

# LINKED DATA LICENSING – DATENLIZENZIERUNG UNTER NETZÖKONOMISCHEN BEDINGUNGEN

Tassilo Pellegrini

Professor (FH), Fachhochschule St. Pölten, Department Medienwirtschaft  
Matthias Corvinus Straße 15, 3100 St. Pölten, Österreich  
[Tassilo.pellegrini@fhstp.ac.at](mailto:Tassilo.pellegrini@fhstp.ac.at); <http://www.fhstp.ac.at>

**Schlagworte:** *Linked Data, Datenlizenzierung, Licensing Policy, Datenbewirtschaftung, Netzökonomie, Semantic Web Standards, IRIS 2014*

**Abstract:** *Ausgehend vom Technologiekonzept „Linked Data“ bespricht der Beitrag die immaterialgüterrechtlichen Aspekte vernetzter Daten und diskutiert die Erfordernisse nach einer „Linked Data Licensing Policy“ im Spannungsfeld von Urheber- und Datenbankrecht. Im Zuge der Darstellungen wird auf den Status Quo der Lizenzierungspraxis und die zukünftige Rolle von Lizenzierungstechnologien zur Verarbeitung von maschinenlesbaren Lizenzinformationen eingegangen.*

## 1. Einleitung

Mit der zunehmenden Interoperabilität von IT-Systemen und -Plattformen, sowie der damit verbundenen Portabilität von Daten rücken neben Content auch technische Artefakte wie Daten, Metadaten, Wissensmodelle und korrespondierende Services ins Vermarktungsportfolio von Unternehmen und öffentlichen Organisationen. Prominente Entwicklungen wie etwa Open (Government) Data, die Renaissance des Data Journalism und der Trend in Richtung Service-Orientierung (Lusch & Vargo 2006) sind Indizien für einen zunehmenden Wertschöpfungsbeitrag der Datenbewirtschaftung. Dies zeigt sich einerseits in der wachsenden Bedeutung von Metadaten in der digitalen Wertschöpfung – speziell im Kontext der stark auf Datenanalytik, Automatismen und Co-Produktion basierenden service-orientierten Logik des eCommerce – (z.B. Haase 2004; Graube et al. 2011; Pellegrini 2012), andererseits in der fortwährenden Entwicklung neuer Standards und Methoden zur Bewirtschaftung von Daten unter netzökonomischen Bedingungen, wie etwa die Semantic Web Initiative des World Wide Web Konsortiums oder die Metadateninitiative schema.org belegen. Begriffe wie „Big Data“ als konzeptioneller Überbau und „Linked Data“ als technologisch-methodische Basis der zunehmenden Bewirtschaftung verteilter Daten helfen die neu entstehenden, teils disruptiven Kulturtechniken zu benennen und zu institutionalisieren. Darunter fällt nicht nur das Beherrschen der Technologie sondern auch die Frage nach der ökonomischen Verwertung der neu geschaffenen, werttragenden Artefakte in der digitalen Wertschöpfung.

Vor diesem Hintergrund behandelt der Beitrag folgende Fragestellungen:

1. Was sind die technischen Enabler und institutionellen (immaterialgüterrechtlichen) Rahmenbedingungen der Datenbewirtschaftung unter netzökonomischen Bedingungen?
2. Welche Auswirkungen hat dies auf die Lizenzierungspraxis von Unternehmen und öffentlichen Organisationen in Bezug auf die Definition von maschinenlesbaren Licensing Policies?

3. Was ist der aktuelle Status der der Linked Data Lizenzierung und welche Handlungsbedarfe erwachsen daraus?

Der Beitrag gliedert sich folgendermaßen: Kapitel 2 stellt das technologische Konzept „Linked Data“ vor und diskutiert dessen Wertschöpfungsbeitrag. Kapitel 3 erläutert ausgehend von unterschiedlichen Linked Data Assets die immaterialgüterrechtlichen Aspekte des Einsatzes von Linked Data Technologien. Kapitel 4 geht auf Linked Data Policies ein und erläutert die Rolle von Rechteausszeichnungssprachen für die Verwertung von Linked Data unter netzökonomischen Bedingungen. Kapitel 5 wirft einen Blick auf die Lizenzierungspraxis von Linked Data und identifiziert aktuelle Problemlagen. Kapitel 6 fasst die Ergebnisse zusammen und gibt einen Ausblick auf weitere Entwicklungen.

