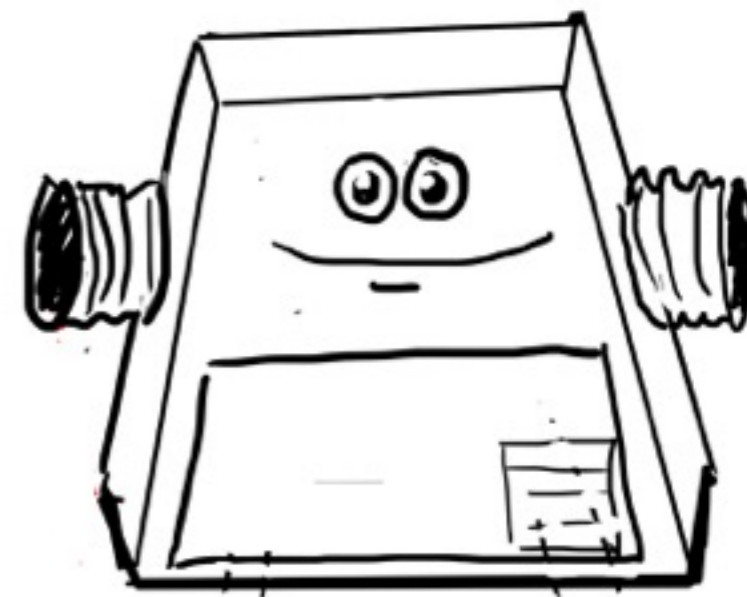


NewsReader

The reading machine



ICT 316404, FP7-ICT-2011-8
www.newsreader-project.eu



<https://youtu.be/rYLaVN3oqLI>



Agenda

- 9:30 - 10:30 General overview
- 10:30 - 10:40 coffee break
- 10:40 - 12:25 demos
- 12:25 - 13:10 lunch
- 13:10 - 14:25 benchmarking
- 14:25 - 14:55 end-users
- 14:55 - 15:05 coffee break
- 15:05 - 15:35 exploitation
- 15:35 - 15:50 management
- 15:50 - 16:20 Reviewers & PO
- 16:20 - 16:50 Comments & discussion
- 16:50 - 17:20 Innovation questionnaire

Overview

- Overall status of the project
- Review recommendations
- Looking forward: NewsReader II

Project objectives

- Scientific: cross-lingual event detection, event coreference, event background modelling, storyline structures, reasoning; innovative models and approaches with state-of-the-art results and beyond for
- Technical: efficient processing of massive streams, scalable cumulative storage of the result, efficient access to the result through visualisations of stories and graphs, decision support tool
- Industrial: end-user evaluation, innovative solutions and new products, 3rd party take-up

Scientific results

- Formal models that combine NLP and Semantic Web data: NAF, SEM, GAF, GRASP
 - Mentions, instances, perspectives
 - ISO WG 5. PWI 2462x Language resource management
- Crosslingual event detection pipelines (13-17 modules per language): English, Spanish, Italian and Dutch (Bulgarian and German being developed)
- Crosslingual resources: ESO as domain event model and PredicateMatrix for mapping ESO to vocabularies across languages
- Software for IDAP (cross-doc event coreference), timelines and storylines
- KnowledgeStore technology: storage, access, reasoning
- MEANTIME (crosslingual) and ECB+ (cross-document and cross-topic) benchmark corpora
- State-of-the-art and beyond state-of-the-art results for 4 languages
- Computational models for storyline and perspectives
- 11 major Journal papers with impact

IDAP method

- **I**dentification: mentions of events are similar if their components are similar —> mentions to instances
- **D**eduplication: similar information is represented only once
- **A**ggregation: complementary information is combined in a single representation
- **P**erspectivation: differences and different view points are traceable through their sources and mentions in text

Event-Centric Knowledge -Graphs (ECKGs)

- Entity centric knowledge graphs such as DBpedia, Google knowledge graph
 - Give static biographies for entities with main events and facts
 - Duplicate information across entities which may lead to inconsistencies:
 - wikipedia:Porsche buys back 10% stake from Qatar
 - wikipedia:Qatar holds 17% stake in Porsche (sales is not mentioned and fact is out of date)
 - Events are properties in triples:
 - subject:Qatar, predicate:sell, object:10%stake
- ECKGs
 - Capture dynamic changes —> knowledge at points in time
 - Events represented once as instance objects —> deduplicates, no inconsistencies
 - Events are subjects in triples:
 - subject:sell#24566, predicate:semActor, object:10%stake#764334.

