

## Capítulo

# 3

## Ecossistemas de Dados na Web: da teoria aos desafios

Marcelo Iury S. Oliveira, Bernadette Farias Lóscio

### *Abstract*

*A number of initiatives have been developed to share and consume data on the Web. The growing interest in these initiatives drove the emergence of Data on Web Ecosystems, which provide an environment for the creation and management of initiatives of data sharing, as well as providing the necessary support to ensure the sustainability of such initiatives. These ecosystems promise a range of benefits for their participants, such as creating new business opportunities, generating innovation and creating value from the publication and consumption of data on the Web. In this course the main concepts related to this new environment are discussed, including theoretical relevant aspects as well as challenges and research opportunities in this area.*

### *Resumo*

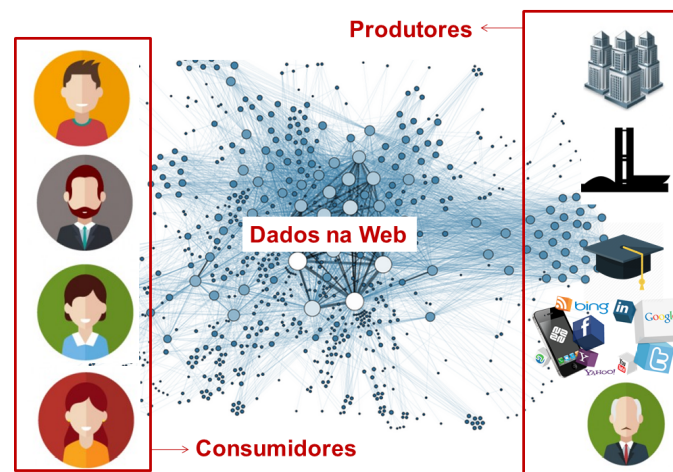
*Uma série de iniciativas vem sendo desenvolvidas em todo o mundo objetivando o compartilhamento e consumo dos dados na Web. O crescente interesse nessas iniciativas motivou o surgimento dos Ecossistemas de Dados na Web, os quais fornecem um ambiente propício para a criação e o gerenciamento de iniciativas de compartilhamento de dados, bem como oferecem o suporte necessário para garantir a sustentabilidade de tais iniciativas. Esses ecossistemas promovem uma série de benefícios para seus participantes, tais como a criação de novas oportunidades de negócios, a geração de inovação e a criação de valor a partir da publicação e do consumo dos dados na Web. Neste contexto, este minicurso discute os principais conceitos relacionados a este novo ambiente, abordando aspectos relevantes, tanto do ponto de vista teórico quanto de desafios e oportunidades de pesquisa nesta área.*

### **3.1. Introdução**

Uma série de iniciativas vem sendo desenvolvidas em todo o mundo objetivando o compartilhamento e consumo dos dados na Web. O crescente interesse nessas iniciativas

motivou o surgimento dos Ecossistemas de Dados na Web, os quais fornecem um ambiente propício para a criação e o gerenciamento de iniciativas de compartilhamento de dados, bem como oferecem o suporte necessário para garantir a sustentabilidade de tais iniciativas [Oliveira and Lóscio, 2018; Zuiderwijk et al., 2014].

Como visualizado na Figura 3.1, Ecossistema de Dados na Web são compostos por redes de atores autônomos que consomem, produzem ou fornecem, direta ou indiretamente, dados e outros recursos relacionados a dados publicados na Web (*e.g.*, software, serviços e infraestrutura) [Oliveira and Lóscio, 2018]. Em um Ecossistema de Dados, um ator pode desempenhar um ou mais papéis e pode estar conectado a outros atores por meio de relacionamentos, de modo que a colaboração e a competição dos atores promovem a autorregulação do ecossistema [Oliveira and Lóscio, 2018].



**Figura 3.1. Ecossistema de Dados na Web. Fonte:Autores**

A pesquisa em Ecossistemas de Dados na Web tem apresentado um crescente interesse, seja por parte da academia, seja pela indústria [Oliveira et al., 2019]. Os governos estão abrindo seus conjuntos de dados com objetivo seja para atender a preceitos democráticos (*e.g.*, transparência e prestação de contas) ou para permitir a melhoria nos seus serviços. Muitas empresas privadas tem liberado dados com objetivo de melhorar sua imagem e até aumentar os lucros, permitindo que grupos de pessoas analisem informações e obtenham resultados valiosos. Outras empresas tem devotado esforços na criação de mercados de dados para facilitar o comércio e o compartilhamento de dados, serviços e outros recursos. Um exemplo é o Bloomberg Marketplace<sup>1</sup> que reúne dados de uma variedade de fontes e provedores, organizando-os e disponibilizando-os para clientes que pagam por serviços de curadoria de dados .

Ecossistemas de Dados na Web não devem ser confundidos com a simples publicação de dados. De acordo com Pollock [2011], usualmente, o modelo básico atual para o fornecimento e uso de dados é uma via de mão única. Não há um retorno por dos parte usuários de dados e consumidores de dados, *i.e.*, os usuários não compartilham dados e conhecimento com os produtores de dados. Em um cenário ideal, os usuários deveriam contribuir com os produtores de dados. Essa contribuição pode incluir um simples

<sup>1</sup><https://www.bloomberg.com/professional/product/market-data/>

*feedback* de avaliação dos dados, sinalização de erros, envio de correções ou mesmo a contribuição com conjuntos de dados limpos e integrados.

De fato, todos podem e devem colaborar, quer sejam usuários ou produtores de dados. Inclusive, Ubaldi [2013] argumenta que os benefícios esperados do compartilhamento de dados somente serão alcançados por meio da criação de um Ecossistema de Dados nos quais produtores e consumidores de dados interagem entre si.

Um exemplo de um Ecossistema de Dados é uma comunidade de instituições relacionada a serviços e pesquisa de saúde. No setor de saúde, um grande volume de conjuntos de dados é produzido [Murdoch and Detsky, 2013]. Por exemplo, sistemas de prontuários eletrônicos coletam uma grande quantidade de dados [Murdoch and Detsky, 2013]. Cada paciente tem seu próprio registro digital, que inclui dados demográficos, histórico médico, alergias, resultados de exames laboratoriais, dentre outros. Os registros são compartilhados por meio de sistemas de informação e estão disponíveis para prestadores de serviços do setor público e privado. Na última década, empresas farmacêuticas e instituições de pesquisa também vêm acumulando dados de pesquisa e desenvolvimento em bancos de dados médicos. Pesquisadores e médicos podem usar essas grandes quantidades de dados a fim de auxiliar diagnósticos ou mesmo encontrar tendências e tratamentos que tenham as maiores taxas de sucesso no mundo real [Lebied, 2018]. Similarmente, hospitais podem usar dados de várias fontes para obter previsões diárias e horárias de quantos pacientes são esperados em um intervalo de tempo específico. Neste exemplo, os sistemas de prontuários eletrônicos, as empresas farmacêuticas e as instituições de pesquisa atuam como produtores de dados. Além disso, os pesquisadores, médicos e hospitais atuam como consumidores de dados. Os dados são, então, fornecidos por organizações públicas e privadas envolvidas em tratamento de saúde, planejamento urbano, monitoramento de tráfego e fiscalização de segurança. Cada um atuando de acordo com os seus objetivos e também esperando alcançar suas próprias metas.

Ecossistemas de Dados na Web promovem uma série de benefícios para os seus participantes, tais como a criação de novas oportunidades de negócios, inovação e a criação de valor a partir da publicação e do consumo dos dados na Web. No entanto, enquanto o potencial desses ecossistemas é real, a realização é malsucedida em muitos casos. Vários ecossistemas não são sustentáveis e, conseqüentemente, o esforço desempenhado pelos seus atores não é utilizado de forma adequada ou é desperdiçado. A falta de comunicação e cooperação entre os produtores de dados e os consumidores é um dos principais obstáculos para a obtenção de Ecossistemas de Dados sustentáveis. Além disso, projetar, desenvolver e manter sistemas adicionais para apoiar os ecossistemas de coleta de dados se tornam tarefas cada vez mais desafiadoras.

Neste contexto, este minicurso discute os principais conceitos relacionados a Ecossistemas de Dados na Web, abordando aspectos relevantes, tanto do ponto de vista teórico quanto de desafios e oportunidades de pesquisa nesta área. De maneira mais específica, serão discutidos os principais conceitos que descrevem um Ecossistemas de Dados na Web, soluções para modelagem dos ecossistemas, arquiteturas e modelos de negócio para Ecossistemas de Dados na Web, o papel de boas práticas para na sustentabilidade dos ecossistemas e os principais desafios de pesquisa existentes na área.

### 3.2. *Background Teórico*

A noção de ecossistemas origina-se nas áreas de Ecologia e Biologia. Os ecologistas usam o termo ecossistema (biológico) como uma unidade natural que consiste em todas as plantas, animais e micro-organismos em uma área que funciona em conjunto com todos os recursos físicos não vivos do ambiente [Wikipedia, 2001]. O conceito de Ecossistema de Dados também abrange ideias de outros ecossistemas, tais como Ecossistema de Negócios, Ecossistema Digital e Ecossistema de Software.

