

Henselmann, Klaus; Seebeck, Andreas; Grümmer, Julian

Research Report

Zielführende Betriebsprüfungen durch Nutzung von „Alternative Data“?

Suggested Citation: Henselmann, Klaus; Seebeck, Andreas; Grümmer, Julian (2021) : Zielführende Betriebsprüfungen durch Nutzung von „Alternative Data“?, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/248286>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Abschlussbericht zum Forschungsprojekt

Zielführende Betriebsprüfungen durch Nutzung von „Alternative Data“?

von Klaus Henselmann / Andreas Seebeck / Julian Grümmer *

Gefördert aus den Mitteln der Stiftung Prof. Dr. oec. Westerfelhaus

zur Interdisziplinären Forschung auf den Gebieten des Steuerrechts
und der Wirtschaftswissenschaften

<https://www.im.nrw/stiftung-prof-dr-oec-westerfelhaus>

Die Autoren bedanken sich sehr herzlich bei der Stiftung Prof. Dr. oec. Westerfelhaus für die finanzielle Unterstützung, welche die Umsetzung dieses Forschungsvorhabens möglich gemacht hat.

* Prof. Dr. Klaus Henselmann klaus.henselmann@fau.de
Dr. Andreas Seebeck andreas.seebeck@fau.de
Julian Grümmer, M.Sc. julian.gruemmer@fau.de

Lehrstuhl für Rechnungswesen und Prüfungswesen
Fachbereich Wirtschafts- und Sozialwissenschaften
Friedrich-Alexander-Universität Erlangen-Nürnberg
Lange Gasse 20
90403 Nürnberg

Inhaltsübersicht

I. Einführung	1
II. Grundlagen	3
A. Klassische Betriebsprüfung	3
1. Aufgabe	3
2. Ziele und Anlässe	4
3. Traditionelle Methoden zur Planung der Betriebsprüfung	6
B. Aktuelle Entwicklungen	7
1. Grundsätzliches.....	7
2. Country-by-Country Reporting	7
a) Regulatorischer Hintergrund und Idee	7
b) Technische Umsetzung & Daten.....	8
c) Aktuelle Entwicklung: Öffentliches CbCR	9
3. Meldepflichten steuerlicher Gestaltungen (DAC6).....	10
a) Regulatorischer Hintergrund & Idee	10
b) Technische Umsetzung & Daten.....	11
4. Continuous Transaction Control (CTC)	12
a) Regulatorischer Hintergrund & Idee	12
b) Technische Umsetzung und Daten	13
c) Beispiel: My Data (Griechenland).....	13
5. Aktuelle Projekte zur Aufdeckung von Steuerhinterziehung, Betrug und Geldwäsche ...	14
a) Common Reporting Standard – CRS.....	14
b) Ausschreibung Forschungsvorhaben des BMBF.....	15
c) EU-Programme in Horizont 2020 (horizon 2020)	16
6. Zwischenfazit	17
C. Chancen einer Nutzung von Alternativen Daten	19
1. Grundlagen	19
2. Aufdecken von Inkonsistenzen	21
3. Verbindungen zu Risikobereichen	24
a) Überblick.....	24
b) Art der Geschäftstätigkeit.....	25
c) Standort und Rechtsform	26
d) Innerhalb des Konzerns tätige Personen.....	28
e) Beziehungen zu Personen außerhalb des Konzerns.....	29
4. Anwendungsbeispiel.....	31
5. Schlussfolgerungen.....	34
III. Alternative Daten	36
A. Systematisierungsansätze.....	36
B. Nützliche Alternative Daten?	41
1. Vorbemerkungen	41
2. Unternehmen und Organe.....	41
a) Unternehmensregister	41
b) Identifizierung und Verknüpfung.....	42
c) Aufbereitete Daten.....	45

3. Gesellschafter und wirtschaftlich Berechtigte.....	47
4. Unternehmensveröffentlichungen	49
5. Steuerleaks.....	50
6. Graue Wirtschaft und Schattenwirtschaft.....	52
7. Nachrichten und Ereignisse	55
8. Soziale Medien.....	56
9. Lexika	57
10.Weitere Sammlungen	57
C. Ergebnis und Ausblick	59
IV. Prototypische Umsetzung.....	60
A. Technische Basis	60
1. Aufgabenstellung.....	60
2. Datensammlung.....	60
3. Einheitliches Datenformat	61
4. Datenintegration.....	65
5. Erforschung des Gesamtdatenbestands.....	67
B. Codierungsbeispiele.....	71
1. Deutsches Unternehmensregister	71
2. Linked Leaks und FactForge.....	79
3. Projekt execGraph	82
V. Zusammenfassung und Ausblick	93
Literatur.....	95

I. Einführung

Im Rahmen des Steuerverfahrens muss die Finanzverwaltung sowohl Unternehmen für die Betriebsprüfung auswählen als auch innerhalb der Unternehmen steuerliche Risikoschwerpunktbereiche identifizieren.

Traditionell stützt die Verwaltung sich dabei auf Veranlagungsdaten, Kennzahlenvergleiche mit anderen Unternehmen sowie ggf. auf E-Bilanz-Daten. Anhaltspunkte können auch die Meldepflichten des sog. Country-by-Country-Reporting (§§ 138a ff. AO) und die Meldepflichten für Steuergestaltungen (EU-Amtshilferichtlinie) bieten.

Meist handelt es sich bei Angaben des Steuerpflichtigen jedoch um (zunächst) ungeprüfte Daten, deren Vollständigkeit und Richtigkeit gerade bei „aggressiven“ Steuerpflichtigen nicht pauschal unterstellt werden kann.

Sinnvoll erscheint somit auch die ergänzende Verwendung anderer Datenquellen, sog. „Alternative Data“. Die Verfügbarkeit alternativer Datenquellen im Internet steigt stetig an. Der Begriff „Alternative Data“ stammt eigentlich aus der Finanzanalyse.¹ Er umfasst dort alle nicht-konventionellen Datenquellen, insbesondere solche, die über Börsenhandelsdaten und Daten aufgrund der Finanzberichterstattung durch das Unternehmen selbst hinausgehen. Alternative Daten werden bei der Finanzanalyse verwendet, um die vom Unternehmen gemachten Aussagen zu ergänzen, zu bestätigen oder eben zu widerlegen. Diesen Grundgedanken kann man auch auf die Besteuerung übertragen. Als Alternative Daten können hier solche Daten bezeichnet werden, die nicht im Zuge des Besteuerungsverfahrens von Steuerpflichtigen erklärt oder durch deren Verarbeitung innerhalb der Finanzverwaltung geschaffen wurden.

Der vorliegende Forschungsbericht geht im Grundlagenteil zunächst der Frage nach, welche Ansätze zur Planung der Betriebsprüfung national und international bereits üblich sind. Ausgehend von Aufgaben, Zielen und Anlässe der klassischen Betriebsprüfung wird hier auf die anlassabhängige und anlassunabhängige Betriebsprüfung und die traditionellen Methoden der Prüfungsplanung eingegangen.

Anschließend werden verschiedene nationale und internationale Entwicklungen vorgestellt, die unterstützend bei der Betriebsprüfungsplanung und -durchführung eingesetzt werden können. Die verschiedenen Projekte befassen sich u.a. mit der Sammlung und Auswertung von Daten zur Aufdeckung von Betrug, Geldwäsche o.ä. und weisen vielfach eine grundsätzliche Eignung zu Betriebsprüfungszwecken auf.

Auf dieser Basis wird in den folgenden Abschnitten abgeleitet, wie eine Erschließung alternativer Daten die Finanzverwaltung auf Unstimmigkeiten zur Steuerdeklaration oder besondere Risikoquellen hinweisen könnte. Alternative Data kann somit nicht nur die risikoorientierte Unternehmensauswahl verbessern, sondern auch die Betriebsprüfungen zielgerichteter und damit möglicherweise schneller und zugleich intensiver machen.

Im Hauptteil geht es ferner um eine Bestandsaufnahme Alternativer Daten. Alternative Daten haben verschiedene Eigenschaften, die ihre Nutzung erleichtern oder erschweren können. Es wird eine umfassende Auswahl von Alternative Data identifiziert, analysiert und beurteilt, um aufzuzeigen, welche

¹ Vgl. Monk et al. (2019).

dieser Daten bei der Unternehmensauswahl bzw. Schwerpunktsetzung im Zuge der Betriebsprüfung dienlich sein können.

Abschließend wird im Hauptteil prototypisch gezeigt, wie konkrete Datenabfragen für ausgewählte Fragestellungen im Kontext der Betriebsprüfung aussehen könnten. Dazu erfolgt zunächst ein Vergleich von unterschiedlichen Konzepten der Datenspeicherung, -verknüpfung und -abfrage. Auf dieser Grundlage wird ein Umsetzungsvorschlag unterbreitet.

Eine Zusammenfassung wichtiger Erkenntnisse und ein Ausblick schließen die Arbeit ab.

Einige Probleme sollen oder können nicht Gegenstand der vorliegenden Untersuchung sein. Hierunter fällt erstens die juristische Frage, inwieweit die vorgestellten Konzepte für Datenabfragen und/oder Datensammlungen nach heutigem Recht bereits zulässig wären. Ob und gegebenenfalls wie hierfür der rechtliche Rahmen geändert werden müsste, soll an dieser Stelle offenbleiben. Die Erfahrung zeigt jedoch, dass das Recht nicht unveränderlich ist, sondern Maßnahmen gegen „Base Erosion and Profit Shifting“ (BEPS)² zu entsprechenden Anpassungen führen können.³

Daneben werden auch praktische Fragen der Implementierung ausgeklammert. Hierunter fallen etwa die organisatorische Verankerung innerhalb der Finanzverwaltung, Anforderungen an die Technik und die Qualifikation der Mitarbeiter, das erforderliche Budget sowie dessen Finanzierung.

In diesem Zusammenhang soll auch nicht hinterfragt werden, wie die innerhalb der Finanzverwaltung bereits vorhandenen Daten durch bessere Auswertungen, umfangreichere Verknüpfungen und vermehrte Abfragen bereits Hinweise für zielgerechte Betriebsprüfungen liefern könnten.⁴

Last but not least verändern sich das Angebot an Alternativen Daten sowie die Bezugsmöglichkeiten ständig. Der im nachfolgenden Text dargestellte Stand bezieht sich grundsätzlich auf die Situation im September 2021.

² Vgl. OECD (2021).

³ Beispielsweise wird das von der Europäischen Kommission vorgeschlagene „öffentliche Country-by-Country-Reporting“ zumindest über bestimmte Länder (Staaten der EU sowie Steueroasen gemäß EU-Liste) künftig Pflicht für große internationale Konzerne. So die Entscheidung der EU-Wirtschaftsminister am 25.2.2021. Das ist bemerkenswert, denn es handelt sich um Daten, die bisher dem Steuergeheimnis unterlagen. Vgl. Kafsak / Schäfers (2021).

⁴ Hinweise auf das große Potenzial gibt ein Bericht des Bundesrechnungshofs zur Verbesserung der Umsatzsteuerbetrugsbekämpfung. Bundesrechnungshof (2020).

II. Grundlagen

A. Klassische Betriebsprüfung

1. Aufgabe

Der Gesetzgeber der AO hat an Stelle des früheren Begriffs „Betriebsprüfung“ bereits im Jahr 1977 den Begriff „Außenprüfung“ gesetzt. Dadurch kommt klar zum Ausdruck, dass die Betriebsprüfung weit auszulegen ist und nicht nur steuerliche Verhältnisse von Betrieben, sondern mitunter auch von Privatpersonen erfassen soll.⁵ Dies erleichtert der Finanzverwaltung den Zugang zu umfassenden Daten, die mitunter außerhalb der betrieblichen Sphäre der Unternehmen liegen. Die Finanzverwaltung selbst behält den Begriff Betriebsprüfung bei, siehe etwa die „Allgemeine Verwaltungsvorschrift für die Betriebsprüfung – Betriebsprüfungsordnung (BpO)“, weshalb auch im Rahmen dieses Forschungsberichts die Begriffe Betriebsprüfung und Außenprüfung synonym verwendet werden.

Eine Betriebsprüfung kann der umfassenden Aufklärung steuerrelevanter Sachverhalte dienen, die im Veranlagungsverfahren von Amts wegen nicht oder nicht zweckmäßig überprüft werden können. Sie ist ein Instrument der umfassenden finanzbehördlichen Sachaufklärung, welches letztlich die Steuergerechtigkeit fördert.⁶

In Abhängigkeit von der Größe des Unternehmens beträgt der **Prüfungsturnus** alle drei bis 20 Jahre. Während Großunternehmen im Bundesdurchschnitt alle drei bis vier Jahre einer Betriebsprüfung unterzogen werden, kommen Kleinbetriebe nur etwa alle 15 bis 20 Jahre in den Genuss einer Prüfung. Zu beachten ist, dass es erhebliche regionale Unterschiede gibt, die nicht zuletzt durch die Personalsituation in den jeweiligen Verwaltungseinheiten bedingt sind.

Der **Prüfungszeitraum** umfasst i.d.R. drei Jahre (§ 4 BpO), kann sich aber in begründeten Fällen, in denen mit nicht unerheblichen Änderungen der Besteuerungsgrundlage zu rechnen ist oder bei Verdacht einer Steuerstraftat oder -ordnungswidrigkeit, ausgeweitet werden. Eine Ausdehnung des Prüfungszeitraums ist dabei nicht bereits darin zu sehen, dass sich ein Prüfer Urkunden außerhalb des Prüfungszeitraums vorlegen lässt, soweit er diese für eine Schätzung für die geprüften Jahre benötigt und dies entsprechend begründen kann.⁷

Prüfungsgegenstand kann der gesamte für die Entstehung und Ausgestaltung eines Steueranspruchs erhebliche Sachverhalt sein – sowohl der Höhe als auch dem Grunde nach. Es werden die erklärten Einkünfte des Unternehmers vollumfänglich überprüft. Zu diesem Zwecke werden alle steuerlich relevanten Sachverhalte gewürdigt. Hierunter fallen alle Einkunftsarten und Besteuerungsmerkmale, selbst wenn sie mit den betrieblichen Verhältnissen in keinem direkten Zusammenhang stehen.⁸

⁵ Vgl. Haarmann (1977), S. 135; sowie Sauer (1988), S. 21f.

⁶ Vgl. BFH v. 9.8.1991, BStBl 1992 II S. 220; Tipke/Lang § 22 Rz. 225.

⁷ Vgl. BFH v. 4.2.1988 – V R 57/83, BStBl 1988 II S. 413.

⁸ Vgl. Harle / Nüdling / Olles (2020), S. 9f.

2. Ziele und Anlässe

Die Betriebsprüfung ist ein spezielles Verwaltungsverfahren, mit dem die Finanzbehörde das ihr in § 85 AO auferlegte **Ziel** durchsetzen kann: Die Steuer nach gesetzlichen Maßgaben gleichmäßig festzusetzen und zu erheben. Übereinstimmend mit dem Gesetzeswortlaut der AO charakterisiert der BFH die Betriebsprüfung als ein Instrument zur Sicherstellung der Gleichmäßigkeit der Besteuerung.⁹ Die Betriebsprüfung hat sich auf die tatsächlichen und rechtlichen Verhältnisse, die für die Steuerpflicht und die Steuerbemessung maßgebend sind – zugunsten wie zuungunsten des Steuerpflichtigen – zu erstrecken (§ 199 Abs. 1 AO).

Die Betriebsprüfung dient letztlich dazu, die Wettbewerbsneutralität der Besteuerung sicherzustellen. Sie ist darauf ausgerichtet die steuerlichen Verhältnisse des Steuerpflichtigen zu überprüfen und umfasst eine oder mehrere Steuerarten, wie Einkommen- und Körperschaftsteuer, Umsatzsteuer und ggf. Gewerbesteuer.

Es kann unterschieden werden zwischen **anlassunabhängigen** und **anlassabhängigen** Außenprüfungen. Die Entscheidung, ob und wenn ja, wann bei einem Betrieb eine Betriebsprüfung durchgeführt wird, liegt gemäß § 5 AO im pflichtgemäßen Ermessen der Finanzbehörde. Bei der Entscheidung ist der verfassungsrechtliche Verhältnismäßigkeitsgrundsatz zu beachten.

Eine besondere Begründung der Prüfungsanordnung oder ein konkreter Anlass als Begründung für die Prüfung ist aufgrund der Rechtsgrundlage des § 193 AO nicht erforderlich. So ist eine weitgehende Begründung gemäß BFH-Urteil vom 12.8.2002 nur dann erforderlich, wenn dies zum Verständnis der Prüfungsanordnung erforderlich ist, wie z.B. bei der Erweiterung des Prüfungszeitraums nach § 4 Abs. 3 BpO.

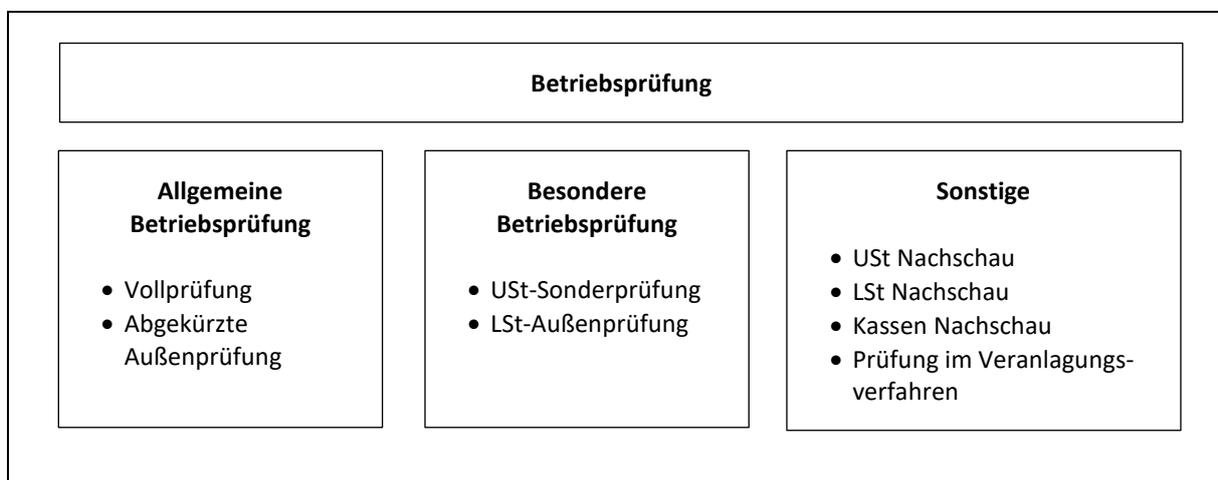


Abb. 1: Durchführungsmöglichkeiten der Außenprüfung

Quelle: Eigene Darstellung in Anlehnung an Harle / Nüdling / Olles (2020), S. 6.

⁹ Vgl. BFH v. 28.9.2011 – VIII R 8/09, BStBl 2012 II S. 395 = BB 2012 S. 3 737.

Die **normale Außenprüfung (Vollprüfung)** ist in § 193 ff. AO geregelt. Es gelten die oben dargestellten Rahmenbedingungen. Eine Schwerpunktbildung nach § 7 BpO ist überdies möglich. Daneben besteht die Möglichkeit der zeitnahen Betriebsprüfung nach § 4a BpO. Sie verfolgt das Ziel möglichst gegenwartsnahe Prüfungszeiträume zu untersuchen und ist häufig durch die Absicht der Verkürzung der Prüfungsdauer bei Großbetrieben oder der Angleichung von Prüfungszeiträumen durch die Finanzämter motiviert. Die wirtschaftlichen Vorteile einer zeitnahen Betriebsprüfung liegen auf der Hand. Neben erhöhter Rechts- und Planungssicherheit für die Unternehmen werden insbesondere die Risiken antizyklischer Steuernachforderungen (in ertragsschwachen Jahren) sowie hohe Nachzahlungszinsen vermieden.

Die **abgekürzte Außenprüfung** (§ 203 AO) beschränkt sich auf die wesentliche Besteuerungsgrundlage. Eine rasche Prüfungsdurchführung wird durch die Beschränkung auf bestimmte Sachverhalte, Zeiträume und Steuerarten ermöglicht. In der Regel bezieht sich die abgekürzte Außenprüfung auf einzelne steuerliche Sachverhalte, die sich auf klar begrenzbare Prüffelder eingrenzen lassen. Die wesentlichen Vorschriften der §§ 193 ff. AO finden Anwendung, mit der Ausnahme, dass kein Anspruch auf Durchführung einer Schlussbesprechung nach § 201 Abs. 1 und § 202 Abs. 2 AO besteht.¹⁰

Der Zweck von **Sonderprüfungen/Nachschau** besteht in der Prüfung einzelner Steuerarten (z.B. Umsatzsteuersonderprüfung, Lohnsteuer-Außenprüfung). Es gelten grundsätzlich die allgemeinen Grundsätze für die Betriebsprüfung nach § 193 AO.

Die **Umsatzsteuersonderprüfung** ist eine besondere Form der Betriebsprüfung. Sie wird aufgrund besonderer Umstände des Einzelfalls angeordnet und erstreckt sich ausschließlich auf umsatzsteuerrelevante Sachverhalte. Eine Besonderheit ist, dass der Prüfungszeitraum nicht nach § 4 Abs. 3 BpO auf drei Jahre beschränkt ist.¹¹

Die **Lohnsteuerußenprüfung** ist in § 42 ff. EStG geregelt. Prüfungsgegenstand ist die Einbehaltung und Abführung der Lohnsteuer durch den Arbeitgeber. Die Lohnsteuer-Außenprüfung wird durch Verwaltungsakt angeordnet. Für die Mitwirkungspflicht des Arbeitgebers bei der Außenprüfung gilt § 200 AO.

Die **Umsatzsteuernachschau** nach § 27b UStG ist keine Betriebsprüfung im Sinne des § 193 AO. Es handelt sich vielmehr um ein besonderes Verfahren zur zeitnahen Aufklärung möglicher steuererheblicher Sachverhalte. Sie findet unangekündigt statt und muss nicht zwingend vom Betriebsprüfer selbst durchgeführt werden. Von der Umsatzsteuernachschau betroffene Unternehmen und Personen haben den Prüfern relevante Unterlagen vorzulegen.

Die ebenfalls unangekündigte **Lohnsteuernachschau** nach § 42g EStG dient der zeitnahen Aufklärung von lohnsteuerrelevanten Sachverhalten. Im Falle von erheblichen Feststellungen kann ohne vorherige Prüfungsanordnung in eine Lohnsteuer-Außenprüfung übergegangen werden.

Die **Kassen-Nachschau** nach § 146b AO wurde mit dem 1.1.2018 eingeführt und ist eine weitere Form der unangekündigten zeitnahen Prüfung zur Sicherung des Steueraufkommens. Dabei werden die Ordnungsmäßigkeit der Kassenaufzeichnungen und die korrekte Übernahme der Kassenaufzeichnungen in die Buchführung überprüft. Auch hier kann im Falle von erheblichen Feststellungen ohne vorherige Prüfungsanordnung zur Außenprüfung übergegangen werden

¹⁰ Vgl. BFH v. 25.1.1989 – X R 158/87, BStBl 1989 II S. 483.

¹¹ Vgl. FG Rheinland-Pfalz v. 25.5.2007 – 6 K 1325/06, n.v.

Die **Prüfung im Veranlagungsverfahren** findet entweder in Form von Inaugenscheinnahmen gem. § 99 AO oder durch Einzelermittlungen nach §§ 85, 88 und 90 AO statt.

3. Traditionelle Methoden zur Planung der Betriebsprüfung

Traditionell erfolgt die Prüfungsauswahl bei Steuerpflichtigen, die von § 193 Abs. 1 AO erfasst sind, u.a. nach den folgenden Kriterien:

- Größenklasse
- Zeitlicher Turnus (Routine)
- Auswertung von Betriebsdaten i.e.S.
- Branchenzugehörigkeit
- Allgemeines Risikoprofil
- Zufallskriterien (Generalprävention und Überraschungsmoment).¹²

Das allgemeine Risikoprofil ergibt sich aus verschiedenen Risikoindikatoren. Diese beinhalten beispielsweise hohe Umsatz- oder Betriebsausgabenschwankungen in den letzten drei Jahren oder außergewöhnlich hohe Betriebsausgaben im Branchenvergleich.¹³ Daneben berücksichtigt die Finanzverwaltung in ihrer Prüfungsplanung aus steuerlicher Sicht typischerweise relevante Sachverhalte wie Betriebsaufspaltungen und -verpachtungen, Abgänge von Wirtschaftsgütern, die regelmäßig stille Reserven beinhalten, und Verträge mit Angehörigen sowie Privatentnahmen oberhalb des Gewinns. Hohe Forderungsverluste, Rückstellungen sowie besonders niedrige oder hohe Rechnungsabgrenzungsposten finden ebenfalls Berücksichtigung bei der Prüfungsauswahl.

Zu einer guten Prüfungsplanung gehört neben dem Studium der Steuerakten, die den Prüfern bereits im Vorfeld über die Unternehmen vorliegen, dass sich die Prüfer umfassend über das zu prüfende Unternehmen informieren. Dazu werden u.a. Unternehmenswebsites, Branchenreports und lokale Medien studiert. Daneben stehen den Finanzbehörden aus verschiedenen Projekten, die im Folgenden kurz vorgestellt werden, potenziell Daten zur Verfügung, die zu Prüfungszwecken genutzt werden können.

¹² Vgl. Harle / Nüdling / Olles (2020), S. 41.

¹³ Vgl. Bittner / Dawid / Metzner (2016), S. 243.

B. Aktuelle Entwicklungen

1. Grundsätzliches

Begrenzte Personalkapazitäten der Finanzverwaltung und Effizienzaspekte gehen einher mit dem Erfordernis Betriebsprüfungen zielorientiert, d.h. risikoorientiert, durchzuführen. Dies betrifft die risikoorientierte Auswahl der zu prüfenden Unternehmen sowie die risikoorientierte Prüfungsdurchführung.

Von der Finanzverwaltung wird (halb-)jährlich ein nicht öffentlicher Prüfungsgeschäftsplan nach § 34 Abs. 1 BpO erstellt, der die zu prüfenden Betriebe enthält. Die Qualität der darin festgelegten Prüfungen hängt von verschiedenen Faktoren wie einer gezielten Fallauswahl, einer zutreffenden Schwerpunktbildung, einer rationellen Prüfungsplanung und -organisation und einer zielführenden Prüfungsmethodik (insb. Stichprobenauswahl) sowie dem sinnvollen Einsatz von IT ab. Eine hohe Prüfungsintensität (Quantität) ist dabei nicht automatisch gleichzusetzen mit einer hohen Qualität. Entscheidend für die Qualität der Prüfung ist eine geeignete Fallauswahl und Schwerpunktbildung.

Verschiedene nationale und internationale Projekte, die ursprünglich zur Aufdeckung von Betrug, Geldwäsche o.ä. gedacht waren, wurden in den vergangenen Jahren initiiert und umgesetzt. Diese können den Finanzbehörden bei der Fallauswahl, Schwerpunktbildung und rationellen Prüfungsplanung und -organisation behilflich sein. Eine Fülle an Daten wird im Rahmen der Projekte gesammelt und ausgewertet. Einige von ihnen weisen eine grundsätzliche Eignung zu Prüfungsplanungszwecken und zur Durchführung von Außenprüfungen auf.

Im Folgenden werden einige ausgewählte Projekte vorgestellt. Dabei wird neben dem regulatorischen Hintergrund und den zugrundeliegenden Ideen auch auf die jeweils erzeugten Datenbestände eingegangen.

2. Country-by-Country Reporting

a) Regulatorischer Hintergrund und Idee

Das Country-by-Country Reporting wurde im Jahr 2016 eingeführt, um den Finanzbehörden eine erste Einschätzung von BEPS- und Verrechnungspreisrisiken zu ermöglichen.¹⁴ Die Analyse der CbCR-Daten durch die Finanzbehörden dient nicht dazu Betriebsprüfungen zu ersetzen und kann auch nicht als alleinige Grundlage für Verrechnungspreisanpassungen herangezogen werden. Aber die CbC-Reports erlauben eine erste überschlägige Risikoabschätzung anhand der Risikoindikatoren und unterstützen damit den zielgerichteten Einsatz der begrenzten Ressourcen der Finanzbehörden.¹⁵ So

¹⁴ Vgl. Lutz / Seebeck (2020), S. 55-59.

¹⁵ Vgl. Lutz / Seebeck (2019b), S. 438-449.

zielt eine erfolgreiche erste Risikoeinschätzung seitens der Finanzbehörden auf die Aussteuerung tatsächlich risikoreicher Unternehmen zur weiteren Nachforschung und auf eine korrekte Einstufung risikoarmer Unternehmen.¹⁶

Das Ziel des CbCR besteht darin, ggf. in Verbindung mit weiteren Informationsquellen Verrechnungspreis- und BEPS-Risiken überschlägig einzuschätzen.¹⁷ Die Interpretation der im Rahmen des CbCR berichteten Kennzahlen bzw. Risikoindikatoren erfordert regelmäßig die Gegenüberstellung mit Referenzwerten, wie Marktdaten, festen Grenzwerten, Vorjahreswerten oder Werten anderer konzerninterner bzw. -externer Unternehmen.¹⁸

Nach der CbCR-Primärpflicht (§ 138a Abs. 1 S. 1 AO) haben solche Unternehmen, die zur Aufstellung eines Konzernabschlusses verpflichtet sind, soweit ihr konsolidierter Vorjahresumsatz 750 Mio. EUR übersteigt und in ihren Konzernabschluss zumindest ein ausländisches Unternehmen oder eine ausländische Betriebsstätte einbezogen wird, ein CbCR für ihren Konzern zu erstellen.¹⁹ Die Berichterstattung besteht dabei aus drei Tabellen. In Tabelle 1 und 2 werden alle voll- und quotenkonsolidierten Konzerngesellschaften, d.h. Kapital- und Personengesellschaften sowie Betriebsstätten, ausgewiesen.²⁰ Die zehn Positionen der Tabelle 1 lassen sich in Erfolgs-, Ertragsteuer- und Substanzkennzahlen unterteilen. In der Tabelle 2 sind für jede Konzerneinheit separat und unterteilt nach den Ländern ihrer Ansässigkeit bzw. Belegenheit die wichtigsten Geschäftstätigkeiten anzugeben. Sie enthält insgesamt zwölf Kategorien wichtigster Geschäftstätigkeiten und eine Auffangkategorie „Sonstige“. Wird die Kategorie „Sonstige“ verwendet, ist hierauf näher in Tabelle 3 einzugehen.²¹ Die Tabelle 3 bietet als dritter Berichtsbestandteil schließlich Platz für weitere qualitative Angaben zu den beiden anderen Tabellen in Form von Freitextfeldern.

b) Technische Umsetzung & Daten

Damit die erste Risikoeinschätzung durch das CbCR ressourceneffizient erfolgen kann²², sind berichtspflichtige Unternehmen verpflichtet, das von der OECD vorgegebene Template im digitalen Berichtsformat eXtensible Markup Language (XML) zu verwenden.²³ XML ist eine Auszeichnungssprache

¹⁶ Vgl. Art. 16 Abs. 6 S. 2, 3 der Richtlinie 2011/16/EU (EU-Amtshilferichtlinie); BT-Drs. 18/9536 v. 5.9.2016, 37; OECD (Fn. 29), Tax Risk, Tz. 11-12, 14, 112-115, 117, 145; OECD (Fn. 2), TPG, Tz. 5.10, 5.25.

¹⁷ Vgl. Lutz / Seebeck (2019a), S. 535-543.

¹⁸ Vgl. OECD (2017) Tz. 37, 40, 48, 52.

¹⁹ Siehe auch Fuchs / Steiner (2016), S. 388; Kahle / Schulz (2016), S. 819.

²⁰ Vgl. BT-Drs. 18/9536 v. 5.9.2016, 38; Drüen (2018), AO § 138a Rz. 14; OECD (2010) Annex III to Chapter V (S. 512). Gegen eine Einbeziehung von quotenkonsolidierten Gemeinschaftsunternehmen, vgl. Grotherr (2016), S. 995.

²¹ Vgl. OECD (2010), Annex III to Chapter V (S. 511, 517).

²² Vgl. OECD (2017), Tz. 21, 120; OECD (2010) Tz. 5.10.

²³ Vgl. BMF v. 11.7.2017 – IV B 5 - S 1300/16/10010 :002, DOK 2017/0558036, BStBl. I 2017, 974 (974).

zur Darstellung hierarchisch strukturierter Daten. Im Zuge des XML-basierten Reportings werden einzelne Berichtselemente, darunter sowohl Zahlen als auch Fließtexte, mit einheitlich definierten Elementen markiert – das sogenannte Tagging. Dadurch werden die CbC-Reports maschinenlesbar.

Das XML-Format schafft somit die Voraussetzungen für eine automatisierte und damit ressourcenschonende Auswertung der Reports. Gleichzeitig ist es vorteilhaft für den (internationalen) Informationsaustausch zwischen den Finanzbehörden der Länder. Es vermeidet typische Schnittstellenprobleme beim Datenaustausch, wie fehleranfällige Medienbrüche, und reduziert damit das Erfordernis manueller Anpassungen auf ein Minimum.

Computer können dank der XML-Struktur verschiedene Risikoindikatoren anhand der übermittelten Erfolgs-, Ertragsteuer- und Substanzkennzahlen aus Tabelle 1 sowie der in Tabelle 2 angezeigten Ansässigkeit bzw. Belegenheit und wichtigsten Geschäftstätigkeit für jede Konzerneinheit automatisiert errechnen. Zudem können die Informationen in den Freitextangaben aus Tabelle 3 in die automatisierte Betrachtung einbezogen werden.²⁴

Die Verwendung von XML ermöglicht einen zielgerichteten Ressourceneinsatz der Finanzbehörden.²⁵ Sie erlaubt beispielsweise, dass die Risikoindikatoren der OECD maschinell ausgewertet werden können. Eine derartige Vorgehensweise wird durch den Untersuchungsgrundsatz nach § 88 Abs. 5 S. 1 AO getragen, wonach automatische Risikomanagementsysteme auf Seiten der Finanzbehörde zum Einsatz kommen dürfen.

Die Möglichkeiten der computergestützten Analyse von den CbCR im XML Format sind vielfältig. Sie reichen von vergleichsweise simplen Data Mining Verfahren wie der datenbankgestützten Analyse einzelner quantitativer Werte über Benchmark- und Zeitreihenanalysen bis hin zu technisch anspruchsvollen Text Mining Verfahren wie Machine Learning-basierten Red Flag Analysen.²⁶

Die OECD erachtet zu Zwecken der ersten Risikoeinschätzung zusätzliche Datenquellen als notwendig, um zutreffende Risikoeinschätzungen abgeben zu können. Diese sollen nach Möglichkeit bereits in die erste Risikoeinschätzung einfließen.²⁷

c) Aktuelle Entwicklung: Öffentliches CbCR

Bisher haben nur die Finanzämter Zugriff auf die CbCR Daten. Die Europäische Kommission hat jedoch bereits einen Entwurf für eine Richtlinie vorgelegt, mit der Unternehmen verpflichtet werden sollen, ihre CbC-Reports zu veröffentlichen (öffentliches CbCR).²⁸ Hierzu soll die europäische Bilanzrichtlinie (Richtlinie 2013/34/EU) geändert werden. Die Kommission geht damit über die Arbeiten der OECD im Rahmen des BEPS-Projekts hinaus.

²⁴ Vgl. Lutz / Seebeck (2020), S. 55.

²⁵ Vgl. OECD (2017) Tz. 21, 26, 120.

²⁶ Vgl. Seebeck / Lutz (2021), S. 260f. sowie Seebeck / Kaya (2021) und Seebeck / Vetter (2021) für mögliche Anwendungsbeispiele von Textminingverfahren im Kontext der Finanzberichterstattung.

²⁷ Vgl. OECD (2017), Tz. 22, 110.

²⁸ Vgl. Europäische Kommission, Vorschlag für eine Richtlinie des Europäischen Parlaments und des Rates zur Änderung der Richtlinie 2013/34/EU im Hinblick auf die Offenlegung von Ertragsteuerinformationen durch

Dem Vorschlag der Europäischen Kommission zufolge ist das öffentliche CbCR auf der Internetseite des Mutterunternehmens fünf Jahre lang öffentlich zugänglich zu machen.²⁹ Die tatsächliche Zugänglichkeit (nicht der Inhalt) ist durch den Abschlussprüfer zu prüfen und entsprechend zu testieren.³⁰

Kritiker argumentieren, dass ein öffentliches Country-by-Country Reporting nicht im Interesse der betroffenen Unternehmen sei, da betriebswirtschaftlich sensible Daten wie Gewinn und Umsatzerlöse öffentlich kenntlich gemacht werden würden. Konkurrenten können darüber direkte Rückschlüsse auf die Profitabilität der betroffenen Unternehmen in einzelnen Ländern ziehen, wodurch Nachteile gegenüber Konkurrenten, die nicht vom öffentlichen CbCR betroffen sind, entstehen können. Im Hinblick auf eine risikoorientierte Prüfungsplanung von Betriebsprüfungen erscheint ein öffentliches CbCR begrüßenswert.

3. Meldepflichten steuerlicher Gestaltungen (DAC6)

a) Regulatorischer Hintergrund & Idee

Die Anzeigepflichten für grenzüberschreitende Steuergestaltungen der Sechsten Änderungsrichtlinie (Richtlinie (EU) 2018/822) zur EU-Amtshilferichtlinie (Richtlinie 2011/16/EU) wurden in Deutschland mit dem Gesetz zur Einführung einer Pflicht zur Mitteilung grenzüberschreitender Steuergestaltungen fristgemäß zum 1.1.2020 in nationales Recht umgesetzt. Im Folgenden werden die Sechste Änderungsrichtlinie (Richtlinie (EU) 2018/822) und die hierdurch geänderte EU-Amtshilferichtlinie (Richtlinie 2011/16/EU) zusammenfassend als DAC6-Amtshilferichtlinie (DAC6) bezeichnet (engl.: DAC6-Directive on Administrative Cooperation).

Die DAC6-Amtshilferichtlinie stellt, wie die Amtshilferichtlinie insgesamt, einen Mindeststandard dar, so dass Mitgliedstaaten darüberhinausgehende Regelungen treffen können. Das deutsche DAC6-Umsetzungsgesetz³¹ setzt die Regelungen aus der DAC6-Amtshilferichtlinie unter den nach § 138 AO normierten Anzeigepflichten insbesondere in den neuen §§ 138d–138k AO sowie im Finanzverwaltungsgesetz und dem EU-Amtshilfegesetz um und entspricht weitgehend einer 1:1-Umsetzung der EU-Vorgaben.

Befeuert durch die auf den zwischenstaatlichen Steuerwettbewerb ausgerichtete Gesetzgebung einiger Staaten, wurden in den vergangenen Jahren von den Marktteilnehmern komplexe Modelle entwickelt, um die Steuerlast von individuellen Steuerpflichtigen und Unternehmen maßgeblich zu reduzieren bzw. zu vermeiden. Systematische und standardisierte Vorgehensweisen wurden entworfen, die darauf abzielen, bestehende Besteuerungslücken und -intransparenz zwischen den einzelnen

bestimmte Unternehmen und Zweigniederlassungen, im Folgenden: EU-Kommission (EU) (2016) 198 v. 12.4.2016.

²⁹ Vgl. EU-Kommission (EU) (2016) 198, Art. 48b Abs. 1.

³⁰ Vgl. EU-Kommission (EU) (2016) 198, Art. 48c Abs. 3.

³¹ Gesetz zur Einführung einer Mitteilungspflicht für grenzüberschreitende Steuergestaltungen v. 21.12.2019, BGBl. 2019 I S. 2875.

Staaten zum eigenen oder im Falle der Intermediäre zum Vorteil der Kunden auszunutzen. Die Meldepflichten der DAC6 dienen dazu, derartige aggressive Steuersparmodelle zu unterbinden.³²

Das erklärte Ziel der EU ist es, die Steuerbehörden frühzeitig über potenziell aggressive grenzüberschreitende Steuergestaltungen zu informieren. Dies ermöglicht etwa für die Finanzverwaltung eine frühzeitige Risikovorsorge zur zielgerichteten Betriebsprüfungstätigkeit (Informationsfunktion). Ein weiteres Ziel ist es, Intermediäre abzuschrecken, bei illegalen Steuerhinterziehungen und aggressiven - wengleich legalen - Steuerumgehungen beratend tätig zu werden (Abschreckungswirkung).³³

DAC6 wohnt eine doppelte Informationsfunktion inne: Erstens erhält der Gesetzgeber frühzeitig die Möglichkeit, auf aggressive Gestaltungen mit geeigneten gesetzlichen Maßnahmen zu reagieren.³⁴ Zweitens kann die Steuerverwaltung bei betroffenen Steuerpflichtigen gezielt ermitteln.³⁵ Die durch die DAC6 Meldepflichten erlangten Informationen über grenzüberschreitende Steuergestaltungen können auch zu Zwecken der Prüfungsplanung bei Betriebsprüfungen verwendet werden.

b) Technische Umsetzung & Daten

Die DAC6 Meldungen sind ausschließlich elektronisch, gemäß amtlich vorgeschriebenem Datensatz an das Bundeszentralamt für Steuern zu übermitteln. Dabei stehen drei Meldewege zur Verfügung:

- Einzeldatenübermittlung über das BZStOnline-Portal (BOP)
- XML-Web Upload im BOP
- Elektronische Massendatenschnittstelle (ELMA).³⁶

Aufgrund der DAC6 Meldepflichten wird jährlich eine fünfstellige Anzahl an Meldungen an das BZSt erwartet.³⁷ Dabei ist anzunehmen, dass Intermediäre eine Vielzahl an Sachverhalten vorsichtshalber melden werden, die von den Finanzbehörden zu prüfen sind. Daraus resultieren erhebliche Bürokratiekosten auf der einen Seite und umfangreiche Datenbestände auf der anderen Seite. Die Kosten in den ersten drei Jahren, die auf Seiten des Bundes (BZSt, Generalzolldirektion und Informationstechnikzentrum Bund) entstehen, schätzte der Regierungsentwurf auf ca. 21-23 Mio. €. ³⁸ Davon entfällt

³² Vgl. von Brocke et al. (2021).

