

Towards Self-Service Analysis and Integration of Publicly Available Data Sources for Corporate Security

Malte Brockmann, 8th of February 2019, Munich, Kickoff Presentation - Master Thesis

Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
www.matthes.in.tum.de



Initial Situation



Motivation



Research Question

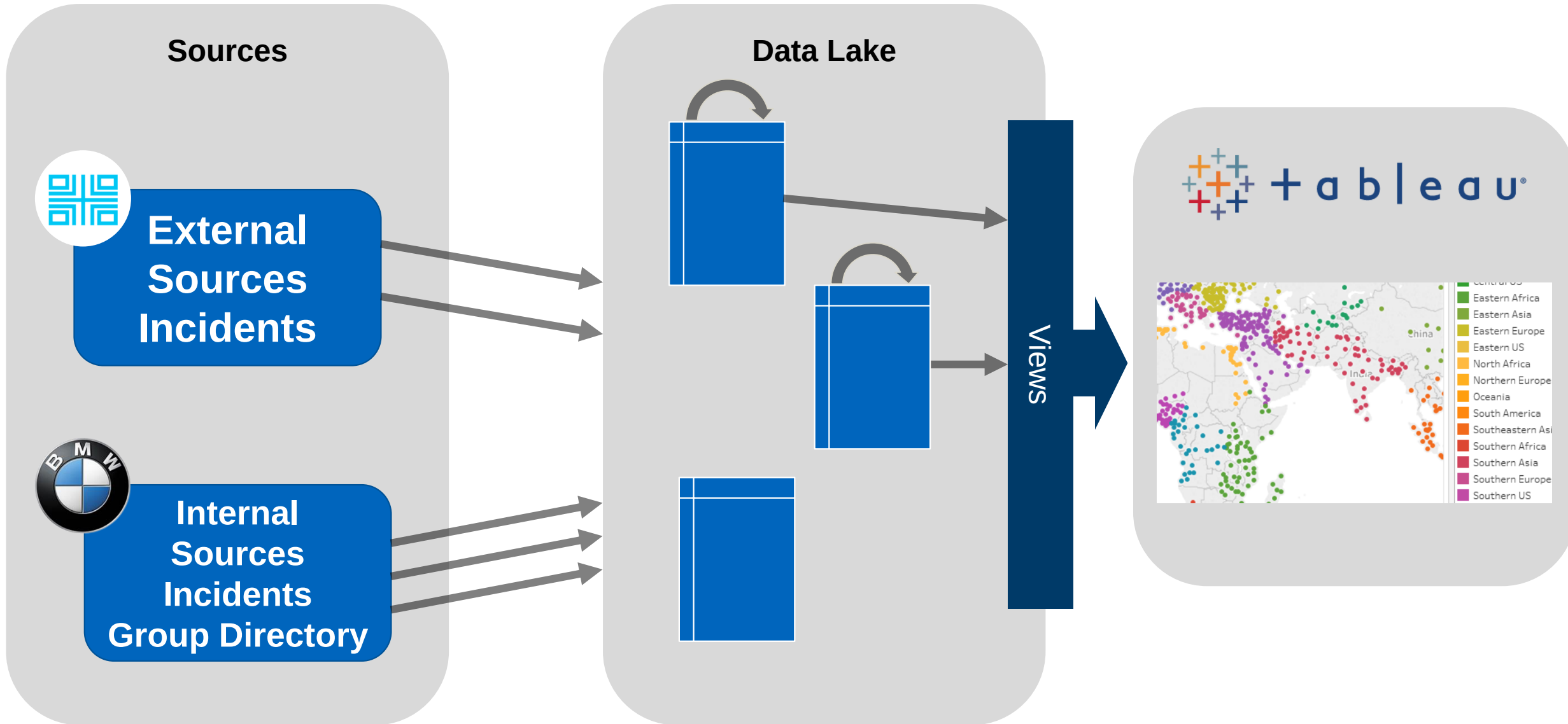


Approach



Timeline

Initial Situation



Current Situation for Changes

Corporate security department

Goals:

- Easy access to data sources
- Flexibility with existing data sources

To implement goals:
contact

IT department

Impediments:

- High complexity
- High costs
- Low process speed



Goal

Self-service 

Corporate security department

- Direct access to data sources
- Work with data independently

To implement goals:
contact

IT department

Motivation



GDELT 2.0	MonthYear		Year		FractionDate	Actor1Code	Actor1Name	Actor1CountryCode	Actor1KnownGroupCode	Actor1EthnicCode	Actor1Religion1Code	Actor1Religion2Code	Actor1Type1Code	Actor1Type2Code	Actor1Type3Code	Actor2Code	Actor2Name	Actor2CountryCode	Actor2KnownGroupCode	Actor2EthnicCode	Actor2Religion1Code	Actor2Religion2Code	Actor2Type1Code	Actor2Type2Code	Actor2Type3Code	IsRootEvent	EventCode	EventBaseCode	EventRootCode	QuadClass	GoldsteinScale	NumMentions	NumSources	NumArticles	AvgTone	:	
	526870433	20160403	201604	2016	20.162.548	LEG	CONGRESS	null	null	null	null	null	LEG	null	null	null	null	null	null	null	null	null	null	null	null	1	190	190	19	4	-10.0	65	4	45	-3.87019321369491	...	
	459554803	20150821	201508	2015	20.156.329	COP	DEPUTY	null	null	null	null	null	COP	null	null	null	null	null	null	null	null	null	null	null	null	1	15	15	1	1	0.0	365	37	345	-7.60546699954661	...	
	294461014	20140422	201404	2014	20.143.068	USA	TWIN FALLS	USA	null	null	null	null	null	null	JUD	PUBLIC DEFENDER	null	null	null	null	null	JUD	null	null	1	84	84	8	2	7.0	5	1	5	0.0	...		



	Edition	GLOBAL ID	Series code	Series	Series unindented	Attribute	ALB
	Edition	GLOBAL ID	Code GCR	Series	Series unindented	Attribute	Albania
1Global Competitiveness Index	2017-2018	EOSQ051	1,01	1.01 Property rights, 1-7 (best)	Property rights, 1-7 (best)	Value	3,70733666
1Global Competitiveness Index	2017-2018	EOSQ051	1,01	1.01 Property rights, 1-7 (best)	Property rights, 1-7 (best)	Period	Weighted average 2016-2017
1Global Competitiveness Index	2017-2018	EOSQ051	1,01	1.01 Property rights, 1-7 (best)	Property rights, 1-7 (best)	Rank	107
1Global Competitiveness Index	2017-2018	EOSQ051	1,01	1.01 Property rights, 1-7 (best)	Property rights, 1-7 (best)	Source	World Economic Forum; Executive Opinion Survey; www.weforum.org/gcr
1Global Competitiveness Index	2017-2018	EOSQ051	1,01	1.01 Property rights, 1-7 (best)	Property rights, 1-7 (best)	Source date	27/09/2017
1Global Competitiveness Index	2017-2018	EOSQ051	1,01	1.01 Property rights, 1-7 (best)	Property rights, 1-7 (best)	Note	Refer to the "Technical Notes and Sources" of The Global Competitiveness Report 2017-2018 for details about the computation methodology
2Global Competitiveness Index	2017-2018	EOSQ052	1,02	1.02 Intellectual property protection, 1-7 (best)	Intellectual property protection, 1-7 (best)	Value	3,38570881
2Global Competitiveness Index	2017-2018	EOSQ052	1,02	1.02 Intellectual property protection, 1-7 (best)	Intellectual property protection, 1-7 (best)	Period	Weighted average 2016-2017
2Global Competitiveness Index	2017-2018	EOSQ052	1,02	1.02 Intellectual property protection, 1-7 (best)	Intellectual property protection, 1-7 (best)	Rank	113
2Global Competitiveness Index	2017-2018	EOSQ052	1,02	1.02 Intellectual property protection, 1-7 (best)	Intellectual property protection, 1-7 (best)	Source	World Economic Forum; Executive Opinion Survey; www.weforum.org/gcr
2Global Competitiveness Index	2017-2018	EOSQ052	1,02	1.02 Intellectual property protection, 1-7 (best)	Intellectual property protection, 1-7 (best)	Source date	27/09/2017
2Global Competitiveness Index	2017-2018	EOSQ052	1,02	1.02 Intellectual property protection, 1-7 (best)	Intellectual property protection, 1-7 (best)	Note	Refer to the "Technical Notes and Sources" of The Global Competitiveness Report 2017-2018 for details about the computation methodology

Which open online sources can be used to gather information regarding company security?

- MIDAS

Which short lived information can be gather by a media monitoring tool?

- Ubermetrics

How can a data model for security data be developed?