Moore [1999] define um Ecossistema de Negócios como uma comunidade econômica apoiada por um ator base que interage com organizações e indivíduos, incluindo clientes, produtores líderes, concorrentes e outras partes interessadas. Essas comunidades podem se unir de uma maneira parcialmente intencional, altamente auto-organizada e até acidental. Bem como, podem ser criadas em torno de atores-chave que são empresas dominantes com forte influência sobre os processos co-evolucionários. Os membros das comunidades também incluem fornecedores, produtores, concorrentes e outras partes interessadas, como organizações de financiamento, associações comerciais, órgãos normativos, sindicatos trabalhistas e instituições governamentais. Além disso, os Ecossistemas de Negócios são baseados em recursos essenciais que são explorados para produzir inovação, negócios disruptivos ou produtos essenciais.

Iansiti and Levien [2004] apresentam outra visão para Ecossistema de Negócios inspirada em Ecossistemas Biológicos. Segundo eles, um Ecossistema de Negócios é caracterizado por uma rede de negócios na qual vários atores interconectados dependem uns dos outros para a sobrevivência e eficiência mútua. Um Ecossistema de Negócios compreende todos os atores que contribuem para o desenvolvimento de processos que influenciarão a integridade de longo prazo da rede [Iansiti and Levien, 2004]. Se o ecossistema é saudável, atores individuais prosperam. Se o ecossistema não é saudável, os atores individuais sofrem profundamente.

O termo Ecossistema (de Negócios) Digitais foi proposto pela primeira vez em [Nachira, 2002; Nachira et al., 2007]. Tendo sido concebido principalmente como uma estratégia ou política de adoção efetiva de Tecnologias de Informação e Comunicação (TICs) na União Européia. Um Ecossistema Digital faria com que indústrias, setores ou regiões se tornassem mais inovadores e competitivos nos mercados globais. Um Ecossistema Digital é considerado uma evolução natural dos Ecossistemas de Negócios.

No entanto, o termo Ecossistema Digital tem sido usado para descrever uma variedade de conceitos. Por exemplo, Fiorina [2000] refere-se à infra-estrutura de rede existente relacionada à Internet, que envolve várias empresas que oferecem serviços digitais, bem como clientes para usar os serviços digitais oferecidos. O termo também está sendo cada vez mais associado à adoção de *e-business* e *e-commerce* para proporcionar maior crescimento, empregos cada vez mais qualificados e maior inclusão social. Outra perspectiva é apresentada em [Sondergaard, 2017] que define um Ecossistema Digital como uma comunidade de empresas e indivíduos interdependentes que compartilham plataformas digitais padronizadas para um propósito mutuamente benéfico (*e.g.*, inovação ou criação de valor).

Ecossistemas de Software também são um termo recente, referindo-se a organiza-

ções de rede ou indivíduos que baseiam suas relações no desenvolvimento, comércio e uso de uma tecnologia de software. Os Ecossistemas de Software quebram as fronteiras internas da linha de produção das organizações, permitindo contribuições de desenvolvedores externos, fornecedores e outras partes externas. Como consequência, os Ecossistemas de Software criam dependências que antes não existiam entre os componentes e as organizações associadas.

Nesse sentido, Jansen et al. [2009] definem Ecossistemas de Software como um conjunto de atores de negócios interagindo com um mercado compartilhado de software e serviços, juntamente com as relações entre eles. Na maioria dos casos, os Ecossistemas de Software são sustentados por plataformas tecnológicas ou mercado comuns com as relações sendo realizadas pela troca de informações, recursos e artefatos. Outra definição proposta por Bosch and Bosch-Sijtsema [2010] define um Ecossistema de Software como “um conjunto de soluções de software que permitem, suportam e automatizam as atividades e transações dos atores nos ecossistemas sociais ou de negócios associados e as organizações que fornecem essas soluções”. Essa definição enfatiza mais a automação da atividade, bem como o interesse comum em software e seu uso.

Assim como os ecossistemas acima apresentados, um Ecossistema de Dados envolve múltiplos elementos, incluindo dados, software e outros recursos computacionais, fluxos de trabalho, pessoas, mercado, governo e infraestrutura. Esses elementos sugerem que um Ecossistema de Dados precisa combinar componentes de diferentes ecossistemas. Desta forma, um Ecossistema de Dados pode ser visto como um ecossistema híbrido, ou seja, que contempla características dos diferentes ecossistemas mencionados.

É preciso ressaltar que, apesar de compartilhar as características de rede e co-evolução, os Ecossistemas de Dados também diferem dos ecossistemas anteriores. Ao contrário de outros ecossistemas, os Ecossistemas de Dados não dependem de uma plataforma comum explícita na qual diferentes atores possam colaborar. A plataforma comum é, na verdade, os vários conjuntos de dados disponibilizados, trocados e consumidos pelos atores. Em particular, os dados não precisam necessariamente ser fornecidos por um único ator. A falta de uma plataforma comum cria uma rede de oferta e demanda mais difusa. Outra diferença está relacionada a como os produtos são negociados entre os atores. Em Ecossistemas de Negócios, operações de negócios e atores são *per se* os produtos [Manikas and Hansen, 2013a]. Em Ecossistemas de Software, os produtos são componentes ou serviços de software. Em Ecossistemas de Dados, o principal produto são os dados e seus derivados.

### 3.3. Definições para Ecossistemas de Dados

Há pouco consenso sobre a nomenclatura e a definição de Ecossistemas de Dados. Embora uma discussão mais aprofundada sobre a terminologia esteja além do escopo deste minicurso, a fim de poder analisar o campo do Ecossistema de Dados, bem como guiar o processo de estudo, revisaremos algumas das definições existentes.

Primeiramente, é preciso destacar que um grande número de estudos não definem o termo Ecossistema de Dados. Segundo Oliveira et al. [2019], 13 dos 29 trabalhos encontrados na literatura não apresentam nenhuma definição para Ecossistemas de Dados. No entanto, alguns desses estudos fazem referência a estudos anteriores. Na maioria dos

casos, isso acontece porque estes estudos focam, de maneira secundária, em Ecossistema de Dados. Em outros casos, os autores consideram um estudo prévio (escrito por eles mesmos ou por outros pesquisadores) como a base para as definições necessárias.

Pollock [2011] fornece a definição mais antiga para um ecossistema de dados.. De acordo com ele, *“Um ecossistema tem ciclos de dados, nos quais consumidores e intermediários de dados (e.g., desenvolvedores de aplicativos) podem compartilhar seus dados limpos, integrados e empacotados no ecossistema de forma reutilizável. Geralmente, esses dados limpos e integrados são mais valiosos do que a fonte original”*. Essa definição enfatiza a necessidade de ciclos de dados para criar Ecossistemas de Dados. Além do ciclo, a visão de Pollock exige que os atores desempenhem papéis, como produtores, intermediários e consumidores.

Similarmente, de acordo com o Zubcoff et al. [2016], um Ecossistema de Dados é composto de muitos atores e pequenas estruturas organizacionais que devem reconhecer dados como a matéria-prima. Os atores devem formar um ciclo a fim de “alimentar” o ecossistema, proporcionando benefícios a todas as partes. Essa visão do Ecossistema de Dados também defende um ciclo, bem como aponta para a existência de múltiplos atores, cada um com suas próprias expectativas.

Por sua vez, Harrison et al. [2012] vislumbram a ideia de um Ecossistema Governamental, que é uma espécie de Ecossistema de Dados Abertos. Segundo eles, *“Um Ecossistema Governamental prevê organizações governamentais como atores centrais, tomando a iniciativa em redes organizadas para atingir objetivos específicos relacionados à inovação e governança”*. Eles também complementam essa definição, afirmando que a metáfora do ecossistema *é frequentemente usada por formuladores de políticas e acadêmicos [...] para transmitir um senso de um sistema social interdependente de atores, organizações, infraestruturas materiais e recursos simbólicos, todos suportados através do uso intensivo de tecnologia e informação*. Assim como Pollock [2011], Harrison et al. [2012] defendem que os papéis devem ser definidos, mas também enfatizam a ideia de um papel fundamental que controla e coordena o ecossistema. Além disso, eles também reconhecem um conjunto de fatores contextuais (e.g., aspecto social) como um elemento-chave de um Ecossistemas de Dados. Além disso, Harrison et al. [2012] combina a metáfora do ecossistema com o conceito de múltiplas e variadas inter-relações entre produtores, usuários, dados, infraestrutura material e instituições.

Uma perspectiva diferente é apresentada por Zuiderwijk et al. [2016], que define Ecossistemas de Dados como *“todas as atividades para liberar e publicar dados na Internet, para quais usuários de dados podem conduzir atividades como pesquisar, localizar, avaliar e visualizar dados e suas licenças relacionadas, limpeza, análise, enriquecimento, combinação, vinculação e visualização de dados e interpretação e discussão de dados e fornecimento de feedback ao produtor de dados e outras partes interessadas”*. Similarmente, o Ding et al. [2011] define um Ecossistemas de Dados como *“um sistema baseado em dados no qual stakeholders de diferentes tamanhos e funções encontram, gerenciam, arquivam, publicam, reutilizam, integram e consomem dados em conexão com ferramentas, serviços e ferramentas online”*. Ambas as definições apresentam a ideia de atividades que desenvolveriam algum valor ou benefício para os atores que usam dados. Essas atividades podem ser atribuídas a funções específicas que serão desempenhadas pelos atores

(i.e., *stakeholders*).

Assim, de acordo com essas definições, os Ecossistemas de Dados contam com um vasto e heterogêneo conjunto de atores, cada um com propriedades, capacidades e expectativas diferentes. Os atores podem produzir e consumir recursos usando atividades diferentes e sob diferentes condições. Além disso, muitos desses elementos são dinâmicos e evoluem com o tempo. Podemos concluir que o panorama do Ecossistema de Dados é composto por um conjunto de atores que relacionam-se entre si por meio da troca de um conjunto de recursos distribuídos, heterogêneos, dinâmicos e em evolução.

