

KI-basierte Datenanonymisierung

*Wie personenbezogene Daten geschützt
und Inhalte nutzbar werden*



Frederick Bednar

Data Analytics & Data Science Consultant, EBCONT

AI & ML | Data Science | Semantics | Knowledge Mgt



Let's connect!



Developing IT-Solutions

Wir sind EBCONT



ENTERPRISE

Custom Business Applications
Software Development Services
Software Integration Services
Mobile Apps



ebcont

PROCONSULT

Project Management & Solution Delivery
Advisory Services, Digitalization
Data Integration, Advanced Analytics, Semantics & AI
Business Analysis, Agile Coaching, Quality Assurance

OPERATIONS

Infrastructure Services
Managed Service
Cloud, IaC & Hybrid
DevOps

COMMUNICATION

Marketing Strategy
Public Relations
UX / UI Design
Prototyping



ÖSTERREICH Wien | Graz | Linz | Hard | Neulengbach | Salzburg | Pinkafeld | St. Pölten

DEUTSCHLAND Frankfurt | Saarbrücken | Hamburg | Berlin

KROATIEN Čakovec

USA New York | Philadelphia

100% Privatbesitz
20+ Jahre Erfahrung
650+ Mitarbeiter:innen

Der Partner für alle IT Belange

Unsere Produktpartner

Ein gut **funktionierendes
Partnernetzwerk** ist uns
besonders wichtig. Mit
langjährigen Partnern bieten wir
vertrauenswürdige und
nachhaltige Lösungen im
IT-Sektor. Einige davon hier:



Wie nutzt man sensible Daten...

...ohne den Datenschutz zu gefährden?

Wie nutzt man sensible Daten...

... ohne den Datenschutz zu gefährden?

- » Aspekte der:
 - » Datenherkunft, Datenqualität
 - » Datenhaltung, Datensicherheit
 - » Performance

id	Name	mail	Geburtsdatum	xAb1f
1	Anna Meier	annameier@example.com	15.02.1985	Premium
2	Ben Schulz	benschulz@example.com	1990-13-01	Standard
3		claudiajung@example.com	28.07.1975	Premium
4	David Müller	davidmuller@example.com	20.05.1980	Deluxe
1	Anna Meier	annameier@example.com	15.02.1985	Premium
5	Eva Linn	evalinn@example.com		Standard

Das BMJ auf dem Weg zur digitalen Justiz

Pain Point:

Keine breite Nutzbarkeit wertvoller Daten

- » hochsensible Daten & strenge Datenschutzregeln
- » Manuelle Datenverarbeitung
- » fehlende Digitalisierung alter Dokumente
- » Unterschiedliche Formate und Dokumentenstruktur
- » Regeln oft unklar
(Kontext: Parteienstellung, Rolle im Verfahren)

Bundesministerium für Justiz

- ★ **Zuständigkeit u.a.:**
Straf- und Zivilrechtliche
Verfahren
- ★ **Infrastruktur:** On Premise
- ★ **Projektziel:** sicherer Zugang
zu Entscheidungen für
Richter:innen

Ist die österr. Rechtsdomäne komplex?

Ja, denn

- » Spezialbegriffe sind nicht in (Standard-)Modellen trainiert
(und könnten sonst als Eigennamen interpretiert werden)
- » Unterschiede in Begriffen zwischen Österreich & Deutschland

„Unsere Entscheidungsorgane benötigten eine technische Unterstützung für die datenschutzkonforme Veröffentlichung von Entscheidungen.“

