



INSTITUT FÜR
DIGITALISIERUNG
IM STEUERRECHT



IDSt-White Paper

Fachausschuss VII

„Process Analytics“: Data Analytics und Process Mining im Steuerbereich

Fachausschuss VII

Innovative Technologien („Hot Shit“)

White Paper – „Process Analytics“: Data Analytics und Process Mining im Steuerbereich ¹

Inhaltsverzeichnis

A. Einleitung.....	2
B. Zielsetzung der Arbeitsgruppe „Process Analytics“	3
C. Definition von Process Analytics.....	4
D. Datenanalyse im Kontext von Process Analytics.....	5
E. Kategorisierung von Analysezielen	6
F. Datentypologien	8
G. Datenbereitstellung und -aufbereitung	9
H. Zusammenfassung und Ausblick.....	10

¹ Das White Paper spiegelt die Auffassungen des gesamten Fachausschusses VII, aus dem heraus es unter der Mitwirkung von Robert Backes, Evrim-Kaya Breuer, Daniel Dallhammer, Thomas Hoppe, Zoran Jotanovic, Roman Kowallik, Franc Scherkamp, Dennis Wiedmann entstanden ist.

A. Einleitung

„Wem die Daten gehören, dem gehört die Zukunft“ – Diese Lektion aus Hararis Bestseller „21 Lektionen für das 21. Jahrhundert“² veranschaulicht die Bedeutung, die Daten heutzutage besitzen. Auch vor dem Steuerumfeld macht diese Entwicklung nicht halt, ganz im Gegenteil: (Big)Tax Data ist in Unternehmen und Kanzleien längst zur Realität geworden. Dabei ist davon auszugehen, dass wir erst am Beginn einer exponentiellen Entwicklung stehen. Auch wenn häufig rasch auf die Risiken hingewiesen wird, Big Tax Data bietet der Steuerfunktion gänzlich neue Möglichkeiten. So verfügen wir heute über die Daten, die es uns ermöglichen, Auswertungen oder Kontrollen zu initiieren, die bis vor kurzem mangels entsprechender Informationen eher der Theorie vorbehalten waren.

Daten jeglicher Art und Herkunft sind zunächst als Rohmaterial anzusehen. Ihr tatsächlicher Mehrwert erschließt sich erst durch die spätere Nutzung. Hierfür eignen sich spezifische Werkzeuge, die im Bereich **Data Analytics** angesiedelt sind. Data Analytics bietet insbesondere die Möglichkeit, steuerlich relevante Datenbestände einer Vollprüfung zu unterziehen, unter dem Blickwinkel von Tax Compliance sowie möglicher Haftungsrisiken ein unschätzbare Mehrwert.

Basis für die Entstehung steuerrelevanter Daten ist die steuerliche **Prozesswelt**. Für deren Erfassung, Management und Optimierung bedarf es stets eines Gesamtkonzepts, um die Zuverlässigkeit der Prozesse und die hiermit eng verwobene Datenqualität sicherzustellen. Entsprechend sollten weniger (steuerliche) Teilprozesse in die Betrachtung einfließen, als vielmehr alle steuerrelevanten Geschäftsprozesse aus einer „End-to-End-Perspektive“. Dies entspricht im Übrigen den in den GoBD konkretisierten verfahrensrechtlichen Grundsätzen (§ 145 Abs. 1 AO), wonach sich Geschäftsvorfälle in ihrer Entstehung und Abwicklung lückenlos nachverfolgen lassen müssen. Entsprechend stellt sich die Frage, wie steuerlich relevante Prozesse und die dort entstehenden steuerlich relevanten Daten zielgerichtet analysiert und transparent dargestellt werden können. Hier treten die Möglichkeiten und Vorteile von **Process Mining** als weiterer steuerlicher Zukunftstechnologie zu Tage.