### 3.4. Principais Componentes de Ecossistemas de Dados na Web

De acordo com Oliveira and Lóscio [2018]; Oliveira et al. [2018], apesar da abstração das características apresentadas na literatura, quatro elementos principais destacam-se (como apresentado na Figura 3.2), sendo eles: (1) atores, (2) papéis, (3) relacionamentos e (4) recursos.

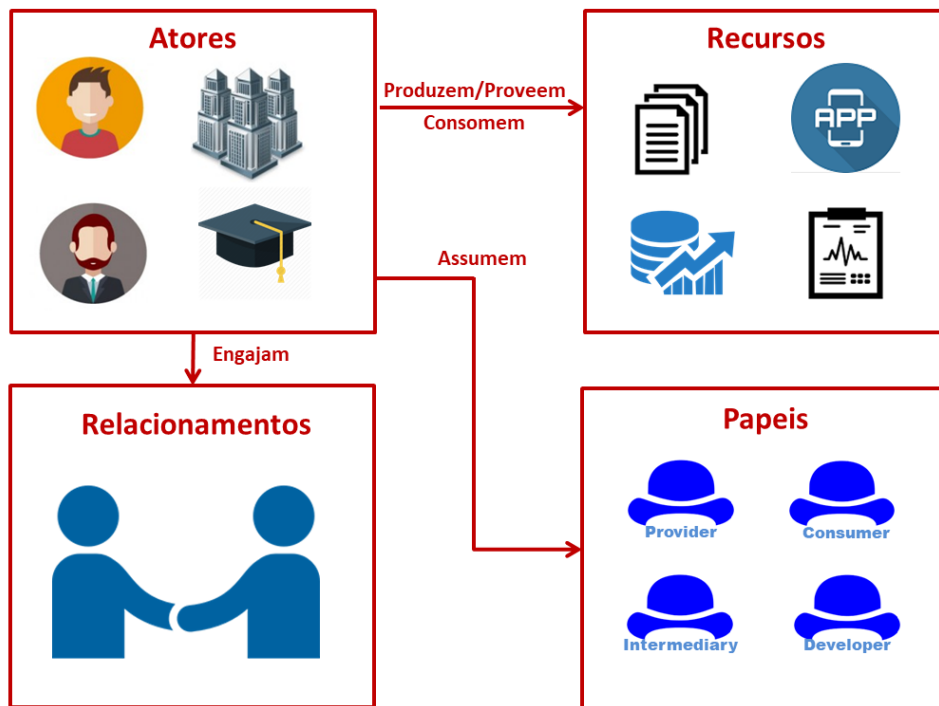


Figura 3.2. Elementos Básicos de Ecossistema de Dados na Web. Fonte: Autores

Um *Ator* é uma entidade autônoma, como uma empresa, instituição ou indivíduo, que desempenha um ou mais papéis específicos em um Ecossistema de Dados. Um conjunto de interesses motiva os atores e cada um deles tem diferentes expectativas e capacidades. Os atores geralmente se comprometem voluntariamente com o ecossistema, mas recebem incentivos para serem atuarem no ecossistema.

Um *Papel* é uma função desempenhada por um ator em um Ecossistema de Dados. Está relacionado a um conjunto de deveres e atividades. Diversos papéis podem ser identificados no Ecossistema de Dados. Em geral, pelo menos os consumidores de dados

e os produtores de dados são identificados nos Ecossistemas de Dados. Porém, existem vários papéis adicionais responsáveis por diferentes tarefas e atividades. Além disso, os papéis podem ou não sobrepor suas responsabilidades.

Atores vinculados a um papel devem possuir a capacidade de cumprir os compromissos que um papel lhes impõe. Além das capacidades, os atores também exigem recursos adequados para fornecer e consumir dados. Exemplos comuns de recursos são conjuntos de dados, serviços, ferramentas, capital financeiro, bem como capital humano, equipamentos, materiais e tecnologia proprietária.

Um *Relacionamento* representa interações entre os atores do Ecossistema de Dados. Os relacionamentos, geralmente, são baseados em um interesse comum ou também estão relacionados ao papel que cada ator desempenha no ecossistema. No mais, eles podem variar de acordo com aspectos econômicos, políticos, culturais e/ou tecnológicos. Finalmente, os relacionamentos são restritos por fatores físicos ou não físicos (*e.g.*, recursos) [Bosch, 2009], assim como produzem valor e envolvem custos.

Um *Recurso* é um produto útil e valioso ou uma capacidade produzida, fornecida, gerenciada ou consumida pelos atores. Em Ecossistemas de Dados, os recursos variam desde conjuntos de dados e software baseado em dados até a infraestrutura. Em particular, o software baseado em dados inclui ativos reutilizáveis (componentes e serviços) ou ativos de software (aplicativos) usados para consumir, produzir ou fornecer dados. Os recursos podem ser trocados individualmente ou em combinação por meio de transações de relacionamentos. Os recursos geralmente estão em conformidade com um conjunto de padrões que também são limitados por um conjunto de licenças, que definem como o recurso pode ser explorado. Além disso, os recursos são geralmente avaliados de acordo com diferentes propriedades de qualidade.

Em relação a recursos, Zuiderwijk et al. [2015] os distinguem entre três categorias: recursos humanos, recursos de dados e recursos de TI. Recursos humanos referem-se a indivíduos que usam seus recursos para explorar dados. Os recursos de dados se referem aos ativos baseados em dados estáticos e dinâmicos, como bancos de dados, bases de conhecimento ou simplesmente conjuntos de dados. Os recursos de TI referem-se ao hardware (*e.g.*, infraestruturas, redes e computadores), plataformas e aplicativos (software). Na verdade, os atores não precisam necessariamente possuir, gerenciar ou operar os recursos subjacentes, mas podem consumir ou contratar esses recursos por meio de outros atores, como prestadores de serviços ou outros tipos de atores intermediários.

Ainda em relação aos elementos que compõem um Ecossistema de Dados, Mercado Lara and Gil-Garcia [2014] agrupam os elementos em três domínios: (i) políticas e práticas governamentais, (ii) inovadores; uma combinação de tecnologia, negócios e governo e (iii) usuários, sociedade civil e negócios. Shin and Choi [2015] também identificam como elementos-chave de um Ecossistema de Dados: (i) infra-estrutura, (ii) software e tecnologias, (iii) serviço e aplicações, (iv) normas, (v) usuários, (vi) fatores sociais e culturais, (vii) governo e (viii) indústria.

De uma forma geral, todos os elementos do Ecossistema de Dados estão interconectados de forma que quando um elemento é alterado, os efeitos podem ser sentidos em todo o sistema [Immonen et al., 2014]. De fato, os atores afetam e são afetados pela



criação e entrega de recursos executados pelos outros atores. Além disso, os interesses dos atores podem levar a conflitos. Por exemplo, os consumidores de dados são fortemente influenciados pela decisão de um produtor de dados de não publicar ou atualizar um determinado dado, alterar o formato em que os dados são publicados, comprometer a qualidade dos dados ou alterar como ele pode ser usado.

### 3.5. Modelagem em Ecossistemas de Dados na Web

As técnicas de modelagem desempenham um papel importante no suporte ao design e ao desenvolvimento de sistemas complexos. Em geral, os modelos permitem compartilhar uma visão e conhecimento comuns entre as partes interessadas técnicas e não técnicas, facilitando e promovendo a comunicação entre elas. Além disso, os modelos tornam o planejamento do projeto mais eficaz e eficiente, proporcionando uma visão mais apropriada do sistema a ser desenvolvido e permitindo que o controle do projeto seja alcançado de acordo com critérios objetivos Brambilla et al. [2012]; OMG [2016].

Na área de Ecossistemas de Dados, modelos conceituais tanto permitem práticas de gerenciamento mais eficientes quanto auxiliam na descrição do conhecimento sobre os recursos e outras características de um Ecossistema de Dados. Além disso, modelos conceituais podem ser usados como linguagem de metamodelos para desenvolver ferramentas CASE (engenharia de software auxiliada por computador) para ajudar os profissionais a construir modelos conceituais derivados que representariam Ecossistemas de Dados específicos (e.g., Ecossistemas de Dados Biomédicos e Ecossistemas de Dados Financeiros).

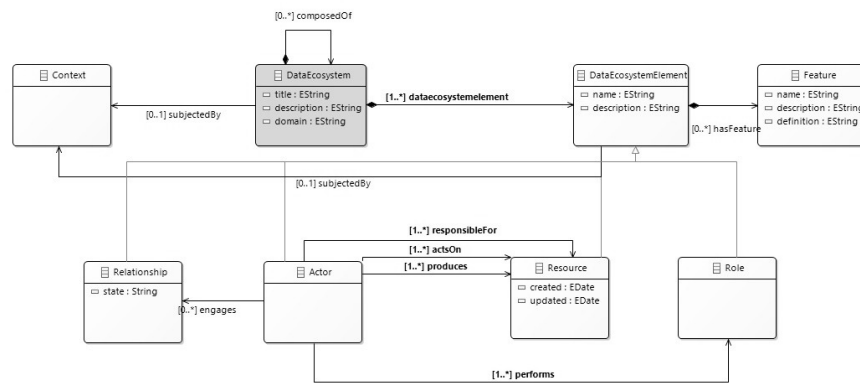
Diante deste contexto, em Oliveira et al. [2018], é apresentada uma proposta de meta-modelo que define os conceitos fundamentais de Ecossistema de Dados e seus relacionamentos para permitir a análise e descrição destes ecossistemas.