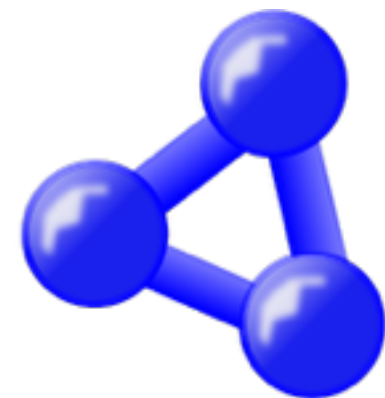
Technical results

- Reading machine: open source and as deployable VMs
- Large batch volumes (Hadoop) and streams (Storm)
- End-to-end system
- 2.3 million articles and 1.2 billion triples in the KnowledgeStore
- Efficient and stable access during hackathons
- Synerscope access to event-centric knowledge graphs as a decision support tool
- Storyteller and Perspective visualisations

Type	Mentions	Instances		
		DBpedia	nonDBpedia	Total
Event	420,010,878	0	42,296,287	42,296,287
Person	16,821,830	126,140	769,401	895,541
<u>Organisation</u>	23,841,719	44,458	1,094,712	1,139,170
Location	11,839,365	76,341	152,104	228,445
News triples	1,146,601,954			
DBpedia triples	94,172,990			
Total	1,240,774,944			

Mentions

Instances



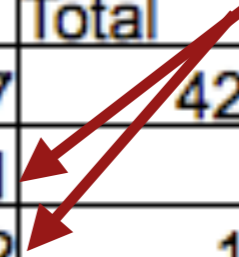
5% entities

All mentions

10% events

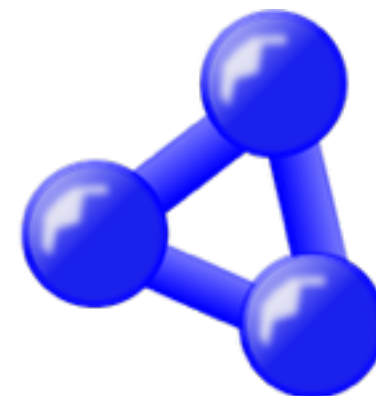
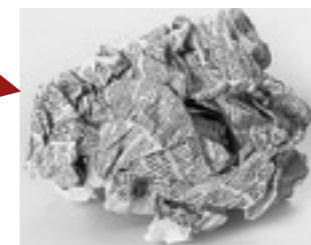
Type	Mentions	Instances		
		DBpedia	nonDBpedia	Total
Event	420,010,878	0	42,296,287	42,296,287
Person	16,821,830	126,140	769,401	895,541
<u>Organisation</u>	23,841,719	44,458	1,094,712	1,139,170
Location	11,839,365	76,341	152,104	228,445
News triples	1,146,601,954			
DBpedia triples	94,172,990			
Total	1,240,774,944			

