todos os membros da comunidade tenham as mesmas oportunidades de uso e que não seja necessária obtenção de permissão para usá-las.

Quando colocamos a informação dentro de um *commons*, temos pelo menos quatro vantagens: ela permanece livre de controle exclusivo; o acesso aumenta e, conseqüentemente, acelera o processo de produção de nova informação; aumenta o potencial de espalhamento e distribuição das idéias; finalmente, cria-se a possibilidade de formação de fortes comunidades em torno dela.

A regulamentação atual do *commons* intelectuais passa pelas legislações de propriedade intelectual, em particular, pelas legislações de direito autoral (Lessig, 1999; Simon, 2000; Boyle, 2003). É um tema de complexidade explosiva, cujos detalhes estão fora do escopo desta dissertação. No entanto, ocorreu uma memorável conferência em Duke sobre o Domínio Público¹ e alguns palestrantes relacionaram esse tema com os *commons*.

O meio de comunicação é um ponto vital no processo de produção de nova informação. Quanto mais eficiente for o processo de comunicação e distribuição da informação, mais facilitado está a produção de nova informação. Meios de armazenamento e preservação eficientes também são aspectos essenciais num *commons* de informação. Eles vão ajudar a garantir o acesso à informação já existente. Há diversas arquiteturas ou tecnologias para implementação dessas etapas de produção de informação.

1- <http://www.law.duke.edu/pd/papers.html>

A tecnologia da imprensa é uma delas. Nela, o processo de produção é fixo, estático, linear, físico, local e centralizado. O conteúdo é limitado, o processo de cópia é demorado e oneroso. Apenas um usuário (ou pequeno grupo) pode acessar o recurso ao mesmo tempo.

Mais recentemente, conquistamos a Internet. Nessa tecnologia, o processo de produção se torna transformativo, interativo, flexível, não-linear, extensível, virtual, global, distribuído e assíncrono. Colapsa espaço e tempo. Temos conteúdo ilimitado e facilidades de multimídia. O processo de cópia é simples e gera cópias idênticas. O custo marginal de distribuição é zero e a disseminação é imediata. Há muitos usuários e produtores concorrentes.

2.3. Commons Digitais

Os *commons* digitais são uma implementação aproximada de *commons* de informação no meio digital. Trata-se de recursos (serviços) compartilhados na Internet, livres para qualquer um utilizar, onde cada usuário tem os mesmos privilégios de uso. As regras e a comunidade variam um tanto, dependendo do tipo de serviço oferecido. Em especial, nesse trabalho nos interessa mais focar os serviços *on-line* e livres que são direcionados para conhecimento público, Ciência aberta, domínio público e a livre troca de idéias. A Wikipedia é um excelente exemplo de *commons* digital para construção colaborativa de conhecimento, como um *commons* intelectual modelado.

“A enciclopédia livre que qualquer um pode editar”, a Wikipedia, apresenta páginas de hipertexto sobre os mais variados assuntos. Conta hoje com mais de 1 milhão de artigos na versão em inglês e 2.6 milhões somando todas as versões (em mais de 200 línguas). Sua comunidade é formada tanto por leitores quanto por contribuidores. É produzida por cerca de 83 mil voluntários, ou seja, usuários que editaram ao menos 10 vezes (Wikipedia Statistics, 2006). É, no momento, o vigésimo segundo domínio mais visitado da Internet, segundo o Alexa¹. Este fato aponta para uma mudança no conceito social e uso de uma enciclopédia, demonstrando o potencial de um *commons* intelectual na forma digital para formação de uma comunidade numerosa.

As regras para contribuição e edição de texto estão formalizadas em Políticas e Guia de Referência, no portal do serviço. Abaixo, um outro exemplo:



O Wikimedia Commons é um projeto que oferece um repositório central para imagens, música, som e vídeo livres. Ele foi lançado em setembro de 2004. No início de março de 2006, continha 463.525 arquivos de mídia (Wikimedia Commons, 2006).

¹ - http://www.alexa.com/data/details/traffic_details?&range=6m&size=large&compare_sites=&y=t&url=http://en.wikipedia.org/wiki/Main_Page, acessada em 2 de março de 2006.

Esse é um dos vários projetos da Wikipedia. Como qualquer *commons*, esse também apresenta regras. O critério de inclusão de arquivos requer que sejam recursos de mídia livres, ou seja, recursos digitais cujas licenças permitam a publicação, a redistribuição e o uso comercial do objeto. O logotipo do serviço (logo acima), criado cooperativamente sugere-nos a idéia de um *commons*, onde as flechas representam os membros da comunidade em torno do serviço, que contribuem fazendo *upload* dos recursos. O Wikimedia Commons os reúne (círculo vermelho) e os distribui (flecha maior) de forma gratuita.

O *software* livre também agrega um *commons*, pois reúne um conjunto de recursos (programas) mantidos pela comunidade do Movimento de Software Livre. Suas regras são formalizadas em licenças como, por exemplo, a GPL (General Public License). Seu acesso é neutro (qualquer um pode participar) e transparente (as regras são claramente especificadas nas licenças) e tem a propriedade anti-rival: quanto mais usuários (testadores), melhor é o *software* e mais valor agregado ele tem (Weber, 2004).

2.4. Commons-based peer production

Nesses últimos anos surgiram na Internet novos mecanismos de produção, radicalmente descentralizados e alheios aos métodos de controle por mercado e hierarquias. Foram impulsionados pelo baixo custo da publicação, comunicação, distribuição e organização da informação na Rede, além da facilidade em se construir conteúdo ou programas cooperativamente nesse ambiente. Esses mecanismos baseiam-se num *commons* intelectual implementado através da Internet. Eles apresentam muitas outras características próprias. Benkler (2002) chamou esse novo modelo de produção por *commons-based peer production*.

1. Colaboração entre grupos de indivíduos (até em largas escalas) sem a necessidade da motivação do mercado financeiro ou hierarquia organizacional;
2. Em geral, utiliza ambiente social para coordenar e motivar;
3. Baixos custos;
4. Pode oferecer mais informações sobre disponibilidade de recursos humanos, ou seja, capital intelectual e criatividade – como os colaboradores normalmente podem escolher em que contribuir, freqüentemente as pessoas fazem as tarefas mais apropriadas para elas;
5. Funciona melhor para motivar contribuições de escala pequena (que tomem pequenas porções de tempo dos voluntários) do que projetos maiores (que demandem mais tempo de contribuições individuais);

6. Utiliza como ferramentas de trabalho material disponível em abundância, como computadores, ciclos ociosos de CPU, etc.

Um exemplo ímpar desse modo de produção, na verdade o revelador desse mecanismo novo de produção, é o fenômeno do *software* livre. Ele não se baseia em mercado nem em hierarquia para organizar produção. A massa crítica de participação nos projetos não pode ser explicada pela presença direta de um comando, preço ou mesmo retorno monetário futuro. Os programadores participam de projetos de *software* livre sem seguir as regras normais dos modelos de mercado (Benkler, 2002).

Outros exemplos marcantes desse modo de produção são:

a) Folding@Home¹ - é um projeto de computação distribuída que estuda a montagem de proteínas, deformações, agregação e doenças relacionadas. Utilizamos métodos de computação novos e computação distribuída de larga escala para simular escalas de tempo de milhares a milhões de vezes mais longas do que aquelas anteriormente conseguidas. Isto nos permitiu simular a formação pela primeira vez e agora direcionar o foco para o exame de doenças relacionadas à formação de proteínas."

"Desde do dia 1o. de outubro de 2000, mais de 1 milhão de CPUs participaram do projeto.". Hoje conta com mais de 160 mil CPUs ativas. Qualquer um pode adicionar a sua contribuição baixando o programa e o rodando no seu computador. Ele utiliza os ciclos ociosos do seu computador para executar o processamento das deformações de proteínas.

b) Clickworkers² (NASA)

De novembro de 2000 a setembro de 2001, nós realizamos um experimento que mostrou que usuários comuns, muitos trabalhando por poucos minutos um pouco aqui, um pouco ali... podem fazer algumas análises científicas rotineiras que normalmente seriam feitas por um cientista ou aluno de pós-graduação ao trabalhar meses a fio.

O trabalho seria extremamente demorado e custoso se fosse realizado por cientistas. Mas ele foi subdividido em pequenas partes que foram realizadas por usuários comuns da Internet. Os voluntários receberam uma interface que possibilitou marcar crateras nas imagens de Marte. Em seis meses, mais de 85.000 usuários visitaram o portal e muitos contribuíram com esforço. O consenso de vários usuários a respeito da presença de uma cratera numa dada posição, automaticamente computadorizado foi equivalente em qualidade ao que se obteria com cientistas treinados. A correteza dos resultados é impressionante.

1- Folding@Home, distributed computing, <http://folding.stanford.edu/>, acessado em 17/11/2005

2- NASA, Welcome to Clickworkers Study, em <http://clickworkers.arc.nasa.gov/top>, visitado em 17 nov 2005

Em seguida, o projeto disponibilizou novas imagens da superfície de Marte, adquiridas recentemente, para os usuários continuarem a marcação de crateras. Também ofereceu aos voluntários a chance de classificar as crateras já catalogadas. Essa tarefa é um pouco mais difícil que a anterior, mas foi disponibilizado um pequeno treinamento no portal do projeto.

c) arXiv¹ – trata-se de um repositório digital de artigos científicos e oferece uma alternativa radical para a publicação de resultados científicos. Os autores enviam seus artigos (cujos assuntos são principalmente de algumas áreas da Física) a um servidor sem nenhum *peer review*², de *preprints*³. As críticas pós-publicação, a facilidade de atualização de versão do artigo e uma comunidade acadêmica, onde os efeitos da reputação são substanciais, oferecem credibilidade e controle de qualidade. Em novembro de 2005, o repositório continha cerca de 350.000 artigos.

1- <http://arxiv.org>, acessado em 16 nov 2005.

2- para conhecer o significado desse termo, veja o glossário no apêndice A.

3- para conhecer o significado desse termo, veja o glossário no apêndice A.

CAPÍTULO 3 - *Commons Científico*

Na obra clássica *The Structure of Scientific Revolutions* (1962), Thomas Kuhn atribui à palavra paradigma mais de um significado. Um dos mais importantes corresponde à matriz disciplinar. Isso quer dizer uma cultura comum em torno de um determinado assunto num certo momento da história; engloba todas as crenças, os valores e as técnicas compartilhadas pelos membros da comunidade. É uma espécie de consenso da comunidade científica sobre uma determinada disciplina.

Aqui, pretendemos relacionar o conceito de paradigma ao *commons* da ciência, onde se reúnem todas as matrizes disciplinares, de todos os campos e especialidades da Ciência nas suas respectivas comunidades. Trata-se de todo o conhecimento científico público acumulado desde os estudos dos primeiros filósofos gregos, passando pelas experiências astronômicas de Tycho Brahe, Kepler e Galileu Galilei, a lei gravitacional de Newton e a relatividade de Einstein, até atingir a Ciência moderna. Trata-se de um processo evolutivo, baseado num *commons* parcialmente material (revistas e artigos impressos) e parcialmente intelectual (idéias e conhecimento científico). Esses *commons* podem ser vistos como segmentados por disciplinas ou como um *commons* global da Ciência.

3.1. Breve história do *commons* científico

Como qualquer *commons*, esse também é definido em termos dos recursos que agrega, da comunidade que os utiliza e das regras para uso. Os primeiros recursos desse *commons* foram manuscritos em rolos de papiro e as primeiras bibliotecas. (É bom ressaltar que, anteriormente aos papiros, escrevia-se em tábuas de pedras, no entanto, não vamos considerá-los como parte do *commons* científico).

Uma das primeiras e mais famosas bibliotecas da Antigüidade foi a biblioteca de Alexandria. Existiram outras poucas bibliotecas anteriores a essa, por exemplo, a de Ugarit (moderna Síria), datada de 1.200 a. C. Estima-se que a biblioteca de Alexandria armazenava entre 40.000 e 700.000 manuscritos. Assume-se que ela tenha sido fundada três séculos antes de Cristo, durante o reino de Ptolomeu II do Egito. Acredita-se que ela tenha sido destruída por fogo (Jochum, 1999).



Fig. 2.1. A reconstrução da sala principal da Biblioteca de Alexandria

(fonte: http://en.wikipedia.org/wiki/Alexandria_library)

Uma das formas de se adicionar manuscritos ao acervo é muito curiosa. Eram confiscados os livros encontrados nos navios do porto e eles eram duplicados através de cópia manual. Esta última era devolvida para o dono do livro original. A biblioteca tinha dez grandes salas cujas paredes amplas eram delineadas por armários espaçosos contendo o conhecimento e informação acumulados pelos helênicos. Nos armários, os rolos de papiro eram empilhados um sobre o outro. As pessoas que tinham acesso a esse acervo eram acadêmicos escolhidos e pagos pelos reis (Jochum, 1999).

Uma das primeiras e mais importantes bibliotecas públicas surgiu por volta de 666 d. C., no sudoeste do Irã, a biblioteca de Gundishapur. Ela foi parte de um complexo científico maior, a Academia de Gundishapur. Nela, o acesso aos livros era público. No mundo ocidental, as primeiras bibliotecas públicas foram estabelecidas sob o Império Romano (Wikipedia, 2006).

Nos primórdios do desenvolvimento do *commons* científico, a comunidade acadêmica era muito reduzida e se restringia aos nobres. Com o surgimento da imprensa, a publicação impressa dos manuscritos antigos e a proliferação das universidades medievais, a comunidade aumentou. Já as regras praticadas pelos cientistas, ou seja, o método científico estava ainda em desenvolvimento. Apesar de existirem documentos anteriores a 1.000 a. C. descrevendo métodos científicos, foi com o advento da Grécia Antiga que os primeiros elementos do método científico foram formulados. A cristalização do método, contudo, deu-se após a Revolução Científica, com a física moderna (entre os séculos XVII e XVIII) (Wikipedia, 2006a, 2006b).

A Revolução Científica é frequentemente considerada um marco divisório entre a fase da pré-ciência e da ciência moderna. Muitos historiadores consideram que o início da Revolução deu-se no momento da publicação de *De Revolutionibus*, de Copérnico. O ápice da Revolução Científica ocorreu com a publicação de *Philosophiae Naturalis Principia Mathematica* em 1687, por Isaac

Newton (Wikipedia, 2006a, 2006b). Peter Salus (2005) resume algumas das atividades científicas dessa época:

Copérnico (1473-1543) observou o céu e registrou suas medições. Em 1563, Tycho Brahe (1546-1601) notou que as figuras de Copérnico não estavam totalmente corretas, então, de 1577 a 1597, Tycho registrou de forma extraordinariamente precisa algumas medições astronômicas. Em 1599, Tycho mudou-se da Dinamarca para Praga, onde Johannes Kepler (1571-1630) foi seu assistente, até sucedê-lo em 1601, quando Tycho morreu.

Copérnico estabeleceu o heliocentrismo. Tycho descobriu que as órbitas circulares não funcionavam e devotou décadas para fazer medições melhores que Kepler mais tarde utilizou para determinar que as órbitas eram elípticas, e não circulares. Então Galileu (1564-1642) aperfeiçoou o telescópio, fez uma variedade de observações astronômicas e formulou a primeira lei do movimento. Newton (1643-1727) nos explicou a força gravitacional. Salus adiciona:

Itália. Dinamarca. Áustria. Alemanha. Inglaterra. Apesar do Papado, os 30 anos de guerra, tumultos nos Países Baixos, na França e na Inglaterra, pensamentos moveram-se via imprensa e correspondência. Embora os países estivessem em guerra e as religiões estivessem em conflito, a troca científica de idéias e o compartilhamento de dados persistiu.

No período da Renascença, levavam-se meses para as descobertas alcançarem aqueles interessados em outros países. No século XVII e XVIII, epístolas volumosas eram distribuídas entre acadêmicos. As revistas científicas então surgiram. Graças ao progresso nas mídias de comunicação, agora levamos segundos para enviar idéias que antes demoravam décadas para se proliferarem. A revolução do conhecimento nos direcionou para a exploração e a descoberta. (...) O computador e a Internet nos levaram a revoluções similares.

Com todos esses exemplos, observamos que a ciência se desenvolveu de forma distribuída, mesmo em épocas onde as dificuldades para se distribuir informação eram enormes. E praticamente sempre, para fazer uma nova pesquisa, os cientistas se baseavam no velho, ou seja, em teorias e experimentos pesquisados pelos seus predecessores. Lessig (2001) observa que “o novo é construído sobre o velho e, portanto, depende em certo grau de acesso ao velho”. Isaac Newton,

numa carta a Robert Hooke em 1676, ressalta indiretamente a importância do acesso a descobertas anteriores: "*If I have seen a little further it is by standing on the shoulders of Giants.*"

3.2. O *commons* científico hoje

Hess e Ostrom (2004) conceituaram o *commons* da comunicação científica. Para tanto, apresentaram a definição de comunicação científica feita pela Association of College and Research Libraries (ACRL, 2003):

Comunicação científica é o sistema através do qual a pesquisa e os manuscritos acadêmicos são criados, avaliados pela qualidade, disseminados para a comunidade científica e preservados para o futuro. O sistema inclui ambos os meios de comunicação, tais como publicação em revistas com peer review e canais informais, tais como listas de discussão eletrônicas.

Segundo Garvey (1977), a comunicação científica incorpora as atividades associadas à produção, disseminação e uso da informação, desde o momento em que o cientista cria uma idéia para pesquisar até que seus resultados são incorporados como constituintes do estoque universal de conhecimentos. A comunicação científica é indispensável à atividade científica, pois permite somar os esforços individuais dos membros das comunidades científicas. Eles trocam continuamente informações com seus pares, emitindo-as para seus sucessores e/ou adquirindo-as de seus predecessores.

De acordo com Menzel (apud Targino, 1998), ainda em 1958, as funções básicas da comunicação na Ciência são:

- a) fornecer respostas a perguntas específicas;
- b) concorrer para a atualização profissional do cientista no campo científico de sua atuação;
- c) estimular a descoberta e a compreensão de novos campos de interesse;
- d) divulgar as tendências de áreas emergentes, fornecendo aos cientistas idéia da relevância de seu trabalho;
- e) testar a confiabilidade de novos mecanismos, diante da possibilidade de testemunhos e verificações;

- f) redirecionar ou ampliar o rol de interesse dos cientistas;
- g) fornecer *feedback* para aperfeiçoamento da produção do pesquisador.

Como as definições acima mostram, um dos objetivos principais da comunicação científica é a disseminação dos resultados de pesquisa. No entanto, de acordo com Thorin (2003) o atual sistema de comunicação científica não completa as necessidades da comunidade acadêmica de forma satisfatória.

De acordo com Hess e Ostrom (2003 e 2004), hoje o *commons* da comunicação científica inclui:

- informação em múltiplos formatos tais como idéias, livros, arquivos de computador, etc;
- os facilitadores (edifícios de bibliotecas, *hardware* de computador, etc);
- os resultados, regras, normas, leis, comportamento e interações;
- questões de apropriação (*enclosure*), perda e degradação;
- comunidade de usuários competitivos e em expansão.

Fig. 2.2. A biblioteca da Universidade da Califórnia em San Diego¹.



O termo *commons* não é mais uma metáfora. Os bibliotecários não são mais capazes de gerenciar sozinhos a coleta e a preservação de todos os registros acadêmicos. As bibliotecas têm tido que fazer parceiras com serviços de computadores da universidade e novos arranjos colaborativos com os acadêmicos. Centenas de novas associações voluntárias locais, regionais e internacionais têm se desenvolvido com o intuito de criar novos padrões de protocolo para a Internet, metadados, etc. A organização, o gerenciamento, a apresentação, a distribuição e a preservação de sucesso dos recursos de informação eletronicamente distribuídos requer esforços de colaboração e ação coletiva sem precedentes (Hess e Ostrom, 2004).

1- Fonte: http://en.wikipedia.org/wiki/University_of_California%2C_San_Diego

3.3. Recursos do Commons Científico

A definição de *commons* em geral, apresentada no capítulo anterior, mostra que qualquer *commons* tem o seu próprio conjunto de recursos. De acordo com Hess e Ostrom (2003) o *commons* científico é composto por recursos como idéias, artefatos e facilitadores. Observe a figura a seguir:

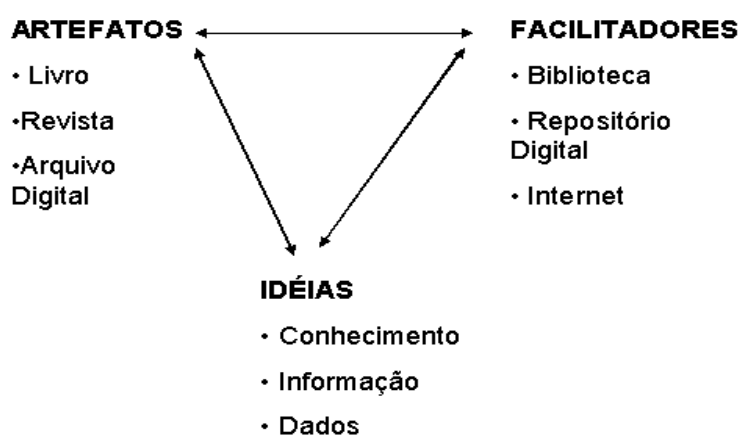


Fig 2.3 – Os recursos do *commons* científico moderno.

Os artefatos são os meios físicos onde se registram as idéias. Os primeiros artefatos foram pedras, argila, depois foi criado o papiro, muitos séculos depois, o papel. Atualmente, não dispensamos o papel, mas freqüentemente registramos nossas idéias em arquivos digitais. Nessa evolução, os artefatos se tornaram mais voláteis e mais propensos à deterioração.

Em se tratando de comunicação científica, os artefatos são artigos, notas de pesquisa, livros, revistas científicas, bases de dados, mapas, arquivos de computador e páginas *web*. Os artefatos variam de acordo com sua durabilidade. Um artefato físico é utilizado por uma pessoa de cada vez. Já os artefatos digitais podem ser utilizados concorrentemente por múltiplos usuários. Pode haver exclusão de usuários do grupo dos que podem utilizar os artefatos físicos, mas a exclusão de usuários em potencial torna-se muito mais complexa e menos transparente no caso dos artefatos digitais (Hess e Ostrom, 2003).

Já um facilitador armazena artefatos e os torna disponíveis. É um sistema de recursos que armazena artefatos e suas idéias. Anteriores ao desenvolvimento dos artefatos digitais, as bibliotecas têm sido facilitadores tradicionais. No entanto, é necessário um esforço para preservar seus recursos. A antiga Biblioteca de Alexandria foi destruída por fogo. As bibliotecas de hoje também estão sujeitas à destruição ou à deterioração se nenhuma manutenção é feita - tragédia dos *commons* (Hardin, 1968).

Porém, as idéias não estão sujeitas a esse tipo de tragédia. Elas podem ser compreendidas como visão criativa, conteúdo intangível, informação inovadora e conhecimento. Idéias são como um fluxo não-físico de unidades, contidas num artefato (Hess e Ostrom, 2003).

3.4. A Comunidade do *Commons* Científico

A comunidade científica compreende os indivíduos que se dedicam à pesquisa científica e tecnológica como um todo. A comunidade global é segmentada por vários critérios como especialidades, línguas, nações e até ideologias políticas, formando sub-comunidades em torno de *commons* de literaturas específicas (confronte com os dois primeiros parágrafos deste capítulo).

De acordo com Hess e Ostrom (2004), hoje em dia, a comunidade em torno do *commons* científico abrange os usuários de informação científica, os provedores de informação científica e os tomadores de decisões relacionadas à informação científica. Os provedores são aqueles que tornam o conteúdo disponível, como bibliotecários, os profissionais de tecnologia da informação, autores que praticam o *self-publishing*, os departamentos, as editoras, as sociedades científicas, etc.

Os usuários são geralmente os acadêmicos, mas também podem ser aquelas pessoas que esporadicamente consultam material científico ou fazem alguma pequena contribuição à Ciência. Num meio distribuído de informação, os usuários são freqüentemente desconhecidos.

Há ainda os tomadores de decisões em cada nível da comunicação científica. Eles podem ser os administradores que definem a quota de *e-mails*, os bibliotecários que escolhem quais pacotes de revistas comprar ou os membros do congresso americano, votando uma nova lei que pode mudar drasticamente o futuro do acesso à informação acadêmica. A comunidade dentro do sistema distribuído da informação tem se expandido largamente desde os dias em que o edifício da biblioteca era o lugar central de acesso à informação (Hess, Ostrom, 2004).

Naqueles dias e não há muito tempo atrás (cerca de 15 anos), Thomas Kuhn caracterizou a comunidade científica da seguinte forma:

Um paradigma é aquilo que os membros de uma comunidade partilham, inversamente, uma comunidade científica consiste em homens que partilham um paradigma. (Kuhn, 1962, p. 219).

De acordo com Kuhn (1962), uma disciplina científica está definida socialmente. Ele a enxerga como uma comunidade científica particular, unida pela educação (isto é, livros, métodos), interação profissional, comunicação (isto é, artigos, convenções), interesses similares em problemas de certos

tipos e aceitação de algumas soluções para tais problemas. A comunidade científica, como outras comunidades, define o que é necessário para ser membro de um grupo.

A afiliação a essa comunidade é geralmente definida em função da educação, status profissional e afiliação institucional. Alguns sociólogos da Ciência também descobriram que o sexo, a raça e a classe social também são fatores fortes para aceitação de entrada dentro de uma comunidade. Já uma pesquisa realizada pelo CEPAC mostra que aptidão é essencial para a carreira científica. Ela inclui: nível intelectual elevado ou mediano, criatividade, a habilidade para integrar conceitos remotamente associados e a capacidade para interpretar o conteúdo semântico nas operações intelectuais (Targino, 1998).

Os membros de uma comunidade nem sempre trabalham juntos. A comunicação entre os membros é estabelecida através da disseminação do trabalho de pesquisa e hipóteses através de artigos de revistas científicas ou através do comparecimento às conferências, onde nova pesquisa é apresentada e novas idéias são trocadas e discutidas. Existem ainda métodos informais de troca de informações como *e-mails* e listas de discussões. E muitos cientistas numa comunidade podem nunca se comunicar, devido a várias questões profissionais (Wikipedia, 2005).

As comunidades científicas não são formalmente organizadas por regras escritas. Seus membros comparam-se a “... cidadãos livres de uma república democrática de erudição” (Ziman, 1984, p. 81). Le Coadic (1996) sustenta que esse modelo vem sendo substituído por estrutura mais complexa. A profissionalização da pesquisa, com a inclusão de remuneração é um dos fatores de mudança, pois a edição de um artigo científico, além de confirmar competência, tem agora a missão de assegurar empregos, prêmios e outras recompensas.

Em termos ideais, é atribuição das comunidades científicas compartilhar conhecimentos com toda a sociedade, até porque, conforme estudiosos do tema, como Garvey(1977) e Meadows (1999), o processo de comunicação científica consiste na interação psicológica entre os interesses individuais e grupais, mediante influência recíproca e permanente.

3.5. Uso e Produção de Recursos no *Commons Científico*



Fig. 3.4. Ciclo da informação na Comunicação Científica. O pesquisador procura insumo para sua pesquisa nos centros de informação (biblioteca, livros, revistas), prepara a seu projeto, comunica seus resultados. Para tal fim, primeiramente utiliza os canais informais como conversas, cartas, *emails* etc. Em seguida, publica pelos canais formais tais como os periódicos, os relatórios técnicos etc. Futuramente seus resultados podem ser incorporados em resumos, monografias, índices e, finalmente em livros. Sua pesquisa poderá servir de insumo para uma nova pesquisa. Fonte: Redmond, Sinclair, Brown (1972).