Zusammenfassend stehen **Data Analytics** und **Process Mining** für zwei Kerntechnologien in der Steuerwelt von morgen. Dabei sollten diese Technologien nicht

² Yuval Noah Harari, 21 Lessons for the 21st Century, 2018.

isoliert voneinander betrachtet und angewandt werden, sondern stets in einem wertstiftenden Zusammenspiel. So basiert auch Process Mining auf Daten und kann damit als eine besondere Form von Data Analytics angesehen werden, mit dem Ziel, die reine Datenperspektive um prozessuale Aspekte zu erweitern. Die detaillierte Analyse von Geschäftsprozessen hilft unter anderem dabei, die für steuerliche Zwecke zielführenden Datenanalysemethoden und -technologien (vgl. Kapitel D bis F) auszuwählen. In der Arbeitsgruppe „Process Analytics“ des IDSt-Fachausschuss VII (Innovative Technologien „Hot Shit“) hat sich für dieses Zusammenspiel der Begriff „**Process Analytics**“ herausgebildet.

Mit diesem White Paper will die Arbeitsgruppe „Process Analytics“ aus dem Fachausschuss VII des IDSt Einblicke in das Thema Process Analytics geben. Zugleich soll deutlich werden, dass die Kombination von Methoden und Technologien auch einen wesentlichen Beitrag zum Aufbau eines wirksamen Tax CMS leisten kann.³ Schließlich erhält der Leser Anregungen, wie sich Data Analytics, Process Mining und schließlich Process Analytics in der steuerlichen Praxis gezielt einsetzen lassen.

B. Zielsetzung der Arbeitsgruppe „Process Analytics“

Das vorliegende White Paper bildet den Ausgangspunkt für eine tiefere Auseinandersetzung mit dem Themenkomplex „Process Analytics“. Insgesamt hat sich die Arbeitsgruppe folgende Ziele für ihre künftige Tätigkeit gesetzt:

- Die Schaffung eines **Verständnisses** und einer breiten **Akzeptanz** für die Themen **Data Analytics, Process Mining** und **Process Analytics** in der Steuerfunktion.
- Das Herausstellen des **Nutzens** von **Process Analytics** für die Steuerpraxis mit konkreten Use Cases.
- Die Darstellung von Möglichkeiten zur **Analyse** und **Optimierung** von steuerlichen Daten und Prozessen als Teil des **steuerlichen IKS** (Tax CMS).
- Das Screening und die Darstellung von **Use Cases, Methoden** und **Technologien** aus dem Beratungs- und Unternehmensumfeld sowie

³ Wir sehen Process Analytics u.a. als möglichen Bestandteil eines wirksamen Tax Compliance Management Systems. Das Thema Tax Compliance selbst wird ganzheitlich innerhalb des IDSt im Fachausschuss V („Steuerliches Kontrollumfeld“) behandelt.

- ein Blick in die **Zukunft** von **Data Analytics** und **Process Mining** im digitalisierten steuerlichen Umfeld.

C. Definition von Process Analytics

Process Analytics ist eine **methodische und technische Symbiose** aus **Data Analytics** und **Process Mining**. Dabei bilden die steuerrelevanten Geschäftsprozesse den Ausgangspunkt der jeweiligen Betrachtung.

Data Analytics bezieht sich vornehmlich auf die Analyse von kleinen und großen Datenmengen bis hin zu Big Data, um nützliche Erkenntnisse und Muster zu gewinnen. Die Zielsetzung besteht darin, strukturierte und unstrukturierte Daten (vgl. Kapitel F) aus verschiedenen Quellen zu sammeln, zu organisieren, zu visualisieren und zu interpretieren, um Trends zu identifizieren sowie Geschäftsentscheidungen und Vorhersagen zu treffen. Analyseziele können dabei von retrospektiv bis prospektiv definiert werden.

Process Mining beschäftigt sich vornehmlich mit der belegorientierten Visualisierung von Geschäftsprozessen (vgl. Abbildung 1) in unterschiedlichen Varianten und Ausprägungen. Mithin werden hierunter Techniken, Werkzeuge und Methoden verstanden, um reale bzw. gelebte Prozesse zu identifizieren, zu beobachten und zu verbessern. Die Informationsbasis hierfür bilden Daten bzw. Ereignislogs in den zugrundeliegenden IT-Systemen, weshalb Process Mining stets die tatsächlich gelebten Prozesse eines Unternehmens widerspiegelt. Dies schafft Transparenz und offenbart Abweichungen vom Sollprozess sowie mögliche Prozessdefizite.



