

INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL

FLAVIO MORETTI

**APLICAÇÃO DAS ESTRATÉGIAS DE APROPRIAÇÃO DAS INOVAÇÕES DE  
TECNOLOGIA DA INFORMAÇÃO PARA A INDÚSTRIA BRASILEIRA DE  
SOFTWARE: O CASO DAS TECNOLOGIAS RELACIONADAS AO  
SEQUENCIAMENTO GENÉTICO HUMANO**

Rio de Janeiro

2020

Flavio Moretti

**Aplicação das estratégias de apropriação das inovações de tecnologia da informação para a indústria brasileira de software: o caso das tecnologias relacionadas ao sequenciamento genético humano**

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Propriedade Intelectual e Inovação, do Instituto Nacional da Propriedade Industrial.

Orientador: Prof. Dr. Celso Lage

Rio de Janeiro

2020

## **AGRADECIMENTOS**

Agradeço à família pelo apoio incondicional nos momentos de dificuldade e pelo incentivo.

Agradeço aos professores pelos ensinamentos em sala de aula e valiosas colaborações a respeito do trabalho.

## RESUMO

Moretti, Flavio. **Aplicação das estratégias de apropriação das inovações de tecnologia da informação para a indústria brasileira de software: o caso das tecnologias relacionadas ao sequenciamento genético humano**. 2020. 77 f. Dissertação (Mestrado em Propriedade Intelectual e Inovação) – Instituto Nacional da Propriedade Industrial, Rio de Janeiro, 2020.

Neste trabalho exploram-se as estratégias utilizadas pelas empresas de tecnologia da informação para a comercialização das inovações ocorridas no setor. A fim de que o estudo trate de inovações relevantes e atuais, busca-se a análise de inovações de tecnologia da informação que são utilizadas nos métodos modernos de sequenciamento genético humano, conhecidos como NGS. Os bancos de dados NoSQL, desenvolvidos a partir de 2006, constituem-se em uma tecnologia de armazenamento de dados capaz de lidar com a necessidade de tratamento de grandes volumes de dados heterogêneos. Esta capacidade vai ao encontro dos requisitos advindos a partir do desenvolvimento dos métodos NGS. Aprofundando-se no caso dos bancos de dados NoSQL, pesquisam-se as estratégias de apropriação adotadas pelas empresas envolvidas com esta inovação. Desta forma, levantam-se as formas de proteção por direitos de propriedade intelectual associadas a estas inovações, assim como os serviços especializados necessários à comercialização dos respectivos produtos comerciais. A partir da análise dos resultados encontrados quanto à proteção por direitos de propriedade intelectual e aos serviços especializados associados à comercialização das inovações dos bancos de dados NoSQL, apontam-se as possibilidades que se aplicam ao mercado brasileiro de software.

Palavras-chave: Estratégia. Inovação. Tecnologia da Informação. NoSQL. Propriedade intelectual.

## ABSTRACT

Moretti, Flavio. **Application of appropriation strategies for information technology innovations for the brazilian software industry: the case of technologies related to human genetic sequencing**. 2020. 77 f. Dissertação (Mestrado em Propriedade Intelectual e Inovação) – Instituto Nacional da Propriedade Industrial, Rio de Janeiro, 2020.

This study aims to explore the strategies used by information technology companies to commercialize innovations. To guarantee that relevant and current innovations are taken into account, the analysis focuses on information technology innovations that are used in modern methods of human genetic sequencing, known as NGS. NoSQL databases, developed since 2006, are a data storage technology, capable of dealing with the need to handle large volumes of heterogeneous data. This capacity meets the requirements arising from the development of NGS methods. Going deeper into the case of NoSQL databases, this work presents the results of the research of the appropriation strategies adopted by the companies involved with this innovation. In this way, the forms of protection for intellectual property rights associated with these innovations are raised, as well as specialized services for the sale of commercial products. From the analysis of the results found regarding the protection by intellectual property rights and the specialized services associated with the commercialization of the innovations of NoSQL databases, the possibilities that apply to the Brazilian software market are pointed out.

Keywords: Strategy. Innovation. Information Technology. NoSQL. Intellectual property.

## LISTA DE ILUSTRAÇÕES

Gráfico 1 —	Custo do sequenciamento do genoma humano.....	30
Quadro 1 —	Marcas registradas para os serviços de bancos de dados NoSQL fornecidos em nuvem.....	54
Gráfico 2 —	Depósitos de patentes que citam o termo BigTable.....	56
Gráfico 3 —	Depósitos de patentes que citam o termo NoSQL.....	59
Gráfico 4 —	Distribuição dos depósitos de patentes que citam o termo BigTable.....	65
Gráfico 5 —	Distribuição dos depósitos de patentes que citam o termo NoSQL.....	65

## LISTA DE ABREVIATURAS

ACID	<i>Atomicity, Consistency, Isolation, Durability</i>
BASE	<i>Basically Available, Soft State, Eventual Consistency</i>
IaaS	<i>Infrastructure as a Service</i>
PaaS	<i>Platform as a Service</i>
PI	Propriedade Intelectual
SaaS	<i>Software as a Service</i>
SNV	<i>Single Nucleotide Variants</i>
TI	Tecnologia da Informação

## LISTA DE SIGLAS

ASF	<i>Apache Software Foundation</i>
APF	Administração Pública Federal
DOE	<i>Department of Energy</i>
FSF	<i>Free Software Foundation</i>
HGMD	<i>Human Gene Mutation Database</i>
INPI	Instituto Nacional da Propriedade Industrial
NCBI	<i>National Center for Biotechnology Information</i>
NHGRI	<i>National Human Genome Research Institute</i>
NIH	<i>National Institute of Health</i>
PGH	Projeto Genoma Humano
P&D	Pesquisa e Desenvolvimento



# SUMÁRIO

INTRODUÇÃO.....	10
QUESTÃO DE PESQUISA.....	12
OBJETIVOS.....	12
OBJETIVO GERAL.....	12
OBJETIVOS ESPECÍFICOS.....	12
METODOLOGIA.....	13
JUSTIFICATIVA.....	14
DESENVOLVIMENTO.....	15
1 ESTRATÉGIAS DE APROPRIAÇÃO DE INOVAÇÕES TECNOLÓGICAS.....	15
1.1 INOVAÇÃO E DESENVOLVIMENTO ECONÔMICO.....	15
1.2 APROPRIAÇÃO DE INOVAÇÕES TECNOLÓGICAS.....	24
2 INOVAÇÕES DE TECNOLOGIA DA INFORMAÇÃO NO CONTEXTO DO SEQUENCIAMENTO GENÉTICO HUMANO.....	28
2.1 CONTEXTUALIZAÇÃO DO SEQUENCIAMENTO GENÉTICO HUMANO.....	28
2.2 INOVAÇÕES DE TI RELACIONADAS AO SEQUENCIAMENTO GENÉTICO HUMANO.....	35
3 ESTRATÉGIAS DE APROPRIAÇÃO DOS BANCOS DE DADOS NOSQL.....	41
3.1 COMERCIALIZAÇÃO DOS BANCOS DE DADOS NOSQL.....	41
3.2 APROPRIAÇÃO DA INOVAÇÃO DOS BANCOS DE DADOS NOSQL.....	49
DISCUSSÃO.....	59
CONCLUSÃO.....	66
REFERÊNCIAS.....	70

## INTRODUÇÃO

O tema do sequenciamento genético humano passa a ser objeto de discussão na década de 1980, com a divulgação pelo *Department of Energy* (DOE) dos EUA do objetivo de mapear o genoma humano. Posteriormente, inicia-se formalmente em 1990 o Projeto Genoma Humano (PGH), uma iniciativa do poder público americano patrocinado pelo *National Institute of Health* (NIH) e pelo DOE.

Em abril de 2003, é anunciado o sequenciamento completo do genoma humano. A partir desse momento, o desafio era tornar viável a comparação entre os genes de milhões de pessoas a fim de interpretar, por exemplo, como as instruções codificadas no DNA influenciavam nas questões de saúde. Assim, em 2004, foi iniciado o programa genoma de mil dólares, executado pelo *National Human Genome Research Institute* (NHGRI) e tendo este valor como o custo alvo do sequenciamento genético humano. O objetivo do programa era financiar projetos de pesquisa básica em novos métodos de sequenciamento, assim como projetos de pesquisa industrial para desenvolver estas novas tecnologias para uso comercial (HAYDEN, 2014).

O National Human Genome Research Institute (2020) mantém atualizados os dados sobre os custos do sequenciamento genético humano. Em 2007, quando predominava a tecnologia Sanger<sup>1</sup> (posteriormente conhecida como tecnologia de primeira geração) para a realização do sequenciamento genético, o montante para o sequenciamento de um indivíduo estava na ordem de 10 milhões de dólares. Com o advento do programa genoma de mil dólares, surge uma série de tecnologias de sequenciamento que ficaram conhecidas como *next-generation sequencing*, ou NGS. Em 2015, com as tecnologias NGS difundidas, o patamar desejado de mil dólares já havia sido atingido.

Uma característica básica das tecnologias NGS é o sequenciamento de milhões de fragmentos de DNA em paralelo. O objetivo desta característica é obter ganho significativo em termos de rendimento, sendo uma consequência a produção de grande volume de dados. A partir de então, tornam-se questões críticas o armazenamento e transferência de dados, a capacidade computacional de análise e

---

<sup>1</sup>Processo de sequenciamento genético o qual consiste na realização de cópias do DNA a ser sequenciado que incluem versões quimicamente modificadas e com marcação fluorescente dos blocos de construção da molécula (HAYDEN, 2014).

métodos de visualização das informações geradas (BREVERN et al., 2015). A adoção de soluções que combinem as tecnologias NGS com bancos de dados NoSQL, uma inovação da área de tecnologia da informação (TI) para o armazenamento de dados heterogêneos e em grande escala, passa a ser objeto de estudo por parte de profissionais ligados à área de bioinformática.

A adoção dos bancos de dados NoSQL como solução de armazenamento de dados além da fronteira da própria TI evidencia a relevância desta inovação. Portanto, o entendimento das estratégias de apropriação adotadas pelas empresas para os bancos de dados NoSQL se constitui em um conhecimento com base no qual pode-se identificar estratégias competitivas adequadas para posicionar-se no mercado de TI.

O primeiro capítulo, intitulado Estratégias de Apropriação das Inovações Tecnológicas, constitui-se na base teórica sobre inovação, desenvolvimento econômico e apropriação de inovações tecnológicas, subsidiando a justificativa da realização do trabalho. O segundo capítulo trata das inovações de TI no contexto do sequenciamento genético humano, de forma a identificar inovações relevantes e atuais por meio de pesquisa bibliográfica publicada em periódicos de artigos científicos. Além disso, delimita o escopo do estudo às inovações relacionadas ao tema do sequenciamento genético humano. Em virtude da grande quantidade de estudos referentes aos bancos de dados NoSQL identificados no segundo capítulo, identificam-se no terceiro capítulo as estratégias de apropriação adotadas pelas empresas de TI para esta inovação, por meio de bibliografia técnica especializada ou publicada em artigos científicos, além de busca em bases internacionais de marcas e patentes. Para a conclusão do objetivo do trabalho, apontam-se as estratégias de comercialização e de proteção por ativos de propriedade intelectual (PI) utilizadas para os bancos de dados NoSQL que se aplicam à indústria brasileira de software.

## **QUESTÃO DE PESQUISA**

Quais estratégias de proteção por direitos de propriedade intelectual e de comercialização por meio de serviços especializados adotadas pelas empresas para os bancos de dados NoSQL podem ser aplicadas no contexto da indústria brasileira de software?

## **OBJETIVOS**

### **OBJETIVO GERAL**

Apontar as possibilidades de aplicação pelo mercado brasileiro de software, das estratégias de proteção por direitos de propriedade intelectual e de comercialização por meio de serviços especializados utilizadas no caso dos bancos de dados NoSQL.

### **OBJETIVOS ESPECÍFICOS**

Levantar as formas de proteção por direitos de PI dos bancos de dados NoSQL para definir o regime de apropriação destas inovações.

Identificar os serviços especializados necessários à comercialização dos bancos de dados NoSQL para definir os ativos complementares destas inovações.

Analisar as estratégias de proteção por ativos de PI e de comercialização por serviços especializados, adotadas pelas empresas de TI para dos bancos de dados NoSQL, com o intuito de identificar as possibilidades de aplicação ao mercado brasileiro de software.

## METODOLOGIA

O levantamento das formas de proteção por direitos de PI dos bancos de dados NoSQL realizou-se mediante buscas em bases de patentes e de marcas.

Adicionalmente, com o intuito de identificar-se as formas de uso dos direitos autoral e de cópia para aquela inovação, efetuou-se pesquisa bibliográfica em artigos científicos e literatura técnica especializada.

Para realizar as pesquisas de patentes, foi utilizado o motor de busca *Lens*, por ser uma ferramenta disponível gratuitamente na Internet, atualizada semanalmente, e que integra as bases de dados de 95% das patentes ao redor do mundo, totalizando acesso a mais de 119 milhões de documentos, incluindo os mais de 850 mil depositados no Brasil. A ferramenta está disponível no endereço <https://www.lens.org/lens/>.

Realizaram-se duas buscas de patentes.

A primeira, utilizando-se a palavra-chave “BigTable” em depósitos que contenham este termo no relatório descritivo, ou que cite o documento “Bigtable: A Distributed Storage System for Structured Data.”. Este constitui-se no primeiro documento publicado acerca de um modelo de armazenamento distribuído que posteriormente foi utilizado como base para o desenvolvimento de diversos produtos comerciais rotulados como NoSQL.

Para a segunda busca, utilizou-se a palavra-chave “NoSQL” contida nos títulos ou relatórios descritivos dos documentos de patentes.

Os resultados foram agrupados por famílias de patentes, pois objetivou-se nestas buscas obter o quantitativo por depositante, não importando, para efeitos deste estudo, a distinção por país de depósito.

Para realizar a pesquisa de marcas registradas, utilizou-se a ferramenta TMView, disponibilizada na Internet pelo Escritório da Propriedade Intelectual da União Europeia (EUIPO), disponível em <https://www.tmdn.org/tmview/welcome>. A ferramenta agrupa mais de 55 milhões de marcas registradas em 70 escritórios de propriedade intelectual no mundo. Neste estudo, objetivou-se com a busca de marcas, identificar a estratégia de proteção por este ativo de PI com relação à comercialização dos bancos de dados NoSQL por meio dos serviços especializados de computação em nuvem.

Identificaram-se os serviços especializados associados à comercialização dos bancos de dados NoSQL por meio de pesquisa de bibliografia técnica especializada e artigos científicos.

## **JUSTIFICATIVA**

De acordo com estudo publicado pela Associação Brasileira das Empresas de Software (ABES), em 2019 foram identificadas 21.020 empresas atuando no setor de software e serviços no Brasil. Estas empresas produziram naquele ano aproximadamente 20,5 bilhões de dólares, o que correspondeu a 1,5% da produção de TI de softwares e serviços em todo o mundo, colocando o país em 11º lugar no ranking mundial (ASSOCIAÇÃO BRASILEIRA DAS EMPRESAS DE SOFTWARE, 2020).

O presente trabalho propõe-se a auxiliar as empresas nacionais dedicadas ao desenvolvimento e produção de software, à distribuição e comercialização ou à prestação de serviços, a identificar as estratégias adotadas pelas empresas de TI na busca pela apropriação das inovações ocorridas no setor. Ao focar nas tecnologias de TI associadas ao tema do sequenciamento genético humano, o estudo concentra-se em tecnologias relevantes e atuais.

Com base nos resultados deste trabalho, as empresas brasileiras poderão desenvolver as próprias estratégias de acordo com o tamanho e especialidade, proporcionando maior competitividade em relevante mercado global.

## DESENVOLVIMENTO

### 1 ESTRATÉGIAS DE APROPRIAÇÃO DE INOVAÇÕES TECNOLÓGICAS

#### 1.1 INOVAÇÃO E DESENVOLVIMENTO ECONÔMICO

As inovações tecnológicas correspondem à aquisição, introdução e aproveitamento de novas tecnologias na produção ou distribuição de bens ou serviços para o mercado. A nova forma de produção ou distribuição, assim como as mercadorias resultantes, são um produto do desenvolvimento de três processos relacionados entre si: a invenção, a inovação e a difusão desta (SZMRECSÁNYI, 2006).

O autor discorre sobre o trabalho de Joseph Schumpeter, indicando que o desenvolvimento econômico caracteriza-se pela ocorrência de inovações radicais, as quais resultam em novas combinações dos recursos disponíveis. O autor define inovação radical como a introdução de um novo produto ou nova qualidade de produto, a introdução de novos métodos de produção e distribuição, fundamentados ou não na descoberta e invenção de novos conhecimentos, a abertura e exploração de novos mercados, a obtenção de novas fontes de abastecimento de matérias-primas, produtos intermediários e insumos produtivos ou o estabelecimento de novas formas de organização econômica, que levam à conquista ou destruição de uma posição de monopólio.

Além disso, complementa que o surgimento e a consolidação das inovações radicais tende a ocorrer fora do âmbito das pessoas e organizações estabelecidas. Normalmente, se dá pela atuação de *outsiders* e novas empresas, que passam a concorrer com aquelas já existentes, passando a disputar mercados e recursos. Assim, o inovador é aquele que introduz ou consolida novos produtos, novas formas de produzir e de comercializar. O inovador detém, portanto, uma posição de liderança por abrir novos caminhos e por atrair imitadores. Estes, por sua vez, concorrem com os primeiros pelos lucros advindos das inovações.

Ainda segundo Szmrecsányi (2006), a concorrência entre as empresas gira em torno de variáveis como a qualidade dos produtos, as estratégias de comercialização, e, principalmente, as inovações e os produtos e processos que se pretendem substituir. Este ciclo constante de concorrência por inovações obriga as empresas a investir continuamente em progresso técnico e organizacional, a fim de resguardar e possivelmente ampliar as posições conquistadas.

Adicionalmente, o autor entende o advento da proteção por patentes como um reconhecimento da importância dos lucros na economia, como também do impacto das inovações na busca pela preservação desses lucros e do nível de empregos nas empresas inovadoras e respectivos ramos de atividade. Considerando esta perspectiva de competição por inovações, Possas (1997) identifica o conhecimento como um elemento de diferenciação entre as firmas, fundamental para a obtenção de lucros que sobreponham os custos de produção. Sendo um fator relevante para geração e apropriação de poder de compra, o conhecimento por si só é uma fonte de rendimento. Algumas formas de obtenção de renda por meio da comercialização do conhecimento são claras e possuem regulamentação legal, como as patentes e a transferência de tecnologia. No entanto, adverte a autora que há dificuldades na transação mercadológica do conhecimento, como a característica não rival deste bem, ou seja, a utilização por um ente não inviabiliza que outro o faça, e o fato de ser inviável para o interessado na aquisição avaliar a utilidade antes da compra.

Deve-se ainda distinguir-se informação de conhecimento, pois este inclui todo o referencial do receptor, que permite, desta forma, a decodificação e utilização da primeira. Esta diferenciação cria ainda maiores dificuldades para a comercialização do conhecimento, como a impossibilidade de plena transmissão deste, e a necessidade de requisitos prévios por parte do comprador. Tais características fazem do conhecimento um bem de difícil reprodução, e justamente por isso, trata-se de um elemento indispensável para a obtenção de lucros e consequentemente de poder de compra (POSSAS, 1997).

Szmrecsányi (2006) ressalta ainda que as inovações são distintas das invenções, as quais são irrelevantes do ponto de vista econômico enquanto não forem incorporadas à produção ou circulação de mercadorias. O desenvolvimento de uma invenção e a realização de uma inovação constituem processos distintos entre si, mesmo quando empreendidos por uma mesma pessoa ou organização.



