

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

Malte Brockmann, 17th of June 2019, Munich, Final Presentation - Master Thesis

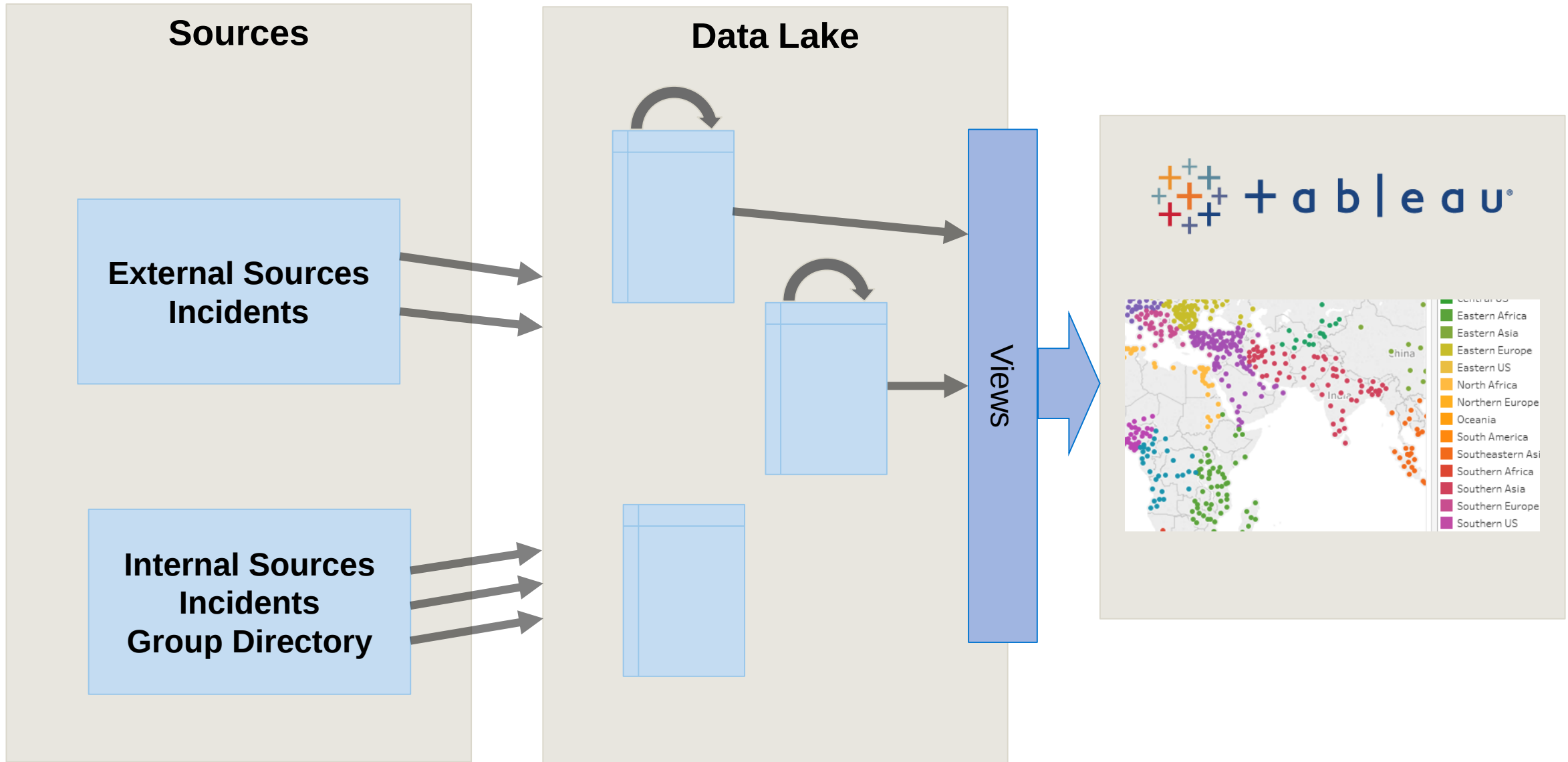
Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
www.matthes.in.tum.de