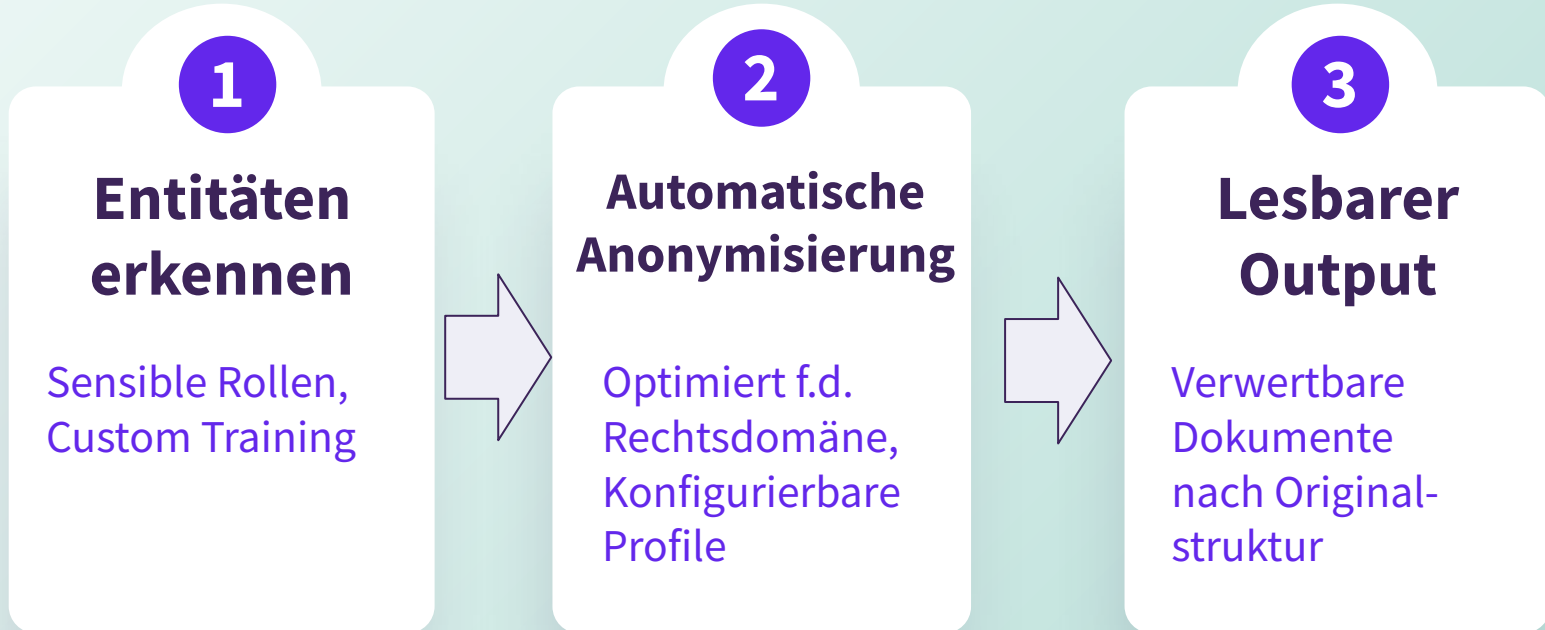
Mag. Christian Gesek, Bundesministerium für Justiz

Leiter der Abteilung für Rechtsinformatik,

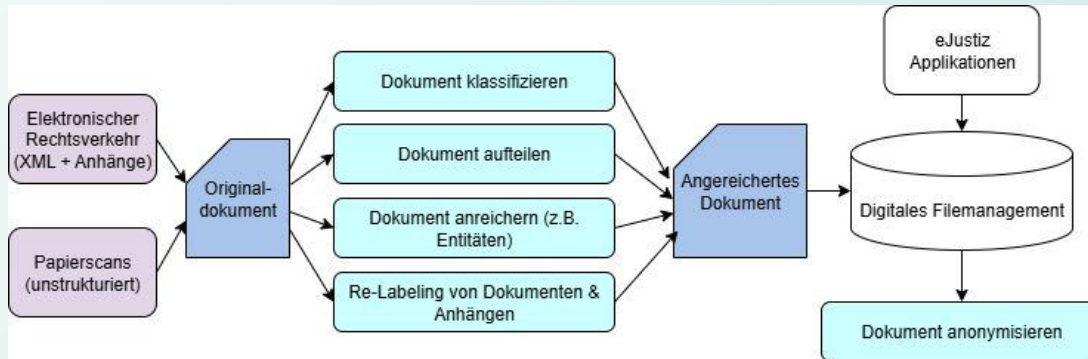
Informations- und Kommunikationstechnologie