- Relevant information for corporate security
- Metadata enrichment of data relevant for corporate security

Approach

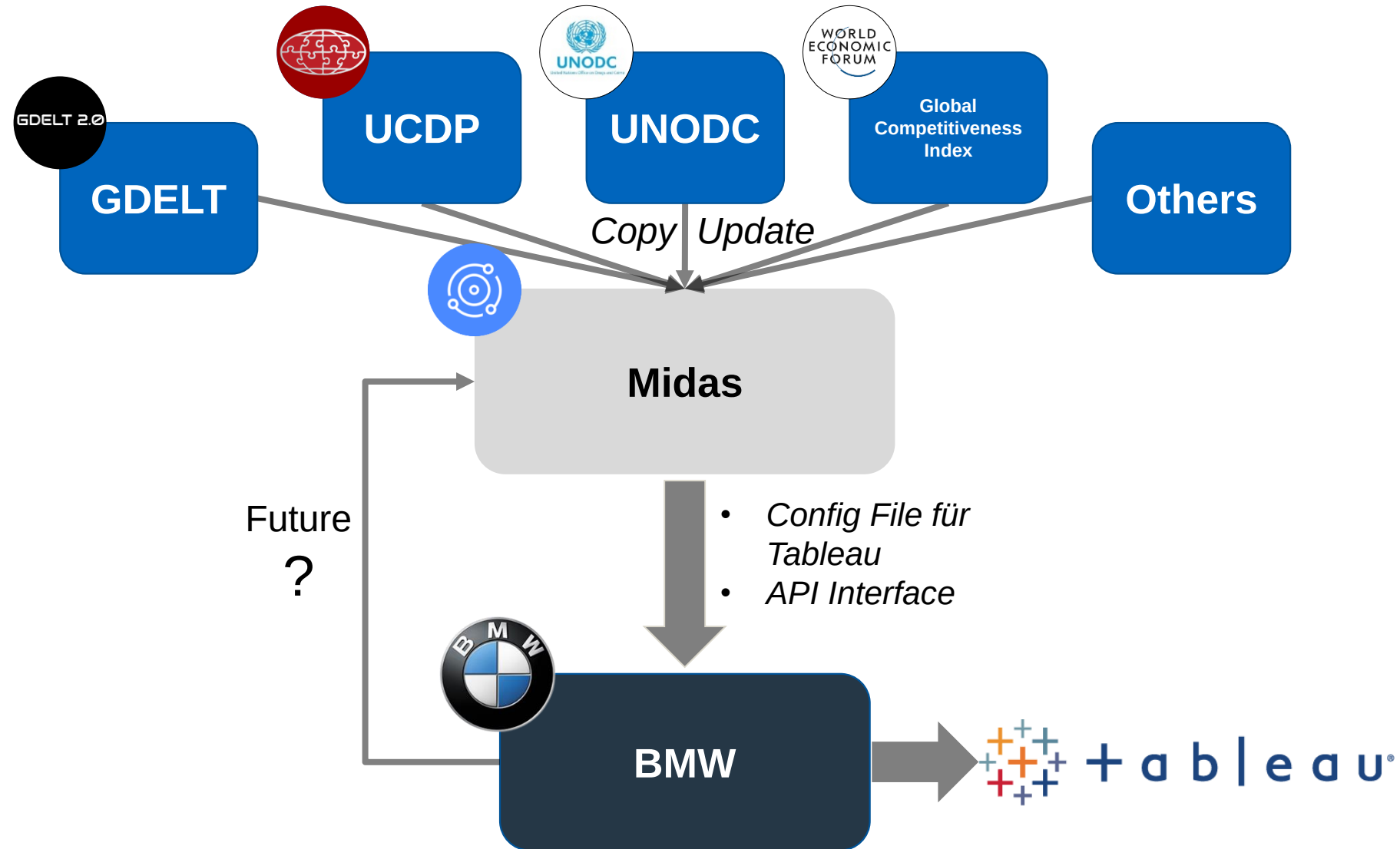
Which open online sources can be used to gather information regarding company security?