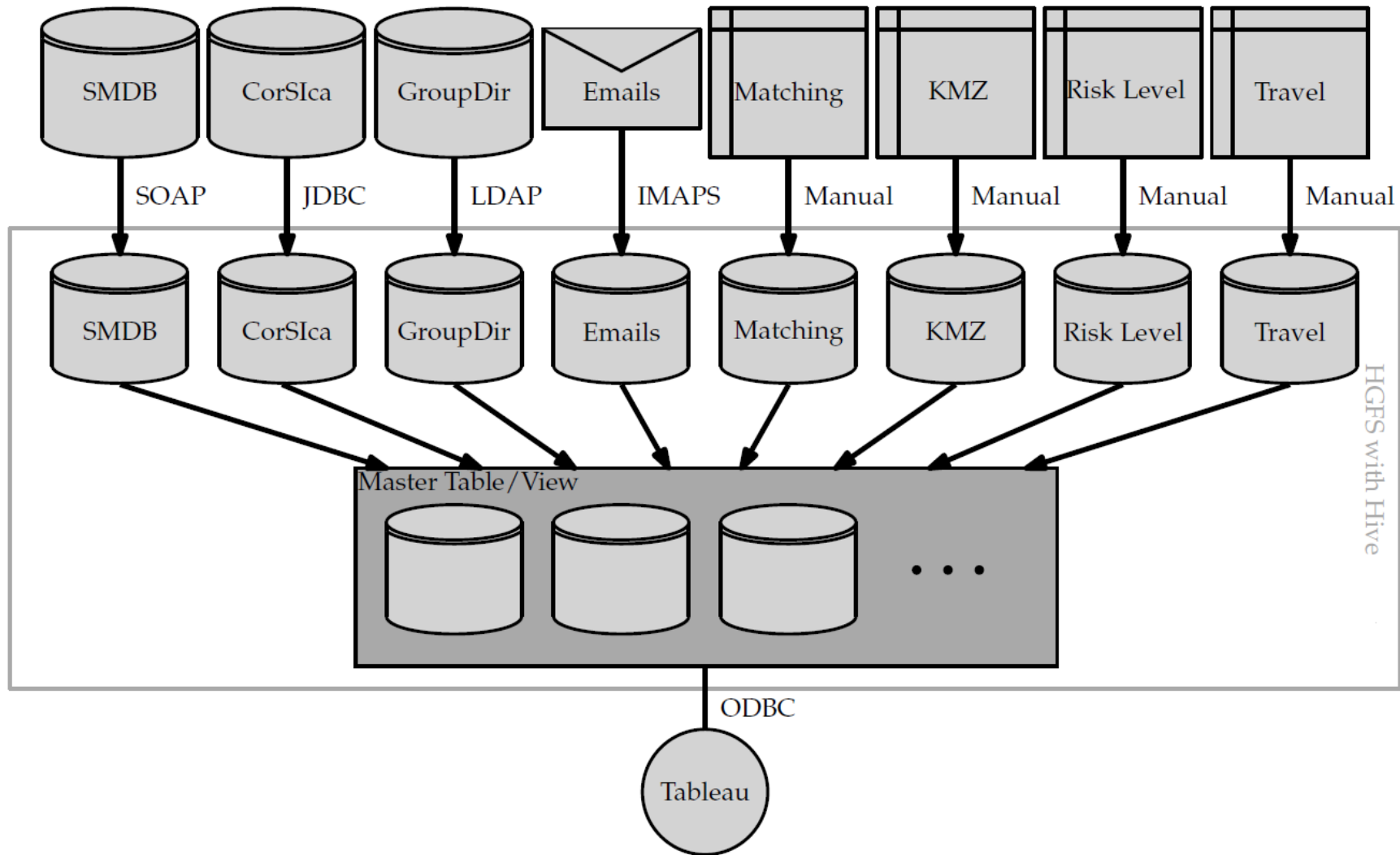
Introduction

Research Questions

Approach

Conclusion





Current Situation for Changes

Corporate security department

Goals:

- Easy access to several external 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

Introduction

Research Questions

Approach

Conclusion

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

Which short-lived information can be gathered by a media monitoring tool for corporate security?

How can the available data be enriched to allow for self-service?