Nelson (2006) define o termo sistema de inovação para abranger uma ampla variedade de instituições envolvidas no suporte e direção da atividade econômica na qual a inovação é a principal força motora. O autor defende que estas instituições, como universidades e sistemas públicos de pesquisa, sociedades técnicas e científicas e programas governamentais, desempenham papel relevante no processo de inovação. Adicionalmente, identifica na literatura que trata do tema de sistemas de inovação uma tendência de foco, especialmente, em instituições envolvidas nos estágios iniciais do processo de inovação, particularmente pesquisa e desenvolvimento (P&D), além de considerações que incluem entre estas instituições o mercado de trabalho, o sistema educacional, as estruturas regulatórias, além de outras capazes de moldar a dinâmica econômica.

Entende o autor que o crescimento econômico é resultado da evolução conjunta de tecnologias, estruturas empresariais e industriais e instituições de suporte e governamentais. Ao dividir a história do crescimento econômico em eras, identifica em cada uma um conjunto relativamente pequeno de tecnologias e indústrias que direcionam aquele crescimento econômico em particular. Para cada era, as tecnologias e indústrias chaves requereram um conjunto distinto de instituições de suporte, e as nações de liderança são as que possuíam ou lograram êxito em construir um conjunto apropriado de instituições, considerando como essenciais para esta dinâmica as políticas públicas e programas e o desenvolvimento de leis.

Assim, entende que, para países em busca de desenvolvimento, o desafio é aprender a dominar novas formas de realizar algo, ou seja, desenvolvimento requer inovação. Para os países em desenvolvimento, a inovação envolve trazer e aprender a dominar algo que pode ter sido utilizado por algum tempo em economias mais avançadas. Na maior parte dos casos, existem modelos oriundos de países avançados que servem como alvos de emulação, e em muitas oportunidades, há disponibilidade de assistência para o desenvolvimento de uma nova capacidade. Em outros casos, aspectos importantes do modelo podem simplesmente ser importados. Salienta o autor que trazer para operação práticas novas ao contexto envolve aprendizado considerável a fim de que novos modos de operação estejam efetivamente sob controle. Neste processo há grande chance de falha, e, portanto, a essência de tentar-se algo novo, ou seja, da inovação, é a incerteza dos acontecimentos futuros, não sendo o sucesso uma garantia. Freeman e Perez

(1988) citam que nos estágios iniciais de inovações tecnológicas radicais prevalece a incerteza, e, no nível da inovação individual, os resultados de estudos empíricos sobre P&D demonstram claramente que investimentos em novos produtos e processos possuem elementos de incerteza, ou seja, os resultados não podem ser conhecidos *ex-ante*.

Enfatiza ainda Nelson (2006) que a inovação bem sucedida requer acesso ao capital físico e humano. No entanto, a inovação e o aprendizado efetivo tendem a atrair suprimentos tanto de um quanto de outro. Assim sendo, se um país não possui a estrutura institucional que permita a atração de capital físico e humano, a inovação ocorrerá de forma escassa. A análise de casos de países asiáticos como Japão, Coreia e Taiwan demonstra que a inovação impulsionou os respectivos processos de desenvolvimento, ocorridos em um ambiente onde os estoques de capital físico e humano estavam disponíveis.

Dessa forma, o autor identifica uma gama de instituições fundamentais para o processo de desenvolvimento, como, por exemplo, o sistema educacional, desde a formação básica até o treinamento avançado em ciência e tecnologia, incluindo corpos de conhecimentos e capacidades específicas. A busca pelo desenvolvimento envolve capacidades de aprendizado e inovação nas empresas, e um sistema robusto de universidades e laboratórios públicos de pesquisa desempenham papel fundamental de suporte neste processo. Também deve ser entendida como instituição fundamental para o desenvolvimento o sistema financeiro, cuja reestruturação deve ser criativa a fim de permitir a transferência de recursos de antigas empresas e indústrias para as inovadoras. Conforme citado, Freeman e Perez (1988) identificam em estudos empíricos a observância de incerteza no nível da inovação individual. No entanto, ressaltam os autores, a análise não pode ser restrita à inovação individual ou à quantidade de inovações. Aspectos qualitativos e a inter-relação de sistemas de inovações devem ser considerados, pois criam condições favoráveis ao investimento a despeito de riscos e incertezas. Tais condições favoráveis incluem a complementaridade entre inovações, infraestrutura apropriada, estabilidade política e instituições que auxiliam na difusão de novas tecnologias.

Por fim, Nelson (2006) entende que os sistemas de inovação nacionais são moldados por políticas públicas, mas de forma alguma devem ser analisados como planejados detalhadamente, mas sim, evoluídos. O entendimento do processo de

evolução institucional, e a interação deste com a evolução de tecnologias e da estrutura empresarial e industrial são essenciais para o processo de desenvolvimento econômico.

Freeman e Perez (1988) arguem que há muita relevância em conhecer-se quais as novas tecnologias importantes e as indústrias de rápido crescimento, conjunto chamado de sistema tecnológico. Este será único e serão distintos os efeitos sobre P&D público e privado, estratégias de investimento, políticas governamentais e mudanças institucionais, os quais são requisitos para desenvolver o próprio sistema tecnológico.

Os autores definem o conceito de paradigma técnico-econômico, como a combinação de inovações de produtos e processos, técnicas, organizacionais e gerenciais, todas inter-relacionadas, que englobam um salto em produtividade para toda a atividade econômica, abrindo uma ampla gama de oportunidades de investimentos e lucros. Esta mudança de paradigma implica em uma combinação única de vantagens técnicas e econômicas. Demonstram os autores que uma das características principais do padrão de difusão de um novo paradigma técnico-econômico é a propagação, a partir das indústrias ou áreas de aplicação para um conjunto muito maior de indústrias e serviços, além da economia como um todo. Ao observarem os principais períodos de maior intensidade da atividade econômica nos séculos XIX e XX, identificam a ocorrência de difusão dos principais novos paradigmas técnico-econômicos de cada época na economia mundial, e as maiores depressões como os períodos nos quais as estruturas social e institucional se ajustam à ascensão de novas tecnologias. Concluem desse estudo que a análise do desenvolvimento econômico envolve o surgimento de novas tecnologias, a ascensão e declínio de indústrias, os principais investimentos em infraestrutura, mudanças na localização internacional de lideranças industriais e tecnológicas, além de outras mudanças estruturais relacionadas, como, por exemplo, de capacidades e composição da força de trabalho e de estruturas gerenciais das empresas.

Quanto aos novos paradigmas técnico-econômicos, Freeman e Perez (1988) entendem que o surgimento ocorre gradualmente como um novo tipo ideal de organização produtiva, abrindo uma variedade sem precedentes de novas oportunidades de investimentos. Isto acarreta em uma mudança radical do senso comum, tendendo a difundir-se tão rapidamente quanto as condições permitirem, substituindo o padrão de investimento do antigo paradigma. Assim, o novo

paradigma técnico-econômico estende-se à mudança técnica por si só, uma vez que acarreta na reestruturação de todo o sistema produtivo.

Tão logo se consolida, o novo paradigma técnico-econômico envolve, dentre outros, uma nova boa prática em termos de organização das firmas, um novo perfil de habilidade da força de trabalho, um novo padrão na localização de investimentos, tanto nacional quanto internacionalmente, uma tendência de surgimento de pequenas empresas inovadoras e empreendedoras capazes de entrar nos novos ramos da economia em rápida expansão, além de, em alguns casos, iniciar setores de produção inteiramente novos, e um novo padrão de consumo de bens e serviços e novos tipos de distribuição e de comportamento do consumidor.

O período de transição entre um paradigma técnico-econômico e outro é caracterizado por uma profunda mudança estrutural na economia, a qual requer uma igualmente profunda transformação no quadro institucional e social. Torna-se necessária uma reacomodação em larga escala do comportamento social e das instituições para adequação aos requerimentos e potenciais de uma mudança que já ocorre em considerável extensão em algumas áreas da esfera técnico-econômica. Esta reacomodação resulta de um processo de experimentação e adaptação, por meio de uma variedade de mudanças sociais e políticas, tanto no nível nacional quanto internacional, e a combinação apropriada de mecanismos regulatórios cria uma onda de novos investimentos, o que possibilitará a implantação completa do novo paradigma. Trata-se de um processo conflituoso e que processa-se de forma muito desigual nos diferentes contextos políticos e culturais de cada país, podendo exercer uma considerável influência na mudança de padrão de liderança tecnológica.

Associa-se as principais mudanças nos paradigmas técnico-econômicos do passado a mudanças na divisão internacional do trabalho e na liderança tecnológica internacional. Tal perspectiva ampara-se na conclusão de que novos entrantes são mais hábeis em realizar as mudanças sociais e institucionais necessárias que os líderes, os quais possuem estruturas estabelecidas. No entanto, ressaltam os autores, países carentes das capacidades educacionais, gerenciais, de P&D e de planejamento mínimas necessárias estão em desvantagem ainda maior na competição internacional.

Dosi (1982) parte do princípio de que o processo inovativo não pode ser explicado por uma motivação unidirecional causadora do surgimento de um novo

paradigma tecnológico. Ao contrário, procura estabelecer um modelo suficientemente geral para definir o processo de seleção de um novo paradigma tecnológico em detrimento de outros possíveis, o qual leva em consideração a interação entre avanços científicos, fatores econômicos, variáveis institucionais e dificuldades não solucionadas pelas tecnologias estabelecidas. O autor parte de duas abordagens básicas: as que consideram as forças de mercado como as principais propulsoras de mudança técnica, as chamadas teorias *demand-pull*, e aquelas nas quais a tecnologia é definida como um fator autônomo, conhecidas como teorias *technology-push*.

Aprofundando-se um pouco mais nas teorias *demand-pull*, o motivador causal da atividade inovativa é o reconhecimento de necessidades pelo mercado, a partir do qual se iniciam os esforços tecnológicos a fim de satisfazê-las. O argumento básico mantém que normalmente conhece-se a direção na qual o mercado direciona a atividade inovativa. Assim sendo, implica-se que as opções de escolha estão dadas e os resultados de cada escolha são conhecidos. De acordo com o autor, três questões básicas dificultam a sustentação das teorias *demand-pull*: a percepção do processo de inovação como reativo frente às condições mercadológicas, a dificuldade de explicar por que e quando um determinado avanço tecnológico ocorreu em detrimento a outros e a negligência quanto às mudanças na capacidade de inovar ao longo do tempo que não apresentam relação direta com mudanças nas condições de mercado. Por outro lado, nas teorias *technology-push*, é desconsiderado o papel dos fatores econômicos na definição da direção do processo inovativo.

Ao criticar as duas abordagens, o autor propõe uma terceira interpretação da atividade inovativa. Inicialmente, define tecnologia como um conjunto que inclui conhecimentos práticos e teóricos, know-how, métodos, procedimentos, experiências baseadas em sucessos e fracassos, além de dispositivos e equipamentos. A partir desta definição, afirma que a tecnologia possui um conjunto limitado de possibilidades e de desenvolvimentos futuros, e conclui que há pouca diferença conceitual entre tecnologia e ciência. Dessa forma, partindo da premissa de que há similaridades entre a natureza e os procedimentos científicos e tecnológicos, o autor determina, paralelamente à paradigma e trajetória científicos, os conceitos de paradigma tecnológico e trajetória tecnológica. O primeiro é um modelo e um padrão de solução para problemas tecnológicos determinados,

baseado em princípios derivados das ciências naturais, como também em tecnologias materiais. O segundo é o padrão de progresso em um determinado paradigma tecnológico.

A fim de entender como um paradigma tecnológico emergiu em detrimento a outras possibilidades, considerando-se a cadeia ciência-tecnologia-produção, são propostos os fatores econômicos, institucionais e sociais como dispositivos seletivos, os quais baseiam-se em critérios como viabilidade, possibilidades de negócios, rentabilidade e capacidade de economia de custos, em especial, neste último caso, o potencial de economia de trabalho. Sobre estas bases bastante gerais, ainda podem existir diversas possibilidades de paradigmas tecnológicos a serem selecionados. Considerando a incerteza associada aos resultados de cada um, tanto em termos de sucesso econômico como tecnológico, dificilmente é possível compará-los e classificá-los *ex ante*. Assim, outras variáveis mais específicas tendem a operar como forças capazes de direcionar o desenvolvimento tecnológico, como os interesses econômicos das organizações envolvidas nos processos de P&D nas novas áreas tecnológicas, a história tecnológica destas organizações e os respectivos campos de especialização, além de variáveis institucionais como agências públicas ou organizações militares.

Ao analisar a influência particular de instituições públicas sobre a ascensão de novas tecnologias, identifica que não existe uma regra geral, mas observa que tais ocorrências demonstram fraqueza nos mecanismos de mercado como fatores seletivos para a direção tecnológica, especialmente nos estágios iniciais de uma indústria. Mesmo onde há intensa influência institucional, há diferentes possibilidades tecnológicas, um processo de busca incerto, diferentes organizações, firmas e indivíduos envolvidos em diferentes soluções tecnológicas. Desta forma, não há apenas uma competição entre a nova tecnologia e a antiga que se pretende substituir, mas também entre diferentes alternativas de abordagens tecnológicas.

De forma mais geral, os padrões de conflito industrial e social também participam do processo de seleção de novos paradigmas tecnológicos, ambos podendo atuar como critérios negativos, influenciando na exclusão de possibilidades de desenvolvimento, ou como critérios positivos, selecionando determinadas tecnologias. Quando ocorre a ascensão de novas tecnologias, observa-se, segundo o autor, novas companhias buscando explorar diferentes inovações tecnológicas. Neste momento, o mercado funciona como um sistema de recompensas ou

penalizações, por meio da seleção de alternativas. A existência de atores dispostos ao risco é fundamental para o processo de tentativa e erro associados à busca por novas trajetórias tecnológicas.

Observa ainda Dosi (1982), que uma característica permanente são as mudanças no ambiente econômico, mas que tais mudanças simplesmente estimulam o progresso técnico dentro de uma trajetória tecnológica. A busca pelo progresso tecnológico surge tanto em relação a novas oportunidades advindas do desenvolvimento científico, quanto à crescente dificuldade de avanço em determinada trajetória tecnológica, seja por razões tecnológicas, econômicas, ou ambas.

Portanto, nas conclusões do autor, o processo de busca e seleção de novos paradigmas tecnológicos é uma interação complexa entre alguns fatores econômicos fundamentais, como a busca por novas oportunidades de lucro e novos mercados e a tendência de redução de custos e de automação, dentre outros, aliados a poderosos fatores institucionais, como interesses e estrutura das firmas existentes e efeitos de agências governamentais, por exemplo. O autor ressalta que a produção dos principais avanços tecnológicos resultaram de esforços organizados de P&D, e não da inventividade individual. Além disso, o período de ascensão de novas tecnologias é normalmente caracterizado pelo surgimento de novas firmas, mesmo nos casos em que os principais avanços tecnológicos são originalmente produzidos em firmas e instituições já estabelecidas.

Posteriormente, há uma fase de maturidade da mudança tecnológica, na qual a trajetória está estabelecida. Nesta fase, o progresso técnico, incluindo a geração de inovações, torna-se mais endógeno à dinâmica da economia. Neste momento de estabelecimento de um paradigma tecnológico definido, ocorre a internalização nas empresas das externalidades relacionadas à atividade de inovação, capitalizando as experiências de tentativas, sucessos e falhas. Portanto, o processo de inovação afeta a estrutura industrial e molda a transformação desta mesma estrutura.

## 1.2 APROPRIAÇÃO DE INOVAÇÕES TECNOLÓGICAS

Teece (1986) define a empresa inovadora como a primeira a comercializar um produto ou processo no mercado. O autor ressalta que a observação das inovações ao longo da história, em diferentes setores industriais, mostra que a vantagem competitiva obtida por meio do pioneirismo pode ser suplantada por empresas seguidoras, ou por aquelas que possuem capacidades necessárias para o inovador. Assim, chegar primeiro ao mercado, desenvolvendo novos produtos que vão ao encontro das necessidades de consumidores, pode ser garantia de sucesso para o produto em si, mas não necessariamente para a empresa inovadora. Portanto, a divisão de lucros advindos de determinada inovação pode ser compartilhada entre o inovador, consumidores, fornecedores e seguidores ou imitadores. Assim sendo, as firmas devem considerar uma estratégia para apropriarem-se de inovações, ou seja, uma série de atividades que envolvem relacionarem-se umas com as outras, tais como empreendimentos conjuntos, acordos de produção, arranjos de distribuição e licenciamento de tecnologia.

Possas (1997) identifica a concorrência por meio da introdução de inovações como um processo seletivo no qual é possível estabelecer estratégias. A finalidade destas é fazer com que o competidor se sobressaia aos demais, ou que possua alguma característica diferenciadora em relação aos concorrentes. Na impossibilidade de sobressair-se, procura-se imitar os demais.

De acordo com Teece (1986), deve haver um processo de convergência entre dois campos de atuação das empresas: a inovação e a estratégia. Este processo é sustentado por três pilares: o regime de apropriação, o paradigma do desenho dominante e os ativos complementares.

O regime de apropriação é definido como o conjunto de fatores relacionados ao ambiente de negócios, excluindo-se a estrutura da firma e do mercado, que determinam a capacidade do inventor em obter os ganhos gerados por uma invenção. Ademais, as dimensões que definem o regime de apropriação são a natureza da tecnologia e a eficácia dos mecanismos legais de proteção, nomeadamente os segredos industriais e os direitos de PI.



Mello (2009) identifica que os fatores que influenciam a natureza da tecnologia são os custos e tempo necessários para introduzir uma imitação da inovação no mercado e o grau em que os conhecimentos relacionados a uma inovação são tácitos ou codificáveis, o que implica em investimentos em P&D e em obtenção de conhecimentos.

Adicionalmente, de acordo com Teece (1986), o nível em que os conhecimentos associados a uma inovação são tácitos ou codificáveis afeta o grau de facilidade de imitação.

Quanto aos direitos de PI, Mello (2009), ao analisar especificamente as patentes, aponta diversos estudos que demonstram que este instrumento é utilizado em diferentes tipos de estratégias empresariais, mesmo em setores da economia nos quais este mecanismo de proteção não é relevante para apropriação direta de lucros a partir de inovações. Algumas formas de benefício no patenteamento de inovações identificadas pela autora são um portfólio de patentes que garanta ao proprietário uma posição estratégica com poder de barganha em negociações de acordos de licenciamento, condição de acesso a mercados, evidência de qualidade dos produtos e estabelecimento de um padrão de mercado para um produto.

Combinando as dimensões de natureza de tecnologia e eficácia dos mecanismos legais de proteção, Teece (1986) determina que um regime de apropriação será mais forte quando há prevalência de conhecimentos tácitos e a proteção legal é eficaz, e será mais fraco quando prevalecem os conhecimentos codificáveis e a proteção legal é ineficaz, havendo a ocorrência de níveis intermediários.

O paradigma do desenho dominante está diretamente atrelado aos conceitos de paradigma tecnológico e trajetória tecnológica, propostos por Dosi (1982) como uma abordagem de interpretação da atividade inovativa. Partindo da premissa de que o processo competitivo por inovações estabelece trajetórias tecnológicas, o autor defende que a competição entre as empresas se manifesta entre os distintos desenhos. Após um considerável processo de tentativa e erro, um limitado conjunto de desenhos, ou até mesmo apenas um, emerge como o promissor desenho dominante. Ressalta o autor que o processo inovativo não é interrompido quando emerge o desenho dominante. Ao contrário, a estabilização de um desenho de um produto pode desencadear, por exemplo, na busca de um processo de redução de custos de produção daquele produto. Para Teece (1986), o paradigma do desenho

dominante possui fundamental relevância na distribuição de lucros entre inovadores e seguidores ou imitadores. O inovador será responsável pelas rupturas fundamentais, além do desenho básico do produto. No entanto, especialmente para as inovações de fácil imitação, os seguidores podem concorrer modificando o produto original, mesmo que mantidas as bases fundamentais do inovador. Sendo assim, ao emergir o desenho dominante, o inovador pode estar em desvantagem em relação ao seguidor. Existe a possibilidade de que o desenho deste seja adotado como padrão de mercado, enquanto o do inovador seja descartado.

