EIN WEG ZUM EINDEUTIGEN DIGITALEN IDENTIFIKATOR JURISTISCHER DOKTRINALER INHALTE

Blaise Dévaud / Franz Kummer

Dr. iur., MA in Wirtschaftsinformatik, Projektleiter bei der Weblaw AG
Weblaw AG, Laupenstrasse 1, 3008 Bern, CH
blaise.devaud@weblaw.ch; http://www.weblaw.ch

Jurist; Lehrbeauftragter Universität Bern; Gründer und Inhaber der Weblaw AG
Weblaw AG, Laupenstrasse 1, 3008 Bern, CH
franz.kummer@weblaw.ch; http://www.weblaw.ch

Schlagworte: ***DOI; ECLI; Identifikator; Zitierbarkeit; juristische Doktrin; bibliographische Metadaten; Indexierung***

Abstract: *Der Beitrag untersucht die Möglichkeiten der Schaffung einer eindeutigen digitalen Bezeichnung juristischer doktrinaler Inhalte im deutschsprachigen Raum. Zu diesem Ziel führen zwei Wege: einerseits die gemeinsame Auswahl und konsequente Nutzung eines bereits existierenden Identifikators, und andererseits die Einführung eines neues Identifikators geschaffen nach Mass für die juristischen doktrinalen Inhalte. Die Analyse zeigt, dass die Anforderungen an einen eindeutigen Identifikator juristischer Doktrin mit dem bestehenden DOI (Digital Object Identifier) gut abgedeckt werden könnten.*

1. Einführung

Die Vielfalt der digitalen juristischen Publikationen bringt das Bedürfnis nach eindeutiger Beschreibung und Kategorisierung mit sich. Die bisherigen Identifikationsmöglichkeiten reichen nicht mehr aus: ISBN bezieht sich nur auf Bücher; ISSN nur auf Zeitschriften, aber nicht auf die Beiträge selber. Auch die standardisierte juristische Zitierweise zeigt in der Praxis viele (kleine) Unterschiede, die eine maschinelle Erkennung und Bearbeitung erschweren. Im juristischen Bereich wurden in der EU eindeutige Identifikatoren für die Rechtsprechung (ECLI[1]) und Gesetzgebung (ELI[2]) konzipiert. Was fehlt ist eine eindeutige Beschreibung der dritten Komponente des juristischen Wissens, der Doktrin. Es existiert zwar ein Standard für digitale Publikationen (DOI – Digital Object Identifier[3]), dieser ist aber in der juristischen Praxis kaum verbreitet.

Der Beitrag untersucht die Möglichkeiten der Schaffung einer eindeutigen digitalen Bezeichnung juristischer doktrinaler Inhalte im deutschsprachigen Raum. Die Bezeichnung soll einerseits eine maschinell lesbare Zitierung der Inhalte ermöglichen. Dies ist eine Basis für eine einfachere Indexierung und Verlinkung. Andererseits soll die Bezeichnung auch einen breiteren Kontext über den Inhalt vermitteln können – nicht nur den Verfasser und das Medium, sondern auch weitere (juristische) Zusatzinformationen wie den Beitrag beschreibende Stichworte, zitierte Erlasse, Entscheide und Publikationen. So soll ersichtlich sein, welche juristischen Referenzen in welchen Texten zitiert sind, ohne auf den Volltext zugreifen zu müssen bzw. ohne einen Zugriff auf den Volltext gewähren zu müssen.

---

1 European Commission, European Case Law Identifier (ECLI), https://e-justice.europa.eu/content_european_case_law_identifier_ecli-175-en.do; (alle Websites zuletzt abgerufen am 3. Januar 2018).

2 European Legislation Identifier, European Legislation Identifier, http://www.eli.fr/de/actualites.html.

3 International DOI Foundation, DOI® Handbook, https://www.doi.org/hb.html.

Eine öffentliche, strukturierte Abrufbarkeit der Metadaten hat viele Vorteile: einerseits für Inhaltsanbieter bzw. Datenbankbetreiber, die nach einem möglichst vollständigen Angebot der Inhalte von allen Verlagen streben, und andererseits für Suchmaschinenbetreiber, die optimale Treffer anbieten möchten, ohne für die Indexierung auf die kommerziell geschützten Volltexten zugreifen zu müssen. Auf diese Weise würden der Zugang zu kommerziellen Inhalten und die Suche nach diesen Inhalten entkoppelt. Die Verlage erzielen eine breitere Streuung ihrer Publikationen und behalten trotzdem die Kontrolle über den kommerziellen Zugang, und die Leser haben ein vollständigeres Angebot beim Kauf und bei der Suche.

