

Towards FAIR Data in Legal Domain

Bakhtiyar Kirgeyev¹, Alsayed Algergawy¹

¹ Chair of Data and Knowledge Engineering, University of Passau, Passau, Germany

Introduction:

FAIR Significance: Ensures interoperable, findable, accessible, reusable legal data that fosters transparency, collaborative research, and standardized datasets.

The Goal: Assess the current FAIR practices in the legal domain to determine whether improvements are needed.

Legal Interoperability: Legal interoperability refers to the ability to exchange and reuse data across different legal frameworks, policies, and jurisdictions.

Challenges:

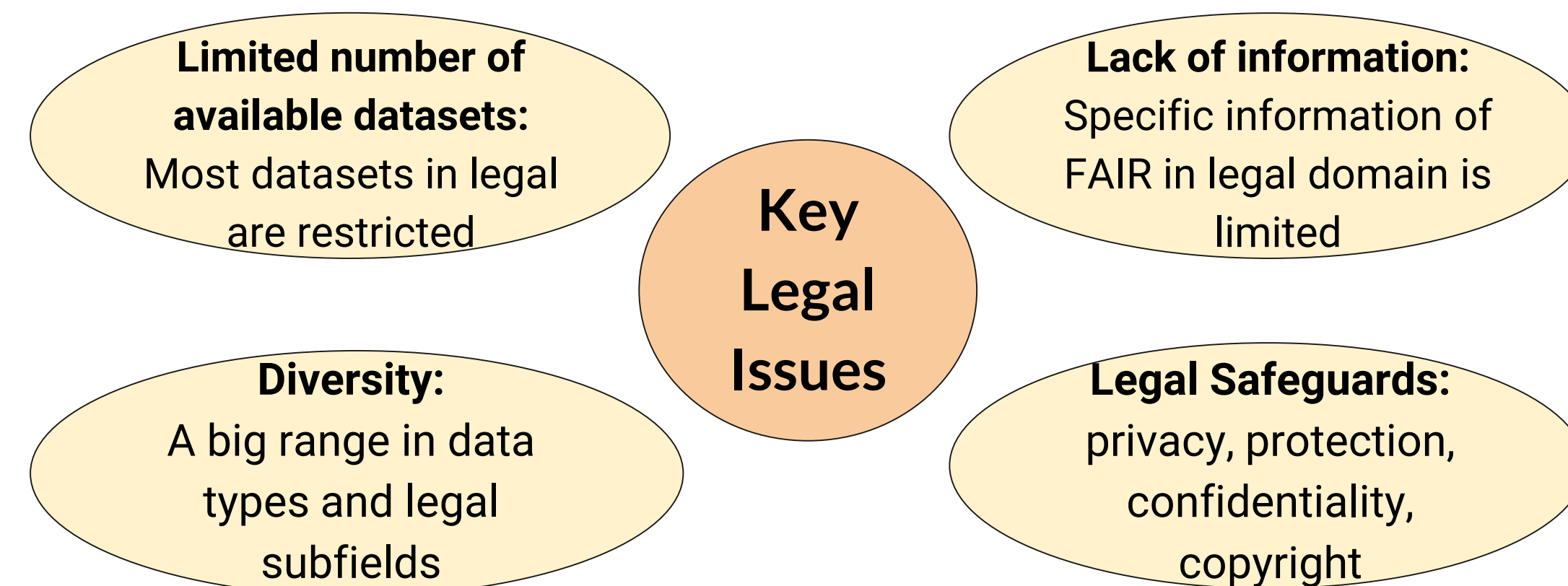


Fig1: Key Challenges

Motivation:

Enhance FAIR practices in the legal domain by addressing provided challenges, aiming for transparency, collaboration, and standardized datasets with ensuring interoperability and compliance.

ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	...	Ward	Community Area	FBI Code	X Coordinate	Y Coordinate
79	HZ250507	05/03/2016 11:33:00 PM	026XX W 23RD PL	0486	BATTERY	DOMESTIC BATTERY SIMPLE	APARTMENT	True	...	28.0	30.0	08B	1159105.0	1888300.0
80	HZ250491	05/03/2016 11:30:00 PM	073XX S HARVARD AVE	1310	CRIMINAL DAMAGE	TO PROPERTY	APARTMENT	True	...	17.0	69.0	14	1175230.0	1856183.0
81	HZ250479	05/03/2016 12:15:00 AM	024XX W 63RD ST	041A	BATTERY	AGGRAVATED: HANDGUN	SIDEWALK	False	...	15.0	66.0	04B	1161027.0	1862810.0
90	HZ250370	05/03/2016 09:07:00 PM	082XX S EXCHANGE AVE	0486	BATTERY	DOMESTIC BATTERY SIMPLE	SIDEWALK	False	...	7.0	46.0	08B	1197261.0	1850727.0
92	HZ250517	05/03/2016 11:38:00 PM	001XX E 75TH ST	5007	OTHER OFFENSE	OTHER WEAPONS VIOLATION	PARKING LOT/GARAGE(NON.RESID.)	True	...	6.0	69.0	26	1178696.0	1855324.0

Fig2: Dataset Sample

Methodology:

Legal Domain Analysis: The method includes literature review.