³³ Vgl. Rat der EU, Verpflichtender automatischer Informationsaustausch im Bereich der Besteuerung über meldepflichtige grenzüberschreitende Modelle, Politische Einigung v. 9.3.2018 - Dok. 6804/48 FISC 103 ECOFIN 206 Nr. 2; Rat der EU, Mitteilung der Kommission über eine externe Strategie für effektive Besteuerung und Empfehlung der Kommission zur Umsetzung von Maßnahmen zur Bekämpfung des Missbrauchs von Steuerabkommen, Schlussfolgerungen des Rates v. 25.5.2016 - Dok. 9452/16 FISC 85 ECOFIN 502 Nr. 12.

³⁴ BMF, Schreiben zur Mitteilung grenzüberschreitender Steuergestaltungen vom 29.3.2021, V A 3 - S 0304/19/10006 :010, Tz. 3.

³⁵ BMF, Schreiben zur Mitteilung grenzüberschreitender Steuergestaltungen vom 29.3.2021, V A 3 - S 0304/19/10006 :010, Tz. 3

³⁶ Vgl. Bundeszentralamt für Steuern (2021b).

³⁷ Regierungsentwurf v. 9.10.2019, S. 25.

³⁸ Regierungsentwurf v. 9.10.2019, S. 25.

ein erheblicher Anteil auf den Aufbau einer leistungsfähigen IT-Infrastruktur, die u.a. auch KI-gestützte Tools zur Auswertung der Mitteilungen beinhaltet. Aus Sicht einer risikoorientierten Betriebsprüfung wäre es vielversprechend an diese IT-Infrastruktur und Datenbestände anzuknüpfen.

Wie bereits im Kontext des CbCR beschrieben, ermöglicht das Übermittlungsverfahren XML ergänzend zur textuellen Beschreibung eine maschinenlesbare, logische Struktur in die Informationen zu bringen. Beispielsweise kann die Unternehmensstruktur in Form eines Baumdiagramms hierarchisch strukturiert werden. So können die Verbindung und Abhängigkeit zwischen Nutzern, verbundenen Unternehmen und betroffenen Personen einschließlich der jeweiligen Besitzverhältnisse (in Prozent) technisch eindeutig vermerkt werden. Die Erfassung der Unternehmensstruktur ist dabei jedoch auf fünf Ebenen begrenzt.³⁹ Derartige Informationen erscheinen im Hinblick auf die Aussteuerung von Risiken in der Betriebsprüfung nützlich.

4. Continuous Transaction Control (CTC)

a) Regulatorischer Hintergrund & Idee

Mit der zunehmenden Digitalisierung der staatlichen Dienste hat die International Chamber of Commerce (ICC) eine Reihe von Praxisgrundsätzen für die Implementierung kontinuierlicher Transaktionskontrollen (engl. Continuous Transaction Controls - CTCs) veröffentlicht.⁴⁰ Diese spielen eine zunehmend wichtige Rolle bei der Festlegung der Höhe der Steuerzahlungen durch die Regierungen - und können darüber hinaus zur Überwachung der Einhaltung in einer Reihe anderer Bereiche eingesetzt werden.⁴¹

Im Zuge der fortschreitenden Digitalisierung haben Regierungen zunehmend cloudbasierte Dienste implementiert, um die Effizienz, Effektivität und Qualität öffentlicher Dienste zu verbessern. Diese als CTCs bezeichneten Verfahren ermöglichen es beispielsweise Strafverfolgungsbehörden aber auch den Steuerverwaltungen, Daten im Zusammenhang mit Geschäftstätigkeiten zu sammeln, die für die Ausübung ihrer Funktion relevant sind. Diese Daten werden direkt von Geschäftsdatenverwaltungssystemen in Echtzeit oder nahezu in Echtzeit abgerufen.

Dadurch ist eine Steigerung in der Effizienz der Steuererhebungsbemühungen zu erwarten. Denn CTCs beheben viele der Ineffizienzen, die mit rückwirkenden Prüfungen verbunden sind, bei denen Prüfer eine Transaktion erst lange nach ihrem Abschluss einsehen können und sich ausschließlich auf Daten stützen, die von den geprüften Unternehmen selbst gespeichert wurden. Waren die Steuerbehörden in der Vergangenheit darauf angewiesen, dass die Unternehmen historische Belege für die Hauptbücher vorlegen konnten, erfassen CTCs relevante Geschäftsinformationen beispielsweise direkt über ein Geschäftsvorgangsbuch, das aus authentifizierten Transaktionsquelldaten besteht. Auf

³⁹ Vgl. BZSt, Kommunikationshandbuch Automatischer Austausch von Steuergestaltungen, Version 1.7, vom 1.04.2021, S. 28.

⁴⁰ Vgl. <https://iccwbo.org/media-wall/news-speeches/continuous-transaction-controls-what-business-needs-to-know/>.

⁴¹ Vgl. <https://iccwbo.org/publication/icc-continuous-transaction-control-ctcs-practice-principles/>.

diese Weise ermöglichen CTCs Steuerverwaltungen, Geschäftstransaktionsdaten in Echtzeit oder nahezu in Echtzeit abzurufen, wodurch die Geschwindigkeit und Genauigkeit der Steuererhebungsmaßnahmen verbessert werden.

b) Technische Umsetzung und Daten

Es werden zwei Arten von CTCs unterschieden: Berichterstattung (Reporting) und Freigabe (Clearance). Beim Reporting handelt es sich um die regelmäßige elektronische Übermittlung von Geschäftsdaten in Echtzeit an staatliche (Steuer-)Verwaltungsplattformen, ohne dass eine Genehmigung der Steuerverwaltung für diese Daten und deren fortgesetzte Verarbeitung aus steuerlicher Sicht erforderlich ist. Beim Clearance handelt es sich ebenfalls um Geschäftstransaktionsdaten, die in Echtzeit elektronisch an (Steuer-)Verwaltungsplattformen übermittelt werden. Anders als beim Reporting ist jedoch eine Genehmigung erforderlich, damit diese Daten und ihre fortgesetzte Verarbeitung aus steuerlicher Sicht gültig sind. Die Verwaltung nimmt folglich bereits im Zeitpunkt der Transaktion eine aktive Rolle ein und validiert die Daten (z.B. Rechnung), bevor die Transaktion abgeschlossen wird. Beim Reporting ist es hingegen Aufgabe der Unternehmen, die Gültigkeit der Informationen (z.B. einer Rechnung) im Nachgang der Transaktion nachzuweisen. Folglich unterscheiden sich auch die Daten je nach Modell.

Beim Reporting können grundsätzlich unterschiedliche Dateiformate zulässig sein. Während einige Länder wie Spanien und Ungarn für diese Zwecke ihren eigenen XML-Standard definiert haben, basieren andere Länder wie Portugal und Polen ihre technischen Anforderungen ganz oder teilweise auf der von der OECD herausgegebenen SAF-T-Spezifikation (Standard Audit File for Tax). Eine SAF-T-Datei enthält Rechnungsdaten sowie verschiedene andere Daten über die zugrunde liegende Lieferung aus dem ERP-System der steuerpflichtigen Unternehmen.⁴²

Beim Clearance ist zwingend erforderlich, dass das berichtspflichtige Unternehmen das i.d.R. eng definierte vorgeschriebene Dateiformat verwendet, um Verzögerungen aufgrund der technischen Verarbeitung durch die Verwaltung zu verhindern.

c) Beispiel: My Data (Griechenland)

Mit Inkrafttreten der Richtlinie 2014/55/EU sind die zentralen öffentlichen Auftraggeber in Griechenland seit April 2019 verpflichtet, Rechnungen in einem EN-konformen Format entgegenzunehmen. Die Verpflichtung gilt seit April 2020 auch für alle anderen öffentlichen Auftraggeber. Das griechische Finanzministerium hatte 2018 zunächst mit einer „Testgruppe“ an Unternehmen begonnen, Rechnungen an die öffentliche griechische Verwaltung (B2G) und an Gesellschaften mit beschränkter Haftung in einem standardisierten Format auszustellen.⁴³

Die elektronische Rechnung ist in Griechenland seit dem 1. April 2021 nunmehr auch für B2B verpflichtend. Damit erweitert das Land die bestehenden Business-to-Government (B2G) E-Invoicing-Mandate um Business-to-Business (B2B) Rechnungen. Die erste Frist für die Einführung dieser neuen

⁴² Vgl. Kowallik / Eismayr / Kirsch (2016), S. 40ff.

⁴³ Vgl. Horák / Bokšová / Strouhal (2020), S. 449.

Verpflichtung war ursprünglich zum Januar 2020 geplant, der Implementierungstermin wurde jedoch aufgrund der Corona-Pandemie mehrfach verschoben.

Sämtliche E-Rechnungen müssen an die MyDATA-Plattform gesendet werden.⁴⁴ Hierauf haben die griechischen Behörden uneingeschränkten Zugriff und können somit alle Daten verarbeiten und daraus anschließend Buchhaltungsdaten für jeden Steuerzahler in Griechenland erstellen. Die Übertragungsmethode für myDATA ist die myDATA REST-API.⁴⁵

Die Rechnungen müssen das gemäß der europäischen Richtlinie vorgeschlagene elektronische Format für die Rechnungsstellung nach EN16931-1 einhalten. Derartige E-Rechnungen müssen mittels zertifizierter Software (ISO-Zertifizierung und AADE-Registrierung) ausgestellt werden.⁴⁶

5. Aktuelle Projekte zur Aufdeckung von Steuerhinterziehung, Betrug und Geldwäsche

Neben den oben beschriebenen Entwicklungen, die den Finanzbehörden bei der Fallauswahl, Schwerpunktbildung und rationellen Prüfungsplanung und -organisation behilflich sein können, gibt es einige aktuelle Projekte und Forschungsvorhaben zur Aufdeckung von Steuerhinterziehung, Betrug und Geldwäsche o.ä., die nachfolgend kurz vorgestellt werden. Diese Projekte machen deutlich, dass weiteres Potenzial besteht, um die Betriebsprüfung durch Berücksichtigung zusätzlicher Informationsquellen und alternativer Daten effizienter zu gestalten.

a) Common Reporting Standard – CRS

Von der OECD wurde der „Common Reporting Standard – CRS“ zum internationalen Austausch von Finanzkonteninformationen, insbesondere Kontensalden und bestimmten Zahlungsströmen, entwickelt.⁴⁷ Der CRS verpflichtet Banken den Kampf gegen Steuerflucht zu unterstützen. Er hat das Ziel grenzüberschreitende Steuersachverhalte aufzudecken und dadurch Steuerhinterziehung zu erschweren. Ihm haben sich inzwischen mehr als 90 Staaten angeschlossen.⁴⁸ Zur Erhebung der Daten sind die Finanzinstitute (z.B. Banken, Versicherungen) verpflichtet.⁴⁹ Sie melden diese in Deutschland an das Bundeszentralamt für Steuern. „Das BZSt leitet die Informationen an die zuständigen Behörden der Partnerstaaten weiter. Gleichzeitig empfängt es die Informationen aus dem Ausland und

⁴⁴ Vgl. Ministry of Economics, Greece (2020).

⁴⁵ Vgl. AADE (2020).

⁴⁶ Richtlinie 2014/55/EU des Europäischen Parlaments und des Rates vom 16. April 2014 über die elektronische Rechnungsstellung bei öffentlichen Aufträgen, Amtsblatt der Europäischen Union L 133, 6.5.2014, S. 1-11.

⁴⁷ Vgl. Bundeszentralamt für Steuern (2021a).

⁴⁸ Vgl. Bundeszentralamt für Steuern (2020).

⁴⁹ Grundlage ist das Gesetz zum automatischen Austausch von Informationen über Finanzkonten in Steuersachen (Finanzkonten-Informationsaustauschgesetz - FKAustG).

übermittelt diese dann an die zuständigen deutschen Finanzämter. Die Auswertung der Daten erfolgt in den Finanzämtern bzw. den jeweiligen Steuerbehörden im Ausland.“⁵⁰

Die zu meldenden Finanzinformationen umfassen diverse Arten von Kapitalerträgen (unter anderem Zinsen, Dividenden, Einkünfte aus bestimmten Versicherungsverträgen und andere ähnliche Erträge), aber auch Kontoguthaben und Erlöse aus der Veräußerung von Finanzvermögen.

Die von den Finanzinstituten an das BZSt zu meldenden Daten umfassen insbesondere:

- Name, Adresse und Steueridentifikationsnummer
- Geburtsdatum und Geburtsort
- Steuerlicher Wohnsitz
- Kontonummer
- Name und Identifikationsnummer des meldenden deutschen Finanzinstituts
- Kontosaldo oder -wert zum Ende des betreffenden Kalenderjahres
- Bei Verwahrkonten jeweils der Gesamtbruttoertrag der Zinsen, der Dividenden und anderer Einkünfte, die mittels der Vermögenswerte dieses Kontos erzielt und diesem gutgeschrieben wurden
- Bei Einlagekonten der Gesamtbruttoertrag der Zinsen, die auf das Konto eingezahlt oder diesem gutgeschrieben wurden
- Bei allen anderen Konten der Gesamtbruttobetrag, der in Bezug auf das Konto an den Kontoinhaber gezahlt oder diesem gutgeschrieben wurde und für den das meldende deutsche Finanzinstitut Schuldner ist. Die Gesamthöhe aller im Meldezeitraum geleisteten Einlösungsbeiträge ist einzuschließen
- Bei Verwahrkonten die Gesamtbruttoerlöse aus der Veräußerung oder dem Rückkauf von Vermögensgegenständen, die auf das Konto eingezahlt oder diesem gutgeschrieben wurden und für die das Finanzinstitut als Verwahrstelle, Makler, Bevollmächtigter oder anderweitig als Vertreter für den Kontoinhaber tätig war.⁵¹

Damit sind vom CRS diverse Informationen betroffen, die auch aus Sicht der Betriebsprüfung von Relevanz sein können.

b) Ausschreibung Forschungsvorhaben des BMBF

Wie in Kapitel II.B.1 beschrieben, besteht das Ziel des CbCR darin, den Finanzbehörden zusätzliche Informationen zu grenzüberschreitenden Unternehmensgruppenstrukturen an die Hand zu geben.

⁵⁰ Bundeszentralamt für Steuern (2021a).

⁵¹ Vgl. § 8 Allgemeine Meldepflichten des Gesetzes zum automatischen Austausch von Informationen über Finanzkonten in Steuersachen (Finanzkonten-Informationsaustauschgesetz - FKAustG).

Die aus den von den Unternehmen gemeldeten Daten entstehende Datenbasis kann und soll für statistische Auswertungen genutzt werden, um das Ausmaß von grenzüberschreitenden Geschäftsbeziehungen innerhalb von Unternehmensgruppen zu analysieren.

Gesetzgeber und Steuerverwaltung benötigen einen wissenschaftlich fundierten und vertieften Einblick in die Datenbestände. Zu diesem Zweck wurde der Forschungsauftrag „fe 4/20: Grenzüberschreitende Geschäftsbeziehungen innerhalb von Unternehmensgruppen – Ausmaß und Reformoptionen“ ausgeschrieben, welcher das Ausmaß von grenzüberschreitenden Geschäftsbeziehungen innerhalb von Unternehmensgruppen untersuchen soll. U.a. soll im Rahmen des Forschungsprojekts untersucht werden,

- wie sich die Geschäftsaktivitäten (qualitativ und quantitativ) einer Unternehmensgruppe auf Aktivitäten zu fremden Dritten und nahestehenden Personen verteilen;
- wie hoch das Ausmaß an potenzieller Gewinnverschiebung ins (niedriger besteuerte) Ausland ist;
- wie hoch der potenzielle Verlust an deutscher Bemessungsgrundlage aufgrund von steuerlichen Gestaltungen ist.

Als Datengrundlage stehen den Forschenden die länderbezogenen Berichte gemäß § 138a der Abgabenordnung zur Verfügung. Dabei sollen sowohl die aus dem Aus- als auch aus dem Inland stammenden Berichte herangezogen werden.⁵²

Wie bereits oben beschrieben, sind die CbCR Daten von potenzieller Relevanz für Betriebsprüfer und somit auch die Ergebnisse des ausgeschriebenen Forschungsvorhabens.

c) EU-Programme in Horizont 2020 (horizon 2020)

Horizont 2020 ist das aktuelle Rahmenprogramm der Europäischen Union für Forschung und Innovation.⁵³ Als Förderprogramm zielt es darauf ab, EU-weit eine wissens- und innovationsgestützte Gesellschaft und eine wettbewerbsfähige Wirtschaft aufzubauen sowie gleichzeitig zu einer nachhaltigen Entwicklung beizutragen. Es gliedert sich in drei Schwerpunkte und vier zusätzliche Teilbereiche. Ein Schwerpunkt ist „Gesellschaftliche Herausforderungen“.⁵⁴ Im Unterpunkt (g) „Sichere Gesellschaften – Schutz der Freiheit und Sicherheit Europas und seiner Bürger“⁵⁵ des Schwerpunkts wird unter anderem das Einzelprojekt „*Multimedia Analysis and Correlation Engine for Organised Crime Prevention and Investigation (MAGNETO)*“ gefördert.⁵⁶

MAGNETO befasst sich mit den erheblichen Bedürfnissen von Strafverfolgungsbehörden (LEAs) bei der Bekämpfung von Terrorismus und organisierter Kriminalität im Zusammenhang mit dem massi-

⁵² Vgl. <https://www.bundesfinanzministerium.de/Content/DE/Standardartikel/Service/Ausschreibungen/2020-08-27-ausschreibung-forschungsvorhaben-fe-4-20.html> (28.08.2020).

⁵³ <https://www.horizont2020.de/einstieg-kurzueberblick.htm>.

⁵⁴ Vgl. <https://cordis.europa.eu/programme/id/H2020-EU.3./de>.

⁵⁵ Vgl. <https://cordis.europa.eu/programme/id/H2020-EU.3.7./de>.

⁵⁶ Vgl. <https://cordis.europa.eu/project/rcn/216142/factsheet/en>.

ven Volumen, der Heterogenität und der Fragmentierung der Daten, die Beamte zur Verhütung, Ermittlung und Verfolgung von Straftaten analysieren müssen. Vergleichbare Herausforderungen im Umgang mit heterogenen Daten bestehen auch für die Finanzverwaltungen der Länder. Daher erscheint das Projekt auch aus Sicht einer effizienten, risikoorientierten Betriebsprüfung interessant.

Die Erforschung und Bereitstellung maßgeschneiderter Lösungen und Tools, die auf ausgefeilter Wissensrepräsentation, fortschrittlichem semantischem Denken und künstlicher Intelligenz basieren wie auch die Entwicklung einer internationalen gemeinsamen modularen Plattform mit integrierten Schnittstellen, welche im Rahmen der internationalen Strafverfolgung vorangetrieben wird, kann wegweisend für die digitale Betriebsprüfung sein. Ein Zugriff auf entsprechende Ressourcen und Technologien wäre für die Finanzverwaltungen zudem von großem Wert.

Mehrere massive heterogene Datenquellen zu verschmelzen und zu analysieren, verborgene Beziehungen zwischen Datenelementen aufzudecken, Trends für die Entwicklung von Sicherheitsvorfällen zu berechnen und letztendlich (und schneller) solide Beweise und beweiskräftige Unterlagen vor Gericht zu erlangen, ist eine Herausforderung, die gleichsam für Strafverfolgungsbehörden und Betriebsprüfer gilt. Auch hier lassen sich folglich Synergien heben.

6. Zwischenfazit

Es wurde deutlich, dass bereits viele Entwicklungen auf nationaler und internationaler Ebene stattfinden, die die Effizienz von klassischen, risikoorientierten Betriebsprüfungen immens steigern können. Gleichzeitig zeigen mit der Betriebsprüfung i.w.S. thematisch verwandte Projekte, dass eine enorme Ressourcenbindung und technische Ausstattung erforderlich sind, um die verfügbaren Datenmengen sinnvoll und zielgerichtet auswerten zu können.

Neben strukturellen Defiziten stellte der Bundesrechnungshof kürzlich im Hinblick auf die Umsatzsteuerbetrugsbekämpfung zahlreiche Vollzugsmängel und sogar Rückschritte bei der Betrugsbekämpfung fest.⁵⁷ Laut Bericht vom 29.10.2020 fehle ein Konzept, wie Zukunftstechnologien zur Umsatzsteuerbetrugsbekämpfung genutzt werden können. Die derzeitige IT-Unterstützung bei der Betrugsbekämpfung sei unzureichend, zentrale IT-Systeme für die umsatzsteuerlichen Kontrollen veraltet. Ferner finde aufgrund mangelnder technischer Infrastruktur kein automatisierter Datenaustausch zwischen den Zentralstellen für Betrugsbekämpfung statt. All dies sind Defizite, die es auch im Hinblick auf eine risikoorientierte, digitale Betriebsprüfung, welche alternative data zu Planungs- und Analysezwecken berücksichtigt, zu beseitigen gilt.

In anderen Ländern wie beispielsweise in Griechenland, Spanien und Italien, ist die Einführung einer elektronischen Echtzeitüberwachung (Continuous Transaction Control) in bestimmten Bereichen bereits fortgeschritten. So erhalten etwa die Finanzbehörden in Spanien oder Italien Umsatzdaten in Echtzeit und können dadurch den zeitlichen Vorsprung von aggressiv-gestaltenden oder gar betrügerischen Steuerpflichtigen minimieren. In Griechenland wurde kürzlich die elektronische Rechnung auch für den B2B Bereich verpflichtend eingeführt. Derartige Entwicklungen sind wegweisend für eine zeitnahe, effiziente Betriebsprüfung.

⁵⁷ Vgl. Bundesrechnungshof (2020).

Die deutsche Finanzverwaltung sollte die Chancen der Digitalisierung nutzen, um die Steuerbetrugsbekämpfung zukunftsfähig zu gestalten. Auch wenn der primäre Schwerpunkt des Umsatzsteuer-Berichts des Bundesrechnungshofs auf Verbrauchsteuern und weniger auf Gewinnverschiebungen (BEPS) liegt, sollten die Anregungen aus den Projekten zur Aufdeckung von Steuerhinterziehung, Betrug und Geldwäsche aufgegriffen werden. Ferner sollten alternative Daten verstärkt in die Prüfung einbezogen werden.

C. Chancen einer Nutzung von Alternativen Daten

1. Grundlagen

Der Begriff „Alternative Data“ stammt eigentlich aus der Finanzanalyse für Kapitalanlagen.

Als „Alternative Daten“ bezeichnet man hier nach verbreiteter Auffassung alle Daten, die nicht konventioneller Weise für Kapitalanlageentscheidungen genutzt werden.⁵⁸ Zu den konventionellen Daten gehören etwa die periodische Finanzberichterstattung des Unternehmens durch Jahres- und Quartalsabschlüsse, aperiodische Berichte des Unternehmens (wie Ad-hoc-Meldungen, Investorenkonferenzen, Wertpapierprospekte, Eintragungen ins Handelsregister), wertpapierspezifische Daten (wie Kursentwicklung und Börsenumsätze) sowie makroökonomische Rahmendaten (Zinsstrukturkurve, Inflationsraten, Wechselkurse, Bautätigkeit, Wachstumserwartungen u.a.).

Wie das Beispiel von CSR-Berichten⁵⁹ zeigt, ist dies keine statische Abgrenzung, sondern durchaus Änderungen unterworfen. Informationen über die Auswirkungen des Unternehmens auf Umwelt, Beschäftigte, Gesellschaft usw. hätte man früher noch eindeutig den Alternativen Daten zugeordnet. Im Lichte der Erkenntnis, dass CSR-Berichte u.a. auch wertvolle Hinweise über die Existenz von teilweise gravierenden Risiken liefern können, werden sie heute zunehmend auch für Kapitalanlageentscheidungen herangezogen.⁶⁰

Als Alternative Daten in der Finanzanalyse können aktuell noch gelten:⁶¹

- Satellitenbilder über wirtschaftliche Aktivität (mit Autos belegte Parkplätze, Schiffsbewegungen, Landwirtschaft, Abbau von Bodenschätzen);
- Streams von Meldungen/Postings auf sozialen Medien, aus denen sich Einstellungen zu Konsum, Politik oder anderen Bereichen ableiten lassen;
- Mikrodaten über das Einkaufsverhalten von Konsumenten (z.B. Kreditkartenkäufe, In-App-Käufe bei Smartphones);
- von Webseiten im Internet extrahierte Daten (z.B. Stellenausschreibungen, um Einstellungsmuster zu erkennen);
- Datenspuren wie Chroniken, Cookies und andere digitale Fußabdrücke, die Benutzer durch ihr Surfen im Netz hinterlassen (einschließlich Geodaten von Suchen auf Smartphones).

Für Entscheidungen zur Planung von steuerlichen Betriebsprüfungen führt das Konzept der Alternativen Daten – im Sinne von konventioneller Weise nicht genutzten Daten – natürlich zu ganz anderen Ergebnissen.

⁵⁸ Vgl. Monk et al. (2019).

⁵⁹ Mit der sog. CSR-Richtlinie verpflichtete die EU bestimmte große Unternehmen zur Abgabe einer „Nichtfinanziellen Erklärung“ bzw. „Konsolidierten nichtfinanziellen Erklärung“ (Richtlinie 2014/95/EU des Europäischen Parlaments und des Rates vom 22. Oktober 2014). Diese wurde in Deutschland mit den §§ 289b-289e HGB und §§ 315b-315c HGB umgesetzt.

⁶⁰ Vgl. Dhaliwal, et al. (2011); Cohen et al. (2017); Eccles et al. (2017); Amel-Zadeh / Serafeim (2018).

⁶¹ Vgl. Monk et al. (2019).

Konventionelle Daten sind für die Betriebsprüfung zunächst einmal Daten, die der Steuerpflichtige im Rahmen seiner Steuererklärungen, Berichte, Anzeigen, Auskünfte den Finanzbehörden aktiv mitteilt oder in der Vergangenheit bereits mitgeteilt hat. Hinzu kommen Daten, welche die Finanzbehörden von anderen Steuerpflichtigen gewinnen können. Dies ermöglicht beispielsweise die Ermittlung von Richtsätzen auf statistischer Basis u.a. für verschiedene Aufwandskategorien im Verhältnis zum Umsatz oder für international übliche Lizenzraten. Ferner können Daten auch von Finanzbehörden anderer Staaten stammen. In gewissem Ausmaß lassen sich auch öffentliche Statistiken (etwa der Bundesbank) als konventionelle Daten einstufen.

Vor diesem Hintergrund ist die Bandbreite Alternativer Daten in der Betriebsprüfung potenziell enorm. Dabei handelt es sich um jedwede Daten aus Quellen außerhalb der Steuererklärung oder der sonstigen oben genannten Quellen der Finanzbehörden, die geeignet sind, die in der Steuererklärung gemachten Aussagen zu bestätigen oder eben zu widerlegen.⁶² Insbesondere bieten in diesem Zusammenhang viele in der Finanzanalyse bereits verbreitet genutzte Daten ein in der Betriebsprüfung noch nicht genügend genutztes Potenzial, das es aus Sicht der Finanzbehörden zu heben gilt. So stellen beispielsweise auch Postings in sozialen Medien, online verbreitete Nachrichten oder Webseiten der Firmen zu weiten Teilen immer noch Neuland für die Betriebsprüfung dar.⁶³

Die Effektivität von Betriebsprüfungen ließe sich durch die Nutzung von Alternativen Daten grundsätzlich auf zwei Ebenen steigern:

Auf **Ebene der Unternehmensauswahl** für die Betriebsprüfung werden nur die wenigsten Unternehmen zeitlich lückenlos anschlussgeprüft. In allen anderen Fällen untersucht die Betriebsprüfung die steuerlichen Verhältnisse nur für ausgewählte Jahre. Je nach Unternehmenskategorie können die zeitlichen Abstände bis zur nächsten Prüfung sehr groß sein. Pro Jahr wird somit eine Stichprobe von Unternehmen für die Betriebsprüfung gebildet. Ziel ist es hier gezielt Unternehmen in die Stichprobe einzubeziehen, bei denen Alternative Daten Indizien liefern, die auf ein höheres Risiko an unabsichtlich oder absichtlich verkürzten Steuern hindeuten.

Auf **Ebene der Durchführung der Betriebsprüfung** können gerade bei größeren Unternehmen und Unternehmensgruppen nicht sämtliche Tatbestände überprüft werden. Hier müssen Prüfungsschwerpunkte gesetzt werden. Alternative Daten ermöglichen hier in gewisser Hinsicht eine Art der (gezielten) Stichprobenbildung. Im Sinne einer bewussten Stichprobenauswahl wären etwa zusätzliche Hinweise auf Bereiche mit erhöhtem steuerlichen Risiko sehr wertvoll. Das gilt selbst dann, wenn der Prüfer im Falle einer zeitlich lückenlosen Anschlussprüfung die Prüfungsschwerpunkte rollierend wechseln kann.

Methodisch bieten sich zwei Wege zur Ableitung von Indizien aus Alternativen Daten an:

- (1) **Inkonsistenzen** zwischen den Daten gemäß den Erklärungen, Berichten, Anzeigen, Auskünften sowie weiteren Unterlagen des Steuerpflichtigen auf der einen Seite und Alternativen Daten auf der anderen Seite.

⁶² Die Idee externe Quellen zur risikoorientierten Prüfungsplanung einzusetzen, findet sich auch in der internen Revision. Vgl. Peemöller / Kregel (2010), S. 211.

⁶³ Peemöller et al. (2020), S. 310, weisen für Accounting Fraud darauf hin, dass die Aufdeckung von Bilanzdelikten in überraschend vielen Fällen nicht durch die interne Revision, Wirtschaftsprüfer oder Aufsichtsräte erfolgte, sondern (neben Whistleblowern) oft auch durch Recherchen von Journalisten, die sich öffentlich zugänglicher Quellen bedienten. Ebenso Endt et al. (2019).

- (2) Durch Alternative Daten aufgedeckte Verbindungen zu Risikobereichen gemäß einer Liste von „**Red Flags**“.⁶⁴ Diese Verbindungen können auch kettenartig über mehrere Schritte wirken.

Beide Wege sollen im Folgenden kurz umrissen werden.

2. Aufdecken von Inkonsistenzen

Von Inkonsistenz soll im Folgenden gesprochen werden, wenn zum gleichen Sachverhalt mehrere – mindestens zwei – Datenbestände vorliegen, die nicht miteinander vereinbar sind. Sie sind also widersprüchlich. Wenn wir davon ausgehen, dass die Daten einheitlich definiert wurden,⁶⁵ muss mindestens einer der Datenbestände falsch sein. Das heißt, es besteht ein Risiko falscher Steuerdaten.

Erste Ansätze zum Aufdecken von Inkonsistenzen (Konsistenzchecks) werden bereits in Einzelfällen durch die Finanzbehörden praktiziert:

- Vergleich der in der Steuererklärung angegebenen Entfernung für den Weg zur Arbeit mit der Entfernung zwischen Adresse der Wohnung und Adresse der Arbeitsstätte gemäß Routenplanung (u.a. Google Maps, Apple Maps, Microsoft Windows Maps App).⁶⁶
- Anbieten einer Wohnung zum Mieten auf airbnb.com mit (fehlenden) Angaben zur Vermietung in der Steuererklärung.⁶⁷
- (Sehr) Viele Bewertungen auf Ebay.de als Folge umfangreicher Verkäufe mit (fehlenden) Angaben zu einem möglichen gewerblichen Handel in der Steuererklärung.⁶⁸

Es ist offenkundig, dass es sich hierbei um eher „kleine Fische“ handelt.

Wesentlich höheres Potenzial versprechen Ansätze, die eventuelle Steuergestaltungen größerer Unternehmen betreffen.⁶⁹ Das betrifft nicht nur große Konzerne, sondern auch viele kleine und mittlere Unternehmen.⁷⁰ Von hoher Relevanz sind insbesondere internationale Sachverhalte. Auf diese Daten hat die Finanzverwaltung keinen unmittelbaren Zugriff und die Sachverhalte betreffen nicht selten auch Jurisdiktionen, deren Behörden keine oder kaum Daten übermitteln. Außerdem kann durch gezielte Steuerplanung und -gestaltung Besteuerungssubstrat von Deutschland weg auf andere Staaten verschoben werden.

⁶⁴ Ein alternativer Begriff wären schwache Signale oder „weak signals“. Vgl. Hofmann (2008), S. 412.

⁶⁵ Diese Annahme ist nicht trivial. Ein Unternehmen könnte z.B. korrekterweise gleichzeitig 300, 430 und 450 Beschäftigte haben, wenn diesen Zahlen verschiedene Definitionen zugrunde liegen: Stand am Jahresende oder im Jahresdurchschnitt? Mit oder ohne Praktikanten? Nur sozialversicherungspflichtige Arbeitnehmer oder alle? Gerechnet nach Köpfen oder Vollzeitäquivalente?

⁶⁶ Vgl. Schönwitz (2014).

⁶⁷ Vgl. Schneider (2020).

⁶⁸ Hierfür nutzt das Bundesamt für Finanzen die Software „XPider“. Vgl. Schaefer-Drinhausen (2021).

⁶⁹ Zum Ausmaß u.a. Zucman (2015).

⁷⁰ Vgl. von Daniels / Wörpel (2019).

Als ein **Anwendungsbeispiel** für Konsistenzchecks kann das oben geschilderte Country-by-Country Reporting von großen Konzernen mit mindestens einer ausländischen Konzerneinheit dienen (§ 138a Abs. 1 Satz 1 AO). Diese Berichte umfassen die Listen aller in den jeweiligen Ländern ansässigen bzw. belegenen Konzerneinheiten⁷¹ sowie u.a. deren wichtigste Geschäftstätigkeiten⁷². Hierbei können als Freitext Erläuterungen gegeben werden.⁷³ Es gibt verschiedene Alternative Daten, mit denen man diese Steuerdaten manuell oder automatisiert abgleichen kann.

Eine erste Quelle bilden **Jahresabschlüsse** (Konzernabschluss und Einzel-Jahresabschlüsse). Der HGB-Einzelabschluss von Kapitalgesellschaften muss u.a. benennen: Alle Beteiligungen (§ 285 Nr. 11 HGB); alle Unternehmen, deren unbeschränkt haftender Gesellschafter die Kapitalgesellschaft ist (§ 285 Nr. 11a HGB); bei börsennotierten Kapitalgesellschaften sogar alle Anteile an großen Kapitalgesellschaften, die 5 % der Stimmrechte überschreiten. Für den HGB-Konzernabschluss wird insbesondere verlangt: Alle Tochterunternehmen (§ 313 Abs. 2 Nr. 1 HGB); alle assoziierten Unternehmen (§ 313 Abs. 2 Nr. 2 HGB); alle quotalkonsolidierten Unternehmen (§ 313 Abs. 2 Nr. 3 HGB); alle Beteiligungen (§ 313 Abs. 2 Nr. 4 HGB); alle Unternehmen, deren unbeschränkt haftender Gesellschafter die Kapitalgesellschaft ist (§ 313 Abs. 2 Nr. 6 HGB); bei börsennotierten Kapitalgesellschaften ebenfalls alle Anteile an großen Kapitalgesellschaften, die 5 % der Stimmrechte überschreiten (§ 313 Abs. 2 Nr. 5 HGB). Für IFRS-Abschlüsse verlangt IFRS 12 „Angaben zu Anteilen an anderen Unternehmen“ weitreichende Hinweise.⁷⁴

Eine zweite Quelle stellen die **Unternehmensregister** (Company Registers) der jeweiligen Länder dar. Für Deutschland ist dies das Unternehmensregister gemäß § 8b HGB (www.unternehmensregister.de), welches auch das Handelsregister mit umfasst (§ 8b Abs. 2 Nr. 1 HGB). In das Handelsregister werden bei Gründung die Namen des Eigentümers bzw. der Gesellschafter eingetragen (Einzelunternehmen § 29 HGB, offene Handelsgesellschaft § 106 Abs. 2 Nr. 1 HGB, Kommanditgesellschaft § 162 Abs. 2 HGB, GmbH § 8 Abs. 1 Nr. 3 GmbHG). Aber auch spätere Eigentümer- bzw. Gesellschafterwechsel müssen mitgeteilt werden. Durch entsprechende Aggregationen lassen sich auf dieser Basis aus dem Gesamtdatenbestand auch alle Unternehmen herausfiltern, bei denen z.B. eine bestimmte natürliche oder juristische Person Gesellschafter ist.

Bei Aktiengesellschaften ist die Situation etwas anders. Zwar ist bei der Gründung noch anzugeben, wie viele Aktien jeder Gründer übernimmt (§ 23 Abs. 2 AktG), aber das kann sich später ändern. Aktien lauten grundsätzlich auf den Namen (§ 10 Abs. 1 Satz 1 AktG). In diesem Fall ergibt sich die aktuelle Liste der Aktionäre aus dem Aktienregister der Gesellschaft gemäß § 67 AktG. Allerdings ist diese Aktionärsliste nicht öffentlich.⁷⁵ An der Stelle von Namensaktien sind unter bestimmten Bedingungen

⁷¹ CbCR-Tabelle 1. Vgl. Lutz (2020), S. 38-45.

⁷² CbCR-Tabelle 2. Vgl. Lutz (2020), S. 55-58.

⁷³ CbCR-Tabelle 3. Vgl. Lutz (2020), S. 58-60.

⁷⁴ Eine entsprechende Auswertung von DAX-Konzernen nimmt Redeker (2020) vor.

⁷⁵ Jedoch müssen in das Transparenzregister nach § 18 GwG alle sog. „wirtschaftlich Berechtigten“ von juristischen Personen des Privatrechts eingetragen werden, soweit sich die geforderten Angaben (Name, Geburtsdatum, Wohnort, Art und Umfang des wirtschaftlichen Interesses; § 19 Abs. 1 GwG) nicht bereits aus einem anderen Register wie dem Unternehmensregister ergeben (§ 20 Abs. 2 GwG). Als wirtschaftlich Berechtigter gilt grundsätzlich, wer als natürliche Person direkt oder indirekt mehr als 25% der Kapitalanteile hält, mehr als 25 Prozent der Stimmrechte kontrolliert oder auf vergleichbare Weise Kontrolle ausübt (§ 3 GwG).

Inhaberaktien möglich, insbesondere natürlich bei Börsennotierung (§ 10 Abs. 1 Satz 2 AktG). In diesem Fall existiert kein Aktienregister, aus dem die Aktionäre hervorgehen.

Allerdings gibt es für Aktiengesellschaften eine dritte Quelle, aus der die einzelnen Aktionäre hervorgehen, zumindest wenn die Aktien an einem organisierten Markt gehandelt werden – die **Stimmrechtsmitteilungen nach WpHG**. Nach § 33 WpHG ist meldepflichtig, wer „durch Erwerb, Veräußerung oder auf sonstige Weise 3 Prozent, 5 Prozent, 10 Prozent, 15 Prozent, 20 Prozent, 25 Prozent, 30 Prozent, 50 Prozent oder 75 Prozent der Stimmrechte aus ihm gehörenden Aktien an einem Emittenten, für den die Bundesrepublik Deutschland der Herkunftsstaat ist, erreicht, überschreitet oder unterschreitet“. Die Meldung muss sowohl an die Aktiengesellschaft wie an die Bundesanstalt für Finanzdienstleistungsaufsicht erfolgen. Dies gilt auch für andere, ggf. potenzielle Erwerbe von Stimmrechten z.B. über Optionen (§§ 38, 39 WpHG). Diese Stimmrechtsmitteilungen sind nach § 40 WpHG an das Unternehmensregister zu übermitteln.

Wenig beachtet wird eine weitere vierte Datenquelle. Durch die Europäische Finanzmarktverordnung⁷⁶ (**Markets in Financial Instruments Regulation**, MiFIR) haben Wertpapierfirmen, die Geschäfte mit Finanzinstrumenten tätigen, der zuständigen Behörde die vollständigen und zutreffenden Einzelheiten dieser Geschäfte zu melden (Art. 26 Abs. 1 MiFIR). Die Meldung muss Angaben zur Identifizierung der Kunden enthalten, in deren Namen die Wertpapierfirma das Geschäft abgeschlossen hat (Art. 26 Abs. 3 Satz 1 MiFIR). Art. 26 Abs. 6 MiFIR führt weiter aus: „In Bezug auf die Angaben zur Identifizierung der Kunden gemäß den Absätzen 3 und 4 verwenden Wertpapierfirmen eine Kennung für Rechtsträger (Legal Entity Identifier), die zur Identifizierung von Kunden eingeführt wurde, bei denen es sich um juristische Personen handelt.“ Die ESMA wurde beauftragt entsprechende technische Regulierungsstandards u.a. für Legal Entity Identifiers (LEI) zu erarbeiten (Art. 26 Abs. 9 MiFIR).