Finalmente, os ativos complementares são serviços ou capacidades especializados como marketing, fabricação e suporte pós-vendas. A inovação pode ser traduzida como um conhecimento tecnológico, o qual pode ser parcialmente tácito e parcialmente codificado. A geração direta de lucros a partir da inovação se manifesta em alguma forma de venda deste conhecimento no mercado. Além disso, complementa o autor que o sucesso na comercialização de uma inovação está associado à conjunção entre o conhecimento relacionado à inovação e a outras capacidades, as quais se traduzem nos ativos complementares. A necessidade de acesso aos ativos complementares para a obtenção de lucros de determinada inovação leva as empresas a estabelecerem alianças estratégicas colaborativas.

Nas conclusões do autor, as indústrias que caracterizam-se por regimes de apropriação fortes são exceções. Portanto, posições estratégicas quanto aos ativos complementares relacionados à uma inovação tendem a determinar os competidores que serão bem sucedidos na obtenção dos lucros advindos daquela inovação, sejam tais competidores inovadores ou imitadores, empresas estabelecidas ou entrantes no mercado. O posicionamento frente às capacidades necessárias para comercialização da inovação é obtido por meio de alianças estratégicas, relações contratuais, ou da internalização daquela capacidade na própria firma, e a seleção de como uma empresa deve posicionar-se depende da especialização do ativo complementar em questão. Além disso, as estratégias adotadas devem levar em conta as diferenças do processo de competição antes e após o surgimento do desenho dominante.

Pisano (2006), considerando o desenvolvimento do ambiente de negócios e as mudanças na natureza dos direitos de propriedade intelectual, propõe uma nova visão quanto à posição das empresas frente ao regime de apropriação e aos ativos complementares. O regime de apropriação, antes exógeno, desafiava a firma a

desenvolver uma estratégia apropriada quanto a posições relativas a ativos complementares, seja por integração vertical ou alianças. Defende o autor que, atualmente, o regime de apropriação não está dado, mas é resultado das próprias estratégias das firmas. Ou seja, a partir do acesso que possuem aos ativos complementares necessários, as empresas buscam influenciar o regime de apropriação a fim de maximizar a utilização de seus ativos na obtenção dos lucros de uma inovação. Exemplos são a estratégia do compartilhamento entre empresas do resultado de pesquisas, ou o interesse por um regime de apropriação fraco por empresas que buscam a exploração de uma inovação por meio da imitação. Assim, conclui o autor que regimes de apropriação vantajosos não são necessariamente aqueles classificados como fortes, como se verifica por exemplo no caso empresas que lucram por meio do uso estratégico de software de código aberto.

A abordagem de Teece (1986) para o regime de apropriação nomeia as patentes e o direito autoral como os direitos de PI que determinam a eficácia dos mecanismos legais de proteção. Desta forma, pode-se discutir se as marcas, que também se constituem em ativos de PI, devem ser entendidas como um ativo complementar ou inseridas no contexto do regime de apropriação. Independente do entendimento adotado, considerando a importância que os direitos de PI desempenham para o ambiente empresarial, as marcas devem ser consideradas na análise do processo de convergência entre inovação e estratégia.

Desta forma, é relevante o entendimento das estratégias de apropriação das inovações adotadas pelo mercado em que determinada empresa atua. Com base neste conhecimento, pode-se adotar uma estratégia competitiva adequada para uma determinada empresa. Para tal, é imprescindível que o estudo seja realizado sobre estratégias adotadas na atualidade.

## 2 INOVAÇÕES DE TECNOLOGIA DA INFORMAÇÃO NO CONTEXTO DO SEQUENCIAMENTO GENÉTICO HUMANO

### 2.1 CONTEXTUALIZAÇÃO DO SEQUENCIAMENTO GENÉTICO HUMANO

O tema do sequenciamento genético humano passa a ser objeto de discussão na década de 1980, com a divulgação pelo Departamento de Energia (DOE) dos EUA do objetivo de mapear o genoma humano. Com este objetivo e previsão de término em 2005, inicia-se formalmente em 1990, o Projeto Genoma Humano (PGH), uma iniciativa do poder público americano patrocinado pelo National Institute of Health (NIH) e pelo DOE. Posteriormente, uniram-se ao projeto laboratórios da Europa, Austrália e Japão, surgindo dessa forma o organismo de coordenação internacional Human Genome Organization, ou HUGO, com o objetivo de organizar o conhecimento adquirido no PGH em um banco de dados centralizado, o *Genome Database*. Ao final da década de 1990, a missão institucional do HUGO era facilitar e coordenar o mapeamento, sequenciamento e análise do genoma humano, além de promover a aplicação dos conhecimentos obtidos ao melhoramento da saúde humana. Posteriormente, a organização desempenha também o papel de disseminação das análises e o fornecimento de diretrizes responsáveis para as aplicações e implicações do genoma humano. (GOLDIM e MATTE, 2000).

Entre o início do projeto e a publicação de um esboço da sequência do genoma humano, em fevereiro de 2001, houve um esforço de mais de 200 cientistas e três bilhões de dólares para que se realizasse a leitura bruta de três bilhões de bases de DNA que compreendem o material genético humano (HAYDEN, 2014). Este esboço foi publicado em um artigo intitulado *Initial sequencing and analysis of the human genome* por um consórcio internacional liderado nos Estados Unidos pelo NHGRI, agência subordinada ao NIH, e pelo DOE. Este consórcio, conhecido como *International Human Genome Sequencing Consortium*, reunia ainda outras vinte instituições entre universidades e centros de pesquisa dos Estados Unidos, Reino Unido, França, Alemanha, Japão e China. O mesmo consórcio anuncia, em abril de 2003, a versão final completa do sequenciamento do genoma humano, que

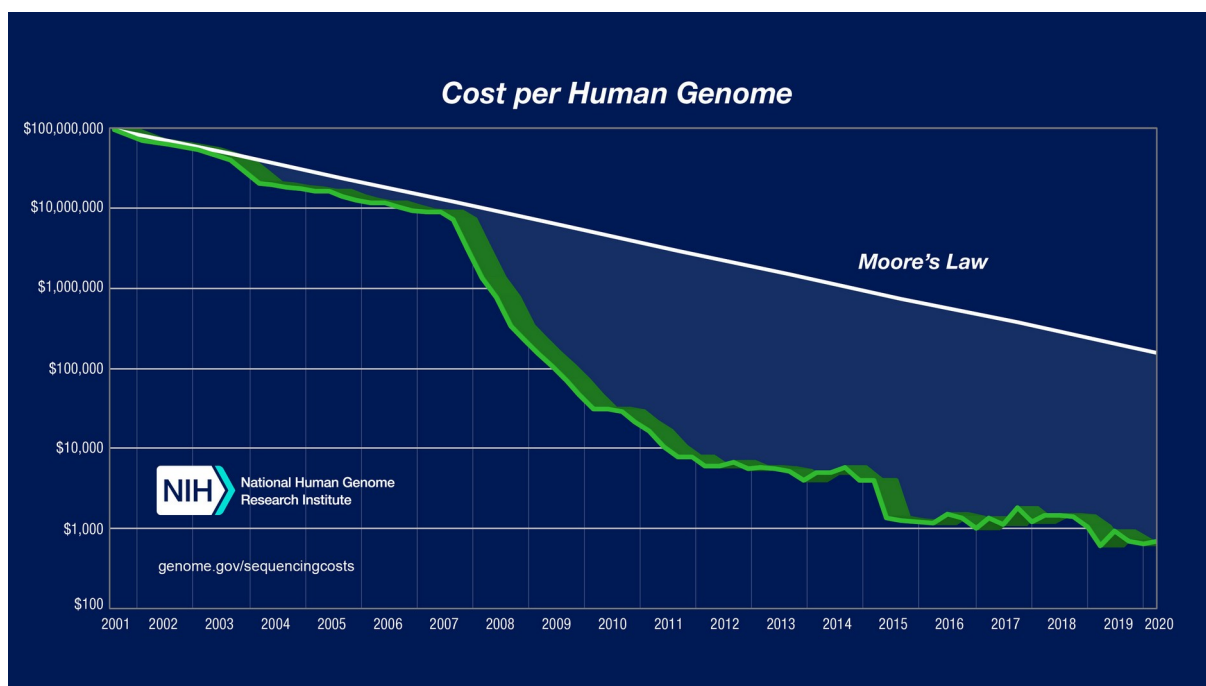
caracteriza formalmente o final do PGH, dois anos antes do previsto (NATIONAL HUMAN GENOME RESEARCH INSTITUTE, 2006; NATIONAL HUMAN GENOME RESEARCH INSTITUTE, 2018).

Prevvia-se que o PGH revelaria o livro de instruções da vida, porém, na realidade, não era ainda possível interpretar como as instruções codificadas no DNA influenciavam, por exemplo, nas questões de saúde. Para tal, seria necessário realizar comparações entre os genes de milhões de pessoas. Naquela época, predominava o método de sequenciamento Sanger, um processo lento e trabalhoso. Além disso, uma única empresa fornecia a maior parte dos sequenciadores a uma limitada gama de clientes, que em geral se constituíam em grandes laboratórios financiados pelo governo, havendo, desta forma, poucos incentivos para a evolução da tecnologia. Como consequência, a partir de então, passa a ser tema de discussão no NHGRI a que custo o sequenciamento genético humano se tornaria uma atividade rotineira tal qual uma solicitação médica para o diagnóstico de um paciente. Assim, foi iniciado em 2004, um programa de premiações oficialmente chamado de Advanced Sequencing Technology Awards. Executado por meio do NHGRI, ficou largamente conhecido como o programa genoma de mil dólares, pois determinou-se esse valor como o custo alvo do sequenciamento genético humano.

Considerando-se o estado da arte da tecnologia Sanger, tratava-se de um objetivo bastante ambicioso e arriscado. Duas características foram fundamentais para que o programa pudesse ser bem sucedido dentro dessa perspectiva. A primeira a agilidade alcançada por meio da concessão de pequenos prêmios a trabalhos considerados arriscados, porém promissores. A segunda a mescla de pesquisa básica e aplicada em um único programa, por meio do financiamento de projetos de pesquisa básica em métodos totalmente novos de sequenciamento, assim como de projetos de pesquisa industrial para desenvolver estas novas tecnologias para uso comercial. Desta forma, o programa forneceu apoio a novas empresas que passaram a competir com o monopólio da tecnologia Sanger. Além disso, por meio de um congresso anual de progresso, obrigatório aos beneficiados pelo programa, forçava-se os concorrentes a compartilharem conhecimentos (HAYDEN, 2014). Para Mardis (2011), a disponibilidade de múltiplos instrumentos representou uma mudança de paradigma comparada à década anterior, durante a qual um único instrumento Sanger fornecido por uma empresa detinha o domínio do mercado. Indica a autora que estas diversas tecnologias inovadoras foram,

inicialmente, desenvolvidas com financiamento do NHGRI, por meio do programa genoma de mil dólares.

O National Human Genome Research Institute (2020) mantém atualizados os dados sobre os custos do sequenciamento genético humano. É possível observar que, em 2007, o montante para o sequenciamento de um indivíduo estava na ordem de 10 milhões de dólares e que, em 2015, já havia atingido o patamar desejado. A relevância destes dados pode ser medida comparando-se a redução do custo no tempo com a tendência empírica que ficou conhecida na indústria de computadores como a Lei de Moore. Em 1975, o então presidente da Intel, Gordon Moore, baseado em simples observação, demonstrou que, desde a década de 1960, a capacidade computacional dobrava enquanto o preço para atingir este nível era reduzido pela metade a cada dois anos. Moore sugeriu ainda que esta tendência seria mantida pelo menos até 1980 (MOORE, 1975). No gráfico 1, o qual demonstra o custo para o sequenciamento de um genoma humano ao longo do tempo, observa-se como a redução do custo deste sequenciamento superou a Lei de Moore a partir de 2007.



**Gráfico 1 — Custo do sequenciamento do genoma humano: National Human Genome Research Institute (2020).**

Ainda de acordo com o National Human Genome Research Institute (2020), até outubro de 2007, conforme mencionado anteriormente, predominava a tecnologia Sanger para a realização do sequenciamento genético, a qual posteriormente ficou conhecida como tecnologia de primeira geração. O custo do

sequenciamento genético apontado no gráfico 1 entre 2001 e 2007 está relacionado à utilização da tecnologia Sanger. Aponta-se que o sucesso do programa genoma de mil dólares ocorreu devido ao financiamento de diferentes empresas ou grupos acadêmicos, cujos projetos buscavam desenvolver tecnologias mais rápidas e de menor custo, ao invés de selecionar uma tecnologia para apostar. Desta forma, realizou-se uma distribuição dos investimentos entre trabalhos acadêmicos e industriais diversos o suficiente para permitir o progresso tecnológico, criando um desenvolvimento paralelo e concorrencial de tecnologias (HAYDEN, 2014). Esta abordagem permitiu o surgimento de uma série de tecnologias mais rápidas e de menor custo para o sequenciamento do genoma humano. A tecnologia Sanger é considerada como sendo a primeira geração de sequenciadores, e as novas tecnologias ficaram conhecidas como *next-generation sequencing*, ou NGS (Metzker, 2010).

O National Human Genome Research Institute (2020) afirma que, a partir de janeiro de 2008, os centros de sequenciamento financiados pelo instituto passaram a adotar as tecnologias NGS em substituição à tecnologia Sanger. A partir deste momento, conforme demonstra o gráfico 1, ocorre uma súbita e profunda queda deste custo.

As tecnologias NGS conseguiram uma mudança significativa em termos de rendimento, permitindo o sequenciamento de milhões de fragmentos de DNA em paralelo a um custo significativamente menor (MARDIS, 2011). De acordo com Metzker (2010), o principal diferencial alcançado pelas tecnologias NGS é a capacidade de produzir um grande volume de dados a baixo custo, sendo possível realizar em alguns casos mais de um bilhão de leituras curtas em uma única execução.

A fim de quantificar este volume de dados, Robison (2014) propõe calcular o tamanho computacional de um genoma humano. A análise mais simples é considerar a utilização de dois bits para cada composto orgânico celular (A, C, G, T). Assim, a sequência completa dos três bilhões de pares de um genoma humano consumiria seis bilhões de *bits*, ou pouco mais de 700 megabytes. No entanto, o autor observa que cada leitura curta de um sequenciador NGS lê, aproximadamente, 100 pares, e que cada par, em média, é coberto trinta vezes ao final do sequenciamento. Além disso, tais leituras não geram apenas uma sequência de letras (A, C, G, T), mas registram também os respectivos índices de qualidade. Por

essa razão, segundo o autor, cada composto consome dois *bytes* (sequência de oito *bits*), e não apenas um *bit*. Assim, assumindo uma cobertura média de trinta leituras de cada par, tem-se 180 bilhões de bytes para um genoma, ou 180 gigabytes.

Para Brevern et al. (2015), a disponibilidade de ferramentas a um custo mais baixo permitiu a utilização massiva das tecnologias NGS, aumentando ainda mais a escala de volume de dados produzido. Os autores identificam ainda que outra característica dos dados biológicos gerados pelas tecnologias NGS é a heterogeneidade, visto que estão distribuídos em distintas bases de dados e cada fornecedor desenvolveu um formato próprio. Desta forma, arguem que tais questões devem ser superadas a fim de possibilitar a extração de informação relevante e confiável.

Para Merelli et al. (2014), há a necessidade de soluções para o armazenamento dos dados gerados em grandes proporções. Além disso, defendem os autores que o objetivo maior é transformar estes dados em informação utilizável e conhecimento. Dessa forma, o desafio é realizar a integração de dados biológicos em tempo real, possibilitando a mineração da informação.

Eisenstein (2015) cita alguns programas lançados pela comunidade científica no Reino Unido, Estados Unidos e China com o propósito de realizar o sequenciamento genético populacional em escala, a fim de obter conhecimento clínico a respeito do genoma humano. O autor defende que não apenas o armazenamento é uma questão a ser resolvida, mas deve-se dar ainda maior importância ao processamento, análise e interpretação destes dados. Segundo o autor, o sequenciamento genômico em escala populacional produziu 40 milhões de *gigabytes* de dados brutos em 2015, além de apresentar estimativas de que esta capacidade pudesse dobrar a cada sete a doze meses. No entanto, o dado bruto por si só, não é suficiente para ser interpretado, e a combinação das variantes entre os indivíduos eleva exponencialmente a questão de armazenamento.

O objetivo a ser alcançado por estes grandes projetos de sequenciamento é possibilitar a análise clínica do genoma humano por meio da identificação das mutações responsáveis por doenças raras, justamente por meio do estudo daquela combinação das variantes entre os indivíduos de diferentes populações. Desta forma, o foco da genômica clínica é a identificação das variantes genéticas de base única (SNV, do inglês *single nucleotide variants*), ou seja, uma alteração do código genético capaz de interromper a função do gene em questão. Um destes grandes



projetos de sequenciamento se intitula Projeto 1000 Genomas (*The 1000 Genomes Project*, no original), lançado em 2008, com o objetivo de caracterizar mais de 95% das variantes genéticas com frequência de pelo menos 1% nas populações estudadas, por meio da análise das sequências genômicas de pelo menos mil participantes anônimos (THE INTERNATIONAL GENOME SAMPLE RESOURCE, 2019; THE 1000 GENOMES PROJECT CONSORTIUM, 2010).

Os resultados obtidos desde a fase piloto do projeto foram disponibilizados à comunidade científica por meio da publicação na base de dados de variantes genéticas humanas de consulta pública de livre acesso conhecida como dbSNP (THE 1000 GENOMES PROJECT CONSORTIUM, 2010). Esta base de dados foi criada pelo Centro Nacional de Informação Biotecnológica (NCBI, do inglês *National Center for Biotechnology Information*), instituição do governo americano, ramo da estrutura do NIH, em resposta à necessidade de um catálogo de variantes genéticas, identificada justamente devido aos projetos de sequenciamento de larga escala (SHERRY, WARD e SIROTKIN, 2001). A maior classe de variantes genéticas contidas no dbSNP são os polimorfismos de nucleotídeo único (SNP, do inglês *single nucleotide polymorphisms*), fato que reflete o próprio nome da base de dados (SHERRY et al., 1999; SHERRY, WARD e SIROTKIN, 2001). A versão 129 do dbSNP, disponibilizada em abril de 2008, era a mais recente no momento do lançamento do Projeto 1000 Genomas, e contém aproximadamente quatorze milhões de SNPs humanos catalogados (THE 1000 GENOMES PROJECT CONSORTIUM, 2010). A fim de dimensionar o crescimento do dbSNP ao longo dos anos, a versão 151<sup>2</sup>, disponibilizada em outubro de 2017, contém mais de 660 milhões de SNPs humanos catalogados (NCBI, 2019).

Além do depósito de resultados, as bases de dados genéticos públicas de acesso livre são utilizadas para efeito de comparação e validação dos resultados obtidos. Entre as publicações dos resultados da fase piloto e do fim do Projeto 1000 Genomas, são citadas como fonte de informações o próprio dbSNP, como também o *Human Gene Mutation Database* (HGMD), o dbVar e o ClinVar. Estas bases de dados são mantidas por institutos de pesquisa em biotecnologia, como o Instituto de

---

2 Após a versão 151, foram disponibilizadas ao público mais duas versões do dbSNP, 152 e 153, respectivamente em dezembro de 2018 e agosto de 2019. A partir da versão 152, adotou-se a nomenclatura dbSNP 2.0 para a base de dados (<https://www.ncbi.nlm.nih.gov/projects/SNP/buildhistory.cgi>). O NCBI publicou em junho de 2018 (<https://ncbiinsights.ncbi.nlm.nih.gov/2018/06/15/dbsnp-updates-json-refsnp-report-api/>) algumas determinações a respeito da base de dados, como o armazenamento apenas de dados genéticos humanos e mudanças no formato dos resultados de busca. A partir da versão 152 não há possibilidade de extração das estatísticas referentes à base de dados.