## **2. Linked Data – Technologische Grundlagen und Wertschöpfungsbeitrag**

Konventionelle Datenbereitstellungsstrategien in Form von (semi-)strukturierten Dokumenten (z.B. HTML, CSV-Dateien) oder proprietären APIs werden nur bedingt den Ansprüchen hoch vernetzter und dynamischer Daten-Ökosysteme gerecht. Mit jeder zusätzlichen Quelle steigen die Integrationsaufwände exponentiell, Veränderungen in der Datenbankstruktur gehen oftmals zu Lasten der Systemintegrität und Aktualisierungen der Datenbasis sind meist nur unter hohen Aufwänden in Echtzeit verfügbar. Hier setzt der Linked Data Ansatz an, der eine höchst-mögliche technische Interoperabilität anstrebt und so die kosteneffiziente und zeitkritische Integrierbarkeit, eindeutige Interpretierbarkeit und Wiederverwendbarkeit von dispersen Daten ermöglicht. Linked Data bedient sich sogenannter Semantic Web Standards<sup>1</sup> um existierende Datenbestände hoch strukturiert aufzubereiten und plattformunabhängig für die Integration und Syndizierung bereitzustellen. Hierbei werden Daten mittels des normierten Datenmodells RDF (Resource Description Framework; siehe W3C 2004) strukturiert und verfügbar gemacht. Die semantisch angereicherten Daten werden im konventionellen Sinne nicht relational sondern als Graph repräsentiert. Sowohl die Knoten als auch die Kanten des Graphen sind über URIs (Uniform Resource Identifiers) eindeutig identifizierbar und referenzierbar. Dieser semantische RDF-Graph kann mittels der normierten Abfragesprache SPARQL (W3C 2008) feingranular und in hoher semantischer Tiefe abgefragt werden. Dies erlaubt die leichte Formulierung expressiver Datenbankabfragen, die mit konventionellen Mitteln entweder gar nicht oder nur mit hohen technischen und personellen Aufwänden machbar wären.

Linked Data zielt insbesondere auf die Erhöhung der Datenportabilität zwischen unterschiedlichen IT-Systemen und -Anwendungen ab. Die gehobene semantische Interoperabilität erlaubt die kosteneffiziente Zusammenführung verteilt vorliegender Datensets, die Entwicklung von service-orientierten Produkten und ermöglicht eine Bewirtschaftung des digitalen Contents entlang der gesamten Wertschöpfungskette (Mitchell & Wilson 2012; Cranford 2009).<sup>2</sup>

Tim Berners-Lee, Direktor des World Wide Web Konsortiums, fasst die technologischen Prinzipien von Linked Data folgendermaßen zusammen (Berners-Lee 2009):

1. Nutze eindeutige Identifikatoren (Uniform Resource Identifiers – URIs) als Name für Dinge.
2. Nutze http-URIs um diese Dinge im World Wide Web auffindbar zu machen.
3. Nutze den RDF-Standard zum Annotieren der URIs mit sinnvoller Kontextinformation.

---

<sup>1</sup> Ein Gesamtüberblick der relevanten Standards findet sich unter <http://www.w3.org/standards/semanticweb/>, aufgerufen am Dezember 20, 2013

<sup>2</sup> Für eine differenzierte Diskussion des volkswirtschaftlichen und betriebswirtschaftlichen Wertschöpfungsbeitrages von Big Data im Allgemeinen und Linked Data im Speziellen siehe Pellegrini (2013). Eine Schematisierung der Linked Data Value Chain findet sich bei Latif et al. (2009).

#### 4. Verknüpfe URIs mit anderen URIs um weitere Informationen auffindbar zu machen.

Das erste Prinzip stellt eine Grundbedingung dar und besagt, dass Ressourcen über einen Uniform Resource Identifier (URI) entsprechend der IETF URI Konventionen (Berners-Lee 1998) ausgezeichnet werden müssen.

Das zweite Prinzip besagt, dass http-URIs als Bezeichnungen für Ressourcen verwendet werden sollen. Wie Berners-Lee anmerkt, wird oftmals übersehen, dass http-URIs im eigentlichen Sinne keine Adressen sondern „Namen“ darstellen, auf deren Protokoll eine mächtige und evolvierende Infrastruktur in Form von Schreib- und Leseautomatismen (sog. REST-Services) aufbaut.

Das dritte Prinzip besagt, dass die maschinelle Verarbeitung verfügbarer Datenquellen, die bereits die URI-Konventionen erfüllen, durch Anreicherung mit interoperablen Metadaten verbessert wird. Strukturierte Annotation auf Grundlage von Wissensmodelle bzw. Ontologien<sup>3</sup> setzt hier an. Die Repräsentation der hierbei verwendeten Vokabulare muss der RDF-Norm genügen.

Als viertes und letztes Prinzip sollen die verfügbaren URIs durch gegenseitige Verweise de-referenziert und dadurch vernetzt werden, so wie es im konventionellen „Web of Documents“ auf Basis des Hypertext-Prinzips erfolgt. Daten werden durch Interoperabilität zu Netzwerkütern und steigern ihren Wert mit dem Grad ihrer Konnektivität und Referenzierbarkeit (Shy 2001) – eine ökonomische Gesetzmäßigkeit, die für den Produktionsfaktor Metadaten bisher unterbelichtet ist. Dieser Aspekt stellt auch den wichtigsten Unterschied zur konventionellen Datenbewirtschaftung dar, wo aufgrund proprietärer Repräsentationsstandards und Schemata kaum Netzeffekte zu erzielen sind.