Die weltweite Vergabe von LEI-Codes wird durch die Global Legal Entity Identifier Foundation (GLEIF) geregelt, dokumentiert und überwacht.⁷⁷ Nur von der GLEIF zugelassene Vergabe-Organisationen dürfen LEIs ausgeben. Unternehmen haben die Wahl, an welche Vergabe-Organisationen sie sich zur Registrierung einer LEI wenden.⁷⁸ Der 20-stellige alphanumerische LEI-Code beruht auf der ISO-Norm 17442. Er ermöglicht eine klare und eindeutige Identifikation der Rechtsträger, die an Finanztransaktionen beteiligt sind (sog. Level 1-Daten „Wer ist wer?“). Darüber hinaus – und das ist im vorliegenden Zusammenhang von besonderer Bedeutung – sind aber für jede LEI noch Informationen über die Eigentumsstruktur des Rechtsträgers hinterlegt (sog. Level 2-Daten „Wer gehört wem?“). Alle Daten sind über die GLEIF-Webseite kostenlos öffentlich zugänglich und können auch komplett heruntergeladen werden.⁷⁹ Hieraus lässt sich (selbstverständlich nur innerhalb der erfassten Bereiche) ein weiteres, globales Verzeichnis von Unternehmensbeteiligungen ableiten.

⁷⁶ Verordnung (EU) Nr. 600/2014 des Europäischen Parlaments und des Rates vom 15. Mai 2014.

⁷⁷ Vgl. zum Folgenden www.gleif.org/de.

⁷⁸ Eine von mehreren zertifizierten Vergabestellen in Deutschland ist etwa der Bundesanzeiger Verlag. Vgl. www.bundesanzeiger-verlag.de/evidenzwesen/leireg (3.3.2021).

⁷⁹ Vgl. www.gleif.org/de/lei-data/access-and-use-lei-data.

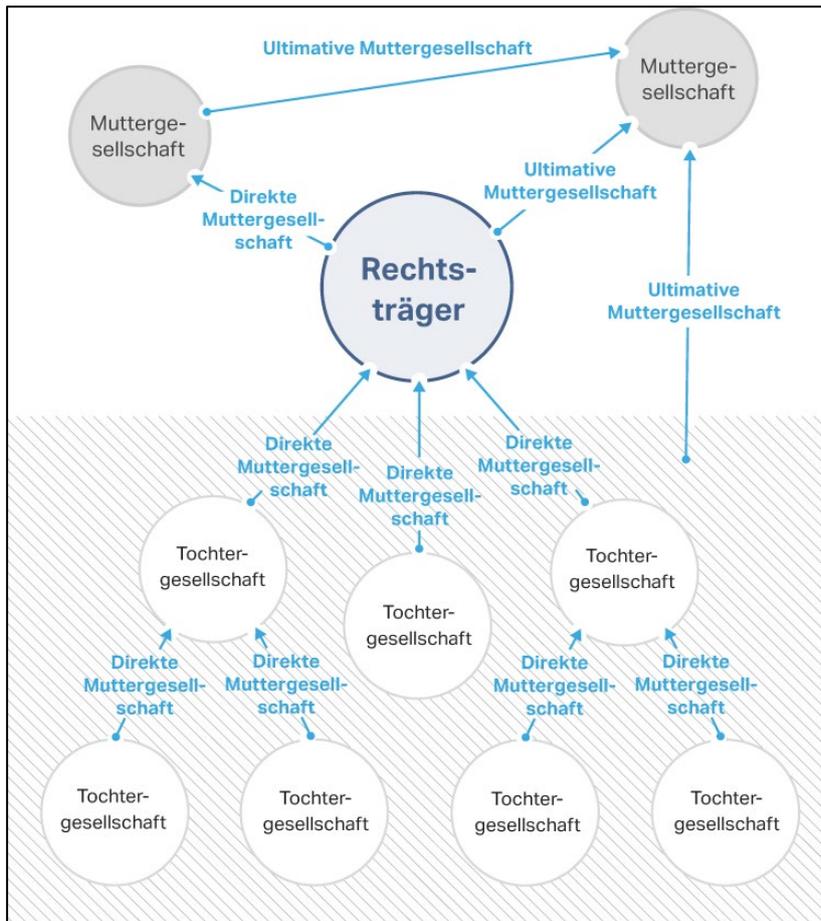


Abb. 2: Beteiligungsstrukturen aus LEI Level 2-Daten

Quelle: GLEIF, www.gleif.org/de/lei-data/access-and-use-lei-data/level-2-data-who-owns-whom (3.3.2021)

Insgesamt gibt es somit mehrere Ansatzpunkte, wie in unserem Beispiel die Liste der pro Land ansässigen bzw. belegenen Konzerneinheiten des CbCR durch Alternative Daten verprobt werden kann. Allerdings wäre eine rein manuelle Vorgehensweise in den allermeisten Fällen so aufwendig, dass dies nicht praktikabel ist.

3. Verbindungen zu Risikobereichen

a) Überblick

Der zweite konzeptionelle Ansatzpunkt geht von sog. „Red Flags“ aus. Hierunter versteht man in der Abschlussprüfung in der Regel Warnhinweise für Betrug (Fraud).⁸⁰ Andere Begriffe sind „Fraud Risk

⁸⁰ Vgl. Henselmann / Hofmann (2010), S. 268 ff.

Factors“ oder „Fraud Risk Indicators“. Die Prüfungsliteratur und Prüfungspraxis kennt umfangreiche Sammlungen möglicher Red Flags.⁸¹

Das Konzept der Red Flags lässt sich auch auf steuerliche Risiken übertragen.⁸² Als Ideenspender für Risikobereiche und zugehörige Red Flags können aber nicht nur die herkömmlichen Ansatzpunkte in Betriebsprüfungen dienen. Das BEPS-Projekt der OECD greift, wie oben beschrieben, verschiedene Bereiche auf. Weitere Anhaltspunkte lassen sich aus den Projekten zur Aufdeckung von Steuerhinterziehung, Betrug und Geldwäsche ableiten. Die Grenze zwischen Verdachtspunkten für eine Betriebsprüfung, für die Steuerfahndung und/oder Geldwäsche ist in einigen Bereichen sicher fließend.

Beim Einsatz Alternativer Daten als Red Flags geht es nicht um das Erbringen eines Beweises, dass Steuern verkürzt oder gar Straftaten begangen wurden, sondern „nur“ um die Entwicklung von Indizien für Verdachtsmomente, deren Auftreten den Einsatz der Betriebsprüfung oder anderer Ermittlungsformen triggern kann. Ob es zu näheren Ermittlungen kommt, ist in der Regel eine Ermessensentscheidung der zuständigen Finanzbehörde. Falls auf der Basis von Red Flags konkrete Ermittlungen angestoßen werden, müssten diese natürlich rechtssichere Belege für tatsächlich verwirklichte – und eventuell falsch deklarierte – Sachverhalte liefern.

Henselmann / Hofmann gruppieren ihre Liste mit verschiedenen „Fraud Red Flags“ in acht sachlich verschiedene Kategorien:⁸³ (1) Economic or industry operating conditions; (2) Entity operating conditions (business development, organizational structure); (3) Control structure (monitoring of management, internal control components); (4) Top management; (5) Relationship between management and auditor; (6) Accounting staff; (7) Journal entries; (8) Irrational ratios.

Während die Punkte (7) *Journal entries* und (8) *Irrational ratios* bereits konventioneller Weise bei der Planung von Betriebsprüfungen genutzt werden, kann man sich in den anderen Fällen fragen, ob eine Übertragung auf steuerliche Risiken möglich erscheint. Als Anknüpfungspunkte bieten sich an:

- die Art der Geschäftstätigkeit (analog zu (2) *Entity operating conditions – business development*);
- Standort und Rechtsform innerhalb der Konzernstruktur (analog zu (2) *Entity operating conditions – organizational structure*);
- innerhalb des Konzerns tätige Personen (analog zu (4) *Top management* sowie (6) *Accounting staff*);
- Beziehungen zu Personen außerhalb des Konzerns (analog zu (5) *Relationship between management and auditor*).

b) Art der Geschäftstätigkeit

Entity operating conditions deutet darauf hin, dass auch für steuerliche Risiken die Art der Geschäftstätigkeit (operations) eine Rolle spielen kann. Grundsätzlich anfällig sind **passive Betätigungen** etwa

⁸¹ Vgl. Appendix 1: Examples of Fraud Risk Factors, International Standard on Auditing (ISA) 240, The Auditor's Responsibilities Relating to Fraud in an Audit of Financial Statements.

⁸² Vgl. Henselmann / Haller (2018).

⁸³ Vgl. Henselmann / Hofmann (2010), S. 268 ff.

im Sinne von § 8 AStG, bei denen die Möglichkeiten zur Gewinnverschiebung besonders groß sind (z.B. Lizenzzahlungen, Verpachtung, Vermietung, Finanzierungen).

Als Red Flags können sicher auch **dubiose Betätigungen** gelten, die sich nicht selten in der Grauzone zur Legalität befinden. Beispiele hierfür sind Erotikbranche, Glücksspiel,⁸⁴ Waffenhandel, Betäubungsmittel.

Alternative Datenquellen, die Auskunft über Tätigkeiten des Unternehmens geben können, sind etwa eigene Webseiten, Nachrichten, Postings in sozialen Medien etc.

c) Standort und Rechtsform

Entity operating conditions umfassen zudem auch die Organisationsstruktur. Damit werden Rechtsformen und Standorte des Konzerngeflechts angesprochen. Unternehmenseinheiten (Gesellschaften und Betriebsstätten) mit Standort in **Niedrigsteuergebieten** bringen die besondere Gefahr von Gewinnverschiebungen ins Ausland mit sich. Es gibt verschiedene Listen von Niedrigsteuergebieten, sog. Steueroasen oder Tax Havens.

§ 2 Abs. 2 AStG definiert „niedrige Besteuerung“ für die Wegzugsbesteuerung⁸⁵ bei natürlichen Personen als (vereinfacht ausgedrückt) Einkommensteuerbelastung, die mehr als ein Drittel geringer ist als im Inland. Dabei spielt es keine Rolle, ob die Minderung wegen einem allgemein geringeren Steuerebene oder wegen einer eingeräumten Vorzugsbesteuerung besteht. Jedoch ist für die Besteuerung von Konzernen eher die Höhe der Körperschaftsteuer an ausländischen Standorten relevant. § 8 Abs. 3 AStG spricht bei Zwischengesellschaften von einer niedrigen Besteuerung „wenn die Einkünfte der ausländischen Gesellschaft einer Belastung durch Ertragsteuern von weniger als 25 Prozent unterliegen.“ Auch hier wäre zwischen der allgemeinen Besteuerung und Sonderregeln für bestimmte Betätigungen zu unterscheiden.

Eine Qualifikation als Steueroasen oder Tax Haven kann sich auch dadurch ergeben, dass aus manchen Ländern oder Jurisdiktionen keine oder nur **sehr eingeschränkte Informationen** über die dort ansässigen Rechtsträger und deren potentiell steuerlich relevanten Merkmale verfügbar sind.

Oxfam (2016b) unterscheidet insbesondere zwei Arten von Tax Havens:

⁸⁴ Siehe etwa § 16 GwG.

⁸⁵ Grundsätzlich beschrieben in Henselmann et al. (2005), S. 225, 502.

<i>Secrecy jurisdiction</i>	<i>Corporate tax haven</i>
Facilitating corruption, money laundering, and avoidance and evasion of taxes on the private wealth of individuals from other countries.	Facilitating avoidance and evasion of taxes on profits of multinationals generated by operations in other countries.
No effective exchange of financial account data or ownership data.	No corporate income tax or low overall corporate tax rate.
No information available about ultimate beneficial owners	Special corporate tax regimes resulting in non-taxation of certain profits or low effective tax rates.
Legislation allowing secretive trusts and other opaque financial structures.	No effective exchange or tax rulings, country-by-country data, or other corporate tax data.
Features of corporate tax havens.	Features of secrecy jurisdictions.

Abb. 3: Tax havens – a diversified industry

Quelle: Oxfam (2016b), S. 3

Aufgrund eines Beschlusses in 2016⁸⁶ führt die EU eine Liste „nicht kooperativer Länder und Gebiete“. Ziel ist es Druck auf Länder dieser Liste aufzubauen und sie zu (mehr) Kooperation zu bewegen, um Steuerbetrug, Steuervermeidung und Geldwäsche zu bekämpfen. Die aktuelle Liste vom Februar 2021 umfasst aber nur 12 Länder.⁸⁷ Insbesondere handelt es sich nur um Gebiete außerhalb der EU.

Oxfam’s Rangliste der Top 15 corporate tax havens umfasst hingegen auch Niederlande, Irland, Luxemburg sowie Zypern in der EU und daneben etwa die Schweiz und Hongkong.⁸⁸

Auch das *Tax Justice Network*,⁸⁹ ursprünglich gegründet als Initiative im Britischen Parlament,⁹⁰ veröffentlicht sowohl einen „Financial Secrecy Index“⁹¹ wie einen „Corporate Tax Haven Index“.⁹² Der Financial Secrecy Index 2020 listet auf den vorderen Plätzen nicht nur typische Kleinstaaten auf, sondern auch: 2. USA; 3. Schweiz; 4. Hongkong; 6. Luxemburg; 8. Niederlande; 12. United Kingdom. Enthalten sind ebenfalls die sog. Britischen Überseegebiete und Kronbesitzungen: 1. Cayman Islands; 9. British Virgin Islands; 11. Guernsey. Das Tax Justice Network führt dazu an: „If the UK and its network of Overseas Territories and Crown Dependencies were treated as a single entity, this UK spider’s web would rank first on the index.“⁹³ Verbindungen zwischen dem britischen Mutterland und Steueroasen

⁸⁶ Rat der Europäischen Union, 9452/16, FISC 85, ECOFIN 502, 25.5.2016.

⁸⁷ Rat der Europäischen Union, 2021/C 66/10, Amtsblatt der Europäischen Union C 66 vom 26.2.2021.

⁸⁸ Oxfam (2016a), S. 4.

⁸⁹ www.taxjustice.net.

⁹⁰ Vgl. Tax Justice Network (2021a).

⁹¹ fsi.taxjustice.net/en/.

⁹² www.corporatetaxhavenindex.org/en/.

⁹³ Tax Justice Network (2021b).

in Britischen Überseegebieten oder Kronbesitzungen schildert sehr anschaulich als „Spinnennetz“ *Shaxson*.⁹⁴

Auch innerhalb eines an sich unverdächtigen Staates können **bestimmte Gebiete** erhöhte Steuerrisiken mit sich bringen, wenn etwa die einzelnen Bundesstaaten oder Kantone eigene Rechtsform-schriften zum Steuer- oder Gesellschaftsrecht erlassen können. Ein bekanntes Beispiel ist der US-Bundesstaat Delaware.⁹⁵ Dort gibt es deutlich mehr Gesellschaften als Einwohner. Auf einige Einkunftsarten fallen keine Steuern an und Unternehmen können mit einem hohen Grad an Anonymität gegründet werden. Auch andere Bundesstaaten wie Nevada und Wyoming zeichnen sich entsprechend aus.⁹⁶ Hinzu kommt, dass die Vereinigten Staaten selbst den „Common Reporting Standard – CRS“ der OECD zum internationalen Datenaustausch nicht unterzeichnet haben, also auch auf diesem Weg keine Informationen weitergegeben werden.⁹⁷ Ähnlich berichtete für die Schweiz die Luzerner Zeitung über auffällig viele Firmen im Verhältnis zu den Einwohnern im Kanton Zug, davon viele ohne Angestellte.⁹⁸

Jedoch können auch **bestimmte konkrete Adressen** den Verdacht einer Briefkastenfirma nahelegen. Die schweizerische und österreichische Presse benennt Gebäude, die unglaublich vielen Firmen als Anschrift dienen. Beispielsweise beherbergt die „Bahnhofstraße 21, Zug, Schweiz“ immerhin 328 aktive Firmen.⁹⁹ Aber auch in Österreich findet man im „Eckhaus Börseplatz 4/Esslinggasse 2, Wien“ 124 aktive und 340 gelöschte Firmen.¹⁰⁰ Merkwürdig erscheint es neben der Häufung, wenn es an diesem Standort für die Firmen keinen Eintrag im Telefonverzeichnis gibt. Das spricht dann wohl wirklich für den namensgebenden Briefkasten.¹⁰¹

Alternative Daten sind hier somit Unternehmensregister und Telefonverzeichnisse.

d) Innerhalb des Konzerns tätige Personen

Die Punkte “(4) *Top management*” und “(6) *Accounting staff*” zeigen, dass die im Konzern tätigen Personen von hoher Relevanz sein können. Das gilt für steuerliche Risiken nicht nur in Hinblick auf eine möglicherweise **zweifelhafte Integrität** (Näheres dazu nachstehend unter e).¹⁰²

Verschiedene Teile einer Unternehmensgruppe lassen sich nämlich nicht nur über Eigentumsrechte zusammenhalten, sondern auch über Personen (Menschen), die in Personalunion mehrere leitende

⁹⁴ Vgl. Shaxson (2012).

⁹⁵ Vgl. Redeker (2020), S. 5. In dieser Studie wurde Delaware als einziger nicht eigenständiger Staat als Steuer-oase eingestuft.

⁹⁶ Vgl. Swanson, Ana (2016).

⁹⁷ Vgl. Bundeszentralamt für Steuern (2020).

⁹⁸ Vgl. Gwerder (2020).

⁹⁹ Vgl. Gwerder (2020).

¹⁰⁰ Vgl. OFR (2020).

¹⁰¹ Vgl. Gwerder (2020).

¹⁰² Vgl. Das Framework der COSO (2013) betont als Teil des Control Environment ein „commitment to integrity and ethical values.“ Vgl. auch Peemöller et al. (2020), S. 345, und Hofmann (2008), S. 417.

oder überwachende Tätigkeiten ausüben. Nicht von ungefähr statuiert § 290 Abs. 2 Nr. 2 HGB, dass auch das Recht, „die Mehrheit der Mitglieder des die Finanz- und Geschäftspolitik bestimmenden Verwaltungs-, Leitungs- oder Aufsichtsorgans zu bestellen oder abzurufen“ einen beherrschenden Einfluss begründet. Ähnliches gilt gemäß IFRS 10.B14 (b).

Unternehmensgruppen in Form eines Konzerns sind abgestufte Gebilde. Um einen engeren Kern aus Muttergesellschaft und Tochterunternehmen (§§ 290, 294 HGB, IFRS 10) gruppieren sich weniger eng eingebundene Einheiten wie Gemeinschaftsunternehmen (§ 310 HGB, IFRS 11), Assoziierte Unternehmen (§ 311 HGB, IAS 28), Beteiligungsunternehmen (§ 271 Abs. 1 HGB) und sonstige Unternehmensanteile.

Die Konzerngrenzen lassen sich dabei nicht scharf ziehen, denn es gibt vielfältige Ermessensspielräume.¹⁰³ Solche nicht konsolidierten Gebilde können ein höheres steuerliches Risiko auslösen, wenn in der Realität der eigentlich bei Konzernfremden vorhandene Interessensgegensatz nicht existiert, da die auf den ersten Blick „fremden“ Unternehmen doch in gewisser Weise „nahestehend“ sind, da sie **von Personen innerhalb des Konzerns gelenkt** werden.

Für die Beurteilung von Zweckgesellschaften (§ 290 Abs. 2 Nr. 4, IFRS 10) kommt es entscheidend auf die wirtschaftliche Tragung von Chancen und Risiken an.¹⁰⁴ Interessant könnten Gesellschaften sein, die für den Konzernabschluss nicht als konsolidierungspflichtig eingestuft wurden. Dieser Sachverhalt wurde möglicherweise vom Konzernabschlussprüfer (§ 316 Abs. 2 HGB) untersucht und in seinem Prüfungsbericht (§ 321 HGB) erwähnt. Die Besteuerung knüpft aber grundsätzlich an den Einzelabschluss an. Selbst wenn man davon ausgeht, dass der Betriebsprüfer bei seiner Planung möglicherweise bereits Einblick in den Prüfungsbericht zum Einzelabschluss nimmt, könnte der Prüfungsbericht zum Konzernabschluss weitere Hinweise für untersuchenswerte Bereiche geben. So etwa auf nicht konsolidierte Gesellschaften, bei denen dennoch vom Konzern beschäftigte Personen als Organ oder in leitender Stellung tätig sind.

Auch außerhalb des Konzernprüfungsberichts geben Unternehmensregister Auskunft über Organmitglieder oder andere Beschäftigte des Konzerns, die zusätzlich bei nicht-konzernzugehörigen Unternehmen tätig sind.

e) Beziehungen zu Personen außerhalb des Konzerns

Ferner können Red Flags auch Hinweise auf eventuelle Geschäftspartner außerhalb des Konzerns geben, die für dolose Handlungen als Komplizen fungieren können. Das können natürliche oder juristische Personen sein.

Konkrete Anknüpfungspunkte für Red Flags wären Geschäftspartner Person, **über die etwas Negatives vorliegt**:

- Personen, die in der Vergangenheit wegen Vergehen (Betrug, Bestechung, Steuerhinterziehung etc.) strafrechtlich verurteilt oder zivilrechtlich belangt wurden.

¹⁰³ Häufig auch als Special Purpose Entities (SPE), Special Purpose Vehicle (SPV), Special Purpose Company (SPC) sowie Variable Interest Entity (VIE) bezeichnet.

¹⁰⁴ Vgl. Henneberger (2012).

- Empfänger oder Leistender von Zahlungen als Person, für die der Verdacht der Geldwäsche besteht.¹⁰⁵
- Beziehungen zu Personen, die unter dem Verdacht stehen andere Straftaten zu begehen oder zu fördern.

Solche Alternative Daten könnten u.a. aus Eintragungen ins Bundeszentralregister¹⁰⁶ oder ins Gewerbezentralregister,¹⁰⁷ von nicht-steuerlichen Behörden, aus dem Datenaustausch mit anderen Staaten oder aus bekannt gewordenen „Steuerleaks“ stammen.

Ein Verdachtsmoment sind daneben auch Beziehungen zu **intransparenten** Geschäftspartnern, insbesondere solchen, die in *secrecy jurisdictions* ansässig sind (bei Unternehmen) oder arbeiten (bei natürlichen Personen). In solchen Fällen lässt sich auch der „wirtschaftlich Berechtigte“ im Sinne des Transparenzregisters nicht ermitteln.¹⁰⁸ Als wirtschaftlich Berechtigte im Sinne von § 3 GwG gelten allgemein natürliche Personen, (1) in deren Eigentum oder unter deren Kontrolle der Vertragspartner letztlich steht, oder (2) auf deren Veranlassung eine Transaktion letztlich durchgeführt oder eine Geschäftsbeziehung letztlich begründet wird. Dazu zählt z.B. bei Kapitalgesellschaften, wer unmittelbar oder mittelbar mehr als 25 Prozent der Kapitalanteile hält, mehr als 25 Prozent der Stimmrechte kontrolliert oder auf vergleichbare Weise Kontrolle ausübt.

Ein gesonderter Punkt sind gegebenenfalls Beziehungen zu **politisch exponierten Personen** (z.B. aktuelle oder ehemalige Minister, hohe Verwaltungsbeamte, Richter und deren nahe Angehörige).¹⁰⁹ Deren Beschäftigung oder Beauftragung hat möglicherweise nicht den Grund, ihr hervorragendes Fachwissen zu nutzen, sondern dient eventuell einem (legalen) Lobbyismus¹¹⁰ oder einer (illegalen) Beeinflussung ihrer Entscheidungen zu Gunsten des Unternehmens.¹¹¹ Insofern sieht auch § 10 Abs.1 Nr. 4 GwG die Prüfung vor, ob es sich „bei dem Vertragspartner oder dem wirtschaftlich Berechtigten um eine politisch exponierte Person, um ein Familienmitglied oder um eine bekanntermaßen nahe-stehende Person handelt“. Hierfür gelten verstärkte Sorgfaltspflichten (§ 15 Abs. 3 Nr. 3 GwG).

¹⁰⁵ Siehe generell das Gesetz über das Aufspüren von Gewinnen aus schweren Straftaten (Geldwäschegesetz – GwG).

¹⁰⁶ Das Gesetz über das Zentralregister und das Erziehungsregister (BRZG) regelt in Deutschland die Führung dieses Registers für natürliche Personen durch das Bundesamt für Justiz (§ 1 BRZG). Neben Straftaten enthält es u.a. die Berechtigung zum Waffenbesitz und Jagderlaubnis, aber auch das Untersagen einer Berufsausübung wegen Unzuverlässigkeit, Ungeeignetheit oder Unwürdigkeit. Finanzbehörden erhalten nach aktuellem Stand (nur) Auskunft für die Verfolgung von Straftaten, die zu ihrer Zuständigkeit gehören (§ 41 Abs. 1 Nr. 4 BRZG).

¹⁰⁷ Nach § 149 GewO werden in das deutsche Gewerbezentralregister u.a. folgende Dinge eingetragen: Untersagung der Ausübung eines Gewerbes wegen Unzuverlässigkeit oder Ungeeignetheit, Bußgelder über 200 Euro im Zusammenhang mit der Ausübung eines Gewerbes oder wegen Steuerordnungswidrigkeiten, bestimmte Verurteilungen wegen Straftaten.

¹⁰⁸ Gemäß § 18 Abs. 1 GwG ist das Transparenzregister ein Register zur Erfassung und Zugänglichmachung von Angaben über den wirtschaftlich Berechtigten. Das Transparenzregister ist grundsätzlich öffentlich (§ 23 Abs. 1 Nr.3 GwG).

¹⁰⁹ In Deutschland definiert nach § 1 Abs. 12 und 13 GwG.

¹¹⁰ Beispielsweise ist der ehemalige Bundeskanzler Gerhard Schröder u.a. Mitglied des Aufsichtsrats des russischen Ölkonzerns Rosneft, vgl. Lobbypedia (2021).

¹¹¹ Peemöller et al. (2020), S. 310, weisen auf negative Wirkungen einer „Filzkultur“ in prominenten Betrugsfällen hin.

4. Anwendungsbeispiel

Diese exemplarische Liste von "Red Flags" klingt in ihren Auswirkungen erst einmal überschaubar. Das täuscht jedoch, denn ausgehend von einem einzelnen Unternehmen lassen sich Netzwerke von Verbindungen zu anderen Unternehmen oder Vermögenswerten bilden.¹¹²

Die konzeptionelle Idee soll an einem **einfachen fiktiven Beispiel** verdeutlicht werden:¹¹³

- Die XYZ-AG hat verschiedene Tochterunternehmen, u.a. die X-GmbH und die Y-GmbH.
- Die X-GmbH ist Gesellschafterin der E GmbH & Co. KG., also eines Enkelunternehmens der XYZ-AG.
- Geschäftsführerin der E GmbH & Co. KG ist Erika Musterfrau.
- Erika Musterfrau ist zugleich Allein-Geschäftsführerin der ABC-GmbH, welche nicht in den Konzernkreis der XYZ-AG einbezogen wurde.
- Die ABC-GmbH ist Lieferant der E GmbH & Co. KG.
- Die E GmbH & Co. KG und die ABC-GmbH haben außerdem die gleiche Adresse, „Hauptstraße 15, 12345 Irgendwo“.
- Die ABC-GmbH hat Zahlungen erhalten, für die der Verdacht von Geldwäsche vorliegt. Die Zahlungen liefen über eine Bank in New York und stammen von einer Person NN aus Brasilien.
- In Brasilien existiert ein Unternehmen, dessen Firmenname den Kurznamen der Konzernspitze XYZ enthält, nämlich die „XYZ Brasil Land & Building Ltd.“ (nicht in den Konzernkreis der XYZ-AG einbezogen).
- Die Y-GmbH besitzt einen 19,9% Anteil an der Z Ltd. mit Sitz in Zypern, also unterhalb der normalen 20% Beteiligungsschwelle.
- Mehrheitsgesellschafterin der Z-Ltd. ist die Kapitalgesellschaft P Ltd. in Panama.
- Max Mustermann als Mit-Geschäftsführer der Y-GmbH hat auf seinem Facebook-Account ein Foto in einem Restaurant in Panama City gepostet. Es zeigt ihn zusammen mit mehreren Herren im Anzug.
- Einer der Herren wird als Pedro Fulano identifiziert, ein bekannter Rechtsanwalt aus Panama.
- Pedro Fulano arbeitet in Panama in der Kanzlei „Fulano, Mengano & Zutano“, von der Steuerleaks bekannt geworden sind, die u.a. Aufschluss über höchstwahrscheinlich illegale, betrügerische Aktivitäten geben.
- Max Mustermann ist zugleich Allein-Geschäftsführer einer Kapitalgesellschaft mit Sitz in Dublin, Irland, der Atlantic Royals Ltd. Die Atlantic Royals Ltd. betreibt selbst kein operatives Geschäft. Der (einzige) Wohnort von Max Mustermann ist München.

¹¹² Vgl. Endt et al. (2019), von Daniels / Wörpel (2019).

¹¹³ Ein Beispiel zu Fraud Detection findet sich in Hodler / Needham (2021), S. 25ff.

Stellt man die oben genannten Sachverhalte bildlich dar (siehe Abb. 4), so ergibt sich durch Verketten untereinander ein Netz. Hierbei bilden Unternehmen, Personen, Adressen und Zahlungen die Knoten. Sie werden untereinander durch Pfeile verbunden, die Beziehungen ausdrücken. Hierzu gehören u.a. auch Eigentumsverhältnisse (besitzt_Mehrheit, besitzt_Anteil), Management-Tätigkeiten (ist_GF), Aktivitäten (reist_nach), persönliche Beziehungen (bekannt_mit) oder andere Merkmale (Name_ähnlich, hat_Tätigkeit). Dieses Netz ist mathematisch ausgedrückt ein „Graph“.

Zum Konzernkreis konsolidierter Unternehmen gehören ursprünglich die XYZ-AG (Mutter), die X-GmbH, Y-GmbH und E GmbH & Co. KG (Töchter bzw. Enkel). Sie sind in der Abb. 4 grau eingefärbt.

Im Beispiel erscheinen jedoch mehrere Verdachtsmomente naheliegend. Die ABC-GmbH residiert an derselben Adresse wie die E GmbH & Co. KG, hat dieselbe Geschäftsführerin und ist zugleich Kundin der E GmbH & Co. KG. Möglicherweise fehlt hier der Interessengegensatz und die Preise wurden entsprechend verzerrt (überhöht?) festgelegt.

Die ABC-GmbH hat aus der Lieferung Zahlungen von der E GmbH & Co. KG erhalten. Selbst leitet sie Gelder an eine Person in Brasilien weiter. Für diese Zahlungen wurden aus dem Ausland Anhaltspunkte für einen Verdacht auf Geldwäsche gemeldet. In Brasilien selbst gibt es die XYZ Brasil Land & Building Ltd., bei der eine teilweise Namensgleichheit zur XYZ-AG besteht, obwohl sie im Konzernabschluss nicht auftritt.

Die Y-GmbH besitzt Anteile an der Z Ltd. in Zypern. Mehrheitsgesellschafter der Z Ltd. ist jedoch die P Ltd. aus Panama, über die selbst nichts weiter bekannt ist. Jedoch trifft sich der Geschäftsführer der Y-GmbH in Panama anscheinend aus geschäftlichem Anlass mit dem Rechtsanwalt einer panamaischen Kanzlei, die im dringenden Verdacht krimineller Aktivitäten steht. Es könnte sich bei der P Ltd. somit auch um eine Scheinfirma handeln, hinter der in Wirklichkeit der XYZ-Konzern steht.¹¹⁴

Derselbe Geschäftsführer der Y-GmbH ist zugleich Geschäftsführer der Atlantic Royals Ltd. aus Dublin in Irland. Mit der Verwaltung von Rechten an immateriellen Wirtschaftsgütern geht sie einer passiven Tätigkeit nach. Auffällig erscheint eine geschäftsführende Tätigkeit in zwei so weit voneinander entfernten Städten, vor allem wenn es sich um Allein-Geschäftsführer handelt. An dieser Stelle ist noch offen, ob die Atlantic Royals Ltd. Geschäftsbeziehungen zum XYZ-Konzern unterhält. Dann wären gegebenenfalls die Lizenzraten zu prüfen.

¹¹⁴ Zu Überschneidungen von Registerdaten mit Datenlecks aus Steueroasen vgl. Endt et al. (2019).

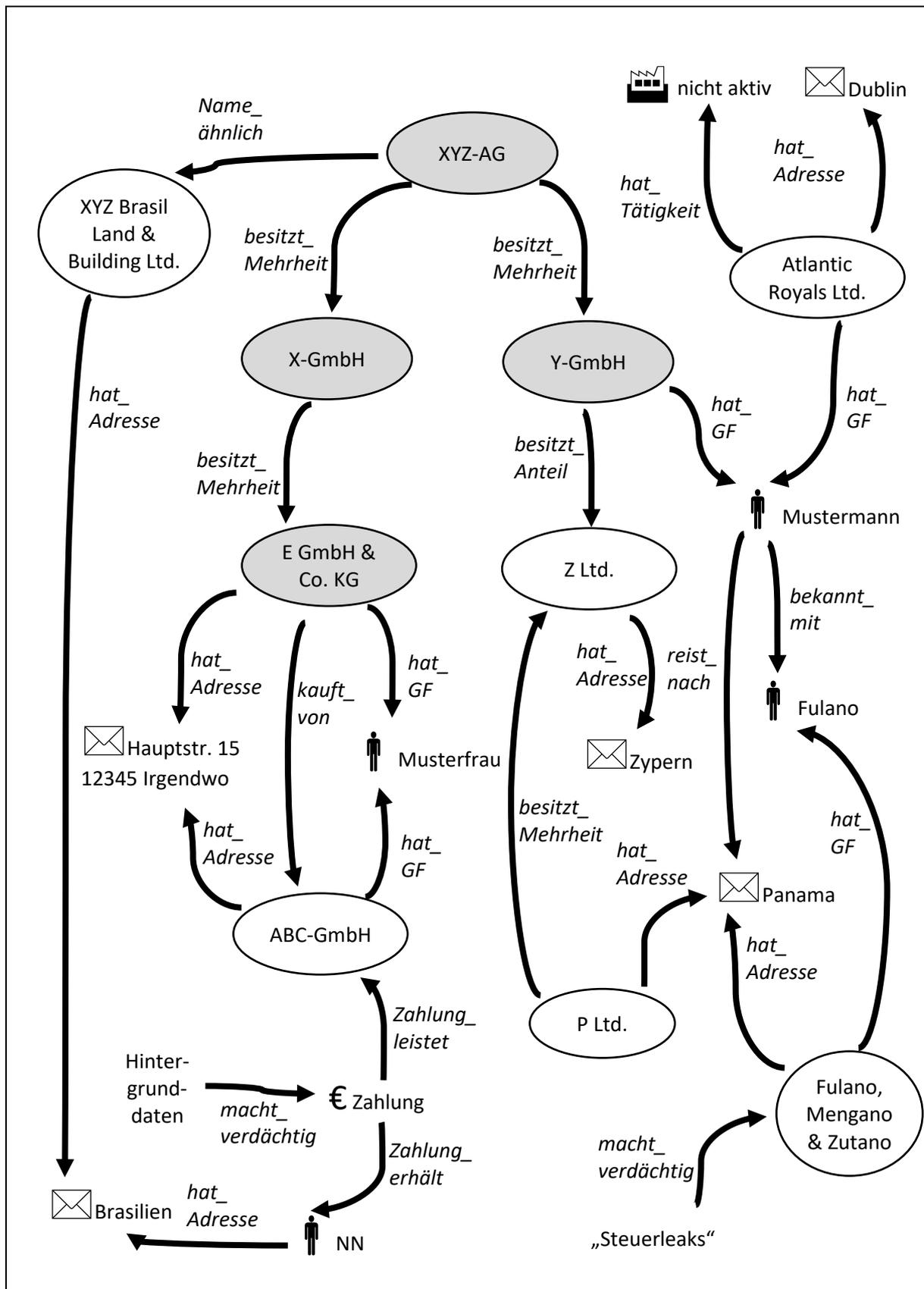


Abb. 4: Daten des Beispiels in Netzwerkdarstellung
Eigene Darstellung

5. Schlussfolgerungen

Die obigen Ausführungen haben Wege skizziert, wie Alternative Daten Hinweise für Betriebsprüfungen liefern können. Im Zentrum stehen als Ausgangspunkt die den Finanzbehörden bekannten Daten, insbesondere aus der Veranlagung, aber auch aus Meldepflichten für steuerliche Gestaltungen sowie aus dem Country-by-Country Reporting.

Durch stufenweise Expansion werden Verbindungen mit Alternativen Daten hergestellt, unter anderem über Unternehmensanteile, Personen, Zahlungsströme oder Adressen. Aus diesen Verbindungen entstehen entweder selbst Verdachtsmomente oder sie weisen auf Daten hin, deren Konsistenz mit den steuerlichen Unterlagen geprüft werden kann.

„Zufallsfunde“ aus Steuerleaks oder durch steuerliche „Wistleblower“ können in dieses System mit eingebaut werden (siehe Abb. 5).¹¹⁵

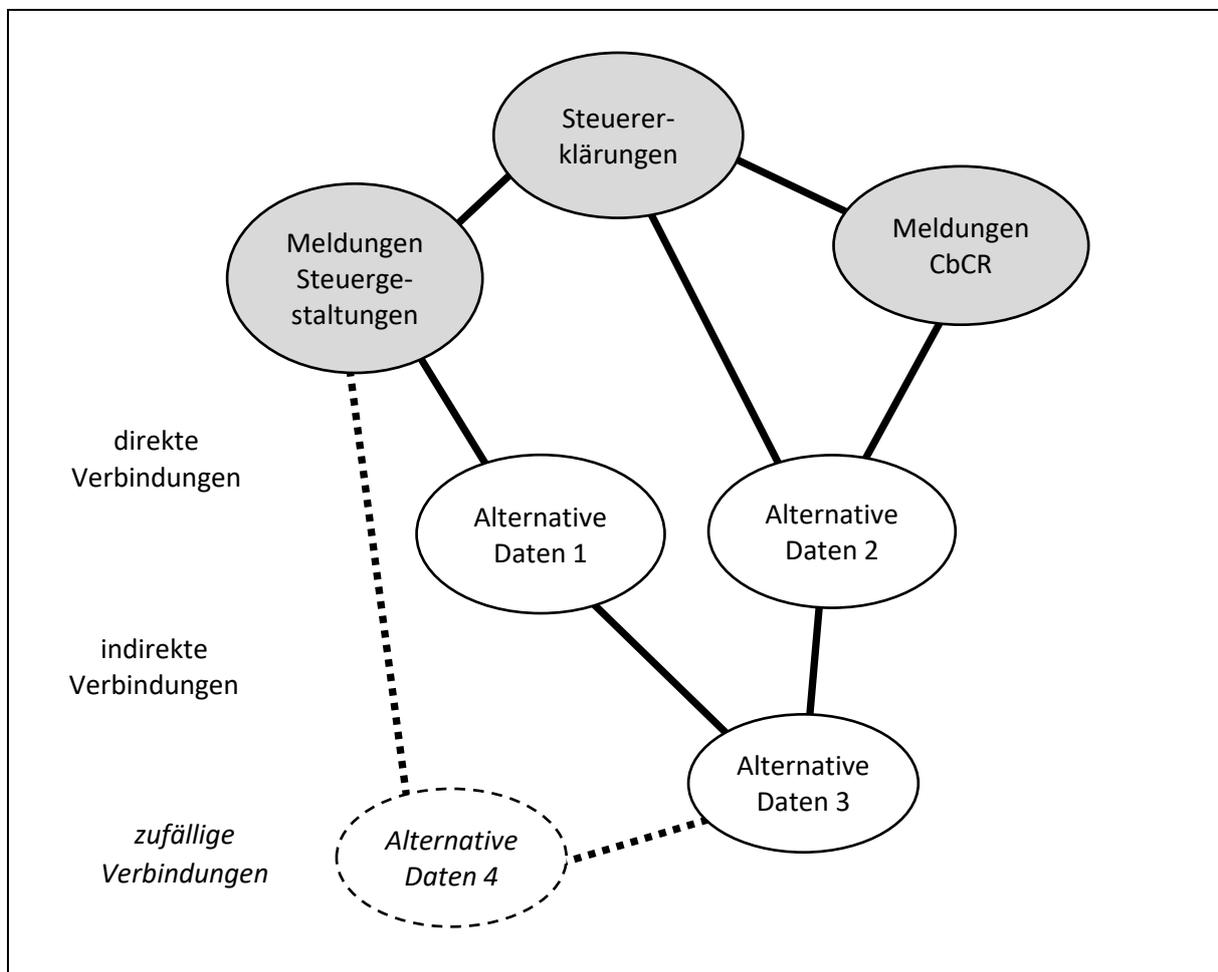


Abb. 5: Verknüpfungen der Datenbestände
Eigene Darstellung

¹¹⁵ Vgl. von Daniels / Wörpel (2019).