Medicina Genética da Universidade de Cardiff, no Reino Unido, responsável pelo HGMD, ou o NCBI, responsável dbVar e ClinVar (THE 1000 GENOMES PROJECT CONSORTIUM, 2010; THE 1000 GENOMES PROJECT CONSORTIUM, 2015).

Estimativas apontadas por The 1000 Genomes Project Consortium (2015) indicam que 86% das variantes genéticas restringem-se a um único grupo continental, e cerca de 10% das variantes observadas em uma determinada população são particulares do grupo em estudo.

Apesar da grande quantidade de variantes genéticas catalogadas pelos projetos de sequenciamento em larga escala, e da grande quantidade de variantes genéticas compartilhadas ao redor do mundo, Ismail et al. (2018) defendem que existem populações ainda pouco estudadas, tanto em projetos de sequenciamento em larga escala quanto individual, que representam uma oportunidade de descoberta de novas variantes genéticas para a comunidade científica. Em estudo realizado pelos autores, o sequenciamento genético de um indivíduo de origem sérvia foi confrontado com as variantes genéticas conhecidas e publicadas em bases de dados como o dbSNP, ClinVar, PharmGKB e HGMD. A partir da análise deste genoma, foram introduzidas dezenas de milhares de variantes genéticas até então desconhecidas.

Outro exemplo do desafio para a interpretação genômica é o fato de que não há uma base de dados genética da população brasileira. Recentemente foi noticiado o projeto DNA do Brasil, por meio do qual se pretende sequenciar o genoma de quinze mil brasileiros, e, por meio da interpretação de três mil destes genomas, descobrir as variantes genéticas características do Brasil (GOMES, 2019).

Assim, tornam-se questões críticas a infraestrutura dos sistemas de TI em termos de armazenamento e transferência de dados, a capacidade computacional de análise e métodos de visualização das informações geradas (METZKER, 2010; BREVERN et al., 2015).

A interação entre TI e as ciências biológicas tornou-se tão evidente e indispensável que a aplicação de técnicas computacionais para a análise de informações biomoleculares em larga escala estabeleceu uma nova disciplina conhecida como bioinformática (LUSCOMBE, GREENBAUM e GERSTEIN, 2001). Dessa forma, são relevantes as inovações tecnológicas ocorridas na área de TI que se aplicam ao contexto do sequenciamento genético humano.

## 2.2 INOVAÇÕES DE TI RELACIONADAS AO SEQUENCIAMENTO GENÉTICO HUMANO

Os sistemas gerenciadores de bancos de dados relacionais constituem a tecnologia predominante na questão de armazenamento e compartilhamento de dados para aplicações de negócios (STRAUCH, 2011). Codd (1970) descreve o modelo relacional de dados, comparando-o a outros tipos de modelos. O autor procura demonstrar algumas características e vantagens do modelo relacional. Por exemplo, a possibilidade de descrever os dados apenas por meio da estrutura natural destes. Ou seja, sem a necessidade de se adicionar qualquer estrutura para fins de representação de máquina. Outra vantagem apontada pelo autor é o fornecimento de uma base para uma linguagem de dados de alto nível, capaz de prover independência entre as aplicações utilitárias de um lado e, do outro, a representação de máquina e a organização dos dados.

A partir deste modelo, desenvolvem-se os bancos de dados relacionais. Esta tecnologia baseia-se no conceito de transação descrito por Eswaran et al (1976), a qual se constitui em uma sequência de ações agrupadas referentes a um processo executado em um banco de dados. Cada transação transforma um estado consistente em um novo estado consistente. Portanto, uma transação obrigatoriamente preserva a consistência do banco de dados.

Header e Reuter (1983) determinam que este conceito impõe que as ações que constituem uma transação sejam executadas de forma indivisível. Ou seja, todas as ações devem refletir no banco de dados ao final da transação, e, caso não se atinja o final da transação, nenhuma ação deve ser refletida. Para que a indivisibilidade da transação seja alcançada, os autores determinam que os bancos de dados devem possuir quatro propriedades: atomicidade, consistência, isolamento e durabilidade.

Atomicidade é o tratamento de cada transação como uma unidade. Desta forma, se ocorre uma falha em uma das ações que compõem a transação, esta falha como um todo, e o estado do banco de dados se mantém inalterado. A transação é bem sucedida apenas quando todas as ações são finalizadas com sucesso.

Consistência é a garantia de que uma transação executada até o final leva o banco de dados de um estado válido a outro, ou seja, nenhuma regra definida para o banco de dados pode ser corrompida por uma transação ilegal. A propriedade do isolamento garante que as transações não são afetadas pelos resultados daquelas executadas de forma concorrente, garantindo assim que o banco de dados possa retornar ao estado inicial. A durabilidade é a garantia de que as transações finalizadas têm resultados confirmados mesmo em caso de falhas no sistema de banco de dados. Os bancos de dados relacionais são caracterizados pelo acrônimo ACID para designar o conjunto das quatro propriedades que garantem a indivisibilidade da transação, visto que tais sistemas de bancos de dados foram desenvolvidos com base nestas características.

O desenvolvimento da Internet como uma plataforma de serviços desencadeou a necessidade para as empresas *Web 2.0*, como ficou conhecido este movimento, do tratamento de grandes e crescentes volumes de dados, além da infraestrutura para lidar com estas questões. Estas empresas, como Google e Amazon, por exemplo, se ocupam então com o desenvolvimento de novas tecnologias para o armazenamento de dados. Além da questão do volume, era necessário desenvolver em tais tecnologias a capacidade de manipular dados não estruturados, como, por exemplo, arquivos de processamento de texto, multimídia ou *e-mails* (LEAVITT, 2010).

Ao final da primeira década do século XXI, começam a surgir novas tecnologias de armazenamento distribuídas e não relacionais, as quais foram rotuladas com o termo NoSQL. SQL é o acrônimo para *Structured Query Language*, e se constitui em uma linguagem padrão estruturada de comunicação com os bancos de dados relacionais. O termo NoSQL já havia sido utilizado ao final da década de 1990 para designar uma solução específica de bancos de dados que omitia a interface SQL, porém ainda relacional.

A difusão do conceito NoSQL ocorre a partir de 2009, por meio de um evento ocorrido em São Francisco, Estados Unidos, com o objetivo de discutir soluções de bancos de dados de código fonte aberto, distribuídas e não relacionais. Posteriormente, houve o surgimento de diversas tecnologias de armazenamento de dados não relacionais distintas entre si, cujo objetivo comum era a solução de problemas para as quais os bancos de dados relacionais não se constituíam em uma tecnologia adequada. Todas estas soluções são caracterizadas como NoSQL,

mesmo que algumas apresentem algum tipo de compatibilidade com a linguagem SQL. Por estas razões, o termo NoSQL foi estendido para *Not Only SQL*, com o intuito de designar as soluções de armazenamento distribuídas com ausência de um modelo estrutural de dados, independentemente da linguagem utilizada para a extração de consultas. Desta forma, a característica principal das tecnologias NoSQL é comportar dados heterogêneos em um mesmo sistema de armazenamento de dados, além de lidar com o tratamento rápido a consultas por meio da omissão de controles estruturais (EVAN, 2009; STRAUCH, 2011; MOHAN, 2013).

Lai (2009) identifica que os bancos de dados NoSQL surgiram por meio do desenvolvimento interno de soluções de armazenamento pelas empresas *Web 2.0*. O objetivo destas soluções era atender às necessidades das próprias empresas *Web 2.0* de armazenar e processar grandes volumes de dados. Os produtos resultantes foram desenvolvidos a partir da influência das tecnologias *Bigtable* e *Dynamo*, publicadas pela Google em 2006 (CHANG et al., 2006) e pela Amazon em 2007 (DECANDIA et al., 2007), respectivamente.

Strauch (2011) lista uma série de motivações para o desenvolvimento e também para o uso destas novas tecnologias de armazenamento de dados. Uma dessas motivações é a necessidade de uma tecnologia com capacidade de processamento do grande volume de dados armazenados superior à dos bancos de dados relacionais. Outra motivação apontada pelo autor é a característica desejável de escalabilidade horizontal, ou seja, a possibilidade de se incluir mais máquinas no sistema de armazenamento de dados conforme o volume destes dados aumenta, visto que a necessidade de tratamento de dados em larga escala excede a capacidade de uma única máquina. O autor cita que a fragmentação de dados entre distintas máquinas no sistema de banco de dados relacional exige esforço operacional, além de uma infraestrutura de custo elevado. O autor sustenta que os bancos de dados NoSQL não se resumem a uma única tecnologia, mas existem distintas soluções para problemas específicos. Neste sentido, contrastam com os bancos de dados relacionais, desenvolvidos sob o conceito de solução única para qualquer tipo de problema (STONEBRAKER e ÇETINTEMEL, 2005).

Os produtos de bancos de dados relacionais desenvolvidos nos anos de 1970 e início dos anos de 1980 tinham o intuito de processar dados de negócios de forma mais eficiente que os sistemas hierárquico e de rede. Naquela época, este tipo de

processamento era a tarefa mais comumente atribuída aos sistemas de bancos de dados. Com o passar do tempo, novos mercados surgiram ou evoluíram, gerando novas necessidades para os sistemas de banco de dados, como, por exemplo, o armazenamento e processamento de grandes volumes de dados e o gerenciamento de textos e multimídia.

Também a forma de interagir com os bancos de dados evoluiu, de operadores efetuando consultas, a partir de terminais para potentes computadores pessoais conectados à Internet. Além disso, características de *hardware*, como capacidade e custo dos processadores e de memória permitiram conceber novas arquiteturas de sistemas de bancos de dados distintas daquela dos bancos de dados relacionais (STONEBRAKER e ÇETINTEMEL, 2005; STONEBRAKER et al., 2007b). A partir de uma série de avaliações comparativas, Stonebraker et al. (2007a) demonstraram que arquiteturas especializadas desenvolvidas para resolver um problema específico possuem desempenho mais eficiente que os bancos de dados relacionais, inclusive para o processamento de dados de negócios.

Obasanjo (2009) identifica a utilização do termo movimento NoSQL para descrever o uso de bancos de dados não relacionais pelas empresas de Internet. O autor identifica como pioneiras do movimento, três empresas de grande porte cuja necessidade de negócio envolve o tratamento de páginas na Internet de larga escala, a saber: Facebook, Google e Amazon, por meio dos respectivos projetos de bancos de dados não relacionais *Cassandra*, *BigTable* e *Dynamo*. O autor identifica outros produtos NoSQL desenvolvidos posteriormente por outras empresas, algumas de pequeno porte. Strauch (2011) complementa que estas empresas desenvolveram soluções internas a fim de solucionar os próprios requerimentos específicos com relação à escalabilidade e características de negócio, e que a maior parte dos produtos de armazenamento de dados NoSQL são adaptações das tecnologias *BigTable* ou *Dynamo*, desenvolvidas e publicadas, respectivamente, pela Google e Amazon, conforme citado anteriormente.

Para Strauch (2011), um dos princípios básicos para o desenvolvimento das tecnologias de armazenamento não relacionais é o Teorema de CAP, por meio do qual Brewer (2000) supõe que somente é possível para um sistema de dados distribuídos prover, simultaneamente, no máximo duas das seguintes propriedades: consistência, disponibilidade e tolerância à partição.

Gilbert e Lynch (2002) formalizaram os conceitos das três propriedades. Para os autores, os sistemas distribuídos foram inicialmente desenvolvidos com a capacidade de prover forte consistência de dados, pois eram dependentes das tecnologias de bancos de dados relacionais, e conseqüentemente, sujeitos à característica de transações ACID. No entanto, reforçam que o conceito de consistência para sistemas distribuídos é distinto daquele relacionado às transações de bancos de dados relacionais. No caso em questão, consistência é a existência de uma ordem sobre todas as operações de uma requisição de tal forma que esta se complete em um único instante. A alta disponibilidade é outra característica desejável nos sistemas distribuídos, especialmente os serviços disponibilizados via Internet. Disponibilidade é definida como a garantia de que toda requisição deve receber uma resposta. Finalmente, em um sistema distribuído em uma rede, é desejável que haja tolerância à falha, ou seja, mesmo que haja quebra de nós da rede, ou falha na comunicação entre nós, o sistema deve permanecer operante. Embora as três propriedades sejam desejáveis e esperadas em sistemas distribuídos, Gilbert e Lynch (2002) provaram a suposição de Brewer. Portanto, de acordo com o teorema de CAP, para o desenvolvimento de um sistema de dados distribuído deve-se realizar a seleção de duas das propriedades inerentes a estes sistemas. Brewer (2000) determina ainda que sistemas para os quais sejam determinantes a consistência e a tolerância à falha, se requer as propriedades ACID. Por outro lado, os sistemas para os quais disponibilidade e tolerância à falha são mais primordiais são caracterizados pelas propriedades BASE (*basically available, soft state, eventual consistency*). Strauch (2011) observa que a tecnologia *Dynamo* da Amazon enquadra-se na categoria BASE, visto que com o intuito de alcançar disponibilidade em um ambiente tolerante à falha, este sistema não oferece consistência estrita, ou seja, as gravações de dados não são visualizadas imediatamente por todos os leitores. Ainda para Strauch (2011), a Google adotou uma abordagem que não se caracteriza nem por ACID nem por BASE ao desenvolver a tecnologia *BigTable*, pois esta garante consistência e disponibilidade, porém não é capaz de operar plenamente quando há falha de nós da rede.

Os bancos de dados relacionais foram a tecnologia naturalmente selecionada para o armazenamento e difusão de informação nos laboratórios de pesquisa dedicados ao tema do sequenciamento genético. Tratava-se da tecnologia

dominante e estável desde o lançamento do PGH até o surgimento das tecnologias NGS (BREVERN et al., 2015).

No entanto, as motivações que levaram ao desenvolvimento dos bancos de dados NoSQL estão presentes na bioinformática. As principais características dos bancos de dados NoSQL que permitem a aplicação desta tecnologia no âmbito da bioinformática são as capacidades de armazenamento de grande volume de dados, de escalabilidade horizontal, de lidar com massivas e frequentes leituras e escritas de forma rápida, confiável e consistente, e um modelo flexível que admite dados heterogêneos.

Brevern et al. (2015), Have e Jensen (2013), Aniceto et al. (2014), Wang et al. (2014), Guimarães et al. (2015), Lima et al. (2016) e Schulz et al. (2016) discutem o uso dos bancos de dados NoSQL na bioinformática, inclusive para o tratamento das necessidades relacionadas aos dados gerados pelas tecnologias NGS. Nesta direção, exemplos da aplicação de bancos de dados NoSQL em projetos de bioinformática relacionados às tecnologias NGS são apresentados por Taylor (2010), O'Connor, Merriman e Nelson (2010), Bonnici et al. (2014), Stajdohar et al. (2017) e Sempéré et al. (2015). Todos estes trabalhos apresentam em comum a conclusão de melhor desempenho dos bancos de dados NoSQL comparados aos bancos de dados relacionais ao processar grandes quantidades de dados gerados pelas tecnologias NGS.

A partir da diversidade de soluções encontradas nos trabalhos mencionados, justifica-se avaliar as estratégias de apropriação adotadas em relação à tecnologia de bancos de dados NoSQL.



### 3 ESTRATÉGIAS DE APROPRIAÇÃO DOS BANCOS DE DADOS NOSQL

#### 3.1 COMERCIALIZAÇÃO DOS BANCOS DE DADOS NOSQL

Duas soluções de bancos de dados NoSQL consolidadas no mercado são o Apache HBase e o Apache Cassandra, dois produtos desenvolvidos pela *Apache Software Foundation* (ASF). Ambas são utilizadas em projetos de bioinformática relacionados às tecnologias NGS ou foram utilizadas em pesquisas a fim de demonstrar os benefícios da utilização da tecnologia NoSQL, conforme relatado em Taylor (2010), O'Connor, Merriman e Nelson (2010), Have e Jensen (2013), Aniceto et al. (2014), Wang et al. (2014) e Brevern et al. (2015). A ASF se constitui em uma comunidade que através da colaboração individual desenvolve e distribui software livre de código aberto, inclusive para uso comercial. A ASF é financiada por meio de doações de pessoas e de patrocínios de corporações. Dentre os patrocinadores da ASF estão a Google e a Microsoft, com aportes de U\$ 100 mil por ano e a IBM, com U\$ 40 mil por ano (THE APACHE SOFTWARE FOUNDATION, 2019).

Todo software livre de código aberto é distribuído por meio de uma licença específica para este tipo de produto. No caso da ASF, a instituição desenvolveu uma licença própria de software livre, desde 2004 em sua versão 2.0, a qual acompanha a cópia de qualquer distribuição de produto, incluindo-se, portanto, os bancos de dados NoSQL Apache HBase e Apache Cassandra. Trata-se de um documento de apenas nove sessões, no qual são tratadas, dentre outras, as questões relativas aos ativos de propriedade intelectual. O segundo item aborda o tema de direito de autor. Pelos termos da licença, todo autor de uma contribuição, definida como qualquer trabalho de autoria sobre a versão original ou suas modificações, concede uma licença perpétua, mundial, não exclusiva, sem custo, livre do pagamento de royalties e irrevogável para qualquer pessoa ou entidade reproduzir, modificar, demonstrar ou executar publicamente, sublicenciar e distribuir o trabalho ou suas modificações. O documento volta à questão das cópias e do direito de autor no quarto item, sob o título de redistribuição. A licença permite a uma pessoa ou entidade reproduzir e distribuir cópias do produto e suas modificações, com ou sem alterações, desde que

acompanhadas por uma cópia da licença, devendo ser mantidos todos os direitos autorais, patentes e marcas da fonte original. Caso sejam efetuadas alterações no produto, novas declarações a respeito de direitos autorais podem ser adicionadas à licença pelo autor. De acordo com o item três do documento, o autor de uma contribuição concede permissão, nos mesmos termos das licenças autorais no item dois, para produzir, vender, importar ou transferir qualquer reivindicação de patente relacionada à contribuição. Finalmente, a licença não concede permissão para o uso das marcas cujo detentor seja o autor da contribuição.

Outro produto de banco de dados NoSQL difundido no mercado é o MongoDB. Desenvolvido pela empresa 10gen, cujo nome posteriormente foi alterado para o mesmo nome do produto, trata-se de um software livre de código aberto para uso comercial, assim como as soluções da ASF. Da mesma forma que os produtos HBase e Cassandra da Apache, a utilização do MongoDB como ferramenta de armazenamento em soluções tecnológicas destinadas à análise de grandes volumes de dados na área da bioinformática é objeto de estudo e aplicação. Exemplos de estudos e aplicações neste sentido encontram-se em Lima et al. (2016), Guimarães et al. (2015), Stajdohar et al. (2017), Sempéré et al. (2015) e Schulz et al. (2016). Especificamente quanto à licença de distribuição, o MongoDB é distribuído sob a *GNU Affero General Public License*. Trata-se de uma licença desenvolvida pela *Free Software Foundation* (FSF), uma instituição não somente adepta do software livre de código aberto, mas manifestamente contrária ao software proprietário. A diferença principal desta licença é o conceito criado pela FSF de *copyleft*, definido pela própria instituição como uma imagem espelhada do *copyright*. Ou seja, a ideia não é ignorar as questões relacionadas aos direitos de autor, cópia e distribuição, mas utilizá-las de uma forma distinta para incentivar a contribuição (GNU AFFERO GENERAL PUBLIC LICENSE, 2016). Em resumo, os termos da licença garantem a qualquer indivíduo os direitos de usar, modificar e redistribuir o software ou qualquer alteração, porém os termos de distribuição não podem ser modificados. Desta forma, a licença não permite que sejam adicionadas novas declarações a respeito de direitos autorais pelo autor de uma modificação. Os termos referentes a marcas e patentes não diferem desta para a licença da ASF.

