Anonymization of court decisions in Austria

Webinar on the use of AI in the justice field

Martin Hackl, Chief Digital Officer
David Steinbauer, Project Lead Anonymization of Court Decisions

29th March 2021

Anonymization of court decisions: Current legal situation

Publication of court decisions

- Judgments made by Austrian courts are published in the Legal Information System of the Republic of Austria and can be found at www.ris.bka.gv.at
- An obligation by law to publish decisions only exists for the Supreme Court (§ 15 OGH-G). Guidelines for anonymization are also fixed here: names, addresses, etc. – everything that can give a conclusion to the specific case. (But not the names of judges, lawyers etc.)
- § 38a GOG: use of the OGH-G according to the personnel and technical conditions and as far as the decisions are of general interest that goes beyond the individual case.

Challenges

- The Anonymization of court decision is a manual task, although supported by specific word functions, but still takes a lot of time
- Insufficient human resources at the courts
- As a result, publications are almost exclusively published by the Supreme Court.

Approach

- Using multiple technologies from the field of Artificial Intelligence to automate the anonymization task

Project Overview from of Technical Perspective

Objective

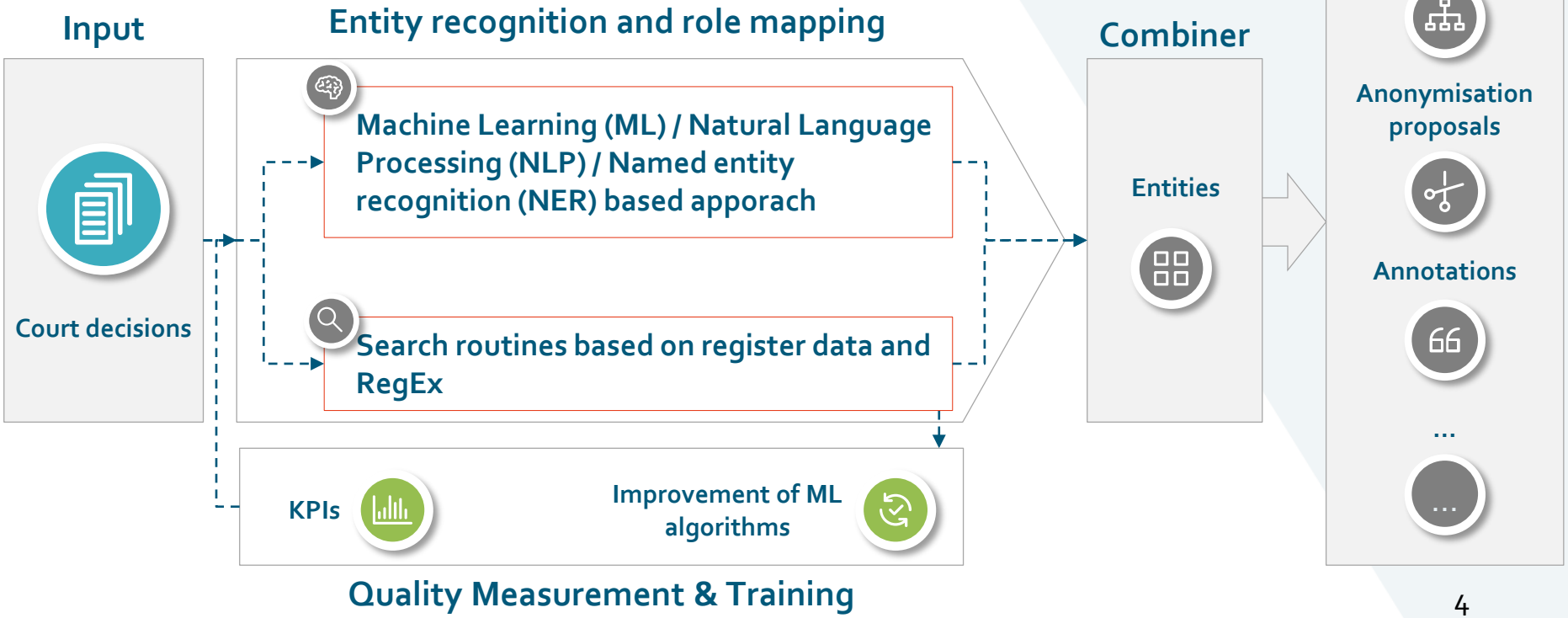
- Identification of entities / metadata in court decisions
- Automation of anonymization in accordance with the respective legal requirements.
- Determination of the anonymization quality through a quality measure