Selbstverständlich bedeutet das Auftreten von „Red Flags“ nicht automatisch, dass hier wirklich eine Steuerverkürzung vorliegt.¹¹⁶ Es handelt sich erst einmal nur um eine Auffälligkeit, für die zu entscheiden ist, ob man ihr weiter nachgehen will und bei der dann das Ergebnis offen ist. In vielen Fällen werden sich einzelne Anzeichen als harmlos herausstellen:

- Beispielsweise waren an einer bestimmten Adresse in der Schweiz verdächtig viele Unternehmen gemeldet. Ursache hierfür war die Tatsache, dass diese Anschrift Sitz eines Schifffahrtsunternehmens war. Für jedes der Schiffe wurde eine eigene Gesellschaft gegründet, deren Sitz dann auch an derselben Adresse lag.¹¹⁷
- Auch personelle Verbindungen können triftige außersteuerliche Gründe haben. So gibt es beispielsweise Vorstände, die bei anderen (nicht verbundenen) Konzernen Mitglied des Aufsichtsrats sind, weil dort ihre Sachkompetenz gefragt ist.

Es ist offensichtlich, dass die Chancen, eine zutreffende „Nadel im Heuhaufen“ zu finden, umso höher ist, je umfassender und umfangreicher die Alternativen Datensammlungen sind. Das steigert auch die Möglichkeiten, einzeln auftretende „Red Flags“ zu ignorieren und sich etwa auf Fälle zu konzentrieren, bei denen sich verschiedene Verdachtsmomente kumulieren.

Daher wird es kaum möglich sein die Datensammlung und Erstverarbeitung manuell vorzunehmen. Manuelle Eingriffe sollten nur nötig sein, um anhand der vom System gesammelten Anhaltspunkte diejenigen Fälle herauszugreifen, welche in der Betriebsprüfung weiterverfolgt werden sollen. Die Analyse potenzieller Alternativer Datenquellen im folgenden Kapitel beschränkt sich daher rein auf digital verfügbare Daten. Letztere ermöglichen automatisierte Verarbeitungsschritte.¹¹⁸ Gewisse Unschärfen in den Daten sind tragbar,¹¹⁹ da wegen des bloßen Verdachtscharakters keine 100%ige Präzision und Verlässlichkeit des Datenmaterials erforderlich ist.

¹¹⁶ In Hinblick auf Bilanzdelikte vgl. Peemöller et al. (2020), S. 323.

¹¹⁷ Vgl. Gwerder (2020).

¹¹⁸ Zu früheren Ansätzen einer wissensbasierten Fraud-Erkennung etwa durch Expertensysteme vgl. Hofmann (2008), S. 418, mit weiteren Nachweisen.

¹¹⁹ Beispielsweise können mehrere Personen denselben Namen haben. Umgekehrt lässt sich der Name einer Person auf mehrere Weisen darstellen.

III. Alternative Daten

A. Systematisierungsansätze

In Abschnitt II. C. 1 wurden bereits Beispiele von Alternativen Daten genannt. Es gibt mehrere Möglichkeiten zur Klassifikation solcher Daten.

Nach dem **Entstehungsprozess** kann man unterscheiden, ob die Daten generiert wurden von Individuen (z.B. Posts in sozialen Netzwerken, Rezensionen auf Amazon, Suchanfragen bei Google, Downloads von SEC-Filings), durch Geschäftsprozesse (z.B. Einkauf von Waren, Käufe oder Verkäufe von Aktien, Veröffentlichung von Konzernabschlüssen, Firmengründungen, Gesellschafterwechsel) oder von Sensoren (z.B. Geo-Bewegungsprofile, Wetterdaten). Jedoch ist diese Unterscheidung für die praktische Anwendung in der Betriebsprüfung von begrenztem Wert.

Relevanter ist die Einteilung nach den **Bezugsmöglichkeiten**:

- Primärdaten werden von der Steuerverwaltung (ggf. in deren Auftrag) selbst erhoben, beispielsweise automatisiert mit Hilfe von Webscraping¹²⁰ oder aber auch durch manuelle Recherchen in Registern und auf sonstigen Webseiten des Internets.
- Bei Sekundärdaten greift man auf Daten zurück, die andere bereits erhoben und i.d.R. aufbereitet haben. Dieser andere „Anbieter“ kann sich auf bestimmte eng begrenzte Inhalte beschränken (sog. Point Vendor) oder aber ein Portal / ein Verzeichnis / einen Marktplatz für viele Daten unterschiedlicher Datenlieferanten bieten (sog. Platform Vendors).¹²¹

Wichtig sind immer die mit den Daten verbundenen **Kosten**.

- Die eigene Erhebung von Daten bindet i.d.R. erhebliche Arbeitszeit. Zusätzlich ist bei maschinellen Recherchen ein gewisses technisches Know-how erforderlich, das gegebenenfalls zunächst durch Schulungen erlangt werden muss.
- Sekundärdaten von kommerziellen Anbietern müssen hingegen entgeltlich erworben werden (Private Data). Die Anforderungen an das Know-how und die benötigte Arbeitszeit sind deutlich geringer.
- Bestimmte Sekundärquellen dürfen aber auch generell oder unter bestimmten Bedingungen kostenlos genutzt werden (sog. Open Data). Dazu zählen beispielsweise Daten öffentlicher Stellen oder durch gemeinnützige Organisationen. Spezialfälle sind Open Government Data (Verwaltung) und Open Science (Wissenschaft).¹²²

Damit ist bereits die Frage der **rechtlichen Nutzbarkeit** angesprochen. Für kommerzielle Anbieter richten sich die Nutzungsmöglichkeiten nach dem zugrundeliegenden Vertrag mit individuellen, maßgeschneiderten Lizenzbedingungen. Im Unterschied hierzu sind „Open Data“ idealerweise Daten, die ohne Einschränkungen von jedem für jeden Zweck sowohl genutzt wie auch weiterverarbeitet und

¹²⁰ Eine gängige Python-Bibliothek ist Scrapy. Siehe <https://scrapy.org/>.

¹²¹ Vgl. Monk et al. (2019). Siehe etwa <https://alternativedata.org>.

¹²² Vgl. Dietrich (2011).

weiterverbreitet werden können.¹²³ Das hängt im Detail von der rechtlichen Ausgestaltung der Lizenzbedingungen ab. Am bekanntesten und verbreitetsten sind die Standardlizenzen von Creative Commons, auch CC genannt.¹²⁴ Jede Creative-Commons-Lizenz ist weltweit gültig und gilt so lange, wie der Schutz des Urheberrechts andauert.¹²⁵

(1)	 <p>Attribution CC BY</p>	<p><i>This license lets others distribute, remix, adapt, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.</i></p>
(2)	 <p>Attribution-ShareAlike CC BY-SA</p>	<p><i>This license lets others remix, adapt, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. ... All new works based on yours will carry the same license, so any derivatives will also allow commercial use.</i></p>
(3)	 <p>Attribution-NoDerivs CC BY-ND</p>	<p><i>This license lets others reuse the work for any purpose, including commercially; however, it cannot be shared with others in adapted form, and credit must be provided to you.</i></p>
(4)	 <p>Attribution-NonCommercial CC BY-NC</p>	<p><i>This license lets others remix, adapt, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.</i></p>
(5)	 <p>Attribution-NonCommercial-ShareAlike CC BY-NC-SA</p>	<p><i>This license lets others remix, adapt, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.</i></p>
(6)	 <p>Attribution-NonCommercial-NoDerivs CC BY-NC-ND</p>	<p><i>This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.</i></p>

Abb. 6: Creative-Commons-Lizenzen

Eigene Abbildung - Logos und Beschreibungstexte von Creative Commons (2021a).

¹²³ Vgl. Dietrich (2011).

¹²⁴ Vgl. Creative Commons (2021b).

¹²⁵ Vgl. Creative Commons (2021a).

Für diese Varianten gilt in Hinblick auf den Einsatz zur Planung von Betriebsprüfungen:

Zu (1) Immer erforderlich ist die Namensnennung des Urhebers (Lizenzgebers), die sog. Attribution, abgekürzt „BY“. Das ist eine nur minimale Einschränkung, die lediglich die Anerkennung der Leistung des Urhebers verlangt. Diese Lizenz genügt vollständig allen Open Data Grundsätzen.

Zu (2) Eine für die Betriebsprüfung unerhebliche Einschränkung wäre, dass die Weitergabe von Daten nur unter denselben Bedingungen erfolgen darf („Share Alike“, SA).

Zu (3) Ebenfalls für die Betriebsprüfung unschädlich wäre das Verbot, die Daten in modifizierter Form zu veröffentlichen („No Derivatives“, ND), da sowieso keine Veröffentlichung beabsichtigt ist.

Zu (4) Finanzverwaltung (einschließlich der Betriebsprüfung und Steuerfahndung) sind hoheitliche, öffentlich-rechtliche Aufgaben und somit nicht-kommerzieller Natur. Daher sind auch Daten, die ausschließlich zu nicht-kommerziellen Zwecken eingesetzt werden dürfen („Non Commercial“, NC), nutzbar.

Zu (5) und (6) Gleiches gilt für die Verbindungen der Anforderungen „Non-Commercial“ mit „Share Alike“ bzw. „No Derivatives“.

Allerdings verhindern oder erschweren in der Praxis **nicht nur rechtliche Grenzen** die Nutzung von Datenbeständen. Die Open Knowledge Foundation schreibt dazu:¹²⁶ *„Open knowledge’ is any content, information or data that people are free to use, re-use and redistribute — without any legal, technological or social restriction. ... The key features of openness are:*

- *Availability and access: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.*
- *Reuse and redistribution: the data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets. The data must be machine-readable.*
- *Universal participation: everyone must be able to use, reuse and redistribute — there should be no discrimination against fields of endeavour or against persons or groups. For example, ‘non-commercial’ restrictions that would prevent ‘commercial’ use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.”*

Auch wenn Open Data grundsätzlich für den Nutzer kostenlos sein sollen, ist es möglich eine kleine Vergütung zu verlangen (siehe oben), welche die Kosten der Bereitstellung der Daten abdeckt. Das fördert die Bereitschaft, auch umfangreichere Datensammlungen zur Verfügung zu stellen. Der Nutzerkreis soll weiterhin universell sein, d.h. ein Ausschluss kommerzieller Zwecke (etwa gemäß der Creative-Commons-Lizenz „CC BY-NC“) oder eine Beschränkung auf Lehr- und Forschungszwecke würde dem entgegenstehen.

Die obige Definition der Open Knowledge Foundation spricht daneben mehrere technische Aspekte an. Erstens stellt sich die Frage nach dem **Volumen**, also ob man nur einzelne konkrete Datensätze herunterladen kann oder ob ein Massendownload des Gesamtdatenbestandes möglich ist. Ersteres gilt beispielsweise für das deutsche Unternehmensregister.¹²⁷ Es enthält u.a. die Jahresabschlüsse

¹²⁶ Open Knowledge Foundation (2021).

¹²⁷ Vgl. www.unternehmensregister.de.

von Kapitalgesellschaften. Jedoch muss das konkrete Unternehmen ausgewählt werden. Eine Sicherheitsabfrage nach Buchstaben/Zahlen, die auf einem Bild enthalten sind, soll den Einsatz von Webscraping-Software verhindern. Damit sind letztlich nur einzelne Abschlüsse kostenlos verfügbar. Ganz im Gegensatz hierzu steht das britische System des „Free Accounts Data Product“ durch das Companies House.¹²⁸ Hier ist es möglich, als zip-Datei alle Jahresabschlüsse zusammen herunterzuladen, die in einem bestimmten Monat veröffentlicht wurden. Durch 12 Downloads erhält man so den gesamten Datenbestand des Jahres.

Zweitens verlangt die Open Knowledge Foundation eine **maschinenlesbare** Fassung der Daten. Dieses Kriterium verletzen natürlich Daten, die nur in Papierform vorliegen. Aber auch von Papierdokumenten angefertigte Scans sind zunächst nur ein Bild (Foto), ohne dass die darin enthaltenen Zeichen des Textes – Buchstaben und Zahlen – für einen Computer sofort ersichtlich sind. Als Aufbereitungsschritt wäre zwar der Einsatz einer OCR-Software denkbar, die den gescannten Text zu erkennen sucht. Jedoch ist in der Regel keine absolut fehlerfreie Texterkennung möglich. Daher sollte es sich um eine Form von Textdateien handeln.

Die Weiterverarbeitung wird erleichtert, wenn die Daten in einer **strukturierten Form** vorliegen. Unstrukturiert wären beispielsweise die Jahresabschlüsse als pdf oder als docx (Microsoft Word), bei denen Abschlusszahlen im Text vorkommen und deren Bedeutung sich nur aus dem Kontext (Beschriftungen etc.) ergibt. Strukturiert wären hingegen Abschlussdaten in tabellarischer Form (csv oder xlsx Microsoft Excel), wenn die Spaltenköpfe und Zeilen den Inhalt eindeutig identifizieren. Weitere strukturierte Erscheinungsformen sind xml oder json-Daten, deren Bedeutung mit Hilfe von Taxonomien festgelegt werden kann sowie RDF-Daten, die auf Ontologien beruhen.¹²⁹

Tim Berners-Lee, der Erfinder des World Wide Web, hat ein „**Fünf-Sterne-Modell**“ vorgeschlagen, um den Grad der Offenheit von Daten zu kategorisieren. Die Anforderungen steigen dabei immer mehr.¹³⁰ Zusätzliche Bedingungen für höhere Stufen sind die Nutzung öffentlicher (nicht firmeneigener) Datenformate, die Nutzung von Internet-Standards zur eindeutigen Identifikation sowie zuletzt die Verlinkung mit anderen Datenbeständen:

★	<i>Available on the web (whatever format) but with an open licence, to be Open Data</i>
★★	<i>Available as machine-readable structured data (e.g. excel instead of image scan of a table)</i>
★★★	<i>as ★★ plus non-proprietary format (e.g. CSV instead of excel)</i>
★★★★	<i>All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff</i>
★★★★★	<i>All the above, plus: Link your data to other people's data to provide context</i>

Abb. 7: Fünf-Sterne-Modell nach Berners-Lee

Quelle: Berners-Lee (2009)

¹²⁸ Vgl. http://download.companieshouse.gov.uk/en_accountsdata.html.

¹²⁹ Näheres hierzu in Kapitel IV.

¹³⁰ Vgl. Dietrich (2011).

Die folgende Tabelle fasst die wesentlichen Kriterien zusammen:

Bezugsquelle	Primärdaten der Finanzverwaltung	Sekundärdaten von anderen	
Kosten	Erhebung (z.B. Webscraping)	Kaufpreis (kommerziell)	kostenlos (Open Data)
Nutzbarkeit	voll	individuelle vertrag- liche Bedingungen	Lizenzbedingungen insbesondere CC
Bezugsmenge	einzelne Datensätze		Massendownload
Maschinenlesbarkeit	keine (Papier)	Bitmap-Daten (z.B. Image-Scans von Dokumenten)	Text-Daten
Strukturiertheit	unstrukturiert (z.B. Fließtext, Foto)		strukturiert (z.B. Tabellen, XML, JSON)
Formatnormierung	proprietär, durch Unternehmen (z.B. Excel)	nicht-proprietär / offen, durch Gremien	
Web-Standard	(nein)	nein (z.B. CSV, XML)	ja (durch W3C) RDF

Abb. 8: Kriterien für Alternative Daten

Eigene Abbildung

B. Nützliche Alternative Daten?

1. Vorbemerkungen

Es folgt eine Bestandsaufnahme Alternativer Datenquellen, die nach Meinung der Verfasser grundsätzlich besonders geeignet sind, Entscheidungen über eine Betriebsprüfung (Außenprüfung) - ergänzend zu den bereits intern vorliegenden steuerlichen Daten - zu unterstützen.

Dabei handelt es sich um eine subjektive Auswahl, die in keinem Falle abschließend zu verstehen ist. Die Auswahl stellt auch eine Momentaufnahme dar, da sich die Datenbestände laufend verändern und somit dynamisch sind. Ändern können sich beispielsweise Datenumfang (in der Regel Zunahme), Qualität, Aktualität, Nutzungsbedingungen oder Kosten.

Der Schwerpunkt wird auf Offene Daten (Open Data) gelegt, aber teilweise erfolgen auch Verweise auf private/kommerzielle Quellen, da hier Inhalte angeboten werden, die kaum oder nur äußerst aufwändig nachzubilden wären.

2. Unternehmen und Organe

a) Unternehmensregister

Ein naheliegender Startpunkt ist ein Abgleich der internen steuerlichen Daten mit den Inhalten von Unternehmensregistern.

Für **Deutschland** ist der Einstieg über die Seiten www.unternehmensregister.de oder www.bundesanzeiger.de möglich. Er beinhaltet unter einer gemeinsamen Oberfläche sowohl das Handels-, Genossenschafts- und Partnerschaftsregister (Eintragungen, eingereichte Dokumente, Registerbekanntmachungen), aber auch Rechnungslegung, Kapitalmarktinformationen im Zusammenhang mit Wertpapieren sowie insolvenzgerichtliche Bekanntmachungen.

Obwohl die Führung von Registern eine hoheitliche Aufgabe ist, wurde diese Aufgabe in Deutschland an die privatwirtschaftliche „Bundesanzeiger Verlag GmbH“ übertragen. Diese ist aufgrund einer Privatisierungsentscheidung seit 2006 im Eigentum der Verlagsgruppe M. Dumont Schauberg.¹³¹ Das ist u.E. eine äußerst fragwürdige Gestaltung, denn ein kostenfreier Zugriff auf die Daten wird nur im nötigen Minimum gewährt. Komfortablere und umfassendere Abfragen sind nur gegen Bezahlung möglich.

Die Bundesanzeiger Verlag GmbH schreibt dazu auf ihrer Webseite in der Rubrik „Kosten der Nutzung“:¹³² „Für die Recherche nach einzelnen Firmen und die Einsicht in Veröffentlichungen und die Unternehmensträgerdaten (UT) entstehen ... keine Kosten. Für jeden Abruf der zu einer Registernummer

¹³¹ https://de.wikipedia.org/wiki/Bundesanzeiger_Verlag.

¹³² <https://www.unternehmensregister.de/ureg/howto1.8.html>. Es handelt sich um Beträge von typischerweise 4,50€ pro Unternehmen. Für alle Mitgliedstaaten der EU gilt, dass Gebühren für die Ausstellung einer vollständigen oder auszugsweisen Kopie die Verwaltungskosten nicht übersteigen dürfen (Artikel 3 Nr. 4 der Richtlinie 2009/101/EG).

angebotenen Daten (Aktueller Abdruck, Chronologischer Abdruck, Historischer Abdruck, Dokumentenabruf) entsteht jeweils eine Gebühr ... Die Höhe der Abrufgebühren richtet sich nach dem Justizverwaltungskostengesetz.“

Die offiziellen Webportale der anderen **Mitgliedstaaten der Europäischen Union** finden sich auf:

https://e-justice.europa.eu/content_business_registers_in_member_states-106-en.do

Die Inhalte variieren teilweise je nach Land, aber immer sind Firma, Rechtsform, Sitz, Kapital und gesetzliche Vertreter (Organe) vorhanden. Auch für diese Staaten gilt, dass gewisse Basisinformationen kostenfrei sind und weitergehende Auskünfte in der Regel bezahlt werden müssen. Die Navigation ist natürlich immer in der Landessprache und teilweise auch in weiteren Sprachen, z.B. Englisch.

Das „**European Business Register**“ (EBR)¹³³ ist eine freiwillige Kooperation von Unternehmensregistern in Europa, erstreckt sich aber nicht nur auf EU-Mitgliedsstaaten. Mitglieder sind u.a. auch das United Kingdom (UK), Gibraltar, die Kanalinseln Jersey und Guernsey, die Isle of Man, Liechtenstein sowie Georgien und Aserbaidschan. Die Daten stammen direkt von den teilnehmenden offiziellen nationalen Registerstellen.¹³⁴

Auf den Seiten der European Business Registry Association findet sich ebenfalls eine Übersicht von **Unternehmensregistern weiterer Staaten**.¹³⁵ Deren Inhalte richten sich nach den nationalen Vorschriften. Auch führt beispielsweise der Link für die USA auf die Webseite der „U.S. Securities and Exchange Commission“ (SEC), die (nur) über Wertpapier-emittierende Unternehmen berichtet.¹³⁶ Die eigentlichen Unternehmensregister werden dort von den Bundesstaaten geführt.

Weitere Links auf Unternehmensregister anderer Länder – ggf. auch deren innerstaatliche Teilregister nach Landesteilen – findet man auf der englischsprachigen Wikipedia-Seite „List_of_official_business_registers“.¹³⁷ Leider ist diese Seite sehr unübersichtlich und enthält auch Verweise auf Portale, die keine öffentlich-rechtlichen Unternehmensregister darstellen wie z.B. kommerzielle Datensammlungen / Auskunfteien.¹³⁸

b) Identifizierung und Verknüpfung

Ein Problem der Verknüpfung von Daten aus verschiedenen Quellen besteht in der **eindeutigen Identifizierung der beteiligten Unternehmen**, etwa wenn Unternehmen A Anteile am Unternehmen B besitzt. Zu einem Unternehmen B* sind möglicherweise bereits weiterführende Informationen ge-

¹³³ <https://ebra.be/> (European Business Registry Association).

¹³⁴ <https://ebra.be/information-distributors/>.

¹³⁵ <https://ebra.be/worldwide-registers/>.

¹³⁶ <https://www.sec.gov/>.

¹³⁷ https://en.wikipedia.org/wiki/List_of_official_business_registers.

¹³⁸ Vgl. Kaya / Seebeck (2019).

speichert. Um die Identität von B und B* zu prüfen, müsste idealerweise ein Datenquellen-übergreifender Schlüssel vorliegen.¹³⁹ Das stellt allerdings ein Problem dar. Beispielsweise ist die Adidas AG beim Amtsgericht Fürth unter der Nummer „HRB 3868“ registriert. Dieselbe Nummer „HRB 3868“ weisen aber auch u.a. die Niehaus Immobilien GmbH am Amtsgericht Freiburg oder die Herzberg Autolackiererei Verwaltungsgesellschaft mbH am Amtsgericht Kaiserslautern auf. Bereits bundesweit müsste also die Registernummer um die Angabe des Registergerichts ergänzt werden. In anderen Ländern gelten vollkommen andere Systematiken. Ideal wären weltweit gültige Schlüssel.

Für dieses Problem gibt es mehrere partielle Lösungsansätze.

Als erstes ist der bereits oben in Kapitel II. C. 2. angesprochene **Legal Entity Identifier (LEI)** zu nennen.¹⁴⁰ Der LEI-Code ist global eindeutig. Das Datenformat wird durch die ISO-Norm 17442 definiert und verlangt strukturierte XML-Daten, welche sich sehr gut maschinell verarbeiten lassen. Unter der LEI werden u.a. folgende Informationen gespeichert:¹⁴¹ offizielle Firma der juristischen Person, Adresse der Hauptniederlassung, Adresse bei Gründung, Verweis auf eine Kennung im Handelsregister (soweit vorhanden). Der LEI ist dadurch mit wesentlichen Referenzdaten verknüpft, die eine Identifikation der Rechtsträger erlauben.¹⁴² Diese sog. „Level 1“-Daten beantworten somit die Frage „Wer ist wer?“. Zusätzlich gibt es jedoch noch die „Level 2“-Daten, die Antworten auf eine für die Planung der Betriebsprüfung auch sehr wichtige Frage geben: „Wer gehört wem?“ (relationship data). Unternehmen, welche einen LEI besitzen, nennen ihre „direkte buchhalterisch übergeordnete Muttergesellschaft“ und ihre „ultimative buchhalterisch übergeordnete Muttergesellschaft“.¹⁴³ Der komplette Datenbestand ist öffentlich zugänglich und kann tagesaktuell als Komplettdatenbestand heruntergeladen werden (ca. 300 MB).¹⁴⁴ Er stellt ein globales Verzeichnis von Unternehmen dar. Allerdings beschränkt sich der Kreis der Unternehmen, die eine LEI benötigen, auf juristische Personen, die Finanzinstrumente ausgeben, kaufen oder verkaufen.

Die sog. **PermID** stellt einen weiteren Ansatz für globale Unternehmensschlüssel dar.¹⁴⁵ Sie wird von dem Unternehmen Refinitiv – ehemals Thomson Reuters – propagiert. Refinitiv ist ein sehr großer

¹³⁹ Für natürliche Personen in Deutschland ist dies die steuerliche Identifikationsnummer (IdNr). Vgl. Bundeszentralamt für Steuern, https://www.bzst.de/DE/Privatpersonen/SteuerlicheIdentifikationsnummer/steuerlicheidentifikationsnummer_node.html.

¹⁴⁰ Vgl. www.gleif.org.

¹⁴¹ Vgl. GLEIF (2017), S. 15.

¹⁴² Vgl. <https://www.gleif.org/de/about-lei/common-data-file-format#>.

¹⁴³ Vgl. <https://www.gleif.org/de/about-lei/common-data-file-format/current-versions/relationship-record-cdf-format#>.

¹⁴⁴ Download-Seite: <https://www.gleif.org/de/lei-data/gleif-concatenated-file/download-the-concatenated-file>.

¹⁴⁵ <https://permid.org/>.

privater Anbieter von Finanzinformationen.¹⁴⁶ Im Unterschied zu sonstigen Datenbeständen wird für die PermID jedoch eine kostenlose und offene Lizenz gewährt.¹⁴⁷ Refinitiv schreibt hierzu:¹⁴⁸

„Refinitiv Permanent Identifier (PermID) is a machine readable identifier that provides a unique reference for data item. Unlike most identifiers, PermID provides comprehensive identification across a wide variety of entity types including organizations, instruments, funds, issuers and people. PermID never changes and is unambiguous, making it ideal as a reference identifier. PermID is also in the center of Refinitiv's own information model and knowledge graph.

PermID is intended to enable interoperability. As a part of our open strategy, we provide users with a set of descriptive metadata for each entity to facilitate disambiguation to PermID with or without explicit mapping. We also provide a set of tools for working with the PermID and our Information Model, which we will continue to update, build and administer.“

Die PermID ist eindeutig und die Daten liegen in einem strukturierten Format. Sie bilden auch einen Linked Data Graph. Dieser enthält neben den Unternehmensdaten im engeren Sinne („Organization“) auch Daten zu Officers & Directors als natürlichen Personen. Ein Massendownload (bulk) ist möglich.¹⁴⁹

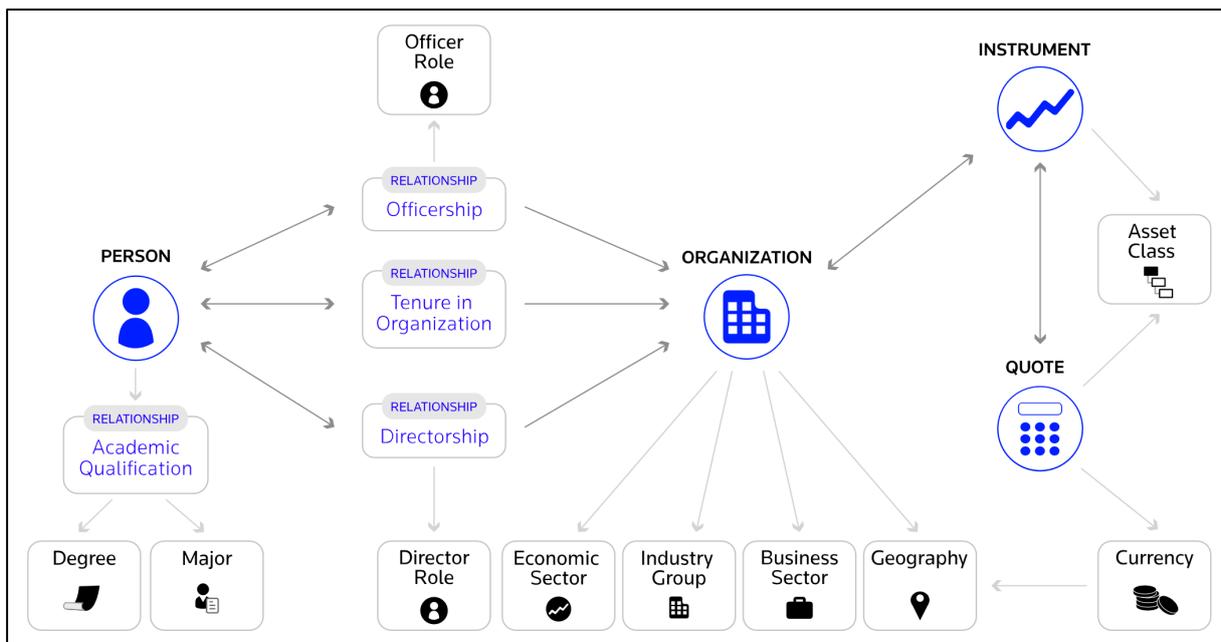


Abb. 9: Linked Data Graph

Quelle: Refinitiv, <https://permid.org/faq#entitiesSupported>

¹⁴⁶ Vgl. <https://www.refinitiv.com/en/about-us>, <https://permid.org/about>.

¹⁴⁷ <https://permid.org/faq>: „The PermID database is licensed under the Creative Commons with Attribution license, version 4.0 (CC-BY). An extended set of fields is also available under the Creative Commons Non-Commercial license (CC-NC 3.0).“

¹⁴⁸ <https://permid.org/>.

¹⁴⁹ <https://permid.org/faq>: „Bulk download enables the retrieval of Refinitiv's entity data in .gz files, one per entity. Currently we support Turtle and Ntriples file formats.“

c) Aufbereitete Daten

Die oben beschriebenen Einschränkungen in der Nutzung von Registerdaten (oft kein Massendownload, originär fehlende Verbindungen zwischen den einzelnen Objekten) haben zu kommerziellen und nicht-kommerziellen Lösungen geführt.

Es gibt eine Vielzahl von privaten Unternehmen, welche **auf kommerzieller Basis** Daten von Unternehmen und deren Organen sammeln, aufbereiten und anbieten. Als Beispiel für nützliche Aufbereitungen soll die North Data GmbH (www.northdata.de) gelten:

„North Data analysiert Pflichtveröffentlichungen europäischer Firmen und viele andere Quellen, um Wirtschaftsinformationen zu gewinnen, insbesondere zu finanziellen Kennzahlen und zu Zusammenhängen zwischen Firmen untereinander sowie zu Personen. ... Die so gewonnenen Wirtschaftsinformationen werden aufgearbeitet, um sie zu vernetzen, übersichtlich darzustellen und interaktiv zu visualisieren. Sie können sie auf dieser Website online recherchieren.“¹⁵⁰

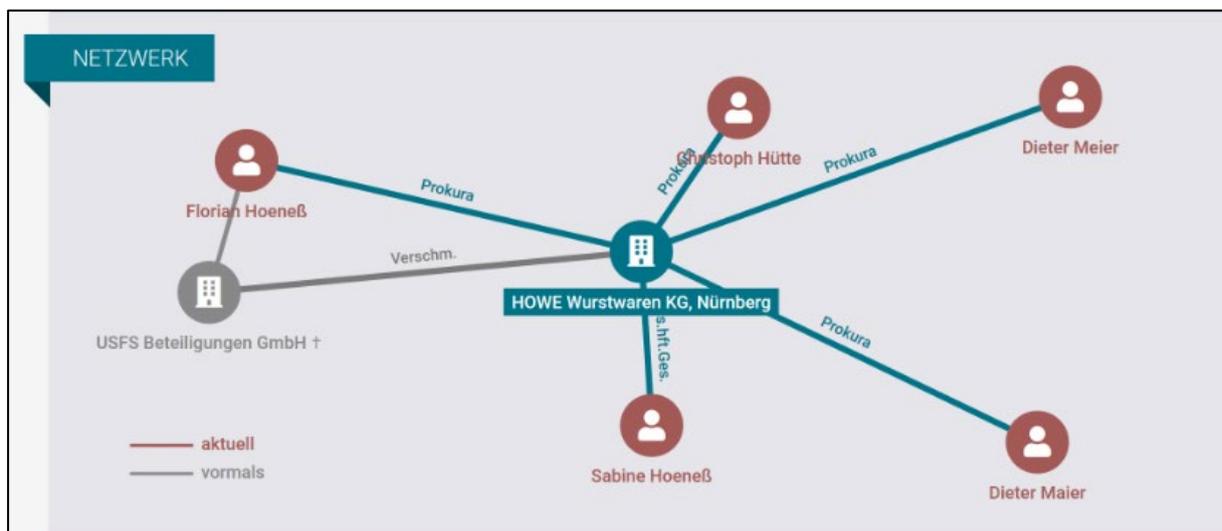


Abb. 10: Beispiel für ein Netzwerk von Unternehmen und Personen

Quelle: North Data, <https://www.northdata.de>

Grunddaten können kostenlos von North Data über ein Webinterface recherchiert werden. Zusätzliche Daten sowie ein Zugriff per Software über ein API sind kostenpflichtig.¹⁵¹

Im Gegensatz dazu stellt **offeneregister.de**¹⁵² ein **nicht-kommerzielles Projekt** des „Open Knowledge Foundation Deutschland e.V.“¹⁵³ dar. Ausgangspunkt ist die Kritik an der Tatsache, dass die Unternehmensdaten auf der offiziellen Webseite des Handelsregisters zwar öffentlich eingesehen werden, aber nur sehr eingeschränkt genutzt werden können. Ziel ist eine Bereitstellung dieser Daten ohne

¹⁵⁰ https://www.northdata.de/_about.

¹⁵¹ Vgl. https://www.northdata.de/_premium.

¹⁵² Vgl. <https://offeneregister.de/>.

¹⁵³ <https://okfn.de/>.

Gebühren und Nutzungseinschränkungen.¹⁵⁴ Die Datensammlung für mehr als 5 Mio. Unternehmen beruht auf den Handelsregisterbekanntmachungen und zu einem geringen Maße den Handelsregistern selbst. Wegen rechtlicher Einschränkungen konnten die Unternehmensbekanntmachungen des Bundesanzeigers sowie Daten aus dem Unternehmensregister nicht genutzt werden.¹⁵⁵ Gemeinsam mit OpenCorporates.com¹⁵⁶ werden die Informationen als Open Data veröffentlicht. Die nachstehende Abbildung zeigt einen Vergleich mit dem offiziellen Handelsregister:

	Handelsregister.de	Open Data
alles durchsuchen	Nein	<input checked="" type="checkbox"/> Ja, z.B. auch nach Personen
Gesamt-Download	Nein	<input checked="" type="checkbox"/> ja, z.B. hier
Freie Weiterverwendung	Nein, in Nutzungsbedingungen ausgeschlossen	<input checked="" type="checkbox"/> ja, unter freier Lizenz
Programmierschnittstelle	Nein	<input checked="" type="checkbox"/> vorhanden, z.B. hier oder hier
Stabile Links auf Unternehmensinfo	Nein	<input checked="" type="checkbox"/> ja, z.B. Open Knowledge Foundation Deutschland e.V.
Stabile und offene Identifikatoren	Nein	<input checked="" type="checkbox"/> wird möglich

Abb. 11: Was ermöglicht Open Data?

Quelle: Open Knowledge Foundation Deutschland e.V., <https://offeneregister.de/>

Damit trägt offeneregister.de auch zur **weltweiten Datenbank opencorporates.com** mit über 190 Mio. Unternehmen aus mehr als 140 Jurisdiktionen bei:¹⁵⁷

„01 Our Purpose

Total corporate transparency is a critical requirement for a fairer society. To ensure that everyone knows exactly who they are working with – and working for. To tackle corruption and criminality. To protect our democracy. To create a trusted business environment we want to work in – and a society we’d all like to live in.

02 Our Data

¹⁵⁴ Es gilt die Creative Commons Lizenz CC BY 4.0 (d.h. Attribution, also lediglich Nennung “by ...”).

¹⁵⁵ Vgl. <https://offeneregister.de/daten/>.

¹⁵⁶ <https://opencorporates.com/>.

¹⁵⁷ Vgl. <https://opencorporates.com/info/our-data/>.

As the largest open database of companies in the world, our business is making high-quality, official company data openly available. Data that can be trusted, accessed, analysed and interrogated when and how it's needed. Data that the world needs.”¹⁵⁸

Rechtlich gesehen wird OpenCorporates von der OpenCorporates Ltd. mit Sitz in London betrieben.¹⁵⁹ Jedoch werden als Unternehmenszweck (mission) Transparenz und freier Zugang zu den Daten vorgegeben und die Einhaltung durch die Mitglieder des „OpenCorporates Trust“ überwacht.¹⁶⁰

Ein Datenzugriff ist möglich über:¹⁶¹

- die Webseite (<https://opencorporates.com/>);
- die Programmschnittstelle OpenCorporates API (<https://api.opencorporates.com/>);
- als Massendownload (bulk).

Kostenlos sind generell Abfragen über die Webseite sowie für „Public Benefit Projects“.¹⁶² Kommerzielle Nutzer können je nach Nutzungsvolumen gestaffelte API-Pläne kaufen oder ganze Massendatenbestände (bulk data).¹⁶³ Zur Identifizierung der Unternehmen werden die offiziellen Schlüssel (Identifiers) der jeweiligen Primärquellen verwendet.

3. Gesellschafter und wirtschaftlich Berechtigte

Wie bereits oben in Abschnitt II. C. 2. beschrieben, bieten sich – außerhalb der den Finanzbehörden bereits vorliegenden steuerlichen Unterlagen – folgende Wege zur Ermittlung der Gesellschafter (Miteigentümer) von Unternehmen an:

- Unternehmensregister (in Deutschland § 29 HGB, § 106 HGB, § 162 HGB, § 8 GmbHG)
- Aktienregister von Aktiengesellschaften (in Deutschland § 67 AktG)
- Anhangsangaben in Einzel- und Konzernabschlüssen (in Deutschland § 285 HGB, § 313 HGB, IFRS 12)
- Stimmrechtsmitteilungen bei Aktiengesellschaften (in Deutschland § 33 WpHG)
- LEI Level-2 Daten

Damit wird zwar eine Reihe von auf Eigentumsverhältnissen beruhenden Verbindungen zwischen Unternehmen sowie zwischen Privatpersonen und Unternehmen offengelegt, es verbleiben aber den-

¹⁵⁸ <https://opencorporates.com/info/about/>.

¹⁵⁹ Company Registration Number 07444723 in England and Wales.

¹⁶⁰ Vgl. <https://opencorporates.com/info/governance/>.

¹⁶¹ Vgl. https://opencorporates.com/legal/public_records_privacy_policy.

¹⁶² Beispielsweise investigative Journalisten, NGOs, öffentliche Forschung, möglicherweise auch staatliche Stellen zur Verfolgung von Straftaten.

¹⁶³ Vgl. <https://opencorporates.com/info/our-data/>.

noch erhebliche Lücken. Insbesondere ist häufig nicht klar, welche natürlichen Personen letzten Endes Nutznießer der Erträge des Unternehmens sind, weil mehrere andere Organisationsstufen zwischengeschaltet sind.

Zentral für diese Frage ist der Begriff des „**wirtschaftlich Berechtigten**“. Als „wirtschaftlich Berechtigter“ im Sinne des Geldwäschegesetzes gilt nach § 3 Abs. 1 GwG grundsätzlich

- die natürliche Person, in deren Eigentum oder unter deren Kontrolle eine juristische Person, sonstige Gesellschaft oder eine Rechtsgestaltung wie Stiftungen, Treuhandverhältnisse o.ä. letztlich steht, oder
- die natürliche Person, auf deren Veranlassung eine Transaktion letztlich durchgeführt oder eine Geschäftsbeziehung letztlich begründet wird.

Als konkrete Fälle nennt § 3 Abs. 2-4 GwG insbesondere unmittelbare oder mittelbare Stimmrechts- oder Kapitalanteile von mehr als 25%; Treugeber, Verwalter, Vorstände oder Begünstigte von Stiftungen o.ä.; sowie Möglichkeiten in vergleichbarer Weise Kontrolle oder beherrschenden Einfluss auszuüben.

Solche Lücken soll seit 2017 das **Transparenzregister**¹⁶⁴ des Geldwäschegesetzes auffüllen. Nach § 18 Abs. 2 GwG wird ein Register zur Erfassung und Zugänglichmachung von Angaben über den wirtschaftlich Berechtigten (Transparenzregister) eingerichtet. Gemäß § 26 Abs. 2 GwG wird das deutsche Transparenzregister mit den Registern anderer Mitgliedstaaten der Europäischen Union im Sinne von Artikel 22 Absatz 2 der Richtlinie (EU) 2017/1132 über die durch Artikel 22 Absatz 1 der Richtlinie (EU) 2017/1132 geschaffene zentrale Europäische Plattform vernetzt.