Sadalage e Fowler (2012) observam algumas características comuns aos bancos de dados denominados como NoSQL, dentre as quais o fato de serem geralmente projetos de software livre de código aberto. Strauch (2011) também

identifica que a maior parte destas soluções de armazenamento possuem esta característica. Mohan (2013) afirma que o desenvolvimento dos bancos de dados NoSQL partiu das empresas Web 2.0, as quais estão envolvidas no movimento de software livre de código aberto. O autor acrescenta ainda que bancos de dados NoSQL desenvolvidos posteriormente, foram fornecidos por empresas de menor porte também adeptas àquele movimento. As estratégias apoiadas em distribuição de software livre de código aberto dos bancos de dados NoSQL diferem totalmente daquelas utilizadas pelas empresas tradicionais com relação aos bancos de dados relacionais. Quando surgem os primeiros produtos NoSQL no mercado, os bancos de dados relacionais estavam consolidados como tecnologia de armazenamento de dados há mais de duas décadas (Mohan, 2013). Empresas como Oracle e IBM adotam para esta tecnologia a estratégia de software proprietário para uso comercial. Desta forma, o *copyright* é utilizado de forma restritiva, ou seja, a empresa fornece o software apenas em seu formato binário, e não seu código fonte. Além disso, a realização de cópias e a distribuição do produto só podem ser realizadas mediante autorização.

May e Sell (2006) realizam uma análise crítica da história dos direitos de PI, e identificam que as evoluções no entendimento destes direitos possuem uma relação estreita entre contexto social, estruturas legais e política econômica. Para os autores, as mudanças que se observam nas leis e no entendimento dos direitos de PI ao longo do tempo são respostas políticas a desafios decorrentes do surgimento de novas tecnologias e de novos usos sociais de tecnologias existentes. Ou seja, a história da propriedade intelectual é produto da interação entre o conceito político e social do papel daquele que gera conhecimento (ideias), a característica da tecnologia que utiliza PI ou está sujeita à proteção por este tipo de direito (mudança tecnológica), e a construção legal da PI (instituições). Defendem ainda que a história da propriedade intelectual se traduz em uma constante competição entre duas formas distintas de legitimação da propriedade do conhecimento. Uma primeira corrente de pensamento entende que os indivíduos devem beneficiar-se de seus esforços intelectuais, limitando o acesso público por meio de direitos privados. Por outro lado, uma segunda corrente argumenta que tais esforços individuais produzem valor tão abrangente que há claro interesse público na relativa disseminação livre do conhecimento resultante. Em linhas gerais, a discussão pode ser resumida como a disputa entre direitos privados, os quais limitam o acesso público, e a intenção do

interesse público em um fluxo de informação livre, ao custo dos direitos do criador. Em meio a esta discussão, surge a perspectiva por meio da qual se percebe que há interesse público no incentivo para a inovação, de forma a balancear as recompensas a entes privados e os benefícios públicos. Dessa forma, a criação de conhecimento em si é um bem público, e os direitos de propriedade sobre este conhecimento são uma forma de garantir a continuidade na busca por novos conhecimentos. Assim, os direitos de propriedade são estímulos à inovação e atrativos a investimentos para a comercialização destas inovações, sendo que este sistema de estímulos atende às necessidades do interesse público quando possui verificações capazes de prevenir abusos provenientes do uso indevido de tais direitos.

A partir da análise histórica dos direitos de PI, os autores observam que o desenvolvimento da PI é um processo político que, sucessivamente, produz novos estabelecimentos e institucionalizações. Dessa forma, percebe-se a evolução estrutural da PI em períodos de contestação nos quais se refletem as tensões entre os direitos privados de propriedade e a livre circulação do conhecimento. Além disso, novas concepções a respeito da propriedade intelectual e pressões decorrentes da mudança tecnológica são os condutores deste contínuo processo político. Neste contexto, arguem que a história dos direitos de PI não se desenvolve por meio de uma única trajetória, mas que, ao contrário, ocorre uma contínua evolução, aberta à contestação e mudança. Sendo assim, em cada estágio dos direitos de PI observa-se o legado de fases anteriores e adaptações às mudanças conceituais e aos desenvolvimentos tecnológicos correntes. Desta forma, a análise de uma sequência particular da história deve considerar as estruturas que refletem os estabelecimentos políticos e econômicos a respeito das regras de PI, e que alguns agentes detêm os benefícios do sistema, ao mesmo tempo em que existem outros em desvantagem. Assim, é possível perceber que a propriedade intelectual, enquanto instituição, é legalmente dependente do Estado, mas o desenvolvimento de questões relativas ao tema acontece por meio das atividades provocadas por agentes que integram o contexto da PI, como os entes privados, por exemplo. Desta forma, o avanço tecnológico e grupos empoderados por novas tecnologias são capazes de desenvolver novas estruturas, contestando aquelas previamente estabelecidas. Este poder reside na capacidade destes agentes de articularem-se em torno de interesses comuns, organizarem-se para agir coletivamente, gerarem

movimentos sociais e exercerem influência na tomada de decisão. Desta forma, estes agentes, conscientes dos objetivos a serem alcançados, articulam-se entre si e organizam-se de forma a alcançar tais objetivos, possibilitando ações coordenadas capazes de reformular as características estruturais em questão. No entanto, a insatisfação com o *status quo* e o reconhecimento das incongruências que reduzem os benefícios e o poder destes agentes não são condições suficientes para desafiar de forma efetiva a grupos previamente formalizados, fazendo-se necessária a capacidade de organização de uma resposta coletiva. Além de organização e habilidade de articulação, os agentes necessitam de capacidade tecnológica, poder político e acesso a recursos, inclusive às instituições que objetivam reformular. Desta forma, estes agentes identificam lacunas na reflexão de ideias dominantes da estrutura estabelecida, estimulados, por exemplo, por mudanças tecnológicas das quais emergem novas possibilidades, ou por mudanças nas ideias do criador de novos conhecimentos. Impulsionados por estes estímulos, tais agentes promovem ações alinhadas aos interesses do grupo, buscando eliminar práticas que reproduzem ou representem a continuação do *status quo* e que dificultam o atingimento dos objetivos que visam alcançar. Assim sendo, novos padrões relativos aos direitos de PI emergem a partir de distintas circunstâncias, como, entre outros, o exercício de poder econômico e o impacto de mudanças tecnológicas. Portanto, para os autores, mudanças tecnológicas e novos entendimentos sobre o uso dos direitos de PI são os estímulos capazes de reestruturar a instituição da propriedade intelectual. Assim, novas ideias sobre a criação do conhecimento podem levar à disparidades entre as instituições e estabelecimentos da PI e as concepções de inovadores e usuários. Para os agentes que estão em situação de desvantagem sob a perspectiva dos benefícios resultantes em virtude das instituições estabelecidas, as mudanças tecnológicas são a forma de alavancagem de alterações estruturais. Por fim, os autores identificam que a primeira corrente de pensamento tem sobressaído na discussão, favorecendo, portanto, os direitos privados sobre a difusão de conhecimento, embora nem sempre a resolução de conflitos seja favorável aos detentores de direitos de PI. Defendem que o estabelecimento corrente dos direitos de PI não é definitivo, mas sim o estágio mais recente do processo político de proteção por direitos de propriedade intelectual.

A partir desta perspectiva, pode-se analisar a questão do direito de autor e de distribuição de software aplicada aos bancos de dados NoSQL. No momento em que

surtem as inovações dos bancos NoSQL, empresas consolidadas como Oracle, IBM e Microsoft dominam o mercado da tecnologia dos bancos de dados relacionais, utilizando uma estratégia restritiva quanto ao *copyright* de software, ou seja, licença e distribuição de uso apenas do código binário dos respectivos produtos comerciais. Para Boyle (2008), a distribuição de software nestas condições exige níveis altos de capital. Como resultado, toda a indústria foi moldada em função da distribuição de software por si só dispendiosa. Esta estratégia é aderente à primeira corrente de pensamento definida por May e Sell (2006) de reforço dos direitos privados de PI como forma de recompensa aos esforços intelectuais dos produtores de conhecimento. No início do século XXI, esta estratégia restritiva de *copyright* estava estabelecida e institucionalizada, sendo as empresas proprietárias dos softwares protegidos as maiores beneficiadas através deste sistema de recompensas. Os bancos de dados NoSQL podem ser entendidos como um novo paradigma tecnológico, sendo uma inovação impulsionada por novos criadores de conhecimentos. A Google e a Amazon, por exemplo, embora já consolidadas no mercado de Internet, não possuíam um produto comercial relacionado à tecnologia de armazenamento de dados. A mesma situação se aplicava às outras empresas de menor porte desenvolvedoras de produtos comerciais de bancos de dados NoSQL, como a MongoDB, por exemplo. Desta forma, as empresas produtoras do novo conhecimento e da nova tecnologia necessitavam desenvolver novas estruturas, contestando aquelas previamente estabelecidas. Neste contexto, o movimento de software livre de código aberto era uma iniciativa utilizada com sucesso na indústria de TI para outros produtos e tecnologias, inclusive por empresas participantes do desenvolvimento dos bancos de dados NoSQL. Para Boyle (2008), o sucesso do software livre de código aberto reside no desenvolvimento e disponibilização de tecnologia explorando-se a redução dos custos de comunicação e distribuição alcançados com o advento da Internet. As comunidades de desenvolvedores promotoras de software livre de código aberto como a ASF e a FSF representam o poder de desenvolvimento de novas estruturas resultante do avanço tecnológico e da capacidade dos agentes produtores do novo conhecimento de articularem-se em torno de interesses comuns, organizarem-se para agir coletivamente, gerarem movimentos sociais e exercerem influência na tomada de decisão. Por meio destas comunidades, novas ideias e possibilidades, especialmente relativas ao uso de direitos de PI, podem ser exploradas pelas empresas responsáveis pelo avanço

tecnológico. Os bancos de dados NoSQL encaixam-se nesta categoria, visto que trata-se de uma nova tecnologia desenvolvida por agentes que não participam do mercado tradicional de tecnologias relacionais de armazenamento. Assim, as empresas produtoras do novo conhecimento optam pelo software livre de código aberto com o objetivo de competirem no mercado de tecnologias de armazenamento de dados. Desta forma, eliminam as práticas de *copyright* que reproduzem a continuação do *status quo* e utilizadas de forma bem sucedida pelas empresas concorrentes que dominam o mercado em questão, adotando novos padrões relativos aos direitos de PI. Duas estratégias adotadas pelas empresas inovadoras foram fundamentais para que obtivessem sucesso no objetivo de introduzirem-se competitivamente no mercado de bancos de dados, buscando novas formas de lucro que não necessariamente garantidos pela utilização restritiva do direito autoral. A primeira foi a criação de padrões de mercado por meio da publicação das tecnologias *BigTable* e *Dynamo*, pelas empresas Google e Amazon, respectivamente. A segunda, a distribuição via licença de software livre de código aberto para uso comercial dos bancos de dados NoSQL. Sem a fonte de renda advinda da exploração do *copyright*, as novas formas de lucro constituíram-se inicialmente no fornecimento de serviço de suporte técnico ao produto, e, mais, recentemente, na disponibilização do produto e toda infraestrutura necessária para sua utilização por meio da Internet, procedimento conhecido como computação em nuvem, por meio do qual se realiza a cobrança de taxas por níveis de utilização. Assim, pode-se afirmar que a distribuição dos bancos de dados NoSQL como software livre de código aberto foi a solução encontrada para a necessidade de entrada no mercado de armazenamento de dados pelas empresas produtoras do novo conhecimento, por meio da racionalidade da disseminação do conhecimento a fim de prover o interesse público, justificativa presente nos movimentos sociais representados, dentre outros, pela FSF e ASF, e aderentes à segunda corrente de legitimação da propriedade do conhecimento, conforme definida por May e Sell (2006).

De volta à análise crítica da história dos direitos de PI, segundo Archer (apud MAY e SELL, 2006, p. 34), o poder dos agentes inovadores está na capacidade de articular interesses em comum, organizando ações coletivas, gerando movimentos sociais e exercitando a influência corporativa. A autora classifica em agentes corporativos aqueles que, conscientes dos objetivos a serem alcançados e a fim de

alcançá-los, conseguem articularem-se e organizarem-se. São estes agentes que se envolvem em ações com a possibilidade de reformular as características estruturais em questão, no caso em estudo os direitos de propriedade intelectual. Além da articulação e organização, tais agentes somente serão bem sucedidos se possuírem capacidade técnica, poder político e acesso a recursos.

O patrocínio às instituições de suporte a projetos de software livre de código aberto como a FSF e a ASF, ou a distribuição de software por meio de licenças criadas por estas instituições, são ações que podem ser entendidas como a articulação e a organização das empresas geradoras do novo conhecimento com o objetivo de disputarem o mercado de tecnologias de armazenamento de dados, dominado por grandes corporações distintas das inovadoras. Estas buscaram como estratégia a reformulação estrutural na forma de uso direito autoral, promovendo o que Boyle (2008) identificou como a utilização dos direitos de propriedade intelectual com o intuito de prevenir a exclusividade, baseada em um novo método inovativo. O autor identifica esta quebra de paradigma quanto ao uso de direitos de PI como uma fonte de inovação e compartilhamento de informação no qual a função dos direitos de PI difere do modelo padrão de exclusão por controle proprietário. A Google e a Amazon, empresas pioneiras no desenvolvimento da tecnologia NoSQL, aparecem entre as dez maiores patrocinadoras da ASF, com aportes que ultrapassam mais de 125 mil dólares por ano (The Apache Software Foundation, 2019). Boyle (2008) cita que empresas participam do movimento tanto por meio do patrocínio formal, como também empregando pessoas que atuam nas comunidades de software livre de código aberto. O patrocínio a projetos de software livre de código aberto ou empregar indivíduos ativos nestas comunidades preenchem os requisitos de capacidade técnica, poder político e acesso a recursos apontados por Archer (apud MAY e SELL, 2006, p. 34) para que os agentes inovadores sejam bem sucedidos na reformulação das características estruturais do padrão vigente, a qual permitirá a estes agentes beneficiarem-se da mudança tecnológica que promoveram.

Por fim, pode-se afirmar que os movimentos de software livre de código aberto estabeleceram uma nova estrutura a respeito do direito autoral de software. O sucesso de tais movimentos evidencia-se pelo apoio tanto de empresas como a Google, referência como inovadora no ramo da Internet e, portanto, historicamente adepta ao software livre, como de empresas como a IBM e a Microsoft, tradicionais representantes da estrutura de direito autoral exclusivo. Boyle (2008) chama atenção



para a IBM, que, mesmo sendo a maior detentora de patentes do mundo, e, portanto, adepta do modelo de controle de proprietário por meio de direitos de PI, patrocina projetos de software livre de código aberto com o intuito de vender serviços para plataformas abertas, investindo, portanto, neste método de inovação e negócios. No entanto, observa-se que esta nova estrutura não significa a substituição da anterior, visto que a Oracle se mantém competitiva utilizando-se de forma restritiva o *copyright*, e a IBM e Microsoft, patrocinadoras formais da ASF, utilizam a estratégia de software proprietário protegido por direito autoral na busca por lucros dos produtos que comercializam.

### 3.2 APROPRIAÇÃO DA INOVAÇÃO DOS BANCOS DE DADOS NOSQL

Tanto o Apache HBase quanto o Apache Cassandra se baseiam no modelo instituído pela tecnologia BigTable, apresentada em um simpósio técnico e publicada pela área de pesquisa da Google em 2006. No documento de publicação, o BigTable é definido como um sistema de armazenamento distribuído para dados estruturados<sup>3</sup> (Chang et al., 2006).

Pode-se inferir a Google como a firma inovadora da tecnologia NoSQL, antes mesmo do surgimento do termo. É interessante observar que ao publicar o modelo do BigTable, foi disponibilizado para o mercado um padrão para uma tecnologia produzida de forma endógena e já utilizada como sistema de armazenamento em alguns projetos ou produtos da própria empresa, tais como a indexação da Internet, o Google Earth e o Google Finance (Chang et al., 2006). A criação de um modelo de mercado utilizando-se de um conhecimento desenvolvido internamente pode ser entendida como uma estratégia consciente de apropriação, visto que não há um produto comercial lançado especificamente para a tecnologia BigTable, seja software livre ou proprietário. Alguns dos produtos comerciais utilizando a tecnologia surgem a partir de projetos de software livre de código aberto como o Apache HBase e o Apache Cassandra da ASF, associação da qual a empresa inovadora participa

---

<sup>3</sup> O documento de publicação da tecnologia não define o BigTable como um banco de dados. Possivelmente a Google tinha o objetivo de diferenciar o BigTable dos bancos de dados relacionais. Como ainda não havia um termo como o NoSQL para essa diferenciação, a tecnologia não foi classificada. No documento de publicação cita-se que o BigTable em muitos aspectos assemelha-se a um banco de dados.

como um patrocinador importante. Ainda assim, a estratégia de apropriação da Google para a tecnologia inclui a proteção por propriedade intelectual. A empresa é a maior detentora de patentes de tecnologias que utilizam o termo BigTable no relatório descritivo, citam o documento de publicação do modelo, ou ambos<sup>4</sup>. Paralelamente, BigTable é um nome protegido por marca registrada em diversos países, como Estados Unidos, Canadá, Reino Unido, França, Alemanha, Índia e Brasil, dentre outros<sup>5</sup>. Trata-se, portanto, de um caso bastante ilustrativo da observação de Mello (2009) de que a proteção por patentes não é a única estratégia de apropriação adotada pelas empresas, e que frequentemente tais estratégias combinam diversos elementos. No caso em estudo, pode-se identificar como estratégia de apropriação a proteção pelos ativos de PI por patentes e marcas juntamente com o pioneirismo e a disponibilização de um conhecimento interno a fim de criar um modelo de mercado.

Este tipo de estratégia para a inovação envolve atividades baseadas em temas como competências essenciais à firma, colaboração, redes e Internet. Segundo Huizingh (2010), tais atividades estão conectadas e integradas pelo amplo termo conhecido como inovação aberta. Considerando-se a amplitude do termo, a partir da análise de outros trabalhos relativos ao tema, o autor identifica quatro perspectivas que comparadas entre si determinam o grau de inovação aberta da estratégia adotada por uma empresa.

A primeira perspectiva é a constatação de que inovação aberta é um processo contínuo com diferentes graus de abertura. Este processo engloba várias atividades, cada qual podendo ser mais ou menos aberta.

A segunda corresponde à comparação entre a entrada de inovação aberta, ou seja, a utilização interna de conhecimento gerado externamente, e a saída de inovação aberta, que se refere à exploração externa de conhecimento interno. Estas duas dimensões são cruzadas com interações pecuniárias e não pecuniárias, formando uma matriz de atividades estratégicas, a saber: adquirir, abastecer-se, vender e revelar.

---

4 Conforme pesquisa pelo termo "Bigtable" na ferramenta Lens.org de busca de patentes, disponível em <https://www.lens.org/lens/>.

5 Conforme pesquisa pelo termo exato "Bigtable" na ferramenta TMView, disponibilizada na Internet pelo Escritório da Propriedade Intelectual da União Europeia (EUIPO), disponível em <https://www.tmdn.org/tmview/welcome>.

A terceira perspectiva trata do fluxo de conhecimento na inovação aberta, considerando que os três processos de conhecimento, quais sejam a exploração, retenção e extrapolação, podem ser desenvolvidos interna ou externamente à firma.

Por fim, a última perspectiva agrupa as práticas adotadas pela empresa de acordo com o processo e os resultados. Tanto estes como aquele podem ser classificados em abertos ou fechados, levando a uma matriz de classificação da inovação em fechada (quando uma inovação proprietária é desenvolvida internamente), inovação aberta privada (quando uma inovação proprietária é desenvolvida por meio de um processo aberto, por exemplo, por meio de insumos de parceiros externos), inovação pública (processo fechado, resultado aberto) e inovação de código aberto (processo e resultados abertos).