Introduction

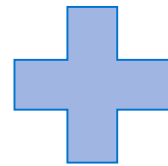
Research Questions

Approach

Conclusion

Access to internal as well as external Data

Enable fast queries through data virtualization



Make data usable to enable Self-Service

Dremel

- “... is a scalable, interactive ad-hoc query system for analysis of read-only nested data.” [1]



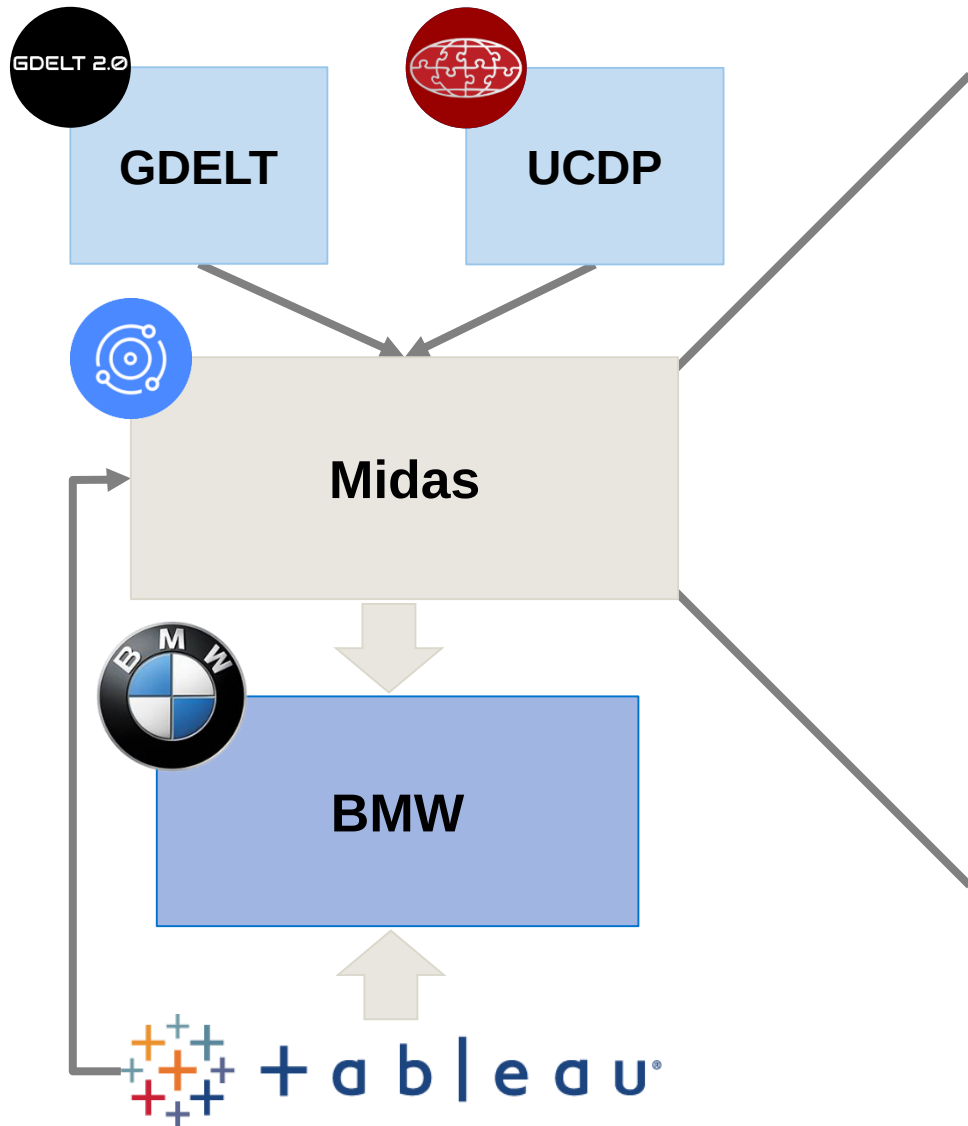
[2]



Midas

Approach

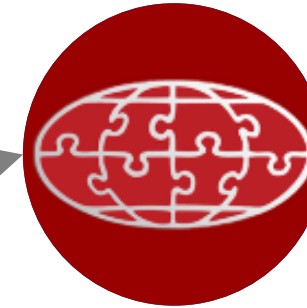
Midas at BMW



Additions to Apache Drill by Midas

- Command & Control
- Network Request Manager
- Adaptive Cache
- Midas Data Hub
- Metadata Repository