Das Transparenzregister ist mittlerweile (nach Registrierung) frei öffentlich zugänglich. Dafür entstehen mit dem Abruf Gebühren.¹⁶⁵ Allerdings müssen zahlreiche Behörden – u.a. die Strafverfolgungsbehörden, das Bundeszentralamt für Steuern sowie die örtlichen Finanzbehörden – für ihre Einsichtnahmen nichts bezahlen.¹⁶⁶

Eine wichtige Änderung erfolgte durch das am 10.6.2021 vom Bundestag verabschiedete Transparenzregister- und Finanzinformationsgesetz (TraFinG).¹⁶⁷ Bislang war das Transparenzregister ein sog. Auffangregister, d.h. per „Eintragungsfiktion“ mussten Unternehmen keine Meldung zum Transparenzregister vornehmen, wenn die in das Transparenzregister einzutragenden Angaben zum wirtschaftlich Berechtigten schon den anderen öffentlichen Registern (z. B. Handelsregister) zu entnehmen waren. Seit dem 1.8.2021 ist das Transparenzregister jedoch ein Vollregister; die Eintragungsfiktion gilt nicht mehr.¹⁶⁸

¹⁶⁴ <https://www.transparenzregister.de>.

¹⁶⁵ § 24 Abs. 1, 3 GwG in Verbindung mit § 1 Transparenzregistergebührenverordnung (TrGebV) in Verbindung mit Nr. 2 Anlage 1 TrGebV.

¹⁶⁶ §§ 23 Abs. 1 Nr. 1, 24 Abs. 2 Satz 3 GwG.

¹⁶⁷ Vgl. Beck (2021).

¹⁶⁸ Für den Vollzug der dadurch zu ergänzenden Eintragungen gelten Übergangsfristen.

4. Unternehmensveröffentlichungen

Neben den Stammdaten wie Firma, Rechtsform, Sitz, gesetzliche Vertreter sind auch **Unternehmensveröffentlichungen in der Form von Finanzberichten** (Jahresabschlüsse, Konzernabschlüsse u.a.) von Bedeutung. Natürlich sind nicht alle Unternehmen publizitätspflichtig. Die Finanzberichte sind je nach Land bereits im Unternehmensregister enthalten oder sie müssen in einem anderen Portal aufgerufen werden.

Den Finanzberichten lassen sich potenziell allerlei nützliche Daten entnehmen:

„Allerdings gibt es in den Jahresberichten noch viel Informationen, die wir noch nicht auswerten, also zur Konzernstruktur, zum Aufsichtsrat, zur Vergütung, und vieles mehr. Das wird alles früher oder später kommen, ergänzt durch attraktive Visualisierungen.“¹⁶⁹

Im Falle **kapitalmarktorientierter Unternehmen** ist die Finanzberichterstattung noch deutlich umfangreicher und aussagekräftiger. Sie haben mehr Berichtspflichten in Form von periodischen Zwischenberichten wie Halbjahres- oder Quartalsberichten sowie einer ereignisbezogenen Berichterstattung wie Ad-hoc-Meldungen,¹⁷⁰ Directors' Dealings¹⁷¹ oder Stimmrechtsmitteilungen.¹⁷²

Diese Stimmrechtsmitteilungen können unübersichtlich werden. Allerdings führt die Bundesanstalt für Finanzdienstleistungsaufsicht eine konsolidierte Datenbank.¹⁷³ Die Angaben zu den Stimmrechtsanteilen beruhen auf den von den Emittenten veröffentlichten Stimmrechtsmitteilungen der Meldepflichtigen. Die Veröffentlichungen nach § 40 WpHG sind im elektronischen Unternehmensregister (www.unternehmensregister.de) abrufbar. Das Datum der Veröffentlichung wird in der Spalte „Veröffentlichung gemäß § 40 WpHG“ angegeben.

Vorteilhaft wirkt sich noch aus, dass auch im Ausland die Börsenzulassungsvorschriften häufig (je nach dem Kapitalmarktsegment, in dem das Unternehmen notiert ist) Berichte in englischer Sprache verlangen. Aber auch wenn dies nicht vorgeschrieben ist, würden börsennotierte Unternehmen Informationen in englischer Sprache veröffentlichen, um den Informationsbedarf internationaler Investoren zu decken. Daher unterhalten sie auf ihren Unternehmenswebseiten eine Rubrik "Investor Relations".

Berichte von kapitalmarktorientierten Unternehmen müssen auch zeitnah erstattet werden. Der Zugang zu diesen Informationen ist außerdem ohne Kosten möglich.

Auf **freiwilliger Basis** veröffentlichen kleine und große Unternehmen weitere Dokumente. Zu diesen gehören Pressemitteilungen, aber auch Imagebroschüren, Standorte/Niederlassungen, Interviews/Reden etc. Viele davon sind auf den Webseiten der Unternehmen verfügbar. Mit Methoden des **Webscraping** können solche Inhalte gesammelt werden.

¹⁶⁹ Interview mit dem Gründer von Northdata, Frank Felix Debatim, in Seebach (2018).

¹⁷⁰ Sog. Veröffentlichung von Insiderinformationen, Art. 17 Verordnung (EU) Nr. 596/2014 (Marktmissbrauchsverordnung).

¹⁷¹ Sog. Eigengeschäfte von Führungskräften, Art. 19 Verordnung (EU) Nr. 596/2014 (Marktmissbrauchsverordnung).

¹⁷² In Deutschland nach § 33 WpHG.

¹⁷³ Vgl. https://www.bafin.de/DE/PublikationenDaten/Datenbanken/Stimmrechte/stimmrechte_node.html.

5. Steuerleaks

Steuerleaks sind „unfreiwillig“ bekannt gewordene Informationen über „graue“ oder (potentiell) illegale Aktivitäten von Unternehmen oder natürlichen Personen. Auf Wikipedia findet sich eine Liste von Leaks zu Steuerdaten.¹⁷⁴ Hierzu gehören¹⁷⁵ die sog.

- Offshore-Leaks
- Luxemburg-Leaks
- Swiss-Leaks
- Panama Papers¹⁷⁶
- Bahamas-Leaks
- Malta Files
- Paradise Papers
- OpenLux
- Pandora Papers¹⁷⁷

In den meisten Fällen wurde der Datensatz an das International Consortium of Investigative Journalists (ICIJ)¹⁷⁸ weitergegeben. Dort wurden sie in die „**ICIJ Offshore Leaks Database**“ eingespeist. Die Datenbank ist für die Öffentlichkeit verfügbar:

„The ICIJ Offshore Leaks Database is licensed under the Open Database License and its contents under Creative Commons Attribution-ShareAlike license. Always cite the International Consortium of Investigative Journalists when using this data.

This database is powered by Neo4j, a graph database that structures data in nodes (the icons you see in the visualization) and relationships (the links between nodes). To make this data easily accessible to everyone, regardless of the technical resources at their disposal, we have converted our original database into several CSV files, one per type of node and one for all the relationships for each project. You may download an archive in zip or torrent format. Please bear in mind that the archive is large, so if you know how to use BitTorrent, we encourage you to use it:

- *Bahamas Leaks [zip] - [torrent]*
- *Offshore Leaks [zip] - [torrent]*

¹⁷⁴ https://de.wikipedia.org/wiki/Liste_von_Leaks_zu_Steuerdaten.

¹⁷⁵ Ohne Datensätze, die die Finanzverwaltungen aufgekauft haben und die nur innerhalb der jeweiligen Finanzverwaltung verfügbar sind.

¹⁷⁶ Ausführlich zur Entstehungsgeschichte vgl. Obermayer / Obermaier (2016).

¹⁷⁷ Ganz frisch im Herbst aufgetaucht handelt es sich um den mit Abstand größten Datenbestand im Umfang von 2,9 TB. Aktuell werden die Daten von den beteiligten Investigativ-Journalisten ausgewertet und sind noch nicht in der ICIJ verfügbar. Vgl. Balbierer et al. (2021).

¹⁷⁸ <https://www.icij.org/>.

- *Panama Papers [zip] - [torrent]*
- *Paradise Papers [zip] - [torrent]*¹⁷⁹

Die durchsuchbare Datenbank des International Consortium of Investigative Journalists (ICIJ) über die Panama Papers und Offshore-Leaks enthält mehr als 300.000 Einträge. Eine Suche nach Verbindungen zwischen Personen, Unternehmen und Ländern kann mühsam und enorm zeitaufwändig sein.¹⁸⁰

Diese Daten wurde von der Firma Ontotext¹⁸¹ – Hersteller des Graphendatenbanksystems GraphDB – genutzt und mit weiteren Open Data Quellen wie DBpedia¹⁸² und GeoNames¹⁸³ zum „**Ontotext Linked Leaks Knowledge Graph**“¹⁸⁴ verbunden.

Er enthält mehr als 22 Millionen RDF Statements.¹⁸⁵ Auf dieser Basis können unterschiedlichste Suchen und Analysen erfolgen. Die Datenbank ist frei zugänglich auf data.ontotext.com/linkedleaks. Auch der Gesamtdatenbestand lässt sich herunterladen.¹⁸⁶ Das Datenbankmanagementsystem GraphDB-Free ist ebenfalls kostenfrei beziehbar.¹⁸⁷

¹⁷⁹ <https://offshoreleaks.icij.org/pages/database>.

¹⁸⁰ Vgl. Kiryakov (2016). Die Software „Datasource“ des ICIJ ist als Open Source verfügbar. Vgl. <https://datasource.icij.org/>.

¹⁸¹ <https://www.ontotext.com/>.

¹⁸² <https://www.dbpedia.org/>.

¹⁸³ <http://www.geonames.org/>.

¹⁸⁴ <http://data.ontotext.com/linkedleaks>.

¹⁸⁵ <https://www.ontotext.com/blog/linked-leaks-a-smart-dive-into-analyzing-the-panama-papers/>.

¹⁸⁶ Download unter <ftp://ftp.ontotext.com/pub/leaks/rdf/rdf.zip>.

¹⁸⁷ Bezug über <https://www.ontotext.com/products/graphdb/graphdb-free/>.

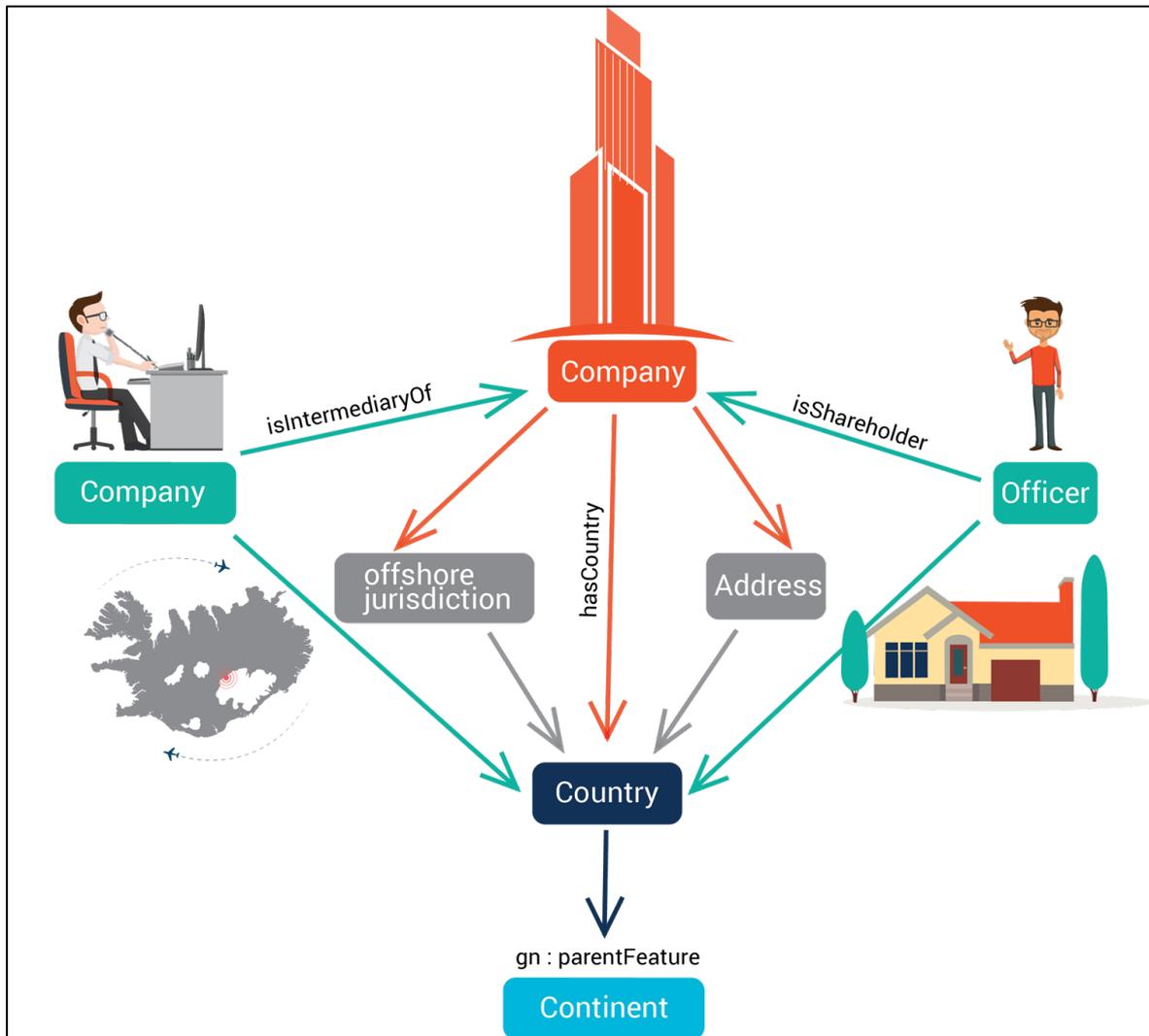


Abb. 12: 'The Game of Queries' in Linked Leaks

Quelle: Kiryakov (2016)

6. Graue Wirtschaft und Schattenwirtschaft

Bereits die Steuerleaks können dem Bereich Schattenwirtschaft zugeordnet werden. Aber auch weitere Daten geben Hinweise für wirtschaftliche Risikobereiche (siehe bereits Kapitel II. C. 3.). Hierzu gehören

- Standorte in Niedrigsteuergebieten oder Niedrigtransparenzgebieten;
- (ehemalige) Straftäter, gesuchte oder sonst wie „verdächtige“ Personen als Eigentümer, Organe oder Geschäftspartner;
- auffällige Geldbewegungen und
- intrasparente Unternehmen als Geschäftspartner;
- (wegen der Möglichkeit einer illegalen Beeinflussung) Kontakte zu sog. „politisch exponierten Personen“.

Umfassender als die Liste „nicht kooperativer Länder und Gebiete“ der EU und als Oxfam’s Rangliste der „Top 15 corporate tax havens“ sind der „**Financial Secrecy Index**“¹⁸⁸ sowie der „**Corporate Tax Haven Index**“¹⁸⁹ das *Tax Justice Network*.¹⁹⁰

Durch die Finanzverwaltung nutzbare Quellen stellen in Deutschland das **Bundeszentralregister**¹⁹¹ und das **Gewerbezentralregister**¹⁹² beim Bundesamt für Justiz dar (siehe auch Kapitel II. C. 3. e).

In Hinblick auf **auffällige Geldbewegungen** hat das Bundesministerium der Finanzen im Jahre 2019 die Ergebnisse der „Ersten Nationalen Risikoanalyse“ veröffentlicht.¹⁹³ Als größte Risikofelder gelten anonyme Transaktionsmöglichkeiten, der Immobiliensektor, der Bankensektor (insbesondere im Rahmen des Korrespondenzbankgeschäfts und der internationalen Geldwäsche), grenzüberschreitende Aktivitäten und das Finanztransfergeschäft wegen der hohen Bargeldintensität. Das Geldwäschegesetz sieht Meldepflichten für Verpflichtete (§ 43 GwG) und Aufsichtsbehörden (§ 44 GwG) vor. Zu den verpflichteten Berufsgruppen (§ 2 GwG) zählen beispielsweise Kreditinstitute, Finanzdienstleistungsinstitute, Zahlungsinstitute, Finanzunternehmen, Versicherungsunternehmen, Versicherungsvermittler, Kapitalverwaltungsgesellschaften, Rechtsanwälte, Notare, Wirtschaftsprüfer, Steuerberater, Immobilienmakler und Glücksspielveranstalter. Die zentrale Auswertung der Verdachtsmeldungen soll durch die „Zentralstelle für Finanztransaktionsuntersuchungen“ (§ 27 GwG) erfolgen. Sie ist aktuell beim Zoll – nicht den Finanzbehörden – angesiedelt und nennt sich „Financial Intelligence Unit (FIU)“.¹⁹⁴ Allerdings gab es starke Kritik, dass die Auswertung und Weiterleitung an Strafverfolgungsbehörden nicht gut funktioniert.¹⁹⁵

Ähnliche Daten und Analysen gibt es auch in anderen Ländern, mit denen teilweise ein Datenaustausch besteht. Nur gelegentlich werden solche Daten in der Öffentlichkeit bekannt, wie im Falle der **FinCEN-Files**. Hierbei handelt es sich um 22.000 Seiten mit Berichten über verdächtige Geldbewegungen, die von Banken weltweit an die US-Finanzaufsicht FinCEN gesendet wurden.¹⁹⁶ Sie wurden durch das International Consortium of Investigative Journalists (ICIJ) ausgewertet.¹⁹⁷ Ein Auszug von 18.153 Transaktionen ist online verfügbar.¹⁹⁸ Der Gesamtbestand kann aus rechtlichen Gründen nicht veröffentlicht werden, da es sich um amtliche US-Dokumente handelt.¹⁹⁹ Die Analysen zeigen aber, dass zahlreiche Gelegenheiten zur Austrocknung von Geldwäsche nicht genutzt wurden.²⁰⁰

¹⁸⁸ fsi.taxjustice.net/en/.

¹⁸⁹ www.corporatetaxhavenindex.org/en/.

¹⁹⁰ www.taxjustice.net.

¹⁹¹ https://www.bundesjustizamt.de/DE/Themen/Buergerdienste/BZR/Inhalt/Uebersicht_node.html.

¹⁹² https://www.bundesjustizamt.de/DE/Themen/Buergerdienste/GZR/GZR_node.html.

¹⁹³ BMF (2019).

¹⁹⁴ https://www.zoll.de/DE/FIU/fiu_node.html.

¹⁹⁵ Vgl. Becker / Diehl / Traufetter (2021), Diehl (2021).

¹⁹⁶ Vgl. Engert (2020).

¹⁹⁷ <https://www.icij.org/investigations/fincen-files/>.

¹⁹⁸ <https://www.icij.org/investigations/fincen-files/explore-the-fincen-files-data/>.

¹⁹⁹ Vgl. Engert (2020).

²⁰⁰ Vgl. Engert / Drepper (2020).

Eine umfangreiche öffentliche Sammlung über Korruption und organisierte Kriminalität wird durch das **OCCRP²⁰¹ (Organized Crime and Corruption Reporting Project)** mit der Datensammlung **Aleph²⁰²** unterhalten. Der OCCRP Aleph wird durch die „Journalism Development Network, Inc.“ mit Sitz in Maryland, USA, einer steuerbefreiten gemeinnützigen Organisation, betrieben.²⁰³

Aus der Selbstdarstellung: „As an investigative reporting platform for a worldwide network of independent media centers and journalists, OCCRP is reinventing investigative journalism as a public good. In the face of rising costs and growing threats to independent media, OCCRP provides media outlets and journalists with a range of critical resources and tools including digital and physical security and allows those covering the most sensitive topics to work in teams with trusted editors. While upholding the highest journalistic ethics and editorial standards, OCCRP develops and deploys cutting-edge tech tools to enable collaborative, secure data-driven investigations. With OCCRP Aleph, an investigative data platform powered by software we developed, journalists can search and cross-reference more than two billion records to trace criminal connections and patterns and efficiently collaborate across borders.“²⁰⁴

Die Datenbank umfasst zurzeit mehr als 300 Millionen Einträge in über 250 Datensätzen aus 140 Ländern.²⁰⁵ Das gemeinsame Datenmodell des Wissensgraphen (Knowledge Graph) wird im Web beschrieben.²⁰⁶

Der Zugriff auf die Daten ist über eine Softwareschnittstelle (API) möglich, wobei die Zugriffsraten für nicht-registrierte Nutzer begrenzt sind. Für die Öffentlichkeit wird auch eine Auswahl Offener Datenbestände zum Download bereitgestellt. Angemeldete Journalisten können für ihre Projekte Massendownloads durchführen. Teil des Aleph-Software Toolkits ist Aleph Data Desktop, ein Werkzeug, mit dem man Netzwerke von Unternehmen, Personen und deren Beziehungen auf dem eigenen Computer abbilden kann. Die Software ist Open Source und wird auch von anderen Organisationen genutzt.²⁰⁷

Das kommerzielle Angebot „Watchlists & Blacklists“ der info4c AG²⁰⁸ aus der Schweiz sammelt internationale Warnmeldungen von Aufsichtsbehörden (Bafin, FMA, CBFA, FINMA, CNMV, MAS, FCA etc.), Fahndungsmeldungen (Interpol, FBI, DEA, DIA etc.), „disqualified directors“ (Personen, denen eine Tätigkeit als Geschäftsführer untersagt ist), Sanktionslisten und manches mehr.²⁰⁹

Zu den „**politisch exponierten Personen**“ (PEP) nach Art. 3 der Richtlinie (EU) 2015/849 des Europäischen Parlaments und des Rates vom 20. Mai 2015 gehören natürliche Personen, die wichtige öffentliche Ämter ausüben oder ausgeübt haben und deren Familienmitglieder sowie bekanntermaßen nahestehende Personen. Die Richtlinie wurde in Deutschland im Geldwäschegesetz umgesetzt (§ 1 Abs.

²⁰¹ <https://www.occrp.org>.

²⁰² <https://aleph.occrp.org>.

²⁰³ Vgl. <https://aleph.occrp.org/pages/terms>.

²⁰⁴ <https://www.occrp.org/en/about-us>.

²⁰⁵ Vgl. <https://aleph.occrp.org>.

²⁰⁶ Auf <https://followthemoney.readthedocs.io/en/latest/entity.html>.

²⁰⁷ Vgl. die Dokumentation unter <https://docs.alephdata.org/guide/getting-started>.

²⁰⁸ <https://www.info4c.net/>.

²⁰⁹ Vgl. <https://www.info4c.net/solutions/watchlists-blacklists/>.

12-14 GWG). Für solche Personen gelten verschärfte Sorgfaltspflichten (§ 15 Abs. 3 Nr. 1, Abs. 4 GWG).

„Bei politisch exponierten Personen geht man aufgrund ihrer einflussreichen Position von einem höheren Risiko der Korruption und Geldwäsche aus. Nicht selten kommt es vor, dass PEPs in die Zahlung von Bestechungsgeldern verwickelt sind, um Entscheidungen oder Auftragsvergaben zu beeinflussen, Terrorismus finanzieren, Steuern hinterziehen oder illegal erworbene Gelder waschen. Die 2016 veröffentlichten Panama Papers brachten beispielsweise Daten zu 140 PEPs zum Vorschein, die Briefkastenfirmer genutzt haben sollen, um Geld zu waschen oder um sich selbst als Eigentümer unsauberen Geldes zu vertuschen.“²¹⁰

Listen politisch exponierter Personen (PEP-Listen) werden von verschiedenen kommerziellen Datenlieferanten geführt, beispielsweise „Nexis Diligence“ von LexisNexis²¹¹ oder „PEP Desk“ von info4c.²¹²

Es gibt jedoch auch frei zugängliche Datensammlungen, etwas durch „opensanctions.org“.²¹³ Hier zu finden ist u.a. eine Aggregation „Politically Exposed Persons (PEPs)“ aus 5 verschiedenen Datenquellen mit insgesamt über 87.000 Einträgen.²¹⁴ Darunter sind beispielsweise auch frühere oder ehemalige Mitglieder des Deutschen Bundestages oder die Mitglieder des Europäischen Parlaments zu finden.

7. Nachrichten und Ereignisse

Ein Vergleich der den Finanzbehörden vorliegenden Daten mit Nachrichten in der Presse oder andere Meldungen über Ereignisse kann ebenfalls Unstimmigkeiten zu Tage treten lassen oder bisher unentdeckte Verbindungen aufdecken.

Aufgrund der Vielzahl von (Presse-)Medien wird es in der Regel keinen Sinn machen, die einzelnen Medienquellen individuell auszuwerten. Sinnvoller ist ein Zugriff auf Aggregatoren, die eine Vielzahl von Medien erfassen. Beispielhaft seien genannt:

- **Google News.**²¹⁵ Die (Links auf) die Original-Nachrichten werden von einer Vielzahl von Quellen zusammengetragen. Es gibt keine Aufbereitung, aber die Möglichkeit einer Personalisierung, d.h. die Wahl eigener Quellen und Themen.
- **Global Entity Graph (GEG) des “GDEL Project”** (Global Database of Events, Language, and Tone).²¹⁶ *“GDEL monitors print, broadcast, and web news media in over 100 languages from across every country in the world to keep continually updated on breaking developments anywhere on the planet. Its historical archives stretch back to January 1, 1979 and update every*

²¹⁰ <https://www.lexisnexis.de/begriffserklaerungen/compliance/pep-politisch-exponierte-personen>.

²¹¹ Vgl. <https://www.lexisnexis.de/loesungen/compliance/geschaeftpartnerpruefung-diligence>.

²¹² Vgl. <https://www.info4c.net/solutions/pep-desk-database/>.

²¹³ Vgl. <https://opensanctions.org>.

²¹⁴ Vgl. <https://opensanctions.org/datasets/peps/>.

²¹⁵ <https://news.google.com>.

²¹⁶ <https://www.gdelproject.org/>. GDEL bzw. GEG wird durch Google Jigsaw unterstützt.

15 minutes. ... The GDELT Translingual platform represents what we believe is the largest realtime streaming news machine translation deployment in the world: all global news that GDELT monitors in 65 languages, representing 98.4% of its daily non-English monitoring volume, is translated in realtime into English and processed.”²¹⁷ Es identifiziert hierin Personen, Orte, Organisationen, Themen, Zahlen und Ereignisse. Die gesamte Datenbank ist kostenlos frei verfügbar, zur Analyse online oder zum Download.²¹⁸

- Von Ontotext stammt das Angebot “**FactForge.net – Open Data and News about People, Organizations and Locations**”:²¹⁹ “*FactForge.net is a hub of Linked Open Data (LOD) and news articles about people, organizations and locations. It includes more than 1 billion facts from popular datasets such as DBpedia, Geonames, Wordnet, the Panama Papers, etc., as well as ontologies such as the Financial Industry Business Ontology (FIBO). It also includes a live stream of news articles and metadata linking news to entities and concepts: about 2000 articles/day tagged by Ontotext’s Publishing platform.*” Es ist also keine reine Nachrichtenseite, sondern verknüpft die Daten bereits mit vielfältigem Hintergrundwissen.²²⁰ Die Speicherung als RDF-Graph ermöglicht eine Verknüpfung von Daten und somit die Generierung neuen Wissens. Abfragen sind beispielsweise über die Seite <http://factforge.net/sparql> möglich.

8. Soziale Medien

Selbstverständlich können Alternative Daten auch aus den Sozialen Medien stammen.²²¹ Das bezieht sich einerseits auf Postings von Unternehmen, aber vor allem auch auf Posts durch die mit diesen verbundenen Personen. Es hängt von der jeweiligen Plattform ab, in welchem Umfang (Menge) bzw. mit welcher Datenrate (Geschwindigkeit) ein Abruf welcher Daten (Herkunft) möglich ist.

Einige Beispiele:

- LinkedIn - LinkedIn’s Economic Graph²²²
- Facebook – Facebook Social Graph API²²³
- Instagram – Instagram Graph API²²⁴
- Twitter – Twitter APIs²²⁵

²¹⁷ <https://www.gdeltproject.org/>.

²¹⁸ Vgl. <https://www.gdeltproject.org/data.html>. Die Daten nur des letzten Jahres umfassen aber bereits mehr als 2,5 TB.

²¹⁹ <https://www.ontotext.com/knowledgehub/demoservices/factforge-explore-linked-open-data/>.

²²⁰ <https://www.ontotext.com/knowledgehub/demoservices/factforge-explore-linked-open-data/>.

²²¹ Vgl. Russel / Klassen (2019).

²²² <https://economicgraph.linkedin.com/>.

²²³ <https://developers.facebook.com/docs/graph-api/>.

²²⁴ <https://developers.facebook.com/docs/instagram-api>.

²²⁵ <https://help.twitter.com/de/rules-and-policies/twitter-api>.

9. Lexika

Lexika bieten oft einen Hintergrund zur Einordnung und somit zum Verständnis von Daten. Neben das für menschliche Leser geschriebene Wikipedia treten auch Formen, die für eine automatisierte maschinelle Auswertung besser geeignet sind. Zu nennen sind insbesondere DBpedia²²⁶ und Wikidata.

DBpedia enthält Informationen (u.a.) aus den Texten von Wikipedia und speichert sie in strukturierter Form ab. Ein Zugriff kann über eine Webanwendung, einen SPARQL-Endpunkt sowie durch einen Massendownload des Datenbestandes (Datenbankdump)²²⁷ erfolgen. Die Inhalte ergeben einen Offenen Wissensgraphen (open knowledge graph; OKG).²²⁸

Wikidata ist quasi eine Art Daten-Äquivalent zu den Textinhalten von Wikipedia. Ähnlich wie Wikipedia ist auch Wikidata frei bearbeitbar.²²⁹ Mit Wikidata soll eine Wissensdatenbank geschaffen werden, deren Inhalte auch in anderen Wikimedia-Projekten, u.a. auch Wikipedia, genutzt werden können.²³⁰ Alle Inhalte des Wissensgraphen sind von jedermann frei nutzbar.

10. Weitere Sammlungen

Im Internet finden sich weitere Sammlungen, die neben überwiegend nicht relevantem Material auch potentiell relevante Inhalte enthalten können.

Auf „**Linked Open Data Cloud**“²³¹ wird auf Datenbestände verwiesen, die gemäß der Linked Data Prinzipien publiziert wurden, also als RDF-Graph.²³² Der Datensatz muss mindestens 1000 Einträge umfassen, vollständig abrufbar sein (durch RDF Crawling, einen Massendownload oder über einen SPARQL Endpunkt) sowie mit dem bisherigen Datenbestand verlinkt werden.

Für alternative Daten aus der klassischen Sicht von Finanzinvestoren gibt es eine Vielzahl **kommerzieller Datenlieferanten**. Einen Einstieg geben beispielsweise ravenpack.com,²³³ alternativdata.org²³⁴ oder datarade.ai.²³⁵

²²⁶ <https://www.dbpedia.org/>. Die Initiative ging von Deutschland aus.

²²⁷ Unter <https://dumps.wikimedia.org/>.

²²⁸ Vgl. <https://www.dbpedia.org/about/>.

²²⁹ Vgl. <https://www.wikidata.org/wiki/Wikidata:Introduction>.

²³⁰ https://www.wikidata.org/wiki/Wikidata:Main_Page.

²³¹ <https://lod-cloud.net/>.

²³² Vgl. <https://www.w3.org/DesignIssues/LinkedData.html>.

²³³ <https://www.ravenpack.com>.

²³⁴ <https://alternativedata.org/data-providers/>.

²³⁵ <https://datarade.ai/data-categories>.

Eine Vielzahl von anderen bereits im Web gecrawlte Daten wird frei über „commoncrawl.org“ zur Verfügung gestellt.²³⁶

Zuletzt genannt sei **Dataset Search**, die Datensuchmaschine von Google.²³⁷ Dabei handelt es sich – im Unterschied zur „normalen“ Google Suche - um eine auf Datensammlungen spezialisierte Suchmaschine.²³⁸ Ihre Nutzer können durch einfache Suchbegriffe Datensätze finden, die in Repositories im weltweiten Web gehostet werden.²³⁹

²³⁶ <http://commoncrawl.org/>.

²³⁷ <https://datasetsearch.research.google.com/>.

²³⁸ Ähnlich wie Google Scholar wissenschaftliche Veröffentlichungen sucht. Siehe <https://scholar.google.com/>.

²³⁹ Vgl. <https://datasetsearch.research.google.com/help>.

C. Ergebnis und Ausblick

Die Auflistung Alternativer Datenquellen belegt, dass bereits heute ein umfangreiches Angebot an automatisiert auswertbaren Datenbeständen vorliegt, die sich zur Verprobung von und Verknüpfung mit steuerlichen Daten eignen.

Dieses Flickwerk an Daten ist beileibe nicht vollständig und nicht für alle Bereiche geeignet. Jedoch lassen sich einige Trends feststellen: Der Umfang der Daten wächst beständig und zwar sowohl in der Tiefe (durch vermehrte Einträge in bestehende Datensätze) als auch in der Breite (durch zusätzliche Datensätze).

Ferner werden die bisherigen „Dateninseln“ immer stärker miteinander verknüpft, was deren Auswertung enorm vereinfacht. So hat die EU als Ziel den Aufbau eines gemeinsamen europäischen Datenraums ausgegeben.²⁴⁰ Das geschieht einerseits innerhalb einzelner Plattformen, andererseits durch wechselseitige Verweise insbesondere mit Techniken aus dem Internet.

Zusätzlich gibt es immer mehr Offene Daten in Deutschland,²⁴¹ in der Europäischen Union und weltweit. Das data.europa.eu-Portal bietet einen Zugang zu offenen Daten aus internationalen, EU-, nationalen, regionalen und lokalen Datenportalen.²⁴²

Daher werden sich die Möglichkeiten zur Nutzung (auch) Alternativer Daten künftig weiter verbessern.

²⁴⁰ EU, Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen „Aufbau eines gemeinsamen europäischen Datenraums“ vom 25.4.2018, COM(2018) 232 final.

²⁴¹ Zu Offenen Daten im Inland am Beispiel von Berlin: Berliner Senat, Verordnung zur Bereitstellung von allgemein zugänglichen Datenbeständen (Open Data) durch die Behörden der Berliner Verwaltung (Open Data Verordnung - OpenDataV) vom 24. Juli 2020; Berlinonline (2021): Das Berliner Open-Data-Handbuch, <https://berlinonline.github.io/open-data-handbuch/> (13.3.2021).

²⁴² Vgl. <https://data.europa.eu/en/about/about-dataeuropa.eu>. Es ersetzt das EU Open Data Portal und das European Data Portal.

IV. Prototypische Umsetzung

A. Technische Basis

1. Aufgabenstellung

Für den Einsatz in der Finanzverwaltung besteht die Aufgabe nicht nur in der Sammlung von Alternativen Daten. Diese müssen vielmehr in einer Form gespeichert werden, die gezielte Abfragen auf den integrierten Gesamtdatenbestand ermöglicht. Daraus ergeben sich folgende Arbeitsschritte:

- Sammlung der Alternativen Daten als Rohdaten
- Überführung in ein einheitliches Datenformat
- Integration der Teil-Datenbestände
- Erforschung des Gesamtdatenbestands durch Abfragen

2. Datensammlung

Die Alternativen Daten können grundsätzlich in ganz verschiedenen **Dateitypen** vorliegen. Hierzu zählen u.a. einfache TXT-Dateien, PDF, MS-Word, MS-Excel, CSV, Bilder/Scans als JPEG oder TIFF,²⁴³ HTML, XML, XBRL, iXBRL, JSON oder RDF-Serialisierungen.

Diese Vielfalt lässt sich technisch gesehen geordnet in einem sog. Data Lake speichern, der sowohl strukturierte wie auch semi-strukturierte und unstrukturierte Daten in ihrem Raw-Format aufnehmen kann. Dieser große See an Rohdaten „verschluckt“ somit zunächst einmal alles. Die künftige Verwendung kann dabei vorläufig offenbleiben.

In vielen Fällen wird den deutschen Behörden **kein kompletter Datensatz** zur Verfügung stehen. Der Zugriff erfolgt durch Abfragen, die sich nach konkreten Suchobjekten (Unternehmen, Personen) richten. Idealerweise ist eine **Programmschnittstelle – API** (Application Program Interface) – vorhanden. Ist eine solche nicht verfügbar, so sind Abfragen (nur) über die Weboberfläche möglich. Solche Abfragen müssen im Einzelfall erstellt und ausgewertet werden, was mit viel Arbeit verbunden ist.

Jedoch gibt es hier ein sehr hohes Automatisierungspotenzial durch sog. „**Robotic Process Automation (RPA)**“. Unter Robotic Process Automation darf man sich keinen körperlichen Roboter vorstellen. Es ist vielmehr ein Programm, das Anstelle des Menschen den Computer und die dort installierte Software über die Benutzeroberfläche bedient. *Moffitt / Rozario / Vasarhelyi* beschreiben diese Vorgehensweise sehr anschaulich wie folgt: „First and foremost, RPA robots conduct work the same way that humans do, through the software presentation layer. Logins, emails, analyses, report building, data entry, and other functions are still completed. RPA robots can be compared to the recorded macros in Excel that automate specific tasks. The primary difference between the two is that RPA

²⁴³ Bilder, aber auch Videos oder Audios werden nicht als alphanumerische Zeichen gespeichert, sondern in Binärformaten. Man spricht auch von Blobs (Binary Large Objects).

'macros' can be recorded to work with virtually any existing desktop or server software. RPA software generally includes an interface with a record button that, when activated, generates a script, or robot, as a user performs the task that is to be automated. With some configuration, robots can be trained to read emails, open PDFs, identify salient information, enter data into ERP systems, and send an email to specific supervisors when ambiguity or errors are encountered."²⁴⁴ Der von der RPA-Software erstellte Programmcode wird auch als „Robot“ oder kurz „Bot“ bezeichnet.

RPA eignet sich besonders für wohldefinierte Aufgaben.²⁴⁵ Hierzu zählen etwa Abfragen ausländischer Register oder Abfragen kommerzieller Datensammlungen (LinkedIn, PEP etc.). In solchen Anwendungsszenarien müssten folgende Bedingungen für den Einsatz von RPA vorliegen:

- (1) Die einzelne Abfrage ist kostenlos oder die mit dem Abruf verbundenen Kosten werden in Kauf genommen. Das könnte beispielsweise bei einer Flatrate von monatlich 1.000€ der Fall sein, wohingegen Kosten von 5€ pro Zugriff schwer kalkulierbar sind.
- (2) Es gibt keine Mechanismen, die eine Abfrage durch Bots (anstelle von Menschen) zu verhindern suchen oder diese Mechanismen werden bewusst überwunden. Hierzu werden sog. CAPTCHA-Tests eingesetzt. CAPTCHA steht für „Completely Automated Public Turing test to tell Computers and Humans Apart“.²⁴⁶ Das kann beispielsweise in Form von 9 oder 12 Fotos geschehen, auf denen der Nutzer diejenigen markieren soll, auf denen Fahrräder (oder Ampeln, Zebrastreifen etc.) zu sehen sind. Auch das deutsche Unternehmensregister versucht Bots auszuschließen.

3. Einheitliches Datenformat

Nach erfolgter Sammlung von Rohdaten müssen die heterogenen Datenbestände in der Folge transformiert werden, um sie besser analysieren zu können. Für die Transformation in ein einheitliches Datenformat sollte dieses gewissen Bedingungen genügen.

In jedem Fall muss es sich um ein **strukturiertes Format** handeln. Gegebenenfalls müssen vorher unstrukturierte Daten in strukturierte Daten überführt werden:

- Im Extremfall liegen Scans von Dokumenten in Binärform wie z.B. JPEG vor. Über diese Bilddaten muss dann eine sog. OCR-Software (Optical Character Recognition) laufen, um die abgebildeten Buchstaben und Zahlen zu identifizieren und alphanumerisch in einer TXT-Datei abzuspeichern.
- Aber auch Fließtexte müssen noch inhaltlich analysiert werden. Ein Text enthält beispielsweise Aussagen zu Unternehmen („ABC GmbH“), Personen („Max Müller“), Funktionen („Prokurist“), Orten („Luxembourg“) usw. Diese sollen erkannt und extrahiert werden. Hierbei hilft eine sog. „Named Entity Recognition (NER)“-Software.²⁴⁷

²⁴⁴ Moffitt / Rozario / Vasarhelyi (2018), S. 2.

²⁴⁵ Vgl. Huang / Vasarhelyi (2019).

²⁴⁶ Vgl. Cloudflare (2021).

²⁴⁷ Vgl. Balogh (2018). Andere Aufgabe wären z.B. das Erkennen von Personen auf Bildern oder in Videos.

Die Struktur der Daten sollte **wesentliche Eigenschaften der interessierenden Objekte** (Entitäten, Entities) erfassen. Zu den Objekten gehören insbesondere Unternehmen und Personen. Sie werden durch verschiedene Eigenschaften (Attribute) sowie ihre Beziehungen (Relationships) untereinander charakterisiert.²⁴⁸

Das Datenmodell sollte **einfache Erweiterungen** des ursprünglichen Datenbestandes erlauben. Darunter ist weniger das Hinzufügen einer zusätzlichen Anzahl derselben Daten gemeint, also wenn sich etwa die Liste von „Politisch Exponierten Personen“ um weitere Personen erweitert. Es geht vielmehr um das Auftreten von Daten mit neuen, zusätzlichen Merkmalen. Beispielsweise können die Aktivitäten eines Tochterunternehmens auch um das Merkmal ergänzt werden, ob es sich um Einkünfte aus passiven Betätigungen im Sinne von § 8 AStG handelt. Daten haben häufig Lücken, weil nicht in allen Fällen sämtliche Eigenschaften bekannt sind.

Das Datenformat soll schließlich eine **problemlose Verknüpfung** der unterschiedlichen Datenquellen erlauben. Dabei ist die Verwendung etablierter und erprobter **Standards** förderlich.