Tais perspectivas podem ser avaliadas na estratégia de inovação que culmina no surgimento da tecnologia BigTable e dos produtos posteriormente desenvolvidos por meio da ASF, estes já classificados como NoSQL e baseados no modelo da primeira. Por meio da publicação do modelo BigTable, a Google explora de forma exógena o conhecimento produzido internamente, revelando-o. Este conhecimento será extrapolado pela adoção do modelo por uma instituição parceira da própria empresa no desenvolvimento de softwares livres de código aberto, no caso a ASF. Pode-se afirmar que a estratégia de apropriação da empresa para a tecnologia se baseia em alto grau de inovação aberta.

A estratégia de apropriação relativa ao MongoDB também utiliza alto grau de inovação aberta. No entanto, há duas diferenças quanto aos produtos da ASF: a estratégia de apropriação de lucros pela empresa e a licença de distribuição do software.

A estratégia de apropriação de lucros provenientes da inovação está na venda do suporte técnico especializado para o software desenvolvido internamente. Trata-se de uma estratégia de apropriação baseada em ativos complementares, visto que está diretamente relacionada ao conhecimento tecnológico específico associado à inovação. Neste caso, pode-se afirmar que a empresa atua como produtora de saída de inovação aberta, explorando o conhecimento endógeno externamente, porém obtendo vantagens pecuniárias, ou seja, adotando uma estratégia de resultados fechados.

Quanto à licença, o MongoDB é distribuído sob a GNU Affero General Public License, a qual restringe novas declarações a respeito de direitos autorais pelo autor

de uma modificação. A empresa garante, desta forma, que outras empresas ou indivíduos possam apropriar-se da venda de produtos modificados a partir do original.

Software livre é comumente confundido com software grátis, especialmente devido à dupla tradução do termo *free software*, a partir do qual se origina o conceito. No entanto, assim como o Apache HBase, o Apache Cassandra e o MongoDB, muitos softwares livres são produzidos em escala comercial. Por isso, as empresas ou instituições que desenvolvem software livre tratam das questões relacionadas aos ativos de PI nas licenças das distribuições disponibilizadas, especialmente direito autoral, marcas e patentes, conforme observado no caso dos produtos da ASF e da MongoDB.

Diferentemente das estratégias apoiadas em alto grau de inovação aberta, há também a estratégia adotada pela Oracle de software proprietário para uso comercial. Tal estratégia de apropriação se apoia no fato de que o produto de banco de dados relacional da empresa é um dos mais consolidados no mercado, devido especialmente à reconhecida qualidade. Assim, a empresa se utiliza de uma posição de liderança mercadológica prévia na busca por lucros com a nova tecnologia, disponibilizando uma solução proprietária nas mesmas condições estabelecidas para o produto de banco de dados relacional. Neste caso, a proteção por marcas é uma estratégia fundamental para a empresa. A marca registrada pela empresa é o próprio nome, e para identificar o banco de dados relacional, é utilizado o termo *Oracle Database*, ou seja, não há uma marca registrada para identificar especificamente o produto. Percebe-se que é um caso típico da definição de marca concebida por Ramello (2006), um sinal distintivo utilizado em atividades econômicas para diferenciar produtos, permitindo a distinção entre estes e o reconhecimento da proveniência dos mesmos. No caso da Oracle, pode-se ainda identificar que a marca, além da função básica de produzir a informação que cria distintividade, possui poder de mercado, situação que o autor classifica como um efeito dinâmico adicional. Baseada neste poder de mercado e na reputação do nome, a empresa mantém esta estratégia para a nova tecnologia, identificando o produto simplesmente como *Oracle NoSQL Database*, sendo marca registrada apenas o primeiro termo.

A partir dos anos 2010, outra estratégia de apropriação tornou-se bastante frequente a partir de um novo paradigma tecnológico da área de TI: a computação

em nuvem. Os bancos de dados NoSQL surgiram para atender as necessidades de armazenamento de dados em grandes volumes e heterogêneos, além de possuírem a característica essencial da escalabilidade. A utilização de softwares por empresas clientes interessadas na tecnologia exige a instalação de um parque tecnológico interno avançado e custoso, tanto na questão do suporte técnico quanto de manutenção, o que envolve também a contratação ou terceirização de técnicos especializados. A computação em nuvem permite que empresas de TI forneçam toda esta estrutura e as atividades inerentes como um serviço na Internet. Os serviços de computação em nuvem são subdivididos em três categorias: *Software as a Service* (Saas), *Platform as a Service* (Paas) e *Infrastructure as a Service* (Iaas). Na primeira categoria, é fornecida ao consumidor a utilização de software em uma infraestrutura remota. O modelo *Paas* se constitui na disponibilização de uma camada de software ou de um ambiente de desenvolvimento para a construção de aplicações de níveis mais altos. Por fim, no modelo *Iaas*, o objetivo é prover capacidades computacionais básicas e armazenamento como um serviço padronizado (Rehman, 2018). De forma geral, os clientes destes serviços não necessitam investir em uma infraestrutura específica para adotar soluções de TI, mas pagam de acordo com o nível de serviço que utilizam dos fornecedores que elegem. Por exemplo, no caso de soluções de bancos de dados NoSQL oferecidas como serviço, são cobradas taxas referentes à taxa de transferência de dados e quantidade de armazenamento. A estratégia de oferecer soluções de bancos de dados NoSQL em nuvem é utilizada por empresas como Amazon, IBM, e Google. A Microsoft fornece por meio da computação em nuvem um serviço mais complexo que envolve, dentre outras tecnologias, o armazenamento em banco de dados NoSQL. A MongoDB também fornece o serviço de banco de dados NoSQL em nuvem. Neste caso, a empresa adota duas estratégias simultâneas, pois mantém a disponibilização do software a clientes que possuam a infraestrutura necessária. A busca por lucros por meio da disponibilização do serviço na Internet trata-se de estratégia de apropriação por meio de ativos complementares, pois se apoia no conhecimento tecnológico associado à inovação. Neste caso, este conhecimento tecnológico está associado a duas inovações: banco de dados NoSQL e computação em nuvem.

A proteção pelo ativo de PI de marcas é relevante para a diferenciação entre os diversos serviços oferecidos na Internet. O quadro 1 agrupa as marcas

registradas em diferentes escritórios de PI, notadamente Estados Unidos, Japão e Europa, e que identificam os serviços de bancos de dados NoSQL em nuvem.

<b>Titular</b>	<b>Marca registrada</b>
Amazon	DynamoDB
IBM	Cloudant
Google	Cloud BigTable
Microsoft	Azure
MongoDB	-

**Quadro 1 – Marcas registradas para os serviços de bancos de dados NoSQL fornecidos em nuvem: TmView, disponível em <https://www.tmdn.org/tmview/welcome#> (2017).**

Pode-se perceber o fenômeno observado por Carvalho (2009) acerca de uma tendência de que as marcas registradas sejam distantes do nome da empresa, sem indicação clara da origem do produto ou serviço. No entanto, como neste caso as marcas correspondem a serviços fornecidos por meio da Internet, e não aos chamados softwares de prateleira, a publicidade ocorre primordialmente nos sítios institucionais das empresas. Desta forma, embora a marca registrada crie a informação que designa distintividade, sem, no entanto, indicar procedência, a utilização da mesma está sempre associada ao nome ou logotipo das empresas.

No caso da Google, a marca Cloud BigTable é uma evidente referência à tecnologia publicada em 2006. Assim, embora a procedência do serviço não esteja clara, um público alvo específico de profissionais de TI pode reconhecer a origem do serviço pela associação à tecnologia publicada pela empresa.

A MongoDB não possui nenhuma marca registrada para o serviço de banco de dados NoSQL fornecido por meio da computação em nuvem. No entanto, no sítio institucional da empresa o serviço sempre é referenciado pelo nome MongoDB Atlas, do qual apenas o primeiro termo é uma marca registrada.

As empresas que buscam beneficiar-se dos lucros provenientes da inovação representada pelos bancos de dados NoSQL, independente da estratégia de apropriação adotada e do grau de inovação aberta, não descartam a proteção por patentes. Não existem patentes para os bancos de dados NoSQL como um produto

comercial. Porém, uma vez que os bancos de dados NoSQL não possuem os controles estruturais existentes nos bancos de dados relacionais, o desenvolvimento de um software desta tecnologia envolve uma série de inovações, como, por exemplo, métodos de busca e de garantia de consistência. Sobre estas inovações recaem as estratégias de apropriação por proteção de patentes pelas empresas de TI.

O gráfico 2 apresenta o resultado da pesquisa por patentes relacionadas ao termo BigTable, distribuídas entre as empresas que apresentam alguma estratégia de apropriação para a tecnologia de bancos de dados NoSQL. Estas patentes citam o documento de publicação de 2006 da tecnologia BigTable, utilizam o termo no relatório descritivo, ou ambos. A Google é a detentora de direitos de 94 patentes. A empresa concentra a maior quantidade de patentes relacionadas à tecnologia, como se pode supor. A pesquisa na base de patentes demonstra também que a tecnologia efetivamente tornou-se uma referência de mercado, visto que empresas concorrentes também buscam a proteção por este ativo de propriedade intelectual sobre tecnologias associadas ao BigTable. Por exemplo, a pesquisa pelo mesmo termo retorna 37, 23 e 20 patentes cujos detentores são, respectivamente, a Microsoft, a IBM e a Amazon. Observa-se em The Apache Software Foundation (2019) que estas três empresas patrocinam em diferentes níveis de aportes financeiros a ASF, instituição esta que desenvolve produtos comerciais livres de código aberto baseados na tecnologia. Dentre as empresas que não aparecem como patrocinadoras da ASF, a MongoDB, fornecedora de uma solução NoSQL livre de código aberto, detém dez patentes utilizando o termo BigTable e a Oracle, cuja estratégia de apropriação reside em software proprietário, é detentora de quinze patentes.

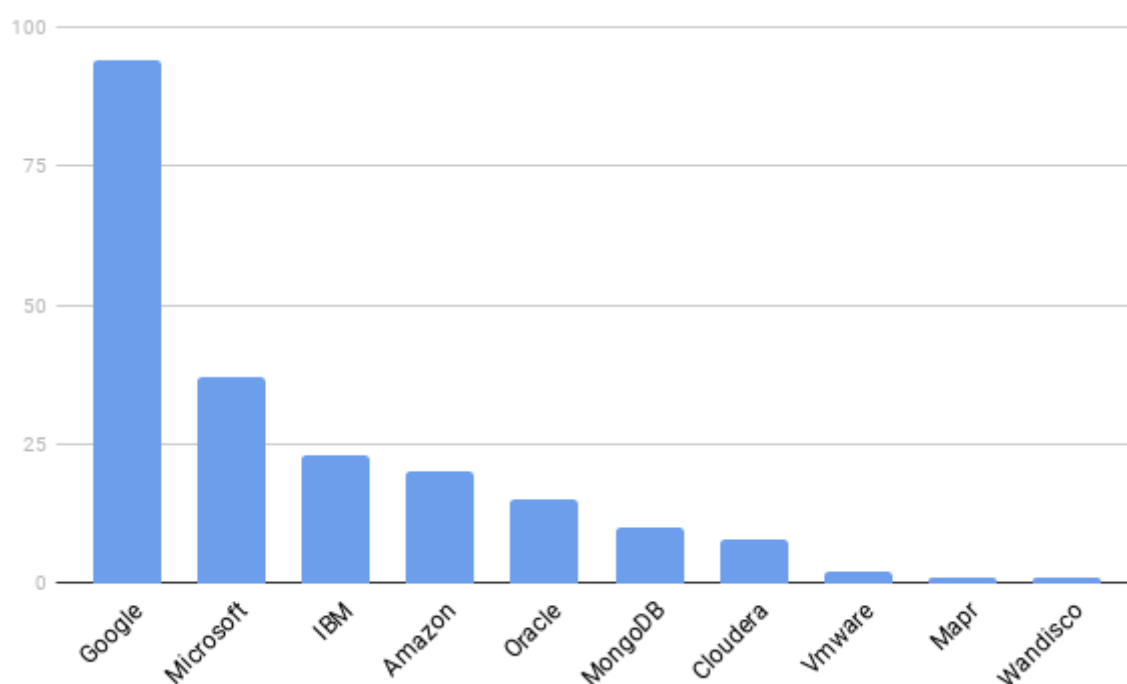


Gráfico 2 – Depósitos de patentes que citam o termo BigTable: Lens.org (<https://www.lens.org/lens/>).

Conforme constatado por Carneiro (2007), a popularização da Internet a partir do movimento Web 2.0 introduziu novos modelos de negócio, inclusive quanto às formas de apropriação dos resultados das inovações e das formas de proteção por ativos de PI. A disponibilização mercadológica de softwares livres de código aberto permite a outras empresas apropriarem-se de parte dos lucros advindos da inovação resultante da tecnologia por meio do serviço especializado de suporte técnico, ou do desenvolvimento de novas soluções de software livre de código aberto derivadas das primeiras. Pode-se entender este nicho de atuação como apropriação por meio de ativos complementares, visto que se trata da prestação de um serviço dependente de conhecimento tecnológico específico associado à inovação.

Este conceito pode ser aplicado no caso específico da tecnologia NoSQL. A ASF mantém na Internet uma página wiki<sup>6</sup> destinada ao projeto *Hadoop* (APACHE HADOOP, 2019) e outra para o projeto *Cassandra* (APACHE CASSANDRA WIKI, 2019). O primeiro consiste em uma coleção de softwares livres de código aberto utilitários cujo objetivo é possibilitar o processamento por meio de uma rede de computadores de problemas que envolvam grandes quantidades de dados, e do qual o Apache HBase é o produto NoSQL utilizado para armazenamento de dados.

<sup>6</sup> Páginas da Internet que permitem edição colaborativa do conteúdo e da estrutura pelos usuários.



O segundo é o projeto que culminou com o banco de dados NoSQL para uso comercial de mesmo nome. Em Apache Hadoop (2019), são mencionados os fornecedores de produtos derivados a partir do Apache Hadoop. Em ambas, são listados os fornecedores de suporte técnico ao respectivo produto, com uma menção de que se trata apenas de uma referência. A análise das listas das empresas envolvidas com aqueles projetos por meio do fornecimento de suporte técnico ou de novos produtos derivados mostra empresas que aparecem também como associadas à ASF por meio de patrocínio financeiro em The Apache Software Foundation (2019), como é o caso da Cloudera, a qual é detentora de oito patentes citando a tecnologia BigTable. Da mesma forma, encontram-se empresas envolvidas nos projetos Hadoop ou Cassandra sem que sejam patrocinadoras da ASF e detentoras de patentes citando o termo BigTable. Incluem-se neste caso a Vmware, detentora de duas patentes, e a Mapr Technologies e a Wandisco, que possuem uma patente cada.

Com o surgimento do termo NoSQL, percebe-se a utilização deste nos relatórios descritivos e até mesmo em alguns títulos de patentes depositadas. No gráfico 3, resultado da pesquisa por patentes que utilizam o termo NoSQL, observa-se a confirmação da importância da proteção por este ativo de PI sobre as inovações associadas aos bancos de dados NoSQL para as empresas que adotam distintas estratégias de apropriação da tecnologia. Percebe-se também que se alteram a ordem de grandeza dos depósitos, como também a liderança entre as empresas quando se compara com o resultado da pesquisa pelo termo BigTable, sendo neste caso a Amazon a maior detentora, com 231 patentes. Em seguida, aparecem outras empresas que, assim como a primeira, possuem estratégia de exploração da tecnologia como serviço: a IBM com 156 patentes e a Microsoft com 129. A Oracle, cuja estratégia baseia-se em software proprietário, aparece com 131. Com números mais modestos, a Google, empresa inovadora na tecnologia e que atualmente também adota a estratégia de disponibilizá-la por meio de serviço, aparece como detentora de 33 patentes. Da mesma forma como se pode verificar na consulta por patentes que utilizam o termo BigTable, empresas especializadas no suporte técnico ou fornecedoras de software livre de código aberto derivado dos bancos de dados desta tecnologia desenvolvidos pela ASF, também depositam patentes pelo termo NoSQL: a Cloudera, que figura como patrocinadora da ASF em The Apache Software Foundation (2019), e detém patentes de acordo com a

pesquisa pelo termo BigTable, é proprietária de oito patentes com termo NoSQL. Da mesma forma, a VMware, a qual aparece na primeira pesquisa e não consta como patrocinadora da ASF, detêm 11 patentes com o termo NoSQL. As empresas MapR Technologies e Wandisco, para as quais a primeira busca retorna resultados, não possuem patentes nesta pesquisa. Caso inverso ocorre com a Pivotal, empresa citada em Apache Hadoop (2019) como desenvolvedora de soluções baseadas no produto Apache Hadoop, a qual não apresenta resultados na primeira busca, sendo detentora de três patentes com o termo NoSQL. Um caso semelhante ao da Pivotal em relação ao depósito de patentes ocorre com a Datastax, a qual é citada como prestadora de suporte técnico em Apache Cassandra Wiki (2019). O diferencial em relação a esta empresa é o fato de ser referenciada como uma empresa que conta com *committers* do próprio projeto Cassandra no quadro de funcionários. O termo *committers* designa os indivíduos que contribuem com código fonte ou documentação nos projetos da ASF, e com os quais a fundação possui um acordo de licença de contribuição. Tal característica confere à empresa maior destaque na lista de prestadoras de serviços de suporte técnico ao Apache Cassandra, sendo a primeira citada em uma seção exclusiva para empregadores de *committers* do projeto. A Datastax não figura na busca por patentes que citam o termo Bigtable, mas é a detentora de três patentes quando a busca é realizada com o termo NoSQL. Finalmente, a MongoDB, cujo produto comercial para a tecnologia é software livre de código aberto, e não figura como patrocinadora da ASF, tampouco é citada em Apache Hadoop (2019) ou Apache Cassandra Wiki (2019), é a detentora de três patentes na pesquisa pelo termo NoSQL.

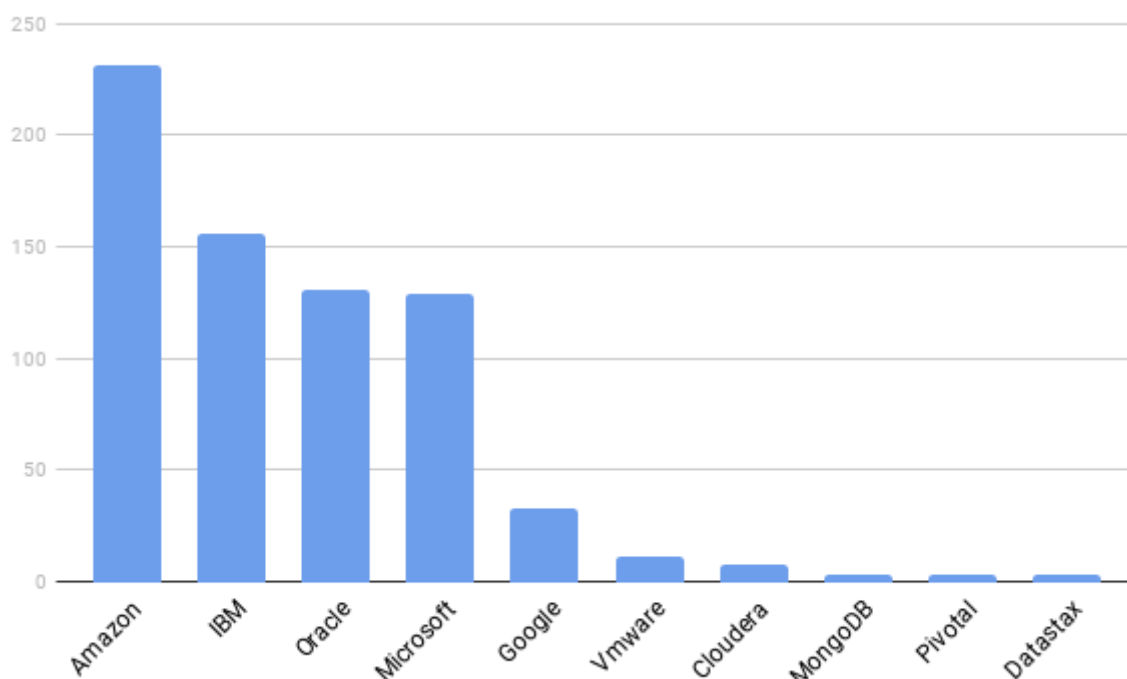


Gráfico 3 – Depósitos de patentes que citam o termo NoSQL: Lens.org (<https://www.lens.org/lens/>).