TO DO

- Integration of relevant external Sources into Midas
→ **Little integration expenses**

Idea

- Because streaming through data lake is not efficient
→ **Proposal of a new Architecture**





Input

Please input your query:

```
1 ( ("München" OR "Berlin" OR "Leipzig") NEAR/12 ("Anschlag" OR "Amoklauf" OR "Erdbeben" OR "Geiselnahme" OR  
"Einsatzkommando" OR "lethal" OR "Massaker" OR "Attentat" OR "Attentäter") AND NOT ( ("Amri" OR "Gericht"  
OR "Prozess") OR (((("PKW" OR "KFZ") NEAR/10 "Unfall")) OR (("ROUNDUP" OR "dpa/AFX" OR "RT @*" OR  
"Wochenzeitschau") :field:title OR "niedelne.de" :field:url) OR "inf" OR "1070" OR (TCPM:ccc credit:trust)
```

Output

Berlin : Angst vor Vulkanausbruch: Ursache für **Erdbeben** geklärt

Jan 27 2019, 19:49:00  www.wolfsburger-nachrichten.de Reach: 13,101  Germany

Ein **Erdbeben** der Stärke 4,2 hat die

Result for:  Ortsliste Allg NEAR Hoch-Risiko_2

Anschlag auf Ordnungsamt: Neun Fahrzeuge gehen in Flammen auf

Jan 21 2019, 10:56:16  www.neuepresse.de Reach: 109,574  Germany

10:56 21.01.2019 Panorama **Berlin - Anschlag** auf

Result for:  Ortsliste Allg NEAR Hoch-Risiko_2

Which short lived information can be gather by a media monitoring tool?



Issues

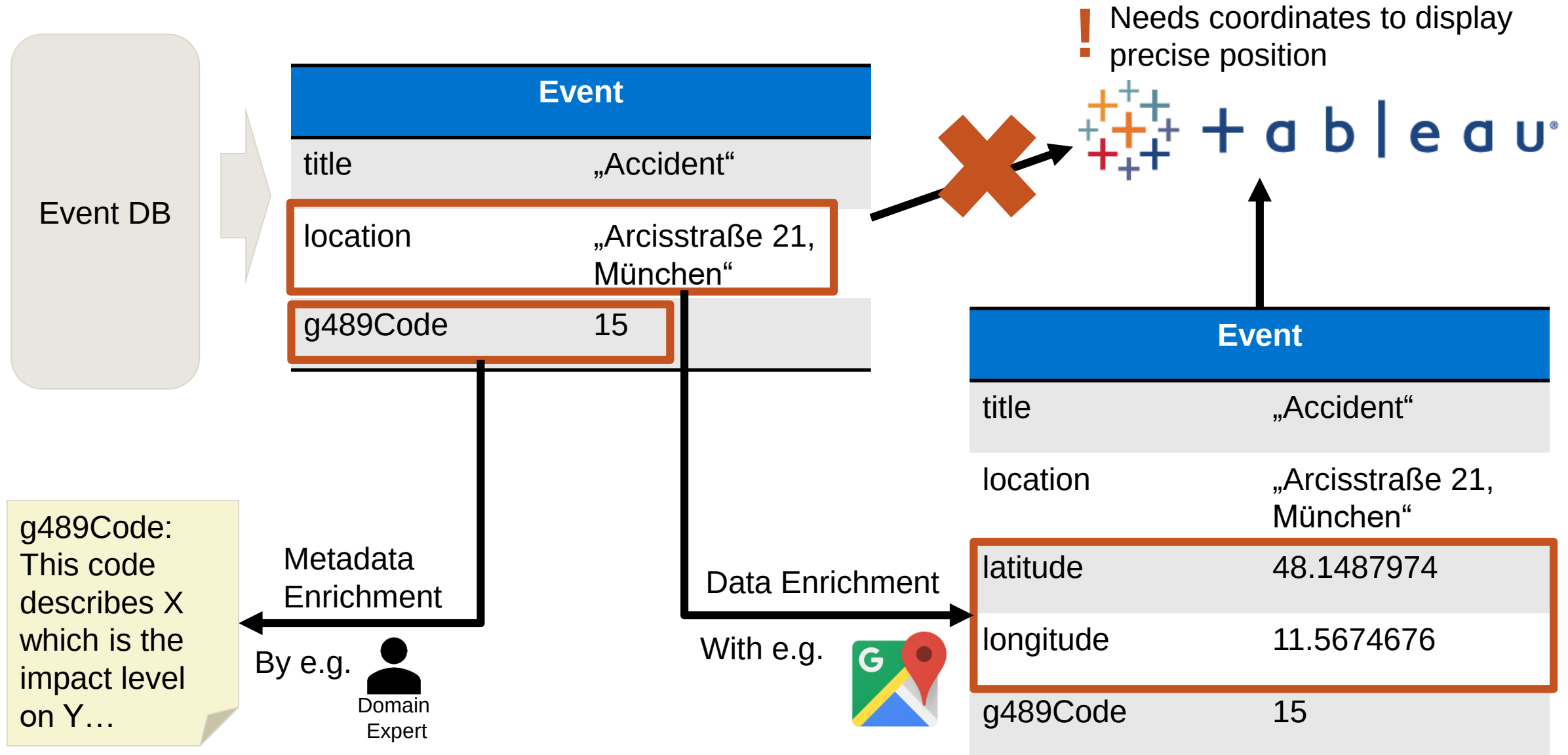
- Content about past events
- False positives through unfortunate hits