Zu diesem Ziel führen zwei Wege: einerseits die gemeinsame Auswahl und konsequente Nutzung eines bereits existierenden Identifikators und andererseits die Einführung eines neues Identifikators geschaffen nach Mass für die juristischen doktrinalen Inhalte. Der Beitrag untersucht die beiden Wege und macht sich anschliessend Gedanken zur möglichen Umsetzung.

2. Nutzung eines existierenden Identifikators

Zwei Arten der existierenden Identifikatoren kommen in Frage: erstens existierende Formate oder Standards aus dem juristischen Umfeld, und zweitens allgemeine Identifikatoren die sich für den gewünschten Zweck einsetzen lassen würden. Existierende Identifikatoren aus anderen Wissensfeldern (wie z.B. PubMedID) werden nicht weiter berücksichtigt.

Bei der ersten Gruppe erscheinen zwei Standards als mögliche Kandidaten. CHJusML ist ein neuer Schweizer Standard noch im Entwurfsstadium, der die Standards CHLexML[4] (gedacht für die normativen Texte) und CHDecML (gedacht für die Entscheide) vervollständigen soll. Akoma Ntoso[5] ist ein internationaler Standard für die Strukturierung juristischer Dokumente, vor allem für normative Texte und parlamentarische Dokumente. Die beiden Formate sind jedoch für die komplette Darstellung eines Dokumentes in einer strukturierten Form gedacht, und nicht für eine eindeutige Identifizierung verschiedener Inhalte in diversen technischen Formaten (Text, Bilder, Ton etc.). Aus diesem Grund erscheint es nicht sinnvoll, den Einsatz dieser beiden Standards als einen eindeutigen Identifikator aller modernen, multimedialen Inhalte weiterzuverfolgen.

Bei der zweiten Gruppe erscheint der DOI als einziger Kandidat. Dieser Standard ist bereits seit Jahren vorhanden, ist nicht Inhalts- oder Fachgebiet-spezifisch und ist in der wissenschaftlichen Praxis bekannt. Darum eignet er sich gut – trotz fehlender Popularität in der juristischen Publikationspraxis – als Ausgangspunkt für die weitere Überlegung.

Der DOI wird von Registrierungsstellen grundsätzlich für jegliche Inhalte vergeben. Er eignet sich somit nicht nur für Texte in verschiedenen Erscheinungsformen, sondern auch für neue multimedialen Inhalte wie z.B. Podcasts. Der DOI kann dabei auch die bisherigen Identifikatoren wie ISBN oder ISSN als Teil der Inhaltsbezeichnung integrieren, und zwar nicht nur für das Gesamtwerk an sich (z.B. ein Buch), sondern auch für Teilinhalte eines Werkes (z.B. Kombination von ISBN und laufender Nummerierung der Beiträge aus einem Sammelwerk). Darüber hinaus kann in den DOI auch eine Bezeichnung des Verlages integriert werden. Das DOI Konzept ist flexibel und erlaubt den einzelnen Registrierungsstellen viel Spielraum.

Der DOI dient nicht nur als eine eindeutige Bezeichnung eines Inhalts, er erlaubt auch eine kontextabhängige Interaktion. Der Basis-Fall ist die Weiterleitung an den Inhalt beim Anklicken des DOI. Eine andere Möglichkeit der Auslösung eines DOI besteht z.B. darin, die Metadaten zum Inhalt zurückzuschicken bzw. an den entsprechenden Eintrag in einer Sammlung der Metadaten weiterzuleiten. Weitere Möglichkeiten bleiben offen und sind von den einzelnen Registrierungsstellen abhängig. Dies bedeutet allenfalls, dass mit der Registrie-

4 Verein eCH, eCH-0095: CHLexML, https://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0095&documentVersion=1.0.

5 Akoma Ntoso, Akoma Ntoso, http://www.akomantoso.org/.

rung eines DOI die nötigen Informationen (wie z.B. URL, bibliographische Metadaten etc.) mitgeliefert und nachträglich gepflegt werden müssen.