A Figura 3.3 apresenta os principais elementos do meta-modelo proposto por Oliveira et al. [2018]. Este meta-modelo foi formalizado por meio do Eclipse Modeling Framework (EMF) <sup>2</sup> e na linguagem de meta-modelagem ECore fornecida pelo EMF. O meta-modelo formaliza os conceitos Ator, Relacionamento, Papel e Recursos apresentados na seção anterior. Além disso, ele introduz novos conceitos como Contexto, Elementos Contextuais, Modelos de Negócios, Transações e entre outros.

O meta-modelo proposto por Oliveira et al. [2018] foi o primeiro voltado especificamente para Ecossistemas de Dados. O mesmo, apesar de formalizar conceitos importantes, ainda carece de diagramas que permitam a praticantes não-técnicos visualizarem de forma gráfica Ecossistemas de Dados. Diante deste contexto, trabalhos de outras áreas poderiam ser adaptados para o contexto de Ecossistemas de Dados de forma a permitir o desenho gráfico do estado de um ecossistemas. Por exemplo, os trabalhos [Brinkkemper et al., 2009; Yu and Deng, 2011] propõem soluções para modelagem de Ecossistemas de Software. Estas soluções permitem identificar os atores e suas relações, assim como fornecem uma visão rápida dos ecossistemas modelados.

Brinkkemper et al. [2009] apresenta um modelo que consiste em dois diagramas: PCD e SSN. O *Product Context Diagram* (PCD) descreve o contexto de um produto de software. Já o diagrama *Supply Network* (SSN) descreve as diferentes partes envolvi-

<sup>2</sup><http://www.eclipse.org/modeling/emf/>



**Figura 3.3. Meta-Modelo para descrição de Ecossistemas de Dados**

das na entrega e implementação de um produto de software ou serviço. Por sua vez, Yu and Deng [2011] apresenta uma abordagem de modelagem estratégica baseada na estrutura de modelagem  $i^*$  para ajudar a entender os ecossistemas de software. Os modelos  $i^*$  são usados para descrever dependências estratégicas entre fornecedores de software, desenvolvedores de terceiros e usuários finais, além de ajudar a explorar e raciocinar sobre formas alternativas de atingir metas estratégicas para cada ator. Ambos os trabalhos podem ser adaptados para mapeamento dos atores de um Ecossistema de Dados na Web.

### 3.6. Papéis em Ecossistemas de Dados na Web

Diversos papéis podem ser identificados em Ecossistemas de Dados. Inclusive, é possível identificar dois ou mais papéis compartilhando as mesmas tarefas. Além disso, as características do Ecossistema de Dados podem levar à necessidade de definição de papéis mais específicos. Por exemplo, em Ecossistemas de Dados baseados em dados médicos/de saúde, geralmente há papéis responsáveis por avaliar questões éticas.

O papel mais frequentemente identificado é o usuário de dados, responsável pelo consumo direto ou indireto de dados. Esse papel é apresentado com uma miríade de nomes nos estudos, por exemplo, usuários finais, consumidores de dados, beneficiários de dados, dentre outros. Os usuários de dados não têm necessariamente a capacidade de consumir dados diretamente dos produtores de dados. Eles geralmente dependem de serviços fornecidos por *Re-Users* (usuários que criam valor a partir dos dados disponíveis na Web), intermediários de dados ou provedores de serviços. Além disso, os usuários de dados geralmente representam os usuários finais de um ecossistema.

O segundo papel mais destacado é o provedor de dados, responsável pela publicação ou fornecimento de dados. Existem também alguns estudos que apresentam papéis menores relacionados ao fornecimento de dados. Por exemplo, o Immonen et al. [2014] apresentam papéis secundários relacionados ao fornecimento de dados, como “Armazenador para coletar e salvar dados, um Desenvolvedor para gerenciar e processar dados, um Agregador para combinar e editar dados de diferentes fontes, um Harmonizador para padronizar e homogeneizar dados de diferentes fontes, um Atualizador para atualizar informações, um Editor para publicar os dados e um Registro para manter a administração

de recursos de dados”.

É importante notar que os provedores de dados não são necessariamente responsáveis pela geração de dados. Essa responsabilidade pode ser atribuída a outro papel, chamado Produtor ou Criador de Dados, responsável pela captura ou geração de dados. Esse papel também pode compilar, agregar e empacotar dados.

Outro papel identificado é um Re-User responsável por agregar valor aos dados a serem reutilizados. De acordo com o Köster and Suárez [2016], o Re-User é responsável pelo uso de dados para desenvolver aplicativos ou serviços destinados a usuários de dados.

O papel de ator-chave (do inglês, *Keystone Actor*) é responsável por impulsionar as forças por trás do ecossistema, além de fornecer estabilidade em ambientes instáveis [Iansiti and Levien, 2004]. Esse papel é muito comum em Ecossistemas de Software. Em Dawes et al. [2016], os pesquisadores afirmaram que, em Ecossistemas de Dados, os atores-chave são responsáveis por fornecer a maioria dos dados, bem como por promover o ecossistema. Lee [2014] afirmam que esse papel deve ser atribuído a atores que lideram os programas de Dados Abertos Governamentais.

Embora tenhamos apresentado alguns papéis, ainda há uma miríade de diferentes papéis na literatura. Oliveira et al. [2019] realizaram um levantamento na literatura de papéis relacionados a Ecossistemas de Dados. No total, foram identificados 13 papéis primários e mais 22 papéis secundários (*i.e.*, um papel especializado que é responsável por alguns dos deveres e atividades de um papel primário). No entanto, os papéis são definidos de forma superficial. Na maioria dos estudos, os papéis sequer são especificados. Vários dos estudos apenas listam os atores do Ecossistema de Dados, deixando para leitor a responsabilidade de identificar e classificar suas funções.

### **3.7. Estruturas de Organização de Ecossistemas de Dados na Web**

Em um Ecossistema de Dados, cada ator é conectado a outros atores por um conjunto de interesses ou modelos de negócios. Toda a rede de relacionamentos pode seguir uma estrutura organizacional, variando de uma abordagem difusa para uma abordagem centrada em um ator-chave. A organização de um Ecossistema de Dados leva em conta tanto a maneira como os atores estão conectados e as propriedades de seus relacionamentos (por exemplo, dependência) [Manikas and Hansen, 2013b]. Estudar a forma de organização dos Ecossistemas de Dados é importante para entender e governar a interação e participação dos atores [Christensen et al., 2014].

De acordo com Oliveira et al. [2019], é possível identificar na literatura 5 diferentes abordagens de organização, sendo elas: centrada em ator-chave, baseada em intermediários, centradas em plataforma, baseada em marketplace e orientada a modelos de negócio.

Na organização centrada em ator-chave, os atores são organizados em torno de um ator-chave, que é direta ou indiretamente responsável por fornecer grande parte dos dados. Contudo, o ator-chave não tem controle completo sobre os outros atores. Eles podem sair (ou entrar) no ecossistema a qualquer momento [Heimstädt et al., 2014]. Já a organização baseada em intermediários depende da presença de atores que atuem como intermediários de dados para gerar valor a partir dos dados.

Na organização centrada em plataforma, uma plataforma fornece infraestrutura e serviços para suportar o fornecimento e o consumo de dados. Dawes et al. [2016]; Ding et al. [2011] enfatizam que os custos de fornecimento de dados são reduzidos quando os dados são liberados por meio da plataforma. A plataforma também pode atenuar problemas de interoperabilidade e usabilidade. Ferramentas de catalogação de dados abertos (e.g, CKAN<sup>3</sup>) tem sido comumente utilizadas para criar Ecossistemas de Dados na Web.

Em organizações baseadas em marketplaces, os marketplaces fornecem infraestrutura, modelos de negócios, regras e serviços necessários para transações de dados e software entre os atores [Smith et al., 2016]. Em geral, os marketplaces abrangem uma plataforma técnica com capacidade de vincular produtores de dados e usuários de dados. Eles também permitem a venda de dados, serviços e aplicativos.

Apesar de não definir como os atores devem ser organizados, alguns estudos apresentam modelos de negócios, que descrevem a lógica de como um ator cria, entrega e captura valor. Em particular, valor refere-se a qualquer benefício que um agente obtém do Ecossistema de Dados, como satisfação, utilidade, solução de problemas ou receita.

### 3.8. Gerenciamento de Ecossistemas de Dados na Web

Enquanto o potencial dos Ecossistemas de Dados é real, a sua realização é malsucedida em muitos casos [Heimstädt et al., 2014; Mercado Lara and Gil-Garcia, 2014; Zuiderwijk et al., 2012]. De acordo com o Dawes et al. [2016], como consequência de inúmeras barreiras e limitações, o desempenho das iniciativas de Ecossistema de Dados tende a ser simplista. Várias iniciativas se concentram em lançar e promover concursos de curta duração, como *hackathons* e *code fests*. Até os aplicativos desenvolvidos para esses concursos apresentam resultados insatisfatórios [Gama and Lóscio, 2014]. A maioria das aplicações nesses cenários acaba sendo rapidamente abandonada.