Hinweis: Eine detailliertere Erläuterung der einzelnen Technologien kann dem Technologie-Glossar des IDSt entnommen werden, abrufbar unter: <https://idst.tax/wp-content/uploads/2022/11/Glossar-TaxTech.pdf>

D. Datenanalyse im Kontext von Process Analytics

Im Hinblick auf die Datenanalyse-Dimension von Process Analytics sind **Anlass** und **Automationsgrad** sowie **Analysefrequenz** und **Anwendungsmethodik** wie folgt zu beschreiben:

(1) Anlass und Automationsgrad:

- **Ad-hoc** (manuelle) Analysen: Individuelle „Self-Service“ Daten- bzw. Prozessanalysen, initiiert durch den Anwender, z.B. nachdem Prozessdefizite oder Fehler im Rahmen der Steuerdeklaration aufgefallen sind.
- **Halbautomatisierte** bzw. **vorgefertigte Analysen**: (Teilweise) standardisierte Daten- bzw. Prozessanalysen, z.B. durch Anwendung von standardisierten, vorgefertigten Skripten oder Prüfmakros zur Vorbereitung auf eine steuerliche Außenprüfung.
- **Automatisierte** bzw. **fortlaufende Analysen**: Analyse von Daten bzw. Prozessen durch vordefinierte Algorithmen, die automatisiert (ggf. basierend auf sog. „Triggering Events“) ausgewählte Datenbereiche oder Prozesse überwachen.
- Eine weitere Ausbaustufe bildet die ergänzende Nutzung von **künstlicher Intelligenz** und **maschinellern Lernen** unter Verwendung spezifischer Algorithmen. Hier ist zu erwarten, dass in den nächsten Jahren diverse Lösungen zur Marktreife gebracht werden.

(2) Analysefrequenzen:

- **Periodische Analysen**: Bestimmter Analyse-Rhythmus, z.B. stündlich, wöchentlich, monatlich, quartalsweise, etc.
- **Real-Time Analytics**: Echtzeit-Analyse individueller operativer Vorgänge und kontinuierliche Echtzeit-Auswertung von Datenströmen und Prozessroutinen (auch als „Continuous Monitoring“ bezeichnet).

(3) Anwendungsmethoden:

- **Tabellenorientierte** Datenanalyse mit Anwendungen wie beispielsweise Excel, IDEA, ACL oder SQL-Werkzeugen.
- **Visual Analytics** mit Anwendungen wie beispielsweise Celonis oder PowerBI: Analyse und Interpretation von Daten bzw. Prozessen auf Basis

computergestützter Visualisierungen, z.B. auf Basis von Grafiken, Charts oder sog. Heatmaps.

E. Kategorisierung von Analysezielen

Neben Anlass, Automationsgrad, Frequenz und Methodik ist nach verschiedenen Analysezielen zu unterscheiden. Je nach Zielsetzung kommen in Bezug auf Process Analytics unterschiedliche Methoden zum Tragen, die aufeinander aufbauen (vgl. *Abbildung 2*)⁴.