Category	Name
Political Stability	Fragile States Index
Political Stability	Weltbank
Political Stability	Bertelsmann Stiftung
Political Stability	EIU – Democracy Index
Political Stability	GCI
Unrest	HIK
Unrest	Fragile States Index
Armed Conflict	HIK
Armed Conflict	UCDP
Armed Conflict	SIPRI
Armed Conflict	Global Peace Index
Armed Conflict	Council of Foreign Relations (US Brille)
Terror	Global Terrorism Index
Terror	Global Terrorism Database
Terror	RAND
Crime	UNODC
Crime	UNODC Intentional Homicide
Crime	OECD Länder Personal Safety
Crime	Interpol
Crime	Europol
Natural Hazard	IFHV - World Risk Report
Natural Hazard	GEM
Natural Hazard	EM-DAT
Transport	WHO
Transport	ETSC- nur Europa
Corruption	Transparency International-Corruption Perception index
Corruption	UNODC
Other	GDELT



Uppsala Conflict Data Program (UCDP)



Global Data on Events, Location and Tone (GDELT)



- Created by Kalev H. Leetaru
- Global geolocated events since 1979
- Updates every 15 Minutes [3]

“The GDELT Project is a realtime network diagram and database of global human society for open research” [4]

GDELT 2.0 Event Database [5]

Dataset containing event data including data about the actors

GDELT 2.0 Global Knowledge Graph (GKG) [5]

Dataset containing extracted numbers and entities from news reports



- Created in the mid 1980
- At the Department of Peace and Conflict Research at the Uppsala University
- Updates once a year for the previous year [6]

“...is the world’s main provider of data on organized violence and the oldest ongoing data collection project for civil war...” [7]

- UCDP Georeferenced Event Dataset (GED) [8]
- UCDP/PRIO Armed Conflict Dataset
- UCDP Conflict Termination Dataset

- UCDP One-sided Violence Dataset
- UCDP Non-State Conflict Dataset
- UCDP Battle-Related Deaths Dataset



- Founded 2011
- 400 million sources
- Queries are build with Boolean logic with some additions [9].
- In use at BMW since 2018 for Corporate Security
- Separated into different sectors

Initial Situation

~ 120 Hits per Week
Close to non-relevant hits
Searched everything

Issues:

Received hits are focused on the past
Many false positives

Goals:

Lower the number of hits
Increased number of relevant hits

Outcome: Goals could be achieved but the future for crime monitoring via ubermetrics is unclear

Challenges for big data by Kanehisa et al. [10]:

- Data Access and sharing of information
- Analytical challenges
- Required Skill

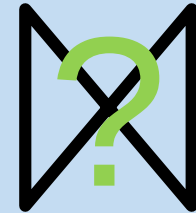
Enrichment through metadata:

...	EventCode	...
...	0874	...
...	036	...
...	...	...