Project Approach

To achieve this goal, several approaches have been implemented and combined:

- Approach 1: Search based (Case Data, RegEx)
- Approach 2: Machine Learning (ML)/Natural Language Processing (NLP)
- Approach 3: Training and improvement of applied custom ML algorithms and quality measurement

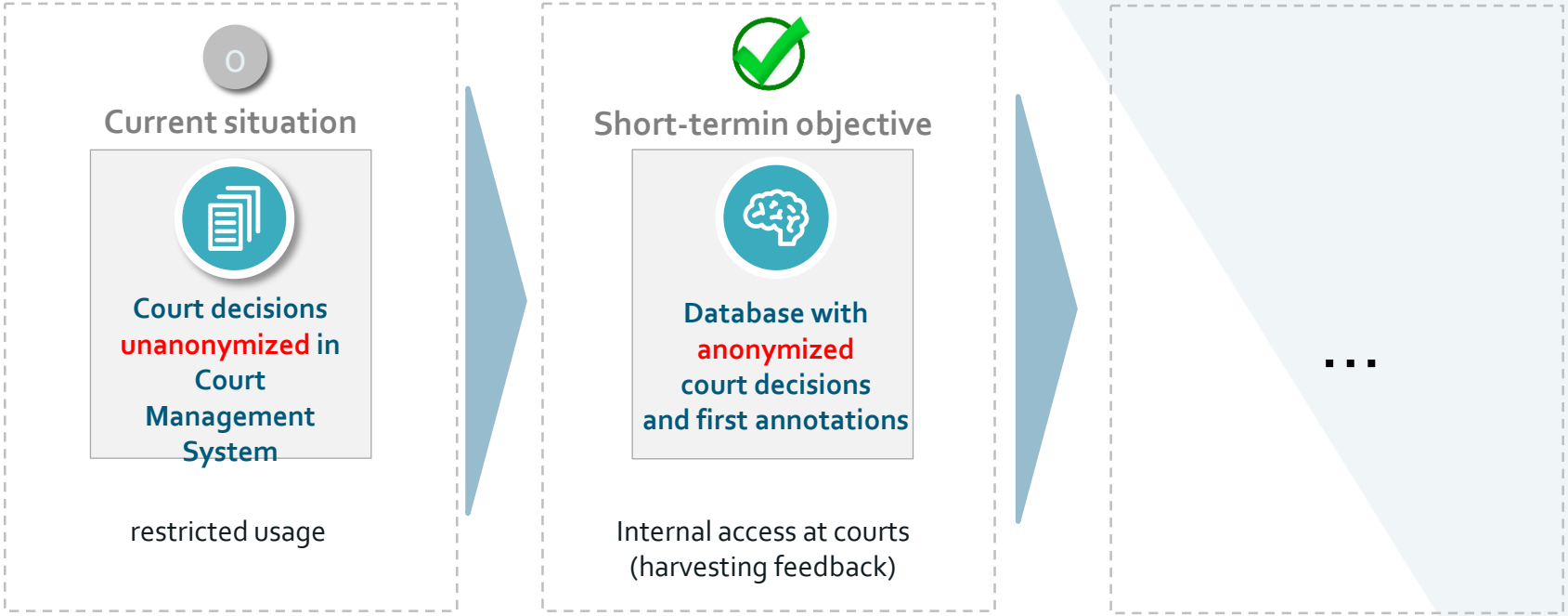
High level architecture



Milestone 1: (Internal) publication of anonymized decisions

- Since December 2020, court decisions have been anonymized for internal judicial purposes using the presented algorithms and are available to the local court.
- After a fully automated anonymization the files are published to an internal full-text search engine for further usage.
- The access to these published and anonmyized decisions is restricted to the reffering court
- Subsequently, suitable classified documents can be downloaded and used for the current case. Otherwise the user can mark the decision and leave feedback for the anonymization quality.