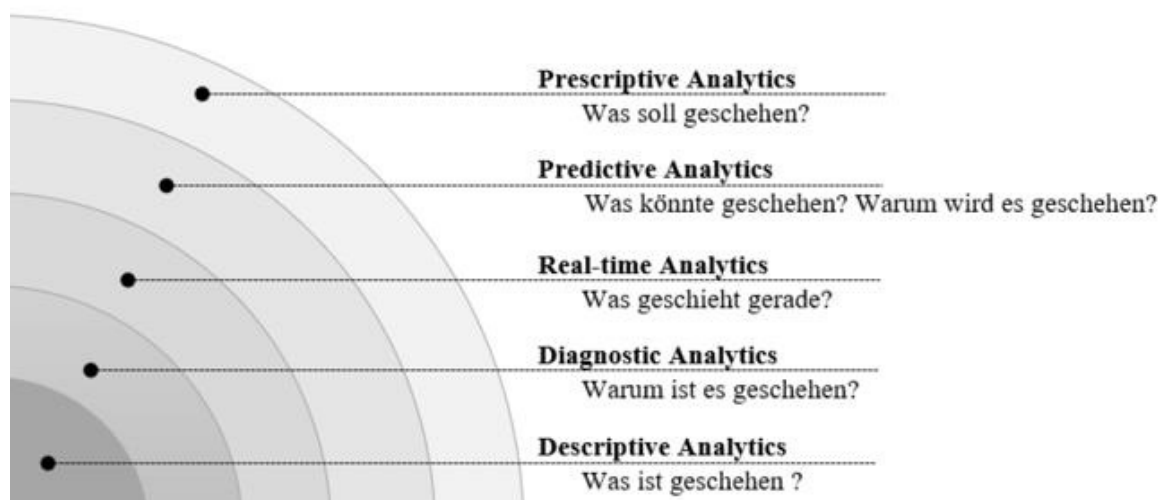


Abbildung 2: Kategorien von Analysezielen

- **Descriptive Analytics** soll die Frage beantworten: „*Was ist geschehen?*“ Die Zielsetzung bildet damit die Sammlung und Zusammenstellung von Daten bzw. korrespondierender Prozessschritte anhand eines vorgegebenen Kontextes zwecks ex-post Analyse bereits realisierter steuerrelevanter Sachverhalte bzw. Prozessschritte (Beispiel: Untersuchung monatlicher Umsatzsteuermeldungen auf Ausreißer).

⁴ Angelehnt an LANQUILLON, C./MALLOW, H. (2015): Advanced Analytics mit Big Data, in: Dorschel, J. (Hrsg.) (2015): Praxishandbuch Big Data – Wirtschaft - Recht - Technik, Wiesbaden 2015, S. 55-89.

- **Diagnostic Analytics** soll die Frage beantworten: „*Warum ist etwas geschehen?*“ Hier geht es – bezogen auf die zugrundeliegenden Daten oder Prozesse – um die Ermittlung und Identifizierung bestimmter Muster und Abhängigkeiten im Rahmen von sog. Assoziations- oder Korrelationsanalysen (Beispiel: Untersuchung der Ursachen für Schwankungen der (Konzern-)Steuerquote).
- **Real-Time Analytics** zielt regelmäßig darauf ab, Daten in (Nahezu-)Echtzeit zu verarbeiten, um die für die Überwachung steuerrelevanter Prozessschritte im transaktionalen Kontrollumfeld relevante Frage „*Was geschieht gerade?*“ zu beantworten. Real-Time Analytics ist insbesondere bei Prozessschritten von Bedeutung, die im Rahmen der Qualitätssicherung eine unmittelbare, prozessintegrierte Reaktion erfordern (Beispiel: Unmittelbare Identifikation von Unstimmigkeiten und Anomalien bei der Auswahl von Steuerkennzeichen für einen Geschäftsvorfall).
- **Predictive Analytics:** Hier geht es um den Blick in die Zukunft: „*Was könnte geschehen und warum könnte dies geschehen?*“ Die Zielsetzung besteht regelmäßig in einer Prognose von Umsatz, Kosten oder anderer betrieblicher Kennzahlen. Auch die Ergebnisse einzelner Prozessschritte oder ganzer Prozessketten können Gegenstand der Prognose sein (z.B. durch Extrapolation ermittelter Trends). Auf dieser Basis lassen sich Entscheidungsparameter für künftige erfolgskritische, unternehmerische Maßnahmen festlegen (Beispiel: Vorhersage der Überschreitung von Umsatzgrenzen, die unter Berücksichtigung sonstiger umsatzsteuerlicher Tatbestandsmerkmale eine steuerliche Registrierung erfordern; Vorhersage der Überschreitung der Mindestbesteuerung im Rahmen von Pillar 2).
- **Prescriptive Analytics** beschäftigt sich mit der Frage „*Was soll geschehen?*“ So sollen Handlungsempfehlungen gegeben werden, durch welche sich ein gewünschtes Ergebnis erzielen lässt. Dies betrifft beispielsweise Vorschläge zur Ausgestaltung von Geschäftsvorfällen zur Optimierung der steuerlichen Auswirkung (Beispiele: Anregungen für zu vereinbarende Incoterms; Empfehlungen betreffend die Gestaltung von Reihengeschäften zur Vermeidung von umsatzsteuerlichen Registrierungen; Vorschläge zur korrekten Verrechnungspreisbestimmung).