Vor dem Hintergrund dieser Anforderungen gibt es eine Reihe von Datenbank-Konzepten, die geeignet sind, strukturierte Daten zu speichern.²⁴⁹

(1) Heute am weitesten verbreitet sind **relationale Datenbankmanagementsysteme (RDBMS)**,²⁵⁰ die Daten aus logischer Sicht in Tabellen organisieren. Die Tabellen bestehen aus vertikalen Spalten (Datenfelder mit festgelegter Datenstruktur) und horizontalen Zeilen (Tupel oder Datensätze). Der gesamte Tabellenentwurf wird als Datenbankschema bezeichnet. Dieses vordefinierte Schema bedeutet ein starres Datenmodell, welches im Vorfeld ein sorgfältiges Design erfordert, denn jede zukünftige Änderung des Datenschemas ist aufwändig und kostspielig. Die Definition der Datenstruktur, Bearbeitungen (Einfügen, Verändern, Löschen) und Abfragen erfolgen über die Sprache SQL.

(2) „**Document Stores**“ sind hingegen Datenmodelle,²⁵¹ die auch für die Verwaltung von unstrukturierten (z. B. Text) oder halbstrukturierten (z. B. XML) Daten geeignet sind und in der Regel eine hierarchische Natur aufweisen.²⁵² Bei dokumentenorientierten Datenbanken können die in jedem Dokument gespeicherten Daten unterschiedliche Datenfelder und Datenfeldtypen haben. Dies ist bei Datensätzen in relationalen Datenbanken nicht möglich. Das macht dokumentenorientierte Datenbanken flexibler bei Änderungen. Umgekehrt ist deshalb eine Analyse des Gesamtdatenbestands durch Abfragen nicht auf einfache Weise möglich.

(3) Einen Ausweg können **Graphen-Datenbanken** bieten. Graphendatenmodelle basieren auf der mathematischen Graphentheorie. Ganz allgemein speichern sie Informationen über Objekte (Entities) und deren Beziehungen (Relations). Die Objekte werden als Knoten und ihre Beziehungen als Kanten gespeichert. Daher zeichnen sich Graphen-Datenbanken durch die Speicherung und Verarbeitung stark miteinander verbundener Datenstrukturen aus, im Unterschied zum relationalen Datenmodell,

²⁴⁸ Vgl. Balogh (2018), S. 3f.

²⁴⁹ Zu anderen Formen von Datenbankmanagementsystemen vgl. die Technik- und Marktübersicht auf <https://db-engines.com>.

²⁵⁰ Zu den relationalen Datenbankmanagementsystemen gehört beispielsweise MySQL. Vgl. <https://www.mysql.com>., <https://db-engines.com/de/ranking/relational+dbms>.

²⁵¹ Das beliebteste dokumentenorientierte Datenbankmanagementsystem ist etwa MongoDB. Vgl. <https://www.mongodb.com>, <https://db-engines.com/de/ranking/document+store>.

²⁵² Vgl. Foote (2018); Nayak / Poriya / Poojary (2013); Sharma / Dave (2012).

bei dem die Tabellen nur lose verbunden sind. Dies passt sehr gut zur Aufgabe einer Verifizierung von steuerlichen Daten durch (ggf. eine Kette von) Alternativen Daten. Neue Knoten und Kanten können zu einer bestehenden Graphen-Datenbank hinzugefügt werden, sobald man sie benötigt, wodurch die Datenbank auch langsam organisch wachsen kann.

Graphen-Datenbanken haben in jüngster Zeit an Beliebtheit gewonnen, eben weil sie heterogene Daten, Integration neuer Datenquellen und ein für die Analytik geeignetes Strukturschema ermöglichen.²⁵³ Beispiele sind die Nutzung von Graphen-Datenbanken u.a. durch Google,²⁵⁴ Facebook, LinkedIn, DBPedia, Wikidata, FactForge, OCCRP Aleph oder das ICJI.²⁵⁵

Im Wesentlichen gibt es zwei Hauptmodelle von Graphen-Datenbanken: **Labeled-Property Graph** und **RDF Stores**.²⁵⁶ Der prominenteste Vertreter des Labeled-Property Graph ist Neo4j,²⁵⁷ zu den RDF Stores gehören u.a. GraphDB von Ontotext (kommerziell) und „Apache Jena – TDB“ (Open Source).²⁵⁸

Wegen der Tatsache, dass es keinen Standard für das Labeled-Property Graph-Datenmodell gibt, existiert leider auch keine gemeinsame Abfragesprache für den Zugriff auf Daten in der Graphen-Datenbank sowie kein Standard zum Importieren von Daten (Upload) in die Graphen-Datenbank. Daher ist es Sache des DBMS-Anbieters, seine eigene Abfragesprache oder Vorgaben für den Datenaustausch zu definieren. Beispiele sind Cypher,²⁵⁹ Gremlin, GraphQL und andere proprietäre Sprachen.²⁶⁰

Für den angestrebten Einsatzzweck haben Graphen-Datenbanken als RDF Store entscheidende Vorteile, denn sie beruhen auf anerkannten Standards.

So gehört RDF zu den **Standards der EU** und bildet einen Teil der Digitalisierungsstrategie. RDF wurde am 20. Juli 2017 von der EU als offizieller technischer Standard (RTS) für Informations- und Kommunikationstechnologien (IKT) angenommen (ähnlich wie XBRL als RTS für IKT am 28. Januar 2016).

RDF ist Teil der internationalen **Standards des W3C** (World Wide Web Consortium) für das Internet:

“RDF is a standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed.

²⁵³ Vgl. Polikoff (2020).

²⁵⁴ Vgl. Singhal (2012) und <https://support.google.com/knowledgepanel/answer/9787176?hl=de>.

²⁵⁵ Siehe die jeweiligen Hinweise auf die Wissensgraphen weiter oben im Text.

²⁵⁶ Vgl. Barrasa (2017), Polikoff (2020).

²⁵⁷ Vgl. <https://db-engines.com/de/article/Graph+DBMS>.

²⁵⁸ Vgl. <https://db-engines.com/de/article/RDF+Stores>.

²⁵⁹ Cypher wurde inzwischen von Neo4j als Open Source frei verfügbar gemacht. Des Weiteren existieren Bemühungen, einen einheitlichen GQL Standard (Graph Query Language Standard) zu etablieren. Vgl. <https://www.gqlstandards.org/>.

²⁶⁰ Vgl. <https://www.gqlstandards.org/existing-languages>.

RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a “triple”). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications.

This linking structure forms a directed, labeled graph, where the edges represent the named link between two resources, represented by the graph nodes. This graph view is the easiest possible mental model for RDF and is often used in easy-to-understand visual explanations.“²⁶¹

Zu den weiteren relevanten Standards des W3C gehören (u.a.) das RDF Schema (RDFS),²⁶² die Web Ontology Language (OWL)²⁶³ sowie die Abfragesprache SPARQL.²⁶⁴

Als **Zwischenergebnis** lässt sich festhalten, dass Alternative Daten am sinnvollsten in einem **Wissensgraphen (Knowledge Graph) als RDF Store** nach den Standards des W3C gespeichert werden sollten, da dies den Anforderungen am besten entspricht. Hierfür müssen auch keine Lizenzierungskosten anfallen, da alle benötigten Bestandteile als Offene Standards und Freie Software verfügbar sind.

Ein Wissensgraph grenzt sich folgendermaßen von anderen Datensammlungen ab: „A knowledge graph (1.) mainly describes real world entities and their interrelations, organized in a graph, (2.) defines possible classes and relations of entities in a schema, (3.) allows for potentially interrelating arbitrary entities with each other and (4.) covers various topical domains.“²⁶⁵

Wie von Kejriwal beschrieben, kann ein Knowledge Graph als eine „graphentheoretische Darstellung menschlichen Wissens beschrieben werden, so dass es von einer Maschine mit **Semantik** aufgenommen werden kann“.²⁶⁶ Das Journal of Web Semantics gibt folgende Definition an: “Knowledge graphs are large networks of entities, their semantic types, properties, and relationships between entities.“²⁶⁷

Unter Semantik versteht man grundsätzlich die Lehre von der Bedeutung von Dingen. Menschen haben bereits eine gewisse Vorstellung, was bestimmte Dinge (z.B. „Geschäftsführer“, „Anteilseigner“, „Sitz“, „Tochtergesellschaft“) bedeuten. Eine Software besitzt das erst einmal nicht. Daher müssen ihr die nötigen Bedeutungen explizit und auf formale Weise mitgeteilt werden. Hierin besteht die Relevanz von sog. Ontologien, welche in Wissensgraphen verwendet werden.

Ontologien sind formale semantische Datenmodelle, die für unsere Domäne (d.h. den uns interessierenden Bereich) die Arten von existierenden Dingen (Objekten, Entitäten) und die Eigenschaften, die zu deren Beschreibung verwendet werden können, definieren.

²⁶¹ <https://www.w3.org/RDF>.

²⁶² <https://www.w3.org/2001/sw/wiki/RDFS>.

²⁶³ <https://www.w3.org/2001/sw/wiki/OWL>. Näheres in Uschold, (2018).

²⁶⁴ <https://www.w3.org/2001/sw/wiki/SPARQL>. Näheres siehe weiter unten unter 5.

²⁶⁵ Vgl. Paulheim (2017).

²⁶⁶ Vgl. Kejriwal (2019).

²⁶⁷ Vgl. Kroetsch / Weikum (2016). <https://www.websemanticsjournal.org/index.php/ps/announcement/view/19>.

Um die im ersten Schritt aus verschiedensten Quellen zusammengetragenen Daten zu vereinheitlichen, sind Ontologien als essenzieller Bestandteil des Prozesses anzuwenden.²⁶⁸

Ontologien stellen in der Informatik das Grundgerüst einer fehlerlosen und eindeutigen Kommunikation dar. Sie sind unabhängig von einer bestimmten Anwendungssoftware und lassen sich daher mehrfach einsetzen.²⁶⁹ Formale Ontologien sollen semantische Strukturen für alle Teilnehmer verständlich machen.²⁷⁰ In diesem Zusammenhang sind Ontologien daher besonders relevant, um Wissen weiterzugeben und zu verarbeiten.²⁷¹

Ein Wissensgraph integriert Daten aus verschiedenen Quellen in einer Graphenstruktur mit Hilfe von Ontologien, um anhand dieser Strukturen Wissen zu modellieren.²⁷² Wissensgraphen werden durch sog. **Triple** ausgedrückt. Ein Triple erlaubt es, Beziehungen (= Kante) zwischen zwei Entitäten (=Knoten) durch drei Elemente auszudrücken:

(1) Subjekt-Entität, (2) Prädikat und (3) Objekt-Entität.



Abb. 13: Beispiel für ein RDF-Triple

Ein komplexeres Beispiel für einen Wissensgraphen mit Tripeln aus jeweils zwei Knoten und einer Kante findet sich in der **Abb. 4** auf Seite 33.

4. Datenintegration

Die Teil-Datenbestände, welche in einen RDF-Graphen überführt wurden, leiden jedoch in der Regel darunter, dass zur Identifikation der individuellen Unternehmen und Personen in den Teil-Datenbeständen unterschiedliche Schlüssel (Identifiers) verwendet wurden.

Wie weiter oben beschrieben kommen u.a. in Frage:

- Handelsregister-Nummer in Verbindung mit dem Registergericht
- entsprechende Nummern aus den Registern andere Länder

²⁶⁸ Vgl. Fensel et al. (2020); Kendall / McGuinness (2019).

²⁶⁹ <https://it-talents.de/it-wissen/ontologie-in-der-informatik/>.

²⁷⁰ Vgl. Allemang / Hendler (2011), S. 13.

²⁷¹ Vgl. Obitko 2007.

²⁷² <https://www.turing.ac.uk/research/interest-groups/knowledge-graphs>.

- (deutsche) Steueridentifikationsnummer
- LEI
- PermID
- QID Wikidata-Kennung²⁷³
- Open Corporates Company Number
- usw.

Zur Schaffung eines Gesamt-Datenbestandes aus den heterogenen Datenbeständen müssen die unterschiedlichen Identifiers (Schlüssel) der Objekte in einzelnen Datenbeständen mit sog. Mappings untereinander verknüpft werden. Darunter versteht man etwa eine Zuordnung, dass

- das Unternehmen mit der LEI 1234568ABCD12345678 und
- das Unternehmen mit der Handelsregister-Nummer Fürth_HRB12345

identisch sind.

Hierzu kann in Wissensgraphen die OWL-Eigenschaft „SameAs“ verwendet werden, wie das folgende Beispiel für Personen (anstelle von Unternehmen) zeigt:²⁷⁴

“The built-in OWL property `owl:sameAs` links an individual to an individual. Such an `owl:sameAs` statement indicates that two URI references actually refer to the same thing: the individuals have the same ‘identity’.

For individuals such as ‘people’ this notion is relatively easy to understand. For example, we could state that the following two URI references actually refer to the same person:

```
<rdf:Description rdf:about="#William_Jefferson_Clinton">
  <owl:sameAs rdf:resource="#BillClinton"/>
</rdf:Description>
```

Aus logischer Sicht werden dann die beiden Einträge „William_Jefferson_Clinton“ und „BillClinton“ als ein und dieselbe Person behandelt, auch wenn sie in getrennten Datensätzen gespeichert wurden. Das ist selbstverständlich auch für verschieden erfasste Unternehmen möglich, also z.B. „BMW“, „Bayerische Motoren Werke Aktiengesellschaft“ oder „Bayerische Motoren Werke AG“.

Eine Identitätsbeziehung kann auch bei gleichzeitiger Nutzung mehrerer Klassifikationsschemata (Ontologien) definiert werden. Beispielweise wird die Sportart „Fußball“ in England „Football“, in den USA „Soccer“ genannt. Für die Fußballmannschaft könnte man definieren:²⁷⁵

“In OWL Full ... we can use the `owl:sameAs` construct to define class equality, thus indicating that two concepts have the same intensional meaning. An example:

```
<owl:Class rdf:ID="FootballTeam">
```

²⁷³ Vgl. <https://www.wikidata.org/wiki/Q43649390>.

²⁷⁴ <https://www.w3.org/TR/owl-ref/#sameAs-def>, Abschnitt 5.2.1.

²⁷⁵ <https://www.w3.org/TR/owl-ref/#sameAs-def>, Abschnitt 5.2.1.

```
<owl:sameAs rdf:resource="http://sports.org/US#SoccerTeam"/>
</owl:Class>
```

Eine Verknüpfung verschiedener Datenbestände wird auch in der EU angestrebt. Beispielsweise handelt es sich beim **euBusinessGraph**²⁷⁶ um ein Forschungsprojekt von Horizon 2020.²⁷⁷ Ziele waren:²⁷⁸

- Ein System von Identifikatoren (Schlüsseln, Identifiers) für unternehmensbezogene Daten und Entitäten. Dies soll die Verknüpfung von Daten zwischen Entitäten in verschiedenen Ländern und über unterschiedliche Sprachen hinweg ermöglichen. Es bildet eine Grundlage für die Verknüpfung von Daten. Dazu wurden in Europa verwendete Identifikatoren genutzt und Mappings erstellt.
- Entwicklung gemeinsamer mehrsprachiger Datenmodelle (Ontologien und Vokabulare) für Unternehmensdaten. Als Grundlage dienten vorhandene Schemata/Vokabulare/Ontologien wie EU Core Vocabs²⁷⁹ (W3C Org, W3C RegOrg, W3C Location, Person - nicht W3C), schema.org,²⁸⁰ GLEI, FIBO,²⁸¹ Wikidata,²⁸² ADMS.^{283, 284} Das Datenmodell ist Open Source und auf GitHub verfügbar.²⁸⁵

5. Erforschung des Gesamtdatenbestands

Nach der Implementierung des generierten Graphen in eine Datenbank kann der bestehende Graph abgefragt werden, um Informationen zu extrahieren. SPARQL (SPARQL Protocol and RDF Query Language) stellt die Abfragesprache für RDF-Daten im Semantic Web dar. Mittels dieser können komplexe Abfragen von RDF-Datensätze im Linked Open Data Format durchgeführt werden, um die spezifischen Informationen aus diesen herauszuziehen.²⁸⁶

Wie schon oben erwähnt, ist einer der Hauptvorteile des Graphen-Datenmodells die Fähigkeit, in Abfragen Daten zu verbinden, Korrelationen zwischen verbundenen Daten zu erkennen und neues Wissen aus den bisher gespeicherten Daten abzuleiten (sog. Inferenz). Diese Inferenz bzw. semantische Generierung von neuem Wissen fügt daher der Verwendung eines einfachen Graphen eine neue „Wissens“-Ebene hinzu, die ihn in einen Wissensgraphen verwandelt.

²⁷⁶ <https://www.eubusinessgraph.eu/>.

²⁷⁷ Fördernummer Grant Agreement No. 732003.

²⁷⁸ Vgl. <https://www.ontotext.com/knowledgehub/current/eubusinessgraph/>.

²⁷⁹ <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/e-government-core-vocabularies/release/20#download-links>.

²⁸⁰ <https://schema.org/>.

²⁸¹ Financial Industry Business Ontology (FIBO), vgl. <https://spec.edmcouncil.org/fibo/ontology/master/latest/>.

²⁸² <https://www.wikidata.org>.

²⁸³ Asset Description Metadata Schema (ADMS), vgl. <https://www.w3.org/TR/vocab-adms/>.

²⁸⁴ Vgl. <https://www.eubusinessgraph.eu/eubusinessgraph-ontology-for-company-data/>,

²⁸⁵ <https://github.com/euBusinessGraph/eubg-data>. Zur Erläuterung Alexiev (2019).

²⁸⁶ Vgl. Auer / Pellegrini / Sack (2014).

Ein Beispiel dafür zeigt **Abb. 14**. Es ist die Graphen-Darstellung eines örtlichen Coffee House („My_Local_Cafe“, My_Local_Cafe), der Eigentümer-Struktur und der Standorte.

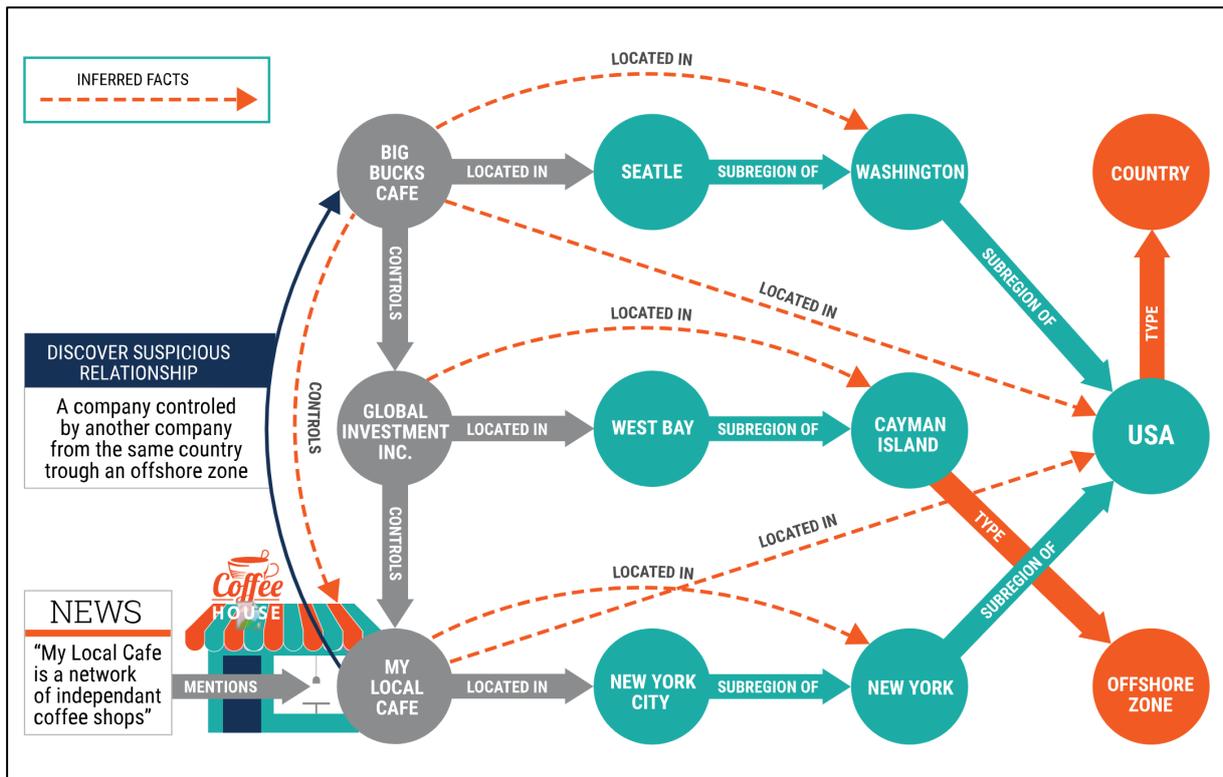
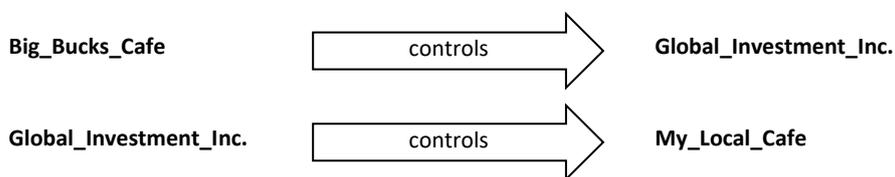


Abb. 14: Erkennen steuerlich risikobehafteter Unternehmensstrukturen

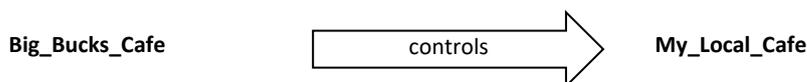
Quelle: Ontotext, <https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/> (25.9.2021)

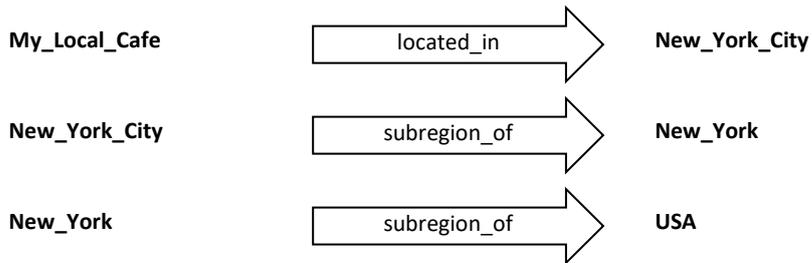
Wenn man das Netz mit seinen verschiedenen Knoten und deren Beziehungen (Kanten) untereinander durchquert, so könnte man auf eine steuerlich verdächtige Beziehung der folgenden Art schließen: „Ein Unternehmen, das von einem anderen Unternehmen aus demselben Land indirekt über eine Offshore-Zone kontrolliert wird“.

Diese Inferenzkette baut sich folgendermaßen auf:

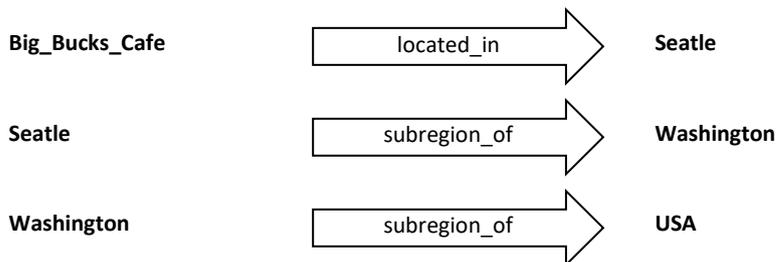
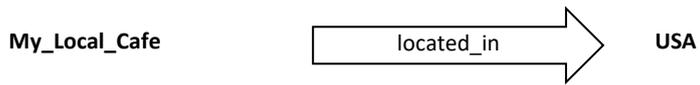


Erste Inferenz:

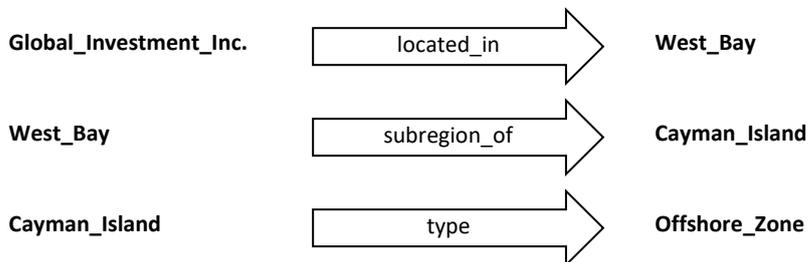
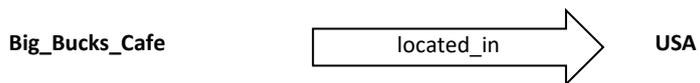




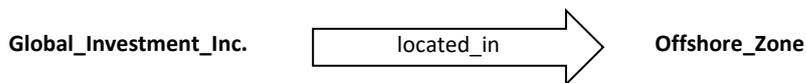
Zweite Inferenz:



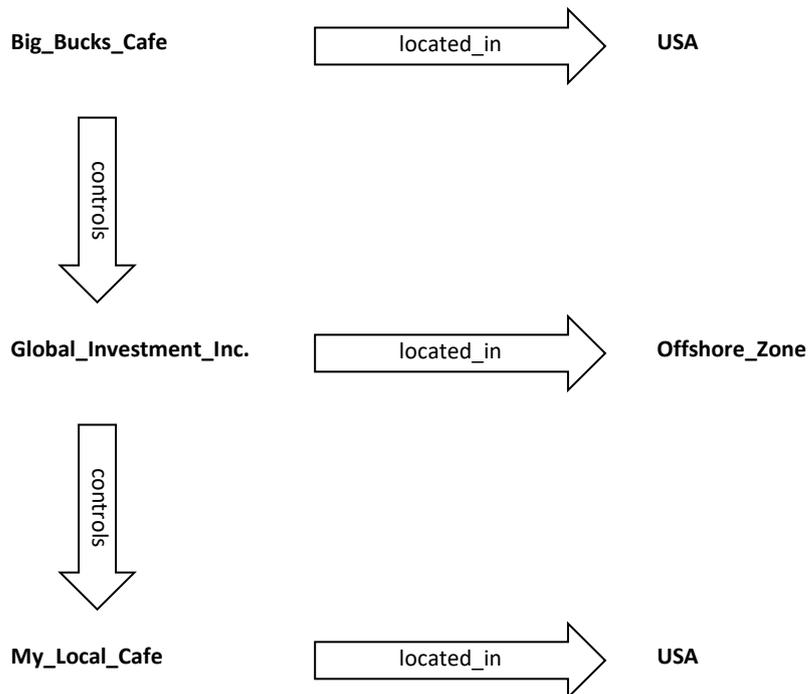
Dritte Inferenz:



Vierte Inferenz:



Zusammen erhält man das folgende „verdächtige“ Muster:



Mithilfe zusätzlicher Konvertierung, je nach verwendetem Tool (z.B. Gephi), können auch komplexere Graphen automatisch visualisiert werden, um weitere Analysen zu ermöglichen. In diesen wird aus dem Subjekt und dem Objekt eines Triples ein Knoten und aus dem Prädikat eine Kante.²⁸⁷ Aufgrund der Visualisierung können dadurch auffällige Muster, z.B. im Zuge der Betrugsaufdeckung, erkannt und dargestellt werden.

²⁸⁷ Arnaout / Elbassuoni (2018).

B. Codierungsbeispiele

1. Deutsches Unternehmensregister

Wie beschrieben verfolgt das Projekt „offeneregister.de“ das Ziel, die faktischen Zugriffsbeschränkungen des offiziellen Handelsregisters zu beseitigen, indem der Inhalt des Handelsregisters aus öffentlich zugänglichen Quellen, insbesondere aus gesetzlich vorgeschriebenen Handelsregisterbekanntmachungen,²⁸⁸ rekonstruiert wird.

Der Gesamtdatenbestand ist hier frei verfügbar. Der Download kann als JSON-Datei oder als SQLite-Datenbank-Datei erfolgen. Es wurde die zweite Form gewählt. Die gezippte Datei hatte ein Volumen von ca. 755 MB. Entpackt umfassen die Daten ca. 2,6 GB. Der Download erfolgt im Februar 2021.²⁸⁹ Damit ergeben sich eine Reihe interessanter Auswertungsmöglichkeiten, die mit dem offiziellen Register so nicht möglich wären.

Von den enthaltenen Daten wurden zwei Tabellen genutzt, die mit „company“ und „officer“ bezeichnet werden. Company sind die im Handelsregister eingetragenen Unternehmen. Entgegen dem Anschein dieser Bezeichnung handelt es sich selbstverständlich nicht nur um Kapitalgesellschaften. Officer umfasst die vertretungsberechtigten Personen, die entweder natürliche oder juristische Personen sein können.

Der Ausgangsdatenbestand von company umfasst 5.305.727 Zeilen und von officer 4.803.514 Zeilen. Jedoch sind hierin sowohl nicht mehr aktive Unternehmen wie auch aktuell nicht mehr vertretungsberechtigten Personen enthalten.

Dies auszuschließen und sich zugleich auf die wesentlichsten Merkmale zu beschränken, wurden durch SQL-Befehle zwei sog. Views (v_company, v_officer) auf die beiden Ausgangstabellen definiert:

```
CREATE VIEW IF NOT EXISTS v_company
AS
SELECT
  id as c_id,
  company_number as c_number,
  name AS c_name,
  registered_office||'_'||registered_address as c_address,
  native_company_number
FROM company
WHERE
  current_status = "currently registered";
```

Damit verblieben 3.341.017 aktive Unternehmen. Einige Datensätze lauten beispielsweise:

²⁸⁸ Vgl. handelsregisterbekanntmachungen.de.

²⁸⁹ Wobei der Datenbestand offensichtlich nicht permanent aktualisiert wird, denn laut Beschreibung handelt es sich um eine Version der Daten vom 15.2.2019. Das ändert aber nichts an den dargestellten grundsätzlichen Auswertungsmöglichkeiten.

c_id	c_number	c_name	c_address	native_company_number
1	K1101R_HRB150148	olly UG (haftungsbeschränkt)	Hamburg_Waidmannstraße 1, 22769 Hamburg.	Hamburg HRB 150148
2	R1101_HRB81092	BLUECHILLED Verwaltungs GmbH	Düsseldorf_Oststr.	Düsseldorf HRB 81092
4	R1101_HRB45109	Albert Barufe GmbH	Hilden_Hans-Sachs-Straße 11, 40721 Hilden.	Düsseldorf HRB 45109
7	K1101R_HRB82839	Verwaltung IFÖ Zweite Immo- bilienfonds für Österreich GmbH	Hamburg_Königstraße 28, 22767 Hamburg.	Hamburg HRB 82839
5	R1101_HRB37996	ITERGO Informationstechnolo- gie GmbH	Düsseldorf_ERGO-Platz 1, 40477 Düsseldorf.	Düsseldorf HRB 37996

Für den View auf officer gilt:

```
CREATE VIEW IF NOT EXISTS v_officer
AS
SELECT
id as o_id,
name AS o_name,
position as o_position,
type as o_type,
company_id as c_number,
city as o_city
FROM officer
WHERE
dismissed ISNULL
```

Es verblieben 3.341.017 „officers“. Ein Auszug aus v_officer:

o_id	o_name	o_position	o_type	c_number	o_city
1	Oliver Keunecke	Geschäftsführer	person	K1101R_HRB150148	Hamburg
2	Christof Wessels	Geschäftsführer	person	R1101_HRB81092	Cloppenburg
3	Christof Wessels	Geschäftsführer	person	R1101_HRB81092	Cloppenburg
11	Joachim Wehrkamp	Geschäftsführer	person	H1101_H1101_HRB18423	Thedinghausen
12	Jörn-Michael Gauss	Geschäftsführer	person	H1101_H1101_HRB18423	Bremen

Auf einen Import in einen RDF Store konnte verzichtet werden, da die wechselseitige Verknüpfung vom Typ „Person – istOfficerBei – Unternehmen“ bzw. „Unternehmen – hatOfficer – Person“ sehr begrenzt sind und sich noch einfach in SQL abbilden lassen.

Im obigen Auszug aus v_officer sind nur natürliche Personen als Geschäftsführer enthalten. Einen Überblick zu den Möglichkeiten und ihrer Häufigkeit (Anzahl der Fälle) gewinnt man durch folgende Datenbankabfrage:

```
SELECT o_position, o_type, count(o_position), count(o_type)
FROM v_officer
GROUP BY o_position, o_type
```

Die Ergebnistabelle zeigt:

o_position	o_type	Anzahl
Geschäftsführer	company	61
Geschäftsführer	person	2192235
Inhaber	company	238
Inhaber	person	126764
Liquidator	company	11990
Liquidator	person	313533
Persönlich haftender Gesellschafter	company	276621
Persönlich haftender Gesellschafter	person	88173
Prokurist	company	4
Prokurist	person	328938
Vorstand	person	2460
Summe		3341017

Interessante Einblicke lassen sich durch komplexere Abfragen gewinnen.

Unter einer sog. „Briefkastenfirma“ versteht man meist ein Unternehmen, das an seiner Adresse keine eigenen Büroräume hat und keine eigenen Mitarbeiter beschäftigt. Die physische Präsenz beschränkt sich somit auf einen Briefkasten bzw. ein Postfach.

Grundsätzlich sind solche Briefkastenfirmen nichts Illegales. Grund hierfür kann die gesellschaftsrechtliche Verkörperung von einzelnen Vermögenswerten wie Grundstücken, Schiffen, aber auch Patenten und anderen Rechten sein. Eine solche Gesellschaft – in der reduziertesten Form eben als Briefkastenfirma – ermöglicht eine Übertragung (Verkauf, Schenkung, Erbe) von Gesellschaftsanteilen anstelle der direkten Übertragung des Vermögenswertes. Hierfür gelten ggf. andere Formvorschriften. Auch lassen sich möglicherweise legale Steuerersparnisse erzielen. Beispiel im deutschen Steuerrecht ist die Übertragung von Anteilen einer grundstücksbesitzenden Kapitalgesellschaft, die keine Grunderwerbsteuer nach sich zieht, wenn Haltefristen (aktuell 10 Jahre) und Beteiligungsgrenzen (aktuell 90%) nicht verletzt werden (§ 1 Abs. 2b GrEStG).

Allerdings mindert der indirekte Besitz von Vermögen grundsätzlich erst einmal die Transparenz. Es ist nicht auf den ersten Blick ersichtlich, wer „wirtschaftlich Berechtigter“ des Vermögens ist, der über das Vermögen letztlich verfügen kann und dem die Erträge zufließen. Das lässt sich durch Einsicht in weitere Register zwar beheben, aber nur mit Zusatzarbeit und selbstverständlich nur, falls dies möglich ist.²⁹⁰

Um Adressen zu finden, die Sitz vieler Unternehmen sind, genügt folgende Abfrage:

```
SELECT c_address, COUNT(c_address)
FROM v_company
GROUP BY c_address
ORDER BY COUNT(c_address) DESC
```

Man erhält erstaunliche Zahlen, hier etwa nur die häufigsten Werte:

²⁹⁰ Insbesondere eben nicht in vielen ausländischen Jurisdiktionen.

c_address	Anzahl
Mainz_Emy-Roeder-Straße 2, 55129 Mainz.	1697
Pullach i. Isartal_Emil-Riedl-Weg 6, 82049 Pullach i. Isartal.	939
München_Leopoldstr.	458
Grünwald_Südliche Münchner Str.	391
Hamburg_Palmaille 67, 22767 Hamburg.	365
Grünwald_Tölzer Straße 15, 82031 Grünwald.	348
München_Landsberger Str.	346
Hamburg_Bleichenbrücke 10, 20354 Hamburg.	336
München_Maximilianstr.	320
Düsseldorf_Königsallee 106, 40215 Düsseldorf.	316
Grünwald_Bavariafilmplatz 7, 82031 Grünwald.	309
Grünwald_Nördliche Münchner Str.	306
Berlin_Charlottenstraße 4, 10969 Berlin	301
Pullach i.Isartal_Emil-Riedl-Weg 6, 82049 Pullach i.Isartal.	296
Hamburg_Elbchaussee 370, 22609 Hamburg.	274
Bremen_Stephanitorsbollwerk 3, 28217 Bremen	268
... usw.	

Hierzu ist freilich Folgendes anzumerken:

Die Einträge im Handelsregister sind teilweise unpräzise. So ist etwa „München_Leopoldstr.“ der dritthäufigste Wert. Da keine Hausnummer enthalten ist, fallen hierunter aber alle Unternehmen aus dieser Straße, welche keine Hausnummer angegeben hatten.

Die Adresse „Pullach i. Isartal_Emil-Riedl-Weg 6, 82049 Pullach i. Isartal.“ kommt scheinbar doppelt vor, einmal an zweiter Stelle (939) und noch weiter unten (296). Erst ein genauer Blick offenbart, dass sich beide Schreibweisen um ein vorhandenes bzw. fehlendes Leerzeichen unterscheiden. Das verdeutlicht ein generelles Problem, was aber ggf. mit mehr Aufwand lösbar wäre.

Selbstverständlich handelt es sich bei manchen Adressen um große Bürogebäude, die Platz für viele Unternehmen bieten. Das ist natürlich nicht immer so.

An 141 „Adressen“ (beachte: für Fälle, in denen keine Hausnummer eingetragen war, ist dies nur die Straße!) sind mindestens 100 Unternehmen gemeldet:

```
SELECT c_address, COUNT(c_address)
FROM v_company
GROUP BY c_address
HAVING COUNT(c_address) > 99
ORDER BY COUNT(c_address) DESC
```

Für 82.225 Adressen gibt es mindestens 10 Unternehmen derselben Anschrift.

Um Fälle ähnlicher Adressen, sei es durch unterschiedliche Schreibweisen oder teilweise fehlende Hausnummern, zu erkennen, kann man sich die Liste auch alphabetisch sortiert ausgeben lassen:

```

SELECT c_address, COUNT(c_address)
FROM v_company
GROUP BY c_address
HAVING COUNT(c_address) > 99
ORDER BY c_address

```

Hier einige Beispiele:

c_address	Anzahl
Pullach i. Isartal_Emil-Riedl-Weg 6, 82049 Pullach i. Isartal.	939
Pullach i.Isartal_Emil-Riedl-Weg 6, 82049 Pullach i.Isartal.	296
...	
Düsseldorf_Mercedesstr.	118
Düsseldorf_Mercedesstraße 6, 40470 Düsseldorf.	221
...	
Hamburg_Große Elbstr.	179
Hamburg_Große Elbstraße 61, 22767 Hamburg.	143
...	
Eschborn_Mergenthalerallee 10 - 12, 65760 Eschborn.	19
Eschborn_Mergenthalerallee 10-12, 65760 Eschborn.	138

Die Auflistung aller Unternehmen, die unter einer bestimmten Anschrift registriert sind, gelingt mit einer Abfrage wie beispielsweise:

```

SELECT c_name
FROM v_company
WHERE c_address = 'Eschborn_Mergenthalerallee 10-12, 65760 Eschborn.'

```

Als Resultat erhalten wir:

c_name
BRND X Sales Solutions GmbH
FPI SW Verwaltungs GmbH
SCD Software Center Deutschland GmbH
Infotexx UG (haftungsbeschränkt)
Mavea Yachts Management GmbH
DSP Immobilienmanager UG (haftungsbeschränkt)
Excelsior GmbH
Kleyer Beteiligungsgesellschaft mbH
WEM Experts GmbH
Scapa Holding GmbH
Hanseatic Venture AG
... usw.

Spannend erscheint die Frage, ob alle oder viele der in einem Gebäude ansässigen Firmen auch denselben Officer haben. Für den Gebäudekomplex 'Eschborn_Mergenthalerallee 10-12, 65760 Esch-

born.' ist das grundsätzlich nicht der Fall. Nur manchmal ist derselbe Officer bei mehreren Unternehmen eingetragen. Den Firmennamen erhält man über einen JOIN-Befehl, der die Daten beider Tabellen zusammenführt:

```
SELECT DISTINCT c_name, o_name
FROM v_company
LEFT JOIN v_officer ON
v_company.c_number = v_officer.c_number
WHERE v_company.c_address = 'Eschborn_Mergenthalerallee 10-12, 65760
Eschborn.'
ORDER BY o_name
```

c_name	o_name
ALFOUZAN Germany Verwaltungs-GmbH	Abdulmohsin Y Al Johaimy
ALSABIQ Germany Verwaltungs-GmbH	Abdulmohsin Y Al Johaimy
ALSABIQ Eschborn GmbH	Abdulmohsin Y Al Johaimy
Excelsior GmbH	Alexander Bauer
NFL Properties Europe GmbH	Alexander Müller
DIJWS Dienstleistungen GmbH	Alexander Richter
DIJWS Dienstleistungen GmbH	Alexander Schleicher
MI Vertriebs GmbH	Alexander Walleczek
ATBO Dienstleistungsgesellschaft mbH	András Zsolt Böröczky
TRE Invest GmbH	Aneta Tartsch
TRE TRACTION REAL ESTATE GMBH	Aneta Tartsch
MST GmbH	Angelika Hundt
Learn To Trade GmbH	Ann Marie Doctor Agius
Anna Lepper UG (haftungsbeschränkt)	Anna Lepper
...	