Tanto para preparar quanto para divulgar seu trabalho, o cientista lança mão de diversos canais de comunicação desde os mais formais até os mais informais. E não são eles excludentes ou antagônicos. Ao contrário, complementam-se e interagem entre si. Ora apresentam características formais, ora informais, pois o comportamento dos cientistas no domínio informal inclui em seu escopo aspectos do comportamento formal. Assim, emerge a tradicional divisão: comunicação formal ou estrutura ou planejada e comunicação informal ou não estruturada ou não planejada [Targino, 1998].

A comunicação científica informal consiste na utilização de canais informais, em que a transferência de informação ocorre através de contatos interpessoais e de quaisquer recursos destituídos de formalismo, como reuniões científicas, participação em associações profissionais e colégios invisíveis. É a comunicação direta pessoa a pessoa. Incorpora formas públicas de troca de informações tais como conferências, colóquios, seminários, conversas, telefonemas, cartas, fax, visitas in loco a centros de pesquisa e laboratórios. Sua grande vantagem é a possibilidade de maior atualização e rapidez, por conseguinte, de menor custo. As informações repassadas informalmente

revestem-se de maior rapidez e redundância [Targino, 1998].

A comunicação científica formal se dá através de diversos meios de comunicação escrita, com destaque para livros, periódicos, obras de referência em geral, relatórios técnicos, revisões de literatura, bibliografias de bibliografias etc. É função do documento formal persuadir e convencer a comunidade científica e a sociedade como um todo de que os resultados então divulgados devem ser aceitos como conhecimento válido e consolidado [Targino, 1998].

Os sistemas formal e informal servem a fins distintos quanto à operacionalização das pesquisas. Ambos são indispensáveis à comunicabilidade da produção científica, mas são utilizados em momentos diversos e obedecem a cronologias diferenciadas. A disseminação através de canais informais precede a finalização do projeto de pesquisa. Em contrapartida, a trajetória da comunicação formal é demorada, como exemplificado, em detalhes, por Garvey e Griffith (1979).

Tanto os canais mais informais quanto os mais formais sofreram algumas mudanças com a chegada da tecnologia digital. Em especial nesse trabalho, estamos interessados em estudar as principais alterações nos canais formais, mais especificamente a publicação de revistas científicas.

A publicação de periódicos científicos, baseada na tecnologia da imprensa, de forma simplificada, funciona assim: o pesquisador vai até o edifício de uma biblioteca (da universidade ou da cidade, geralmente), procura nas revistas do seu campo de pesquisa o assunto de interesse, junta ao insumo (idéias das revistas) as suas próprias idéias e estudos, escreve um artigo e o envia à revista. Então, o documento é remetido via correio aos *referees*. Depois de vários meses, o documento pode ser diretamente enviado à publicação ou mandado de volta ao pesquisador para correções. Esse processo todo pode durar até anos. Algumas vezes, quando o artigo é finalmente publicado, seu conteúdo já está desatualizado. Depois da publicação, o documento pode ser adquirido pelas bibliotecas que assinem a revista onde ele foi publicado. Para um pesquisador ter acesso a esse documento, ele deve ir até o prédio de uma biblioteca que guarde a revista onde ele está contido.

Esse processo impõe algumas dificuldades no acesso. Existem algumas estatísticas, coletadas por Don King e seus colaboradores que mostram que conforme aumenta a distância física a uma biblioteca, o uso dela decresce dramaticamente (Odlyzko, 2001). Nesse mesmo artigo, Odlyzko observou a reação dos matemáticos na Universidade do Estado da Pensilvânia quando todas as revistas publicadas antes de 1973 foram enviadas para outro lugar de armazenamento por causa de limitações de espaço. Esta mudança não foi apreciada, ainda que qualquer volume pudesse ser obtido dentro de um dia. O ponto interessante é que a comunidade de pesquisa de Matemática passou a procurar muito menos esses itens.

Hoje, estamos convivendo tanto com a tecnologia da imprensa quanto a da Internet. No entanto, no meio digital, o processo de comunicação científica se tornou um tanto mais ágil. Os documentos são enviados via *e-mail* para a revista, *referees* e autores. Assim são recebidos quase instantaneamente.

O processo de *peer review* ainda é um pouco demorado, mas se estão experimentando formas alternativas de realizá-lo (veja, por exemplo, a revista ACP¹).

Quando o documento é publicado numa revista eletrônica, a forma de acesso pode ser muito diferente do jeito como seria realizada em papel. Se a revista oferece acesso aberto, então qualquer usuário, em qualquer parte do planeta pode lê-lo e utilizá-lo. O uso por parte de um não concorre com o uso de outro. Como as cópias digitais têm custo praticamente zero, muitos usuários podem acessá-la ao mesmo tempo. Porém, se a revista é de acesso fechado, então para um usuário ler o artigo em que está interessado, alguém precisa pagar pelo serviço. Pode ser o próprio usuário ao comprar o acesso a um único artigo. Também pode ser a biblioteca (ou a instituição de pesquisa que financia o seu trabalho), ao assinar pacotes que incluam essa revista. A política de restrição de acesso hoje é inserida dentro do código de programas, por isso muitas vezes não são muito claras.

Por outro lado, a acessibilidade no meio digital foi muito enriquecida com recursos como, por exemplo, busca por palavras nos títulos, resumo ou até mesmo texto completo. Porém, se quisermos fazer uma busca com as mesmas palavras-chaves nas revistas Science Direct e em revistas de outras editoras, é preciso repetir a busca porque as interfaces são diferentes. Observamos falta de integração entre os serviços existentes. No entanto, muitas iniciativas (CrossRef², OpenURL³, entre outras) estão se desenvolvendo atualmente para diminuir esse problema.

Apesar dessas diferenças entre a era do papel e a era digital nas formas de uso e produção da literatura acadêmica, alguns aspectos não mudaram. Ainda é possível reconhecer o padrão apresentado na figura 2.4. Do *commons*, o cientista retira para o seu trabalho as idéias, o conhecimento e as teorias contidos nos artefatos. Esse material serve de base para o seu trabalho.

Ele associa a esse insumo suas idéias e desenvolve experiências ou estudos. A partir dos resultados da pesquisa, escreve artigos e os envia às revistas científicas, onde são validadas as informações através do *peer review*. Se o artigo é publicado, então nova informação é acrescentada ao *commons* científico. Esses resultados podem servir futuramente de insumo para o desenvolvimento de outras pesquisas, feitas tanto pelo próprio autor como por outros cientistas. Trata-se de um ciclo. E quanto maior o acesso à informação, potencialmente mais informação pode ser gerada e refinada.

1- Um novo modelo está sendo implementado pela revista Atmospheric Chemistry and Physics (ACP), uma revista extremamente respeitada, com pesquisadores dos Max Planck Institute no corpo editorial. Nela, um *peer review* inicial garante a relevância e coerência com assuntos da revista além de qualidade científica e técnica. Depois acontece uma discussão interativa e pública *on-line* pelos *referees*, autores e outros membros da comunidade científica. Se não há comentários da comunidade, uma revisão tradicional é aplicada, mas de forma mais transparente. A sequência de passos para o processo de *peer review*, publicação e discussão interativa é: a) submissão eletrônica; b) *peer review* inicial - o editor aceita ou rejeita a publicação para discussão; c) correções técnicas - o autor tem a oportunidade de fazer correções técnicas; d) publicação do artigo para discussão no *site* Atmospheric Chemistry and Physics Discussion (ACPD) - a disponibilidade do trabalho é garantida permanentemente; e) discussão aberta (8 semanas); f) resposta final (até 4 semanas) - comentários dos autores; g) submissão de um artigo revisado para publicação não é obrigatória, mas encorajada; h) o editor aceita ou rejeita o artigo revisado ou consulta *referees*; i) publicação final do artigo revisado no ACP. Todos os comentários e discussão pública ficam permanentemente disponíveis no *site* da ACPD. Para saber mais detalhes, veja <http://www.copernicus.org/EGU/acp/>

2- <http://www.crossref.org/>

3- http://www.crossref.org/02publishers/openurl_info.html

3.6. Regras do *Commons Científico*

No capítulo sete do livro *Code* (Lessig, 1999), publicado recentemente de forma aberta na Rede, o autor define o que regula as coisas. Ele afirma que quase tudo é regulado por **mercado, normas sociais, arquitetura e leis**. Ele dá como exemplo o cigarro. Este é regulado por preços de mercado, normas sociais como, por exemplo, não fumar dentro de um carro quando se está acompanhado, a arquitetura do cigarro, ou seja, o material a partir do qual é feito. Há cigarros, por exemplo, cujo teor de nicotina é menor e isso traz menores prejuízos à saúde. Há ainda leis que regulam o uso do cigarro. Por exemplo, é proibido fumar dentro de salas de aula nas escolas.

De acordo com Lessig, a Internet também é regulada por mercado, normas, arquitetura e leis. Através de nosso estudo, também concluímos que esses quatro fatores regulam o *commons* científico. Esses fatores se relacionam direta e respectivamente com os aspectos econômicos, sociais, tecnológicos e legais do *commons* em questão.

Quanto ao mercado, algumas regras são ditadas por estruturas corporativas e mentalidade corporativa na Ciência ou por conselhos de financiamento poderosos, agências e fundações. Essas agências controlam a distribuição dos fundos de pesquisa entre os cientistas e, conseqüentemente influenciam quais pesquisas são feitas e a quem é permitido fazê-las. Indiretamente, isso determina o que será publicado no *commons* científico.

Aliás, a publicação dos resultados científicos tem sido bastante cobrada pelas instituições, principalmente devido à profissionalização da pesquisa e à introdução da remuneração do cientista. Hoje, essa publicação serve também para assegurar empregos. Se o artigo publicado do cientista alcança impacto internacional, isso ainda pode trazer ao pesquisador prêmios e promoções.

As leis de propriedade intelectual regulam como os artefatos do *commons* científico são utilizados e como trabalhos derivados podem ser criados. O direito autoral dá ao autor direitos exclusivos de distribuição do seu trabalho. Assim o autor pode obter lucros através dele. Também pode indiretamente controlar quem tem acesso ao recurso. No entanto, as editoras de revistas científicas requerem do autor a transferência desses direitos exclusivos e não remuneram os autores nem as instituições de pesquisa pelo trabalho. Ao invés disso, elas utilizam esse direito para permitir que apenas quem pague um valor estipulado seja capaz de ter acesso ao recurso científico. A tecnologia digital tem auxiliado as editoras a reforçarem essa regra através de portais que solicitam senhas de acesso.

A arquitetura do meio digital, seus protocolos, padrões e código de computador freqüentemente tornam as regras do *commons* científico mais ocultas. Ainda mais problemático que isso é a

ocorrência de uma inversão causada pela tecnologia, por causa das capacidades tecnológicas poderem, em princípio e silenciosamente, contraporem-se a missões, a valores e até mesmo a direitos constitucionais. Além disso, as leis de direito autoral americanas anteriores a 1998 faziam claras exceções através dos princípios de *fair use* e *first sale*. No entanto, os legisladores do Digital Millenium Copyright Act não transferiram esses direitos de forma apropriada para o meio digital (Hess e Ostrom, 2004).

Voltaremos aos aspectos legais, econômicos e sociais do *commons* científico no capítulo cinco, ao tratarmos da transição. Aos aspectos tecnológicos do Acesso Aberto, dedicamos o capítulo seis.

3.7. Crise no Acesso ao *Commons* Científico

A mudança gradativa de publicação sem fins lucrativos para publicação com fins lucrativos desde os anos 50 trouxe alguns problemas para a publicação científica. As editoras comerciais começaram a aplicar políticas de preço elevados às assinaturas de revistas. Por outro lado, o orçamento das bibliotecas não acompanhou o aumento de preços. No sistema de publicação científica atual, há basicamente duas perspectivas: a perspectiva da comunidade acadêmica, onde tanto a produção quanto o consumo da informação científica são parte de uma cultura de troca; e a perspectiva das editoras comerciais, onde aparecem dois mercados nos quais as trocas são governadas por contratos - o mercado dos autores e o dos consumidores (principalmente bibliotecas acadêmicas) (Lyman, 1997).

No mercado dos consumidores, ocorreu um aumento de 227% nas despesas das bibliotecas da ARL (Association of Research Libraries) com aquisições de livros e revistas científicas entre 1986 e 2002. Os administradores das universidades estão começando a reconhecer que esse não é um problema criado pelas bibliotecas. Como sinal de insatisfação com os preços elevados, a Universidade Cornell cancelou cerca de 200 títulos da Reed Elsevier. Nos últimos anos, a Universidade de Wisconsin, Harvard e a Universidade da Califórnia também estiveram se manifestando contra os altos custos das assinaturas da publicação comercial (Library Committee Recommendation, 2004).

Mesmo com o aumento de preço, a academia ainda subsidia a economia de mercado, e, notoriamente aqueles que subsidiam a pesquisa não requerem direito autoral da propriedade intelectual que sustentam. Os subsídios incluem, por exemplo, o investimento federal em pesquisa, subsídios institucionais e trabalho voluntário aos serviços editoriais. É importante ressaltar que, em geral, o autor não é remunerado pelas editoras. Frequentemente, os editores ou *referees* também

prestam gratuitamente seus serviços às editoras. Os autores entregam seus artigos às editoras apenas para disseminá-los. Da forma como o sistema está estabelecido, as idéias da academia se transformam em propriedade intelectual e são vendidas de volta à academia. E ainda: as publicações acadêmicas são consumidas dentro das bibliotecas, que são um bem aberto a suas comunidades e subsidiado, dentro das quais o conhecimento surge para o leitor de forma gratuita.

Outro desafio que a ciência enfrenta hoje é a dificuldade no gerenciamento do crescimento exponencial da literatura científica. O número de artigos publicados anualmente tem dobrado a cada 10-15 anos nos últimos dois séculos (Odlyzko, 1995). A duplicação na publicação de artigos a cada década corresponde a um crescimento exponencial de 7% ao ano. Além disso, a ciência tem se especializado cada vez mais. E hoje é impossível que uma biblioteca, mesmo entre as mais ricas do mundo, consiga adquirir todos os títulos. Há ainda o problema da falta de espaço para armazenar os livros e revistas nas bibliotecas das universidades. Na biblioteca da Universidade de Nova Iorque, onde o metro quadrado de terreno é caríssimo, a questão do espaço torna-se crítica.

CAPÍTULO 4 - Acesso Aberto

De acordo com Tenopir e King (2001), aproximadamente dois terços das revistas acadêmicas atuais migraram para o formato digital e continuaram a publicar em papel e cerca de 1000 revistas com *peer review* estão somente na forma *online*. Nessa mudança, muitos passos são necessários. Willinsky (2003) ressalta que em cada passo, escolhas significantes feitas pelo tomadores de decisões durante o período de transição vão determinar e modelar o futuro do acesso ao *commons* científico.

Apesar do meio digital oferecer muitas possibilidades de formas de publicação, várias revistas simplesmente migraram para o formato digital e conservaram a mesma estrutura e forma das revistas tradicionais. Outras enxergaram as possibilidades de inovação desse novo formato e criaram formas alternativas de publicação. Uma dessas formas é o modelo de disseminação do conhecimento em acesso aberto, ou seja, o fluxo livre da informação científica.

Ao redor desse novo modelo, tem se intensificado o movimento social de Acesso Aberto. Sua mola propulsora não foi apenas a Internet, mas também o desejo dos cientistas de fazerem uso das possibilidades novas de disseminação do seu trabalho através das tecnologias oferecidas pelo advento e desenvolvimento da Internet.

As revistas científicas de acesso aberto, de acordo com Willinsky (2003), são aproximadamente 10-20% do total das revistas científicas. Além desse modelo de publicação aberta, há também a possibilidade de depositar um artigo num repositório digital. O movimento de Acesso Aberto à literatura científica é uma iniciativa dentro de uma “onda” muito maior em defesa do interesse público na Internet, que pretende garantir o direito democrático do saber.

4.1. Vantagens

O modelo de acesso aberto tem muitas vantagens sobre o molde tradicional. Algumas delas são exibidas a seguir:

1. **Redução de custos** – os modelos de publicação de acesso aberto são consideravelmente mais baratos do que os convencionais.
2. **Maior acessibilidade ao *commons* científico** - qualquer pessoa (seja um acadêmico ou um leigo) que tenha um computador conectado a Internet pode acessar os artigos científicos abertos a qualquer hora, em qualquer lugar. O CiteSeer é um dos exemplos mais bem

sucedidos nessa direção (Simon, 2002).

3. **Maior visibilidade dos artigos** – de acordo com Lawrence (2001) e Harnad (2004), os artigos científicos nesse modelo tem maior número de citações do que os publicados no modelo fechado.
4. **Maior integração da informação no *commons* científico** – com o auxílio de ferramentas computacionais avançadas, os dados bibliográficos e texto completo podem ser digeridos e relacionados de modo a melhorar enormemente a associação de conceitos correlatos, bem como a navegação.
5. **Maior velocidade na dinâmica das descobertas científicas** – se largamente aplicado, esse modelo tem a capacidade de melhorar significativamente a comunicação acadêmica e, conseqüentemente, as descobertas da Ciência aconteceriam mais rapidamente.
6. **Interoperabilidade via protocolo OAI** – através de uma interface produzida pela equipe da Open Archives Initiative (OAI), é possível coletar automaticamente dados de artigos científicos armazenados em bibliotecas distribuídas na Rede (essa questão é detalhada no capítulo 6).

4.2. O que é

O Acesso Aberto (A. A.) refere-se ao acesso *on-line* sem barreiras e permanente de textos completos de artigos científicos, de forma livre para qualquer um que esteja conectado à Internet. Esse modo de publicação tem duas propriedades essenciais: os materiais estão gratuitamente disponíveis e o detentor do direito autoral permite leitura irrestrita, o privilégio de fazer busca, *download*, indexar, copiar e imprimir o texto completo; algumas vezes também é permitido redistribuir o texto completo. Os metadados do documento estão normalmente disponíveis e podem ser copiados, redistribuídos, indexados, etc.

Existem duas principais definições do Acesso Aberto. A primeira está na Declaração de Budapeste (2002) e diz o seguinte:

Por ‘acesso aberto’ a essa literatura, nós queremos dizer sua livre disponibilidade na Internet pública, permitindo a todos os usuários lerem, fazerem o download, copiar, distribuir, imprimir, buscar, ou fazer links ao texto completo desses artigos e o crawling para indexação (...). A única restrição na reprodução e distribuição e o único papel para o direito autoral nesse domínio devem ser dar ao autor o controle sobre a integridade de seu

trabalho e o direito de ser apropriadamente reconhecido e citado (...)

Essa definição implica no acesso livre aos textos completos através da Internet, seu uso e distribuição respeitando o direito autoral. A definição seguinte consta na Declaração de Bethesda e é semelhante à anterior, porém adiciona a condição de depósito do artigo num repositório institucional, imediatamente depois da publicação na revista.

Uma publicação de Acesso Aberto é aquela que atinge as seguintes condições:

- 1. O autor e o detentor de direito autoral garantem a todos os usuários o direito de acesso perpétuo no mundo todo, irrevogável e livre, bem como uma cópia de licença para copiar, utilizar, distribuir, transmitir e mostrar o trabalho publicamente e fazer e distribuir trabalhos derivados, em qualquer meio digital para qualquer propósito respeitável, sujeito a atribuir apropriadamente o direito autoral e também o direito de fazer um pequeno número de cópias impressas para uso pessoal.*
- 2. A versão completa do trabalho e também de todo material suplementar, incluindo a cópia de permissão como estipulada acima, num formato eletrônico de padrão conveniente é depositada imediatamente após a publicação em ao menos um repositório on-line mantido por uma instituição ou sociedade acadêmica, agência governamental ou outra agência bem estabelecida que procure disponibilizar o acesso aberto, a distribuição irrestrita, a interoperabilidade e o arquivamento por longo tempo.*

Outra adição importante feita pela declaração acima é sobre a publicação da licença. É essencial que as permissões estejam claras e que a licença esteja facilmente disponível para consulta tanto por robôs quanto por humanos. Um meio fácil, efetivo e crescentemente comum para os detentores do direito autoral manifestarem seu consentimento sobre o Acesso Aberto é utilizar uma das licenças Creative Commons¹. É possível compor a sua própria licença de permissões e anexá-la a seus trabalhos (Suber, 2005).

De forma geral, o pré-requisito legal do Acesso Aberto é que o trabalho esteja livre de restrições legais que impeçam o Acesso Aberto. A base legal do A. A. é o consentimento do portador do direito autoral para publicação do trabalho de forma aberta. Por isso, não requer a abolição, a reforma ou a violação do direito autoral. O pré-requisito econômico do A. A. é encontrar meios para pagar pelos pré-requisitos físicos e o *peer review*. Os pré-requisitos físicos para o Acesso Aberto são: o trabalho precisa ser digital e residir num servidor da Internet. (Suber, 2005)

1- <http://creativecommons.org>

4.3. Breve Histórico

O acordo entre autor e editora de revista científica que era feito quando a única forma de distribuição era impressa em papel (transferência do direito autoral) favoreceu as editoras cujos lucros são feitos através da restrição de acesso (somente àqueles que pagam pelo conteúdo). Com a Internet, agora o autor não depende mais tanto das editoras e formas alternativas de publicação *on-line* estão se proliferando.

O Acesso Aberto se tornou visível principalmente através do arXiv¹, um repositório de temas variados, concentrados principalmente em algumas áreas da Física, originalmente produzido no Laboratório de Los Alamos. Em 1999, a Open Archives Initiative (OAI)² apareceu para facilitar a troca de metadados entre repositórios desse tipo. Foi produzido um protocolo para coleta de dados de portais distribuídos. Isso trouxe a possibilidade de estender o modelo de *self-archiving* do arXiv para outros repositórios digitais institucionais. (Brody, Hitchcock, 2005)

O protocolo foi desenhado para se aplicar a todos os tipos de materiais que poderiam ser encontrados numa biblioteca digital. A OAI acabou perdendo seu objetivo inicial, o acesso *on-line* gratuito a artigos científicos. Esse foco foi recuperado e formalizado como Acesso Aberto (A. A.) na Budapest Open Access Initiative, onde foram identificadas duas rotas para o A. A.: o depósito de um artigo feito pelo próprio autor (*self-archiving*) num repositório digital institucional ou temático e a submissão de artigos a revistas científicas de acesso aberto (Brody, Hitchcock, 2005).

Para maiores informações sobre a história e desenvolvimento do Acesso Aberto, veja as seções 4.3 e 6.1 e também a Linha do Tempo do Acesso Aberto (Suber, 2006b).

4.4. Rota 1: Repositórios Institucionais ou Temáticos

Os repositórios de Acesso Aberto podem ser institucionais, aqueles que contém a publicação da comunidade científica de uma instituição de pesquisa. Também podem ser temáticos, ou seja, repositórios que colecionam artigos científicos sobre um certo assunto ou especialidade de pesquisa.

1- <http://arxiv.org>

2- <http://www.openarchives.org>

Um dos primeiros repositórios, o arXiv, era tanto institucional quanto temático. Agregava a publicação do Laboratório de Los Alamos e no início, continha apenas artigos da área de Física de Altas Energias. Com o tempo se expandiu para outras áreas principalmente algumas outras áreas da Física, a Matemática e a Ciência da Computação. Hoje é mantido pela Universidade de Cornell e contém muito mais do que a publicação da própria instituição. Há cerca de 20 espelhos esparramados pelo mundo.

Os repositórios temáticos e os institucionais podem conter tanto *preprints* (artigos depositados antes de serem publicados numa revista científica) quanto *postprints* (que já foram aceitos por uma revista após passarem pelo processo de *peer review*). Ainda podem ser encontrados nesses repositórios teses e dissertações, materiais de curso, objetos de aprendizado, arquivos de dados, áudio ou vídeo ou outro tipo qualquer de arquivo digital relacionado a atividades acadêmicas.

Os autores de artigos que desejam depositar seu *preprint* num repositório de Acesso Aberto não precisam de permissão de ninguém para fazê-lo. Porém, se o autor transfere o direito autoral para uma revista, então ele precisa da permissão do detentor do direito autoral para poder depositar o seu trabalho numa biblioteca digital de acesso aberto. A maior parte das revistas já permite essa prática. Para uma busca num banco de dados de políticas de revistas científicas relativas ao direito autoral e o *self-archiving* (a prática onde o autor deposita um *eprint* numa biblioteca digital de Acesso Aberto), veja Project SHERPA¹.

Existem alguns *softwares* livres que implementam a funcionalidade necessária para esses repositórios. Os dois mais importantes são o EPrints² e o DSpace³ e ambos são compatíveis com o protocolo de coleta de metadados OAI.

Os dois maiores repositórios que tem um crescimento maior e mais rápido em termos de textos completos são o repositório de CERN⁴ e repositório de Eletrônica e Ciência da Computação de Southampton⁵. Os dois implementam políticas de *self-archiving* obrigatório, ou seja, a instituição de pesquisa requer que toda sua publicação seja depositada no seu repositório (Brody, Hitchcock, 2005).

1- <http://www.sherpa.ac.uk>

2- <http://software.eprints.org>

3- <http://dspace.org>

4- <http://cdsweb.cern.ch/>

5- <http://eprints.ecs.ac.uk/>

Devido à lentidão em se preencher os repositórios institucionais com artigos científicos, desenvolveu-se a idéia do *self-archiving* obrigatório. Essa idéia existe pelo menos desde 1998, mas foi a partir de 2003 que esforços maiores têm sido feitos pelos financiadores de pesquisa como governos, agências e universidades. Em 2005, as duas maiores agências financiadoras de pesquisa médica, o NIH (National Institute of Health, EUA) e a Wellcome Trust (Inglaterra), adotaram políticas de, respectivamente, recomendação e exigência para fornecer acesso aberto aos resultados das pesquisas financiadas pelas mesmas. Assim, os artigos devem ser depositados na Pubmed Central¹ ou no repositório central inglês de Biomédicas (em construção), quando este estiver disponível.

A Pubmed Central é um repositório temático de periódicos de Ciências da Vida muito importante e grande (conta com mais de 370 mil artigos científicos). Não aceita publicações diretamente dos autores; ao invés disso, as editoras contribuem tornando o texto completo da sua pesquisa disponível através do arquivo do PMC.

A iniciativa foi do Dr. Harold Varmus, então diretor do National Institutes of Health (NIH), em 1999, quando apresentou uma ambiciosa proposta para o NIH desenvolver e operar um *site* de publicação eletrônica que forneceria acesso sem barreiras à literatura de ciências médicas e biológicas (*life sciences*). O resultado, PubMed Central (PMC), foi lançado em 2000 com conteúdos do Proceedings of the National Academy of Sciences e do Molecular Biology of the Cell. Hoje conta com mais de 150 títulos de revistas científicas com *peer review*. No entanto, os materiais da PMC só se tornam disponíveis livremente depois de algum tempo da publicação da revista. Por isso, não completa suficientemente os requisitos para ser considerado um repositório de Acesso Aberto (Brody, Hitchcock, 2005).