Aus den Ergebnissen einer jeden der vorstehend beschriebenen Datenanalysekategorien lassen sich (teil-)automatisierte Handlungen ableiten. Diese können von der einfachen

Ausgabe systemseitiger Meldungen bzw. Warnungen oder dem automatischen Versand von Reports über ein Verhindern von Geschäftsvorfällen mit bestimmten steuerschädlichen Parametern bis hin zur vollautomatischen Korrektur von als fehlerhaft erkannten Buchungen reichen.

F. Datentypologien

Während steuerliche Datenanalysen bislang vornehmlich auf **strukturierte** Daten rekurrieren, schreitet nun auch die „Eroberung“ **unstrukturierter** Daten im Steuerbereich voran. Die wesentlichen Merkmalsausprägungen stellen sich wie folgt dar:

(1) Strukturierte Daten

- Strukturierte Daten sind Daten, die in einem standardisierten Format vorliegen und eine klare und vordefinierte Struktur aufweisen. Sie entsprechen einem spezifischen Datenmodell, folgen einer konsistenten Reihenfolge und sind sowohl für Programme als auch für Menschen leicht zugänglich, soweit sie für letztere lesbar gemacht werden.
- Dieser Datentypus liegt in normalisierten Datenstrukturen vor und lässt sich einfach in eine Datenbank exportieren, speichern und dort organisieren.
- Mit strukturierten Daten lassen sich Trends erkennen und Entwicklungen abschätzen.

(2) Unstrukturierte Daten

- Bei unstrukturierten Daten liegt die Information in Form von nicht-normalisierten Datenstrukturen vor. Dabei handelt es sich u.a. um Bilder, Texte, Audio- und Videodateien. Die Auswertung der Daten in dieser Form ist regelmäßig mit einem größeren Aufwand verbunden und erfordert meist den Einsatz von KI oder Deep-Learning Algorithmen.
- Die Aufbereitung von unstrukturierten Daten erfolgt über Technologien bzw. Techniken wie Text Mining, Indexierung und Strukturierung (Crawling), Tokenisierung und syntaktische Aufbereitung oder semantische Auswertung (Parsing Rules, Natural Language Processing).

Bei der Fortentwicklung steuerlicher Data Analytics-Aktivitäten hin zu einem Process Analytics-Ansatz werden unstrukturierte Daten voraussichtlich an Bedeutung gewinnen, weil mit Hilfe ihrer Auswertung Prozesse, die bislang lediglich manuell bearbeitet werden konnten, zumindest teilweise automatisierbar werden (z.B. die Auswertung von Quellensteuerklauseln in Verträgen und darauf basierend die Verwaltung von Freistellungs- oder Ansässigkeitsbescheinigungen).

G. Datenbereitstellung und -aufbereitung

“Um Daten auszuwerten, benötigt man Daten.” Essenziell für den Einsatz von Process Analytics ist mithin stets das Vorhandensein ausreichender und repräsentativer Daten. Was banal klingt, stellt sich in der Praxis jedoch vielfach als nicht unwesentliche Herausforderung dar. So entstehen Daten häufig in unterschiedlichen Quellsystemen und wandern – in unterschiedlichen Aggregationsstufen – über Schnittstellen durch die Systemlandschaft des Unternehmens. Dies spielt immer dann eine wesentliche Rolle, wenn steuerrelevante Prozesse, die sich über unterschiedliche Systeme erstrecken, Ende-zu-Ende identifiziert, analysiert und auf Basis der Analyseergebnisse optimiert werden sollen (systemübergreifendes Process Mining).

So muss es regelmäßig zunächst darum gehen, die für die Auswertungen erforderlichen Daten nebst den datenführenden Systemen zu identifizieren. In der Praxis haben sich dabei zwei unterschiedliche Vorgehensweisen herausgebildet, die auch in Kombination oder abwechselnd zur Anwendung kommen können: Ein **technologie- bzw. datenzentrierter** sowie ein **Use Case-zentrierter** Ansatz.