Im Kern der oben beschriebenen Entwicklungen steht die technische Herstellung von semantischer Interoperabilität zwischen Datenbanken, Repositorien und anderen werthaltigen Informationsquellen. Die Vorteile von Linked Data gegenüber konventionellen Integrationstechnologien lassen sich laut Auer (2011) folgendermaßen beschreiben:

*De-Referenzierbarkeit:* Die Verwendung von URIs erlaubt nicht nur Dinge am Web eindeutig zu identifizieren sondern auch diese inklusive der angereicherten Zusatzinformation abzurufen.

*Kohärenz:* Die Verwendung von RDF als universelles Datenmodell erlaubt die kohärente Vernetzung von Informationen aus unterschiedlichen Namensräumen und ermöglicht auf diese Weise die semantische Anreicherung von Information durch sogenannte typisierte Links.

*Integrierbarkeit:* Das normierte RDF-Datenmodell erlaubt – aus technischer Perspektive – die niedrigschwellige Integration von syntaktischen und semantischen Informationen aus den vernetzten, dispers vorliegenden Datenquellen. Mittels Schema-Mapping (z.B. von RDF Vokabularien) und Instance Matching können in Folge semantisch hoch expressive Informationsbestände aggregiert und abgefragt werden.

*Aktualität:* Die Datenquellen inklusive ihrer Netzstruktur können aufgrund des geteilten Datenmodells leicht aktualisiert und veröffentlicht werden, ohne daraus resultierender zusätzlicher Integrationsaufwände oder Performanceverluste, wie sie üblicherweise bei konventionellen Extraktionsmaßnahmen oder Datentransformationen entstehen. Dies garantiert unter anderem eine hohe Aktualität der Daten und darauf aufbauender Dienste.

Als konkrete Manifestation des Linked Data Paradigmas lässt sich die seit 2007 stetig wachsende „Linked Data Cloud“<sup>4</sup>, eine dezentrale und kollaborativ gewachsene Infrastruktur aus RDF-Daten,

---

<sup>3</sup> Saumure & Shiri definieren Ontologien folgendermaßen: „Ontologies are being considered valuable to classifying web information in that they aid in enhancing interoperability - bringing together resources from multiple sources.“ (Saumure & Shiri 2008, S. 657)

<sup>4</sup> Siehe <http://linkeddata.org/>, aufgerufen am Dezember 26, 2013

anführen. Diese Data Cloud umfasste mit Stand 2013 mehrere hundert Milliarden Fakten aus unterschiedlichsten Themenfeldern und mittlerweile tausenden Datenquellen.<sup>5</sup> Diese Daten sind vorwiegend offen lizenziert und werden bereits aktiv kommerziell genutzt. So veröffentlichen Unternehmen und öffentliche Organisationen ihre Datensets in der Linked Data Cloud und nutzen gleichzeitig deren Daten um hauseigene Datenbestände anzureichern und Rich Content Anwendungen darauf aufzusetzen. Insbesondere Unternehmen aus der Pharma-Industrie (z.B. Roche, Merck, Elly Lilly) und der Medienbranche (z.B. BBC, NY Times, Reuters, Reed Elsevier, Wolters Kluwer, Pearson Publishing, Springer Verlag, ACM, Agence France Press, Google, Facebook) haben sich als Early Adopter von Linked Data Technologien hervorgetan.<sup>6</sup>

### 3. Immaterialgüterschutz von Linked Data

#### 3.1. Asset-Typen in der Bewirtschaftung vernetzter Daten

Eine differenzierte Betrachtung der technischen Prozessierung von semantischen Metadaten offenbart ein komplexes Gefüge aus Assettypen, die als technische Artefakte einen wertschöpfenden Beitrag in der Content-Produktion leisten und durch entsprechende Schutzinstrumente auch Kapitalcharakter haben. Dies ergibt sich im Wesentlichen aus dem gewerblichen Rechtsschutz geistiger Schöpfungen durch Rechtsinstrumente wie das Urheberrecht, das Datenbankrecht, das Patentrecht u.a.m. Im Kontext von Linked Data treten neben die klassische Vermarktungseinheit „Dokument“ auch sogenannte Metadata-Assets ins Zentrum der Schutzwürdigkeit. Je nach Reichhaltigkeit und Expressivität der semantischen Aufbereitung lassen sich folgende Metadata-Assets unterscheiden (Tabelle 1):

Metadata Assets	Strukturelle und technische Artefakte für die Erzeugung von Linked Data
Datensatz	Strukturierte Sammlung und Aufbereitung von Rohdaten
URIs (Uniform Resource Identifiers)	Eindeutige Identifikatoren als Bezeichner und Adresse für Entitäten eines Datensatzes
Namespaces	Eindeutige Namensräume zur Dereferenzierung von URIs
Vokabulare	Eindeutige domänen- und funktionsspezifische Begrifflichkeiten zur Annotation für deskriptive, strukturelle oder administrative Zwecke
Schemata	Formales Modell zur Strukturierung von Daten durch Selektion, Kombination und Mapping von Vokabularen
Ontologien	Formale Modelle um Beziehungen zwischen und Eigenschaften von Metadaten abzubilden
Regeln	Logische Operationen zur automatischen Erschließung von Information aus Ontologien