Em se tratando de repositórios e bibliotecas importantes, o CiteSeer² precisa ser mencionado devido principalmente a seus recursos de análise automática de textos e impacto na Ciência da Computação. Foi desenvolvido pelo NEC Institute e pela Universidade da Pensilvânia. Ele conta hoje com mais de 700 mil artigos nessa área. Além de armazenar e organizar os *eprints*, também os indexa, gera estatísticas e permite navegação na bibliografia e nos artigos semelhantes. Ele localiza por meio de robôs artigos na Rede, extrai informações como as citações, o contexto das delas, o título do artigo, etc., e também faz indexação do texto completo.

1- <http://www.pubmedcentral.nih.gov/>

2- <http://citeseer.ist.psu.edu/>

Ao invés de oferecer apenas uma outra biblioteca digital, ele dispõe os seguintes recursos:

1. A criação automática de um índice de apontadores entre artigos científicos, semelhante ao Science Citation Index;
2. A implementação de algoritmos que localizam documentos relacionados, baseados nas informações de citação e também de palavras usuais;
3. Análise automática de citações e elaboração de grafos de citação;
4. As informações bibliográficas do artigo no formato BibTeX, preparadas automaticamente;
5. Estatísticas e gráficos exibindo detalhes do artigo como seu ciclo de vida.
6. A possibilidade de navegação nos artigos da referência bibliográfica de um determinado artigo e também dos artigos que o citam - assim é possível não apenas rastrear de onde veio determinada idéia, mas principalmente quais as idéias derivadas de um certo artigo.

Se fosse suficientemente completo em termos de literatura, poderíamos afirmar que o CiteSeer nos daria uma visão do paradigma atual da Ciência da Computação. E melhor do que isso: com suas ferramentas sofisticadas de busca e indexação, ele poderia nos ajudar a detectar as teorias mais importantes e aceitas no momento.

4.5. Rota 2: Revistas de Acesso Aberto

As revistas de Acesso Aberto são semelhantes às revistas científicas tradicionais: têm *peer review*, editoração, formatação, etc. A diferença principal é a do livre acesso *online*, ou seja, os usuários (quer sejam bibliotecas, instituições de pesquisa ou indivíduos) não precisam pagar pelo acesso. Basta que tenham o equipamento correto, ou seja, computadores conectados a Internet. Outra diferença é que esse tipo de revista costuma deixar o direito autoral com o próprio autor.

Existem algumas revistas A. A. que têm fins lucrativos (por exemplo, as revistas da Biomed Central¹ ou BMC) e outras não (as revistas da PLoS²). Quanto à forma de financiamento, algumas revistas A. A. pagam suas despesas por meio de subsídios de universidades ou de sociedades científicas. Outras aplicam uma taxa sobre os artigos aceitos. Ela pode ser paga pelo próprio autor ou agência financiadora de pesquisa do autor. As revistas que tem subsídios institucionais em geral não aplicam taxa de aceitação do artigo. Também existe a possibilidade de uma instituição pagar uma taxa anual à revista A. A. e então todos os seus pesquisadores podem publicar nessa revista.

1- <http://www.biomedcentral.com>

2- <http://www.plos.org/journals/>

Um mal entendido comum é achar que todas as revistas A. A. utilizam o modelo em que o “autor paga”. Existem muitos modelos. Na verdade apenas 47% das revistas de A. A. aplicam a taxa de aceitação. E ainda, quando uma revista aplica essa taxa, normalmente é paga pela instituição financiadora de pesquisa e não pelo próprio bolso do autor (Suber, 2005).

Há um serviço agregador de revistas bastante famoso, o Diretório de Revistas de Acesso Aberto (Directory of Open Access Journals¹ – DOAJ) da Biblioteca da Universidade Lund. Ele oferece livre acesso a 1980 revistas com mais de 83 mil artigos (DOAJ, 2006). Aceita revistas em qualquer área da Ciência, em qualquer língua e de qualquer lugar do mundo. A idéia da criação desse diretório foi discutida na Primeira Conferência Nórdica em Comunicação Científica em Lund/Copenhague e a conclusão foi de que seria um serviço de grande valia para a pesquisa global.

Algumas revistas de Acesso Aberto importantes têm sido publicadas pela Public Library of Science (PLOS). Elas estão incluídas no DOAJ e são cinco: PLoS Biology, PLoS Computational Biology, PLoS Genetics, PLoS Medicine e PLoS Pathogens. Apenas dois anos após o lançamento da primeira revista, o PLoS Biology tem um fator de impacto de 13.9 de acordo o Thomson ISI. Isso a caracteriza como uma das revistas mais citadas na área de Ciências da Vida.

A Biomed Central (BMC) é uma editora comercial que provê acesso livre de toda a sua literatura. Conta com mais de 130 revistas científicas. Os autores que publicam na BMC retêm o direito autoral de seus trabalhos, licenciando-os sob a Creative Commons Attribution License. Essa licença permite o *download* do texto completo, o seu reuso e sua redistribuição sem restrições, desde que o trabalho original seja corretamente citado.

Em 2002, a BMC começou a cobrar 500 dólares para publicar artigos aceitos. Alternativamente, instituições começaram a pagar uma taxa de afiliação. Para promover a publicação em revistas A. A., o JISC (Joint Information Systems Committee) fez um acordo com a BMC para cobrir todas as assinaturas institucionais na Inglaterra a partir de julho de 2003, estendendo o acordo até setembro de 2005. Com a taxa de autores e instituições e também com propagandas no portal, os custos foram redirecionados dos indivíduos para as instituições que os utilizam para publicação.

Uma iniciativa brasileira e de considerável destaque internacional é o Scielo¹. Trata-se de um sistema eletrônico de portais para acesso a coleções digitais de textos completos de cerca de 200 revistas científicas. Existem muitos *sites* Scielo em operação, oferecendo acesso a revistas do Brasil, América Latina e Caribe. Tem desempenhado uma função importante na disseminação do conhecimento técnico-científico desses países, possibilitando acesso a uma literatura que antes só

1- <http://www.doaj.org>

2- <http://scielo.org>

podia ser alcançada dentro das fronteiras. De acordo com Meneghini, um dos principais responsáveis pelo repositório, o fator de impacto das revistas científicas do Scielo indexadas pelo ISI aumentou em média 42% entre 1998 até 2001.

4. 6. O *Commons* do Acesso Aberto

É interessante caracterizarmos a literatura de acesso aberto como um *commons*. Dessa forma, enfatizamos os recursos, a comunidade, as regras, bem como as questões de rivalidade de recursos, neutralidade e transparência no acesso.

Os recursos desse *commons* são os artigos, as revistas e os facilitadores de acesso aberto, ou seja, bibliotecas ou repositórios digitais como o arXiv, o CiteSeer, o Scielo, entre outros. Esses recursos são praticamente não-rivais por serem digitais e não-excludentes estarem livremente disponíveis na Rede. Por causa desta primeira propriedade, os recursos de acesso aberto não estão sujeitos a problemas relacionados à exploração excessiva (tragédia dos *commons*). No entanto, necessita de incentivos para que cresça. Esse é um ponto central em se tratando de dificuldades para uma transição do sistema atual de publicação científica para uma publicação predominantemente de acesso aberto.

Por definição, a literatura de acesso aberto não exclui ninguém, ou ao menos ninguém com uma conexão da Rede. Assim, o acesso é neutro, não sendo necessário pedir permissão a ninguém para utilizar os recursos. Por outro lado, as revistas eletrônicas que não são de acesso aberto optam quase sempre por excluir os não-assinantes. Os mecanismos de controle custam dinheiro para as editoras. Esse é um outro motivo pelo qual a literatura A. A. torna-se mais barata: ela dispensa as técnicas de DRM (Digital Rights Management). Um segundo custo é o gerenciamento de contratos e assinantes (Suber, 2004).

A combinação do acesso neutro a esses recursos mais a implementação dos mesmos na Internet abrem largamente as portas de disseminação do conteúdo e das idéias científicas contidas neles. Elas podem se espalhar por vários pontos do globo e assim agregar de forma mais coesa e completa uma comunidade científica em torno deles. As oportunidades de trocas de conhecimento, interação, reunião, fortalecimento e até mesmo de formação de uma comunidade científica específica em torno de certos recursos crescem enormemente. Estes valores são determinantes para o entusiasmo com que uma parte da comunidade científica persegue os objetivos do A. A. Infelizmente, esse entusiasmo não contagiou ainda a comunidade toda.

O arXiv é um serviço que conquistou muitas comunidades em torno dos temas abordados no seu

portal. Tão grande é a sua importância em algumas áreas da Física que, para garantirem o reconhecimento de suas descobertas, os cientistas sentem-se obrigados a publicar nele. Assim, eles recebem o crédito por serem os primeiros a estudar ou descobrir determinadas novidades no seu campo de pesquisa.

O Pubmed Central também é um serviço que tem reunido muitas comunidades da área de saúde em torno dele. Infelizmente, tanto o arXiv, o Pubmed e outros serviços relacionados oferecem pouco suporte para integração das comunidades em torno dos seus recursos e de seus temas.

Até o momento, nesta seção, referimo-nos a comunidades em torno de campos de pesquisa específicos. Podemos pensar na comunidade de Acesso Aberto também sob outra perspectiva, mais geral. A comunidade de Acesso Aberto abrange pesquisadores, leitores esporádicos, bibliotecários, programadores, editoras científicas, legisladores, entre outros; resumindo: todas as pessoas e instituições que direta ou indiretamente estão realizando o Acesso Aberto. Esses agentes formam uma comunidade muito heterogênea e ativa que, em um âmbito geral, faz progressos constantes no sentido de liberar o acesso a parcelas crescentes da literatura científica.

Algumas regras praticadas pela comunidade em torno da literatura de acesso aberto estão formalizadas nas declarações de Acesso Aberto, nas leis de direito autoral e nas licenças utilizadas para publicação. No entanto, nem sempre as regras são transparentes. Muitas vezes os portais não deixam claras questões sobre o que se pode fazer com os recursos em termos de trabalho derivativo, distribuição e construção de novos serviços a partir da coleta de metadados e de texto completo. Essa falta de clareza dificulta a criação de novos facilitadores de acesso que poderiam ajudar enormemente os pesquisadores e servir de incentivo para mais cientistas lerem e publicarem artigos em acesso aberto.

As regras do *commons* de Acesso Aberto estão atualmente em negociação. Talvez esse seja um dos motivos pelo qual elas não sejam tão transparentes. Há uma agitada interação entre os valores e interesses de toda a comunidade científica. Esta interação (cujo resultado é imprevisível) determinará se o Acesso Aberto será um paradigma para a publicação científica e de que forma ele será realizado. Maiores detalhes serão vistos no capítulo cinco.

4.7 – A questão dos incentivos para produção de recursos do *commons* A. A.

Um dos maiores incentivadores de publicação científica é o impacto. O Acesso Aberto tem explorado seu já comprovado aumento de impacto para atrair mais autores. No entanto, há muita

falta de incentivo também nesse tipo de publicação.

4.7.1 – O fator de impacto no A. A.

Os autores de artigos científicos quase na sua totalidade não escrevem pelo dinheiro. Esses autores escrevem pelo impacto (Suber, 2004). Se esses autores virem um aumento de impacto através da publicação A. A., então eles estarão mais propensos a publicarem nesse modelo. Brody (2004) concorda que:

Maior acesso proporciona maior impacto.

Mais especificamente, em média, artigos de acesso aberto revisados por pares são mais citados que aqueles textos completos que estão disponíveis apenas através de assinaturas. Encontramos alguns estudos que demonstraram esse fato.

O primeiro foi realizado por Lawrence (2001). Ele analisou as taxas de citações na área de Ciência da Computação, comparando os artigos disponíveis livremente *on-line* contra os artigos que não foram disponibilizados *online*. Relatou o seguinte:

... clear correlation between the number of times an article is cited and the probability that the article is online. More highly cited articles, and more recent articles, are significantly more likely to be online. The mean number of citations to offline articles is 2.74, and the mean number of citations to online articles is 7.03, an increase of 157 percent (Lawrence, 2001)

Antelman (2004) estudou o impacto de artigos livremente disponíveis nas disciplinas Filosofia, Ciências Políticas, Matemática, Engenharias Elétrica e Eletrônica em diferentes estágios de adoção do Acesso Aberto. Os resultados mostram que os artigos A. A., nessas quatro disciplinas, tiveram uma taxa média de citação maior do que os que não estavam livremente disponíveis na Rede em acesso aberto. Harnad (2004) revelou na Física taxas de citação de 2.5 a 5.8 vezes maiores que as dos artigos da mesma área que não são de acesso aberto.

Já para as revistas A. A., a vantagem de impacto do Acesso Aberto não foi amplamente demonstrada ainda. O Thompson ISI, utilizando uma seleção de revistas de Acesso Aberto na área de Ciências Naturais encontrou que as revistas A.A. e as outras têm fatores de impacto similares (Pringle, 2004). No entanto, dado o pequeno número de revistas A. A. em comparação com as

revistas que não são de acesso aberto, e o período curto de tempo analisado, esse resultado pode ser visto como positivo para o Acesso Aberto. Crescentes fatores de impacto têm sido relatados também pela PLoS (PLoS, 2006).

4.7.2. A questão dos desincentivos

De acordo com Anderson (2004), convencer as pessoas a ler a literatura de Acesso Aberto provavelmente não será um grande problema, principalmente se houver conteúdo de qualidade. Guédon (2004) argumenta que as facilidades de acesso, ou melhor, ferramentas sofisticadas de busca, navegação e indexação podem ser cruciais para atrair o leitor. E quanto mais leitores houver, mais autores vão querer publicar no Acesso Aberto. Além disso, quanto mais completa for a literatura de um repositório numa dada disciplina (tema), muito maior será o seu valor para busca dos resultados científicos de pesquisa.

No atual mercado de revistas científicas de Acesso Aberto, há revistas científicas que atraem muitos autores devido ao seu grande prestígio. Porém, para as revistas novas alcançarem esse nível pode levar alguns anos. Enquanto isso, os autores preferirão publicar em outras revistas.

Porém, o movimento de Acesso Aberto não se dá apenas através das revistas de Acesso Aberto, mas também através do *self-archiving*, onde o autor deposita *preprint* ou *postprint* (com a permissão da editora) numa biblioteca digital. Esse caminho tem sido apontado por Stevan Harnad como o meio mais rápido de alcançar acesso aberto de toda a literatura hoje publicada.

Todavia, essa questão não é tão simples. Para reduzir a demanda pelas revistas comerciais é um longo caminho. O *self-archiving* progride gradual e anarquicamente, e as revistas não podem ser canceladas enquanto apenas algumas partes aleatórias do seu conteúdo estão disponíveis abertamente. Outra ameaça ao *self-archiving* é a possibilidade das editoras repentinamente quererem rescindir a permissão para essa prática, e assim fazer desmoronar boa parte do Acesso Aberto.

Muitas das bibliotecas digitais de Acesso Aberto são implementadas em instituições de pesquisa por meio de *softwares* livres como DSpace (do MIT) ou EPrints (da Universidade de Southampton). Como há poucas instituições com boas políticas para encher esses repositórios, então frequentemente são armazenados neles materiais de aula, relatórios técnicos, *preprints* e *postprints*. E as disciplinas são as mais variadas possíveis. Existem provedores de serviços como o OAIster, que coletam os metadados (alguns dados bibliográficos) dos artigos depositados nesses repositórios institucionais. O resultado é um portal que dá acesso a cerca de 6 milhões de registros com diferentes níveis de qualidade, que são difíceis de se distinguir.

Além disso, não há padrões de classificação eficientes. Portanto, a busca se torna um tanto pobre e se tem pouca acessibilidade. E o que é mais fácil para se trabalhar: o OAIster ou o Science Direct da Elsevier? Se o autor tiver essa escolha, ele em geral preferirá o último.

Na prática, um cientista tem bastante literatura na biblioteca - felizmente, é uma fração significativa da literatura científica. Como resultado, e da perspectiva do usuário, os artigos genuinamente de Acesso Aberto competem com outros documentos que, embora muito mais caros, aparecem em acesso aberto para o pesquisador dentro das bibliotecas (Guédon, 2004). Pode ser através das revistas armazenadas nas estantes ou de portais de acesso subsidiado, como o oferecido pela CAPES no Brasil.

CAPÍTULO 5 - Transição

A comunicação científica está atravessando um período de transição, onde as comunicações antes feitas em papel estão sendo substituídas ou suplementadas por mecanismos eletrônicos. Aqui não se trata apenas de uma mudança de meio físico onde se armazena e distribui informação, mas implica também em transformações nos aspectos econômicos, sociais e legais associados à Comunicação Científica. Resulta um quadro complexo de múltiplas transições interdependentes que se desenrolam no meio de permanentes contestações. Como ferramenta de auxílio ao entendimento dessas questões, vamos utilizar algumas idéias da obra clássica de Thomas Kuhn (1962) sobre as mudanças de paradigmas na Ciência.

5.1. A Estrutura das Revoluções Científicas

Na sua obra clássica, a *Estrutura das Revoluções Científicas* (1962), Kuhn chama de ciência normal à ciência cumulativa. Essa ciência sobrepõe-se às novidades fundamentais que são subversivas a suas crenças básicas. Nessa ciência, a pesquisa é uma tentativa devotada de forçar a natureza a se encaixar nos conceitos supridos pela educação formal. Para entendermos melhor o significado da ciência normal, é importante conhecer a caracterização de um paradigma.

A palavra paradigma aparece com mais de um significado na obra de Kuhn, porém, um dos mais importantes corresponde à matriz disciplinar. Isso quer dizer uma cultura comum em torno de um determinado assunto num certo momento da história; engloba todas as crenças, valores e técnicas compartilhadas pelos membros de uma certa comunidade. É uma espécie de consenso da comunidade científica sobre uma determinada disciplina. Porém, o caminho para estabelecê-lo é extraordinariamente árduo.

Um paradigma é essencial para o questionamento científico – nada pode ser interpretado na ausência de ao menos algum corpo implícito de crença teórica e metodológica, que permita a seleção, a avaliação e as críticas. O paradigma guia todo um grupo de pesquisadores. Ele recebe esse status porque foi o que obteve mais sucesso em relação aos seus competidores na solução de alguns problemas considerados importantes.

Inicialmente, um paradigma oferece a promessa de sucesso. A ciência normal consiste na atualização dessa promessa. Isso é alcançado através da extensão do conhecimento daqueles fatos que o paradigma exhibe como particularmente reveladores e também através da articulação do

paradigma. A pesquisa baseada em paradigmas é uma tentativa de forçar a natureza a entrar em caixas pré-construídas e relativamente inflexíveis que o paradigma oferece.

Talvez essas características sejam defeitos, porém essas restrições trazem o próprio desenvolvimento da Ciência. Por focar a atenção numa faixa pequena de problemas, o paradigma força o cientista a investigar alguma parte da natureza em detalhe e profundidade. E essa investigação aprofundada numa especialidade às vezes traz à tona anomalias capazes de colocar o paradigma em crise. Kuhn ressalta que uma crise ocorre quando surgem discrepâncias entre a teoria e o fato. É uma falha da teoria existente em resolver problemas definidos por ela mesma.

As novidades fundamentais de fato e de teoria trazem uma mudança de paradigma. Esta vem através do reconhecimento de que a natureza violou as expectativas do paradigma que governa a ciência normal. Ou então pode vir através de um fenômeno para o qual o paradigma não havia estabelecido um investigador. A área da anomalia é então explorada. São criados modelos e realizados exaustivos testes para tentar adequar a anomalia à teoria atual e assim pôr um fim à crise.

De acordo com Kuhn (1962), todas as crises terminam de uma das três formas:

- a ciência normal prova ser capaz de lidar resolver os problemas que provocaram a crise e tudo retorna ao normal.
- o problema persiste e se percebe que faltam as ferramentas para solucioná-lo, então os cientistas deixam a questão para as próximas gerações resolverem.
- um novo candidato a paradigma emerge e se inicia uma batalha para a sua aceitação.

O ciclo de uma revolução é mostrado na figura a seguir:

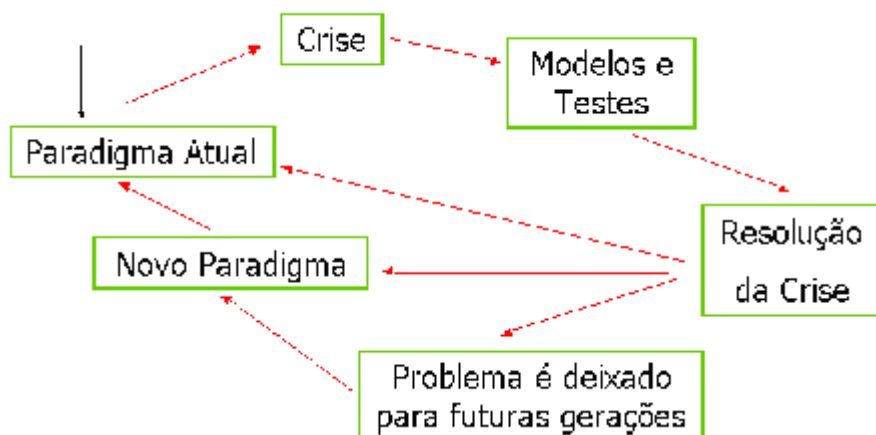


fig 4.1. Surge algum fato novo que mexe com conceitos fundamentais do paradigma, então a matriz disciplinar entra em crise. Aparecem novos modelos para explicar o fato. Estes são testados exaustivamente. Finalmente, chega-se a uma resolução pra crise que pode ser: o fato novo encaixa-se no paradigma atual e tudo volta ao normal; ou não é possível solucionar o problema e ele é deixado para as futuras gerações ou ainda um dos modelos competidores vence e emerge um novo paradigma (e então se dá a transição de paradigmas).

A fase de transição de paradigmas não é rápida e nem calma. Como Max Planck disse e Kuhn citou:

"a new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it."

A transição de um paradigma em crise para um novo, no qual uma nova tradição de ciência normal pode emergir, não é um processo cumulativo. É a reconstrução do campo a partir de novos fundamentos. A mudança de paradigma não apenas transforma uma teoria única, ela muda a forma como as palavras são definidas, o jeito como os cientistas encaram seus assuntos e, talvez, o mais importante, as questões que são consideradas válidas e as regras usadas para determinar a verdade de uma teoria particular. As novas teorias não são apenas extensões das novas teorias, mas novas visões de mundo.

Como vimos no capítulo quatro, os artigos científicos publicados em acesso aberto são livres e *online*. Para entendermos a natureza de uma possível transição do modelo vigente na publicação científica para o Acesso Aberto, é necessária uma breve reflexão sobre algumas características do meio *online*. Estas são transmitidas para a publicação de Acesso Aberto, por ela ser definida basicamente nesse ambiente. No entanto precisamos enfatizar que a referida transição é algo altamente complexo, com uma diversidade enorme de agentes cujos interesses são muitas vezes divergentes. Esse processo é pouco entendido até mesmo entre os acadêmicos. Não é possível prever como o processo irá se desenrolar e é difícil enxergá-lo de forma abrangente.

5.2. Rivalidade, Controle e Liberdade na Rede Mundial

Como mostrado no capítulo três, a tecnologia da imprensa trouxe imensos benefícios para o armazenamento, a distribuição e a preservação da informação. Ela proporcionou mecanismos para aumentar dramaticamente o acesso ao *commons* científico. Tal aumento é parceiro inquestionável do crescente desenvolvimento científico que a humanidade experimentou nos últimos 500 anos. No entanto, a Internet implementa um esquema de armazenamento e principalmente distribuição de informação enormemente mais eficiente do que a tecnologia da imprensa. Lessig diz algo ainda mais profundo. Ele afirma que o mundo digital está mais próximo do mundo das idéias do que do mundo físico.

Parafraseando Thomas Jefferson, Lessig (2001) escreveu:

No ciberespaço as coisas podem “livremente espalhar-se de um a outro ponto sobre o globo, para a instrução moral e mútua do homem, e melhoramento de sua condição”, porque nós (pelo menos originalmente) construímos o ciberespaço de tal forma que o seu conteúdo é “como fogo, expansível sobre todo o espaço, sem diminuir sua densidade em nenhum ponto, e como o ar o qual respiramos, no qual nos movemos, e o qual sustenta nossa existência física, incapaz de confinamento, ou apropriação exclusiva.”

O mundo digital está mais próximo das idéias do que das coisas, mas ainda não está bem lá. As idéias tem características não-rivais. Não é possível confiná-las uma vez que elas já foram expressadas para outras pessoas. Apesar de ser parcialmente rival, o meio digital apresenta algumas manifestações não-rivais. Por exemplo, muitos usuários podem compartilhar a leitura de um arquivo digital ao mesmo tempo.

A Rede apresenta a dicotomia mundo dos *bits* versus mundo dos átomos (Nicholas Negroponte). O *software* da Rede está descrito em *bits* (ou melhor, algoritmos codificados em *bits*). No entanto, a infra-estrutura é *hardware* (mundo físico). Apesar de haver escassez no mundo físico, a Rede como um todo, com cerca de 400 milhões de máquinas conectadas em 2006, apresenta uma abundância enorme no número de máquinas, na quantidade de ciclos computacionais, na quantidade de banda e no espaço de armazenamento. Além disso tudo, ainda proporciona uma infra-estrutura robusta e ampla para distribuição das informações (idéias) pelo mundo todo (entre as várias técnicas, o *peer to peer* destaca-se como uma tecnologia inovadora e capaz de distribuir também os custos de armazenamento e distribuição da informação). Mesmo Apesar de depender do mundo físico, a Rede apresenta pouca escassez. E essa escassez pode diminuir ainda mais conforme a capacidade dos computadores e das redes aumenta e seus preços diminuem. Por tudo isso, não é apropriado analisar a Rede como um recurso rival. Nem é uma boa prática aplicar as mesmas regras de controle necessárias no mundo físico.

A criatividade pode ser incentivada pela proteção que sistemas de controle estabelecem, mas também é severamente limitada pelas cercas que esses sistemas impõem. Por exemplo, um filme, uma música ou um livro que se produza só será publicado se a produtora, a gravadora ou a editora permitir. No entanto, na Rede, é como um *commons*. Eu posso colocar o que quiser, sem pedir a permissão pra ninguém (Lessig, 2001, p.115-116).

Esses sistemas de controle são necessários para gerenciar recursos que por serem físicos são,

conseqüentemente, rivais. No entanto, quando transformamos esses recursos em *bits*, não podemos mais tratá-los da mesma forma que antes. Muitas vezes, esses sistemas de controle são dispensáveis. De acordo com Benkler (2003), com menor controle, a Rede proporciona oportunidades maiores para o discurso livre, e conseqüentemente, para a democracia. Lessig (2001) destaca que a liberdade na Rede proporciona um ambiente rico para a inovação. No entanto, a arquitetura da Rede está mudando à medida que os governos e o comércio aumentam a habilidade de controlar comportamentos no ciberespaço.

Com isso, não queremos dizer que o controle é maléfico ou desnecessário, mas sim que o equilíbrio entre a liberdade e o controle devem ser repensados à luz das novas tecnologias digitais. Não podemos simplesmente deixar que os vencedores do passado (Hollywood, as grandes produtoras de filmes, gravadoras e editoras) ditem as regras do futuro e transfiram o controle para o meio digital.

As economic policy, letting yesterday's winners dictate the terms of economic competition tomorrow is disastrous. As social policy, missing an opportunity to enrich democracy, freedom, and equality in our society, while maintaining or even enhancing our productivity, is unforgivable. (Benkler, 2001)

5.3. Transição para o Acesso Aberto

Finalmente focando a comunicação científica no ambiente digital, observamos que tendências contraditórias, mas paralelas, onde existe acesso sem precedentes à informação científica e, concomitantemente, maiores restrições no acesso através da legislação de propriedade intelectual, técnicas de Digital Management Rights, abuso de patentes, falta de preservação e monopólios de grandes editoras de revistas científicas, tudo isso indica a profunda e complexa transição. Há muitos agentes envolvidos com interesses distintos. São acadêmicos, governos, leitores esporádicos, bibliotecários e os donos das editoras. Todos eles estão repensando seus negócios e interesses e os adaptando para o ciberespaço.