Dark entities



Mentions

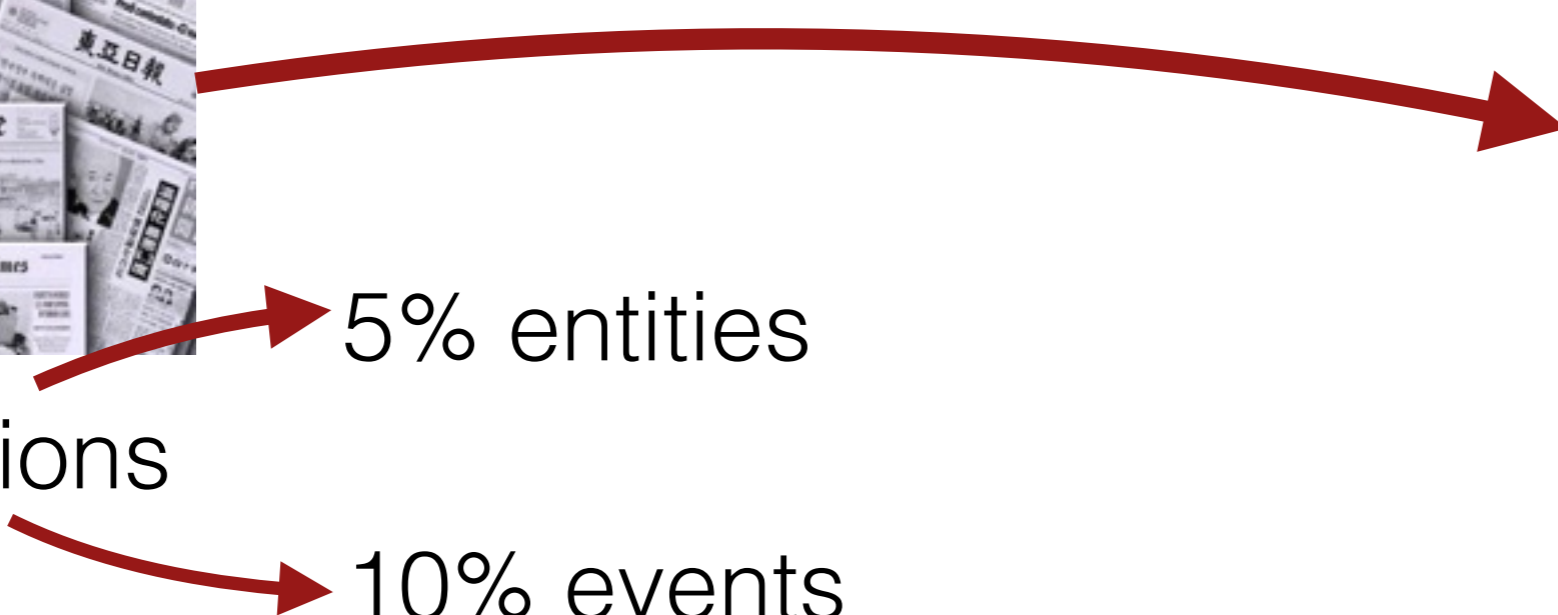
Instances



5% entities

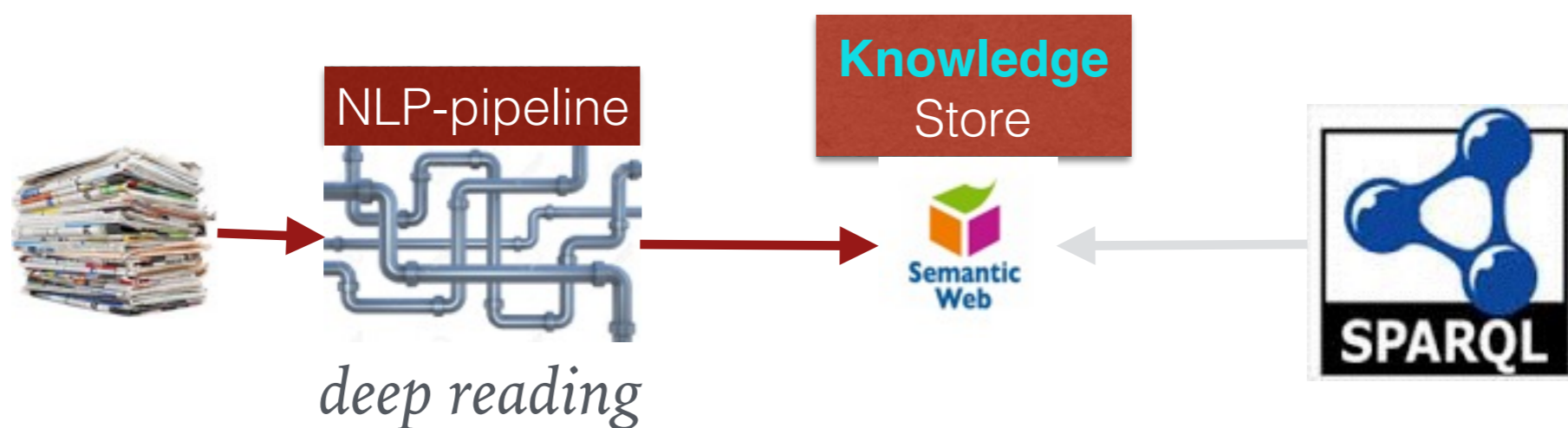
All mentions

10% events



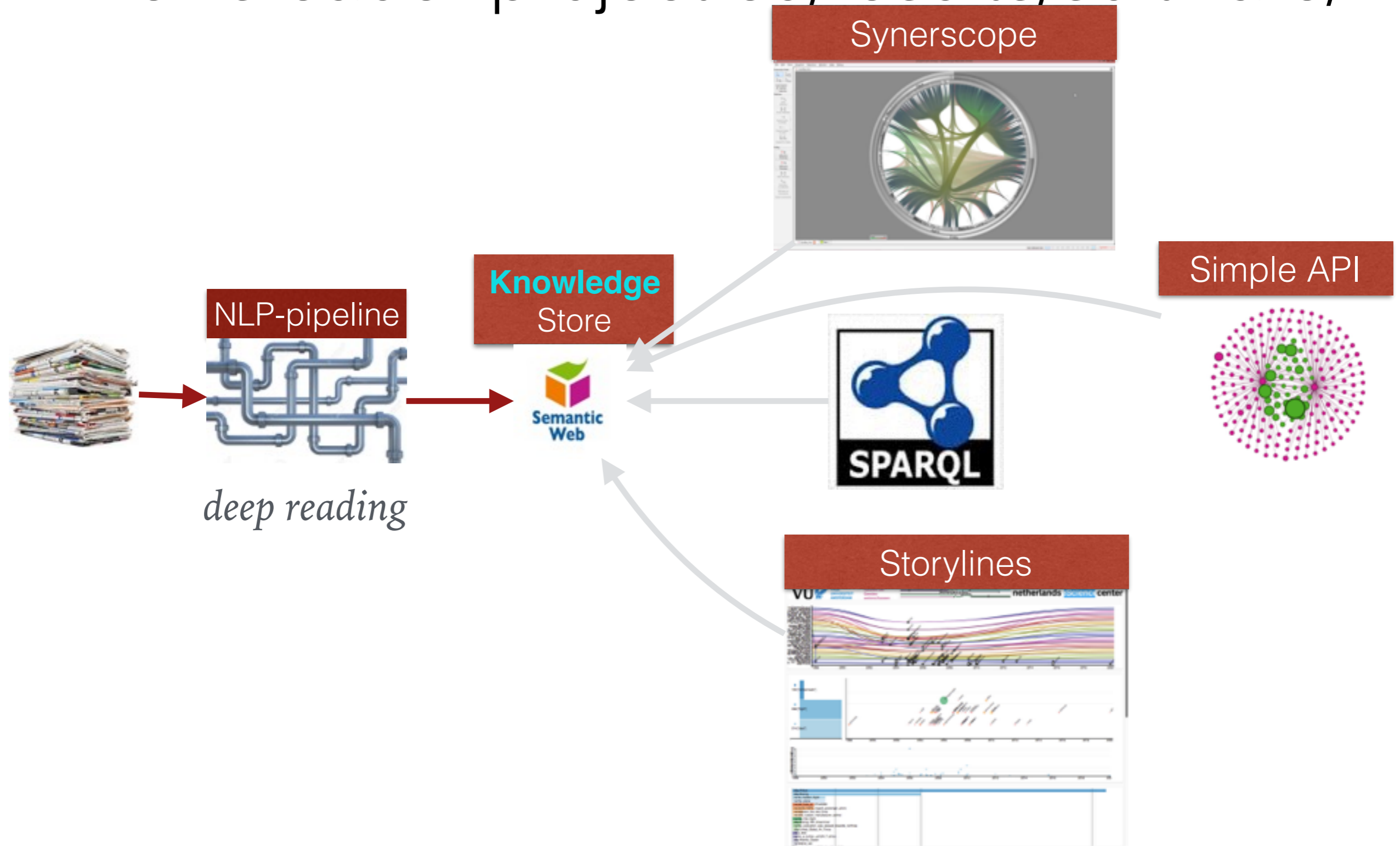
End-to-End system

newsreader-project.eu/results/software/



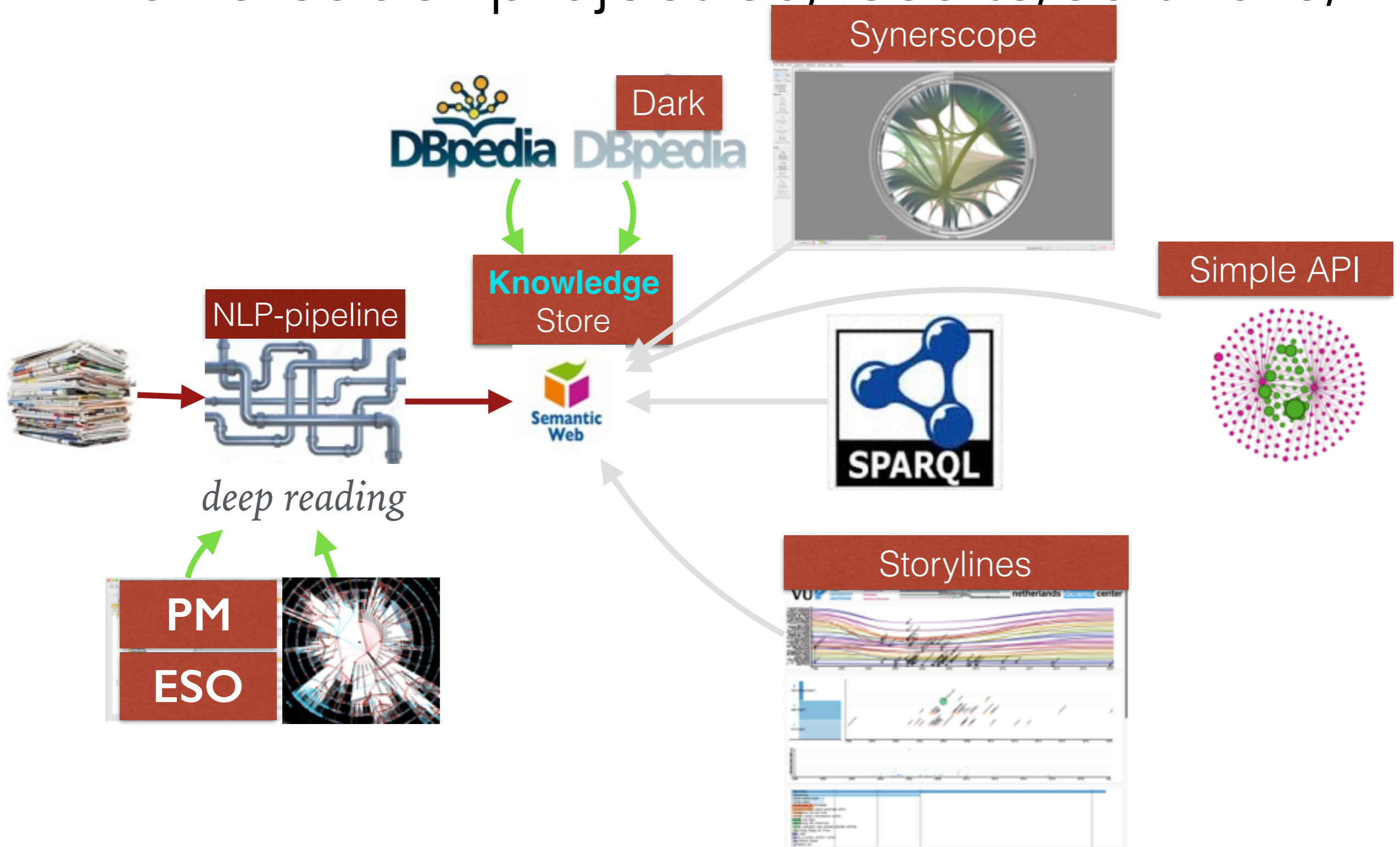
End-to-End system

newsreader-project.eu/results/software/



End-to-End system

newsreader-project.eu/results/software/



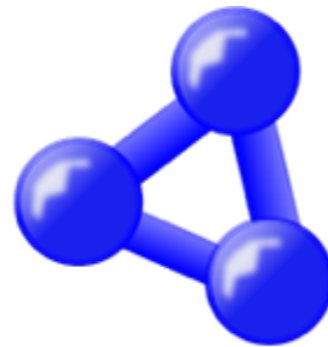
Industrial results

- Evaluation sessions with end-users and positive results