O estabelecimento correto de Ecossistema de Dados significa a coordenação adequada de várias categorias de atores, a provisão de apoio às empresas, o estímulo ao desenvolvimento e uso/reúso de recursos [Koznov et al., 2016]. Outros elementos essenciais para um Ecossistema de Dados de sucesso são a colaboração entre atores, a integração de informações, a preservação dos processos e o gerenciamento adaptativo [Zuiderwijk et al., 2014].

De fato, o gerenciamento de Ecossistemas de Dados é importante para facilitar ativamente o funcionamento efetivo e cumprir as metas dos ecossistemas [Mercado Lara and Gil-Garcia, 2014]. Além disso, se um Ecossistema de Dados não possui uma estrutura de gerenciamento, torna-se difícil impulsionar o ecossistema, construir e aprender com experiências passadas [Lee, 2014].

Os Ecossistemas de Dados funcionarão bem somente se forem projetados considerando sua complexidade por completo. De acordo com Mercado Lara and Gil-Garcia [2014], o gerenciamento de um Ecossistema de Dados requer o esboço de alguns tópicos básicos, que se concentram em (i) identificar os atores mais ativos que atuam como componentes essenciais do ecossistema; (ii) analisar a natureza das transações que ocorrem entre esses atores; (iii) reconhecer quais recursos são necessários para cada ator e como

---

<sup>3</sup><https://ckan.org/>

eles envolvem as transações; e (iv) estudar os indicadores que sinalizam o status da atividade do ecossistema [Mercado Lara and Gil-Garcia, 2014]. Portanto, essas considerações exigem uma abordagem mais sistêmica para planejar, projetar e coordenar Ecossistemas de Dados.

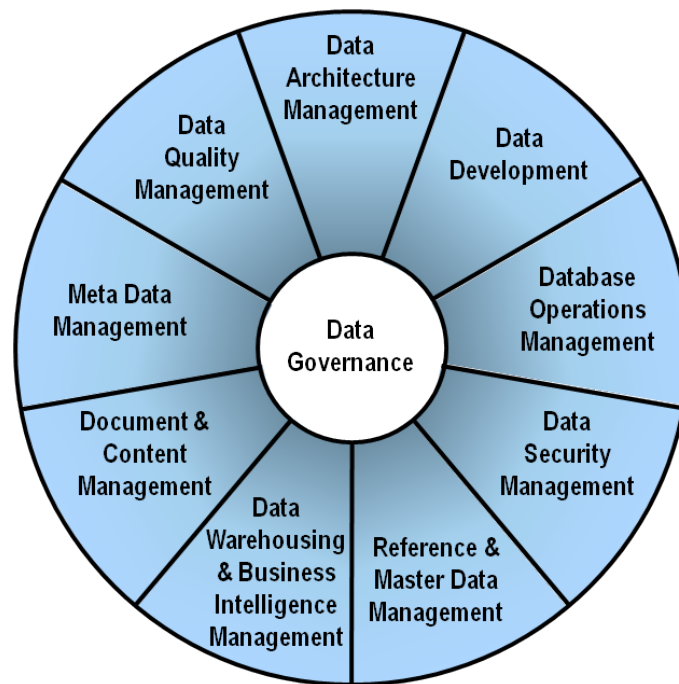
No entanto, até o momento, as iniciativas de gerenciamento de Ecossistemas de Dados são simplistas. Por exemplo, vários governos desenvolveram programas e políticas que visam promover o fornecimento e o uso de dados do setor público [Koznov et al., 2016; Mercado Lara and Gil-Garcia, 2014]. Tais políticas muitas vezes se concentram em garantir a disponibilidade e a qualidade dos recursos de dados. Essas políticas de gerenciamento ajudaram a expandir os Ecossistemas de Dados e melhorar o fornecimento de dados. No entanto, essas políticas não incluem outros atores-chave, como consumidores de dados e intermediários de dados, que realmente demandam o fornecimento. Por isso, é crucial incluir, desde o início, o ponto de vista de todos os atores do ecossistema. Uma gestão integrada e colaborativa deve garantir que as metas incluídas em uma agenda de Ecossistema de Dados atendam às necessidades, direitos e interesses de todos os atores que fazem parte do ecossistema [Köster and Suárez, 2016].

Soluções como modelos de maturidade, *frameworks* e metodologias propostas para outras áreas podem ser adaptadas para o contexto de Ecossistemas de Dados. Tais soluções propõem diversos instrumentos de gerenciamento que poderiam ser usados para um melhor monitoramento e sistematização das atividades de Ecossistemas de Dados. Entre essas soluções destacamos: DMBOK, ISO 8000 e DMM.

O DMBOK (do inglês, *Data Management Body of Knowledge*) consiste de um corpo de conhecimento que compila um conjunto de processos e práticas para servir como um guia abrangente para as atividades de gerenciamento de dados. O DMBOK foi desenvolvido em 2009 com a colaboração de mais de 120 profissionais. Ele fornece uma visão geral do gerenciamento de dados, além de fornecer definições de processos de gerenciamento de dados, funções e resultados de entrega e sua terminologia padrão.

O DMBOK possui 10 funções principais de gerenciamento de dados, conforme apresentado na Figura 3.4. Tomando a governança de dados como o núcleo, ela narra o escopo de cada função pela rotação no sentido horário. Cada uma das funções é detalhada como [Mosley et al., 2010]:

- **Data Governance:** é responsável pelo planejamento, supervisão e controle de alto nível do gerenciamento de ativos de dados;
- **Data Architecture Management:** é responsável por definir o plano de gerenciamento de dados para atender às necessidades de dados corporativos. Essa função inclui o desenvolvimento e a manutenção da arquitetura de dados corporativos, dentro do contexto de toda a arquitetura corporativa;
- **Data Development:** é responsável por projetar, implementar e manter soluções para atender às necessidades de dados da empresa. Inclui análise de demanda de dados, implementação, teste, manutenção e outras soluções;
- **Data Operation Management:** é responsável pelo planejamento, controle e suporte



**Figura 3.4. Mandala de Funções do DMBOK. Fonte:[Mosley et al., 2010]**

do ciclo de vida dos dados, desde a aquisição de dados ao arquivamento e eliminação de dados;

- **Data Security Management:** é responsável por planejar políticas e medidas de segurança para garantir a confidencialidade dos dados e os direitos de acesso hierárquico.
- **Reference and Master Data Management:** é responsável pelas atividades de planejamento, implementação e controle para garantir a consistência dos valores dos dados.
- **Data Warehouse e o Business Intelligence Management:** é responsável pelo planejamento, implementação e controle de processos para fornecer dados de suporte a decisões e suporte para trabalhadores do conhecimento envolvidos em relatórios, consultas e análises;
- **Document and Content Management:** é responsável pelo gerenciamento de arquivos eletrônicos e registros físicos (incluindo texto, gráficos, imagens, áudio e vídeo);
- **Metadata Management:** é responsável por integrar, controlar e fornecer metadados de alta qualidade;
- **Data Quality Management:** é responsável pelo planejamento, implementação e controle de atividades que aplicam técnicas de gerenciamento de qualidade para medir, avaliar, melhorar e garantir a adequação dos dados para uso.

Por sua vez, a família de padrões ISO 8000 fornece um conjunto de estruturas para melhorar a qualidade dos dados para tipos específicos de dados [ISO, 2011]. A ISO 8000 está sendo aplicada em vários setores industriais e países em todo o mundo <sup>4</sup>. A ISO 8000 abrange características de qualidade de dados industriais durante todo o ciclo de vida do produto, desde a concepção até o descarte [ISO, 2011]. Ele também descreve o vocabulário e os recursos, bem como define os requisitos para troca padrão e garantia de qualidade dos dados. A força da ISO 8000 reside no fato de que suas definições sobre dados de qualidade são baseadas em acordos internacionais.

Outra solução de gerenciamento que pode ser utilizada para a criação de métodos e processos de gerenciamento é o modelo DMM (do inglês, *Data Management Maturity*) Institute [2014]. O DMM é uma estrutura abrangente de práticas de gerenciamento de dados em seis categorias principais que ajuda as organizações a avaliar suas capacidades, identificar pontos fortes e lacunas e alavancar seus ativos de dados para melhorar o desempenho dos negócios. O DMM se concentra em 5 áreas de processo: Estratégia de Gerenciamento de Dados, Governança de Dados, Qualidade de Dados, Operações de Dados e Arquitetura de Dados.

Usando como base soluções desenvolvidas para outros tipos de ecossistemas, Baars and Jansen [2012] e Albert [2014] propuseram *frameworks* voltados para habilitação de governança em Ecossistemas de Software. Baars and Jansen [2012] descrevem um *framework* para a análise da governança de Ecossistemas de Software. O *framework* permite que organizações possam analisar e melhorar sua governança de Ecossistemas de Software de maneira estruturada, levando a um melhor desempenho e saúde do ecossistema. Albert [2014] propõem a abordagem SECOGov (Software Ecosystems Governance) que permite gerir e analisar informações sob a visão de Ecossistemas de Software. Apesar destas soluções não terem sido concebidas especificamente para gestão de Ecossistemas de Dados, parte de suas práticas podem ser adaptadas para o contexto de Ecossistemas de Dados.