Em meio a essas disputas, surgiu o Acesso Aberto. Ele foi impulsionado fortemente pelos altos preços das assinaturas de revistas científicas, a redução no orçamento das bibliotecas acadêmicas e as facilidades que o meio digital proporciona para a publicação e a distribuição da informação.

Algumas iniciativas do Acesso Aberto, como o *self-archiving*, a publicação de revistas científicas de acesso aberto, o modelo do CiteSeer, entre outras, provaram que é possível aumentar a liberdade e eficiência na publicação acadêmica através da Internet. Ela tem oferecido a acadêmicos, autores, leitores, bibliotecários e universidades formas alternativas de se expor, organizar e obter a informação científica. Permitiu o resgate de parte do controle que está excessivamente concentrado

nas mãos das grandes editoras científicas. Melhor ainda, criou um ambiente onde as idéias da academia podem fluir livremente, associarem-se e se transformarem em novas idéias, de forma mais rápida e intensa do que no ambiente da tecnologia da imprensa.

Num âmbito geral, a transição para o Acesso Aberto está se dando de forma bastante lenta. Kuhn, na sua supracitada obra *The Structure of Scientific Revolutions*, destaca a comum resistência para a mudança de paradigmas. Isso é bastante nítido na transição atual. Há uma considerável resistência dentro da própria comunidade científica. Percebe-se pouca informação e motivação a respeito do Acesso Aberto. Ele enfrenta diversas críticas. Há uma parcela razoável de cientistas extremamente satisfeitos com as estabelecidas hierarquias que o sistema atual sustenta. O movimento de A. A. também se defronta com resistência de robustas forças econômicas de editoras comerciais, que lucram demasiadamente com o sistema vigente.

Por outro lado, há muitos motores propulsores a favor do Acesso Aberto. Inúmeras declarações de apoio têm sido feitas por comunidades de cientistas, renomadas instituições de pesquisa, associações de bibliotecas, agências financiadoras de pesquisa, nações e até mesmo organizações internacionais como a UNESCO. Segundo Kuhn, conforme a crise se desenvolve, as questões se tornam mais conhecidas e mais atenção é devotada por autoridades eminentes.

De acordo com o mesmo historiador da Ciência, outra marca importante de uma crise é a proliferação de modelos e testes dos mesmos. Na atual transição, variados modelos estão sendo implementados por centenas de repositórios distribuídos pelo globo, milhares de revistas científicas e provedores de serviços em geral, quer sejam de acesso aberto ou não. Os modelos de sustentação econômica são também bastante variados, bem como as políticas de inclusão de recursos no *commons* científico.

Do lado do Acesso Aberto, o volume de revistas científicas está aumentando. Hoje o Directory of Open Access Journals reúne quase duas mil. A prática do *self-archiving* também está se disseminando. As editoras científicas também têm feito pequenas concessões para este lado, permitindo a prática do *self-archiving* ou oferecendo publicação de acesso aberto, como a Springer, ou então disponibilizando parte do seu material de forma aberta para atrair novos consumidores, como a HighBeam.

Nessa transição, a implementação do *commons* científico em *bits* tem demonstrado alta eficiência. Permite redução de custos e oferece uma capacidade de disseminação jamais vistas, tornando possível a transmissão instantânea da informação científica para qualquer ponto do globo. Essa grande vantagem do meio digital tem sido explorada tanto pelo modelo de publicação A. A. quanto pelos demais. É interessante notar que a mesma tecnologia que permite a livre e intensa circulação das idéias da academia possibilita também a circulação controlada por autenticação nos portais das revistas eletrônicas. As editoras científicas, em geral, estão simplesmente transportando o modelo

antigo para o meio digital.

A respeito da resistência das associações de editoras científicas diante desse novo modelo e das dificuldades para motivar os cientistas com as novidades do Acesso Aberto, é interessante refletir sobre essas questões à luz do seguinte trecho de O Príncipe (1505), de Maquiavel:

*Innovation makes enemies
of all those who prospered under the old regime,
and only lukewarm support is forthcoming
from those who would prosper under the new.
Their support is indifferent
partly from fear and
partly because they are generally incredulous,
never really trusting new things
unless they have tested them by experience.*

Segundo Kuhn, o significado das crises consiste exatamente no fato de que indicam que é chegada a ocasião para renovar os instrumentos. A crise no paradigma atual da publicação científica indica que os instrumentos atuais (revistas científicas tradicionais) precisam ser renovados. De acordo com Kuhn: a transição de um paradigma em crise para um novo não é um processo cumulativo; é uma reconstrução do campo a partir de novos fundamentos. Se essa crise realmente terminar em mudança de paradigma (para publicação de Acesso Aberto somente), podemos esperar a emergência de um paradigma radicalmente novo para a Comunicação Científica e uma reconstrução de regras e práticas da comunidade científica.

Apesar de não podermos prever o desenrolar da crise, é possível enxergar alguns caminhos de resolução dessa crise através da teoria de Kuhn (ver figura 4.1). A primeira possibilidade é chegarmos à conclusão de que a forma mais efetiva de publicação e preservação da informação científica é a revista em papel. No caso dessa possibilidade remota, tudo (ou quase) voltará a ser como antes. Há a possibilidade de nossa geração não conseguir resolver a crise e os problemas serem deixados para a próxima geração. No entanto, apostamos mais em uma das seguintes resoluções da crise:

- 1) atingirmos um equilíbrio entre a publicação de acesso aberto e fechado – os dois modelos conviveriam “pacificamente”, cooperando para possibilitar o melhor aproveitamento do ambiente digital pelas comunidades interessadas;
- 2) o acesso aberto se torna o modelo vigente de publicação científica;
- 3) o acesso fechado domina a publicação científica.

Caso 1:

Pode ser que haja espaço para os dois tipos de publicação científica. Na realidade, hoje convivemos com os dois, no entanto, eles estão em disputa. Talvez exista um lugar bem definido para cada um desses tipos e uma forma de gerenciamento que não prejudique os interesses do cientista.

Caso 2:

O Acesso Aberto triunfa sobre o fechado. Toda a publicação científica torna-se disponível *online*. Serviços sofisticados de indexação e cruzamento de referência e dados podem ser construídos de forma a abranger toda a produção científica, aumentando abruptamente a acessibilidade dos resultados científicos. A riqueza informacional da Ciência seria multiplicada enormemente. A comunicação entre cientistas aumentaria e se tornaria mais rápida e eficiente. Conseqüentemente, as transformações científicas poderiam se dar de forma mais veloz. A Medicina poderia descobrir mais rapidamente curas e tratamentos para muitas doenças.

Caso 3:

O acesso fechado vence. Toda a publicação científica concentra-se na mão de editoras de acesso fechado. Assim o acesso torna-se restritivo e controlado. Conseqüentemente, o desenvolvimento da Ciência caminha em ritmo mais lento, devido à necessidade da satisfação de interesses financeiros.

5.4. Quatro Aspectos da Transição

A possível transição para o Acesso Aberto é tão ampla e abrangente que a forma mais correta de analisá-la parece ser em termos de um repensamento amplo de toda a instituição milenar da comunicação científica, ponto de vista que adotamos nesta dissertação. Seja qual for a resolução da crise, há quatro aspectos da transição de paradigmas dentro da Comunicação Científica. Observe a figura abaixo:

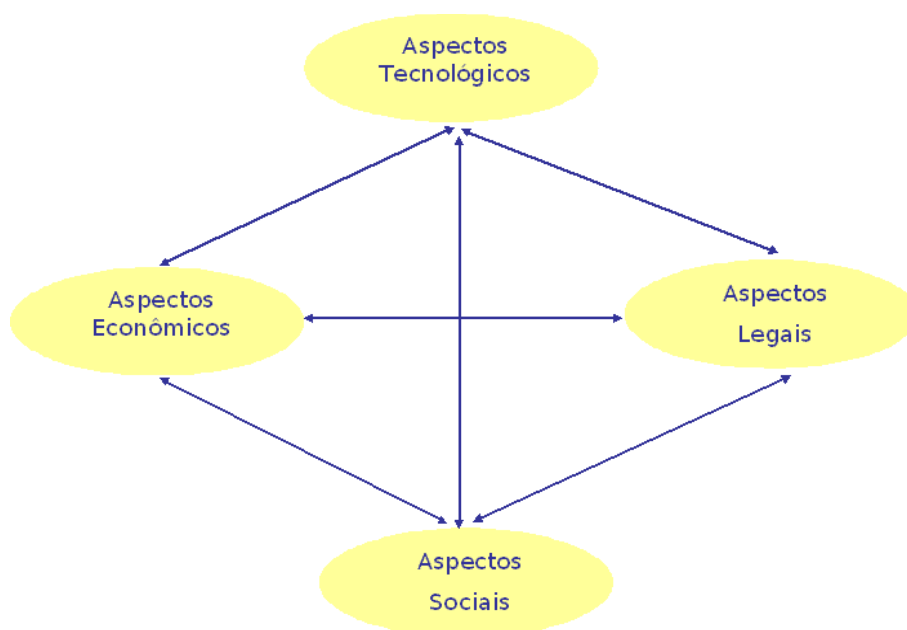


Fig. 4.2. Os aspectos tecnológicos englobam principalmente as tecnologias da imprensa, papel, computadores, redes, a Internet e o protocolo OAI-PMH. Os aspectos legais incorporam as leis, com ênfase no direito autoral. Os aspectos econômicos constituem as questões de mercado, incluindo as de custos e preços da publicação científica. Os aspectos sociais referem-se a como a sociedade científica, as editoras, os governos, os bibliotecários estão promovendo, incentivando ou desincentivando a publicação de Acesso Aberto, bem como a renegociação de seus interesses.

Esses quatro aspectos estão continuamente interagindo entre si e reformulando o acesso ao *commons* científico. As setas na figura indicam que um influencia o outro. Os aspectos econômicos, por exemplo, são fortemente influenciados pela tecnologia disponível e também pela forma como utilizamos essa tecnologia. Se a tecnologia é a imprensa, temos a predominância de recursos rivais. A economia aqui se volta para o custo de impressão, papel, serviços editoriais e correio. Se a tecnologia é a Internet, então há basicamente custos para *hosting*, *hardware*, *software*, conectividade e serviços editoriais.

As leis também precisam ser adaptadas quando surgem novas tecnologias. Em especial, as leis de direito autoral precisam se adaptar à realidade da Internet. Por outro lado, a economia também interage com o aspecto legal. Por exemplo, algumas grandes editoras científicas requerem via contrato a transferência dos direitos exclusivos de distribuição que, a princípio são dos autores. E como exemplo de interação entre o legal e o social, o direito autoral foi criado como uma forma de incentivo à criação para beneficiar a sociedade. O econômico também influencia o legal. Por exemplo, são criados *lobbies* junto a congressos nacionais para que as leis estejam de acordo com os interesses financeiros dos economicamente poderosos.

Esses são apenas alguns exemplos da interação entre os quatro aspectos que caracterizamos. Todos esses aspectos são muito importantes na transição. No entanto, os aspectos sociais têm um papel determinante para que o Acesso Aberto se torne um novo paradigma para a Comunicação Científica. Os verdadeiros agentes do sistema de comunicação científica estão na comunidade científica. É verdade também que há no sistema tomadores de decisões, como legisladores, presidentes de editoras científicas, bibliotecários, etc. No entanto, os legisladores não têm um conhecimento profundo sobre comunicação entre cientistas e são influenciáveis por *lobbies*. As editoras científicas, principalmente as comerciais, têm pouco interesse no Acesso Aberto por estarem lucrando bastante com o sistema atual. Portanto, uma transição para o modelo de Acesso Aberto somente ocorrerá se a comunidade científica e os bibliotecários impulsionarem o movimento. Até agora, embora a intensidade dos movimentos de A. A. esteja em crescimento, ainda não conseguiu criar uma sinergia típica de outros processos baseados no *commons-based peer production* (seção 2.4) como o *software* livre ou a Wikipedia. Somos da opinião de que as causas dessa lentidão merecem uma atenção acadêmica e tentativas de compreendê-la e explicá-la devem ser elaboradas no futuro.

5. 5. Aspectos Legais

É interessante notar que tecnologia que proporciona meios de construção de serviços como Pubmed Central, arXiv e CiteSeer, essa mesma tecnologia é utilizada para algumas editoras científicas reforçarem o seu controle sobre o material que publicam. A Internet está questionando o outrora estabelecido balanço entre os direitos públicos e privados em relação ao direito autoral. De um lado extremo, podemos ter um mundo onde “todos os direitos reservados” é a norma. Do outro lado, temos uma visão de anarquia, onde os criadores desfrutam uma ampla liberdade, mas o sistema é vulnerável à exploração.

No momento, há alguns estudiosos das leis bastante preocupados com a renegociação de direitos. De acordo com Boyle, Benkler e Lessig as mudanças recentes na lei do direito autoral (incluindo o Digital Millenium Copyright Act) tendem a limitar as oportunidades de inovação na Rede e também restringir a liberdade de expressão. Boyle se refere a essas mudanças como um segundo movimento de apropriação. Agora a apropriação do *commons* intangível das mentes.

Para ajudar a promover um equilíbrio entre a liberdade e o controle na criação artística, Lessig, Boyle, entre outros, criaram as licenças Creative Commons¹, que foram inicialmente inspiradas na GNU General Public License (GPL). No entanto são um tanto diferentes, principalmente porque não foram arquitetadas para *software*, mas sim para música, filmes, vídeos, fotografias, material

acadêmico, *courseware*, etc. O principal objetivo é tornar o acesso a esses materiais mais barato e fácil. As licenças também são legíveis por robôs, possibilitando assim mais inovação.

Recentemente foi criada o Science Commons, um projeto-ramo do Creative Commons e que tem por objetivo remover barreiras desnecessárias ao fluxo de conhecimento científico e informação técnica. O projeto encoraja, através de licenças padronizadas, “um commons da Ciência construído sobre acordos privados, não debaixo deles”². Recentemente a Science Commons (veja nota de rodapé 2), iniciou um programa chamado de Open Access Law. Ele visa montar uma sistemática onde o autor retém os direitos de pelo menos uma das licenças da Creative Commons na disponibilização do seu trabalho. Quase 30 revistas renomadas da área de Direito já aderiram ao programa.

Atualmente, podemos dizer que estamos sob os acordos privados. É prática comum as editoras científicas requererem dos autores a transferência dos direitos exclusivos de distribuição. Willinsky (2005, p. 41) afirma que no momento em que recebem o formulário de transferência, os autores deveriam parar e refletir um pouco nas formas alternativas de publicação que hoje existem.

A economia atual das editoras científicas parece estar fazendo algo mais do que protegendo o direito dos autores para beneficiá-los a partir da publicação do seu trabalho intelectual e criativo original. O direito autoral tem sido freqüentemente utilizado para proteger o direito da editora de cobrar preços elevados pelas assinaturas de revistas científicas (Willinsky, 2005, p. 47).

As editoras abriram uma exceção para o crescente controle dos materiais que publicam. A concessão refere-se a permitir que os autores façam o *self-archiving* dos seus trabalhos em repositórios de Acesso Aberto. Entre 131 editoras listadas numa pesquisa recente, 93% delas já permitem alguma forma de *self-archiving* (veja Romeo³).

1- <http://creativecommons.org/>

2- <http://sciencecommons.org/>

3- <http://romeo.eprints.org/stats.php>

5. 6. Aspectos Econômicos

Já existem muitas revistas com publicação de Acesso Aberto e também há algumas evidências de que esse modelo é economicamente sustentável. A primeira evidência é a de que os custos da publicação de Acesso Aberto são significativamente menores que os custos da publicação tradicional. A segunda razão é a de que o dinheiro para cobrir esses custos reduzidos pode ser encontrado através do redirecionamento do dinheiro que hoje paga os altos custos da publicação tradicional. Frequentemente, os custos mais elevados na publicação digital de revistas científicas é o *peer review* e também a disponibilização permanente dos artigos.

As revistas de Acesso Aberto economizam significativamente por sua publicação ser digital e também por dispensar os custos de manutenção das assinaturas e do aparato que restringe o acesso aos usuários autorizados. Também existem programas de livre distribuição para automatizar as operações da revista *online*, exceto pelo exercício do julgamento por editores e *referees*.

Qualquer modelo que não cobre uma taxa dos leitores pode ser aplicado aqui. Atualmente, há discussões a respeito de taxas para os autores. Além disso, há muitas possíveis instituições financiadoras como as universidades, laboratórios que empregam pesquisadores, entidades governamentais, amigos da causa do Acesso Aberto, etc.

A Biblioteca Pública de Ciência (Public Library of Science, ou PLoS) lançou a sua própria revista de Acesso Aberto, a PLoS Biology. Com essa revista, espera-se demonstrar a viabilidade do Acesso Aberto como uma alternativa à publicação comercial. A situação atual é descrita como um estado de transição em relação à disseminação da pesquisa científica e à economia de novas formas de publicação. Como nas reações químicas, as transições podem precisar de grande quantidade de “energia” inicial, e menos depois que a reação “decola”.

Foram recebidas grandes somas de dinheiro de doações e de instituições de pesquisa. O PLoS está utilizando várias estratégias de apoio financeiro à revista PLoS Biology, incluindo taxas para os autores, assinaturas tradicionais à versão impressa, redução de custos no gerenciamento dos manuscritos, etc.

O Prof. Henry Hagedorn (da Universidade do Estado de Arizona) acredita que a biblioteca acadêmica deveria ser responsável pela publicação e ele não acredita na publicação comercial. Como um acadêmico, ele confia mais na biblioteca para desempenhar o papel da publicação, pois afirma que as bibliotecas: a) entendem o valor do acesso gratuito como parte do processo de crescimento acadêmico do indivíduo e do grupo de pesquisa; b) preocupam-se com a preservação e manutenção da coleção bibliotecária; c) são mais direcionadas pela sua missão de serviço do que pela demanda de mercado.

Ele descreve sua experiência com a Revista da Ciência do Inseto, uma publicação de Acesso Aberto feita em colaboração com a biblioteca da Universidade do Estado de Arizona. O primeiro volume da revista foi publicado em janeiro de 2001. Ela tem uma taxa de aceitação de artigo de cinquenta por cento, tem peer review, é gratuita aos autores e leitores, e os autores retêm o direito autoral. O Prof. Hagedorn espera que depois que a revista estiver bem conhecida e conceituada, os autores tenham que pagar uma taxa para publicação dos seus trabalhos.

A revista científica Journal of Clinical Investigation (JCI) provê acesso gratuito e irrestrito ao conteúdo de seus artigos desde que ele se tornou on-line em 1996. Os custos do Acesso Aberto estão sendo supridos através de taxas aplicadas às submissões e às assinaturas da versão impressa, que tem sido publicada desde 1924.

O Acesso Aberto não só trouxe à revista uma oportunidade de expor seu material a um público enorme como também mostrou à revista que essa é uma política viável. Entretanto, o problema principal em se manter um modelo gratuito *on-line* é que ele adiciona despesas à revista, que são consideráveis para ao JCI. Em 2003, o custo total de publicação foi em torno de 2.580.000 dólares. O JCI publicou 308 artigos de pesquisa em 2003 (em torno de 90% das submissões foram rejeitadas).

5.7. Aspectos Sociais

Recentemente, muitas declarações de apoio ao movimento de Acesso Aberto foram feitas. No dia 26 de agosto de 2004, 25 cientistas ganhadores do Prêmio Nobel enviaram uma carta¹ ao congresso americano, pedindo apoio à iniciativa do NIH (National Institute of Health) em requerer que todos os cientistas que recebem financiamento desse instituto forneçam seus artigos ao repositório de acesso aberto PubMed Central. A PLoS também promoveu a circulação de uma carta aberta que coletou assinatura de mais de 34.000 cientistas a favor do Acesso Aberto.

A BOAI é uma iniciativa que surgiu num encontro do Open Society Institute (OSI) em dezembro de 2001. O propósito desse encontro foi a aceleração do processo de tornar os artigos de pesquisa gratuitamente acessíveis na Internet. A iniciativa tem sido aderida por participantes do encontro em Budapeste e um crescente número de indivíduos e organizações do mundo todo que representam pesquisadores, universidades, laboratórios, bibliotecas, fundações, revistas, editores e iniciativas do Acesso Aberto. A BOAI quer o acesso aberto da literatura que passou pela análise de peer review. A única exceção é para os *preprints*.

1- Veja a carta em <https://mx2.arl.org/Lists/SPARC-OAForum/Message/991.html>.

O AmSci de um fórum de apoio ao Acesso Aberto, criado em 1998, onde são discutidas questões sobre preservação, comunicação e publicação da literatura científica, direito autoral, modelos economicamente viáveis e como implantar o modelo de Acesso Aberto e self-archiving. O moderador é o Professor Dr. Stevan Harnad cujo discurso dentro do fórum enfatiza que “o self-archiving é o meio mais rápido e eficiente para atingir o acesso aberto de toda a literatura científica hoje publicada”.

Além das iniciativas aqui mencionadas, existem muitas outras. Num âmbito geral, o movimento está crescendo (ainda que lentamente) e já envolve milhares de pesquisadores pelo mundo. Porém, para uma mudança de paradigma realmente acontecer na Comunicação Científica, é necessário que praticamente toda a comunidade científica (de todas as áreas e sub-áreas da Ciência) estejam envolvidas e motivadas com o modelo de Acesso Aberto. Infelizmente existe um consenso de que o progresso é muito lento nessa direção.

Para tentar vitalizar o progresso, existe um número crescente de congressos que focalizam as questões do A. A. (Suber, 2006). Existe também um esforço crescente, mas que enfrenta muitas dificuldades levantadas por algumas editoras e sociedades científicas, para que todos os trabalhos científicos sejam disponibilizados em A. A. em repositórios institucionais. Para tanto, buscam-se implementações de *mandating open access* para a publicação subsidiada por instituições acadêmicas, de órgãos de financiamento e até mesmo de países, como é o caso da Índia.

É muito importante que os cientistas conheçam essas questões, pois podem afetar o seu dia-a-dia como pesquisador, consumidor e autor de informação. É necessário que se manifestem e participem ativamente da renegociação para defender os interesses da comunidade científica e da sociedade em geral. Se não o fizerem, as editoras comerciais continuarão a ditar as regras, reforçarão o controle e algumas barreiras ao acesso do conhecimento fundamental.

Nós estamos no meio de uma transformação organizacional, econômica e tecnológica que permite renegociar os termos da liberdade, justiça e produtividade na sociedade da informação. Como nós viveremos nesse novo ambiente dependerá grandemente das escolhas que fizermos nas próximas duas décadas. Para sermos capazes de entendê-las e as escolhermos bem, nós precisamos compreender que elas são escolhas sociais e políticas. (Benkler, 2003).

CAPÍTULO 6 – Aspectos Tecnológicos: Open Archives Initiative

Visando a facilitação da disseminação eficiente de conteúdo, a Open Archives Initiative¹ (OAI) desenvolve e promove padrões de interoperabilidade entre repositórios digitais. Assim, ela criou o protocolo OAI-PMH, um meio simples de se compartilhar metadados entre servidores distribuídos.

O movimento OAI tem as suas raízes no esforço de ampliar o acesso a repositórios de artigos científicos como um meio de aumentar a disponibilidade da comunicação científica. Entretanto, os principais padrões e ferramentas tecnológicas desenvolvidos são independentes do tipo de conteúdo oferecido. Hoje a OAI não se restringe apenas a repositórios científicos. No registro oficial de provedores de dados² da OAI, há também repositórios de músicas, URLs e vídeos.

Um repositório OAI, ou seja, que implementa o protocolo OAI não é necessariamente um repositório de acesso aberto. Isso porque o protocolo insiste na abertura dos metadados e não necessariamente na abertura do texto completo dos trabalhos presentes nos repositórios. O contrário também é verdade, isto é, existem repositórios de acesso aberto que não implementam o protocolo OAI-PMH.

Para apoiar a iniciativa OAI, criamos o Clube OAI². Trata-se de um portal, hospedado na Incubadora da FAPESP. Esse portal explica conceitos básicos relacionados a OAI como o funcionamento do protocolo e algumas definições técnicas.

6.1. Breve Histórico

No final dos anos 90 já existiam na Internet alguns repositórios de artigos científicos, como o arXiv (na época limitava-se à Física de Altas Energias), o CogPrints⁴ da área de Neurociência, o Repec⁵ com conteúdo acadêmico da área de Economia, a biblioteca digital de teses e dissertações NDLTD⁶ e o NCSTRL⁷, repositório digital de relatórios técnicos de Ciência da Computação. No entanto cada repositório implementava um protocolo próprio. Havia muitas dificuldades para se compartilhar metadados entre esses servidores.

1- <http://www.openarchives.org>

2- <http://clube-oai.incubadora.fapesp.br>

3- <http://www.openarchives.org/Register/BrowseSites>

4- <http://cogprints.org>

5- <http://repec.org>

6- <http://www.ndltd.org/index.en.html>

7- <http://www.ncstrl.org/>

Foram identificadas duas questões que deviam estar atrapalhando o impacto dos repositórios de *eprints*: as diferentes interfaces que o usuário encontrava tornavam o processo de busca mais difícil e não havia um método automático de compartilhamento de dados. Em julho de 1999, Paul Ginsparg, Rick Luce e Herbert Van de Sompel, todos do Laboratório de Los Alamos, convidaram um grupo restrito de especialistas para um encontro em Santa Fé, Novo México, em outubro desse mesmo ano.

O objetivo do encontro foi discutir questões de interoperabilidade, promover um protótipo de serviço de biblioteca digital baseado nos repositórios de *eprints* já existentes e estabelecer um fórum para continuidade do trabalho. Nesse encontro, eles propuseram a criação de um serviço universal de repositórios de literatura científica, o UPS (Universal Preprint Service), o qual coletaria dados de servidores de *preprint* distribuídos. Esse serviço deu origem ao ARC¹, um agregador de mais de 7 milhões de registros de metadados desenvolvido pela Old Dominion University.

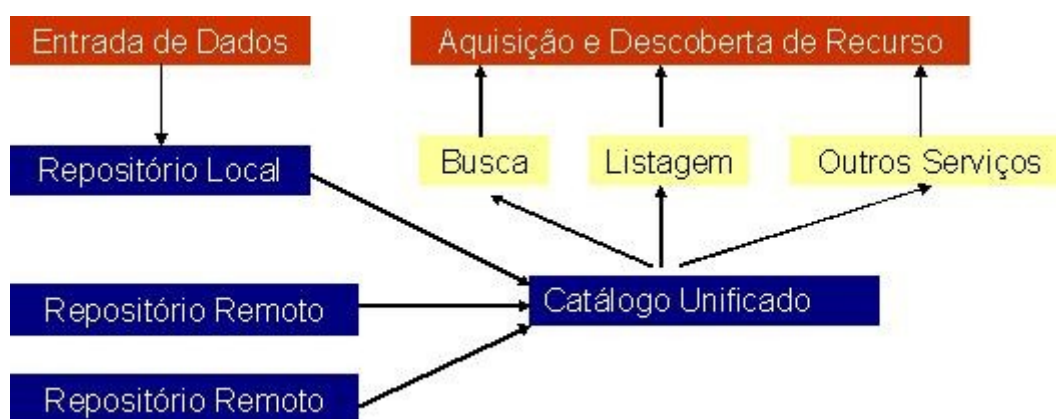


Fig. 5.1. A figura acima mostra um dos principais benefícios do protocolo OAI: a facilitação para agregação de metadados de bases locais e remotas e o subsequente fornecimento desses dados unificados através de serviços úteis ao usuário.