**Tabelle 1: Metadata Assets in der technischen Bewirtschaftung von Online-Content**

Aus der Prozessierung semantischer Metadaten lassen sich in Folge weitere Assettypen ableiten, die als „2<sup>nd</sup> Order Information“ bezeichnet werden sollen. Hierbei lassen sich vier Kategorien unterscheiden: 1) *Referenzen* als Sammlung von Verweisen innerhalb und zwischen Dokumenten (z.B. als Indices); 2) *Inferenzen* als automatische Erschließung impliziter Information aus semantisch verknüpften Datensets (z.B. als Queries); 3) *Präferenzen* als gebrauchsspezifische Muster der Interaktion mit digitalen Artefakten (z.B. als anonymisierte Nutzerprofile); und 4)

<sup>5</sup> Einen Überblick über verfügbare Datenquellen bietet z.B. <http://datahub.io>, aufgerufen am Dezember 31, 2013

<sup>6</sup> Vertiefende Fallbesprechungen siehe z.B. Rayfield (2012) für die BBC oder Dodds & Davis (2009) für guardian.co.uk.

*Konfidenzen* als personenbezogene Profile aus Transaktionsspuren, Sozial-, Stimmungs- und Meinungsmustern. Tabelle 2 erläutert die unterschiedlichen Assettypen.

<b>2<sup>nd</sup> Order Information</b>	<b>Information, die aus der Prozessierung semantischer Metadaten entsteht</b>
Referenzen	Aggregation kontextrelevanter Ressourcen in Form von semantischen Indices. Diese können sowohl Dokumente als auch Instanzdaten enthalten.
Inferenzen	Queries zur logik-basierte Erschließung impliziter Information im semantischen Graphen.
Präferenzen	Gebrauchssensitive Empfehlung und Filterung von Ressourcen auf Basis konstitutiver, regulativer und generativer Regeln (Beer 2009, S. 994).
Konfidenzen	Beobachtung und Analyse nutzerbezogener, bewusster und unbewusster Transaktionsspuren, Interessen und Stimmungsmuster.

**Tabelle 2: 2nd Order Information in der Prozessierung semantischer Metadaten**

Wie Beer (2009) herausstreicht, weckt insbesondere der letzte Assettyp in Form von 2<sup>nd</sup> Order Information Begehrlichkeiten für die Bewirtschaftung von semantischen Metadaten, zumal sich diese Daten hervorragend für Zwecke der nutzerspezifischen Informationsaggregation nutzen lassen. Die sich daraus ergebenden datenschutzrechtlichen Sachverhalte sind bis dato nur unbefriedigend gelöst und institutionalisiert (Kulathuramaiyer & Maurer 2009; Nagenborg 2009).

### 3.2. Rechtsschutz von Linked Data

Die Lizenzierungsfrage von Linked Data ist nicht trivial, zumal unterschiedliche Bestandteile eines semantischen Ordnungssystems mit unterschiedlichen Rechtsinstrumenten geschützt werden können. Zur Anwendung kommen in der folgenden Übersicht das Urheberrecht, das Datenbankrecht und das Recht gegen unlauteren Wettbewerb.<sup>7</sup> In Anlehnung an Sonntag (2006) lassen sich folgende Schutzobjekte unterscheiden (Tabelle 3):

	<b>Urheberrecht</b>	<b>Datenbank-Recht</b>	<b>Wettbewerbsrecht</b>
Dokument	Ja	Ja	Ja
Datensatz	Nein	Ja	Ja
Identifikatoren	Nein	Ja	Nein
Namensräume	Ja	Nein	Ja
Vokabulare	Teilw.	Ja	Ja
Schemata	Teilw.	Ja	Ja
Ontologie	Teilw.	Ja	Ja
Regeln	Teilw.	Ja	Ja
Inferenzen	Ja	Ja	Ja
Referenzen	Ja	Ja	Ja
Präferenzen	Ja	Ja	Ja
Konfidenzen	Ja	Ja	Ja

**Tabelle 3: Rechtsschutz von Linked Data (in Anlehnung an Sonntag 2006)**

Die Tabelle offenbart ein dicht gewobenes Schutzregime. Während das Urheberrecht den kreativen Werkcharakter schützt, stellt das Datenbankrecht einen Leistungs- bzw. Investitionsschutz dar.

<sup>7</sup> Das Patentrecht wird an dieser Stelle ausgespart, da – zumindest nach europäischer Rechtsprechung – die genannten Assets nur indirekten Schutz in Kombination mit einer technischen Erfindung erlangen können. Ebenfalls nicht Bestandteil der Betrachtung ist das Markenrecht und das Domainrecht.

Fragestellungen zum missbräuchlichen Gebrauch geschützter Assets werden meist im Wettbewerbsrecht behandelt. Diese drei Rechtsbereiche spielen in Folge die wichtigste Rolle in der Spezifizierung von Licensing Policies für Linked Data und darauf aufbauender Verwertungsmodelle.