### 3.9. Criação de Valor e Modelos de Negócio em Ecossistemas de Dados

Em Ecossistemas de Dados, os atores são obrigados a empregar um conjunto de capacidades e recursos para gerar valor. De acordo com o Magalhaes et al. [2014], muitas vezes o ônus é reservado aos consumidores que devem extrair valor dos recursos disponíveis. Isso cria um problema, já que o consumidor médio usualmente não possui as habilidades necessárias [Zuiderwijk et al., 2012]. Devido a essas barreiras, o valor não deve ser criado apenas por um único ator, mas sim por uma cadeia de valor (*i.e.*, em uma rede de atores). Uma cadeia de valor é um conjunto de atividades independentes de valor agregado que é usada para explorar um conjunto de recursos. Além disso, uma cadeia de valor consiste em diferentes atores conduzindo uma ou mais atividades (*e.g.*, provisão de dados, curadoria de dados, análise de dados), e cada atividade pode consistir em um número de ações ou técnicas de criação de valor (*e.g.*, coleta, visualização, criação de serviço).

Em Ecossistemas de Dados, a cadeia de valor mínima consiste em produtores de dados, intermediários de dados e consumidores de dados [Heimstädt et al., 2014]. Como as atividades de valor agregado oferecem complexidade diferente, é possível que cada

<sup>4</sup><https://www.dataqualitypro.com/iso-8000-data-quality-certification-options/>

ação possa consistir em uma ou mais cadeias de valor [Attard et al., 2016].

A introdução de incentivos e recompensas também pode estimular o fluxo de recursos e geração de valor em um Ecossistema de Dados [Lindman et al., 2016; Moiso and Minerva, 2012]. De fato, a produção, provisão e exploração de recursos do Ecossistema de Dados precisam de investimentos [Lindman et al., 2016; Moiso and Minerva, 2012]. Moiso and Minerva [2012] afirmam que, sem recursos financeiros, torna-se muito difícil sustentar iniciativas de Ecossistema de Dados. No entanto, há pouco incentivo para investir em recursos e capacidades. A falta de conhecimento sobre os benefícios do compartilhamento de dados e a falta de novos modelos de operação são os principais impedimentos que explicam porque os atores, principalmente as empresas privadas, não estão motivados a se engajar em Ecossistemas de Dados [Immonen et al., 2014].

Portanto, é importante desenvolver modelos de negócios sustentáveis que proporcionem um incentivo para manter os dados atualizados e acessíveis e, além disso, criar aplicativos e ferramentas comerciais sustentáveis [Immonen et al., 2014]. Modelos de negócios apoiam a proposição de valor para atores em um ecossistema. Como apresentado na Figura 3.5, um modelo de negócio envolve diversos aspectos, tais como estrutura de custo, capacidade, modelo de lucro, dentre outros aspectos.

Diversos modelos de negócios aplicáveis a dados foram descritos na literatura [Teece, 2010; Zott and Amit, 2010], tais como cobrança por suporte, modelo de assinatura e oferta de serviços/consultoria. Serviços e aplicativos podem ser cobrados com base em funcionalidades, custo ou por meio do modelo *pay-per-use* [Immonen et al., 2014]. Os recursos de dados também podem ser precificados usando o modelo de assinatura, no qual o consumidor paga um preço fixo por um determinado período de tempo. Outro modelo adequado é o fluxo de receitas múltiplas do Flickr [Teece, 2010], pois envolve a coleta de taxas de assinatura, cobrança de publicidade contextual para os anunciantes e o recebimento de patrocínio e taxas de participação nos lucros das parcerias.

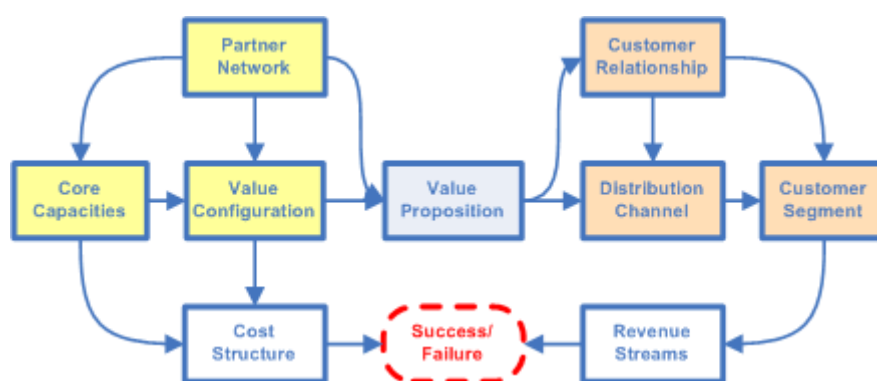


Figura 3.5. Modelo de Negócio. Fonte: <https://bit.ly/2U89mLM>

Outro exemplo é o modelo de negócios em nuvem, que é adequado no caso de grandes conjuntos de dados pelo fato do armazenamento, processamento e análise usualmente exigir uma grande quantidade de recursos. Além disso, como Ecossistemas de Dados são dinâmicos, os modelos de negócios devem tolerar a imprevisibilidade. Isso se deve ao fato de que os Ecossistemas de Dados ainda são incertos por natureza e não



é possível prever ou garantir que alguns recursos estarão disponíveis por tempo longo o suficiente [Zuiderwijk et al., 2015].

### 3.10. Boas Práticas para Publicação de Dados na Web

O sucesso de Ecossistemas de Dados, entre outros fatores, dependem diretamente da qualidade dos dados que estão providos e consumidos pelos atores. Além disso, segundo Lóscio et al. [2015], a heterogeneidade dos dados e a falta de padrões para descrição e acesso aos conjuntos de dados tornam o processo de publicação, compartilhamento e consumo uma tarefa complexa. De maneira geral, torna-se necessário publicar dados de forma que possam ser facilmente compreendidos e utilizados por consumidores, além da disponibilização dos dados em formatos que possam ser facilmente processados por aplicações.

Desse modo, em busca de alternativas que possibilitem um entendimento comum entre os atores desse contexto, o W3C criou um grupo de trabalho denominado *Data on the Web Best Practices*. Esse grupo teve como objetivo propor uma recomendação que servisse como um guia para a publicação e o consumo dos dados na Web.

**Tabela 3.1. Desafios da publicação de dados na Web. Fonte: Derilinx et al. [2015]**

Desafio	Descrição
Metadados	Permitir que os seres humanos entendam os metadados, interpretando a natureza e a estrutura dos dados, e que as máquinas também possam processá-los
Licença	Permitir que os seres humanos compreendam as informações da licença e que as máquinas possam detectar automaticamente
Proveniência	Permitir que os seres humanos conheçam a origem ou o histórico do conjunto de dados e que as máquinas possam processar automaticamente tais informações
Qualidade	Documentar a qualidade dos dados, para facilitar o processo de seleção dos conjuntos de dados e chances de reutilização
Versionamento	Permitir que versões dos dados sejam geradas e seja possível o acesso a cada versão
Identificação	Fornecer identificadores únicos para os conjuntos de dados e distribuições
Formato	Escolher formatos que permitam o uso e o reuso
Vocabulários	A fim de melhorar a interoperabilidade e manter terminologia comum entre os produtores e consumidores
Acesso	Permitir o fácil acesso aos dados usando a infraestrutura da Web tanto para seres humanos quanto para máquinas
Preservação	A fim de indicar corretamente se os dados foram removidos ou arquivados
Feedback	Receber feedback dos consumidores e assegurar que os dados atendam as suas necessidades
Enriquecimento	Enriquecer, melhorar ou refinar os dados brutos agregando valor
Republicação	Permitir que os dados utilizados possam ser republicados

Para alcançar esse objetivo, o grupo de trabalho do W3C, selecionou um conjunto de casos de uso<sup>5</sup> que representam cenários de como os dados são publicados e consumidos na Web. Com esses casos de uso, foi possível identificar os principais desafios enfrentados por produtores e consumidores de dados (ver Quadro 3.1), assim como um conjunto de requisitos necessários para a publicação. A partir dos desafios e requisitos encontrados nesses casos de usos, foi desenvolvido o documento de Boas Práticas para Publicação de Dados na Web (*Data on the Web Best Practices - DWBP*).

De acordo com Lóscio et al. [2016], as Boas Práticas para Dados na Web foram

<sup>5</sup><https://www.w3.org/TR/dwbp-ucr>

desenvolvidas para oferecer orientação técnica para a publicação de dados na Web, contribuindo para melhorar a relação entre publicadores e consumidores de dados. Além disso, elas são independentes de domínio e aplicação, ou seja, é aplicável a todos os domínios, podendo ainda ser estendidas ou complementadas com outros documentos ou normas mais especializadas. Para cada desafio apresentado no Quadro 3.1 foram propostas uma ou mais práticas. No total, são 35 boas práticas que discursam sobre diferentes aspectos relacionados à publicação e consumo de dados, como acesso aos dados, identificadores, metadados, formatos, dentre outros. O Quadro 3.2 apresenta as boas práticas estipuladas para cada desafio encontrado.