Anonymisierung mit KI - maßgeschneidert für die Justiz



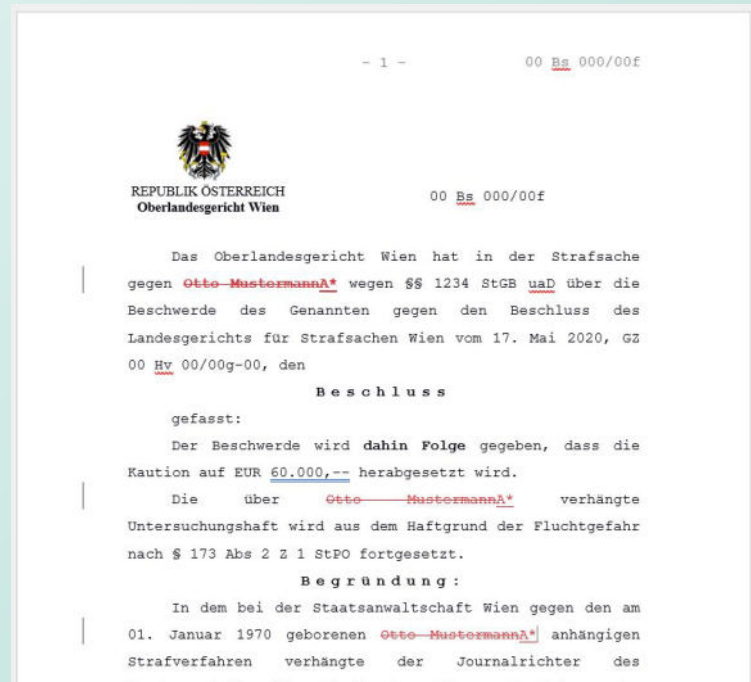
Cognitive Services Plattform



- » **Machine Learning & KI-Plattform** für das Bundesministerium für Justiz
- » **Korrekte Zuordnung von Fällen**, ohne vorherige Themenvorgabe
- » **Automatische Dokumententrennung und Zuweisung** bzw. **Generierung von Titeln** für hochgeladene PDFs / gescannte Dokumente
- » Anonymisiert automatisch Urteile
- » **Inhaltsanreicherung & Harmonisierung**
- » **Entitätenextraktion und -verknüpfung**
- » **Ergebnisgenauigkeit > 90%**

Anonymisierung mit System

- » Zuordnung nach Entitäten
- » Fortlaufende Pseudonymisierung oder Schwärzung von Passagen
- » Anonymisierungsprofile
- » Struktur & Lesbarkeit im Dokument sollen gewährleistet bleiben



Kombination für bessere Treffsicherheit

- » Unscharfe Suche, RDF-Triples & Suchkandidaten
- » NLP & Custom Deep Learning (Transformer-Modelle)
- » OCR Services für Verarbeitung von Text in Bildern
- » Recognizer für maßgeschneiderte Entitäten
- » Combiner für alle aktiven Services, Gruppierung von Token



Von geschützten Daten zu wertvollen Erkenntnissen

Vorteile:

- » Datenschutz, Effizienz, breite Nutzung des Outputs

Messbare Erfolgskriterien:

- » “Klassische” KPI wie Accuracy, aber auch z.B. WER & BLEU
- » Thresholds für kombinierte Zwischenergebnisse
- » Direkte Qualitätsmessung (regelbasiert, Konfidenz)
- » Benchmarking mit repräsentativen Daten (skalierbar)

„Mit der Lösung können wir auf Knopfdruck Anonymisierungsvorschläge erzeugen und damit ohne zusätzliche Personalressourcen den Anteil an veröffentlichten Entscheidungen erhöhen.“

Mag. Christian Gesek, Bundesministerium für Justiz

Leiter der Abteilung für Rechtsinformatik,

Informations- und Kommunikationstechnologie



Übertragbarkeit auf den Public & Private Sector



Best Practice für viele Bereiche: Skalierbar und sicher

Einsatzbereiche

- » in Justiz, Verwaltung, Finanz, Polizei, Gesundheitswesen etc.
- » Aber auch Einsatzgebiete in der Privatwirtschaft

Normen und Compliance

- » Anpassbarkeit an andere Rechts- und Fachdaten
- » Compliance mit DSGVO, Datensouveränität gewährleistet

Was wir gelernt haben – und was als nächstes kommt

» **Starke Kundennähe:**

Enge Zusammenarbeit ist unerlässlich für tiefes Verständnis

» **Technisches Enablement:**

verbessert Entscheidungsfindung bei Zielsetzungen (was /wann möglich?)

» **Gezielte Weiterentwicklung:**

mit generativen Methoden, multimodale Services

» **Gelebte Interaktivität:**

Rechtsakte durch höhere Beteiligung lebendiger gestalten

Vielen Dank!

**Besuchen Sie uns auch gerne
an unserem EBCONT Stand!**



Frederick Bednar

Data Analytics & Data Science
Consultant, EBCONT