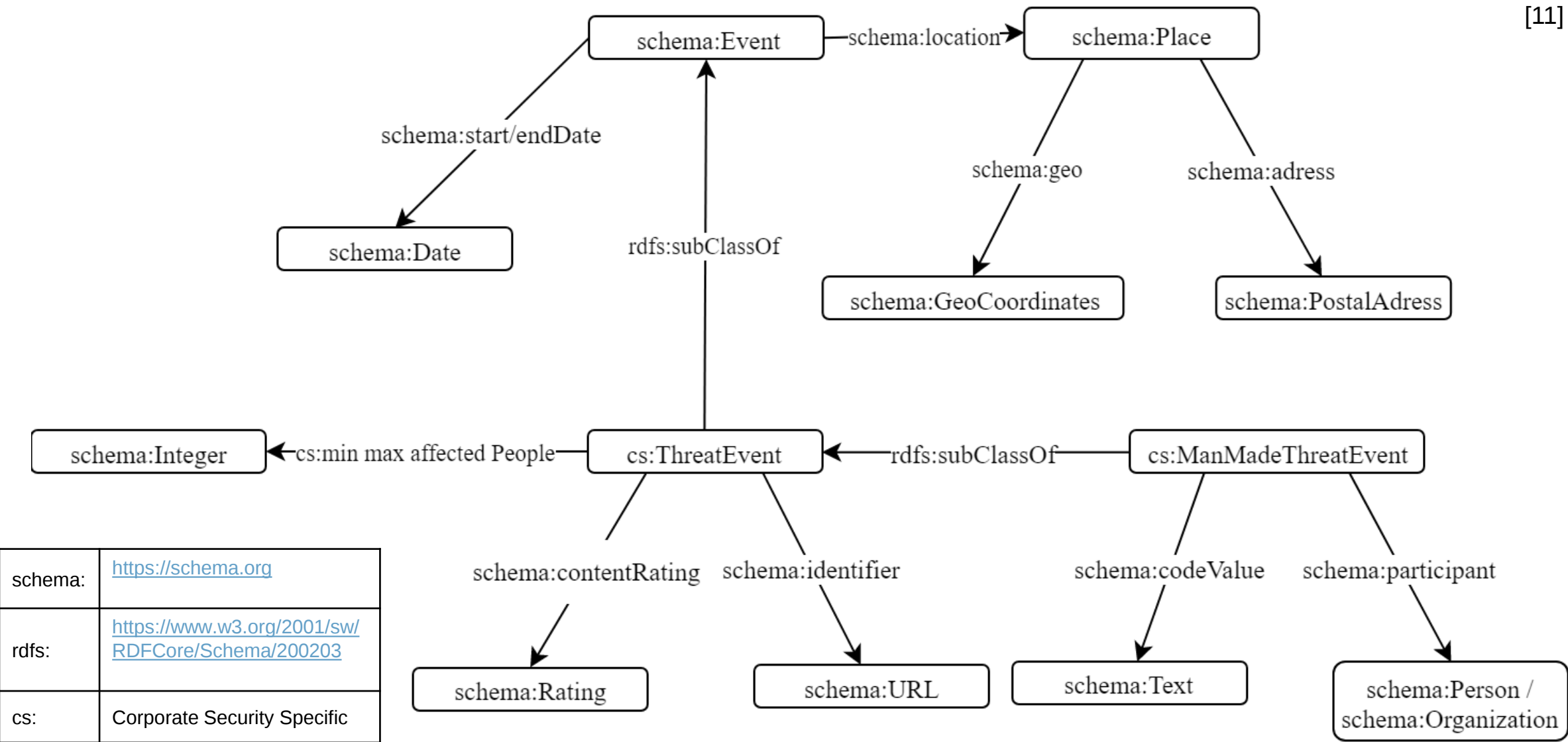
“This is the raw CAMEO action code describing the action that Actor1 performed upon Actor2. NOTE: it is strongly ...”

Enrichment through a data model:

...	ActionGeo_CountryCode	...
...	US	...
...	GM	...
...	...	...



...	country	...
...	United Stats of America	...
...	Germany	...
...	...	...



Introduction

Research Questions

Approach

Conclusion

The idea of virtualization of internal and external data can be extended for self-service through features like metadata or the data hub

The proposed ontology does provide the foundation for future development regarding corporate security.

Long term goal:

- Building a data driven Ontology
- Automated join recommendation

- [1] Sergey Melnik, Andrey Gubarev, Jing Jing Long, Geoffrey Romer, Shiva Shivakumar, Matt Tolton, and Theo Vassilakis. Dremel: Interactive analysis of web-scale datasets. In Proc. of the 36th Int'l Conf on Very Large Data Bases, pages 330–339, 2010. URL <http://www.vldb2010.org/accept.htm>.
- [2] The Apache Software Foundation. Apache drill. Website, 2019. Online available at <https://drill.apache.org/>; Retrieved February 26, 2019.
- [3] The GDELT Project. The gdelt story. Website, 2019. Online available at <https://www.gdeltproject.org/about.html>; Retrieved March 18, 2019.
- [4] The GDELT Project. Querying, analyzing and downloading. Website, 2019. Online available at <https://www.gdeltproject.org/data.html>; Retrieved March 18, 2019.
- [5] The GDELT Project. The gdelt event database data format codebook v2.0. Document, 2015. Online available at http://data.gdeltproject.org/documentation/GDELT-Event_Codebook-V2.0.pdf; Retrieved March 26, 2019.
- [6] Uppsala Conflict Data Program. About background. Website, 2019. Online available at <https://www.pcr.uu.se/research/ucdp/about-ucdp/>; Retrieved April 01, 2019.
- [7] European Commission. Uppsala conflict data program. Website, 2019. Online available at https://ec.europa.eu/knowledge4policy/dataset/ds00121_en; Retrieved March 18, 2019.
- [8] Ralph Sundberg and Erik Melander. Introducing the ucdp georeferenced event dataset. Journal of Peace Research, 50(4):523–532, 2013.
- [9] Ubermetrics Technologies GmbH. Ubermetrics - Über uns. Website, 2019. Online available at <https://www.ubermetrics-technologies.com/de/uberuns/>; Retrieved February 12, 2019.
- [10] Minoru Kanehisa, Susumu Goto, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. Data, information, knowledge and principle: back to metabolism in kegg. Nucleic acids research, 42(D1):D199–D205, 2013.
- [11] Natalya F Noy, Deborah L McGuinness, et al. Ontology development 101: A guide to creating your first ontology, 2001.