## DISCUSSÃO

O paradigma tecnológico das tecnologias NGS revolucionou o desempenho do sequenciamento genético humano em termos de tempo e custo. A premissa básica desta inovação é o sequenciamento paralelo de milhões de fragmentos de DNA, o que produz um grande volume de dados. Como consequência, tornam-se necessárias novas tecnologias de armazenamento e transferência de dados. Estas tecnologias são disponibilizadas por meio de inovações da área de TI.

Os bancos de dados relacionais surgiram a partir da década de 1970 com o intuito primordial de processar dados de negócios. Desenvolvidos sob o conceito *one size fits all*, ou seja, de solução única para qualquer tipo de problema, se constituíram na tecnologia consolidada para armazenamento e compartilhamento quando surge o movimento Web 2.0 no início do século XXI. Este foi o termo utilizado para o desenvolvimento da Internet como uma plataforma de serviços, o qual desencadeou novas necessidades para os sistemas de banco de dados, como, por exemplo, o armazenamento e processamento de dados em grandes volumes

(Big Data), além de não estruturados. Estas questões motivaram também uma mudança de característica estrutural, pois, como o tratamento de dados em larga escala excede a capacidade de uma única máquina, passa a ser desejável a possibilidade de escalabilidade horizontal (incluir mais máquinas no sistema) em detrimento à vertical (aumento de recursos de uma máquina). Devido a essas questões, inicia-se nos anos 2000 o desenvolvimento de novas tecnologias de armazenamento de dados. Tais tecnologias foram posteriormente denominadas de NoSQL, em uma alusão direta à linguagem de comunicação dos bancos de dados relacionais, posicionando-as de forma oposta.

A primeira inovação referente à tecnologia NoSQL foi a publicação pela Google em 2006 do modelo BigTable de armazenamento de dados, o qual já era utilizado pela empresa em alguns produtos. Desta forma, conclui-se que a estratégia de apropriação da empresa inovadora foi tornar uma tecnologia desenvolvida e utilizada internamente em um padrão de mercado. Dessa forma, a empresa explora externamente um conhecimento interno. Este conhecimento é extrapolado pela adoção da tecnologia revelada em produtos comerciais por empresas ou organizações que desenvolvem software livre de código aberto. Estas atividades denotam uma estratégia com alto grau de inovação aberta pela empresa inovadora.

A Google utiliza a estratégia de publicar a tecnologia, criando um modelo de mercado, em conjunto com a proteção por ativos de propriedade intelectual (PI). Embora a empresa não possua nenhum software comercial de banco de dados NoSQL, o termo BigTable, utilizado inicialmente para designar a tecnologia, foi posteriormente protegida como uma marca registrada em diversos países. Paralelamente, a empresa é a maior detentora de patentes que citam a tecnologia ou o documento de publicação. Exemplifica-se desta forma que a proteção por patentes é uma estratégia relevante, porém não se resume à única forma de apropriação sobre uma tecnologia, sendo utilizada juntamente com outros elementos.

Os primeiros produtos comerciais utilizando como base o modelo BigTable são os softwares livres de código aberto Apache HBase e Apache Cassandra, desenvolvidos de forma colaborativa pela comunidade Apache Software Foundation, da qual a Google é uma das patrocinadoras com maior volume de aportes por ano.

A adoção por empresas de TI de estratégias com alto grau de inovação aberta tem relação direta com softwares livres de código aberto. Esta forma de distribuição de software permite que outras empresas adotem como estratégia de apropriação a comercialização do serviço especializado de suporte técnico. Trata-se de uma estratégia que explora o conhecimento tecnológico específico associado à inovação, e, portanto, pautada em apropriação por ativos complementares.

Outra estratégia com alto grau de inovação aberta e apropriação por ativos complementares é o desenvolvimento de software livre de código aberto e comercialização de suporte técnico especializado pela própria empresa, como adotado pela MongoDB. Neste caso, a firma produz inovação aberta, explorando externamente o conhecimento endógeno, e apropria-se dos lucros por meio de um serviço baseado no conhecimento tecnológico associado à inovação.

A disponibilização de software livre de código aberto no mercado requer a observação de questões relativas aos direitos de propriedade intelectual. As licenças de distribuição de software, desenvolvidas pelas instituições que aderem ao movimento do software livre, são disponibilizadas em conjunto com o produto. São o documento no qual estão caracterizadas as tratativas quanto a direito de autor, patentes e marcas cujo detentor seja o indivíduo responsável pela colaboração que resulta em um novo produto ou na modificação de um existente.

Além da questão da licença, que trata de direitos de PI dos colaboradores individuais, a marca registrada da própria instituição é um ativo relevante para a disponibilização de software livre de código aberto. Por exemplo, os nomes Apache, HBase e Cassandra e os logotipos associados a estes nomes são marcas registradas de titularidade da ASF. Em um ambiente colaborativo de código fonte aberto, a proteção por marcas é uma estratégia fundamental para evitar o uso indevido de termos e imagens por pessoas ou instituições que inapropriadamente queiram beneficiar-se da reputação de mercado da instituição e dos produtos por esta disponibilizados. O objetivo deste tipo de proteção de PI é evitar o uso indevido de nomes e figuras por pessoas ou instituições que busquem obter algum benefício relacionando-se inapropriadamente da origem, qualidade ou reputação dos softwares disponibilizados.

A publicação de tecnologia desenvolvida internamente e a disponibilização de software livre de código aberto constituem-se em atividades que evidenciam uma reformulação estrutural na forma de uso do direito autoral e de cópia. No entanto, não compreende a eliminação do software proprietário protegido por *copyright*. Neste contexto, o direito de autor é utilizado de forma a restringir a cópia e distribuição do produto, portanto, de maneira oposta às estratégias de distribuição de software livre de código aberto. No caso da Oracle, que mantém a estratégia de software proprietário, a proteção por marcas é essencial, pois, por meio do nome consolidado no mercado e da qualidade e reputação do produto de banco de dados relacional, a empresa obtém poder de mercado ao indicar a origem do produto para a nova tecnologia.

A utilização de bancos de dados NoSQL pelos clientes das empresas de TI remete à necessidade de infraestrutura e pessoal técnico especializado para lidar com a tecnologia. A computação em nuvem, um novo paradigma tecnológico da área de TI, permite a disponibilização destes bancos de dados como um serviço na Internet. Isto se constitui em uma nova estratégia de apropriação, na qual os lucros são provenientes de taxas cobradas por nível de serviço, como, por exemplo, quantidade de armazenamento e taxa de transferência de dados. Esta é outra estratégia de apropriação por ativos complementares, já que se baseia no conhecimento tecnológico associado à inovação, sendo a mais utilizada atualmente pelas empresas de TI, adotada, por exemplo, pela Google, Amazon, IBM, Microsoft e MongoDB.

As marcas registradas cumprem a função da diferenciação entre estes serviços no mercado. Os nomes registrados para os serviços repetem a tendência recente do distanciamento entre as marcas registradas e a indicação de origem do produto ou serviço. No entanto, por tratar-se de serviços comercializados e disponibilizados por meio da Internet, estas marcas são visualizadas principalmente nos sítios institucionais das empresas fornecedoras. Desta forma, são marcas apresentadas ao público consumidor em associação aos nomes e figuras registradas da empresa.

A maior parte das estratégias de apropriação das empresas de TI para os bancos de dados NoSQL está baseada no conceito de ativos complementares, independente do grau de inovação aberta. No entanto, observa-se que estas

mesmas empresas também se utilizam da estratégia de apropriação por meio da proteção por propriedade intelectual.

A apropriação por meio da proteção por marcas é notória, pois todos os produtos de mercado, independente da estratégia de software livre, proprietário ou serviço em nuvem, são marcas registradas em diferentes partes do mundo, associadas diretamente ao nome da empresa, inclusive os softwares desenvolvidos pela ASF. Desta forma, as empresas induzem reputação e promovem o reconhecimento imediato da origem de um produto de uma tecnologia recente.

Também se observa que a apropriação por patentes são outra estratégia de proteção por ativos de propriedade intelectual relevante para a tecnologia de bancos de dados NoSQL. O termo NoSQL representa um conceito que, resumidamente, se refere a armazenamento e transferência de grande volume de dados, cuja implementação envolve uma série de diferentes tecnologias. Estas tecnologias constituem o alvo das estratégias de apropriação por proteção por patentes pelas empresas de TI. A análise dos depósitos mostra que são detentoras de patentes as empresas que desenvolvem e distribuem software livre de código aberto, as que se apropriam dos lucros provenientes da tecnologia por meio do suporte técnico especializado, as que comercializam software proprietário e as que disponibilizam a tecnologia por meio de serviço em nuvem. Desta forma, conclui-se que independente da estratégia de apropriação dos lucros provenientes dos bancos de dados NoSQL e do grau de inovação aberta adotado, a proteção por patentes é parte importante da estratégia de apropriação das empresas de TI para esta inovação.

Embora haja coexistência entre o desenvolvimento de software livre de código aberto e de software proprietário, no caso específico da comercialização dos bancos de dados NoSQL, percebe-se que esta última estratégia foi adotada pela Oracle. Neste caso, trata-se de empresa consolidada no mercado internacional de armazenamento de dados e com expressiva reputação com referência ao produto de banco de dados relacional. Por outro lado, a primeira estratégia foi adotada por uma variedade de empresas, desde gigantes multinacionais, como Google e Amazon, as quais buscavam se inserir no mercado de armazenamento de dados a partir de uma nova tecnologia, como também por empresas menores as quais desenvolveram seus produtos baseando-se nas tecnologias *BigTable* e *Dynamo* publicadas pelas primeiras. Empresas que se posicionam desta forma no mercado produzem

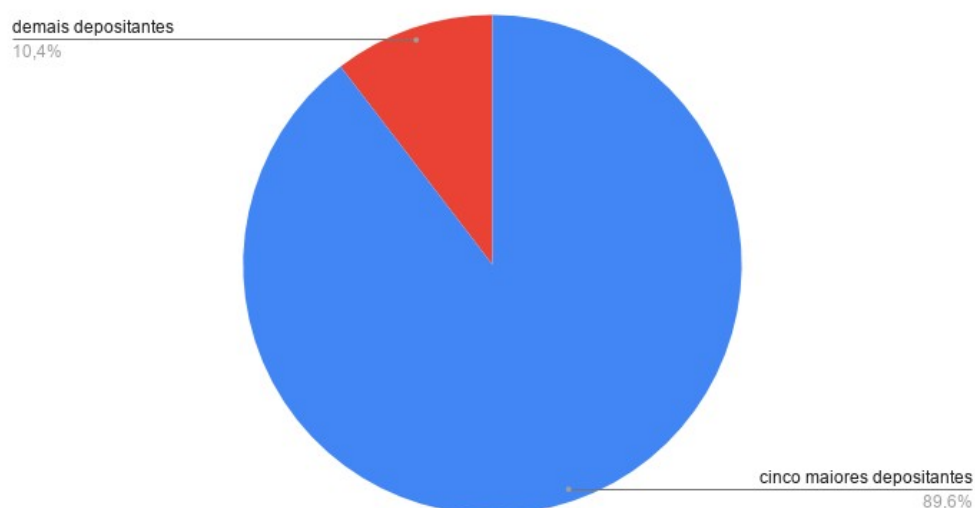
inovação aberta, utilizando uma estratégia de apropriação que explora o conhecimento tecnológico gerado internamente por meio de um serviço técnico especializado associado à inovação.

Conforme citado anteriormente, as estratégias de apropriação de uma inovação envolvem os serviços especializados necessários à comercialização, como também as formas de proteção por direitos de PI.

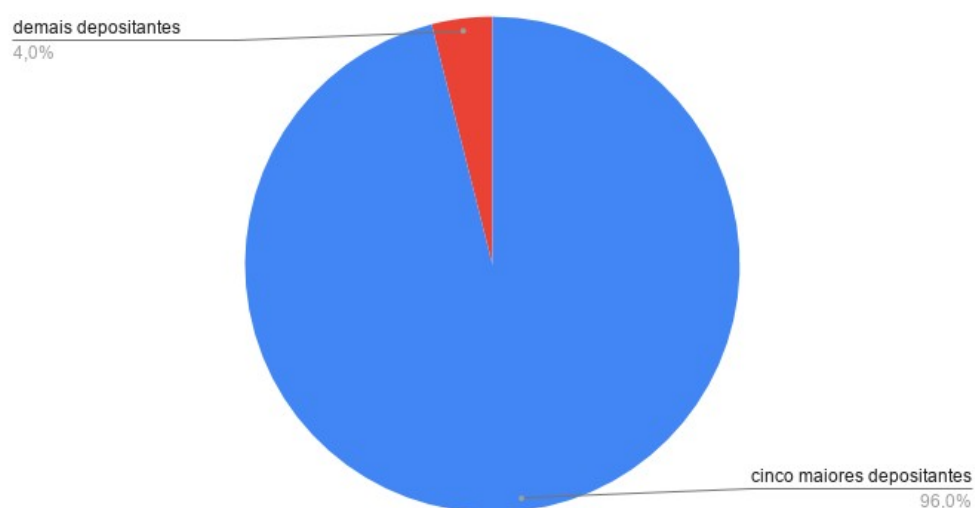
Dessa forma, cabe avaliar as estratégias das empresas quanto à proteção por patentes de invenção no caso dos bancos de dados NoSQL. A partir do estudo dos depósitos que citam os termos BigTable e NoSQL, pode-se identificar algumas características em relação ao uso deste ativo de PI pelas empresas de software.

Buscou-se identificar entre as depositantes de patentes para a tecnologia de banco de dados NoSQL as empresas reconhecidas como as geradoras do conhecimento, aquelas consolidadas no mercado de bancos de dados relacionais, e as empresas desenvolvedoras de bancos de dados NoSQL comerciais que tenham sido apontados em estudos de implementação da tecnologia para a solução de questões relacionadas ao sequenciamento genético humano. No caso dos bancos de dados NoSQL distribuídos pela ASF e citados em artigos de pesquisa científica relacionados ao sequenciamento genético humano, buscou-se identificar neste estudo as empresas que possuem estreita relação com os projetos de desenvolvimento ou prestadoras do serviço de suporte técnico destes produtos. Neste conjunto, foram identificadas 12 empresas, totalizando 211 depósitos de patentes citando o termo Bigtable, e 708 citando NoSQL. No entanto, percebe-se uma grande discrepância na distribuição dos depósitos destas patentes entre cinco maiores depositantes e as demais. Na busca pelo termo Bigtable, as empresas Google, Amazon, Microsoft, IBM e Oracle totalizam, juntas, 189 depósitos, o que representa quase 90% dos depósitos analisados. Esta situação se repete na busca quando utilizado o termo NoSQL, sendo estas empresas detentoras de 680 patentes dentre as analisadas, ou mais de 96% do total, demonstrando ainda maior preponderância destas empresas frente às demais.





**Gráfico 4 – Distribuição dos depósitos de patentes que citam o termo BigTable: Lens.org (<https://www.lens.org/lens/>).**



**Gráfico 5 – Distribuição dos depósitos de patentes que citam o termo NoSQL: Lens.org (<https://www.lens.org/lens/>).**

Assim, pela distribuição dos depósitos de patentes das tecnologias associadas aos bancos de dados NoSQL, infere-se que este ativo de PI possui grande relevância para as grandes corporações multinacionais. Especificamente quanto a esta tecnologia, estas empresas dividem-se entre as geradoras do conhecimento tecnológico e as consolidadas no mercado da tecnologia anteriormente dominante – os bancos de dados relacionais.

Por fim, a partir do estudo das tecnologias de sequenciamento genético humano conhecidas como NGS, chegou-se aos bancos de dados NoSQL como uma

das inovações de TI relevantes relacionadas ao tema. Posteriormente, identificaram-se as estratégias de proteção por direitos de propriedade intelectual e de comercialização por meio de serviços especializados adotadas pelas empresas no mercado internacional de software para esta inovação.

A partir da análise das estratégias de apropriação adotadas pelas empresas internacionais no caso da inovação dos bancos de dados NoSQL, é possível identificar algumas possibilidades de aplicação pelo mercado brasileiro de software quanto às estratégias de proteção por direitos de propriedade intelectual e de comercialização por meio de serviços especializados.

## CONCLUSÃO

Considerando-se que 95,3% das 5.519 empresas dedicadas ao desenvolvimento e produção de software no país classificam-se como micro ou pequena empresa (ASSOCIAÇÃO BRASILEIRA DAS EMPRESAS DE SOFTWARE, 2020), pode-se afirmar que uma estratégia com alto grau de inovação aberta, investindo no desenvolvimento de software livre de código aberto e na comercialização de suporte técnico especializado ao produto pela própria empresa, é uma alternativa com maior potencial de sucesso para o mercado brasileiro que o desenvolvimento de código proprietário protegido por *copyright*. Como demonstrado no caso dos bancos de dados NoSQL, empresas em busca de inserção no mercado e de menor porte, adotam a primeira estratégia, enquanto empresas maiores e consolidadas no mercado, exploram a segunda alternativa.

Da mesma forma como se pode observar especificamente no caso dos bancos de dados NoSQL, investir no suporte a produtos de software livre de código aberto desenvolvidos por terceiros, tanto nacionais quanto estrangeiros, apresenta-se como outra possibilidade viável às empresas brasileiras. Nestes casos, o foco é a comercialização do serviço especializado de suporte técnico. Portanto, cabe à empresa interessada identificar quais produtos de mercado apresentam uma perspectiva promissora, a fim de investir na contratação e qualificação da força de trabalho.

Tais estratégias baseiam-se na exploração do conhecimento tecnológico específico associado à inovação, e, portanto, pautada em apropriação por ativos complementares. Estes se constituem em um dos pilares do processo de convergência entre inovação e estratégia, suportado também pelo conceito de regime de apropriação. Este é definido pela natureza da tecnologia e pela eficácia da proteção por direitos de PI. Assim, as estratégias de apropriação de inovações a serem adotadas pelas empresas nacionais de software devem levar em consideração estes direitos. No caso da indústria de TI no Brasil, os ativos protegidos por direitos de PI incluem o direito de cópia e distribuição, o registro de marcas, o depósito de patentes e o registro de programa de computador.

Quanto ao direito de cópia e distribuição, adotar a estratégia de desenvolvimento de software livre de código aberto significa abdicar da proteção do código fonte e distribuição apenas do código binário através de *copyright*. No entanto, o conceito de software livre não elimina as questões relacionadas aos direitos de autor, cópia e distribuição, mas as utiliza de uma forma distinta a fim de permitir a colaboração no desenvolvimento do produto. Desta forma, cabe às empresas nacionais estudar e entender as licenças de distribuição de software livre disponíveis no mercado para adotar a mais adequada aos objetivos pretendidos.

O registro de marca é um destes ativos de PI, cuja definição é um sinal distintivo utilizado em atividades econômicas para diferenciar produtos ou serviços, permitindo a distinção entre estes e o reconhecimento da proveniência. A partir de uma reflexão da análise das páginas *wikis* mantidas pela ASF para os projetos Hadoop e Cassandra, pode-se identificar uma modalidade do uso das marcas registradas além daquela tradicional definição. Ao associarem-se a projetos da ASF, seja por meio da comercialização do suporte técnico ou do desenvolvimento de produtos derivados, é possível às empresas alcançar maior visibilidade e reputação para as respectivas marcas registradas. Tanto a primeira quanto a segunda acabam concedidas pela própria ASF. Além disso, cabe ressaltar a maior relevância dada pela fundação às empresas que porventura empreguem indivíduos com as quais a própria fundação possua um contrato de contribuição, os chamados *committers*, naturalmente concedendo a estas empresas ainda maior visibilidade aos citá-las de forma preferencial. Ou seja, um dos possíveis benefícios ao se associar a um ou mais projetos da ASF é criar uma identificação mercadológica da marca registrada da empresa com a marca registrada da ASF, elevando a primeira à mesma

reputação da segunda. Deve-se considerar que este fenômeno não está necessariamente relacionado a algum tipo de aporte financeiro, pois se conclui da análise das empresas listadas em Apache Cassandra Wiki (2019) e Apache Hadoop (2019) que a maior parte não consta como patrocinadora em The Apache Software Foundation (2019). É possível observar que esta modalidade aplica-se tanto às empresas que prestam o serviço de suporte técnico aos bancos de dados NoSQL livres de código aberto desenvolvidos pela ASF, quanto àquelas que desenvolvem produtos da mesma forma livres de código aberto derivados dos bancos de dados desenvolvidos pela instituição.