Aus dieser kurzen Übersicht ergibt sich bereits, dass der DOI alle gewünschten Funktionalitäten anbieten kann: eindeutige Bezeichnung aller Inhalte der juristischen Verlage und Abrufbarkeit der Metadaten zu den Inhalten ohne Zugang auf den vollständigen Inhalt. Der DOI hat jedoch auch Nachteile.

Der erste Nachteil ist mit der zwingenden Abhängigkeit von den Registrierungsstellen verbunden. Einerseits kann das Kosten generieren: die Registrierungsstellen können sich als kommerzielle Service Betreiber verstehen und von den Verlagen Geld verlangen, bzw. eine entgeltlose Registrierung nur für akademische bzw. nur für open access Inhalte reservieren. Andererseits können die Registrierungsstellen auch diverse Einschränkungen vorsehen, sei es was den Typ der registrierbaren Inhalte betrifft (je nach Medium, Typ des Inhalts etc.), sei es was die Interaktion mit den bei sich registrierten DOIs anbelangt (Verfügbarkeit der registrierten Inhalte, zwingende Angaben bestimmter Metadaten, Möglichkeit und Form der Metadatenabrufe etc.).

Dieser erste Nachteil könnte so gelöst werden, dass sich die Verlage aus dem DACH Raum zusammenschliessen und gemeinsam eine DOI Registrierungsstelle einrichten. So könnten sie die Registrierung für die Mitglieder unentgeltlich machen und für Verlage ausserhalb der Federation kostenpflichtig. Die Verlage würden nur die Infrastruktur und Betreibung der Registrierungsschnittstelle mit einem fixen Betrag finanzieren, ohne Zusatzkosten für die einzelnen Registrationen. Je mehr Verlage mitmachen würden, desto kleiner wären diese Kosten pro Verlag. Weiter könnten die Verlage selber über alle Modalitäten der Registrierung bestimmen, insbesondere in Bezug auf die Inhalte. So könnten nicht nur die «klassischen» Texte (Doktrin), sondern auch neue Publikationsformen (wie z.B. Blogs, Podcasts etc.) mitregistriert und eindeutig zitierbar werden.

Der zweite Nachteil bestünde darin, dass die Kombination von einem eindeutigen, weltweit bekannten Identifikator und der Möglichkeit, die Inhalte mit einem einfachen Klick abzurufen, zu automatischem en bloc Hochladen von Inhalten führen könnte. Die Inhalte würden diesfalls ausserhalb der Kontrolle des herausgebenden Verlages abgespeichert und ggf. zur Verfügung gestellt.

Dieser zweite Nachteil ist schwieriger zu umgehen. Eine mögliche Lösung wäre, die interaktive Auslösung des DOI auf den Abruf von Metadaten zu beschränken. Diese Einschränkung ist natürlich erst dann möglich, wenn die Verlage selber eine Registrierungsstelle betreiben. Zudem ist diese Einschränkung nicht universell: jeder Verlag und jede Drittdatenbank können intern den DOI für den direkten Abruf der eigenen Inhalte einsetzen, was wiederum die gleiche Gefahr in sich birgt. Das bedeutet, dass dieses Risiko nicht vollständig unterbunden werden kann – dies gilt aber für den Einsatz jeglicher eindeutiger Identifikatoren.

3. Schaffung eines neuen Identifikators

Die Schaffung eines neuen Identifikators für juristische doktrinale Texte wirft eine ganze Reihe von Grundsatzfragen auf.

Erstens stellt sich die Frage, ob der Identifikator lediglich ein «Handle» sein soll, d.h. eine Nummer, die an sich keine bzw. nur wenige Informationen trägt und lediglich als Weiterleitung an eine Datenbank dient (analog DOI), oder ob der Identifikator in sich die vollständige Bezeichnung des Inhaltes fragen soll (analog ECLI). Die erste Option hat den Vorteil, dass der Identifikator kürzer ist, was vor allem in den Fussnoten von Bedeutung ist; der Nachteil ist, dass die Informationen über den Inhalt erst nach dem Auslösen des Identifikators und der Abfrage der Metadatendatenbank verfügbar sind. Bei der zweiten Option ist der Identifikator selbstbeschreibend und benötigt keine zusätzliche Datenbankabfrage; allerdings macht die Menge der kodierten Informationen den Identifikator viel länger, was sich wiederum bei der Zitierung in den Fussnoten negativ auswirkt.