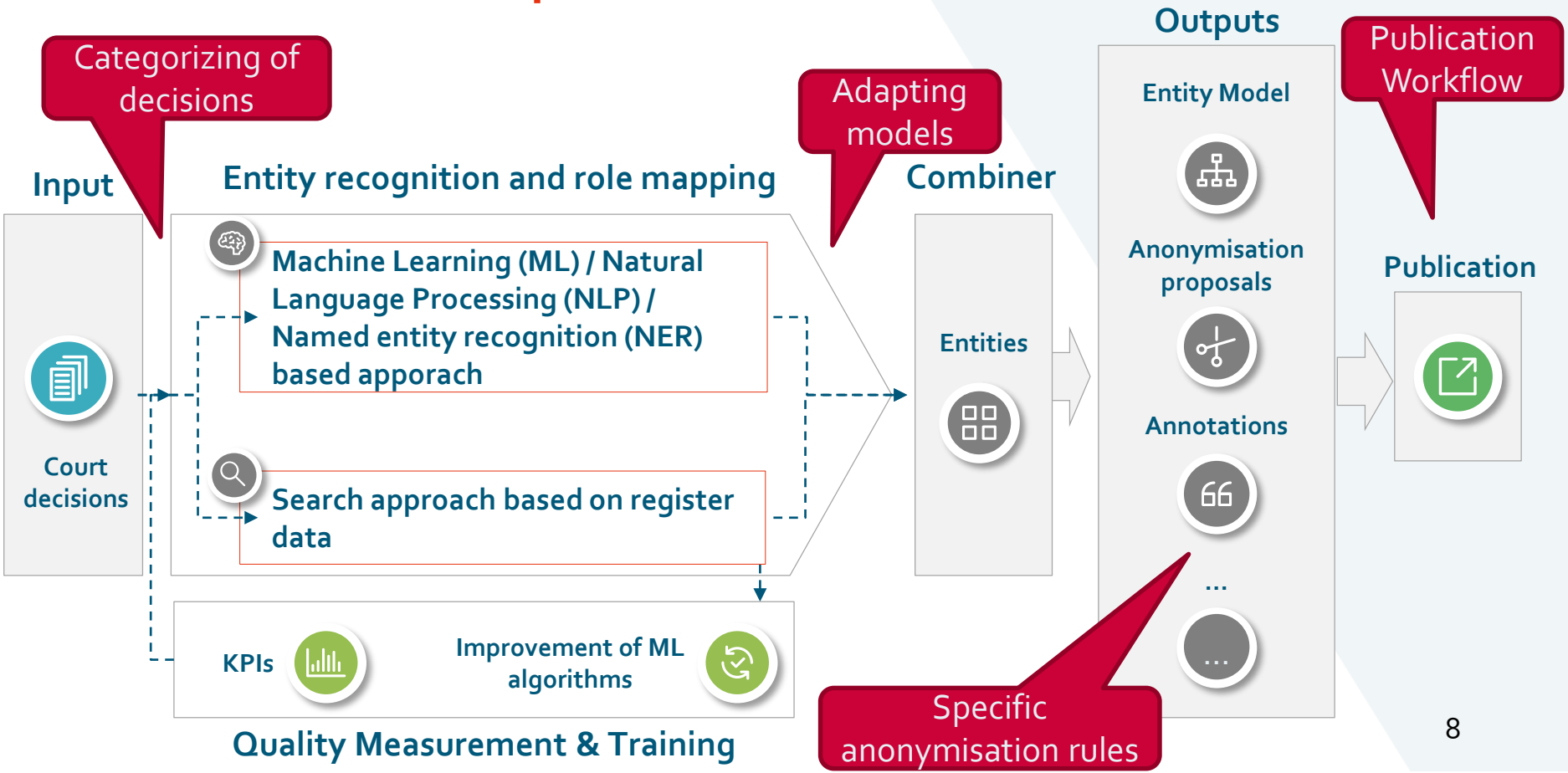
From vision to reality



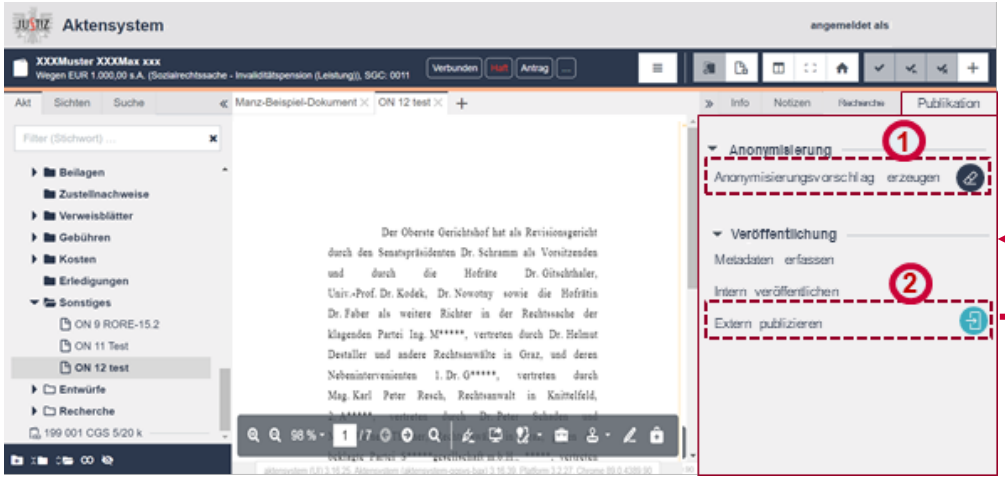
Project approach - General

- Pre-analysis of database (consisting of approx. 66k anonymised and 120k non-anonymised court decisions + metadata) and data-structure
- Selection of potential methods and algorithms
- Design of a search based model using register data of legacy core application
- Integration of already existing and on demand created lists of technical terms and definition of a domain model
- Design of further models using NLP/NER (named entity recognition) and machine learning for performing specialised model training
- Combination and evaluation of results

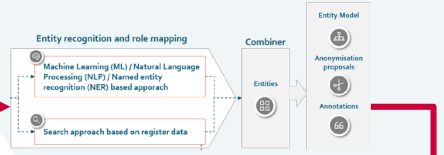
Identified Opimization Points to Architecture