Data Preparation: Public datasets related to crimes in US cities.

Data Collection Analysis: Practical evaluation for primary data, data type, metadata, and legal interoperability.

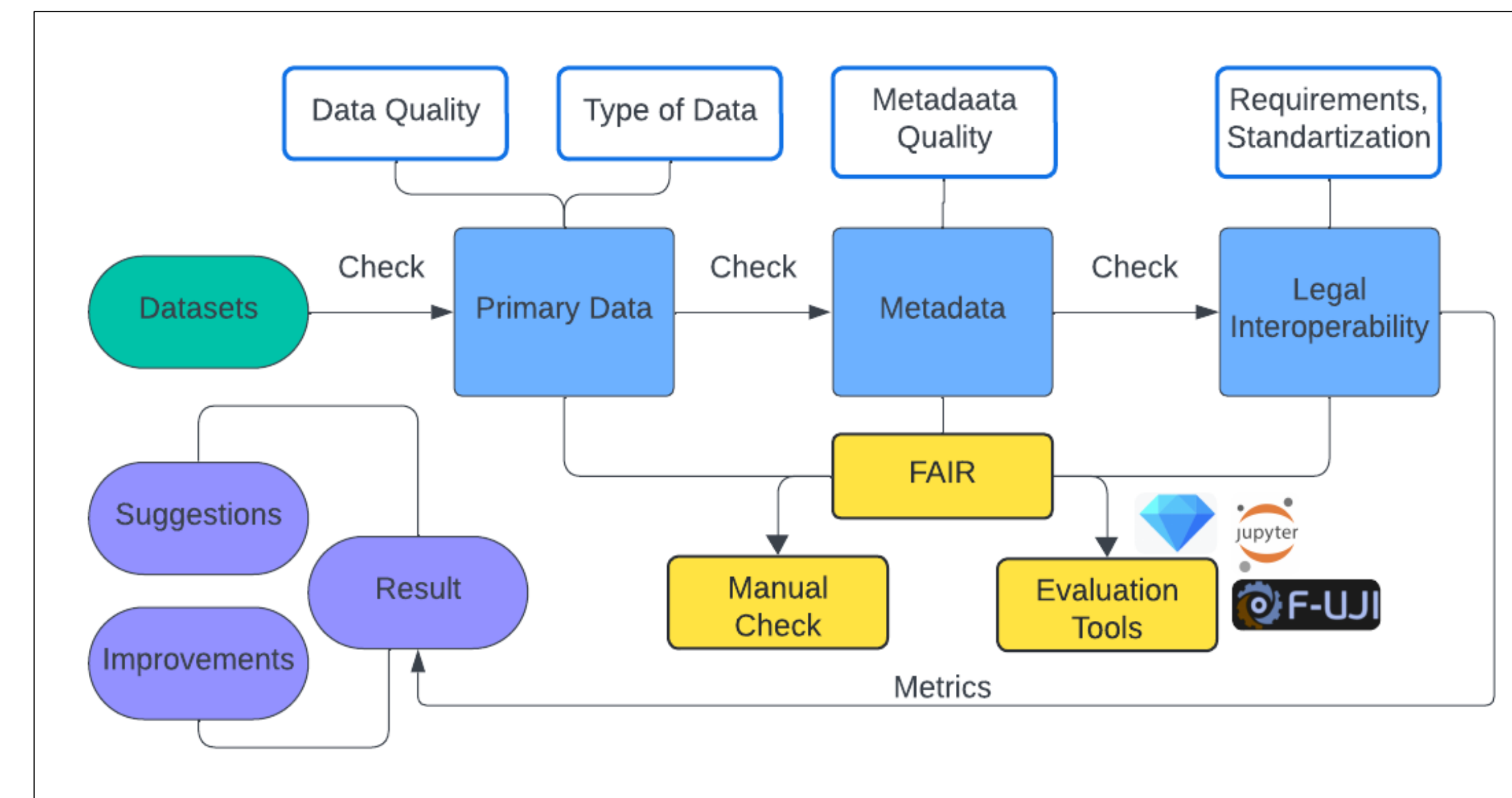


Fig3: Experiment Pipeline

Evaluation:

	Boston (B)	New York (N)	Chicago (C)	Orlando (O)
1	Police Department			Kaggle
2	18	24	23	13
3	726 650	2 838 250	187 129	198 370
4	Yes	Yes	Not Found	Not Found
5	49%	47%	52%	60%
6	CSV, TSV, JSON, XML	CSV, TSV, RDF, RSS, XML	CSV, TSV, RDF, RSS, XML	CSV