### 3.3. Rechtsschutz unter netzökonomischen Bedingungen

Während Metadaten bisher aufgrund ihrer oftmals proprietären Strukturen und Repräsentationsstandards nicht dazu geeignet waren, auf Basis netzwerkökonomischer Prinzipien bewirtschaftet zu werden, ändert sich dies mit Linked Data grundlegend. Asset Creation durch Linked Data bedeutet, die werthaltigen Artefakte differenziert mit Lizenzen zu versehen, welche die skalenökonomischen Effekte von Netzgütern stützen oder beschränken.<sup>8</sup> Hierbei gilt, dass die traditionell vorherrschende Philosophie „starker Eigentumsrechte“ mit dem viralen, selbstorganisierenden und dezentralen Charakter des World Wide Web nur eingeschränkt kompatibel ist. Deshalb kommen vermehrt Commons-basierte bzw. offene Lizenzmodelle – oft in Kombination mit geschlossenen Lizenzmodellen in Form eines Dual Licensings – zum Einsatz.<sup>9</sup>

Im Bereich des Urheberrechts hat sich mit Creative Commons<sup>10</sup> eine tragfähige Alternative für den Schutz von Dokumenten etabliert. So ermöglicht es die CC0-Lizenz, auf alle Urheber- und Urheberpersönlichkeitsrechte sowie all ihre verwandten Schutzrechte an dem betreffenden Werk zu verzichten und auf diesem Weg Nutzungsrechte an den Daten der Öffentlichkeit zu übertragen. Ergänzend steht ein Lizenzbaukasten zur Verfügung, der die feingranulare Definition von Nutzungsrechten mit unterschiedlichen Freiheitsgraden auch für kommerzielle Nutzung zulässt.

Im Bereich des Datenbankrechts arbeiten unterschiedliche Initiativen parallel zu Creative Commons an sogenannten Data Commons<sup>11</sup>, einem Set von Lizenzen, das für die Spezifika der Datenlizenzierung optimiert ist. Mit aktuellem Stand werden zusätzlich zur GNU Documentation License drei Lizenzmodelle angeboten: Die Lizenz ODBL (Open Data Commons Open Database License) bringt vergleichbar der CC0-Lizenz einen völligen Verzicht auf alle Nutzungsrechte mit sich. Die Lizenz Open Data Commons Attribution License verlangt nach einer Nennung des Urhebers. Die Lizenz PDDL (Open Data Commons Public Domain Dedication and License) erlaubt die offene Definition von Nutzungsrestriktionen.

---

<sup>8</sup> Je nach strategischem Ziel eines Linked Data Providers könnte dies eine kostengünstige (bis kostenlose) Bereitstellung von Metadaten-Assets auf Basis offener Lizenzen oder eine Hochpreisstrategie auf Basis geschlossener Lizenzmodelle bedeuten. Bildlich gesprochen könnte ein service-orientiertes Unternehmen seinen Produkt-Index im Netz veröffentlichen und anderen die Möglichkeit geben, auf dessen Basis Applikationen zu entwickeln, wie es etwa die Metadateninitiative <http://schema.org> (bestehend aus Google, Yahoo und Bing) bzw. der Anbieter <http://geonames.org> praktizieren. Der konkrete Nutzungskontext entscheidet über die zur Anwendung kommende Lizenz, die Version des Index, die Service Levels und das Bepreisungsmodell.

<sup>9</sup> Ein Blick auf die Lizenzierungspraxis der BBC zeigt, dass durch den kombinierten Einsatz offener und geschlossener Lizenzmodelle bestimmte Datenbestände der Öffentlichkeit für die Weiterverwendung zur Verfügung gestellt werden. So bediente sich die BBC (mit Stand Februar 2013) neben dem klassischen Urheberrecht folgender Lizenzmodelle: GNU Free Documentation Licence für Content, der aus der Wikipedia bezogen wird, Creative Commons Public Domain and Attribution-NonCommercial-ShareAlike für Content, der aus der MusicBrainz Datenbank bezogen wird, und Attribution-NonCommercial-ShareAlike 3.0 Unported für die Besprechungen der Musikalben der BBC. Zusätzlich wird die Verwendung der Datenschnittstellen über Geschäftsbedingungen geregelt, die eine uneingeschränkte, nichtkommerzielle Nutzung der BBC Music Beta-Daten erlauben. Siehe [http://backstage.bbc.co.uk/archives/2005/01/terms\\_of\\_use.html](http://backstage.bbc.co.uk/archives/2005/01/terms_of_use.html), aufgerufen am Februar 20, 2013

<sup>10</sup> Siehe <http://creativecommons.org/>, aufgerufen am Dezember 5, 2013

<sup>11</sup> Siehe <http://www.opendatacommons.org/>, aufgerufen am Dezember 5, 2013

## 4. Licensing Policies und Rights Expression Languages für Linked Data

Entsprechend den diversen Rechtsschutzaspekten sollte eine Linked Data Licensing Policy aus normativer Perspektive drei Komponenten umfassen: 1) eine maschinenlesbare Lizenz, die die urheberrechtlichen Aspekte abdeckt; 2) eine maschinenlesbare Lizenz, die die datenbankrechtlichen Aspekte abdeckt; und 3) eine Community Norm, welche die verwendeten Lizenzen und Nutzungsrechte für den Human User leicht verständlich aufbereitet und im Sinne des Gesetzes gegen unlauteren Wettbewerb transparente Nutzungsbedingungen und „Good Conduct“ definiert.