Beim **technologie- bzw. datenzentrierten Ansatz** besteht bereits ein mehr oder weniger umfangreich definierter Datenumfang, einschließlich der Möglichkeit des entsprechenden Zugriffs auf die datenführenden Systeme (z.B. ein steuerliches Data Warehouse, einen steuerlichen Data Lake oder einen GoBD-Datenkranz für steuerliche Daten). Dem liegt regelmäßig die Annahme zugrunde, dass der vorliegende Datenumfang hinreichend valide und repräsentativ ist. Ausgehend von den Daten werden schließlich Analyseziele (Use Cases) festgelegt und initiiert. Stellt sich die Annahme der Validität und Konsistenz der Daten als unzutreffend heraus, so existieren Aufbereitungsmethoden und -tools, um die erforderlichen Voraussetzungen für eine zielführende Datenanalyse zu schaffen.

Umgekehrt knüpft der **Use Case-zentrierte Ansatz** beim Analyseziel an, welches zunächst definiert wird. Entsprechend gilt es, davon ausgehend, valide und repräsentative Daten in geeigneter Granularität zu akquirieren und/oder aufzubereiten. Je nach Analyseziel benötigt man in einem nächsten Schritt entsprechende Real-Time-Zugriffe auf die Daten, regelmäßige Datenextrakte oder auch nur sporadische Zugriffe auf die datenführenden Systeme.

Was die Datenbereitstellung und -aufbereitung angeht, ergeben sich generisch folgende Herausforderungen:

- Datenidentifikation
- Beurteilung der Datenqualität
- Berechtigung unter Berücksichtigung von Datenschutzerfordernungen
- Datenbeschaffung (aus z.T. heterogenen IT-Systemlandschaften)
- Datenaufbereitung
- Datenbereitstellung in angemessener Frequenz (passend zum Ziel).

Hinweis: Die letzten drei Punkte sind der Methodik „ETL“ (Extraction – Transformation – Load) zuzuordnen. Mit dem Thema „Standardisierte Datenmodelle“ beschäftigt sich im Detail die IDSt-Arbeitsgruppe „Common Data Model“.

H. Zusammenfassung und Ausblick

Der Einsatz von Technologie darf kein Selbstzweck sein. Für den Steuerbereich ergeben sich aus den Möglichkeiten von Process Analytics vornehmlich folgende Mehrwerte:

- Systematische Identifizierung von Ineffizienzen, Prozessschwächen, Prozessverstößen und Prozessverbesserungspotenzialen.
- Identifizierung kontext-abhängiger Beziehungen, die bei alleiniger Betrachtung von Stichproben (aufgrund des mangelnden Zusammenhangs mit anderen Transaktionen ähnlicher Art) nicht feststellbar sind.
- Erhöhung der Leistungsfähigkeit der Steuerfunktion durch Automatisierung steuerlicher Kernprozesse und verbesserten Zugang zu allen übrigen steuerrelevanten Geschäftsprozessen („Mehr Arbeitsleistung bei gleichbleibenden Ressourcen“).

- Reduzierung von Risiken durch den Nachweis eines angemessenen und prozessorientierten steuerlichen IKS (angemessene interne Kontrollen sind wesentliche Elemente der Tax Compliance).
- Erhöhung der Wirksamkeit eines Tax CMS durch Auswahl von Datenanalysemethoden und -technologien, die bei den jeweiligen steuerrelevanten Prozessschritten effektiv und effizient (automatisiert) angewandt werden können.

Die Arbeitsgruppe „Process Analytics“ des IDSt-Fachausschuss VII (Innovative Technologien „Hot Shit“) wird die technologischen Entwicklungen weiter beobachten und dieses White Paper fortschreiben. Darüber hinaus plant die Arbeitsgruppe im nächsten Schritt weitere Hilfestellungen für die Praxis zu erarbeiten, insbesondere:

- Entwicklung von Use Cases und Praxisbeispielen für den Einsatz von Process Analytics,
- Identifizierung und Bewertung von Lösungsanbietern im Bereich Process Analytics.