- 1 - Source
- 2 - Attributes
- 3 - Missing Data
- 4 - Inconsistency

- 5 - F-UJI FAIR Metrics Result²
- 6 - Available Formats

² <https://www.fairsfair.eu/>

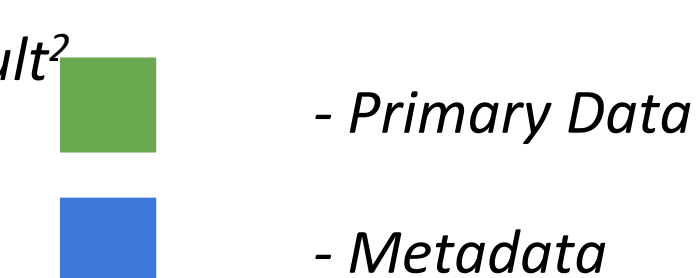


Table 1: Crimes Datasets

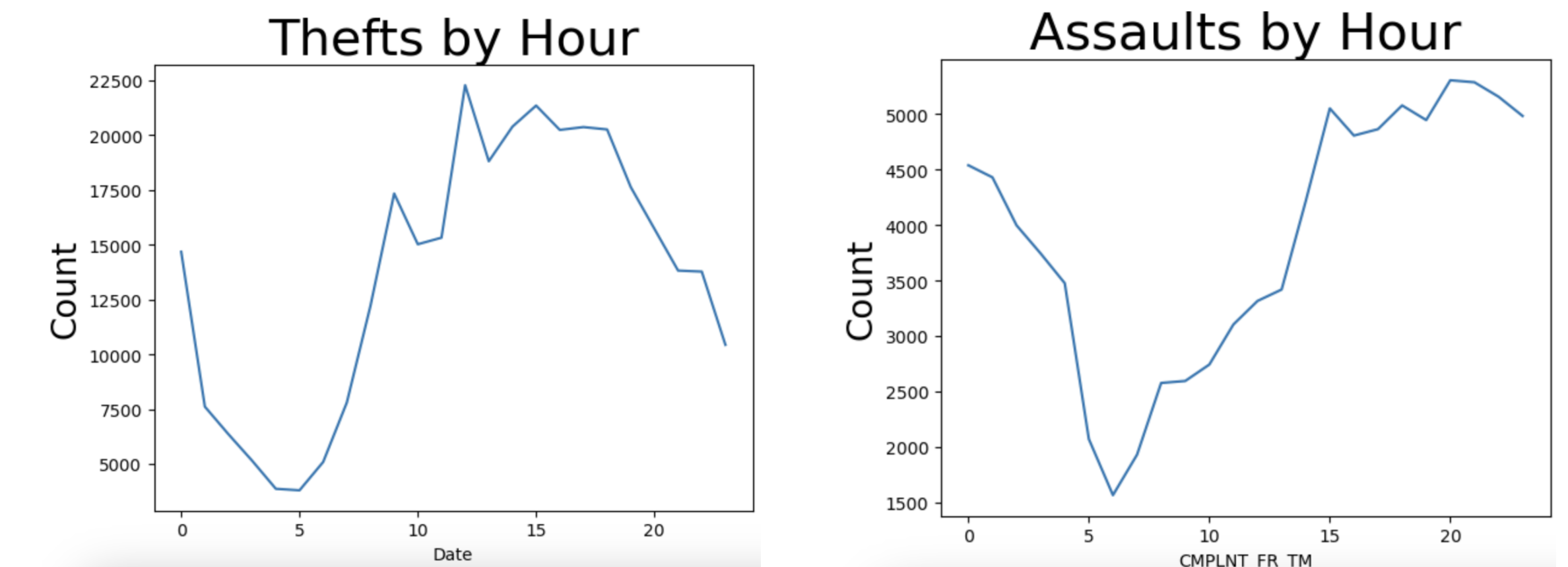


Fig4: Visualization for Finding Anomalies

Name of Attributes	B	N	C	O
Time and Date	+	+	+	+
Offence Description	+	+	+	+/-
Offence Classification Code	+	+	+/-	-
District	+	+/-	+	+
Location (Latitude, Longitude)	+	+	+	+

Table 2: Pattern of Common Attributes

Suggested Improvements:

1. Implement appropriate metadata standards that should be applied to various subfields.
2. Metadata should be devoid of any legal restrictions.
3. Data should follow anonymization to protect sensitive data.
4. Metadata should cover a standardized declaration of the "legal safeguards" in both readable formats.

Conclusion:

The suggested measures will contribute to improving FAIR data practices in the legal domain, fostering interoperability, transparency, and effective dataset utilization.