- Event centric knowledge graphs as a new paradigm
- 3rd party take-up during project: Dutch House of Parliament, Spanish Ministry, Trafficking project, Dasym investment company, Dutch National Bank, Financial Times, Commerz Bank, Dutch broadcasting (KRO/NCRV)
- Almost 40 take up projects, 18 with non-academic partners

User scenario

SEMANTIC QUERIES

Events



Entities



eso:FinancialTransaction
34,664

eso:IntentionalEvent
2,927,973

nwr:Organisation
1,993,411

nwr:Person
1,526,551

buying stake

sue

dbp:VW

dbp:Wiedeking

dbp:Porsche

dbp:Piëch

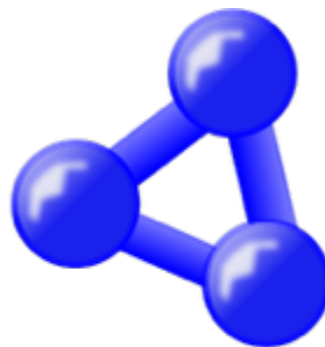


1995 96 97 98 99 2000 01 02 03 04 2005 06 07 08 09 2010 11

Events



Entities



eso:FinancialTransaction
34,664

eso:IntentionalEvent
2,927,973

nwr:Organisation
1,993,411

nwr:Person
1,526,551

eso:Selling
1,000,692

eso:Buying
596,517

fn:Court_Examination
34,882

dbp:Bank
897

dbo:KeyPerson
29,576

dbo:CEO
746

Automotive_industry
983

buying stake

sue

dbp:VW

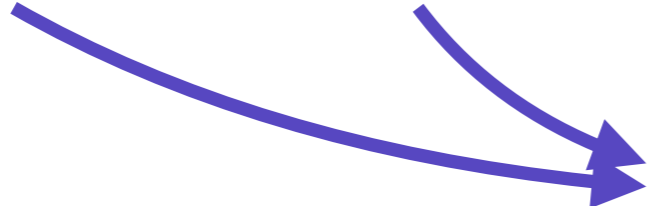
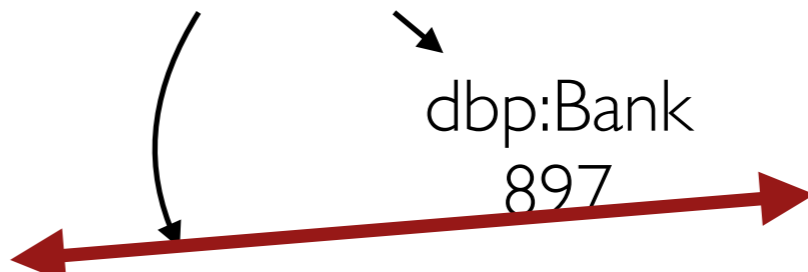
dbp:Wiedeking

dbp:Porsche

dbp: Piëch



1995 96 97 98 99 2000 01 02 03 04 2005 06 07 08 09 2010 11



Finding 'hidden events'

- Motivation for looking for an event: something happened with impact that you care about
- People are involved, responsible or have some control over the situation
- What are the events in the past that can be connected to this events but remained unnoticed: so-called 'hidden events'?

Triple haystack method

“Finding hidden events”



1. Choose an entity or class of entities
2. Ask for Events with significant impact
3. What did you miss from the past that helps explaining the impact event?