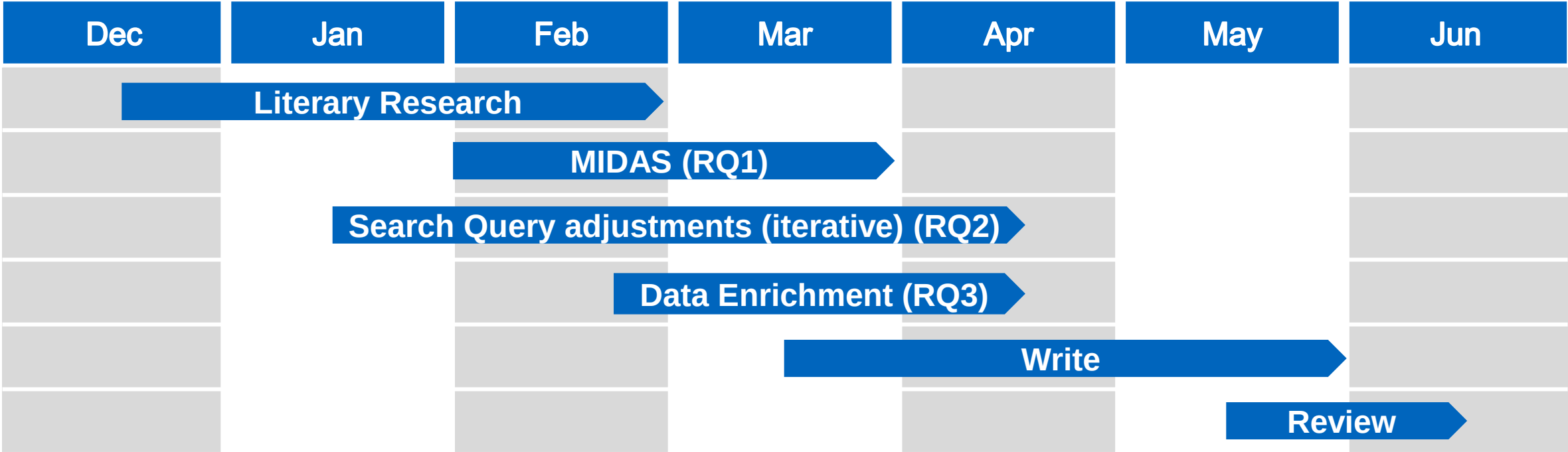
Iterative Approach

- Literary research on effective search
- Finding pattern in false positive

Approach

How can a data model for security data be developed?





GDELT, “The gdel project,” Website, 2019, online available at <https://www.gdelproject.org/>; Retrieved January 29, 2019.

World Economic Forum, “Global competitiveness index 2017-2018,” Website, 2019, online available at <http://reports.weforum.org/global-competitiveness-index-2017-2018/downloads/>; Retrieved January 29, 2019.

U.S. National Science Foundation, “Cameo event data codebook,” Website, 2013, online available at <http://eventdata.parusanalytics.com/data.dir/cameo.html>; Retrieved January 30, 2019.

P. A. Schrod, “Cameo: Conflict and mediation event observations event and actor codebook,” Pennsylvania State University, 2012.

P. Hitzler, M. Krötzsch, S. Rudolph, and Y. Sure, Semantic Web: Grundlagen. Springer-Verlag, 2007.



B.Sc.

Malte Brockmann

Technische Universität München
Faculty of Informatics
Chair of Software Engineering for Business
Information Systems

Boltzmannstraße 3
85748 Garching bei München

malte.brockmann@tum.de
www.matthes.in.tum.de