Für die Anschrift 'Pullach i. Isartal_Emil-Riedl-Weg 6, 82049 Pullach i. Isartal.' wird beispielsweise Herr Christian Floth als Geschäftsführer von sieben Unternehmen genannt. Eine Internetsuche nach diesem Namen führt auf die Floth Real Estate GmbH,²⁹¹ deren Geschäftszweck die Akquirierung und Entwicklung von Immobilien ist. Christian Floth ist der Geschäftsführer und Gründer der Floth Real Estate GmbH.²⁹² Aus dem Unternehmenszweck kann man ableiten, dass die Existenz von mehrfachen Zweckgesellschaften hier keine Besonderheit darstellt und per se keinen Hinweis für besondere steuerliche Risiken gibt.

In eine andere Richtung geht die Frage, welche Personen auffällig häufig als „officer“ in Erscheinung treten. Näheres erfährt man durch den folgenden Befehl:

²⁹¹ <https://floth-realestate.com/>.

²⁹² Vgl. <https://floth-realestate.com/christian-floth/>.

```

SELECT o_name, o_position, o_type, COUNT(c_number)
FROM v_officer
WHERE o_type = 'person'
GROUP BY o_name
ORDER BY COUNT(c_number) DESC

```

Das sind verblüffende Größenordnungen:

o_name	o_position	o_type	Anzahl
Katja Gogalla	Geschäftsführer	person	4124
Antje Borchardt	Geschäftsführer	person	2069
Angelika Hundt	Geschäftsführer	person	1935
Andreas Koglin	Geschäftsführer	person	1328
Christian Goldbrunner	Geschäftsführer	person	1279
Kerstin Zander	Geschäftsführer	person	1232
Julia Vieth	Geschäftsführer	person	1177
Nicole Lotz	Geschäftsführer	person	1118
Heinz Günter Höhne	Geschäftsführer	person	1117
Cornelia Wendt	Geschäftsführer	person	989
Steffen Kurt Holderer	Liquidator	person	980
Nicole geborene Liebera Lotz	Geschäftsführer	person	892
Thomas Doctor Naumann	Liquidator	person	891
Carsten Eckert	Liquidator	person	864
Achim Bönninghaus	Geschäftsführer	person	824
... usw.			

Grundsätzlich man muss sich schon die Frage stellen, wie man realistischer Weise eine Geschäftsführungstätigkeit bei so vielen Unternehmen einnehmen kann. Allerdings lösen sich manche Fälle bei näherem Hinsehen plausibel auf. Spitzenreiterin ist Frau Katja Gogalla mit sage und schreibe 4.124 Geschäftsführungspositionen.

Eine weitere Abfrage ergibt:

```

SELECT *
FROM v_officer
WHERE o_name = 'Katja Gogalla'
ORDER BY o_city

```

Die Ergebnistabelle deutet mit vier verschiedenen Städten an, dass es sich zunächst einmal vermutlich um vier verschiedene Personen handelt:

o_id	o_name	o_position	o_type	c_number	o_city
2085031	Katja Gogalla	Prokurist	person	R2402_HRB7401	Dortmund
2830440	Katja Gogalla	Geschäftsführer	person	R2402_HRB21749	Dortmund
3374822	Katja Gogalla	Geschäftsführer	person	M1201_HRB80658	Frankfurt am Main
4503612	Katja Gogalla	Geschäftsführer	person	M1201_HRB80659	Frankfurt am Main
3710508	Katja Gogalla	Geschäftsführer	person	K1101R_HRB149869	Hamburg
6639	Katja Gogalla	Geschäftsführer	person	M1201_HRB89553	München
7666	Katja Gogalla	Geschäftsführer	person	M1201_HRB87991	München
8054	Katja Gogalla	Geschäftsführer	person	M1201_HRB94131	München
... usw.					alle München

Dominierend bleibt allerdings Katja Gogalla aus München.

Eine Internetrecherche führt zu folgendem Treffer:²⁹³

“Katja Gogalla

Senior Beraterin

Frau Gogalla ist seit dem Jahr 2004 bei der Blitzstart tätig. Sie ist Industrie- und Handelskauffrau (IHK) und verfügt über mehr als 20 Jahre Erfahrung im Vertrieb.“

Die Dame arbeitet also bei der Blitzstart Holding AG aus München, die kommerziell sog. Vorratsgesellschaften als Alternative zur eigenen Unternehmensgründung anbietet.²⁹⁴ Damit wird in diesem Fall die extrem hohe Anzahl von Firmen verständlich, denn die Firmen sind nicht wirklich werbend tätig. Vielmehr „schlafen“ sie eher vor sich hin und warten auf einen passenden Erwerber.

Er versteht sich von selbst, dass solche plausiblen Erklärungen nicht immer vorliegen. In solchen Fällen können Indizien für besondere steuerliche Risiken gegeben sein.

Besonders aufschlussreich könnten solche Auswertungen sein, wenn sie sich nicht auf das deutsche Unternehmensregister beschränken, sondern auch soweit möglich entsprechende **Register anderer Staaten** einbeziehen. Bereits innerhalb der EU gibt es ja mehrere Staaten, die nach weit verbreiteter Einschätzung als „Steuroasen“ gelten (ausführlich dazu Kapitel II., Abschnitt C. 3. c) Standort und Rechtsform auf Seite 26). Hierzu zählen u.a. Luxemburg, Niederlande, Irland, Malta und Zypern. Beispiele für Industrieländer (also keine reinen Steuroasen) außerhalb der EU sind USA, Schweiz, Hongkong und das United Kingdom.

Es liegt auf der Hand, dass die Erschließung ausländischer Datenquellen die Möglichkeiten vervielfältigt. Insbesondere wäre auch ein **Abgleich des deutschen Transparenzregisters** mit den Inhalten ausländischer Register möglich, wenn der unmittelbare Anteilseigner in einem ausländischen Unternehmen besteht.

²⁹³ <https://www.blitzstart.com/de/ueber-uns/team/>.

²⁹⁴ Vgl. <https://www.blitzstart.com/de/>.

2. Linked Leaks und FactForge

Linked Leaks ist eine Datensammlung, die – wie der Name schon andeutet – auf den Enthüllungen zu den Panama Papers aufbaut. Letztere wurden ja vom International Consortium of Investigative Journalists (ICIJ) veröffentlicht (siehe Abschnitt III. B. 5. Steuerleaks auf den Seiten 50ff.).

Die Firma Ontotext hat die öffentlichen Daten der Panama Papers zu Demonstrationszwecken in ihr Graphen-DBMS „GraphDB“ eingelesen.²⁹⁵ Diese Daten wurden jedoch noch um weitere Daten angereicht. Dazu zählen geographische Angaben aus DBPedia und aus GeoNames. Bereits hiermit werden umfangreichere Analysen möglich, da beispielsweise in den Panama Papers erwähnte Städte automatisch ihrem Land zugeordnet werden können.

Linked Leak lässt sich über eine Webseite als SPARQL Endpoint abfragen.²⁹⁶ Alternativ ist auch der Download des kompletten Datensatzes als zip-Datei möglich. Das DBMS „GraphDB Free“ wird ebenfalls kostenlos zur Verfügung gestellt:²⁹⁷

Es handelt sich somit um Linked Open Data:

“Linked Leaks, an Ontotext portal, publishes this data as a knowledge graph, according to the Linked Open Data principles. This allows one to enter the URL identifier of the an entity or a person (say, <http://data.ontotext.com/resource/leaks/entity-123456>) in a web browser to see all the information available in for this entity in the database. Applications can retrieve the relevant information for a resource making HTTP GET with its URL.

Data allows for all sorts of discovery and analytics queries such as:

- *Companies that have more than one shareholder in common with a given one;*
- *Companies related to a given shareholder (be it person or organization), including control relationships;*
- *Companies that control other companies in the same country, through company in an offshore zone;*
- *Most popular offshore jurisdictions.”²⁹⁸*

Die Webseite verweist dazu auf einige Beispielabfragen.²⁹⁹

“LL2: Country pairs by ownership statistics” zeigt für die Panama Papers, welche Länder am häufigsten über Eigentumsverhältnisse verbunden sind. Die SPARQL-Abfrage dazu lautet:³⁰⁰

²⁹⁵ Vgl. hierzu und zum Folgenden <http://data.ontotext.com/linkedleaks>.

²⁹⁶ <http://data.ontotext.com/linkedleaks>.

²⁹⁷ <ftp://ftp.ontotext.com/pub/leaks/rdf/rdf.zip>.

²⁹⁸ <http://data.ontotext.com/linkedleaks>.

²⁹⁹ Siehe auch Code und Erläuterungen auf Github. <https://github.com/Ontotext-AD/leaks>.

³⁰⁰ <http://data.ontotext.com/sparql?savedQueryName=LL2:%20Country%20pairs%20by%20ownership%20statistics&execute=true>.

```

PREFIX onto: <http://www.ontotext.com/>
PREFIX leak: <http://data.ontotext.com/resource/leak/>
PREFIX leaks: <http://data.ontotext.com/resource/leaks/>
SELECT ?owner_country ?owner_country_name ?entity_country ?entity_country_name (COUNT(*) as ?count)
FROM onto:disable-sameAs
{
  ?owner leak:OWNER_TRANSITIVE ?entity .
  ?owner leak:hasCountry ?owner_country .
  FILTER(?owner_country != leaks:country-XXX)
  ?entity leak:hasCountry ?entity_country .
  FILTER(?entity_country NOT IN (leaks:country-XXX, ?owner_country ))
  ?owner_country leak:name ?owner_country_name .
  ?entity_country leak:name ?entity_country_name .
}
GROUP BY ?owner_country ?owner_country_name ?entity_country ?entity_country_name
ORDER BY DESC(?count)

```

Für Deutschland erhält man folgende Ergebnisstabelle:

Land 1	Name 1	Land 2	Name 2	Anzahl
leaks:country-DEU	Germany	leaks:country-VGB	Virgin Islands, British	"176"^^xsd:integer
leaks:country-DEU	Germany	leaks:country-SGP	Singapore	"282"^^xsd:integer
leaks:country-DEU	Germany	leaks:country-IDN	Indonesia	"175"^^xsd:integer
leaks:country-DEU	Germany	leaks:country-HKG	Hong Kong	"109"^^xsd:integer
leaks:country-DEU	Germany	leaks:country-IND	India	"73"^^xsd:integer
leaks:country-DEU	Germany	leaks:country-MYS	Malaysia	"73"^^xsd:integer
leaks:country-DEU	Germany	leaks:country-CHE	Switzerland	"46"^^xsd:integer
leaks:country-DEU	Germany	leaks:country-THA	Thailand	"38"^^xsd:integer
leaks:country-DEU	Germany	leaks:country-CYM	Cayman Islands	"35"^^xsd:integer
leaks:country-DEU	Germany	leaks:country-LUX	Luxembourg	"31"^^xsd:integer
leaks:country-DEU	Germany	leaks:country-RUS	Russian Federation	"24"^^xsd:integer
leaks:country-DEU	Germany	leaks:country-TWN	Taiwan, Province of China	"22"^^xsd:integer
leaks:country-DEU	Germany	leaks:country-GBR	United Kingdom	"19"^^xsd:integer
leaks:country-DEU	Germany	leaks:country-ITA	Italy	"19"^^xsd:integer

Die einzelnen Treffer, die hinter der aggregierten Anzahl stecken, ließen sich selbstverständlich auch auflisten.

Ein Webinar mit dem Titel „Diving in Panama Papers and Open Data to Discover Emerging News“ stellt verschiedene Auswertungsmöglichkeiten dar.³⁰¹ Es wurde aufgezeichnet und ist frei auf Youtube verfügbar.³⁰²

Ein weiteres Demonstrationsbeispiel von Ontotext neben LinkedLeaks ist **FactForge** (siehe auch Abschnitt III. B. 7. auf Seite 55).³⁰³ Hier werden noch umfangreichere Datenquellen integriert wie GLEI, WorldFacts,³⁰⁴ WordNet³⁰⁵ und Now News.³⁰⁶ Die dahinter stehenden Ontologien sind sehr umfangreich und extrem komplex.³⁰⁷ Auch hier gibt es einen SPARQL Endpoint.³⁰⁸

Die SPARQL-Abfrage für das Beispiel “# F05: Suspicious control chain through off-shore company” lautet etwa:

```
PREFIX onto: <http://www.ontotext.com/>
PREFIX fibo-fnd-rel-rel: <http://www.omg.org/spec/EDMC-FIBO/FND/Relations/Relations/>
PREFIX ff-map: <http://factforge.net/ff2016-mapping/>
SELECT *
FROM onto:disable-sameAs
WHERE {
  ?c1 fibo-fnd-rel-rel:controls ?c2 .
  ?c2 fibo-fnd-rel-rel:controls ?c3 .
  ?c1 ff-map:primaryCountry ?c1_country .
  ?c2 ff-map:primaryCountry ?c2_country .
  ?c3 ff-map:primaryCountry ?c1_country .
  FILTER (?c1_country != ?c2_country)
  ?c2_country ff-map:hasOffshoreProvisions true .
}
```

Hier werden Einflussketten aufgelistet, bei denen ein Unternehmen (quasi der Enkel) indirekt von einem anderen Unternehmen (quasi die Mutter) beherrscht wird, wobei ein weiteres Unternehmen (Tochter) mit Sitz in einer Steueroase zwischengeschaltet ist.

³⁰¹ https://www.ontotext.com/knowledgehub/webinars/discover-emerging-news-with-open-data/?utm_source=LinkedLeaks&utm_medium=refferal&utm_campaign=LinkedLeaks.

³⁰² <https://www.youtube.com/watch?v=wKLAAafAdY>.

³⁰³ <https://www.ontotext.com/blog/exploring-linked-open-data-factforge/>.

³⁰⁴ <http://worldfacts.us/>.

³⁰⁵ <https://wordnet.princeton.edu/> für Synonyme.

³⁰⁶ <http://now.ontotext.com> für aktuelle Nachrichten (laufend aktualisiert).

³⁰⁷ Siehe <http://factforge.net/relationships> und <http://factforge.net/hierarchy>.

³⁰⁸ <http://factforge.net/sparql>.

3. Projekt execGraph

Das dritte Beispiel stammt mit der Wirtschaftsprüfung aus einem Gebiet, das nur indirekt mit der steuerlichen Betriebsprüfung verbunden ist. Gemeinsamkeit ist jedoch die Frage, ob Indikatoren für besondere, erhöhte Risiken vorliegen.

Konkreter aktueller Bezug ist die Wirecard AG. Angesichts der Unregelmäßigkeiten der Rechnungslegung und der mangelnden Aufsicht über das deutsche Finanzdienstleistungsunternehmen Wirecard AG, die schließlich in der Insolvenz und der Verhaftung des Vorstandsvorsitzenden Markus Braun endeten, wurden die Forderungen nach einer umfassenderen Finanzaufsicht lauter.³⁰⁹ Journalisten und Wissenschaftler waren erstaunt, wie ein Unternehmen, das im Deutschen Aktienindex (DAX) unter den 30 größten börsennotierten Unternehmen Deutschlands gelistet ist, Aktionäre, Finanzaufsichtsbehörden und die Öffentlichkeit über einen so langen Zeitraum täuschen konnte.³¹⁰

Das Forschungsprojekt „execGraph“ wurde begonnen, um Risikohinweise in zwei Richtungen zu gewinnen:

Um als Abschlussprüfer tätig zu sein, darf keine Besorgnis der Befangenheit vorliegen (§ 328 Abs. 2 HGB), welche die Unabhängigkeit des Wirtschaftsprüfers bzw. der Wirtschaftsprüfungsgesellschaft gefährden könnte. Hierzu hat der Gesetzgeber umfangreiche Regelungen explizit kodifiziert (§ 318 Abs. 3 und 4 HGB, § 319a HGB, § 319b HGB), welche allerdings von Zeit zu Zeit verändert – im Sinne von zunehmend verschärft – wurden. Generell könnten also auch „Beziehungen“ schwächerer Art graduell die Unabhängigkeit schwächen. Hinzu kommt, dass die angewandten Kriterien sich in den allermeisten Fällen nur auf die Gegenwart beziehen.³¹¹ Frühere Ausschlussgründe, die heute nicht mehr bestehen, sind rechtlich unbeachtlich, auch wenn sie möglicherweise noch nachwirken. Die Existenz von schwächeren und/oder vergangenen Beziehungen zwischen dem Abschlussprüfer und den Organen – Vorstände, Aufsichtsräte – des zu prüfenden Unternehmens sollen als erstes im ExecGraph erfasst werden.

Daneben spielt auch die Qualifikation des Aufsichtsrates in seiner Gesamtheit (§ 107 AktG, § 100 Abs. 5, 2. HS), des Prüfungsausschusses bei kapitalmarktorientierten Unternehmen (§ 324 Abs. 1 HGB) sowie des sog. Finanzexperten im Aufsichtsrat (§ 100 Abs. 5, 1. HS) eine wichtige Rolle. Auch hierfür gibt es kodifizierte Mindestanforderungen, die jedoch nicht alleine Maßstab sein können. Der ExecGraph soll daher auch das Qualifikationsgefüge im Aufsichtsrat so gut wie möglich mehrdimensional – Kompetenzportfolio, Erfahrungswissen – abbilden.

Dem Beispiel liegt die Annahme zugrunde, dass Risiken in Hinblick auf die Korrektheit der externen Rechnungslegung auch von Bedeutung für steuerliche Risiken sein können.

Die Aufsicht von Unternehmen steht somit im Mittelpunkt des execGraph. Ziel ist es, Möglichkeiten zur Verbesserung der Rechenschaftspflicht und Transparenz im Bereich der Regulierung zu erkunden.

³⁰⁹ Vgl. Krahen / Langenbucher (2020); Wortham / Liebscher / Christian (2020); Véron (2020).

³¹⁰ Vgl. Alderman / Schuetze (2020); Bartz et al. (2020).

³¹¹ Ausnahme ist etwa die Regelung nach § 319 Abs. 3 Nr. 5 HGB (hoher Honoraranteil von dem Mandanten innerhalb der letzten 5 Jahre).

Fraglich ist, inwiefern Mitglieder des Vorstands und Aufsichtsrats miteinander verknüpft sind. Verknüpfungen können bspw. aufgrund gemeinsamer Tätigkeiten, Mitgliedschaften in denselben Organisationen oder auch gemeinsamer Ausbildung bestehen.

Eine Möglichkeit, diese Informationen zu strukturieren, ist ein Knowledge Graph. Die Knoten stellen reale Entitäten und Ressourcen dar, während die Kanten ihre Beziehung zueinander beschreiben.³¹² Die Stärke eines Wissensgraphen liegt in seiner Fähigkeit, die begrenzten Informationen in der Datenbank durch Links zu externen, im Internet verfügbaren offenen Graphen-Datenbanken anzureichern. Da Wissensgraphen die Beziehungen zwischen Entitäten betonen, ermöglichen sie die Analyse und Visualisierung dieser Beziehungen.

Das Gesamtziel des Projekts ist die Erforschung des Einsatzes von Wissensgraphen als Instrument zur Untersuchung der Beziehungen zwischen Vorständen, Aufsichtsräten und Wirtschaftsprüfern in den deutschen DAX-30-Unternehmen. In diesem Sinne gibt es fünf Teilziele: (1) Datenbankgenerierung (2) Graphenmodellierung (3) Graphengenerierung (4) Graphenanalyse und (5) Graphenvisualisierung.

Die *Datenbankgenerierung* umfasst das Bereinigen vorhandener Daten, das Extrahieren neuer Daten aus verschiedenen Quellen und das Erstellen einer zentralen Datenbank, die den üblichen Datenbanknormalisierungsregeln entspricht und als Grundlage für den Wissensgraphen dient. Die *Graphenmodellierung* umfasst den Entwurf eines Modells auf der Grundlage der verfügbaren Daten, das eine sinnvolle Durchquerung und Analyse des Graphen ermöglicht. Die *Graphengenerierung* beschreibt eine Instanziierung des Graphenmodells. Die *Graphenanalyse* beschreibt alle Aufgaben, die mit der Suche nach aussagekräftigen Erkenntnissen und der Evaluierung des Projekts selbst zusammenhängen. Die *Graphvisualisierung* zielt darauf ab, eine visuelle Darstellung des Graphen zu liefern, die Interaktion und Navigation ermöglicht. Das Ziel ist die Verwendung von Programmen (VisiNav und Tarsier), die eine grafische Benutzeroberfläche (GUI) zur Verfügung stellen und somit die Einstiegshürden für Benutzer mit weniger technischem Know-how beseitigen.

Speicherung der Daten

Daten in einem Excel- oder csv.-Format sind für den Knowledge Graph nicht geeignet. Dementsprechend müssen die Daten erst normalisiert werden. Der Schwerpunkt liegt dabei auf der Benutzerfreundlichkeit für Benutzer ohne technisches Fachwissen, indem MS Excel als Speicherformat beibehalten wird, während gleichzeitig versucht wird, die in relationalen Datenbanken üblichen Prinzipien zu befolgen. Folglich wird eine einzige MS Excel-Datei erstellt, die alle erforderlichen Informationen in 11 Registerkarten enthält, wie in Tabelle 2 zu sehen ist. Allen Registerkarten, die sich auf Ressourcen beziehen (in Tabelle 2 blau hervorgehoben), wird eine eindeutige ID zugewiesen, die bei der Erstellung des Diagramms in eine URI umgewandelt wird, sowie gegebenenfalls Links zu externen Datenbanken (z. B. *wikidata_sameAs*).

³¹² Vgl. Ehrlinger / Wöß (2016).

Registerkarte	Beschreibung
Person	Weist jeder Person (<i>personEntity</i>) eine eindeutige ID zu. Enthält persönliche Informationen zu jeder Person im Diagramm (z. B. Name, Geburtsjahr).
Unternehmen	Weist jedem Unternehmen eine eindeutige ID zu (<i>companyEntity</i>). Enthält Informationen zu jedem Unternehmen im Diagramm (z. B. Name, ISIN, DAX-Mitgliedschaft).
Universität	Weist jeder Hochschule eine eindeutige ID zu (<i>universityEntity</i>). Enthält Informationen zu jeder Hochschule im Diagramm (z. B. Name, Link zu anderen Datenbanken).
Städte	Weist jeder Stadt eine eindeutige ID zu (<i>cityEntity</i>). Enthält Informationen zu jeder Stadt im Diagramm (z. B. Name, Link zu anderen Datenbanken).
Länder	Weist jedem Land eine eindeutige ID zu (<i>countryEntity</i>). Enthält Informationen zu jedem Land im Diagramm (z. B. Name, Link zu anderen Datenbanken).
Position	Enthält Informationen über Mitarbeiterpositionen, die eine Person in ihrer Karriere in einem Unternehmen innehatte oder derzeit innehat. Dient als Bindeglied zwischen einem Unternehmen und einer Person.
Bildung	Enthält Informationen über Bildungsprogramme oder Abschlüsse, die eine Person erworben hat. Dient als Bindeglied zwischen einer Universität und einer Person.
Prüfung	Enthält Informationen über Prüfungen, die ein Unternehmen durchlaufen hat. Dient als Bindeglied zwischen einer Wirtschaftsprüfungsgesellschaft, dem geprüften Unternehmen und den Wirtschaftsprüfern (einer Person).
Geschäfte	Enthält Informationen über Eigengeschäfte, die eine Person mit einem Unternehmen getätigt hat. Dient als Verbindung zwischen einem Unternehmen und einer Person.
Beziehungen	Enthält Informationen über Beziehungen zwischen Personen (z. B. Ehepartner, Geschwister). Dient als Verbindung zwischen zwei Personen.
<i>Organisation</i>	<i>Wurde aufgrund der geringen Datenmenge eingestellt: Enthält Informationen über die Mitgliedschaft einer Person in Organisationen. Dient als Bindeglied zwischen einer Person und einer Organisation.</i>

Abb. 15: Übersicht der Registerkarten der Datenbank

Eigene Abbildung

Herkunft der Daten

Die Daten wurden aus einer Vielzahl von Quellen extrahiert, wobei die meisten Daten von den DAX-Unternehmen selbst zur Verfügung gestellt wurden, um Transparenz für ihre Aktionäre zu schaffen. Allerdings waren die Lebensläufe teilweise direkt auf der Webseite oder im PDF-Format sowie strukturiert (Liste aller Positionen) oder unstrukturiert (reine Textbeschreibung) enthalten. Daher ist es

eine Herausforderung, Daten zu scrapen oder den Prozess zu automatisieren. Zusätzlich werden offene Graph-Datenbanken wie DBpedia³¹³ und WikiData³¹⁴ genutzt, allerdings waren die Informationen hier nur unvollständig oder veraltet. Außerdem werden Daten aus einer Bloomberg-Datenbank entnommen. Eine weitere wertvolle Datenquelle ist LinkedIn, wo die Personen selbst Informationen bereitstellten und pflegen.

Umfang der Daten

Die Datenbasis besteht aus 745 Personen darunter Führungskräfte, Aufsichtsratsmitglieder und Wirtschaftsprüfer. Mit Hilfe von OpenRefine³¹⁵ kann etwa ein Drittel (32 %) der Personen mit Wikidata verknüpft werden. Darüber hinaus finden sich 1.203 Unternehmen in der Datenbank. Hier kann mehr als die Hälfte (53 %) mit Wikidata verknüpft werden. Insgesamt umfasst unsere Datenbank 5.116 Stellen, von denen 1.830 im Dezember 2020 besetzt waren, sowie 1.128 Bildungsabschlüsse (zu den Formen siehe **Abb. 16**) und Studiengänge/Bildungskategorien (siehe **Abb. 17**) an 341 Hochschulen, die von den gesuchten Personen absolviert wurden.

Grad
Hauptschule
Mittlere Reife
Abitur
Ausbildung
Bachelor
Diplom
MBA
Meister
Staatsexamen
PhD
Weiterbildung

Abb. 16: Abschlüsse

Eigene Abbildung

³¹³ <https://wiki.dbpedia.org/>.

³¹⁴ https://www.wikidata.org/wiki/Wikidata:Main_Page.

³¹⁵ <https://openrefine.org/>.

Feld 1/ Feld 3	Feld 2/ Feld 4
Wirtschaftswissenschaften	VWL BWL FACT Betriebswirtschaftslehre Kommunikationswissenschaft Marketing Wirtschaft
Ingenieurwissenschaften	Elektrotechnik Informatik Maschinenbau Maschinenbau Luft- und Raumfahrttechnik Agrartechnik
Naturwissenschaften	Biologie Chemie Physik Mathematik Medizin
Geisteswissenschaften	Theologie Philosophie Journalismus Politikwissenschaften Soziologie Geschichte Psychologie Anglistik Kunst Sport
Rechtswissenschaften	
Pädagogik	
Linguistik	Englisch, etc.
Andere	

Abb. 17: Studiengänge/Bildungskategorien

Eigene Abbildung

Verknüpfung der Daten

Während die Verknüpfung von Entitäten im Graphen mit Datenbanken wie WikiData oder DBpedia manuell erfolgen kann, ist OpenRefine ein nützliches Werkzeug, um diesen Schritt zu automatisieren. Nachdem eine Excel- oder csv-Datei geöffnet wurde, kann ein Projekt erstellt werden. OpenRefine gleicht dann die Namensstrings in ihrer Datei mit den Namen der Entitäten in der ausgewählten Datenbank ab.

Graph Modellierung

Die Erstellung des Graphen erfolgte mit tarql³¹⁶, einem in Java geschriebenen Kommandozeilen-Tool zur Konvertierung von CSV-Dateien in RDF mit SPARQL-Syntax. Tarql benötigt drei Dinge: (1) einen Befehl, (2) eine csv-Datei und (3) eine SPARQL-Konstruktabfrage.

Der Befehl weist tarql im Wesentlichen an, eine bestimmte SPARQL-Konstruktabfrage für eine bestimmte csv-Datei durchzuführen. Variablen ('?example') werden mit den Spaltennamen in der csv-Datei abgeglichen. Um den Befehl auszuführen, muss zuerst die Kommandozeile im tarql-1.2 Ordner geöffnet und das tarql Unix Executable (siehe Pfad in Beispiel 3) ausgeführt werden.

Die csv-Datei kann für jede Registerkarte in MS Excel erstellt werden, allerdings traten mehrere Probleme bei der Formatierung auf. Es ist wichtig, dass keine Bezeichner (BeispielEntity-Spalten) Sonderzeichen (,ü', ,ö', ,ä', ,ß' oder Leerzeichen dazwischen) enthalten und dass keine Zelle Doppelpunkte oder Semikolons enthält. Wir haben einen Workaround für die Konvertierung in UTF-8 csv-Dateien gefunden, der bisher gut funktioniert hat. Man erstellt einfach eine leere Google-Tabelle und fügt alle Daten aus einer Registerkarte als Werte ein und speichert die Tabelle als csv-Datei.

```
sh target/appassembler/bin/tarql -ntriples ../Erstellung/Person.rq ../Erstellung/Person.csv
```

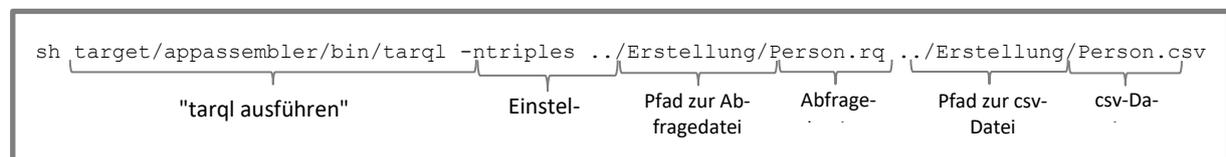


Abb. 18: tarql Befehl

Eigene Abbildung

Die SPARQL-Konstruktabfragen können in jedem Texteditor geschrieben und als .rq-Datei gespeichert werden. Zu Beginn kann es hilfreich sein, einen Blick auf etablierte Abfragen zu werfen, um ein Verständnis für Anforderungen und Syntax zu bekommen, wie in Beispiel 4 zu sehen ist.

³¹⁶ <https://tarql.github.io/>.

```

PREFIX ex: <http://execgraph.org/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX-Schema: <http://schema.org/>
PREFIX dbpprop: <http://dbpedia.org/property/>

KONSTRUKTION {
    ?person ex:hasRole [
        eine schema:EmployeeRole;
        schema:roleName ?jobTitle;
        schema:startDate ?startgYear;
        schema:endDate ?endgYear;
        ex:roleWith ?company;
        ex:inDivision ?divisonName;
        ex:isCurrent ?currentBool;
        ex:isSupervisoryBoard ?positionARBool;
        ex:isManagementBoard ?positionVOBool
    ]
}
FROM <Datei:Position.csv>
WHERE {
    BIND (URI(CONCAT('http://execgraph.org/', ?positionID)) AS ?position)
    BIND (URI(CONCAT('http://execgraph.org/', ?personEntity)) AS ?person)
    BIND (URI(CONCAT('http://execgraph.org/', ?entityCompany)) AS ?company)
    BIND (xsd:gYear(?startYear) AS ?startgYear)
    BIND (xsd:gYear(?endYear) AS ?endgYear)
    BIND (xsd:boolean(?current) AS ?currentBool)
    BIND (xsd:boolean(?positionAR) AS ?positionARBool)
    BIND (xsd:boolean(?positionVO) AS ?positionVOBool)
}

```

Abb. 19: SPARQL-Konstrukt-Abfrage

Eigene Abbildung

Nach dem Ausführen des Befehls werden die Tripel automatisch generiert und können kopiert und eingefügt werden. Durch Einfügen aller Tripel aus allen Registerkarten haben wir dann den fertigen Graphen erstellt. Das Ergebnis wird als Turtle-Datei (.ttl) gespeichert.

Graphische Analyse

Die Abfrage `birthPlaces_map` veranschaulicht die Nutzung offener Graphen-Datenbanken über unsere eigenen Daten hinaus, da sie Stadtkoordinaten aus Wikidata abfragt und uns somit erlaubt, Geburtsorte von DAX-Mitgliedern wie unten in **Abb. 20** gezeigt abzubilden. In Hinblick auf die Corporate Governance kann ein gewisses Ausmaß an geographischer – insbesondere internationaler – Diversität dazu beitragen, einseitig verengte Blickwinkel zu vermeiden. Sehr eng örtlich konzentrierte Herkunftsorte können auf persönliche Beziehungen hindeuten, die der geforderten Unabhängigkeit zuwiderlaufen. Es wären selbstverständlich Ausschnitte der Karte vergrößerbar. Auch ließen sich aus den Stadtkoordinaten die Entfernungen berechnen ohne dass eine Karte dargestellt werden muss

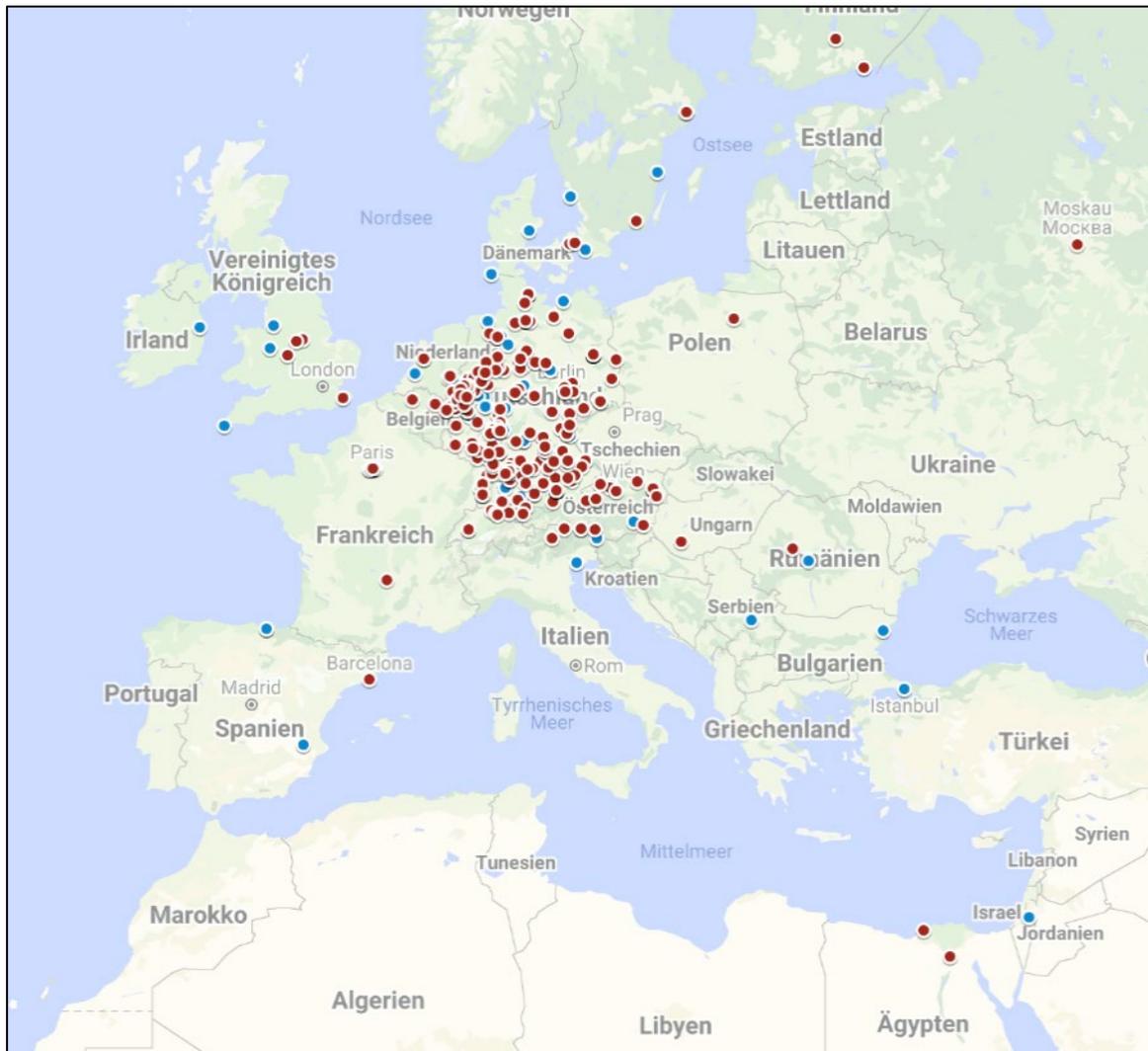


Abb. 20: Karte mit Geburtsorten der DAX-Vorstände (blau) und Aufsichtsräte (rot)
Eigene Abbildung

Visualisierung

Gephi³¹⁷ ist ein intuitives und bewährtes Werkzeug zur Visualisierung von Graphen und bietet eine Vielzahl verschiedener Funktionen zur Netzwerkanalyse. Der Nachteil ist, dass das Tool nicht sehr gut skalierbar ist und Schwierigkeiten bei der Darstellung großer Diagramme hat. Zunächst muss der Turtle-Graph (oder Subgraph) in das .gexf-Format konvertiert werden. Ein Beispiel, wie das geht, finden Sie in den Skripten mit der Endung `_to_gexf` in `execgraph/python_conversion`. In Gephi können die .gexf-Dateien importiert und in verschiedenen Formen (z.B. Graphvisualisierung, Datentabelle) und personalisiert (z.B. Farbschema, Knotengröße) dargestellt werden. Darüber hinaus können verschiedene Metriken zur Netzwerkanalyse berechnet werden, wobei die nützlichste wahrscheinlich PageRank ist. Die von Page et al. (1999) vorgeschlagene PageRank-Metrik berechnet die wahrgenommene Wichtigkeit eines Knotens in einem gerichteten Graphen und kann durch den folgenden vereinfachten Algorithmus ausgedrückt werden:

³¹⁷ <https://gephi.org/>.

$$R(u) = c \sum_{v \in B_u} \frac{R(v)}{N_v}$$

Wenn u ein Knoten im Graphen ist, ist v ein Knoten in der Menge B_u ist, die alle Knoten enthält, die auf die Seite u verlinken, und N_v die Anzahl der Links vom Knoten v ist. Der PageRank des Knotens u wird dann iterativ berechnet, indem der PageRank jedes Knotens, der auf den Knoten u verlinkt, geteilt durch seine ausgehenden Links summiert wird. Dabei werden Links von Knoten mit höherem PageRank stärker gewichtet als Links von solchen mit niedrigerem PageRank. Der anfängliche PageRank-Wert für jeden Knoten in Gephi wird auf $\frac{1}{n}$ gesetzt, wobei n die Anzahl der Knoten im Graphen ist.

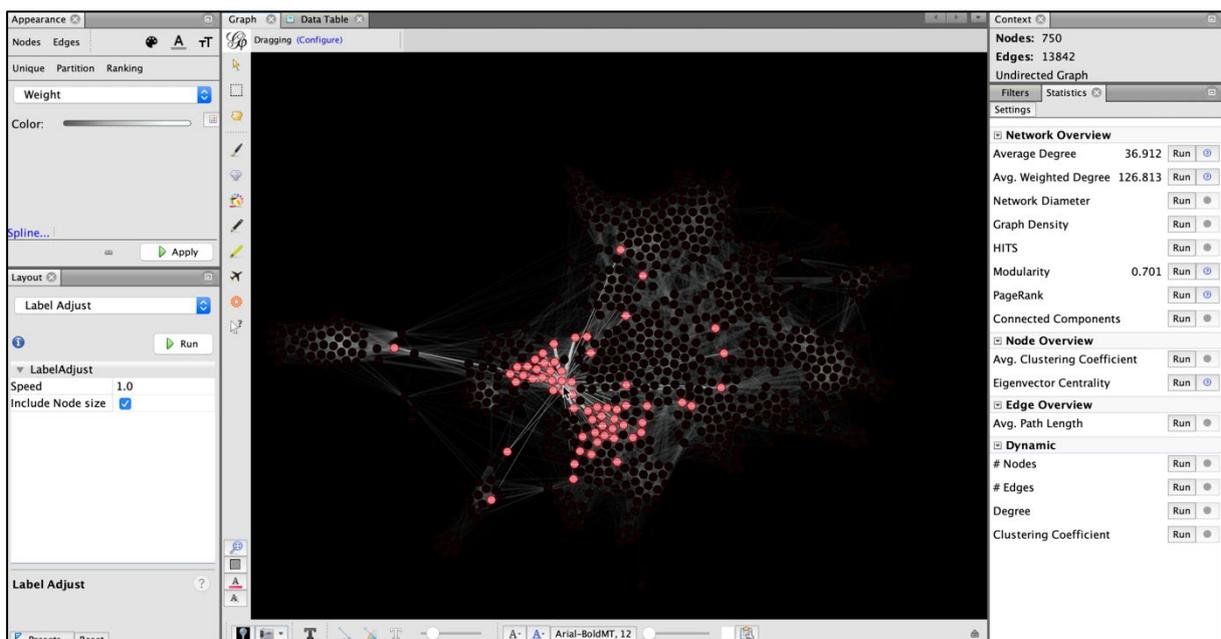


Abb. 21: Screenshot der Gephi-Visualisierung
Eigene Abbildung

Graphistry³¹⁸ ist ein fortschrittlicheres Tool zur Visualisierung von Graphen, das in Python verwendet werden kann und die beste Leistung der in diesem Abschnitt genannten Tools aufweist. Der Nachteil ist, dass das Tool nicht so viele Funktionen zur Netzwerkanalyse bietet wie Gephi. Graphistry bietet allerdings die Möglichkeit einer dynamischen Visualisierung.