Sollte der neue Identifikator nach dem ECLI-Vorbild geschaffen werden, müssten auch seine Bestandteile definiert werden. Dieser Entscheid stünde in einem Spannungsfeld zwischen Knappheit und Vollständigkeit der Beschreibung. Soll die Eindeutigkeit des Identifikators aus einer Kombination verschiedener informativer Fel-

der bestehen, dann müsste sich die Auswahl der Felder am Muster der gängigen Zitierung juristischer Inhalte orientieren. Dies bedeutet eine Kombination von mindestens 4 Elementen: Autor, Titel, Publikationsdatum, Verlag. Diese Kombination an sich ist aber nicht immer ausreichend eindeutig, da z.B. bei Zeitschriftenbeiträgen die Zeitschrift, Heft- und Seitenangaben fehlen würden. Weiter müssten alle vier Felder kodiert werden, da sonst der Identifikator gleich lang wie eine normale Zitierung wäre, was seinem Zweck widersprechen würde. Die Kodierung des Datums und der relativ kleinen Liste der Verlage scheint weniger problematisch, aber die Kodierung der Autoren und der Titel in einer kurzen Form könnte nur auf die Kosten der menschlichen Lesbarkeit gemacht werden, was wiederum zu einem Widerspruch führen würde. Diese Überlegungen sprechen zu Gunsten eines DOI-ähnlichen Identifikators.

Der neue Identifikator soll auch eine maschinelle Abfrage der Metadaten ermöglichen, was einerseits eine technische Infrastruktur und andererseits ein Metadatenformat benötigt.

Die Infrastruktur könnte entweder zentralisiert oder dezentralisiert eingerichtet werden: eine zentrale Datenbank braucht eine stärkere Zusammenarbeit und Koordination der Verlage, während eine dezentrale Lösung den Verlagen mehr Spielraum lassen würde. So könnte jeder Verlag eine Datenbank der eigenen Inhalte nach eigenem Ermessen führen, so lange eine standardisierte Schnittstelle eine einheitliche Abfrage in allen separaten Datenbanken ermöglichen würde. Auch bei der dezentralen Lösung wäre jedoch zu überlegen ob eine zentrale Abfrage-Stelle nötig wäre, insbesondere um die Abfragen entgegenzunehmen und ggf. an die zuständigen Verlagsdatenbanken weiterzuleiten. Zu klären wäre ebenfalls, welche Lösung günstiger ist und wie die Kosten verteilt werden sollten (Parität, Verhältnis zur Verlagsgrösse, zur Anzahl publizierter Inhalte etc.).

Was das Metadatenformat anbelangt, scheint es am einfachsten, ein bestehendes bibliografisches Metadatenformat zu nutzen und eine Mindestzahl an obligatorischen Feldern zu definieren, die bei jedem Eintrag ausgefüllt werden sollten. Weitere Felder, wie z.B. Abstract, Stichwörter oder zitierte Referenzen könnten optional bleiben, wobei allerdings gerade diese Felder den Mehrwert des neuen Identifikators ausmachen würden.

4. Fazit und weiteres Vorgehen

Die Analyse der Option der Nutzung eines existierendes Identifikators zeigt klar, dass die in der Einführung gesetzten Anforderungen mit DOI gut abgedeckt werden könnten. Sinnvollerweise würden sich die Verlage zusammenschliessen und selber eine Registrierungsstelle einrichten. In dieser Hinsicht stellt sich die Frage, ob die Schaffung eines neuen Identifikators wirklich nötig ist, insbesondere weil dies darin bestehen würde, das bereits bei DOI Bestehende nochmals nachzubauen.

Es könnte auch vorgängig versucht werden die Frage zu beantworten, warum sich DOI bisher nicht durchgesetzt hat. Zu klären wäre, ob die Nachteile des DOI wirklich unüberwindbar sind bzw. ob sich die Kosten und Aufwände für eine eigene Registrierungsstelle rechtfertigen lassen.

Dieser Beitrag hat keinen Anspruch an konzeptuelle Vollständigkeit oder an Entscheidungsreife – er kann nur Ansporn liefern. Was er zu bewirken versucht ist vor allem eine Diskussion, in der konkrete weitere Schritte definiert werden. Ein möglicher Weg wäre, eine informelle Struktur aller interessierten Verlage zu bilden, welche die verfügbaren Optionen konkret definiert und analysiert, bevor sich alle Verlage zu einer formellen Struktur entschliessen und über den gewünschten Weg entscheiden.