Zur maschinellen Auszeichnung von Licensing Policies wurden seit den 1990er Jahren sogenannte Rights Expression Languages (REs) entwickelt, die dem Bereich der Digital Rights Management Technologien zuzurechnen sind (Prenafeta 2010). REs unterstützen die Identifikation, Filterung, Syndizierung und Modifikation von Content, der sich aus mehreren Quellen unterschiedlicher Rechteinhaber speist, und sie bilden die Grundlage für eine differenzierte automatische Prozessierung und Verwertung von Content. REs sind damit eine zentrale technologische Komponente in service-orientierten, hoch automatisierten, vernetzten Verwertungsstrukturen.

### 4.1. Open Digital Rights Language (ODRL)

Seit dem Jahr 2011 entwickelt die ODRL Arbeitsgruppe der W3C Community and Business Group<sup>12</sup> ein hoch expressives RDF/XML Vokabular zur Auszeichnung von Policies für die automatisierte Interaktion mit Online Content. ODRL baut auf einem Entity-Attribute-Modell auf, das eine feingranulare, maschinenlesbare Definition von Nutzungsrechten für digitale Assets erlaubt. Die Version 2.0 enthält 50 Ausprägungen in Bezug auf Rechte und Pflichten, 27 Verbotstypen und 10 Operatoren. ODRL eignet sich aufgrund seiner hohen Expressivität ideal zur Definition von Community Normen entlang unterschiedlicher Assets, Nutzertypen und Szenarien.

Die hohe Expressivität und damit verbundene Implementierungskomplexität des ODRL-Vokabulars hemmte bisher die Adaption des Standards für kommerzielle Zwecke. Im Jahr 2013 begann das International Press and Telecommunications Council (IPTC) unter der Bezeichnung RightsML<sup>13</sup> an einer leichtgewichtigen Adaption von ODRL für Zwecke der Lizenzierung von News-Content zu arbeiten.<sup>14</sup>

### 4.2. Creative Commons Rights Expression Language (CCREL)

Komplementär zu ODRL hat sich die Creative Commons Rights Expression Language (CCREL)<sup>15</sup> für urheberrechtsrelevante Schutzaspekte etabliert. Sie ist das Ergebnis einer informellen W3C Arbeitsgruppe, die ihre RDF-Spezifikationen im Jahr 2008 veröffentlichte und seither von der Creative Commons Foundation als Standard für die maschinelle Auszeichnung von Creative Commons Lizenzen empfohlen wird.

CCREL bietet ein kondensiertes, hierarchisch strukturiertes Set an Attributen zur Definition von Nutzungsrechten mit Online-Content, das komplementär zu ODRL steht. Diese Attribute können nahtlos in das ODRL-Vokabular übernommen und mittels ODRL weiter ausdifferenziert und spezifiziert werden. Jedoch eine Kombination von ODRL mit CCREL ist nicht zwingend. Die

---

<sup>12</sup> Siehe <http://www.w3.org/community/odrl/>, aufgerufen am Jänner 2, 2014

<sup>13</sup> Siehe <http://dev.iptc.org/RightsML>, aufgerufen am Jänner 2, 2014

<sup>14</sup> Ein Überblick über existierende Use Cases aus der Nachrichtenbranche findet sich unter <http://dev.iptc.org/RightsML-Use-Cases>, aufgerufen am Jänner 2, 2014

<sup>15</sup> Siehe <http://www.w3.org/Submission/ccREL/>, aufgerufen am Jänner 2, 2014

semantische Expressivität von CCREL ist ausreichend für die simple Annotation von digitalen Assets mit CC Lizenzinformationen.

### 4.3. Open Data Commons

Da Open Data Commons bisher kein eigenes Vokabular zur Auszeichnung von Policies anbietet, ist die Einbindung von datenbankrechtlichen Aspekten in eine Licensing Policy zum aktuellen Stand nur durch eine Adaption des ODRL- bzw. CCREL- Vokabulars möglich, sofern der relevante Datensatz über eine dereferenzierbare URI verfügt. Sämtliche damit verbundene Interaktionsszenarios lassen sich jedoch einwandfrei durch ODRL bzw. CCREL auszeichnen.