**Tabela 3.2. Boas Práticas para publicação de dados na Web. Fonte: Lóscio et al. [2017]**

Desafio	Boas Práticas
Metadados	BP1 - Fornecer metadados BP2 - Fornecer metadados descritivos BP3 - Fornecer metadados estruturais
Licença	BP4 - Fornecer informações de licenciamento de dados
Proveniência e Qualidade	BP5 - Fornecer informações sobre a proveniência dos dados BP6 - Fornecer informações sobre a qualidade dos dados
Versionamento	BP7 - Fornecer um indicador de versão BP8 - Fornecer histórico de versão
Identificação	BP9 - Utilizar URIs constantes como identificadores de conjuntos de dados BP10 - Utilizar URIs constantes como identificadores dentro dos conjuntos de dados BP11 - Designar URIs para versões e séries de conjuntos de dados
Formato	BP12 - Utilizar formatos de dados padronizados inteligíveis por máquinas BP13 - Utilizar representações de dados de localidade neutra BP14 - Fornecer dados em formatos múltiplos
Vocabulários	BP15 - Reutilizar vocabulários preferencialmente padronizados BP16 - Escolher o nível correto de formalização
Acesso	BP17 - Fornecer download em massa BP18 - Fornecer subconjuntos para conjuntos de dados extensos BP19 - Utilizar a negociação de conteúdo para disponibilizar dados em formatos múltiplos BP20 - Fornecer acesso em tempo real BP21 - Fornecer dados atualizados BP22 - Fornecer uma justificativa para dados não disponíveis BP23 - Disponibilizar dados por meio de uma API BP24 - Utilizar padrões da Web como base para as APIs BP25 - Fornecer a documentação completa para sua API BP26 - Evitar modificações que quebrem sua API
Preservação	BP27 - Preservar os identificadores BP28 - Avaliar a cobertura do conjunto de dados
Feedback	BP29 - Coletar feedback de consumidores de dados BP30 - Disponibilizar feedback
Enriquecimento	BP31 - Enriquecer dados por meio da geração de novos dados BP32 - Fornecer apresentações complementares
Republicação	BP33 - Fornecer feedback ao editor original BP34 - Seguir os termos de licenciamento BP35 - Citar a publicação original

Conforme apresentado por Lóscio et al. [2017], cada boa prática (BP) tem um resultado esperado com sua aplicação e possíveis formas de implantação da prática. Além disso, são descritas a motivação para o seu uso e quais testes podem ser realizados para verificar se a prática foi implementada de forma adequada. Ao final, ainda apresenta as evidências que comprovam a relevância da prática e os benefícios que serão alcançados com o seu uso.

### 3.11. Desafios em Ecossistemas de Dados na Web

O objetivo deste minicurso foi apresentar pesquisas relevantes e atuais sobre Ecossistemas de Dados, bem como fornecer uma visão geral do campo. Além de fornecer uma visão geral do campo Ecossistema de dados, também consideramos importante apresentar vários aspectos que ainda não são abordados na literatura.

Por exemplo, há poucos estudos apresentando e analisando possíveis modelos de negócios que permitam a criação e agregação de valor em Ecossistemas de Dados. Também há uma falta de linguagens de modelagem ou diagramas para representar os Ecossistemas de Dados. Atores de negócios e usuários não-técnicos podem enfrentar algumas dificuldades ao avaliar um modelo de negócios ou a organização estrutural de um Ecossistema de Dados. Uma nova linguagem de modelagem baseada em uma notação gráfica pode fornecer a capacidade de entender aspectos e processos importantes de um Ecossistema de Dados, além de oferecer aos atores a capacidade de comunicar esses aspectos e processos de maneira padrão. Por enquanto, as únicas opções existentes foram projetadas para outros tipos de ecossistemas, o que pode dificultar o uso pleno.

Modelos e soluções de avaliação para validar a saúde dos Ecossistemas de Dados são outra lacuna de pesquisa. Esses modelos devem fornecer os meios para avaliar a funcionalidade e o status dos elementos em um Ecossistema de Dados. A saúde de um ecossistema depende em parte de uma variedade de fatores, incluindo os atores e como eles agem, relacionamentos, políticas e a infraestrutura disponível. De fato, parte da definição de eficácia e sucesso de um Ecossistema de Dados está na determinação e utilização de métricas para medir sua saúde.

O desenvolvimento de métodos para governança e controle mais eficazes de Ecossistemas de Dados é outra lacuna na pesquisa. O funcionamento de um Ecossistema de Dados depende da atividade e interação de um conjunto de diferentes atores. Esse comportamento difuso tem como consequência a diminuição do controle e o aumento resultante dos desafios associados ao planejamento e manutenção. Portanto, os métodos de gerência de Ecossistema de Dados podem fornecer uma estrutura comum na forma de regras, procedimentos, protocolos e processos bem definidos para desenvolver, coordenar e evoluir os Ecossistemas de Dados.

Existem ainda vários desafios técnicos/não técnicos relacionados ao uso de dados em um Ecossistema de Dados, incluindo a complexidade das atividades necessárias para identificar, entender e usar dados, falta de recursos e conhecimento técnico entre os atores Janssen et al. [2012]; Zuiderwijk and Janssen [2014]. Desafios adicionais incluem proveniência de dados, qualidade de dados (*e.g.*, validade, integridade e pontualidade), fornecimento de metadados e interoperabilidade, bem como preocupações com privacidade e confidencialidade Janssen et al. [2012]; Zuiderwijk and Janssen [2014]. De um modo geral, a proposta de soluções para enfrentar alguns desses desafios pode aliviar o ônus para os atores, principalmente os que consomem dados, e, conseqüentemente, promover sua participação nos Ecossistemas de Dados.

### 3.12. Conclusão

Este minicurso teve como principal objetivo apresentar e contribuir para disseminação do conceito de Ecossistemas de Dados. De forma a atender este objetivo, apresentamos tanto pesquisas relevantes sobre Ecossistemas de Dados, bem como fornecemos uma visão geral do campo Oliveira et al. [2019]. A pesquisa em Ecossistemas de Dados, de uma forma geral, ainda está em estágio inicial, tendo recebido mais atenção em blogs, relatórios, poucas dezenas de artigos. Além disso, a literatura em Ecossistemas de Dados é focada em certos tipos de dados (por exemplo, dados governamentais ou dados científicos), ou não fornece uma descrição clara sobre como um Ecossistema de Dados deve ser.

Como já observado, o campo de Ecossistemas de Dados é inspirado no trabalho realizado em outros ecossistemas, como os Ecossistemas de Negócios e Software. De fato, é difícil definir os limites entre esses diferentes tipos de ecossistemas. Apesar da relação, a literatura em Ecossistema de Dados não parece examinar o trabalho realizado em ecossistemas de outras áreas. Por exemplo, em [Oliveira et al., 2019], foram identificados apenas 2 trabalhos analisando conceitos dos Ecossistemas de Software. Nesse sentido, a análise de outros campos permitiria o uso de suas teorias e soluções que poderiam, por sua vez, alavancar as pesquisas no Ecossistema de Dados.

O correto estudo de Ecossistemas de Dados abrange uma ampla variedade de disciplinas, incluindo projeto, técnica, orquestração, gerenciamento e avaliação. Neste sentido, identificamos vários aspectos que não são abordados na literatura, entre eles destacamos: esquemas para classificação de papéis dos atores, métodos de governança, soluções de engenharia e gerenciamento de Ecossistemas de Dados, plataformas e outras ferramentas de suporte, dentre outras lacunas. É preciso destacar que não existem muitos trabalhos acadêmicos relacionados aos Ecossistemas de Dados. Na maioria dos casos, eles estão focados em algum aspecto específico, como modelos de negócios ou uma solução que reflete apenas um pequeno fragmento de toda a área de pesquisa.

### Referências

- B Albert. *SECOGov: Um Modelo de Governança de Ecossistemas de Software para Apoiar Atividades de Arquitetura de TI*. PhD thesis, Dissertação. COPPE/UFRJ, Rio de Janeiro, Brasil, 2014.
- Judie Attard, Fabrizio Orlandi, and Sören Auer. Data value networks: Enabling a new data ecosystem. In *Web Intelligence (WI), 2016 IEEE/WIC/ACM International Conference on*, pages 453–456. IEEE, 2016.
- Alfred Baars and Slinger Jansen. A framework for software ecosystem governance. In *International conference of software business*, pages 168–180. Springer, 2012.
- Jan Bosch. From software product lines to software ecosystems. In *Proceedings of the 13th international software product line conference*, pages 111–119. Carnegie Mellon University, 2009.
- Jan Bosch and Petra Bosch-Sijtsema. From integration to composition: On the impact of software product lines, global development and ecosystems. *Journal of Systems and Software*, 83(1):67–76, 2010.

- Marco Brambilla, Jordi Cabot, and Manuel Wimmer. Model-driven software engineering in practice. *Synthesis Lectures on Software Engineering*, 1(1):1–182, 2012.
- Sjaak Brinkkemper, Ivo Van Soest, and Slinger Jansen. Modeling of product software businesses: Investigation into industry product and channel typologies. In *Information Systems Development*, pages 307–325. Springer, 2009.
- Henrik Bærbak Christensen, Klaus Marius Hansen, Morten Kyng, and Konstantinos Manikas. Analysis and design of software ecosystem architectures—towards the 4S telemedicine ecosystem. *Information and Software Technology*, 56(11):1476–1492, 2014.
- Sharon S Dawes, Lyudmila Vidiasova, and Olga Parkhimovich. Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, 33(1):15–27, 2016.
- Deirdre Lee Derilinx, Bernadette Farias Lóscio, and Phil Archer. Data on the web best practices use cases and requirements. <https://www.w3.org/TR/dwbp-ucr/>, feb 2015. Acessado em 31 de agosto de 2019.
- Li Ding, Timothy Lebo, John S Erickson, Dominic DiFranzo, Gregory Todd Williams, Xian Li, James Michaelis, Alvaro Graves, Jin Guang Zheng, Zhenning Shangguan, et al. Twc logd: A portal for linked open government data ecosystems. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(3):325–333, 2011.
- Carleton Fiorina. The digital ecosystem. <https://goo.gl/eSWuxN>, 2000. Acessado em 31 de agosto de 2019.
- Kiev Gama and Bernadette Farias Lóscio. Towards ecosystems based on open data as a service. In *International Conference on Information Systems*, pages 659–664, 2014.
- Teresa M Harrison, Theresa A Pardo, and Meghan Cook. Creating open government ecosystems: A research and development agenda. *Future Internet*, 4(4):900, 2012.
- Maximilian Heimstädt, Fredric Saunderson, and Tom Heath. Conceptualizing open data ecosystems: A timeline analysis of open data development in the UK. In *CeDEMI4: Conference for E-Democracy an Open Government*, page 245. MV-Verlag, 2014.
- Marco Iansiti and Roy Levien. *The keystone advantage: what the new dynamics of business ecosystems mean for strategy, innovation, and sustainability*. Harvard Business Press, 2004.
- Anne Immonen, Marko Palviainen, and Eila Ovaska. Requirements of an open data based business ecosystem. *IEEE Access*, 2:88–103, 2014.
- CMMI Institute. Data management maturity, 2014. URL <https://cmmiinstitute.com/data-management-maturity>. Acessado em 31 de agosto de 2019.
- ISO. Iso 8000 – the international standard for data quality, 2011. URL <https://www.iso.org/standard/50798.html>. Acessado em 31 de agosto de 2019.