Tarsier³¹⁹ bietet eine weitere Möglichkeit, RDG Knowledge Graphs in 3D zu visualisieren. Der Vorteil von Tarsier liegt darin, dass verschiedene Ebenen erstellt werden können. In unserem Beispiel etwa für:

- Unternehmen
- Aufsichtsrat
- Vorstand

³¹⁸ <https://www.graphistry.com/>.

³¹⁹ Entwickelt von Fabio Viola, Universität Bologna.

- Wirtschaftsprüfer.

Auswertungen können auf diese Weise vielfältig vorgenommen werden. Es können bspw. die folgenden Verflechtungen vorliegen:

- Tätigkeit in mehr als einem Gremium (z.B. Aufsichtsrat zeitgleich bei 2 (konkurrierenden) Unternehmen).
- Mitglieder eines Gremiums oder Wirtschaftsprüfer und Aufsichtsratsmitglied studierten zeitgleich an selber Universität.
- Ehemaliger Wirtschaftsprüfer einer Gesellschaft wechselt in Aufsichtsrat oder Vorstand der Gesellschaft.
- Verwandtschaften oder private Freundschaften zwischen Vorstand und Aufsichtsrat bzw. Vorstand und Wirtschaftsprüfer bzw. Aufsichtsrat und Wirtschaftsprüfer oder innerhalb eines Gremiums.
- Erfahrung der Mitglieder in Position in anderer Gesellschaft oder anderem Gremium.
- Aufsichtsrat oder Vorstand war ehemaliger Vorstand oder Aufsichtsrat in der gleichen Gesellschaft.
- Wurden Personen wegen ihrer Eignung/Qualifikation eingestellt oder möglicherweise aufgrund personeller Verflechtung.
- Gleiche Nationalität, Herkunft oder Geburtsort von Aufsichtsräten, Vorständen oder Wirtschaftsprüfern.

Die **Analyse der Wirecard AG** führt zu einigen aufschlussreichen Erkenntnissen.³²⁰ Wirecard hatte im Vergleich zu anderen DAX-Gesellschaften wie beispielsweise der Allianz SE nur eine kleine Anzahl von Verbindungen zu anderen Unternehmen. Es war eine Art „closed shop“. Über die Mitglieder von Vorstandes war nur wenig öffentlich bekannt. So enthält die Datenbank zum Wirecard-Vorstand Jan Marsalek nur 48 Triples, zu Oliver Bäte, Vorstandsvorsitzender der Allianz, hingegen 123 Triples. Der Aufsichtsrat von Wirecard umfasste nur wenige Personen (anfangs drei, später fünf, dann sechs). Diese hatten keine Aufsichtsratserfahrung in (anderen) DAX30-Unternehmen und ganz überwiegend zeitlich nur sehr kurze Aufsichtsratserfahrung (). Ein Mitglied verließ den Aufsichtsrat bereits nach einem Jahr wieder, was Fragen nach dem Grund hierfür aufwirft.

Die umseitige Abbildung zeigt die allgemeine **Struktur des Graphen**, unabhängig von Wirecard.

Die Datensammlung wird laufend erweitert. Ein **probeweiser Zugriff** (Betaversion) auf Daten des execGraph ist möglich unter: <http://tc.ontologycentral.com/execgraph/>.

³²⁰ Grümmer (2021).

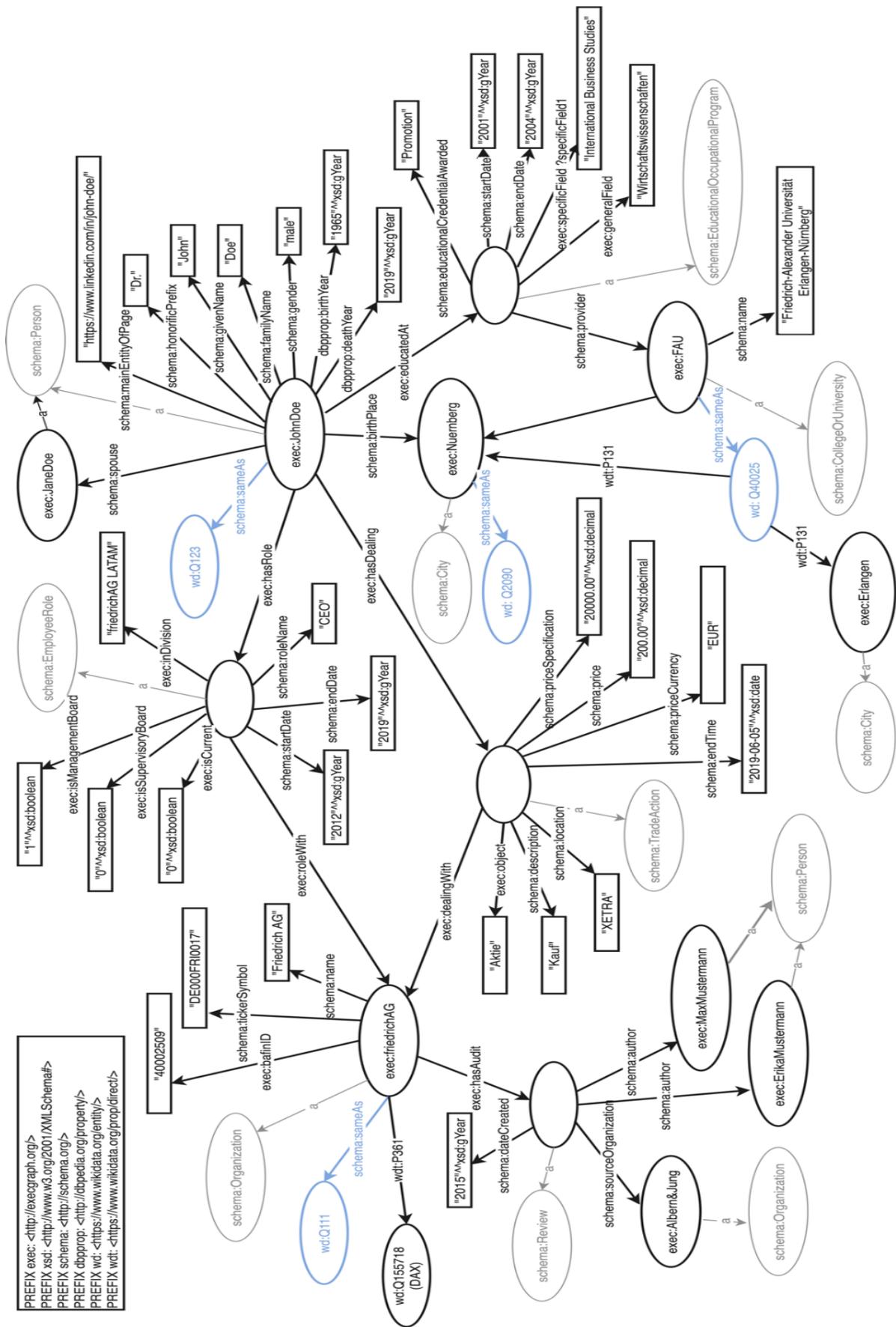


Abb. 22: Struktur des execGraph
Eigene Abbildung

V. Zusammenfassung und Ausblick

Die wesentlichen Erkenntnisse dieser Arbeit lassen sich in den folgenden Thesen zusammenfassen:

(1) Es gibt bereits heute umfangreiche technische Möglichkeiten zur Nutzung von Alternativen Daten, die bei weitem noch nicht ausgeschöpft werden. Da der Rechtsrahmen langfristig als veränderbar angesehen wird, blendet die vorliegende Untersuchung juristische Fragen aus. Augenscheinliche Einsatzmöglichkeiten bestehen einerseits in Konsistenzprüfungen (der bislang intern über die Steuerpflichtigen vorliegenden Daten mit externen Informationen aus den Alternativen Daten), andererseits in der Identifikation von Risikobereichen, bei denen „Warnsignale“ (Red Flags) für versehentlich oder gar absichtlich falsche bzw. fehlende Steuererklärungen vorliegen.

(2) Für Konsistenzprüfungen (etwa von CbCR Berichten) lassen sich außerhalb der Finanzverwaltung keine konkreten Auswertungen zeigen, da die Daten dem Steuergeheimnis unterliegen. Jedoch illustrieren drei beispielhafte Auswertungen das Potenzial für die Gewinnung von „Warnsignalen“ aus Alternativen Daten.

(3) Die Beispiele verdeutlichen jedoch auch, dass es sich häufig um „unscharfe“ Daten handelt. Probleme für die Auswertung entstehen aufgrund von lückenhaften Angaben (z.B. nur die Straße ohne Hausnummer als Adresse), verschiedenen Schreibweisen (z.B. „Manfred Mustermann“, „M. Mustermann“, „Dr. M. Mustermann“, „Manfred G. Mustermann“, Mustermann Manfred“, „Manni Mustermann“), Tippfehlern („Manferd Mustermann“), Namensgleichheit unterschiedlicher Personen (viele Datensätze haben keinen eindeutigen Identifier/Primärschlüssel) und vieles andere.

(4) Die Verknüpfung mehrerer verschiedener Datensätze wird ebenfalls dadurch behindert, dass – selbst wenn sie jeweils einen eindeutigen Identifier besitzen – diese Identifier/Primärschlüssel nicht dieselben sind. Eine wechselseitige Zuordnung ist daher nur mehr oder weniger genau über Hilfsgrößen (wie z.B. die Kombination von Namen und Geburtsdatum) möglich.

(5) Die Aufwertung des Transparenzregisters zu einem Vollregister – also unter Einbezug derjenigen Daten, die eigentlich schon an anderer Stelle verfügbar sein müssten – ist ein wichtiger Schritt. Allerdings gelten aktuell noch Übergangsfristen. Grundsätzlich bestehen bleibt das Problem, dass die Einträge aus Selbstauskünften der Berechtigten beruhen. Immerhin wird die Transparenz erhöht.

(6) Geldwäscheverdachtsmeldungen werden normalerweise nicht öffentlich bekannt. Indes belegen partielle Enthüllungen, dass dieser Datenschatz allenfalls sehr bruchstückhaft gehoben wird. Es besteht anscheinend ein gravierendes Vollzugsdefizit.

(7) Bedauerlich ist, dass bereits innerhalb Deutschlands die öffentlichen Register nicht die Bedingungen für „Offene Daten“ im Sinne der Open Knowledge Foundation erfüllen: *Data available as a whole; at no more than a reasonable reproduction cost; in a convenient and modifiable form*. Abfragen lassen sich in der Regel nur einzelne Datensätze und damit entstehen erhebliche Kosten. Die Tatsache, dass öffentliche Register von einem privaten (privatisierten!) Unternehmen betrieben werden, ist ein Schildbürgerstreich ersten Grades. Dies sollte dringend geändert werden, wobei realistischer Weise höchstens die rechtlichen Entwicklungen auf EU-Ebene Anlass zu Hoffnung bietet.

(8) Entsprechende öffentliche Registerdaten wären zumindest in der gesamten EU erforderlich. Das würde die Datenerhebungen für die Finanzverwaltungen aller Staaten erheblich erleichtern und zielgerichteter machen. Immerhin bieten Staaten wie Luxemburg, Niederlande, Irland, Malta und Zypern bewusst Möglichkeiten für legale Steuergestaltungen an, die aber auch in illegaler Weise missbraucht

werden können. Nicht mehr Mitglied der EU ist Großbritannien. Allerdings erlaubt das dortige „Companies House“ bislang sehr weitreichende Datenzugriffe.

(9) Je weniger Offene Daten vorliegen, desto aufwändiger und potentiell fehlerträchtiger ist die Sammlung und Aufbereitung der Daten. Diese Arbeiten kann die Finanzverwaltung natürlich selbst durchführen. Jedoch wäre für bestimmte Daten der Zukauf von schon fertig aufbereiteten Datensammlungen von kommerziellen Anbietern eine Alternative. Dieser Markt ist stark im Wachsen begriffen. Sollte es nicht möglich sein, komplette Datensätze zu kaufen, so stellt *Robotic Process Automation* (RPA) einen Weg für massenhafte automatisierte Einzelabfragen dar.

(10) Für die Sammlung, Integration und Analyse der Alternativen Daten bietet es sich an, auf etablierte und erprobte internationale Standards zurückzugreifen. Nach Auffassung der Autoren eignet sich am besten eine flexible Graphen-Datenbank in der Version eines RDF-Stores. RDF gehört zu den technischen Standards der EU als Element ihrer Digitalisierungsstrategie. Das RDF-Konzept ist auch Teil der internationalen Standards des W3C (World Wide Web Consortium) für das Internet. Im Unterschied zu (vielen) privaten Datenformaten fallen hierbei keine Lizenzgebühren an.

(11) Es lassen sich im Zeitvergleich drei eindeutige Trends erkennen: Erstens nimmt die Menge der potentiell verfügbaren Alternativen Daten permanent und exponentiell zu. Zweitens gibt es verstärkte Bemühungen staatliche Daten als Offene Daten bereit zu stellen. Und drittens greift der Gedanke von „verknüpften“ Linked Open Data langsam um sich. Damit ist ein wohl unumkehrbarer Prozess in Gang gesetzt, der die Nutzbarkeit von Alternativen Daten für die Betriebsprüfung stetig ausweiten wird.

(12) Jedoch müssen diese Potenziale auch gehoben werden. Dafür braucht es den Willen, ein entsprechendes System von Menschen und Technik einzurichten. Und last not least, auch Kapazitäten und Knowhow seitens der Betriebsprüfer (und Steuerfahnder), den entdeckten Verdachtsmomenten nachzugehen.

Literatur

AADE (2020): myDATAElectronic Books AADE - Technical documentation of REST API interface for submitting and retrieving data Version 0.6 – March 2020, https://www.aade.gr/sites/default/files/2020-04/myDATA%20API%20Documentation%20v0%206b_eng.pdf (8.9.2020).

Alderman, Liz / Schuetze, Christopher (2020): In a German Tech Giant's Fall, Charges of Lies, Spies and Missing Billions, The New York Times vom 26.06.2020, <https://www.nytimes.com/2020/06/26/business/wirecard-collapse-markus-braun.html> (01.07.2020).

Alexiev, Vladimir (2019): euBusinessGraph Semantic Data Model, Version 1.4., https://docs.google.com/document/d/1dhMOTIIOC6dOK_jksJRX0CB-GIRoiYY6fWtCnZArUUhU/edit#heading=h.dskoqyysgc2l (17.9.2021).

Allemang, Dean / Hendler, Jim (2011): Semantic Web for the Working Ontologist, 2. Aufl., Morgan Kaufmann, Waltham 2011.

Amel-Zadeh, Amir / Serafeim, George (2018): Why and How Investors Use ESG Information: Evidence from a Global Survey, Financial Analysts Journal, Vol. 74(3), S. 87-103.

Arnaout, Hiba / Elbassuoni, Shady (2018): Effective searching of RDF knowledge graphs, Journal of Web Semantics 48, S.66-84.

Auer, Sören / Pellegrini, Tassilo / Sack, Harald (2014): Linked Enterprise Data: Management und Bewirtschaftung vernetzter Unternehmensdaten mit Semantic Web Technologien. Springer-Verlag.

Balbierer, Thomas et al. (2021): Neues Steueroasen-Leak belastet Hunderte Politiker, 3.10.2021, <https://projekte.sueddeutsche.de/artikel/politik/pandora-papers-geheimgeschaefte-von-politikern-enttarnt-e500259/> (8.10.2021).

Balogh, Krisztian (2018): Entity-Oriented Search, Springer Open Verlag, <https://rd.springer.com/content/pdf/10.1007%2F978-3-319-93935-3.pdf> (22.9.2021).

Bartz, Tim / Böcking, David / Hesse, Martin / Traufetter, Gerald (2020): Wirecard-Skandal: Wirtschaftsprüfer EY und Aufsichtsbehörde machen sich gegenseitig Vorwürfe, Der Spiegel vom 09.12.2020, <https://www.spiegel.de/wirtschaft/unternehmen/wirecard-skandal-wirtschaftspruefer-ey-und-aufsichtsbehoerde-machen-sich-gegenseitig-vorwuerfe-a-f011387f-8e19-4c82-99f7-037b27076b3f> (23.12.2020).

Beck, Wolf (2021): Reform des GWG - Transparenzregister wird zum Vollregister, Haufe News vom 27.7.2021, https://www.haufe.de/compliance/recht-politik/reform-des-gwg-transparenzregister-wird-vollregister_230132_548132.html (17.9.2021).

Becker, Markus / Diehl, Jörg / Traufetter, Gerald (2021): Scholz übergibt Warnungen zu Geldwäsche-Fahndern, Spiegel Online vom 10.9.2021, <https://www.spiegel.de/wirtschaft/zoll-spezialeinheit-fiu-olaf-scholz-uebergibt-warnungen-zu-geldwaesche-fahndern-a-4f9ef6fc-faf6-48e1-9a25-7a3350a526ae> (11.9.2021).

Berliner Senat, Verordnung zur Bereitstellung von allgemein zugänglichen Datenbeständen (Open Data) durch die Behörden der Berliner Verwaltung (Open Data Verordnung - OpenDataV) vom 24. Juli 2020.

Berlinonline (2021): Das Berliner Open-Data-Handbuch, <https://berlinonline.github.io/open-data-handbuch/> (13.3.2021).

Berners-Lee, Tim (2009): Linked Data, <https://www.w3.org/DesignIssues/LinkedData.html> (31.7.2021).

Bittner, Thomas / Dawid, Roman / Metzner, Susann (2016): Typische Problemfelder in Betriebsprüfungen. In: Dawid, Roman (Hrsg.) Verrechnungspreise. Springer Gabler, Wiesbaden 2016, https://doi.org/10.1007/978-3-658-09377-8_6 (25.9.2021)

BMF (2019): Erste Nationale Risikoanalyse. Bekämpfung von Geldwäsche und Terrorismusfinanzierung. 2018/2019, https://www.bundesfinanzministerium.de/Content/DE/Downloads/Broschueren_Bestellservice/2019-10-19-erste-nationale-risikoanalyse_2018-2019.pdf?__blob=publicationFile&v=17 (18.9.2021).

BMF, Schreiben zur Mitteilung grenzüberschreitender Steuergestaltungen vom 29.3.2021, V A 3 - S 0304/19/10006 :010.

Bundesrechnungshof (2020): Bericht nach § 99 BHO über Maßnahmen zur Verbesserung der Umsatzsteuerbetrugsbekämpfung – Chancen der Digitalisierung nutzen, <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/sonderberichte/langfassungen-ab-2013/2020/massnahmen-zur-verbesserung-der-umsatzsteuerbetrugsbekämpfung-chancen-der-digitalisierung-nutzen> Bonn, 29.10.2020.

Bundeszentralamt für Steuern (2020): Aktuelle Liste der teilnehmenden Staaten. Stand: 1. Juli 2020, https://www.bzst.de/SharedDocs/Downloads/DE/CRS/crs_teilnehmende_staaten_2020.pdf?__blob=publicationFile&v=12 (8.3.2021).

Bundeszentralamt für Steuern (2021a): Common Reporting Standard – CRS, https://www.bzst.de/DE/Privatpersonen/Selbstauskuenfte/CommonReportingStandard/commonreportingstandard_node.html (8.3.2021).

Bundeszentralamt für Steuern (2021b): Kommunikationshandbuch - Automatischer Austausch von Steuergestaltungen (DAC 6), https://www.bzst.de/DE/Unternehmen/Intern_Informationsaustausch/DAC6/Handbuecher/handbuecher.html#js-toc-entry1 (25.9.2021).

Cloudflare (2021): How CAPTCHAs work | What does CAPTCHA mean?, <https://www.cloudflare.com/de-de/learning/bots/how-captchas-work/> (23.10.2021).

Cohen, Jeffrey / Holder-Webb, Lori / Khalil, Samer (2017): A Further Examination of the Impact of Corporate Social Responsibility and Governance on Investment Decisions, Journal of Business Ethics, Vol. 146, S. 203-218.

COSO (Committee of Sponsoring Organizations of the Treadway Commission) (2013): Internal Control - Integrated Framework (updated), <https://www.coso.org> (12.3.2021).

Creative Commons (2021a): About The Licenses, <https://creativecommons.org/licenses/?lang=en> (31.7.2021).

Creative Commons (2021b): What is Creative Commons and what do you do?, <https://creativecommons.org/faq/#what-is-creative-commons-and-what-do-you-do> (31.7.2021).

Christian, Christopher / Liebscher, Marc / Wortham, Leah (2020): Wirecard, Europe's Enron? - Auditor Liability to Investors Corporate Fraud, Other Lectures & Events. 33, https://scholarship.law.edu/other_lectures/33/ (13.12.2020).

Dhaliwal, Dan S. / Li, Oliver Zhen / Tsang, Albert / Yang, Yong George (2011): Voluntary Nonfinancial Disclosure and the Cost of Equity Capital: The Initiation of Corporate Social Responsibility Reporting, The Accounting Review, Vol. 86(1), S. 59-100.

Diehl, Jörg (2021): Länder erneuern Kritik an Zoll-Spezialeinheit, <https://www.spiegel.de/panorama/justiz/zoll-spezialeinheit-fiu-kampf-gegen-geldwaesche-laender-erneuern-kritik-a-48f32de2-2288-4c34-88bf-a786ef17d16d> (8.10.2021).

Dietrich, Daniel (2011): Was sind offene Daten, <http://www.bpb.de/gesellschaft/digitales/open-data/64055/was-sind-offene-daten> (12.3.2021).

Drüen Klaus-Dieter in Tipke/Kruse, AO/FGO, 154. EL 10.2018, AO § 138a Rz. 14;

Eccles, Robert G. / Kastrapeli, Mirtha D. / Potter, Stephanie J. (2017): How to Integrate ESG into Investment Decision-Making: Results of a Global Survey of Institutional Investors, Journal of Applied Corporate Finance, Vol. 29(4), S. 125-133.

Ehrlinger, Lisa / Wolfram Wöß (2016): Towards a Definition of Knowledge Graphs, SEMANTiCS, 48(1-4), 2.

Endt, Christian / Munzinger, Hannes / Obermaier, Frederik / Prugger, Daniela / Wormer, Vanessa (2019): Der Eigentümer bleibt geheim, Süddeutsche Zeitung, 5.2.2019, <https://www.sueddeutsche.de/wirtschaft/transparenzregister-firmeneigentuemer-eu-1.4317342> (13.3.2021).

Engert, Markus (2020): Was sind die FinCEN-Files?, BuzzFeedNews vom 20.9.2020, <https://www.buzzfeed.com/de/marcusengert/was-sind-die-fincen-files> (18.9.2021).

Engert, Markus / Drepper, Daniel (2020): Die FinCEN-Files: Wie Großbanken an Oligarchen, Drogendealern und Terroristen verdienen, BuzzFeedNews vom 20.9.2020, <https://www.buzzfeed.com/de/marcusengert/fincen-files-banken-verdienen-an-kriminellen> (18.9.2021)

EU, Durchführungsbeschluss (EU) 2017/1358 der Kommission vom 20. Juli 2017 zur Festlegung der technischen Spezifikationen im IKT-Bereich, auf die bei der Vergabe öffentlicher Aufträge Bezug genommen werden kann, Amtsblatt der Europäischen Union L 190 vom 21.7.2017, S. 16-19.

EU, Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen „Aufbau eines gemeinsamen europäischen Datenraums“ vom 25.4.2018, COM (2018) 232 final.

EU, Rat der Europäischen Union, Schlussfolgerungen des Rates zur Mitteilung der Kommission über eine externe Strategie für effektive Besteuerung und Empfehlung der Kommission zur Umsetzung von Maßnahmen zur Bekämpfung des Missbrauchs von Steuerabkommen, 9452/16, FISC 85, ECOFIN 502 vom 25.5.2016.

EU, Rat der Europäischen Union, Schlussfolgerungen des Rates zur überarbeiteten EU-Liste nicht kooperativer Länder und Gebiete für Steuerzwecke (2021/C 66/10), Amtsblatt der Europäischen Union C 66 vom 26.2.2021, S. 40-45.

EU, Richtlinie 2014/55/EU des Europäischen Parlaments und des Rates vom 16. April 2014 über die elektronische Rechnungsstellung bei öffentlichen Aufträgen, Amtsblatt der Europäischen Union L 133, 6.5.2014, S. 1-11.

EU, Richtlinie 2014/95/EU des Europäischen Parlaments und des Rates vom 22. Oktober 2014 zur Änderung der Richtlinie 2013/34/EU im Hinblick auf die Angabe nichtfinanzieller und die Diversität betreffender Informationen durch bestimmte große Unternehmen und Gruppen, Amtsblatt der Europäischen Union L 330 vom 15.11.2014, S. 1-9.

EU, Richtlinie (EU) 2015/849 des Europäischen Parlaments und des Rates vom 20. Mai 2015 zur Verhinderung der Nutzung des Finanzsystems zum Zwecke der Geldwäsche und der Terrorismusfinanzierung, zur Änderung der Verordnung (EU) Nr. 648/2012 des Europäischen Parlaments und des Rates und zur Aufhebung der Richtlinie 2005/60/EG des Europäischen Parlaments und des Rates und der Richtlinie 2006/70/EG der Kommission, Amtsblatt der EU L 141 vom 5.6.2015, S. 73-117.

EU, Verordnung (EU) Nr. 596/2014 des Europäischen Parlaments und des Rates vom 16. April 2014 über Marktmissbrauch (Marktmissbrauchsverordnung) und zur Aufhebung der Richtlinie 2003/6/EG des Europäischen Parlaments und des Rates und der Richtlinien 2003/124/EG, 2003/125/EG und 2004/72/EG, Amtsblatt der Europäischen Union L 173 vom 12.6.2014, S. 1-61.

EU, Verordnung (EU) Nr. 600/2014 des Europäischen Parlaments und des Rates vom 15. Mai 2014 über Märkte für Finanzinstrumente und zur Änderung der Verordnung (EU) Nr. 648/2012, Amtsblatt der Europäischen Union L 173 vom 12.6.2014, S. 84-148.

EU Kommission (EU) (2016) 198: Vorschlag für eine RICHTLINIE DES EUROPÄISCHEN PARLAMENTS UND DES RATES zur Änderung der Richtlinie 2013/34/EU im Hinblick auf die Offenlegung von Ertragsteuerinformationen durch bestimmte Unternehmen und Zweigniederlassungen.

Fensel, Dieter et al. (2020): Knowledge Graphs, Springer Nature Switzerland Verlag, Cham 2020.

Footnote, Keith D. (2018): What Is a Document Database? <http://www.dataversity.net/tag/document-store/> (17.10.2021).

Fuchs, Christina / Steiner, Gerhard (2016): Verrechnungspreisdokumentationsgesetz: Hinweise zum länderbezogenen Bericht SWI 2016, 388.

Gesetz zur Einführung einer Pflicht zur Mitteilung grenzüberschreitender Steuergestaltungen, Regierungsentwurf v. 9.10.2019 <https://dip.bundestag.de/vorgang/.../253781> (17.10.2021).

GLEIF (2017): Common Data File Formats – Questions and Answers, <https://www.gleif.org/de/about-lei/common-data-file-format#> (19.9.2021).

Grotherr, Siegfried (2016): Anwendungsfragen bei der länderbezogenen Berichterstattung – Country-by-Country Reporting IStR 2016, 991-1008.

Grümmer, Julian (2021): What can we learn from knowledge graphs? - A Wirecard perspective, Vortrag auf der Knowledge Graph Conference 2021, May 3-6 2021, <https://knowledgegraphconference.vhx.tv/packages/kgc-2021/videos/julian-grummer-what-can-we-learn-from-knowledge-graphs-a-wirecard-perspective> (24.10.2021).

Gwerder, Zoe (2020): Zug und die Briefkastenfirmen: die Schattenseiten des Erfolgs, Luzerner Zeitung, 4.4.2020, <https://www.luzernerzeitung.ch/zentralschweiz/zug/zug-und-die-briefkastenfirmen-die-schattenseiten-des-erfolgs-ld.1210007> (25.2.2021).

Haarmann, Wilhelm (1977): Abgabenordnung - Einführung und Begriffsbestimmungen, Inf. 1977.

Harle, Georg / Nüdling, Lars / Olles, Uwe (2020), Die moderne Betriebsprüfung, 4., aktualisierte und erweiterte Auflage, NWB Verlag, Herne 2020.

Henneberger, Bernd (2012): Die Abgrenzung des Konsolidierungskreises unter IFRS, Shaker Verlag, Aachen 2012.

Henselmann, Klaus / Haller, Stefanie (2018): Potentielle Risikofaktoren in der E-Bilanz-Taxonomie für die Erhöhung der Betriebsprüfungswahrscheinlichkeit, Deutsches Steuerrecht DStR 20/2018, S. 1033-1039.

Henselmann, Klaus / Hofmann, Stefan (2010): Accounting Fraud. Case Studies and Practical Implications, Erich Schmidt Verlag, Berlin 2010.

Henselmann, Klaus / Schmidt, Lutz / Sigloch, Jochen (2005): Internationale Steuerlehre. Steuerplanung bei grenzüberschreitenden Transaktionen, Gabler Verlag, Wiesbaden 2005.

Hodler, Amy / Needham, Mark (2021): Graph Data Science (GDS) for dummies, Wiley, Hoboken 2021.

Hofmann, Stefan (2008): Handbuch Anti-Fraud-Management, Erich Schmidt Verlag, Berlin 2008.

Horák, Josef / Bokšová, Jiřina / Strouhal, Jiří (2020): Electronic Invoicing Adoption within the European Union, International Advances in Economic Research, Volume 26, S. 449–450.

Huang, Feiqi / Vasarhelyi, Miklos A. (2019): Applying robotic process automation (RPA) in auditing: A framework, International Journal of Accounting Information Systems, Volume 35, 2019, 100433, DOI: 10.1016/j.accinf.2019.100433.

Kafsak, Hendrik / Schäfers, Manfred (2021): Konzerne müssen Gewinn für jedes EU-Land veröffentlichen, FAZ vom 25.2.2021, <https://www.faz.net/aktuell/wirtschaft/country-by-country-reporting-konzerne-muessen-gewinn-fuer-jedes-eu-land-veroeffentlichen-17216429.html> (12.3.2021).

Kahle Holger / Schulz Sebastian (2016): BEPS-1-Gesetz: Einführung einer drei-stufigen Verrechnungspreisdokumentation DStZ 2016, S. 823-824.

Kaya, Devrimi / Seebeck Andreas (2019): The dissemination of firm information via company register websites – Country level empirical evidence, *Journal of Accounting and Organizational Change*, 15(3), S. 382–429.

Kejriwal, Mayank (2019): *What Is a Knowledge Graph?*, Domain-Specific Knowledge Graph Construction, Springer International Publishing, S. 1–7.

Kendall, Elisa F. / McGuinness, Deborah L. (2019): *Ontology Engineering*. Morgan & Claypool, Williston 2019.

Kiryakov, Atanas (2016): *Linked Leaks: A Smart Dive into Analyzing the Panama Papers*, <https://www.ontotext.com/blog/linked-leaks-a-smart-dive-into-analyzing-the-panama-papers/> (17.9.2021)

Kowallik, Andreas / Eismayr, Rainer / Kirsch, Andreas (2016): Globale Entwicklungen bei der Automation von Besteuerungsprozessen, *Der Betrieb*, Beilage 04 zu Heft Nr. 47, 25.11.2016, S. 40-46.

Krahen, Jan Pieter / Langenbacher, Katja (2020): The Wirecard lessons: A reform proposal for the supervision of securities markets in Europe, No. 88, *SAFE Policy Letter*, <https://www.econstor.eu/handle/10419/222230> (23.10.2021).

Kroetsch, M / Weikum, Gerhard (2016): Special issue on knowledge graphs, *Journal of Web Semantics*, Vol. (37)38, S. 53-54.

Lobbypedia (2020): Eintrag „Gerhard_Schröder“, https://lobbypedia.de/wiki/Gerhard_Schröder (25.2.2021).

Lutz, Fabian / Seebeck, Andreas (2019a): Country-by-Country Reporting: Herausforderungen und Möglichkeiten einer automatisierten ersten Risikoeinschätzung, *Internationales Steuerrecht (IStR)*, 14/2019, S. 535-543.

Lutz, Fabian / Seebeck, Andreas (2019b): US-amerikanische Bundessteuerbehörde IRS veröffentlicht aggregierte Country-by-Country Reports: Erste Berechnungen und Hinweise auf Problematiken bei der Analyse, *Steuer & Wirtschaft International (SWI)* 9/2019, S. 438-449.

Lutz, Fabian / Seebeck, Andreas (2020): OECD veröffentlicht aktualisierte Guidance zum Country-by-Country Reporting, *Internationales Steuerrecht (IStR)* 29/2020, S. 55-59.

Lutz, Fabian (2020) *Eignung des Country-by-Country Reportings der OECD zur Einschätzung von ausgewählten BEPS-Risiken und Ableitung eines Ansatzes zur Verbesserung des Country-by-Country Reportings*, Duncker & Humblot, Berlin 2020.

Ministry of Economics, Greece (2020): The myDATA platform for electronic pricing is ready, <https://www.vatupdate.com/wp-content/uploads/2020/06/2020-06-13-mydata-en.pdf>, 12. Juni 2020 (8.9.2020).

Monk, Ashby / Prins, Marcel / Rook, Dane (2019): Rethinking Alternative Date in Institutional Investment, *The Journal of Financial Data Science*, Winter 2018, S. 14-31.

Moffitt, Kevin C. / Rozario, Andrea M. / Vasarhelyi, Miklos A. (2018): Robotic Process Automation for Auditing, Journal of Emerging Technologies in Accounting, Vol. 15, No. 1, Spring 2018, S. 1-10, DOI: 10.2308/jeta-10589.

Nayak, Ameya, Poriya, Anil / Poojary, Dikshay (2013). Type of NOSQL databases and its comparison with relational databases. International Journal of Applied Information Systems, 5(4), S. 16-19.

Obermayer, Bastian / Obermaier, Frederik (2016): Panama Papers. Die Geschichte einer weltweiten Enthüllung, 2. Aufl., Kiepenheuer & Witsch, Köln 2016.

OECD (2017): Country-by-Country Reporting: Handbook on Effective Tax Risk Assessment, 29.9.2017, www.oecd.org/tax/beps/country-by-country-reporting-handbook-on-effective-tax-risk-assessment.pdf (17.10.2021).

OECD (2018): Guidance on the Implementation of Country-by-Country Reporting, BEPS Action 13 (Guidance), <https://www.oecd.org/tax/beps/guidance-on-country-by-country-reporting-beps-action-13.htm> (13.9.2018).

OECD (2021): What is BEPS? (12.3.2021).

OECD (2010), OECD Transfer Pricing Guidelines for Multinational Enterprises and Tax Administrations 2010 (TPG), OECD Publishing, Paris 2010.

Offenregister.de (2021): German company register data: README, <https://offenregister.de/daten/> (13.3.2021).

ORF (2020): Was Wien für Briefkastenfirmen attraktiv macht, ORF.at, 22.6.2020, <https://orf.at/stories/3170200/> (25.2.2021).

Open Knowledge Foundation (2021): What is open?, <https://okfn.org/opendata/> (31.7.2021)

Opencorporates (2019): German company data now open for all, <https://blog.opencorporates.com/2019/02/05/german-company-data-now-open-for-all/> (13.3.2021).

Oxfam (2016c): Data in Excel File, https://oi-files-d8-prod.s3.eu-west-2.amazonaws.com/s3fs-public/file_attachments/tb-race-to-bottom-methodology-spreadsheet-121216-en.xlsx (7.3.2021).

Oxfoam (2016a): Tax Battles. The dangerous global Race to the Bottom on Corporate Tax, <https://oi-files-d8-prod.s3.eu-west-2.amazonaws.com/s3fs-public/bp-race-to-bottom-corporate-tax-121216-en.pdf> (7.3.2021).

Oxfoam (2016b): Technical Methodology Document. How Oxfam identified the world's worst corporate tax havens, https://oi-files-d8-prod.s3.eu-west-2.amazonaws.com/s3fs-public/file_attachments/tb-race-to-bottom-methodology-note-121216-en.pdf (7.3.2021).

Paulheim, Heiko (2017): Knowledge graph refinement: A survey of approaches and evaluation methods, Semantic web, Vol. 8(3), S. 489-508.

Peemöller, Volker H. / Kregel, Joachim (2010): Grundlagen der internen Revision, Erich Schmidt Verlag, Berlin 2010.

Peemöller, Volker H. / Krehl, Harald / Hofmann, Stefan / Lack, Jana (2020): Bilanzskandale, 3. Aufl., Erich Schmidt Verlag, Berlin 2020.

Redeker, Steffen (2020): Der DAX in Steueroasen. Studie für die Fraktion DIE LINKE. im Bundestag, <https://www.fabio-de-masi.de/de/article/2757.studie-der-dax-in-steueroasen.html> (3.3.2021).

Russel, Matthew / Klassen, Mikhail (2019): Mining the Social Web, 3. Auflage, O'Reilly Verlag, Sebastopol 2019.

Sauer, Otto (1988): Steuerliche Außenprüfung, Vahlen Verlag, München 1988.

Schaefer-Drinhausen, Heinrich (2021): Steuerhinterziehung bei privaten Verkäufen, <https://www.rae-koeln.de/steuerhinterziehung-bei-privaten-verkaufen/> (13.3.2021).

Schneider, Katharina (2020): Per Airbnb zum Steuerhinterzieher: Warum Vermieter jetzt dringend handeln müssen, Handelsblatt, 16.09.2020 (<https://www.handelsblatt.com/finanzen/steuern-recht/steuern/kurzvermietung-per-airbnb-zum-steuerhinterzieher-warum-vermieter-jetzt-dringend-handeln-muessen/26192082.html>) (13.3.2021).

Schönwitz, Daniel (2014): Mit diesen Tricks machen Sie sich fast zum Hoeneß, Wirtschaftswoche, 10.3.2014, <https://www.wiwo.de/finanzen/steuern-recht/steuerhinterziehung-mit-diesen-tricks-machen-sie-sich-fast-zum-hoeness/9583386-all.html> (13.3.2021)

Seebach, Nils (2018): Bundesanzeiger 2.0 oder Stalking für Firmen, <https://digitalkaufmann.de/entrepreneur-radar/bundesanzeiger-2-0-stalking-fuer-firmen/> (17.9.2021)

Seebeck, Andreas / Kaya Devrimi (2021): The Power of Words: An Empirical Analysis of the Communicative Value of Extended Auditor Reports, European Accounting Review, im Erscheinen.

Seebeck, Andreas / Lutz, Fabian (2021): Das XML-basierte Country-by-Country-Reporting - Neuerungen, Möglichkeiten und Herausforderungen, Steuer und Wirtschaft International – Tax and Business Review 5/2021, S. 256–268.

Seebeck, Andreas / Vetter, Julia (2021): Not Just a Gender Numbers Game – How Board Gender Diversity Affects Corporate Risk Disclosure, Journal of Business Ethics, im Erscheinen.

Sharma, Vatika / Dave, Meenu (2012): SQL and NoSQL Databases, International Journal of Advanced Research in Computer Science and Software Engineering 2(8), S. 20-27.

Shaxson, Nicholas (2012): Treasure Islands: Tax Havens and the Men who Stole the World, Vintage Verlag, London 2012.

Singhal, Amit (2012): Introducing the Knowledge Graph: things, not strings, 16.5.2012, <https://blog.google/products/search/introducing-knowledge-graph-things-not/> (8.10.2021)

Swanson, Ana (2016): How the U.S. became one of the one of the world's biggest tax havens, The Washington Post, 5.3.2016, <https://www.washingtonpost.com/news/wonk/wp/2016/04/05/how-the-u-s-became-one-of-the-worlds-biggest-tax-havens/> (8.3.2021).

Tax Justice Network (2021a): Our History, <https://www.taxjustice.net/our-history/> (7.3.2021).

Tax Justice Network (2021b): Financial Secrecy Index, <https://fsi.taxjustice.net/en/> (7.3.2021).

Usschold, Michael (2018): Demystifying OWL for the Enterprise, Morgan & Claypool, Williston 2018.

Véron, Nicolas (2020): The Wirecard debacle calls for a rethink of EU, not just German, financial reporting supervision, Bruegel vom 30.06.2020, <https://www.bruegel.org/2020/06/the-wirecard-debacle-calls-for-a-rethink-of-eu-not-just-german-financial-reporting-supervision/> (23.12.2020)

von Brocke, Klaus / Nonnenmacher, Roland / Przybilka, Stefan (2021): Anzeigepflichten für grenzüberschreitende Steuergestaltungen, NWB Verlag, Herne 2021.

von Daniels, Justus / Wörpel, Simon (2019): Mieten unter Palmen, <https://correctiv.org/aktuelles/wem-gehoert-hamburg/2019/02/05/mieten-unter-palmen/> (13.3.2021).

Wortham, Leah / Liebscher, Marc / Christian, Christopher D. (2020), Wirecard, Europe's Enron? - Auditor Liability to Investors Corporate Fraud, https://scholarship.law.edu/other_lectures/33 (24.10.2021).

Zucman, Gabriel (2015): The Hidden Wealth of Nations, The University of Chicago Press, Chicago 2015.