## 5. Status Quo der Linked Data Lizenzierung – Diskrepanz zwischen Theorie und Praxis

Eine Untersuchung der verwendeten Lizenzen in der Linked Data Cloud (Pellegrini & Ermilov 2013) offenbart eine aus mehreren Perspektiven unbefriedigende Situation. Tabelle 4 veranschaulicht die Lizenzmodelle jener Linked Data Sets, die über das Datenportal <http://datahub.io> zur Verfügung gestellt werden.<sup>16</sup>

License	Number of Datasets
License Not Specified	251
Creative Commons Attribution	135
Creative Commons CCZero	72
Creative Commons Attribution Share-Alike	71
Creative Commons Non-Commercial (Any)	49
Other (Attribution)	38
UK Open Government Licence (OGL)	36
Open Data Commons Open Database License (ODbL)	28
Open Data Commons Public Domain Dedication and Licence (PDDL)	27
Other (Not Open)	26
Other (Open)	25
Other (Public Domain)	25
Open Data Commons Attribution License	14
GNU Free Documentation License	9
Other (Non-Commercial)	9

<sup>16</sup> Eine vergleichbare Erhebung und kritische Reflexion findet sich bei auch bei Jain et al. (2013).



ukcrown-withrights	6
W3C	1
apache	1
gpl-2.0	1
gpl-3.0	1

**Tabelle 4: Überblick Lizenzen auf <http://datahub.io> (Stand Juli 10, 2013)**

Der Status Quo lässt sich folgendermaßen zusammenfassen: Noch hat sich keine Konvention zur Deklaration von Policies, die das vollständige Rechtsspektrum von Linked Data abdecken, ausgebildet. Bisher werden hauptsächlich urheberrechtlich relevante Aspekte lizenziert. Die Verwendung von datenbankrelevanten Lizenzmodellen hat sich noch kaum etabliert. Allerdings ist zu berücksichtigen, dass mit der Creative Commons Version 4.0 auch datenbankrechtliche Aspekte stärker in der CC-Lizenzierung Berücksichtigung finden und dadurch eine bessere Rechtsabdeckung gewährleistet ist. Die Situation wird weiters durch regional abweichende Rechtsregime kompliziert. Während das Datenbankrecht ein EU-Spezifikum darstellt, werden in den USA datenbankrechtliche Sachverhalte durch den Copyright Act gedeckt. Hinzu kommt, dass im Gegensatz zu Europa Datensätze, die über keine Lizenz verfügen, automatisch der Public Domain zugerechnet werden, wohingegen dies in Europa explizit deklariert werden muss. Ein Blick auf verfügbare Community Normen zeigt, dass diese in Umfang, Formulierung und Zugänglichkeit stark voneinander abweichen. Auch hat sich die Verwendung von Rights Expression Languages kaum etabliert, wodurch nur sehr eingeschränkte Möglichkeiten existieren, Datensätze auf Basis ihrer maschinenlesbaren Lizenzinformation einer automatischen Prozessierung zuzuführen, etwa für Zwecke der Aggregation, Versionierung und Servicierung von Content.

## 6. Conclusio und Ausblick

Während Metadaten bisher aufgrund ihrer oftmals proprietären Strukturen und Repräsentationsstandards nicht dazu geeignet waren, auf Basis netzwerkökonomischer Prinzipien bewirtschaftet zu werden, ändert sich dies durch Linked Data radikal. Der Schlüssel zu gewerblichen Diversifikation ist eine Strategie der Rechtediversifikation, die im Sinne der Versionierung sowohl die Bedienung des Web-Ökosystems als auch des Corporate Marktes ermöglicht und gleichzeitig Rechtssicherheit schafft. Die notwendige Kulturtechnik im Sinne der maschinellen Bereitstellung von interoperablen Lizenzinformationen mittels Rights Expression Languages entlang des Urheber- und Datenbankrechts ist allerdings noch sehr schwach ausgeprägt, was zum Einen auf eine fehlende technische Infrastruktur in Form leicht bedienbarer, systemisch integrierten Tools zur Kompilation und Annotation von Lizenzen, zum Anderen auf fehlende ökonomische Incentives zur Veröffentlichung von Linked Data zurückzuführen ist. Vor dem Hintergrund der weiteren technologischen Ausdifferenzierung von Linked Data Technologien zur maschinellen Datenbewirtschaftung ist jedoch von der inkrementellen Herausbildung einer Kulturtechnik der Linked Data Lizenzierung auszugehen. Die technologischen und juristischen Voraussetzungen dafür sind geschaffen.