- Slinger Jansen, Sjaak Brinkkemper, and Anthony Finkelstein. Business network management as a survival strategy: A tale of two software ecosystems. In *Proceedings of the 1st International Workshop on Software Ecosystems*, page 34, 2009.
- Marijn Janssen, Yannis Charalabidis, and Anneke Zuiderwijk. Benefits, adoption barriers and myths of open data and open government. *Information systems management*, 29(4):258–268, 2012.
- Victoria Köster and Gustavo Suárez. Open data for development: Experience of uruguay. In *Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance*, pages 207–210. ACM, 2016.
- D. Koznov, O. Andreeva, U. Nikula, A. Maglyas, D. Muromtsev, and I. Radchenko. A survey of open government data in Russian Federation. In *IC3K 2016 - Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, volume 3, pages 173–180, 2016.
- Mona Lebid. 12 examples of big data analytics in healthcare that can save people. <https://www.datapine.com/blog/big-data-examples-in-healthcare/>, 2018. Acessado em 31 de agosto de 2019.
- Deirdre Lee. Building an open data ecosystem: an Irish experience. In *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, pages 351–360. ACM, 2014.
- Juho Lindman, Tomi Kinnari, and Matti Rossi. Business roles in the emerging open-data ecosystem. *IEEE Software*, 33(5):54–59, 2016.
- Bernadette Lóscio, Caroline Guimarães, and Newton Calegari. Data on the web best practices: Challenges and benefits. *Open Data Reserach Symposium (ODRS 2016)*, 2016.
- Bernadette Farias Lóscio, Marcelo Iury S Oliveira, and Ig Ibert Bittencourt. Publicação e Consumo de Dados na Web: Conceitos e Desafios. *Tópicos em Gerenciamento de Dados e Informações (Mini Cursos - SBBB 2015)*, pages 39–69, 2015. URL <http://dex1.lncc.br/sbbd2015/anais/ShortCourses.pdf>.
- Bernadette Farias Lóscio, C Burle, and N Calegari. Data on the Web Best Practices. <https://www.w3.org/TR/dwbp/>, Janeiro 2017. Acessado em 31 de agosto de 2019.
- Gustavo Magalhaes, Catarina Roseira, and Laura Manley. Business models for open government data. In *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, pages 365–370. ACM, 2014.
- Konstantinos Manikas and Klaus Marius Hansen. Reviewing the health of software ecosystems—a conceptual framework proposal. In *Proceedings of the 5th International Workshop on Software Ecosystems (IWSECO)*, pages 33–44, 2013a.
- Konstantinos Manikas and Klaus Marius Hansen. Software ecosystems—a systematic literature review. *Journal of Systems and Software*, 86(5):1294–1306, 2013b.

- Eunice Mercado Lara and J Ramon Gil-Garcia. Open government and data intermediaries: the case of AidData. In *Proceedings of the 15th Annual International Conference on Digital Government Research*, pages 335–336. ACM, 2014.
- Corrado Moiso and Roberto Minerva. Towards a user-centric personal data ecosystem the role of the bank of individuals' data. In *Intelligence in Next Generation Networks (ICIN), 2012 16th International Conference on*, pages 202–209. IEEE, 2012.
- James F. Moore. *Creating Value in the Network Economy*, chapter Predators and Prey: A New Ecology of Competition, pages 121–141. Harvard Business School Press, Boston, MA, USA, 1999. ISBN 0-87584-911-3.
- Mark Mosley, Michael H Brackett, Susan Earley, and Deborah Henderson. *DAMA guide to the data management body of knowledge*. Technics Publications, 2010.
- Travis B Murdoch and Allan S Detsky. The inevitable application of big data to health care. *Jama*, 309(13):1351–1352, 2013.
- Francesco Nachira. Towards a network of digital business ecosystems. [www.digital-ecosystems.org/doc/discussionpaper.pdf](http://www.digital-ecosystems.org/doc/discussionpaper.pdf), 2002. Acessado em 31 de agosto de 2019.
- Francesco Nachira, Paolo Dini, and Andrea Nicolai. *A network of digital business ecosystems for Europe: roots, processes and perspectives*. European Commission, Bruxelles, 1 edition, 2007. ISBN 92-79-01817-5. Acessível em <http://www.digital-ecosystems.org/book/>.
- Marcelo Iury S. Oliveira and Bernadette Farias Lóscio. What is a data ecosystem? In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, dg.o '18, pages 74:1–74:9, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-6526-0.
- Marcelo Iury S. Oliveira, Lairson Emanuel R. A. Oliveira, Marlos G. Ribeiro Batista, and Bernadette Farias Lóscio. Towards a meta-model for data ecosystems. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, dg.o '18, pages 72:1–72:10, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-6526-0.
- Marcelo Iury S Oliveira, Glória de Fátima Barros Lima, and Bernadette Farias Lóscio. Investigations into data ecosystems: a systematic mapping study. *Knowledge and Information Systems*, pages 1–42, 2019.
- OMG. Meta object facility (mof) core specification v2.5.1. <http://www.omg.org/spec/MOF/2.5.1/PDF>, 2016. Acessado em 31 de agosto de 2019.
- Rufus Pollock. Building the (open) data ecosystem. <https://bit.ly/2DddJ1k>, 2011. Acessado em 31 de agosto de 2019.

- Dong-Hee Shin and Min Jae Choi. Ecological views of big data: Perspectives and issues. *Telematics and Informatics*, 32(2):311–320, 2015.
- Göran Smith, Hosea Ayaba Ofe, and Johan Sandberg. Digital service innovation from open data: exploring the value proposition of an open data marketplace. In *System Sciences (HICSS), 2016 49th Hawaii International Conference on*, pages 1277–1286. IEEE, 2016.
- Peter Sondergaard. Ecosystems are the future of digital. <http://www.gartner.com/technology/topics/digital-ecosystems.jsp>, 2017. Acessado em 31 de agosto de 2019.
- David J Teece. Business models, business strategy and innovation. *Long Range Planning*, 43(2-3):172–194, 2010.
- Barbara Ubaldi. Open government data: Towards empirical analysis of open government data initiatives. Technical report, Organisation for Economic Cooperation and Development (OECD), 2013. Acessado em 31 de agosto de 2019.
- Wikipedia. Ecosystem. <https://en.wikipedia.org/wiki/Ecosystem>, 2001. Acessado em 31 de agosto de 2019.
- Eric Yu and Stephanie Deng. Understanding software ecosystems: A strategic modeling approach. In *IWSECO-2011 Software Ecosystems 2011. Proceedings of the Third International Workshop on Software Ecosystems. Brussels, Belgium*, pages 65–76, 2011.
- Christoph Zott and Raphael Amit. Business model design: an activity system perspective. *Long Range Planning*, 43(2-3):216–226, 2010.
- Jose Jacobo Zubcoff, Llorenç Vaquer, Jose-Norberto Mazón, Francisco Maciá, Irene Garrigós, Andrés Fuster, and Jose Vicente Carcel. The university as an open data ecosystem. *International Journal of Design & Nature and Ecodynamics*, 11(3):250–257, 2016.
- Anneke Zuiderwijk and Marijn Janssen. Barriers and development directions for the publication and usage of open data: A socio-technical view. In *Open Government*, pages 115–135. Springer, 2014.
- Anneke Zuiderwijk, Marijn Janssen, Sunil Choenni, Ronald Meijer, R Sheikh Alibaks, and R Sheikh\_Alibaks. Socio-technical impediments of open data. *Electronic Journal of e-Government*, 10(2):156–172, 2012.
- Anneke Zuiderwijk, Marijn Janssen, and Chris Davis. Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, 19(1, 2):17–33, 2014.
- Anneke Zuiderwijk, Marijn Janssen, Kostas Poulis, and Geerten van de Kaa. Open data for competitive advantage: insights from open data use by companies. In *Proceedings of the 16th Annual International Conference on Digital Government Research*, pages 79–88. ACM, 2015.



Anneke Zuiderwijk, Marijn Janssen, Geerten van de Kaa, and Kostas Poulis. The wicked problem of commercial value creation in open data ecosystems: Policy guidelines for governments. *Information Polity*, 21:223–236, 2016.