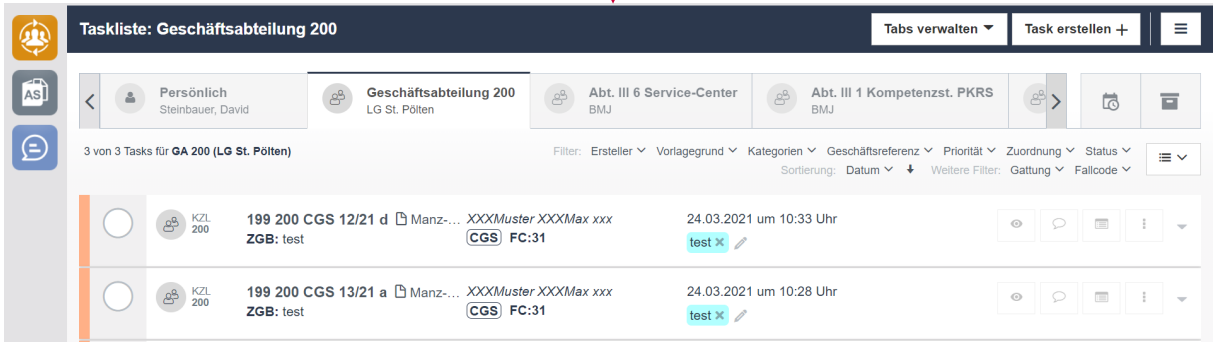
Target state for publication & research



1 Initiating anonymisation propsal directly from file management system

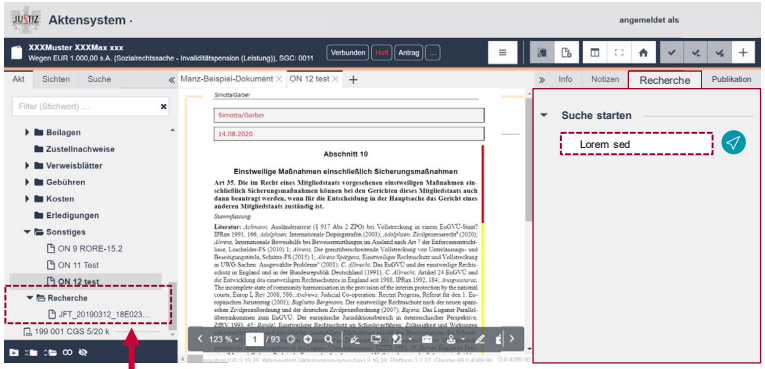


2 Sending task for human file check before internet publicaton

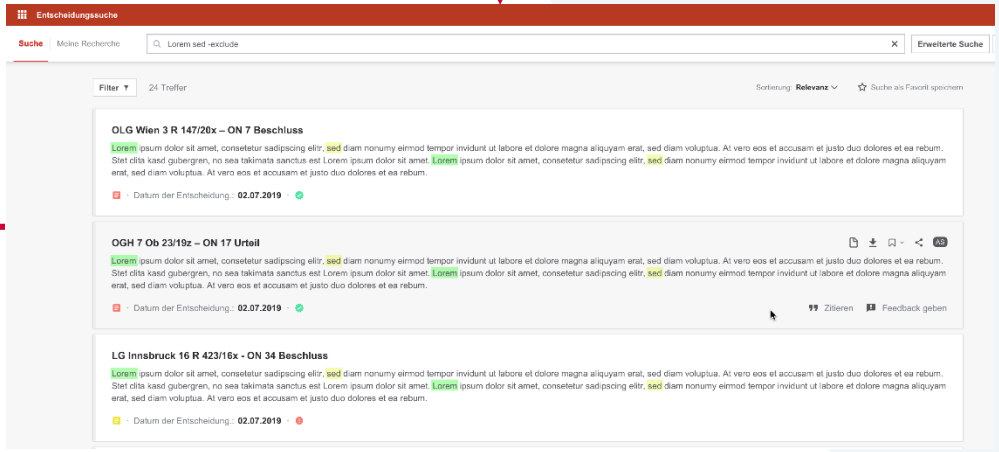


- Task inbox with reference to decision to check anonymisation quality before publication
- Anonymisation quality is estimated by system upfront to offer some guidance

Target state for publication & research



While working on a digital file, users can directly start a research for relevant cases in the anonymized decision collection



After researching users can transfer relevant search results directly in the research folder within the digital file

Lessons learned and outlook

- Combining many models brings improved results comparing to single solution approaches
- Shifting focus from recall-optimisation to precision-maximisation

Search based approach:

- Further pre-processing / sanitizing of the register data
- Expansion of correspondence (dependency parsing, RDF structure) and mapping for synonyms, etc.

Machine Learning approach:

- Additional trainings for specific corpora and better differentiation of roles
- Evaluation of other corpora and constant re-training of ML models based on feedback
- Quantitative benchmarking tool to objectively evaluate model re-training effects

- Seamless integration between anonymisation and everyday-work in the file digital management system offering considerable benefits
- Intelligent tagging to enable smart search offerings with related content and decisions

From vision to reality

0

Current situation

Court decisions unanonymized in Court Management System

restricted usage

1

Short-termin objective

Database with anonymized court decisions and first annotations

Internal access for all Austrian courts (harvesting feedback)

2

In 2022

Database with anonymized court decisions and extended annotations

Public available and free of charge

Digitization: a transition to exponential times

Thank you for your attention!

Mag. Martin Hackl (CDO)
martin.hackl@bmj.gv.at

Mag. David Steinbauer, MSc.
david.steinbauer@bmj.gv.at