## 7. Literatur

Auer, Sören (2011). Creating Knowledge Out of Interlinked Data. In: Proceedings of WIMS'11, May 25-27, 2011, p. 1-8

- Beer, David* (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. In: *new media & society*, 11/6, 2009, p. 985 - 1002
- Berners Lee, Tim* (1998). Uniform Resource Identifiers (URI): Generic Syntax. In: IETF Network Working Group, Request for Comments: 2396, See also: <http://www.ietf.org/rfc/rfc2396.txt>, aufgerufen am Februar 20, 2013
- Berners-Lee, Tim* (2009). Linked Data Design Issues. In: <http://www.w3.org/DesignIssues/LinkedData.html>, aufgerufen am 25.05.2013
- Cranford, Steve* (2009). Spinning a Data Web. In: Price Waterhouse Coopers (Ed.). *Technology Forecast, Spring 2009*. <http://www.pwc.com/us/en/technology-forecast/spring2009/index.jhtml>, aufgerufen am September 20, 2013
- Dodds, Leigh; Davis, Ian* (2009). MP Data SPARQL Editor. In: <http://www.guardian.co.uk/open-platform/apps-mp-data-sparql-editor>, visited April 20, 2012
- Graube, Markus; Pfeffer, Johannes; Ziegler, Jens; Urbas, Leon* (2011). Linked Data as integrating technology for industrial data. In: 2011 Int. Conference on Network-Based Information Systems, 7-9 Sept. 2011, p. 162-167
- Haase, Kenneth* (2004). Context for Semantic Metadata. In: *Proceedings of MM'04*, October 10–16, 2004, New York, USA. ACM
- Jain, Prateek; Hitzler, Pascal; Janowicz, Krzysztof; Venkatramani, Chitra* (2013). There's No Money in Linked Data. In: <http://knoesis.wright.edu/faculty/pascal/pub/nomoneylod.pdf>, aufgerufen am Dezember 18, 2013
- Kulathuramaiyer, Narayanan; Maurer, Hermann* (2009). Implications of Emerging Data Mining. In: Blumauer, Andreas; Pellegrini, Tassilo (Hg.). *Social Semantic Web*. Berlin: Springer Verlag, S. 469 – 484
- Latif, Atif; Us Saeed, Anwar; Höfler, Patrick; Stocker, Alexander; Wagner, Claudia* (2009). The Linked Data Value Chain: A Lightweight Model for Business Engineers. In: *Proceedings of I-Semantics 2009, the 5th International Conference on Semantic Systems*. Graz: *Journal of Universal Computer Science*, p. 568–577
- Lusch, Robert F.; Vargo, Stephen L.* (2006). Service-dominant logic: reactions, reflections and refinements. In: *Marketing Theory*, September 2006; vol. 6, 3: pp. 281-288
- Mitchell, Ian; Wilson, Mark* (2012). Linked Data. Connecting and exploiting big data. Fujitsu White Paper, March 2012. <http://www.fujitsu.com/uk/Images/Linked-data-connecting-and-exploiting-big-data-%28v1.0%29.pdf>, aufgerufen am September 12, 2013
- Nagenborg, Michael* (2009). Privacy im Social Semantic Web. In: Blumauer, Andreas; Pellegrini, Tassilo (Hg.). *Social Semantic Web*. Berlin: Springer Verlag, S. 484 – 506
- Pellegrini, Tassilo* (2012). Semantic Metadata in the News Production Process. Achievements and Challenges. In: Lugmayr, Artur et al. (Eds). *Proceeding of the 16th International Academic MindTrek Conference 2012*. ACM SIGMM, p. 125-133
- Pellegrini, Tassilo* (2013). The Economics of Big Data: A Value Perspective on State of the Art and Future Trends. In: Akerkar, Rajendra (Ed.). *Big Data Computing*. New York: Chapman and Hall/CRC, p. 343-371
- Pellegrini, Tassilo; Ermilov, Ivan* (2013). Guide and Best Practices to Licensing Interlinked Data. Public Deliverable 7.4. EU-Project LOD 2. Grant Agreement No: 257943. Siehe auch <http://svn.aksw.org/lod2/WP7/D7.4/public.pdf>, aufgerufen am 03.01.2014
- Prenafeta, Javier* (2010). Protecting Copyright Through Semantic Technology. In: *Publishing Research Quarterly*, 26/4, pp 249-254
- Rayfield, Jem* (2012). Sports Refresh: Dynamic Semantic Publishing. In: BBC Internet Blog, [http://www.bbc.co.uk/blogs/bbcinternet/2012/04/sports\\_dynamic\\_semantic.html](http://www.bbc.co.uk/blogs/bbcinternet/2012/04/sports_dynamic_semantic.html), aufgerufen am Mai 5, 2012
- Saumure, Kristie; Shiri, Ali* (2008). Knowledge organization trends in library and information studies: a preliminary comparison of pre- and post-web eras. In: *Journal of Information Science*, 34/5, 2008, p. 651–666
- Shy, Oz* (2001). *The Economics of Network Industries*. Cambridge: Cambridge University Press
- Sonntag, Michael* (2006). Rechtsschutz für Ontologien. In: Schweighofer, Erich; Liebwald, Doris; Drachsler, Matthias; Geist, Anton (Eds.). *e-Staat und e-Wirtschaft aus rechtlicher Sicht*. Stuttgart: Richard Boorberg Verlag, p. 418-425
- W3C (2004). RDF – Resource Description Framework. In: <http://www.w3.org/RDF/>, aufgerufen am Dezember 10, 2013
- W3C (2008). SPARQL Query Language for RDF. in: <http://www.w3.org/TR/rdf-sparql-query/>, aufgerufen am Dezember 20, 2013