6.2. Algumas definições

Provavelmente a maior conquista da OAI foi a elaboração do protocolo OAI-PMH. Para entendermos seu funcionamento, precisamos conhecer o significado de alguns conceitos:

- Metadados

De forma literal, metadados significa dados sobre dados. Um exemplo comum é um cartão de catálogo de uma biblioteca, que contém dados sobre os conteúdos e localização de um livro. Campos comuns de metadados incluem o título, a fonte ou autor, a descrição de como o recurso pode ser acessado, licença para uso, onde e em que ano foi publicado.

1- <http://arc.cs.odu.edu/>

De forma mais geral, trata-se de informação estruturada sobre recursos (digitais e não-digitais). Os metadados podem ser utilizados para viabilizar uma ampla série de operações nos recursos. No contexto da OAI, as operações mais comuns são a de descoberta e aquisição dos recursos. E os recursos podem ser, por exemplo, imagens, livros, obras de arte, músicas, URLs, artigos científicos, entre muitos outros. O formato de metadados mais amplamente utilizado pela OAI é o **Dublin Core**.¹

Title	Contributor
Creator	Date
Subject	Type
Source	Language
Description	Format
Relation	Coverage
Rights	Identifier
Publisher	

Tabela 5.1. Elementos Dublin Core.

De forma resumida, algumas recomendações para implementação dos campos do Dublin Core que são obrigatórios no protocolo OAI seguem a seguir (Powell, Day e Cliff, 2004):

- *title* – título do artigo.
- *creator* – autor do artigo.
- *subject* – o tópico a que se refere o artigo (assunto) ou palavras-chaves.
- *description* – um sumário do artigo, normalmente em forma de *abstract*.
- *publisher* – o nome da editora ou da instituição do autor.
- *date* – a última data de modificação do artigo e/ou a data de disponibilização do artigo no repositório.
- *type* – tipo de artigo (livro, capítulo, artigo de conferência, *preprint*, *postprint*, etc).
- *identifier* – identificador único para o artigo.
- *language* – língua em que o artigo foi escrito.
- *relation* – URL(s) para o texto completo.

1- <http://dublincore.org>

- **Provedores de Dados**

Os provedores de dados¹ são sistemas que utilizam o protocolo OAI-PMH como um meio de expor metadados. Esses provedores também podem (mas não necessariamente) oferecer acesso aberto a textos completos e a outros recursos. Por exemplo, no Diálogo Científico² há um servidor de artigos digitais implementado através do *software* EPrints³, da Universidade de Southampton. Outro exemplo é o repositório Open Video⁴, que contém imagens digitalizadas e expõe os metadados desses arquivos via protocolo OAI-PMH.

O repositório eScholarship⁵, financiado pela Biblioteca Digital da Califórnia, é implementado por ferramentas do Berkeley Electronic Press. A Universidade da Califórnia (UC) licencia essas ferramentas de publicação para as suas unidades. Essa infra-estrutura possibilita criação rápida a baixos custos, gerenciamento e disseminação de periódicos, artigos de trabalho, artigos de discussão e outras formas eletrônicas de comunicação entre cientistas da UC.

Alguns dos maiores e mais importantes provedores de dados são os já mencionados arXiv, Pubmed Central, CiteSeer e o CERN Document Server.

- ***Harvesters***

São programas que utilizam a interface oferecida pelo protocolo OAI-PMH para coletar metadados. Eles importam os metadados dos provedores de dados e oferecem assim, a viabilidade de se construir novos serviços sobre esses dados coletados.

- **Provedores de Serviços**

Os provedores de serviços utilizam os metadados coletados pelos *harvesters* via protocolo OAI-PMH como base para construção de novos serviços. Um exemplo interessante é o Citebase⁶ – uma ferramenta de cienciometria automática que explora e demonstra o potencial dos metadados e do protocolo OAI-PMH. Apresenta uma medida de impacto de citação de artigos e serviços de busca para dados coletados de repositórios grandes de *eprints*.

1- Para testar se a sua implementação de provedor de dados OAI está funcionando adequadamente, você pode utilizar a ferramenta Repository Explorer.

2- <http://dici.ibict.br/>

3- <http://www.eprints.org>

4- <http://www.open-video.org>

5- <http://repositories.cdlib.org/escholarship/>

6- <http://citebase.eprints.org/cgi-bin/search>

O OAster¹ é outro provedor de serviços bastante famoso – coleta e expõe, através do protocolo, mais de 6,5 milhões de registros de cerca de 600 repositórios de material relacionado a atividades acadêmicas. Trata-se de um projeto do Serviço de Produção da Biblioteca Digital da Universidade de Michigan (OAster, 2006).

6. 3. Protocolo OAI-PMH

O protocolo OAI-PMH² é um mecanismo para transferência de dados entre repositórios digitais. É uma interface que um servidor de rede pode empregar para que os metadados de objetos residentes no servidor estejam disponíveis para aplicações externas que desejem coletar esses dados. Essa interface tem duas propriedades: interoperabilidade e extensibilidade.

A interoperabilidade decorre da obrigatoriedade embutida no protocolo para implementação do padrão Dublin Core. Assim todos os repositórios que utilizam o protocolo OAI podem trocar metadados. Já a extensibilidade advém da oportunidade de se criar ou utilizar também padrões de metadados diferentes do Dublin Core. Descrições específicas para uma comunidade ou especificidade de metadados para satisfazer necessidades especiais podem ser criadas ou adaptadas de forma a funcionarem com o protocolo OAI.

O protocolo utiliza um ferramental de transações *http* muito simples, baseadas em *request-response*. A resposta deve ser no formato XML. Um *harvester* pode selecionar os metadados que deseja coletar de forma a pegar somente os novos ou os que foram modificados desde a última interação de coleta com o repositório. Pode também restringir os metadados que deseja coletar ao indicar qual é conjunto de dados de interesse (alguns servidores separam os metadados por assunto ou por ano de publicação).

Há seis verbos de requisição do protocolo. São eles:

- **Identify**: traz as principais informações do repositório como nome, identificador, *e-mail* do administrador, informações sobre a propriedade intelectual dos dados contidos no repositório, etc.
- **ListMetadataFormats**: lista os formatos de metadados implementados pelo sistema do repositório; é obrigatória a implementação de pelo menos o formato Dublin Core.

1- <http://oaister.umdl.umich.edu/o/oaister/>

2-<http://www.openarchives.org/OAI/openarchivesprotocol.html>

- **ListSets**: traz a árvore de assuntos que classificam os documentos no repositório ou outro conjunto de classificação.
- **ListIdentifiers**: lista todos identificadores de registros do repositório.
- **ListRecords**: lista todos registros do repositório.
- **GetRecord**: dado um identificador, lista o registro correspondente.

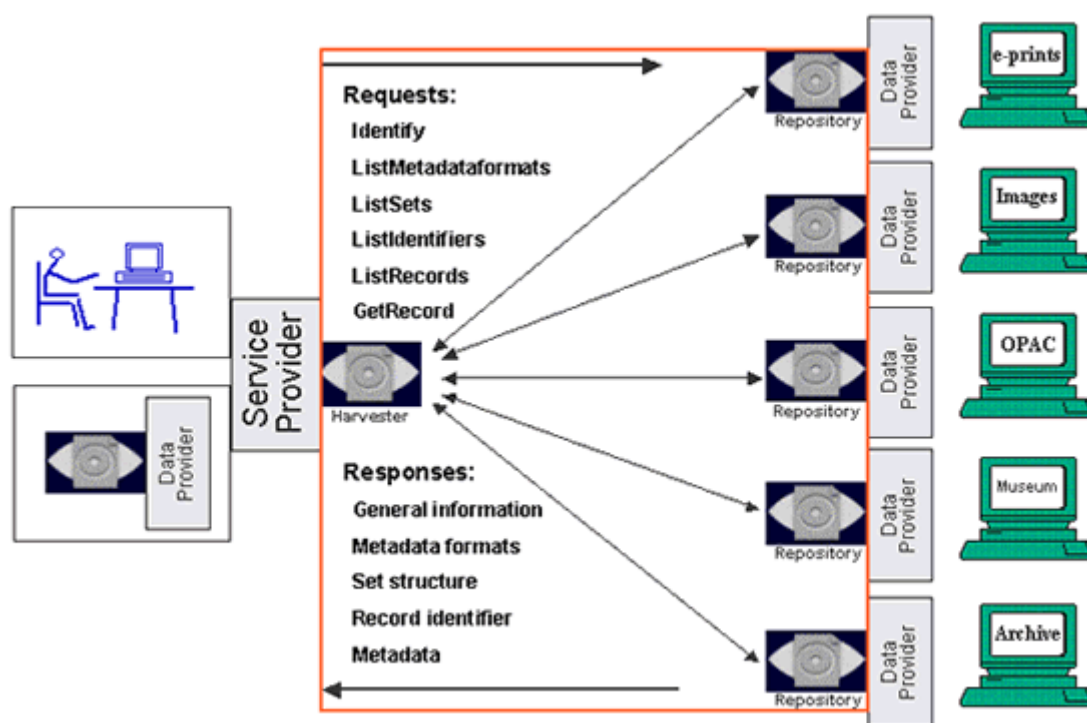


Fig. 5.2. Fonte: tutorial do OAForum (em <http://www.oaforum.org/tutorial/english/intro.htm>).

```
<record>
  <header>
    <identifier>oai:cogprints.soton.ac.uk:2</identifier>
    <timestamp>1998-03-13</timestamp>
    <setSpec>7374617475733D707562</setSpec>
    <setSpec>7375626A656374733D7068696C:7068696C2D6D696E64</setSpec>
    <setSpec>7375626A656374733D6E6575726F:6E6575726F2D70687973696F</setSpec>
    <setSpec>7375626A656374733D70737963:636F672D707379</setSpec></header>
  <metadata>
    <oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:dc="http://purl.org/dc/elements/1.1/">
      <dc:title>The Limits of Neuropsychological Models of Consciousness</dc:title>
      <dc:creator>Velmans, Max</dc:creator>
```

```

<dc:subject>Cognitive Psychology</dc:subject>
<dc:subject>Neurophysiology</dc:subject>
<dc:subject>Philosophy of Mind</dc:subject>
<dc:description>...</dc:description>
<dc:date>1995-01-01</dc:date>
<dc:type>Journal (Paginated)</dc:type>
<dc:identifier>http://cogprints.org/2/</dc:identifier>
<dc:identifier>Velmans, Max (1995) The Limits of Neuropsychological Models of
Consciousness. Behavioral and Brain Sciences 18(4):pp.&#xA0;702-703.</dc:identifier>
<dc:format>html http://cogprints.org/2/00/velman10.html</dc:format>
</oai_dc:dc>
</metadata>
</record>

```

Fig. 5.3. Resposta a http://cogprints.org/perl/oai2?verb=GetRecord&identifier=oai:cogprints.soton.ac.uk:2&metadataPrefix=oai_dc. Esse comando serve para listar o metadado do registro de identificador `oai:cogprints.soton.ac.uk:2`, utilizando o padrão de metadados `oai_dc` (Dublin Core adaptado no protocolo OAI). O conteúdo do campo `description` (abstract do documento) foi omitido por questão de espaço.

Para entendermos a funcionalidade do protocolo é igualmente importante enfatizar o que ele não faz. A autenticação e o gerenciamento de acesso não estão implementados no protocolo. Os repositórios podem por conta própria implementar o controle de acesso no código do programa. Independente do controle de acesso sobre os metadados, o servidor inclui um apontador (como URL) para o objeto descrito pelos metadados; o controle de acesso pode ou não existir no nível do recurso. Por exemplo, o texto completo de um artigo pode ou não estar disponível para *download* (para humanos ou para robôs). O protocolo também não inclui um ponto importante: como os *harvesters* identificarão os repositórios que eles desejam coletar nem provêem informação para ajudar a determinar quando a coleta deve ocorrer ou com que frequência (Lynch, 2001).

6. 4. O Protocolo na prática

A principal finalidade do protocolo é facilitar o compartilhamento de metadados, ou seja, a coleta e agregação desses dados para criação de novos serviços, onde o mais comum é a busca federada. Existem alguns passos a serem seguidos para uma coleta eficiente e também há várias dificuldades em cada uma dessas etapas, as quais são:

1. Selecionar os repositórios dos quais se deseja coletar os metadados e descobrir as respectivas URLs bases de coleta (ou seja, as URLs a partir das quais é possível aplicar os verbos de requisição).

2. Escolher uma ferramenta apropriada de *harvesting*.
3. Iniciar a coleta, fornecendo uma lista com as URLs bases dos repositórios.
4. Normalizar os metadados coletados de forma a se tornarem úteis para implementação do serviço pretendido.

A seguir, detalhamos as dificuldades em cada uma dessas fases.

6.4.1. Seleção

Decidir quais os repositórios cujos metadados participarão do seu serviço é uma tarefa manual e exaustiva. A partir da resposta do verbo Identify, é possível descobrir algumas informações sobre a natureza dos dados contidos no repositório, porém, frequentemente esses dados são insuficientes para se saber se é de interesse a coleta desses metadados. Para tal decisão, é necessário navegar nos recursos do repositório e também checar a qualidade da implementação dos metadados.

Desenvolver ferramentas computacionais que automatizem essa etapa é uma proposta bastante complexa, principalmente porque o tema e a qualidade do repositório são critérios subjetivos e sujeitos a debate. Os diversos repositórios não só implementam o padrão Dublin Core de forma diferente, mas também apresentam qualidade variando desde o impecável até o decepcionante.

6.4.2. Ferramentas para *Harvesting*

Um serviço bastante útil para testar se o repositório de interesse está respondendo adequadamente aos verbos OAI é o Repository Explorer¹. Basta fornecer a ele a URL base do repositório e preencher alguns campos relativos aos verbos de requisição. Depois de testados, é necessário fazer a coleta dos metadados. Para tal fase, é imprescindível escolher uma ferramenta adequada as suas necessidades.

No portal de ferramentas² da OAI, podemos encontrar alguns *harvesters* de licença livre. Suas funcionalidades e desempenho variam, por isso devem ser bem testados de acordo com seus requisitos. Uma decisão mal feita aqui pode atrapalhar bastante a implementação do serviço pretendido.

1- <http://re.cs.uct.ac.za/>

2- <http://www.openarchives.org/tools/index.html>

6.4.3. Coleta de Metadados

Apesar da interface OAI-PMH oferecer um conjunto de verbos que facilitam a coleta automática dos metadados, ainda surgem alguns problemas como:

1. Mudança de URL base do repositório - para coletar via protocolo OAI, é preciso fornecer ao programa *harvester* uma URL a partir da qual ele aplicará os verbos de requisição; porém, às vezes, a URL do repositório muda e é difícil descobrir a nova URL.
2. Servidor fora do ar – quando se utiliza *softwares* livres como EPrints ou DSpace, os repositórios institucionais são baratos e fáceis de se montar; por isso, algumas vezes são implementados por instituições pobres ou pouco comprometidas, que oferecem pouca manutenção e sustentação do serviço.
3. Falhas na implementação da interface OAI - alguns repositórios não respondem corretamente aos seis verbos do protocolo.
4. Interrupção na coleta por sobrecarga do servidor - por receberem muitas requisições de *harvesters*, alguns servidores caem e interrompem a transferência dos metadados antes de ser finalizada.

Esses pequenos empecilhos tornam-se mais incômodos à medida que cresce o número de repositórios dos quais fazemos a coleta, tornado a etapa de *harvesting* um tanto trabalhosa e exaustiva.

6.4.4. Heterogeneidade nos metadados e necessidade de normalização

Devido à heterogeneidade na implementação do padrão Dublin Core, para serem úteis, os metadados precisam ser padronizados de forma a se adequarem ao serviço. Como os metadados nesse padrão não pretendem descrever apenas *eprints*, a definição de alguns campos apresentada no *site* oficial do Dublin Core é um tanto vaga. Por exemplo, o campo *relation* - de acordo com a definição dada no Dublin Core - é "uma referência a um recurso relacionado".

Na realidade, o protocolo OAI não implementa exatamente o Dublin Core, mas sim o *oai_dc*. Trata-se de um subconjunto de campos do Dublin Core. A principal diferença entre o Dublin Core e o *oai_dc* é que no último, a implementação de alguns campos se torna obrigatória. São eles: *title*, *creator*, *subject*, *description*, *publisher*, *date*, *type*, *identifier*, *relation*, *language*.

Existem algumas recomendações de implementação desses campos para *eprints*, feitas por Andy Powell, Michael Day e Peter Cliff (2004). Por exemplo, recomenda-se colocar a URL para o texto

completo no campo *relation*. As recomendações são bastante boas do ponto de vista das informações necessárias para que, tanto um robô quanto um ser humano, sejam capazes de extrair dados importantes para avaliação do documento. Porém, elas são muito pouco seguidas, mesmo pelos maiores e mais importantes repositórios de artigos científicos de Acesso Aberto.

Verificamos que, em repositórios famosos como arXiv, Pubmed Central, CiteSeer, Cogprints, o repositório institucional do Departamento de Computação de Southampton, entre outros, os campos *title* e *description* são freqüentemente bem implementados, seguindo as recomendações de Powell, Day e Cliff. Eles contêm respectivamente o título e o *abstract* do artigo. No entanto, há muita confusão no campo *date*, que pode se referir à data de publicação na revista, à disponibilização do artigo no repositório digital ou até mesmo à data de criação do documento.

Além desses quatro campos importantes mencionados no parágrafo anterior, para a criação de uma biblioteca temática, os campos *subject* e *set* (assunto e conjunto) assumem um papel especial para ajudar na seleção e separação dos artigos por assunto. Porém, surgem muitas dificuldades devido à ausência de padrões compartilhados entre os repositórios. Os desenvolvedores da biblioteca digital NSDL (National Science Digital Library), financiada pela NSF (National Science Foundation), comentam seus problemas com os campos *set*:

The need for sets and the usage thereof differ greatly among providers. Small-collection providers find them unnecessary and confusing. Their utility for managing the demands of accessing metadata from larger collections is compromised by their inconsistent use.

(...) Most semantic mismatches come from two aspects of the protocol: identifiers and sets. People are confused by the concepts and use them differently. The OAI needs to have very clear descriptions of these concepts, standards of good practice and, as far as possible, validation tests (Arms et al., 2002).

Podemos observar formas de utilização do campo *set* bastante diferentes ao observar o repositório institucional de Southampton, o Cogprints e o DSpace do MIT. O repositório de Southampton, por exemplo, implementa o difundido padrão de classificação *Library of Congress*. O Cogprints utiliza um esquema especializado nos ramos das suas disciplinas (Ciências Cognitivas e Ciência da Computação). Já o DSpace hospedado pelo MIT utiliza nos campos *subject* as palavras-chaves que o autor escolheu para o documento. Somente nestes três exemplos temos três esquemas completamente diferentes.

Nas suas últimas versões, o *software* EPrints tem colocado como incorporado o padrão de classificação *Library of Congress*. Trata-se de um dos melhores e mais famosos. Porém, para muitas áreas especializadas da Ciência, essa árvore não é suficientemente ramificada. Ela abrange

apenas os assuntos mais gerais. Para a disciplina Ciência da Computação, por exemplo, ela não é tão detalhada quanto ao necessário.

Essas deficiências podem ser diminuídas através de ferramentas computacionais sofisticadas. Uma possível solução seria, a partir do texto completo, encontrar automaticamente as palavras-chaves do artigo e utilizá-las para inferir o assunto. Ainda não existe uma ferramenta desse tipo entre as ferramentas de apoio da OAI.

Sobre a qualidade dos metadados e as dificuldades de se implementar um agregador, os desenvolvedores do NSDL comentam:

Our experience matches that of other organizations harvesting metadata via the OAI protocol: quality among metadata records is extremely variable requiring a non-trivial amount of human effort to make it usable. The ultimate practicality of metadata harvesting lies in our ability to improve automatically the quality of the metadata (Arms et al., 2002).

6.5. Melhorias no Protocolo

Da forma como foi desenhado o protocolo, é possível o desenvolvimento de padrões mais adequados às necessidades das variadas comunidades. Esse é um processo social, onde se determinam práticas descritivas, de estruturação e de classificação. Ainda está muito lento e é caro o desenvolvimento desses padrões. Também é cedo para saber quais ganharão mais adeptos por períodos longos de tempo (Lynch, 2001).

Além do progresso nos metadados descritivos, o que é associado à atividade intelectual humana, também é possível a criação de ferramentas automáticas para se aperfeiçoar a qualidade dos metadados. A partir de uma página HTML de descrição do recurso ou da leitura automática do próprio recurso, é possível descobrir propriedades e características do mesmo e embuti-las nos metadados.

CAPÍTULO 7 – A Evolução do Acesso Aberto

Alguns proponentes do Acesso Aberto afirmam que, Apesar de nos últimos anos o movimento ter crescido bastante tanto em volume de artigos, de revistas e de cientistas envolvidos, uma parcela bem reduzida da literatura científica tem sido publicada em acesso aberto. Harnad (2004) afirma que de 10% a 20% dos 2.5 milhões de artigos publicados anualmente estão em acesso aberto, quer seja através da publicação em revistas de acesso aberto ou através do depósito em repositórios digitais (*self-archiving*).

7.1. Crescimento dos Repositórios

O OAIster agrega mais de 6.5 milhões de registros de mais de 610 repositórios distribuídos pela Rede e compatíveis com o protocolo OAI-PMH. No entanto, nem todos esses registros são metadados de artigos científicos. Há muitos arquivos de imagens, áudio e *software* (OAIster, 2006). Há outros dois portais oferecendo informação estatística sobre o crescimento dos repositórios. São eles o Directory of Open Access Repositories¹ (DOAR) e o Registry of Open Access Repositories² (ROAR). O DOAR (2006) tem uma lista de 349 repositórios institucionais, temáticos ou sustentados por agências financiadoras de pesquisa (como NIH e Wellcome Trust, por exemplo). Tem controle de qualidade realizado por uma equipe de especialistas e classifica os repositórios por assunto. Nesse serviço também é possível visualizar os repositórios por país e por tipo de conteúdo (artigos, livros, capítulos, teses, *preprint*, etc).

O ROAR (2006) tem 635 repositórios de artigos eletrônicos (*preprints*, *postprints*, teses, etc). Apresenta listas de repositórios separadas por país e por tipo de *software* como EPrints, DSpace, Bepress, etc (veja apêndice C). Também lista por tipo de conteúdo e é capaz de ordenar sua lista pelo nome ou pelo total de registros do repositório. O ROAR (2006) aponta que o país com mais repositórios são os Estados Unidos com 176, seguido pela Inglaterra com 68, a Alemanha com 60 e o Brasil com 42. Dentre os repositórios do ROAR, há 196 utilizando o *software* GNU Eprints³, 131 implementando o DSpace³ e 43 rodando o bepress³. Segundo esse mesmo serviço, há 63 repositórios de teses, 314 com conteúdo de pesquisa institucional ou departamental, 63 de pesquisa inter-institucional e 66 periódicos eletrônicos.

1- <http://www.opendoar.org/doar?func=browse>

2- <http://archives.eprints.org>

3- para maiores detalhes sobre este *software*, veja apêndice C.

Os cinco maiores repositórios listados no ROAR são: o CiteSeer¹, com 703.654 artigos científicos em texto completo e abertos; a PubMed Central¹ com 487.064 registros; o arXiv¹ com 356.199 *eprints* em texto completo e abertos; o American Memory (Library of Congress)¹ com 231.533 registros e cerca de 30% em acesso aberto; e o Networked Digital Library of Theses and Dissertations¹ com 217.520 teses e dissertações com o texto completo aberto. Observe também que os quatro maiores repositórios são temáticos, ou seja, abordam temas específicos e não toda a publicação de uma instituição, como o fazem os repositórios institucionais. A seguir, apresentamos um gráfico de crescimento dos repositórios **institucionais** no ROAR:

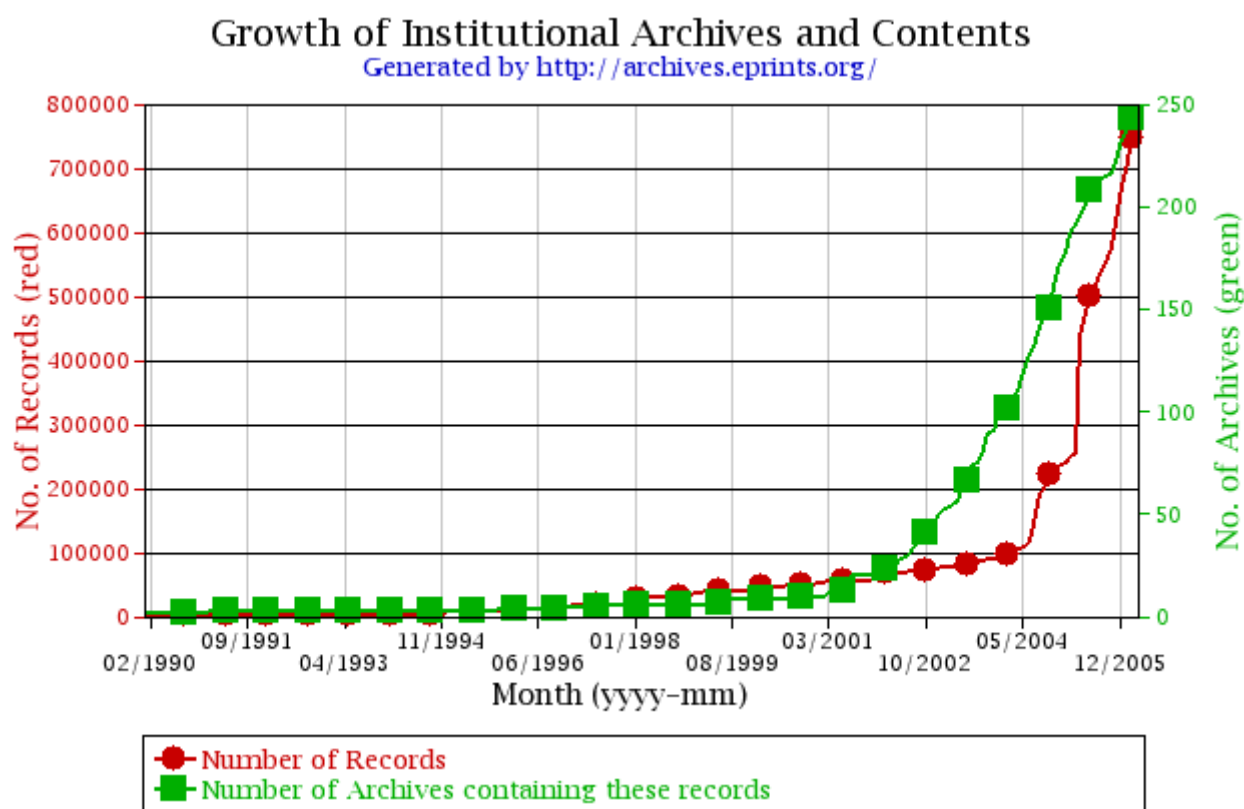


Figura 6.1. O gráfico aponta apenas para o crescimento dos repositórios institucionais. Eles são cerca de 250 atualmente, agregando aproximadamente 800 mil registros.