Outro ativo de PI a ser considerado na estratégia das empresas de software brasileiras são as patentes de invenção. A partir do estudo dos depósitos que citam os termos BigTable e NoSQL, identificou-se uma disputa entre grandes corporações multinacionais, divididas em dois grupos: empresas inovadoras e previamente consolidadas no mercado. Mesmo que não se possa afirmar, apenas por este resultado, que esta disputa seja um padrão para o universo de patentes das empresas de software, no âmbito de avaliação deste estudo, a análise dos depósitos de patentes relacionadas à tecnologia dos bancos de dados NoSQL demonstra que investir na proteção por este ativo de PI não é uma atividade a qual a grande parte das empresas brasileiras de software deva dispendar maiores recursos. Tal conclusão baseia-se no perfil majoritário de micro e pequenas empresas nacionais, identificado pela Associação Brasileira das Empresas de Software (2020), contrastante com o perfil de grandes corporações multinacionais das depositantes daquelas patentes.

Ainda assim, cabe às empresas brasileiras perceberem os documentos de patentes como fonte de informação tecnológica a fim de realizarem prospecção e identificação de tecnologias relevantes na indústria de software. Além disso, o estudo do uso que as empresas de menor porte fazem das patentes das quais são proprietárias pode indicar de que forma este ativo de PI pode propiciar um diferencial mercadológico para o negócio.

Ainda no âmbito da proteção por ativos de PI, outra estratégia a ser considerada pelas empresas nacionais é o registro de software. No Brasil, a proteção por direitos de propriedade intelectual de programa de computador está prevista na Lei nº 9.609 de 19 de fevereiro de 1998. Define a referida Lei no artigo 2º que tal proteção é a mesma conferida às obras literárias pela legislação de direitos

autorais. No artigo 3º, estipula-se que trechos do programa de computador devem fazer compor o pedido de registro, e que estas informações são sigilosas, somente podendo através de ordem judicial ou por requerimento do próprio titular. Ao abordar o tema do desenvolvimento de software pelas empresas nacionais à luz da legislação brasileira de registro de programa de computador, pode-se concluir que o conceito de software livre de código aberto vai de encontro à previsão de sigilo de código fonte disposto na lei. No entanto, esta legislação é atualizada e consolidada pela Lei nº 9.610 de 19 de fevereiro de 1998, que dispõe sobre a questão de direitos autorais. No capítulo reservado aos direitos patrimoniais, o artigo 29 estipula que depende de autorização do autor da obra a disposição e uso por quaisquer das modalidades previstas, dentre as quais se incluem a edição, a adaptação e a distribuição. Importante lembrar que obra, neste sentido, também se refere aos programas de computador, dada a proteção por direito de autor conferida pelo artigo 2º da Lei 9.609/1998. Assim, conforme identificado por Mendes (2006), o autor pode registrar o programa de computador e licenciá-lo como software livre, não havendo nenhuma incompatibilidade entre o registro de software no Instituto Nacional da Propriedade Industrial (INPI) e o licenciamento. Para a autora, a importância do registro junto ao INPI é a garantia da autoria e da anterioridade, integrando o software no portfólio de ativos intangíveis da empresa, os quais podem, inclusive, ser utilizados em negociações com outros agentes. Assim sendo, garantindo-se a autoria e anterioridade do software livre de código aberto, o registro de programa de computador pode ser utilizado em conjunto com a marca registrada da empresa com o intuito de se obter reputação no mercado nacional de empresas de software.

Também verificada na comercialização dos bancos de dados NoSQL, outra possibilidade a ser explorada pelas empresas nacionais é prestação de serviços por meio da computação em nuvem. De acordo com a Associação Brasileira das Empresas de Software (2020), ocorre um crescimento significativo na disponibilização deste tipo de solução no mercado brasileiro desde 2015, chegando a produzir mais de um bilhão de dólares em 2018. Esta estratégia pode ser adotada tanto por empresas que desenvolvam software livre de código aberto, disponibilizando o produto por meio do modelo *SaaS*, como também por empresas que ofereçam a estrutura necessária à produção de software pelas primeiras, por meio do modelo *PaaS*.

Especificamente quanto à prestação de serviços por meio da computação em nuvem, as empresas brasileiras inseridas neste mercado devem atentar para a Norma Complementar número 14 do Departamento de Segurança da Informação e Comunicações do Gabinete de Segurança Institucional da Presidência. Tal norma estabelece os princípios e diretrizes relacionados à segurança da informação para o tratamento da informação em ambiente de computação em nuvem para os entes da Administração Pública Federal (APF), direta e indireta. Em vigor desde 19 de março de 2018, determina que os dados e informações produzidos ou custodiados pelos órgãos ou entidades da APF residam exclusivamente em território nacional. Portanto, esta exigência de caráter obrigatório cria uma vantagem competitiva para as empresas nacionais que possuam capacidade computacional para a prestação de serviços através da computação em nuvem para o governo federal.

## REFERÊNCIAS

ANICETO, Rodrigo et al. Genomic Data Persistency on a NoSQL Database System. In: **IEEE International Conference on Bioinformatics and Biomedicine (BIBM)**, 2014, Belfast, Reino Unido, IEEE, p. 8-14.

APACHE CASSANDRA WIKI. Disponível em: <<https://cwiki.apache.org/confluence/display/CASSANDRA2/ThirdPartySupport>>. Acesso em: 5 nov. 2019.

APACHE HADOOP. Disponível em: <<https://cwiki.apache.org/confluence/display/HADOOP2/Distributions+and+Commercial+Support>>. Acesso em: 5 nov. 2019.

ASSOCIAÇÃO BRASILEIRA DAS EMPRESAS DE SOFTWARE. **Mercado Brasileiro de Software: panorama e tendências**. São Paulo, 2020.

BONNICI, Vincenzo et al. Comprehensive reconstruction and visualization of non-coding regulatory networks in human. **Frontiers in Bioengineering and Biotechnology**, v. 2, artigo 69, dez. 2014.

BOYLE, James. **The Public Domain**. New Haven: Yale University Press, 2008.

BREVERN, Alexandre G. de et al. Trends in IT Innovation to Build a Next Generation Bioinformatics Solution to Manage and Analyse Biological Big Data Produced by NGS Technologies. **BioMed Research International**, v. 2015, artigo 904541, 2015.

BREWER, Eric: Towards robust distributed systems. In: **Principles of Distributed Computing**, 19., 2000, Portland, Oregon., Estados Unidos, Association for Computing Machinery.

CARNEIRO, Ana Maria. **Proteção de ativos na indústria de software: estratégias e tendências de Propriedade Intelectual**. 2007. 215f. Tese (Doutorado em Política Científica e Tecnológica) – Instituto de Geociências, Universidade Estadual de Campinas, Campinas.

CARVALHO, Nuno Pires de. **A Estrutura dos Sistemas de Patentes e de Marcas – Passado, Presente e Futuro**. Rio de Janeiro: Lumen Juris, 2009.

CHANG, Fay et al. Bigtable: A Distributed Storage System for Structured Data. In: **USENIX Symposium on Operating Systems Design and Implementation (OSDI)**, 7., 2006, Berkeley, Califórnia, Estados Unidos, USENIX, p. 205-218.

CODD, E.F.. A Relational Model of Data for Large Shared Data Banks. **Communications of the ACM**, v.13, n.6, p. 377-387, jun. 1970.

DeCANDIA, Giuseppe et al. Dynamo: Amazon's Highly Available Key-value Store. In: **ACM Symposium on Operating Systems Principles (SOSP)**, 21., 2007, Stevenson, Washington, Estados Unidos, Association for Computing Machinery, p. 205-220.

DOSI, Giovanni. Technological paradigms and technological trajectories. **Research Policy**, v. 11, p. 147-162, 1982.

EISENSTEIN, Michael. The power of petabytes. **Nature**, v. 527, p. S2-S4, nov. 2015.

EVAN, Eric. NoSQL: What's in a name? Disponível em: <[http://blog.sym-link.com/2009/10/30/nosql\\_whats\\_in\\_a\\_name.html](http://blog.sym-link.com/2009/10/30/nosql_whats_in_a_name.html)>. Acesso em: 28 out. 2018.

ESWARAN, K.P. et al. The notions of consistency and predicate locks in a database system. **Communications of the ACM**, v. 19, n. 11, p. 624-633, nov. 1976.

FREEMAN, Christopher; PEREZ, Carlota. Structural crises of adjustment, business cycles and investment behaviour. In: DOSI, Giovanni et al. (eds.) **Technical Change and Economic Theory**. London: Pinter Publishers, 1988.

GILBERT, Seth; LYNCH, Nancy. Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services. **ACM SIGACT News**, v. 33, n. 2, p. 51-59, jun. 2002.

GNU AFFERO GENERAL PUBLIC LICENSE. Disponível em:  
<<https://www.gnu.org/licenses/agpl.html>>. Acesso em: 14 ago. 2017.

GOMES, Helton Simões. Pela 1ª vez, Brasil criará banco com mapeamento do genoma da população. **Uol**, São Paulo, 10 dez. 2019. Disponível em:  
<https://www.uol.com.br/tilt/noticias/redacao/2019/12/10/pesquisadores-anunciam-1-banco-de-sequenciamento-do-dna-dos-brasileiros.htm>. Acesso em: 19 dez. 2019.

GOLDIM, José Roberto; MATTE, Úrsula. Projeto Genoma Humano (HUGO), 4 ago. 2000. Disponível em: <<http://www.ufrgs.br/bioetica/genoma.htm>>. Acesso em: 10 ago. 2015.

GUIMARÃES, Valéria et al. A Study of Genomic Data Provenance in NoSQL Document-Oriented Database Systems. In: **IEEE International Conference on Bioinformatics and Biomedicine (BIBM)**, 2015, Washington D.C., Estados Unidos, IEEE, p. 1525-1531.

HAERDER, Theo; REUTER, Andreas. Principles of transaction-oriented database recovery. **Computing Surveys**, v. 15, n. 4, p. 287-317, dez. 1983.

HAVE, Christian Theil; JENSEN, Lars Juhl. Are graph databases ready for bioinformatics? **Bioinformatics**, v. 29, n. 24, p. 3107–3108, out. 2013.

HAYDEN, Erika Check. The \$1,000 Genome. **Nature**, v. 507, p. 294-295, mar. 2014.

HUIZINGH, Eelko K.R.E. Open innovation: State of the art and future perspectives. **Technovation**, v. 31, n. 1, p. 2-9, 12 nov. 2010. DOI 10.1016/j.technovation.2010.10.002. Disponível em:  
<https://doi.org/10.1016/j.technovation.2010.10.002>. Acesso em: 19 dez. 2019.



ISMAIL, Wazim Mohammed et al. The sequencing and interpretation of the genome obtained from a Serbian individual. **Plos One**, v. 13, n. 12, 19 dez. 2018. DOI 10.1371/journal.pone.0208901. Disponível em: <https://doi.org/10.1371/journal.pone.0208901>. Acesso em: 19 dez. 2019.

LAI, Eric. No to SQL? Anti-database movement gains steam, Computerworld, 1 jul. 2009. Disponível em: <<https://www.computerworld.com/article/2526317/database-administration/no-to-sql--anti-database-movement-gains-steam.html>>. Acesso em: 30 out. 2018.

LEAVITT, Neal. Will NoSQL Databases Live Up to Their Promise?. **IEEE Computer Society**, v. 43, n. 2, p. 12-14, fev. 2010.

LIMA, Iasmini et al. An Evaluation of Data Replication for Bioinformatics Workflows on NoSQL Systems. In: **IEEE International Conference on Bioinformatics and Biomedicine (BIBM)**, 2016, Shenzhen, China, IEEE, p. 896-901.

LUSCOMBE, N.M.; GREENBAUM, D.; GERSTEIN; M. What is bioinformatics? An introduction and overview. **Yearbook of Medical Informatics**, p. 83-100, 2001.

MARDIS, Elaine. A decade's perspective on DNA sequencing technology. **Nature**, v. 470, p. 198-203, fev. 2011.

MAY, Cristopher; SELL, Susan K. **Intellectual Property Rights: A Critical History**. Londres: Lynne Rienner Publishers, 2006.

MENDES, Cássia Isabel Costa. **Software Livre e Inovação Tecnológica: Uma Análise Sob a Perspectiva da Propriedade Intelectual**. 2006. 282f. Dissertação (Mestrado em Desenvolvimento Econômico, Espaço e Meio Ambiente) – Instituto de Economia, Universidade Estadual de Campinas, Campinas.

MELLO, Maria Tereza Leopardi. Propriedade Intelectual e Concorrência. **Revista Brasileira de Inovação**, v. 8, n. 2, p. 372-402, jul./dez. 2009.

MERELLI, Ivan et al. Managing, Analysing, and Integrating Big Data in Medical Bioinformatics: Open Problems and Future Perspectives. **BioMed Research International**, v. 2014, artigo 134023, 2014.

METZKER, Michael L. Sequencing technologies - the next generation. **Nature Reviews Genetics**, v. 22, p. 31-46, jan. 2010.

MOHAN, C. History repeats itself: sensible and NonsenSQL aspects of the NoSQL hoopla. In: **International Conference on Extending Database Technology**, 16., 2013, Gênova, Itália, EDBT, p. 11-16.

MOORE, Gordon E. Progress in digital integrated electronics. In: **International Electron Devices Meeting**, 1975, Washington D.C., Estados Unidos, IEEE, p. 11-13.

NATIONAL HUMAN GENOME RESEARCH INSTITUTE. DNA Sequencing Costs: Data, 2020. Disponível em: <<https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data/>>. Acesso em: 2 fev. 2020.

NATIONAL HUMAN GENOME RESEARCH INSTITUTE. Human Genome Project Results, 2018. Disponível em: <<https://www.genome.gov/human-genome-project/results>>. Acesso em: 2 fev. 2020.

NATIONAL HUMAN GENOME RESEARCH INSTITUTE. International Human Genome Sequencing Consortium, 2006. Disponível em: <<https://www.genome.gov/11006939/ihg-sequencing-centers>>. Acesso em: 2 fev. 2020.

NCBI. dbSNP Short Genetic Variations. Disponível em: <[https://www.ncbi.nlm.nih.gov/projects/SNP/snp\\_summary.cgi?view+summary=view+summary&build\\_id=151](https://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi?view+summary=view+summary&build_id=151)>. Acesso em: 18 dez. 2019.

NELSON, R. Richard. Economic development from the perspective of evolutionary economic theory. In: **Working Papers in Technology Governance and Economic Dynamics no. 2**. The Other Canon Foundation, Norway; Tallinn University of Technology, Tallinn, 2006.

O'CONNOR, Brian D.; MERRIMAN, Barry; NELSON, Stanley F. SeqWare Query Engine: storing and searching sequence data in the cloud. **BMC Bioinformatics**, v. 11, suplemento 12, artigo S2, 2010.

OBASANJO, Dare. Building Scalable Databases: Denormalization, the NoSQL Movement and Digg. Disponível em:

<<http://www.25hoursaday.com/weblog/2009/09/10/BuildingScalableDatabasesDenormalizationTheNoSQLMovementAndDigg.aspx>>. Acesso em: 24 nov. 2018.

PISANO, Gary. Profiting from innovation and the intellectual property revolution. **Research Policy**, v. 35, p. 1122-1130, 2006.

POSSAS, Silvia. Conhecimento e atividade econômica. **Economia e Sociedade**, v.6, n.1, p. 85-100, jun. 1997.

RAMELLO, Giovanni B. What's in a sign? Trademark law and economic theory. **Journal of Economic Surveys**, v. 20, n. 4, p. 547-565, 2006.

REHMAN, T.B. **Cloud Computing Basics**. Mercury Learning and Information, 2018.

ROBISON, Reid J.. How big is the human genome?. Disponível em: <<https://medium.com/precision-medicine/how-big-is-the-human-genome-e90caa3409b0>>. Acesso em: 14 out. 2019.

SADALAGE, Pramod J.; FOWLER, Martin. **Nosql Distilled: A Brief Guide to the Emerging World of Polyglot Persistence**. Nova Jersey: Addison-Wesley, 2012.

SCHULZ, Wade L. et al. Evaluation of relational and NoSQL database architectures to manage genomic annotations. **Journal of Biomedical Informatics**, v. 64, p. 288-295, 2016.

SEMPÉRÉ, Guilhem et al. WIDDE: a Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. **BMC Genomics**, v. 16, artigo 940, 2015.

SHERRY, Stephen T.; WARD, Minghong; SIROTKIN, Karl. dbSNP-Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. **Genome Research**, v. 9, n. 8, p. 677-679, ago. 1999.

SHERRY, S. T. et al. dbSNP: the NCBI database of genetic variation. **Nucleic Acids Research**, v. 29, n. 1, p. 308-311, jan. 2001.

STAJDOHAR, Miha et al. dictyExpress: a web-based platform for sequence data management and analytics in Dictyostelium and beyond. **BMC Bioinformatics**, v. 18, artigo 291, 2017.

STONEBRAKER, Michael; ÇETINTEMEL, Uğur: One Size Fits All: An Idea whose Time has Come and Gone. In: **International Conference on Data Engineering**, 21., 2005, Washington D.C., Estados Unidos, IEEE Computer Society, p. 2–11.

STONEBRAKER, Michael et al. One Size Fits All? – Part 2: Benchmarking Results. In: **Conference on Innovative Data Systems Research**, 3., 2007, Asilomar, Califórnia, Estados Unidos, CIDR, p. 173–184

STONEBRAKER, Michael et al. The end of an architectural era: (it's time for a complete rewrite). In: **International Conference on Very Large Data Bases**, 33., 2007, Viena, Áustria, VLDB Endowment, p. 1150–1160.

STRAUCH, Cristof. NoSQL Databases, 2011. Disponível em: <<http://www.christof-strauch.de/nosql dbs.pdf>>. Acesso em: 24 out. 2018.

SZMRECSÁNYI, Tamás. A herança schumpeteriana. In: PELAEZ, Victor.; SZMRECSÁNYI, Tamás. (orgs.). **Economia da Inovação Tecnológica**. São Paulo: Hucitec, 2006.

TAYLOR, Ronald C. An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics. **BMC Bioinformatics**, v. 11, suplemento 12, artigo S1, 2010.

TEECE, J. David. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. **Research Policy**, v. 15, p. 285-305, 1986.

THE APACHE SOFTWARE FOUNDATION. Disponível em: <<http://www.apache.org/>>. Acesso em: 6 nov. 2019.

THE INTERNATIONAL GENOME SAMPLE RESOURCE. Providing ongoing support for the 1000 Genomes Project data. Disponível em: <<https://www.internationalgenome.org/about>>. Acesso em: 14 dez. 2019.

THE 1000 GENOMES PROJECT CONSORTIUM. A map of human genome variation from population-scale sequencing. **Nature**, v. 467, p. 1061-1073, out. 2010.

THE 1000 GENOMES PROJECT CONSORTIUM. A global reference for human genetic variation. **Nature**, v. 526, p. 68-74, out. 2015.

WANG, Shicai et al. High dimensional biological data retrieval optimization with NoSQL technology. **BMC Genomics**, v. 15, suplemento 8, artigo S3, 2014.