B.Sc.

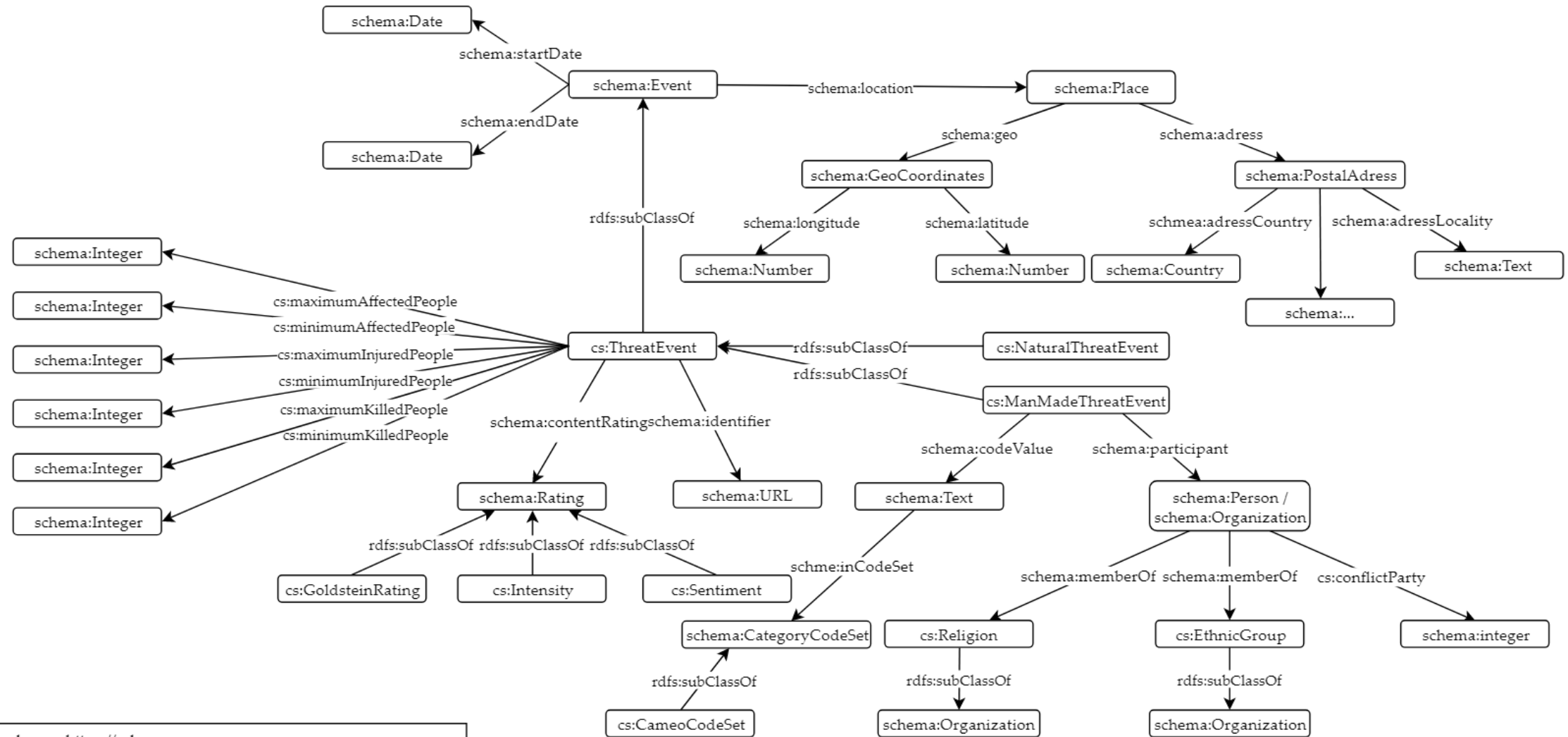
Malte Brockmann

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





```
schema: https://schema.org
rdfs:    https://www.w3.org/2001/sw/RDFCore/Schema/200203
cs:      Corporate Security Specific
```

Backup

Midas Architecture