Apesar dos comentários no fórum eletrônico American Scientist Open Access sobre um crescimento lento do A. A., o gráfico acima mostra um crescimento exponencial. Esse tipo de crescimento é lento no início, mas depois dobra dentro de determinados intervalos de tempo. O número de repositórios institucionais tem dobrado mais ou menos a cada 1 ano e meio. Já a quantidade de registros parece estar crescendo de forma mais acelerada ainda.

1- para maiores detalhes sobre este repositório, veja apêndice C

Abaixo, apresentamos gráficos de crescimento de alguns repositórios importantes, já mencionados nesse trabalho. Selecionamos dois repositórios temáticos e dois institucionais.

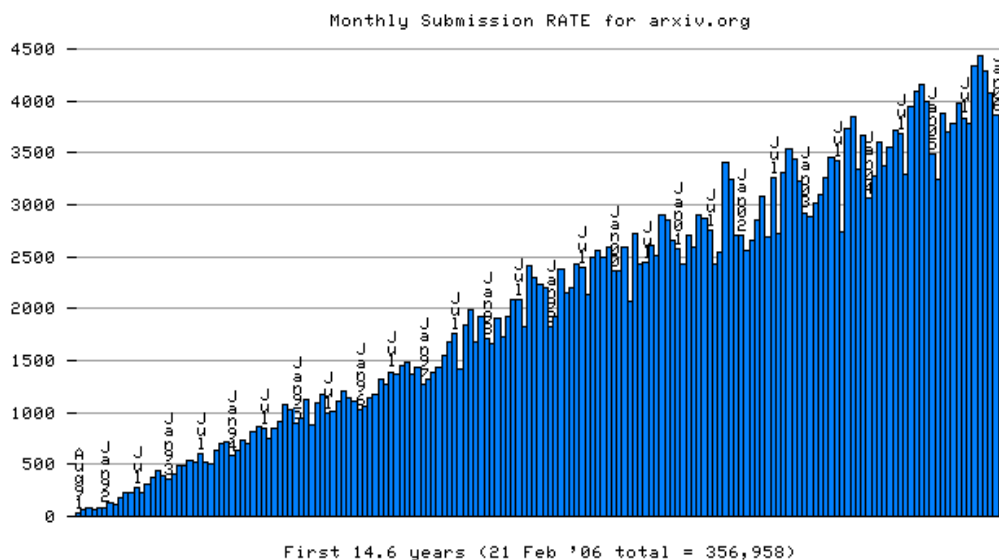


Figura 6.2. O gráfico do arXiv demonstra crescimento linear.

As três áreas mais amplas em volume de artigos no arXiv são Física de Altas Energias, Física da Matéria Condensada e Astrofísica e Matemática. No total, elas contêm aproximada e respectivamente 100 mil, 70 mil e 65 mil artigos em texto completo.

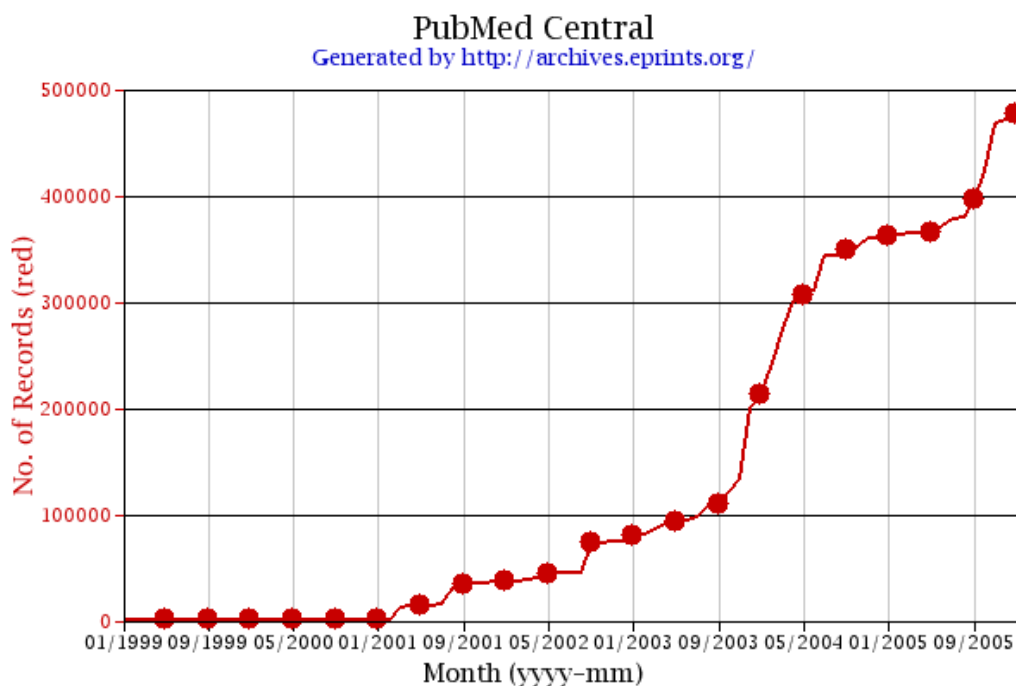


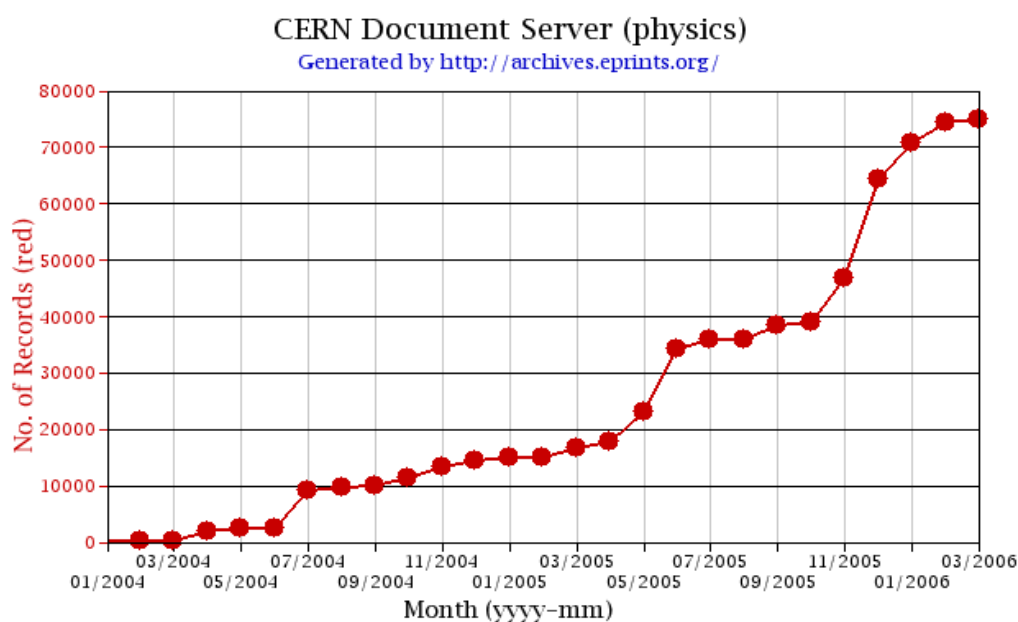
Figura 6.3. O PubMed mostra crescimento bastante acelerado.

O PubMed hoje conta com mais de 400 mil textos completos e mais de 200 revistas científicas na

área de Ciências da Vida e Biomédicas. Cerca de 85% das revistas dão acesso imediato ao texto completo. As demais oferecem o acesso ao texto completo com atraso de seis meses.

De acordo com Harnad (2005), os dois repositórios com crescimento mais rápido em termos de crescimento anual são o CERN Document Server e o ECS Eprints Service da Universidade de Southampton.

Figura 6.4. O CERN Document Server é o décimo maior repositório listado no ROAR em quantidade de registros OAI.



University of Southampton: Department of Electronics and Computer Science

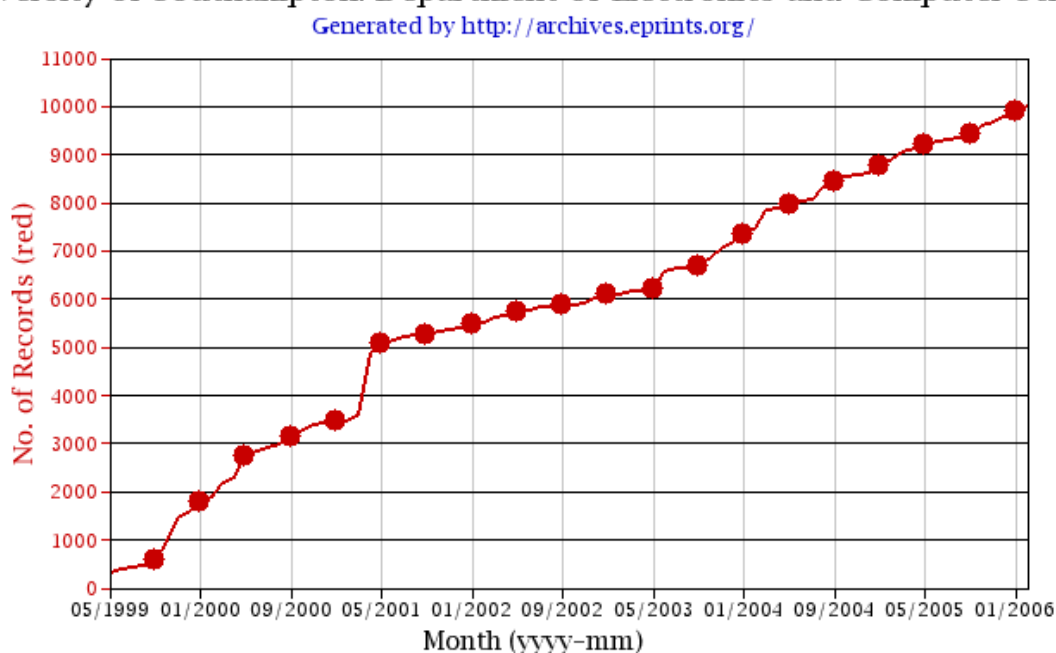


Fig. 6.5. Crescimento do ECS Eprints Service, praticante da política *mandating open access*.

O Cern Document Server agrega mais de 800 mil registros, incluindo 360 mil textos completos. Os tipos de documentos armazenados nesse serviço são artigos, preprints, livros, apresentações, teses, periódicos, relatórios técnicos, fotos, vídeos, etc. No entanto, ele exporta os metadados de uma parcela muito menor, como mostra o gráfico acima. Já o ECS Eprints Service reúne mais de 10 mil registros relacionadas à pesquisa do Departamento de Eletrônica e Ciência da Computação da Universidade de Southampton, onde 50% desses registros tem apontadores para texto completo aberto.

7.2. Revistas de Acesso Aberto

De acordo com Harnad (2004), havia cerca de 24 mil revistas com *peer review* no mundo e mais de 5% delas são de acesso aberto (1.200 revistas aproximadamente). Em 2006, o Directory of Open Access Journals (DOAJ) já reúne cerca de 2.000 revistas de acesso aberto (DOAJ, 2006).

O Institute for Scientific Information (ISI) estimou que cerca de 2% das 8.700 revistas científicas hoje indexadas no seu serviço tem acesso aberto. Esse fato mostra que realmente há revistas A. A. de alto nível de qualidade. Outro dado interessante é o fato de mais de 40% das revistas de acesso aberto indexadas pelo ISI serem da América do Sul e Central (McVeigh, 2004).

As duas mais importantes editoras exclusivamente de Acesso Aberto são a Biomed Central e a Public Library of Science (PLOS). Hoje a Biomed Central (2006) tem 140 revistas de acesso aberto e 504 instituições-membros, onde seus pesquisadores podem publicar diretamente sem precisar pagar pela taxa de publicação do artigo. As instituições-membros já pagam uma taxa anual por todos os seus pesquisadores afiliados. Já a PLOS (2006) tem 6 revistas de acesso aberto, cerca de 145 instituições afiliadas, e experimenta um crescimento muito rápido no fator de impacto da revista PLOS Biology (13.9 de acordo com o ISI).

7.3. Crescimento Social

Um bom indicador do crescimento da comunidade envolvida com o Acesso Aberto é através das conferências. Peter Suber (2006b) mantém uma lista de conferências passadas e futuras que tratam as questões do A. A. De acordo com sua lista, em 2002 aconteceram cerca de 170 conferências. Já em 2004, ocorreram cerca de 270 conferências. Outro indicativo do crescimento da comunidade envolvida é a quantidade de assinaturas individuais e institucionais coletadas pela BOAI, 3.968 e 338, respectivamente.

Quanto às políticas para prática do *self-archiving*, entre as 129 editoras listadas no Romeo e suas 8.698 revistas científicas:

- 69% das revistas não-AA permitem que os autores façam o depósito dos seus *postprints* (Eprints Statistics, 2006)
- 93% das revistas não-AA permitem que os autores façam o depósito dos seus *preprints*, *postprints* ou ambos (Eprints Statistics, 2006)

De acordo com a Wikipedia (2006c), são 14 as revistas científicas que fizeram a sua declaração de independência (ou seja, a resignação do corpo editorial para lançar uma outra revista de mesmo tema numa editora mais amigável).

O Institutional Self-Archiving Policy Registry (ISPR, 2006) é um serviço que tem contabilizado políticas de práticas de implementação do *self-archiving*. Há duas principais políticas de implementação, baseadas na Declaração de Berlim: uma só recomenda a prática do depósito do artigo num repositório institucional e a outra é de *mandating open access*. Ainda nesse serviço, 18 instituições já assinaram o ISPR e são 5 as instituições praticando o *mandating open access*. Recentemente, a agência financiadora de pesquisa na área médica Wellcome Trust adotou essa política para toda a publicação de resultados de pesquisa apoiados pela mesma.

CAPÍTULO 8 - Bibliotecas Digitais Temáticas no Acesso Aberto

Stevan Harnad, um dos principais proponentes do Acesso Aberto, afirma que o meio mais rápido e fácil de atingir o Acesso Aberto aos 2.5 milhões de artigos científicos publicados anualmente é através do *self-archiving*. Para tanto, existe hoje um movimento internacional muito forte para estabelecer, de forma universal, o *self-archiving* institucional "obrigatório". Essa prática, se bem sucedida, levará a um grande número de repositórios institucionais contendo uma riqueza incalculável de artigos científicos livremente acessíveis com o seu texto completo.

Porém, o conjunto desses repositórios institucionais não tem sido tão útil aos pesquisadores quanto o esperado. São muitos repositórios, de funcionalidades muito variadas, tornando muito trabalhosa a localização de alguma informação procurada. Por isso, alguns cientistas ainda preferem utilizar os serviços das grandes editoras e das sociedades científicas. Elas oferecem bibliotecas digitais com muitos serviços atraentes ao pesquisador e os ajudam a selecionar melhor o insumo procurado para a sua pesquisa científica.

Nesse capítulo, apresentamos um meio de tornar o conteúdo dos repositórios livremente acessíveis mais interessante ao pesquisador, e assim, motivá-lo mais quanto à prática do Acesso Aberto. Sugerimos a construção de bibliotecas temáticas a partir dos repositórios institucionais, utilizando o protocolo OAI-PMH. Descrevemos também os passos necessários para atingir tal objetivo, as funcionalidades importantes para essas bibliotecas, algumas ferramentas para auxiliar este trabalho, além das muitas dificuldades técnicas a serem vencidas.

8. 1. O Porquê das bibliotecas temáticas a partir dos repositórios institucionais

Até mesmo os proponentes do *self-archiving* admitem o seguinte:

*Institutional archives are being created, but need to be filled more quickly, by authors, with research journal papers. Attracting authors and their papers requires **evidence of services** that will improve the visibility and impact of their works.* (Hitchcok et al., 2003)

Atualmente, o Acesso Aberto está competindo com outros documentos que, embora muito caros, são subsidiados por bibliotecas de algumas instituições, e assim, estão facilmente acessíveis aos cientistas. Esses artigos são oferecidos pelos portais de grandes editoras (como o Science Direct e o site da Springer) e por sociedades de pesquisa (como ACM e AMS, por exemplo). Além de contar com excelentes artigos, esses portais implementam serviços de busca, índices de citação, navegação na bibliografia do artigo, etc.

Argumentando a favor de serviços como as bibliotecas temáticas das sociedades de pesquisa sobre

os repositórios institucionais, no fórum "AmSci Open Access" o Dr. Les Carr, outro importante proponente do A. A., disse o seguinte:

(...) I (as a researcher) am interested in "the literature" and not "just a specific paper". Publishers (and societies) have an opportunity to provide better services and facilities over a coherent collection of papers than a repository. That is one of the reasons why I stick with my society and its digital library!

Jean-Claude Guéron (2004), um dos mais importantes líderes da BOAI, diz o seguinte sobre aproveitar o conteúdo dos institucionais na criação de serviços mais atraentes:

*More fundamentally, we must find a way to move from **institutional to disciplinary and even specialty** repositories. This is important because it is easier to create the former than the latter and they are presently multiplying. (...) It becomes important therefore to move beyond simple, isolated, institutional repositories. This means aggregating and repacking the information that is contained in these institutional repositories along subject lines.*

Além disso, quando um cientista procurar insumo para sua pesquisa, ele não está interessado em conhecer o repositório ou a publicação de certa instituição, mas sim em encontrar artigos sobre um tema específico. Ainda, verificamos que a maior parte dos repositórios institucionais são muito pequenos (veja tabela abaixo). Por essas razões, argumentamos que é preciso tirar os institucionais do seu isolamento, complementá-los e agregar os seus conteúdos em bibliotecas temáticas.

Faixa qtd. registros	Qtd. reps.
Maiores de 100 mil	2
Entre 100 mil e 10 mil	6
Entre 10 mil e 1 mil	32
Entre 1 mil e 100	68
Menores que 100	112

Tabela 8.1. Distribuição de tamanho dos repositórios institucionais em agosto de 2005. Dados calculados automaticamente por robô que coletou informações do portal do ROAR nessa mesma data. A primeira coluna apresenta cinco faixas de distribuição da quantidade de registros e a segunda mostra quantos repositórios estão dentro da faixa.

8.2. Características Desejáveis nas Bibliotecas Temáticas

O modo de publicação do Acesso Aberto pode trazer muitas vantagens para as bibliotecas temáticas e o destacamos como um dos aspectos mais relevantes para extração do máximo aproveitamento dos serviços propostos nessas bibliotecas. Essa forma de publicação de informação pode facilitar a obtenção do texto completo, o cruzamento de referências, a navegação na bibliografia e a contagem de citações, uma vez que o documento é livremente acessível para qualquer um ou qualquer serviço. Além do Acesso Aberto, há serviços e funcionalidades que enriquecem grandemente o valor de uma biblioteca digital. Entre elas, destacamos:

- A. Mostrar quem são os artigos mais citados, seja através de uma lista dos artigos mais citados do portal ou pela ordenação de resultados de busca pelos artigos mais citados.
- B. Implementar busca no texto completo – recurso bastante importante visto que a busca no metadado oferece menos oportunidades do usuário encontrar o que procura.
- C. Crivo dos pares – algum critério de crivo dos pares é importante para se manter um controle de qualidade dos artigos. Por exemplo, o portal pode expor o nome da revista (responsável pelo *peer review* do artigo) ou mostrar o número de citações do artigo, ou ainda o número de *downloads*. Assim, o usuário pode ter um indicador da qualidade do documento.
- D. Mostrar quem são os autores mais citados num certo assunto ou área de pesquisa – isso pode ajudar a encontrar os pesquisadores e os pensamentos mais influentes em determinada área do conhecimento.
- E. Apresentar referência bibliográfica completa - fornecer o título do documento, os nomes dos autores, o nome da revista, a série, o volume e as páginas onde foi publicado. Essa informação é muito importante para se procurar o texto completo, caso somente o metadado esteja disponível de forma aberta ou se simplesmente se queira ter acesso ao artigo em papel.
- F. Possuir formação temática – como já comentado, isso auxilia o pesquisador a encontrar o documento desejado.

G. O portal também pode exibir os artigos semelhantes por co-citação ou similaridade de texto, como o CiteSeer o faz.

I. Identificação dos *hubs* e *surveys*, artigos com ampla bibliografia (Kleinberg, 1998).

O CiteSeer é um dos serviços mais próximos do que consideramos importante e útil numa biblioteca temática para um pesquisador. Ele agrega artigos e é uma biblioteca digital que se foca principalmente na literatura de Ciência da Computação e da Informação. Coleta artigos científicos encontrados na Rede, indexa seus textos completos, cruza referências, gera índices para facilitar a análise do artigo e redistribui mais de 700 mil textos completos.

O Citebase é um tanto semelhante ao CiteSeer. Porém, não disponibiliza ainda a busca no texto completo e abrange, além da Computação, outras áreas de pesquisa, principalmente algumas vertentes da física, a matemática, as Ciências Cognitivas, a Ciência da Informação e as Ciências Biomédicas. Outra diferença importante entre esses dois serviços é o fato do Citebase fazer a coleta a partir do protocolo OAI-PMH e o CiteSeer não.

Esses serviços podem ser ainda mais interessantes e úteis se a abrangência do tema for completa o suficiente para refletir parte das matrizes disciplinares. Usamos aqui o termo matriz disciplinar com o significado semelhante ao dado por Thomas Kuhn na obra clássica *The Structure of Scientific Revolutions* (Kuhn, 1962). Queremos dizer uma cultura comum em torno de um determinado assunto. Ela engloba todas as crenças, valores, técnicas, vocabulário específico e metodologias compartilhadas pelos membros da comunidade.

Os serviços oferecidos pelo CiteSeer já expõem numa interface de computador parte da matriz disciplinar na área de Computação. Ele mostra as sub-áreas, as autoridades, os artigos mais citados e as tendências, além de possibilitar a navegação nas idéias que produziram o artigo (através da navegação nas bibliografias). O CiteSeer não é completo na literatura do seu tema, mas quanto mais se aproximar da completude e inserir serviços que mostrem os principais aspectos dessa literatura, mais próximos estaremos de ter uma ferramenta que auxilie na identificação dos valores, técnicas e metodologias principais do paradigma da Ciência da Computação (Simon, 2002).

A seguir, definimos um conjunto mínimo de serviços que nossa biblioteca temática precisa oferecer:

1. Boa interface de busca no texto completo;
2. Navegação para trás e para frente na bibliografia, ou seja, navegação nas referências bibliográficas (trabalhos que influenciaram o artigo selecionado) e navegação nos artigos que citam o artigo selecionado (trabalhos influenciados pelo mesmo).
3. Mecanismo para implementação de algum crivo dos pares, seja através do nome da revista responsável pelo *peer review* ou pelo número de citações ou até mesmo pela

contagem de *downloads*; notamos que é essencial algum mecanismo que ajude o cientista a filtrar artigos de qualidade.

Além da implementação de mecanismos de avaliação de qualidade nos sistemas das bibliotecas temáticas, também é muito desejável que o *software* seja livre e facilmente adaptável para as diversas áreas de pesquisa e especialidades. A proliferação de bibliotecas de temas variados poderia ser facilitada, uma vez que o *software* fosse re-utilizável.

8.3. Como construir bibliotecas temáticas

A seguir, apresentamos uma síntese dos passos para se criar repositórios temáticos a partir dos institucionais via protocolo OAI-PMH. Depois, detalhamos as dificuldades de cada passo.

1. Coletar os metadados – para tanto, siga os passos descritos na seção 6.4.
2. Coletar o texto completo;
3. Indexar os metadados e o texto completo;
4. Disponibilizar uma boa interface para busca e navegação.

De forma muito resumida, esses são os passos. As dificuldades do passo 1 estão detalhadas na seção 6.4. Há muitos outros detalhes relacionados com os passos 2, 3 e 4, os quais detalhamos a seguir.

8.4. O Texto Completo

O texto completo tem um papel muito importante dentro de uma biblioteca digital. Se ele está disponível, é possível implementar mecanismos de busca muito mais eficientes do que a busca nos metadados. Ele é essencial para extrairmos informações como as referências bibliográficas e para fazer a análise de documentos similares. Ainda pode ser muito útil para extração do nome da editora e revista onde foi publicado, o volume e as páginas. Muitas vezes, essas informações não estão disponíveis nos metadados.

A respeito da importância do texto completo, os desenvolvedores do National Science Digital Library (NSDL)¹, financiada pela National Science Foundation (NSF), comentaram o seguinte:

Little can be done with a minimal Dublin Core record alone. To build effective NSDL services it is essential for service providers to have access to some combination of: (a) full content of items, (b) extensive item-level metadata, or (c) contextual information - such as annotations, citations or link analyses - from which additional characteristics about the item may be inferred. (Arms et al., 2002)

1- <http://nsdl.org/>

Apesar da importância do texto completo na implementação de serviços úteis, freqüentemente ele não está disponível para coleta por meio de robôs. Há diversas razões que podem impedir ou dificultar a obtenção do documento completo por pessoas ou por robôs, por exemplo:

1. Os metadados ou a página HTML fornecida no campo *identifier* (ver seção 6.2) podem estar em línguas pouco usuais.
2. Há uma heterogeneidade muito grande na forma de implementação dos metadados da OAI - por exemplo, alguns repositórios colocam o URL para o texto completo no campo *format*, outros em *source*, outros diretamente no *identifier* e outros em lugar nenhum.
3. Muitas vezes a URL para o texto completo não está nos metadados OAI do documento; nesse caso, é preciso ir até uma página HTML fornecida no campo *identifier* dos metadados, e a partir dessa página descobrir se o link para o texto completo está presente.
4. Pode não haver URL para o texto completo nem no metadado nem na página HTML de identificação do documento. O texto completo pode não estar *online*.
5. Quando finalmente obtemos a URL para o texto completo, ela pode ser um apontador quebrado. Também podemos não conseguir acessar o texto completo pelo servidor estar temporariamente fora do ar, entre outros problemas de conexão. Ainda, o acesso ao texto completo pode estar protegido por mecanismos de autenticação digital, restrito a quem paga pelo serviço do portal da revista.
6. O repositório disponibiliza o texto completo, mas restringe o seu uso para os usuários humanos. Robôs não são permitidos em bibliotecas como o arXiv e o Google Scholar.

8.4.1. Disponibilidade do Texto Completo em Repositórios Grandes ou Importantes

A tabela a seguir mostra a disponibilidade do texto completo aberto em alguns repositórios grandes ou famosos. Estes praticam uma boa política de implementação do protocolo, são estáveis e fornecem de alguma forma o acesso ao texto completo.

Tabela 8.3. Disponibilidade do Texto Completo nos Repositórios Famosos

Repositório	Regs	Textos Comp.	Tipo Acesso Txt
arXiv	323512	100%	B
CiteSeer	703660	100%	A
PubmedCentral	372255	100%	B-
BiomedCentral	18023	100%	B
CogPrints	2370	100%	A
ECS Eprints	9202	25%	A,C
e-Prints Soton	4950	20%	A,C,D
Cern	68794	35%	A,B,D
eScholarship	7532	96%	B
Scielo	50676	100%	B

A segunda coluna exibe a quantidade de registros do repositórios, a terceira mostra a quantidade de textos completos disponíveis em relação ao número de registros, a quarta coluna mostra o(s) tipo(s) de acesso ao texto completo. Dados extraídos dos portais desses serviços e das informações contidas nos metadados OAI expostos por esses repositórios em agosto de 2005.

A = URL para texto completo aberto já no metadado

B = URL para texto completo aberto a partir de uma página HTML de identificação do documento (dc:identifier)

- = significa que alguns artigos podem demorar cerca de 1 ano após a publicação para serem disponibilizados

C = URL para texto completo fechado a partir do metadado

D = URL para texto completo fechado a partir de uma página HTML de identificação do documento (dc:identifier)

Dentre os repositórios listados na tabela acima, os cinco primeiros (em vermelho) são temáticos, os quatro seguintes (em azul) são institucionais e o último é uma coleção de revistas científicas de Acesso Aberto latino-americanas e caribenhas. Apenas 2 repositórios dos repositórios listados expõem a URL de todos os seus textos completos no metadado e disponibilizam todos esses documentos em Acesso Aberto. Seis dos repositórios listados forçam o leitor ou o robô a irem até suas páginas HTML dentro dos seus portais para obterem o texto completo.

Observamos também pela tabela que a menor porcentagem de textos completos abertos está entre os institucionais Cern e os repositórios de Southampton (sexta e sétima linhas na tabela). Nas próximas tabelas, poderemos confirmar que o texto completo não está tão presente nos repositórios institucionais quanto está nos cinco primeiros repositórios listados (em vermelho).

Os repositórios institucionais ECS Eprints Service e o e-Prints Soton disponibilizam em Acesso Aberto o texto completo de apenas cerca de 20% dos seus registros. Esses dois repositórios são da Universidade de Southampton, universidade onde alguns dos principais proponentes do *self-archiving* trabalham.

8.4.2. Disponibilidade do Texto Completo nos Institucionais

Dentre os 220 repositórios institucionais listados no ROAR, escolhemos em agosto de 2005 os 8 maiores para uma primeira análise. Coletamos seus metadados e contamos a quantidade de textos completos (em pdf ou ps) presentes nesses repositórios grandes e apresentamos o resultado na tabela abaixo.

Tabela 8.4. Disponibilidade do Texto Completo nos Maiores Institucionais

Repositório	Qtd. Registros	Txt Completos*
UVA Pub ¹	68697	57%
Thessaloniki ²	24947	100%
Scientific Production ³	28799	0%
VTT Publications ⁴	10000	0%
WUR Publications ⁵	109000	-
Dspace at ANU ⁶	42532	2.8%
Cern ⁷	68794	35%
Dspace at MIT ⁸	ERRO/15886	

* artigo aberto em formato PDF ou PS encontrados por um robô simples

– não foi possível calcular devido à demora no *harvesting* dos metadados

A tabela acima mostra na primeira coluna o nome dos 8 repositórios com maior volume de registros; na segunda coluna, mostra a quantidade de registros desses repositórios; na terceira, a porcentagem de registros onde o texto completo do artigo está disponível em acesso aberto.

Os metadados do repositório VTT Publications não contém nenhuma URL, nem para página HTML de identificação do artigo, nem para texto completo. Esse repositório disponibiliza o texto completo de algumas publicações eletrônicas a partir do seu portal. O repositório Scientific Production só permite acesso ao texto completo via Virtual Protocol Network (VPN). O repositório Thessaloniki disponibiliza 100% de texto completo em Acesso Aberto a partir das páginas HTML de identificação do documento. E o DSPACE do MIT não está respondendo ao verbo *ListRecords* (ver seção 6.3).

Esses 8 repositórios representam 3% do número total de institucionais, porém contém cerca de 60% do total de registros dos repositórios institucionais. No entanto, observamos que há pouco texto completo aberto. Para compreendermos melhor o comportamento dos repositórios de tamanho médio, fizemos um estudo semelhante numa amostra aleatória de 6 repositórios institucionais (listados no ROAR) cuja quantidade de registros está entre 1 mil e 10 mil registros.

Ao calcular a média ponderada de textos completos, constatamos que 44% dos documentos contidos nos repositórios da amostra têm o texto completo aberto (em PS ou PDF).

1- <http://dare.uva.nl>

2- <http://cds.lib.auth.gr/>

3- <http://arno.uvt.nl>

4- <http://www.otalib.fi/vtt/jure/search.html>

5- <http://library.wur.nl>

6- <http://dspace.anu.edu.au>

7- <http://cdsware.cern.ch>

8- <http://dspace.mit.edu>

8.4.3. Licenças para a Coleta do Texto Completo

Para criarmos nossa biblioteca temática a partir de repositórios institucionais, precisamos de permissão para coletar o texto completo, principalmente por causa da questão do direito autoral. O estatuto do direito autoral dá ao seu detentor o direito exclusivo de reproduzir em cópias os trabalhos protegidos, sujeito apenas a algumas exceções.

A princípio, o detentor do direito autoral é o próprio autor. No caso dos artigos científicos, é prática comum as editoras requererem dos autores a transferência dos direitos de distribuição do documento. No entanto, segundo Stevan Harnad, mais de 90% das editoras dão o direito do *self-archiving* ao autor. Muitas vezes esse "direito" não está explicitado numa licença ou contrato, mas apenas numa página do portal da editora. Ademais, em vários casos não está claro se o texto completo disponibilizado através do *self-archiving* pode ser coletado por robôs e em caso positivo, se ele pode ou não ser redistribuído, como é feito, por exemplo, pelo CiteSeer.

Além disso, algumas editoras oferecem permissão para o depósito do *postprint* somente em repositórios pré-determinados. Nesse caso, não é possível redistribuir o texto completo através das nossas bibliotecas temáticas construídas a partir dos institucionais. No caso dos *preprints*, legalmente é preciso pedir autorização aos autores (detentores do direito autoral) para poder distribuí-los.

Segundo Litman (2001), como na lei americana não há exceções para usos relacionados a Internet, o detentor do direito autoral tem o direito legal de controlar ou cobrar dinheiro por cada aparição única de um trabalho na memória de um computador em qualquer lugar. Isso então, a rigor, implica na necessidade de obtenção de permissão do detentor do direito autoral para coletar e indexar o texto completo. Porém, os motores de busca da Internet não o fazem.

A seguir, temos duas tabelas que mostram se os repositórios dão permissão ou não para a coleta do texto completo. A primeira tabela lista repositórios importantes e famosos e a segunda representa os institucionais com maior volume de registros.

Tabela 8.5. Permissão para Coleta do Texto Completo por Robôs nos Repositórios Importantes¹

Repositório	Permite	Parcialmente	Com autorização	Não permite	Não diz
arXiv			X		
CiteSeer		X			
Pubmed Central		X			
Biomed Central	X				
Cogprints	X				
Cern				X	
eScholarship				X	
Scielo					X
e-Prints Soton					X
ECS Eprints					X

O arXiv requer autorização para poder coletar automaticamente seus textos completos. Já o CiteSeer diz que é preciso respeitar os direitos de cada trabalho individualmente. A Pubmed Central disponibiliza cerca de 3% do total de artigos para coleta por robôs. Já a Biomed Central permite a coleta automática do texto completo de todos os seus trabalhos. O Cogprints oferece essa permissão desde que a fonte, a referência bibliográfica e as notas de direito autoral sejam citadas. O Cern não permite a coleta automática do texto completo. O Scielo e os dois últimos repositórios (de Southampton) não se posicionaram sobre essa questão.

Tabela 8.6. Permissão para Coleta Automática do Texto Completo nos Repositórios Institucionais (informações retiradas do verbo Identify dos metadados desses repositórios, coletados em agosto de 2005).

Repositório	Permite	Não diz	Não permite
UVA Pub		X	
Scientific Production		X	
VTT Publications			X
WUR Publications		X	
Thessaloniki			X
Dspace em ANU		X	
Dspace no MIT		X	
Cern			X

1- O arxiv, Cern, CiteSeer e Pubmed Central foram considerados importantes pela sua imensa quantidade de registros. O Cogprints, o ECS Eprints Service e o e-Prints Soton são repositórios pioneiros. O Scielo é um importante portal brasileiro de revistas científicas. Já o eScholarship é um serviço da Universidade da Califórnia, o maior complexo universitário do mundo.

Acima estão listados os oito maiores repositórios institucionais, onde a soma de seus registros representa cerca de 60% do total de registros de todos os repositórios institucionais. Entre esses repositórios, podemos observar na tabela acima que nenhum deles permite explicitamente a coleta do texto completo por robôs. Três deles o proíbem e os outros cinco não se manifestam sobre essa questão.

Comumente, os direitos são expressos na resposta ao verbo *Identify*, nos campos *metadataPolicy* e *dataPolicy*. Também há repositórios que especificam os direitos pra cada registro individualmente, através do campo *rights*. Recentemente, foi criado pela equipe da OAI um documento¹ bastante bom sobre como expressar os direitos de propriedade intelectual. Elas mostram como é possível expressar direitos tanto no nível de repositório, conjunto ou registro, de forma que até robôs possam lê-las facilmente.

Sobre as dificuldades em se obter licenças úteis para a construção do NSDL, comenta-se o seguinte:

The NSDL brings together groups from different cultures. Libraries have a long tradition of sharing catalog records openly, while commercial indexing and abstracting services see their metadata records as an asset to which access is tightly controlled. Many of the collections that contribute to the NSDL fall between these two extremes. They want their metadata to be widely used for educational and non-commercial purposes, but they do not wish to give up all control of its distribution. Sometimes they themselves are restricted by the terms under which they obtained the metadata.

Currently, the NSDL Policy Committee is developing guidelines for redistribution of metadata. As an example, we wish to present the NSDL records in a way that search engines as Google or Ask Jeeves, which are commercial organizations, can index them. It is challenging to write a policy that permits such use, yet retains some control for the suppliers of the metadata (Arms et al., 2002).

Devido ao direito autoral e à falta de clareza sobre as licenças, observamos que é desafiador construir as bibliotecas temáticas. Mais desafiador ainda pode ser a redação de uma licença para redistribuição do texto completo que satisfaça todos os repositórios, os autores e as editoras, envolvidos direta ou indiretamente.

1- <http://www.openarchives.org/OAI/2.0/guidelines-rights.htm>

8.5. Resultados

Devido à falta de padrões de implementação do protocolo OAI-PMH e devido à fragilidade de algumas das ferramentas computacionais já difundidas para lidar com esse protocolo, surgem muitas dificuldades para a implementação de provedores de serviços na sistemática da OAI. Os serviços mais significativos existentes, como o Citebase, restringem-se à coleta de poucos repositórios muito bem implementados. Ao todo, existem apenas cerca de vinte provedores de serviços registrados no site da OAI. Assim, a OAI tem de certa forma falhado no objetivo de fornecer consolidação de repositórios científicos mundiais e também de fornecer uma interface robusta, flexível e eficiente entre o provedor de dados e o provedor de serviços.

O tema principal deste capítulo, bibliotecas digitais temáticas de acesso aberto, poderia perfeitamente ser implementado dentro do protocolo OAI-PMH, mas o seu sucesso dependeria preliminarmente da possibilidade de contornar as dificuldades técnicas que existem para o uso efetivo do protocolo e que foram amplamente apontadas no corpo deste trabalho.

Um outro fator muito importante para a consecução dos objetivos da nossa proposta é a ampla disponibilidade e a facilidade de implementação e de adaptação dos *softwares* necessários para a construção de bibliotecas digitais temáticas a partir dos repositórios institucionais. A melhor forma de obter estas características seria através do licenciamento dos programas como *software* livre. Infelizmente, não existem, ainda, programas nestas condições.

Quanto aos metadados, não é só a falta de padrões para o seu uso que tem dificultado a construção de uma biblioteca temática, mas também a falta de completude das informações registradas. Para a implementação dessas bibliotecas temáticas, é muito importante que existam informações como a referência bibliográfica completa da publicação do trabalho, quem realizou o *peer review* e, principalmente, a URL para o texto completo aberto, além da permissão para a coleta, a indexação e a possível redistribuição do texto completo do trabalho. Constatamos que poucos repositórios oferecem todas essas informações nos metadados, trazendo assim várias dificuldades para a construção das bibliotecas propostas neste. Lembramos, ainda, a importância de uma codificação padronizada destes metadados, a ponto deles poderem ser entendidos por robôs que serão, em última instância, os responsáveis pela coleta dos dados.

As dificuldades tecnológicas são desafiadoras, mas como o CiteSeer, Citebase, NSDL e Google Scholar estão demonstrando, é possível transpô-las, principalmente se estiver disponível o texto completo do documento. Sem ele, é bastante difícil adicionar, cruzar e/ou indexar informações para a criação de serviços úteis ao usuário.

As dificuldades legais também podem ser vencidas à medida que licenças mais claras e mais

completas sejam difundidas, abordando a coleta do texto completo por robôs, a indexação automática do texto completo e especificando, também, os direitos de redistribuição do texto completo. Mais uma vez, essas informações devem ser codificadas nos metadados de forma que sejam entendíveis por robôs.

Quando essas dificuldades tecnológicas e legais forem superadas, e quando o *software* para a construção das bibliotecas temáticas a partir dos repositórios institucionais estiver amplamente disponível, então poderemos ver surgir um ecossistema de bibliotecas temáticas criadas a partir dos repositórios institucionais. Esses novos serviços poderão trazer a motivação necessária para que mais cientistas promovam e apoiem o Acesso Aberto à literatura científica.

Por enquanto, constatamos que é difícil criar um serviço desse tipo, a não ser que os repositórios institucionais selecionados abordem todos um mesmo tema, tenham uma boa implementação do protocolo OAI, ofereçam algum mecanismo de credibilidade e dêem licença para a coleta automática e indexação do texto completo.

CONCLUSÃO

Concluimos que caracterizar a Comunicação Científica como um *commons* é um excelente exercício de análise dos problemas hoje enfrentados pela mesma, pois:

- chama a atenção para as decisões tomadas sobre os recursos que interessam a comunidades numerosas;
- promove um estudo interdisciplinar;
- incentiva a construir capital social entre comunidades de usuários hoje dispersas;
- elucida conexões entre as mudanças tecnológicas, a preservação da informação, a adaptação entre novos padrões de escolha e as interações entre participantes;
- ajuda focar alguns dilemas urgentes defrontados pela Comunicação Científica.
- fornece um foco principal e relevante para a análise das fortes transições pelas quais a Comunicação Científica atravessa no momento.

Percebemos que entender bem o *commons* científico requer um profundo conhecimento das suas regras, da sua comunidade e da natureza física dos seus recursos. O *commons* científico abrange um ecossistema inteiro que reflete interações complexas entre humanos e recursos. As análises precisam examinar como a tecnologia dos recursos e os valores e interesses da comunidade influenciam as ações e os comportamentos subsequentes, que levam a saídas sustentáveis ou menos satisfatórias. As consequências das decisões a serem tomadas precisam de cuidadosa avaliação sob diversos prismas. As questões de gerenciamento são complexas, conflitos estão se desenvolvendo e o desenrolar da transição é incerto.

A Comunicação Científica hoje não implementa completamente um *commons*. Há excessiva exclusão sócio-intelectual no processo de produção de recursos do *commons*. As dificuldades começam a partir das barreiras ao acesso impostas pelo modelo vigente, de acordo com interesses sócio-econômicos. Há excessivo domínio de valores financeiros sobre o *commons*. Não faz sentido a academia sustentar (ainda que parcialmente) os altos lucros das editoras comerciais. Isto dá-nos a impressão de que as universidades, talvez sem perceber, estão se desviando das suas missões primordiais de gerar, disseminar e preservar o conhecimento. A responsabilidade da disseminação e até mesmo da preservação parece ter sido terceirizada para as editoras, que nem sempre têm os mesmos objetivos e valores da comunidade científica. Essa divergência de interesses é um dos principais geradores de conflito na crise.

Em meio a esses problemas, surgiu a Internet, apresentando-se também com capacidade e ambiente propício para a resolução de vários desses impasses. Ela tem demonstrado ser o melhor mecanismo já inventado de implementação física do *commons* intelectual (exceto pelo nosso cérebro), principalmente devido a sua natureza pouco rival e mais próxima das idéias do que qualquer

tecnologia até agora empregada. O caráter aberto dos seus protocolos básicos de comunicação, seus baixos custos transacionais, sua crescente capacidade de memória, de processamento e de distribuição, seu decrescente custo de implementação, sua natureza distribuída e a crise da Comunicação Científica impulsionaram o surgimento do modelo de Acesso Aberto.

Esse novo modo de publicação científica pode tornar o desenvolvimento da Ciência um processo muito mais eficiente. Numerosos resultados positivos sociais e econômicos podem ser alcançados, inclusive sob de efeitos de rede e em bases exponenciais, por meio da disseminação aberta de resultados científicos na Internet. Por outro lado, a comercialização de informações e de dados científicos de forma exclusiva produz monopólios, que tem ineficiências econômicas e são contrários ao interesse público em termos éticos e sociais.

Os interesses públicos são servidos de forma melhorada através da máxima disponibilidade e disseminação da pesquisa pública, ao menor custo possível, com as menores restrições no uso e promoção do reuso e integração dos resultados existentes em novas pesquisas. Além disso, em se tratando principalmente de pesquisa realizada com dinheiro público, o Acesso Aberto à literatura científica é uma questão de ética social.

Ainda, o Acesso Aberto evita a duplicação de pesquisa e promove:

- desenvolvimento nos países pobres;
- pesquisa interdisciplinar;
- mais pesquisa e novos tipos de pesquisa;
- verificação dos resultados científicos;
- educação de novos pesquisadores;
- maximização do potencial da pesquisa.

Para conquistarmos o acesso aberto de toda a literatura científica, é imprescindível motivarmos a comunidade científica. Esta, no momento, de forma geral, demonstra pouco interesse por essas questões fundamentais. Há também algumas barreiras de cunho legal a serem superadas. No entanto, é a comunidade quem agrega os agentes do sistema. Ela é a força motriz para mudanças. Ao tomar decisões na direção do Acesso Aberto, também é preciso se preocupar com infra-estrutura de preservação e tratar os recursos da informação como um importante investimento.

Nós estamos no meio de um processo de transição. Em termos econômicos, as grandes editoras científicas (principalmente as comerciais) querem fazer todo o possível para continuarem sendo as vencedoras. Para tanto, utilizam até mesmo a legislação e a opinião judicial para reter a velha estrutura e continuar controlando o mercado. Por outro lado, a Internet nos propiciou um meio novo e muito eficiente de implementação do *commons* científico. Não podemos consentir que interesses comerciais sufoquem a inovação e a oportunidade de termos uma Ciência mais aberta, mais eficiente e mais promotora de interesses sociais.

Apêndice A - Glossário

Digital Rights Management (DRM).

É o *software* entre a informação *on-line* e o usuário. Se o usuário atinge o critério do programa (por exemplo, pagando pela assinatura), então é permitido a ele acesso à informação. Para os demais usuários, o acesso será bloqueado. A tecnologia DRM é implementada de muitas formas diferentes, incluindo criptografia, dados de identificação de usuários (como nome e senha) ou *plug-ins* no navegador do usuário.

Eprint.

É um *preprint* ou *postprint* eletrônico. O termo *e-print* (com o hífen) foi criado em 1992 por Greg Lawler, e originalmente se referia apenas a *preprints* eletrônicos. Paul Ginsparg mais tarde generalizou a expressão para significar *preprints* ou *postprints* eletrônicos que foram depositados em repositórios digitais.

Extensible Mark-up Language (XML).

É uma linguagem para indicar os tipos de informação num texto. É uma linguagem comumente utilizada para codificar metadados. Pode facilitar a descoberta e a coleta de informação, a visualização em vários formatos e a preservação digital. Algumas referências importantes são:

- A página W3C sobre XML, <http://www.w3.org/XML/>.
- A página OASIS sobre XML, <http://www.oasis-open.org/cover/>.

Impacto.

A influência de um trabalho acadêmico entre outros estudiosos da área. Se maior acesso leva a maior impacto (não necessariamente através de uma função simples), então tornar os trabalhos científicos livremente acessíveis na Internet aumenta seu impacto. Essa conclusão é controversa, visto que há vários métodos para medidas de impacto.

Interoperabilidade.

A cooperação, a interação ou o compartilhamento entre diferentes repositórios, editoras, formatos de texto ou setores de informação (como as bibliotecas universitárias e as revistas comerciais) permitida pelo uso de padrões de metadados comuns, isto é, para armazenamento, acesso, catalogação e comunicação. Uma forma de interoperabilidade é a através de *links* de referências, ou a ligação de citações aos seus recursos, por exemplo, o CrossRef. Uma outra forma é o suporte de diferentes repositórios para serviços de dados, como busca, por exemplo, na OAI.

Metadados.

Dados sobre um arquivo de dados. Por exemplo, se um artigo de um periódico online é o arquivo de dados primário, então seus metadados podem incluir autor, título, editora, data e URL. Diferentes padrões de metadados colocam diferentes elementos nessa lista e, às vezes, identificam o mesmo elemento por nomes diferentes. Por exemplo, para alguns propósitos, a língua, o nome do detentor do direito autoral e o número de classificação podem ser relevantes, enquanto para outros não. Metadados podem ser utilizados para ajudar a localizar documentos, catalogá-los, preservá-los, navegar entre eles, abrir ou bloquear acesso àqueles usuários corretos e colocar um link entre as citações e as suas fontes. Como o arquivo de metadados é freqüentemente menor que o arquivo de dados correspondente, ele pode aumentar a velocidade de processamento de catalogação.

Peer review.

Escrutínio de um trabalho por outros acadêmicos do campo. Um método essencial de controle de qualidade realizado pela maior parte dos periódicos científicos e das editoras de livros acadêmicos. Existem novas possibilidades de condução de *peer review* na forma eletrônica.

Postprint.

Uma versão de um texto produzido após ele ter sido publicado numa revista científica e usualmente também depois de ele ter passado pelo *peer review* e ter sido revisado pelo processo editorial. Veja *eprint* e *preprint*.

Preprint.

Uma versão de um texto antes de ele ter sofrido *peer review* ou ter sido publicado num periódico. Veja *eprint* e *postprint*.

- A biblioteca eletrônica da Universidade do Estado de Iowa tem uma lista de coleções de *preprints* gratuitos *on-line* em <http://www.lib.iastate.edu/services/ref/preprint.html>.

Self-Archiving.

A prática dos acadêmicos de colocarem seus trabalhos em repositórios *on-line* compatíveis com o protocolo OAI-PMH. Os autores primeiramente colocam um preprint. Então eles submetem o artigo a uma revista científica (eletrônica ou impressa) com peer review. Se o artigo é aceito e o autor pode negociar o direito de fazer o depósito, então ele pode colocar o *postprint* revisado no repositório. Se ele é aceito mas a editora não permite o *self-archiving*, então o autor pode colocar uma corrigenda (as diferenças entre o *preprint on-line* e a versão publicada do artigo) no repositório. Veja também:

- Stevan Harnad's case for self-archiving, <http://www.cogsci.soton.ac.uk/~harnad/Tp/resolution.htm>.
- The eprints Self-Archiving FAQ, <http://www.cogsci.soton.ac.uk/~harnad/selfaq.htm>.

Apêndice B – Principais Provedores de Dados

arXiv.

Um repositório compatível com a interface OAI, de *preprints* eletrônicos em Física, Matemática, Ciência da Computação, Ciências Não-Lineares e Biologia Quantitativa. Criado por Paul Ginsparg em 1991, ele é um dos mais antigos repositórios acadêmicos de acesso aberto. É financiado atualmente pela Universidade Cornell.

- Home page, <http://arxiv.org>.

Cern Document Server.

Cobre mais de 800 mil registros bibliográficos, incluindo 360 mil textos completos, de interesse para pessoas trabalhando na área de Física de Partículas e outros campos relacionados. Reúne *preprints*, artigos, livros, revistas, fotografias e muito mais. Parte do seu conteúdo está disponível em acesso aberto.

- Home page, <http://cdsweb.cern.ch>

CiteSeer.

Uma biblioteca de acesso aberto de literatura científica. Oferece buscas e cruzamento de referências entre os textos completos que são também distribuídos em diferentes formatos, além de estatísticas de citações, entre outros serviços. É atualmente financiado pela Microsoft, as agências americanas National Science Foundation (NSF) e NASA. É hospedado pela Escola de Ciências da Informação e Tecnologia da Universidade da Pensilvânia.

- Home page, <http://citeseer.ist.psu.edu/>.

CogPrints.

Um repositório EPrints¹ que agrega publicação das áreas de Psicologia, Neurociência, Linguística, Ciência da Computação, Filosofia e Biologia. Conta com cerca de 2.500 artigos com acesso aberto ao texto completo.

- Home page, <http://cogprints.org/>

1- veja mais sobre o *software* EPrints no apêndice C

DSpace@MIT.

Implementado pelo *software* Dspace¹, esse repositório armazena a pesquisa do Massachusetts Institute of Technology (MIT) em forma digital, incluindo *preprints*, relatórios técnicos, artigos de trabalho, teses, conferências, imagens, etc. Cobre mais de 11 mil teses do MIT.

- Home page, <https://dspace.mit.edu/index.jsp>

ECS Service.

Implementado pelo *software* EPrints¹, esse repositório digital agrega toda a publicação de Eletrônica e Ciência da Computação da Universidade de Southampton. A política de preenchimento é o *mandating open access*. Tem crescido rapidamente e hoje conta com quase 10 mil registros.

- Home page, <http://eprints.ecs.soton.ac.uk/>

e-Prints Soton.

Criado a partir do *software* Eprints¹, esse repositório armazena parte da pesquisa científica da Universidade de Southampton. Hoje agrega cerca de 9.500 registros.

- Home page, <http://eprints.soton.ac.uk/information.html>.

eScholarship.

Financiado pela California Digital Library, provê acesso persistente e torna o conteúdo fácil de ser descoberto. Agrega mais de 10 mil artigos da Universidade da Califórnia e é mantido pelo *software* e serviços da Bepress¹. Já ultrapassou a marca dos 45 mil *downloads* de texto completo.

- Home page, <http://repositories.cdlib.org/escholarship/>

Library of Congress.

Biblioteca nacional dos Estados Unidos. Além de colocar uma grande quantidade de informação em acesso aberto, tem sido um líder no movimento das bibliotecas digitais.

- Home page, <http://www.loc.gov/>.

1- veja mais sobre este *software* no apêndice C

National Science Digital Library (NSDL).

É um repositório de informação acadêmica e de materiais interativos de ensino em vários campos da Ciência, desenhado para dar suporte à educação em diferentes níveis. É financiado pela agência americana National Science Foundation (NSF).

- Home page, <http://nsdl.org>

Networked Computer Science Technical Reference Library (NCSTRL).

Um repositório para buscas e listagem de relatórios em Ciência da Computação. É de acesso aberto e compatível com o protocolo OAI-PMH. Trata-se de um projeto colaborativo entre a NASA, a Old Dominion University, a Universidade da Virgínia e o Virginia Tech.

- Home page, <http://www.ncstrl.org>.

Networked Digital Library of Theses and Dissertations (NDLTD).

Um repositório de teses e dissertações digitais internacionais, todas acessíveis sem taxas. É financiado pela UNESCO e pela Adobe Systems Inc. Conta com mais de 217 mil teses e dissertações.

- Home page, <http://www.ndltd.org>.

PubMed Central.

Um repositório digital de literatura de revistas de Ciências da Vida no Instituto Nacional de Saúde (U. S. National Institute of Health), desenvolvido e gerenciado pelo National Center for Biotechnology Information (NCBI) em National Library of Medicine(NLM). Com o PubMed Central, o NLM está levando a liderança em questões de preservação e manutenção de acesso irrestrito à literatura eletrônica, da mesma forma que fez por décadas com a literatura impressa de Ciências Biomédicas.

A PubMed Central oferece dois serviços à comunidade científica mundial: o GenBank, um repositório de seqüência de dados, e o PubMed, a base de dados de citações e abstracts da literatura biomédica e de ciência da vida. Este último serviço oferece facilidades de busca e o texto completo da maior parte do seu material. Cerca de 15% das revistas indexadas podem demorar alguns meses após a publicação para disponibilizar o texto completo.

- PubMed Central home-page, <http://www.pubmedcentral.nih.gov/>.

Research Papers in Economics (RePEc).

É um esforço colaborativo de centenas de voluntários em 53 países para aumentar a disseminação da pesquisa em Economia. O centro do projeto é um banco de dados descentralizado de artigos de trabalho e artigos de revistas e componentes de software. Todo esse material está livremente disponível. No entanto, o RePEc não contém o texto completo dos artigos; ele provê *links* para muitos textos completos de artigos, mas é preciso uma assinatura pessoal ou institucional para seguir esses *links*. O serviço hoje conta com cerca de 171 mil artigos de trabalho e 187 artigos de revistas.

- Home page, <http://www.repec.org/>.

Scientific Electronic Library on-line (SciELO).

Um sistema eletrônico de portais para acesso a coleções digitais de textos completos de cerca de 200 revistas científicas. Existem muitos *sites* em operação, oferecendo acesso a revistas do Brasil, América Latina e Caribe. Tem desempenhado uma função importante na disseminação do conhecimento técnico-científico desses países, possibilitando acesso a uma literatura que antes só podia ser alcançada dentro das fronteiras. Foi fundado pela FAPESP em parceria com a Bireme (Centro Latino-americano e do Caribe de Informação em Ciências da Saúde).

- Home page, <http://www.scielo.br/>.

Apêndice C – Principais *Softwares* para Provedores de Dados

Berkeley Electronic Press (bepress).

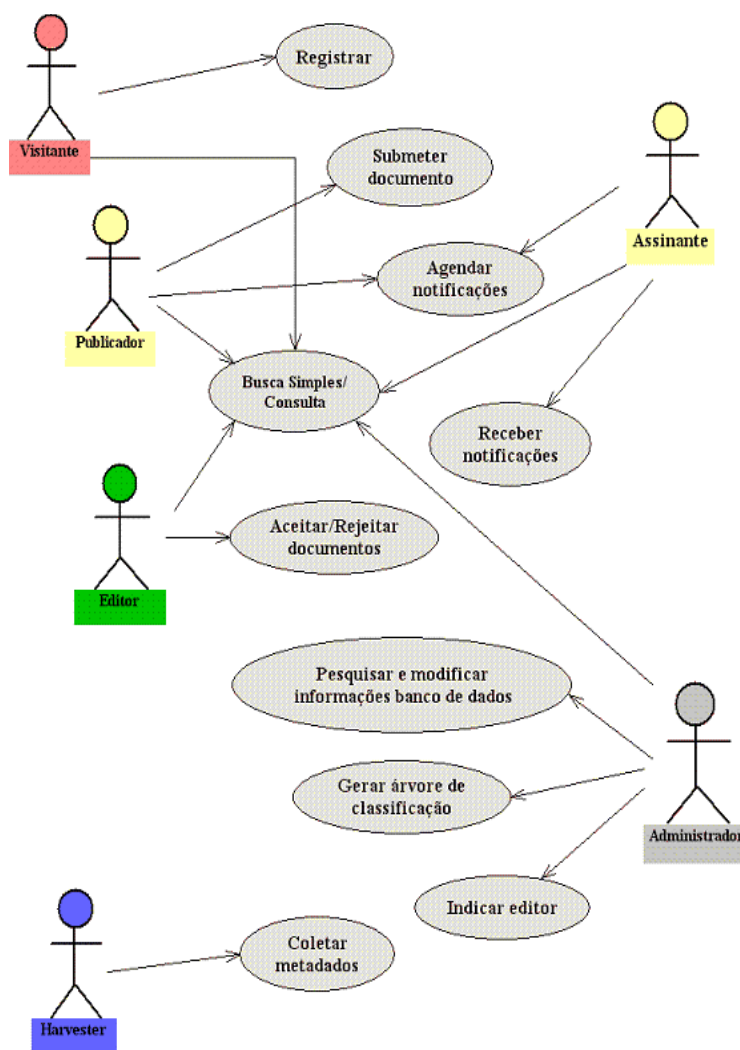
Fundada por acadêmicos em 1999, a Berkeley Electronic Press é tanto uma editora de revistas eletrônicas com *peer review* quanto uma desenvolvedora de *software*. Produz serviços para revistas científicas, repositórios institucionais ou temáticos, séries de artigo e *software* de gerenciamento editorial. Atualmente mantém e desenvolve a biblioteca digital eScholarship, da Universidade da Califórnia.

- O portal da bepress, <http://www.bepress.com>

GNU EPrints.

É uma *software* livre para criar repositórios compatíveis com o protocolo OAI-PMH com conteúdo de uma instituição ou departamento, desenvolvido pela Universidade de Southampton. Seu principal desenvolvedor é Christopher Gutteridge. Foi baseado inicialmente no CogPrints. Muitas das especificações foram dadas por Stevan Harnad.

Seu principal objetivo é configurar um repositório para artigos de pesquisa e sua configuração inicial reflete isso, mas ele pode também armazenar figuras, dados de pesquisa, arquivos de áudio – qualquer coisa que possa ser digitalmente armazenada. Ele foi desenvolvido sob a tecnologia GNU/Linux.



A figura a seguir ajuda a visualizar as funcionalidades do programa. Existem basicamente dois tipos principais de usuários no sistema: o visitante e o registrado. O visitante pode realizar consultas e buscas nos metadados e no texto completo dos artigos. Já o usuário registrado pode assumir vários papéis no sistema. Ele pode ser um usuário publicador, um usuário assinante, um editor ou administrador do sistema. Pode assumir qualquer papel simultaneamente. O usuário publicador é aquele que submete para publicação seus próprios documentos. Já o usuário assinante é aquele que assina serviços de alertas de chegada de novos documentos no repositório. O editor é aquele que aprova ou rejeita os documentos enviados para publicação. O administrador é o usuário capaz de modificar a árvore de classificação de assuntos, alterar dados na base e indicar editores. Ainda existe um tipo especial de usuário-robô: o *harvester*, um sistema externo ao EPrints que coleta automaticamente os metadados armazenados no repositório via protocolo OAI-PMH.

- Veja mais sobre este programa no portal do EPrints, <http://www.eprints.org/>.

DSpace.

Sistema de biblioteca digital para capturar, armazenar, indexar, preservar e redistribuir

materiais de pesquisa acadêmica em formatos digitais. Permite compartilhamento de descobertas de pesquisa de forma rápida com uma audiência global e preserva seus materiais.

É um *software* semelhante ao EPrints, no entanto, apresenta suporte para comunidades.

Atualmente, algumas universidades que estão trabalhando com o DSpace são o Massachusetts Institute of Technology (MIT), a Universidade de Cambridge, a Universidade Cornell, etc.

- Veja mais sobre este *software* no portal do DSpace, <http://www.dspace.org>
- FAQ do DSpace, <http://dspace.org/faqs/index.html>

Apêndice D - Serviços Importantes e Paralelos ao Acesso Aberto

CrossRef.

Uma corporação sem fins lucrativos para automatizar o cruzamento de referências através das editoras, isto é, para conectar citações bibliográficas e de rodapés à fonte do texto completo. CrossRef foi a primeira aplicação do sistema DOI. O objetivo do CrossRef é interligar toda a literatura acadêmica *online*. Operado pelo PILA (Publishers International Linking Association), foi lançado em junho de 2000.

- Home page, <http://www.crossref.org/>

Digital Object Identifier (DOI).

Um número identificador permanente para uma revista, artigo, imagem, citação, *applet*, *script* ou outro objeto digital. Pelo fato de identificar o objeto, difere de uma simples URL e permanece o mesmo quando a URL muda. Para sermos mais precisos, o DOI inclui um número identificador, metadados associados e um sistema para entregar o objeto digital aos usuários que conhecem ou clicam no número identificador.

A International DOI Foundation (IDF) mantém um diretório conectando DOI a URL e metadados correspondentes. O padrão DOI inclui sintaxe própria. Aqueles que detêm direitos de propriedade intelectual dos objetos digitais controlam o DOI do seu objeto. Se os usuários ligam documentos através do DOI e as editoras atualizam o diretório DOI quando seus documentos mudam de lugar, então os *links* não quebram e os objetos têm endereços permanentes.

- Home page, <http://www.doi.org/index.html>.
- DOI FAQ, <http://www.doi.org/faq.html>.

Google Scholar.

Esse serviço do Google oferece buscas para literatura científica especificamente, incluindo artigos revisados por mesa editorial, teses, livros, *preprints*, resumos (*abstracts*) e relatórios técnicos de muitas áreas do conhecimento.

"Ele funciona de forma semelhante ao Google Web Search, ordenando os resultados pela relevância à sua *query*. Assim, as referências mais úteis devem aparecer no topo da página.

Esta ordem de relevância considera o conteúdo completo de cada artigo bem como o autor, a publicação em que o artigo apareceu e quão frequentemente é citada na literatura científica."

"O Google Scholar automaticamente analisa e extrai citações e as apresenta como resultados, mesmo se os documentos a que eles se referem não estão disponíveis *online*. Isso significa que os resultados da sua busca podem incluir citações de trabalhos mais antigos e artigos que aparecem apenas em livros e outras publicações *offline*." (Google Scholar Beta, <http://scholar.google.com/scholar/about.html>, acessada em 20nov2004).

O conteúdo inclui arranjos entre editoras e agregadores com os quais o Google já negociou, por exemplo, a IEEE, o Open WorldCat, Elsevier, etc. No entanto, ele ainda está um tanto incompleto. Com esse serviço, o Google criou um grande competidor para serviços como Web of Science do ISI e para o Scopus da Elsevier.

Lots of Copies Keeps Stuff Safe (LOCKSS).

Uma rede *peer-to-peer* auto-corretora de *caches* desenhados para manter múltiplos e idênticos *backups* do conteúdo no *cache*. Quando o *cache* principal recebe novo conteúdo, os outros nós são automaticamente atualizados.

Quando um nó é danificado ou apagado, ele é automaticamente reparado. Isso pode garantir acesso de longo prazo a periódicos eletrônicos, mesmo se uma editora sai dos negócios. O *software* foi desenvolvido na Universidade Stanford com suporte do NSF, Sun Microsystems e Mellon Foundation.

- Home page, <http://lockss.stanford.edu/>

OpenResolver.

Uma demonstração de resolvedor para OpenURLs.

- Home page, <http://www.ukoln.ac.uk/distributed-systems/openurl/>.

OpenURL.

Uma URL composta contendo um endereço de rede e algum metadado sobre um arquivo de dados. O propósito é fornecer serviços além do sistema de *links* comum tais como: persistência, de forma que a URL permaneça válida mesmo se o destino do arquivo se muda para uma nova localização; resolução múltipla, de forma que clicando num *link*, seja trazido

um menu de opções de acesso ao arquivo com *links* para diferentes versões do arquivo, nos diversos locais onde ele é armazenado ou foi publicado, metadados sobre o arquivo, etc; sensibilidade ao contexto, de forma que as opções do menu sejam determinadas pela identidade do usuário, sua localização e sua instituição. Isso ajuda a determinar se o usuário tem direito de acesso ao arquivo de dados. OpenURL é compatível com o DOI. Dois dos vários resolvedores que existem são: SFX e OpenResolver.

- O padrão, <http://www.sfxit.com/openurl/openurl.html>.

SFX.

Um ferramental para *links* de contexto, ou para solucionar o problema das cópias. Se um usuário clica numa OpenURL associado à citação bibliográfica, então um menu aparece dando ao usuário escolhas para acessar as versões do arquivo em revistas, repositórios e bancos de dados abertos ou assinados pela instituição do usuário.

- Home page, <http://www.sfxit.com/>.

Scopus.

Um serviço de busca semelhante ao ScienceDirect da Elsevier e produzido pela mesma editora. No entanto enquanto o produto ScienceDirect cobre 1.800 revistas com textos completos da Elsevier, o Scopus cobre cerca de 15.000 títulos de revistas. Conforme um representante da Elsevier afirmou, o Scopus complementa, mas não substitui as capacidades únicas de bases de dados especialistas. Scopus é compatível com OpenURL.

- Home page, <http://www.scopus.com/scopus/home.url>.

Web of Science.

Um serviço do Scientific Thompson que oferece acesso à informação multidisciplinar atual e retrospectiva de aproximadamente 8.700 títulos dos mais prestigiosos e de alto impacto existentes na literatura científica. Oferece um método de busca baseado em referências de citação. Os usuários podem navegar para frente e para trás pela literatura.

- Home page, <http://scientific.thomson.com/products/wos/>

REFERÊNCIA

ACRL Scholarly Communications Committee. 2003. Principles and Strategies for the Reform of Scholarly Communication.

URL: <http://www.ala.org/ala/acrl/acrlpubs/whitepapers/principlesstrategies.htm> (acessado em 04/03/06).

Anderson. 2004. Author disincentives and open access. *Serials review*.

Antelman, K. 2004. Do open-access articles have a greater research impact? *College & Research Libraries*. 65(5), 372-382.

URL: <http://eprints.rclis.org/archive/00002309/> (acessado em 04/01/06)

Apt, Krzysztof. 2001. Towards Free Access to Scientific Literature.

URL: <http://www.math.leidenuniv.nl/~naw/serie5/deel02/sep2001/pdf/apt.pdf> (acessado em 15 dez 2004).

Arms, William Y; Naomi Dushay, Dave Fulker; Carl Lagoze. 2002. A Case Study in Metadata Harvesting: the NSDL.

URL: <http://www.cs.cornell.edu/lagoze/papers/Arms-et-al-LibraryHiTech.pdf> (acessado em 15/05/03).

Benkler, Yochai. 2001. The Battle Over the Institutional Ecosystem in the Digital Environment, *44 Communications of the ACM* No.2 84.

Benkler, Yochai. 2002. Coase's Penguin: Linux and The Nature of The Firm. *Yale Law Journal*.

URL: < <http://www.benkler.org/CoasesPenguin.html> > (acessado em 18 jan 2004).

Benkler, Yochai. 2003. Freedom in the Commons: Towards a Political Economy of Information. *Duke Law Journal*.

URL: <http://www.law.duke.edu/journals/dlj/articles/dlj52p1245.htm> (acessado em 10 fev 2005).

Benkler, Yochai. 2004a. Intellectual Property: Commons-Based Strategies and the Problems of Patents. *Science*, vol. 305, 1110-1111. URL:

<http://www.sciencemag.org/cgi/content/summary/305/5687/1110?siteid=sci&ijkey=IJEyaBYYPm9zI&keytype=ref>

(acessado em 17 Set 2004)

Benkler, Yochai. 2004b. Sharing Nicely. *Yale Law Journal*.

URL: http://www.yalelawjournal.org/pdf/114-2/Benkler_FINAL_YLJ114-2.pdf. (acessado em 5 jan 2005)

Boyle, James. 2003. The Second Enclosure Movement and The Construction of Public Domain. *Duke Conference on Public Domain*.

URL: <http://www.law.duke.edu/pd/papers/boyle.pdf>

Brody, T., Kampa, S., Harnad, S., Carr, L. and Hitchcock, S. 2003. Digitometric Services for Open Archives Environments. In *Proceedings of European Conference on Digital Libraries 2003*, pp. 207-220, Trondheim, Norway.

Brody, Tim. 2004. Citation Analysis in the Open Access World.

URL: http://eprints.ecs.soton.ac.uk/10000/01/tim_oa.pdf (acessado em 05/05/2005).

Brody, Tim; Stevan Harnad. 2004. Earlier web usage statistics as predictors of later citation impact, *Journal of the American Association for Information Science and Technology*. URL: <http://www.ecs.soton.ac.uk/~harnad/Temp/timcorr.doc> (acessado em 10/10/2005).

Brody, Tim; Steve Hitchcock; Rachel Hardy; Charles Oppenheim. 2005. Open Access Citation. Final Report: Extended Version – JISC Scholarly Communications Group. URL: http://www.jisc.ac.uk/uploaded_documents/iv%20OA%20Citation%20Information%20FINAL%20Extended%20Report.DOC (acessado em 20/11/05).

de Sompel, Hebert Van; Sandy Payette, John Erickson, Carl Lagoze, Simeon Warner. 2004. Rethinking Scholarly Communication, Building the System that Scholars Deserve. *D-Lib Magazine*, vol. 10, número 9. URL: <http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html> (acessado em 15/11/04)

DOAJ. 2006. URL: <http://www.doaj.org/> (acessado em 20/02/06)

Garvey, William. 1977. *Communication: The Essence of Science*. Pergamon Press (Johns Hopkins University). 231-255; 300-320.

Guédon, Jean-Claude. 2004. The "Green" and "Gold" Roads to Open Access: The Case for Mixing and Matching. *Serials Review*. 30, 315-328.

Hardin, Garret. 1968. The Tragedy of the Commons. *Science*, vol. 162. URL: <http://www.sciencemag.org/cgi/reprint/162/3859/1243.pdf?ijkey=W762Xr9TwfD4g> (acessado em 07/03/05).

Harnad, Stevan; Tim Brody. 2004. Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine*. URL: <http://www.dlib.org/dlib/june04/harnad/06harnad.html>(acessado em 04/09/04).

Harnad, Stevan. 2005. Fast-Forward on the Green Road to Open Access: The Case Against Mixing Up Green and Gold. URL: <http://www.ariadne.ac.uk/issue42/harnad/>

Hess, Charlotte, Elinor Ostrom. Ideas, Artifacts, and Facilities: Information as a Common-Pool Resource. 2003. *Law and Contemporary Problems*, vol. 66. URL: [https://www.law.duke.edu/shell/cite.pl?66+Law+&+Contemp.+Probs.+111+\(WinterSpring+2003\)](https://www.law.duke.edu/shell/cite.pl?66+Law+&+Contemp.+Probs.+111+(WinterSpring+2003)) (acessado em 05/02/06).

Hess, Charlotte; Ostrom, Elinor. 2004. Studying Scholarly Communication: Can Commons Research and the IAD Framework Help Illuminate Complex Dilemmas? *Workshop in Political Theory and Policy Analysis*. Indiana University. URL: http://dlc.dlib.indiana.edu/archive/00001411/00/Hess_Studying_040512_Paper192.pdf (acessado em 15/10/04).

Hitchcock, Steve; Tim Brody, Christopher Gutteridge, Les Carr; Stevan Harnad. 2003. The Impact of OAI-based Search on Access to Research Journal Papers. URL: <http://opcit.eprints.org/serials-short/serials11.html> (acessado em 07/03/04)

Jochum, Uwe. 1999. The Alexandrian Library and its Aftermath. *Library History*. 15, 5-12. URL: <http://www.ub.uni-konstanz.de/fi/ger/alexandria-aftermath.pdf> (acessado em 07/01/06).

Kleinberg, Jon M. Authorative Sources in a Hyperlinked Envirommnet. 1998. *Proc. 9th ACM-SIAM Symposium on Discrete Algorithm*.

URL: <http://www.cs.cornell.edu/home/kleinber/auth.pdf> (acessado em 15/02/03)

Knuth, Donald. 2003. Public letter to fellow members of the editorial board of the Journal of Algorithms. URL: <http://www-cs-faculty.stanford.edu/~knuth/joalet.pdf> (acessado em 07/11/03).

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. University of Chicago.

Lawrence, Steve. 2001. on-line or Invisible? *Nature*, vol. 411, 6837, p. 521.

URL: <http://citeseer.ist.psu.edu/online-nature01/> (acessado em 15 Out 2004).

Le Coadic, Y. 1996. *A Ciência da Informação*. Brasília: Briquet de Lemos/livros. 119 p.

Lesk, Michael. 1997. *Practical Digital Libraries: Books, Bytes & Bucks*. Editora Morgan Kaufmann. San Francisco, California. p. 1-8.

Lessig, Lawrence. 1999. *Code and other laws of cyberspace*. New York: Basic Books.

Lessig, Lawrence. 2001. *The Future of Ideas: The Fate of The Commons in a Connected World*. New York: Random House.

Lessig, Lawrence. 2004. *Free Culture*. The Penguin Press, New York.

Lessig, Lawrence. 2006. Code and the Other Laws of Cyberspace.

URL: <http://codebook.jot.com/Book/Chapter7/Ch7Part1> (acessado em 03/01/06).

Library Comittee Recommendation. 2004.

URL: <http://oregonstate.edu/dept/senate/agen/2004/library.html> (acessado em 10/01/06)

Licklider, Joseph Carl Robnett . 1965. *Libraries of the Future*. Cambridge, MA..

Liu, Xiaoming; Kurt Maly, Mohammad Zubair, Michael L. Nelson. 2001. Arc - An OAI Service Provider For Cross-Archive Searching. *JCDL*.

URL: <http://www.ils.unc.edu/~mln/jcdl-arc.pdf> (acessado em 15/10/03).

Litman, Jessica. 2001. *Digital Copyright*. Amherst, NY: Prometheus Books.

Lyman, Peter. 1997. Scholarly Communication and Technology. Conferência sobre Documentos Digitais e o Futuro da Comunidade Acadêmica, Emory University.

URL: <http://www.arl.org/scomm/scat/lyman.html> (acessado em 20/11/05).

Lynch, Clifford. 2001. Metadata Harvesting and the Open Archives Initiative.

URL: <http://www.arl.org/newsltr/217/mhp.html> (acessado em 12/01/06)

Meadows, A. J. 1999. *A Comunicação Científica*. Brasília: Briquet de Lemos/livros. 268 p.

Newman, M. E. J. 2001. The structure of scientific collaboration network. *Proceedings of National Academy of Sciences. Applied Mathematics*.

URL: <http://intl.pnas.org/cgi/content/full/98/2/404> (acessado em 20/11/04).

OAister, 2006. URL: <http://oaister.umd.umich.edu/o/oaister/> (acessado em 22/02/06).

Odlyzko, Andrew M. 1995. Tragic Loss and Good Riddance: The Impeding Demise of Traditional Scholarly Journals. *Notices of the American Mathematical Society*.

URL: <http://www.dtc.umn.edu/~odlyzko/doc/tragic.loss.long.pdf> (acessado em 22/11/04)

Odlyzko, Andrew. 2000a. Peer and Non-peer Review.

URL: <http://www.dtc.umn.edu/~odlyzko/doc/peer.review.txt>. (acessado em 20/11/05).

Odlyzko, Andrew M. 2000b. The Future of Scientific Communication. AT&T Labs. *NIWI*, pp. 273-278.

URL: <http://www.dtc.umn.edu/~odlyzko/doc/future.scientific.comm.pdf> (acessado em 20/12/04)

Odlyzko, Andrew M. 2002. Rapid Evolution of Scholarly Communication. AT&T Labs. *Learned Publishing*, 15(1), pp. 7-19.

URL: <http://www.dtc.umn.edu/~odlyzko/doc/rapid.evolution.pdf> (acessado em 20/12/04)

Powell, Andy, Michael Day, Peter Cliff. 2004. Using simple Dublin Core to describe eprints. UKOLN, University of Bath. Version 1.2.

URL: <http://www.rdn.ac.uk/projects/eprints-uk/docs/simpledc-guidelines> (acessado em 07/07/05).

PLoS, 2006. URL: <http://biology.plosjournals.org/perlserv/?request=index-html&issn=1545-7885> (acessado em 22/02/06)

Pringle, Jim. 2004. New Opportunities in Open Access. Thompson Scientific.

URL: <http://scientific.thomson.com/news/newsletter/2004-11/8254463/> (acessado em 20/02/2006)

Redmond, Donald A.; Sinclair, Michael P.; Brown, Elinore. 1972. University libraries and university research. *College & Research Libraries*, v. 33, n. 6, p. 447- 453.

Reichman, J. H., Paul Uhler. 2003. A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment. *Law & Contemp. Probs*, vol. 66.

URL: <http://www.law.duke.edu/journals/lcp/downloads/LCP66DWinterSpring2003P315.pdf> (acessado em 12 /07/04).

Salus, Peter. 2005. A History of Free and Open Source. URL:

<http://www.groklaw.net/articlebasic.php?story=20050327184603969> (acessado em 05/03/05)

Simon, Imre. 2000. Propriedade Intelectual na Era da Internet.

URL: <http://www.ime.usp.br/~is/papir/direitos/direitos-dgz.html> (acessada em 02/03/06).

Simon, Imre. 2002. Oráculo Bibliográfico: Sonhos de Um Pesquisador.

URL: <http://www.ime.usp.br/~is/papir/oraculo/oraculo-rusp/oraculo-rusp.html> (acessado em 10/04/02)

Suber, Peter. 2004. Creating an Intellectual Commons through Open Access.

URL: <http://dlc.dlib.indiana.edu/archive/00001246/01/suberrev052804.pdf> (acessado em 07/01/06).

Suber, Peter. 2005. Open Access Overview.

URL: <http://www.earlham.edu/~peters/fos/overview.htm>(acessado em 29/11/05).

Suber, Peter. 2006. Conference and Workshops related to Open Access Movement.

URL: <http://www.earlham.edu/~peters/fos/conf.htm> (acessado em 15/02/06).

Suber, 2006b. Open Access Timeline.

URL: <http://www.earlham.edu/~peters/fos/timeline.htm> (acessado em 05/01/06).

Targino, Maria das Graças. 1998. Comunicação Científica: Uma Revisão de Seus Elementos Básicos. URL: <http://www.informacaoesociedade.ufpb.br/pdf/IS1020002.pdf>. (acessado em 05/01/06).

Tenopir, Carol; Donald W. King. 2001. Lessons for the future of journals. *Nature* 413 (672-674), doi:10.1038/35099602 (acessado em 19/11/03).

Thorin, Suzanne. Global Changes in Scholarly Communication. 2003. *Presented at e-Workshops on Scholarly Communication in the Digital Era, Feng Chia University*. Taichung, Taiwan.

URL: <http://www.arl.org/scomm/disciplines/Thorin.pdf> (acessado em 19/12/05).

Wikipedia. Library. 2006. URL: <http://en.wikipedia.org/wiki/Library> (acessado em 10/02/06).

Wikipedia. History of Science. 2006a. URL: http://en.wikipedia.org/wiki/History_of_science (acessado em 07/02/06).

Wikipedia. Scientific Method. 2006b. URL: http://en.wikipedia.org/wiki/Scientific_method (acessado em 07/02/06).

Wikipedia. Scientific community. 2005. URL: http://en.wikipedia.org/wiki/Scientific_community (acessado em 13/12/05).

Wikipedia Statistics. 2006. URL: <http://stats.wikimedia.org/EN/TablesWikipediaZZ.htm> (acessada em 02/03/06)

Willinsky, John. 2003. Scholarly associations and the economic viability of open access publishing. *Journal of Digital Information* 4(2).

URL: <http://jodi.ecs.soton.ac.uk/Articles/v04/i02/Willinsky/> (acessado em 3/02/05).

Willinsky, John. 2005. *The Access Principle. The Case for Open Access to Research and Scholarship*. London: MIT Press.

Wikimedia Commons. 2006.

URL: http://commons.wikimedia.org/wiki/Main_Page (acessado em 2/03/06).

Ziman, J. 1979. *Conhecimento Público*. Belo Horizonte: Itatiaia. 164 p.

Ziman, J. 1984. *An introduction to science studies: the philosophical and social aspects of science and technology*. Cambridge: Cambridge University, 203 p.