Triggers for event impact

- Volume of news, searches, tweets (per topic)
- Strong sentiment e.g. in Social Media
- Unexpected or critical trends in structured data (graphs)

Main Menu -

- Top Stories
- 24 Hours Overview
- Events Detection
- Most Active Themes
- Help about EMM
- Overview
- Advanced Search
- Sources list
- Web Site Map

EU Focus +

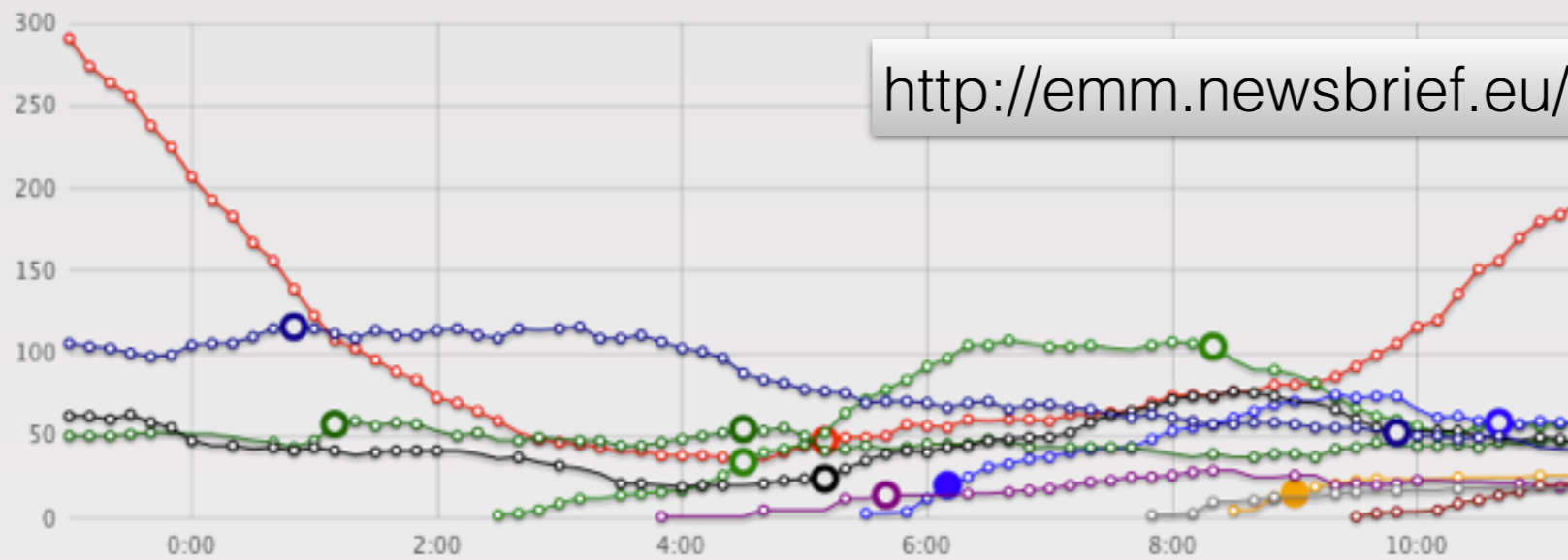
EU Policy Areas +

Themes +

The World +

Offices & Agencies +

Current top 10 stories
Language: en Period: Feb 17, 2016 11:20 PM - Feb 18, 2016 11:20 AM



Turkey: Syrian man behind deadly Ankara car bomb attack
Articles : 708 | Last update : Feb 18, 2016 11:24:00 AM | Start : Feb 17, 2016 5:55:00 PM | Sources : 225 | Peak : 1 | Current rank : 1

Ankara bomb: Syrian linked to PKK posed as refugee to carry out suicide attack in Turkey
ibtimes-uk Thursday, February 18, 2016 11:24:00 AM CET | info [other]

Turkish army service buses burn after an explosion in Ankara, Turkey Defne Karadeniz/ Getty Images. A Syrian man posing as a refugee, identified as Saleh Nejar with links to the militant Kurdistan Worker's Party (PKK), has been confirmed as the suicide bomber behind a massive explosion in the.....

[More articles...](#)

Obama plans historic trip to Cuba to further ties
Articles : 174 | Last update : Feb 18, 2016 11:23:00 AM | Start : Feb 18, 2016 2:33:00 AM | Sources : 118 | Peak : 1 | Current rank : 2

Obama to visit Cuba in 'coming weeks'
manilatimes Thursday, February 18, 2016 11:13:00 AM CET | info [other]

WASHINGTON, D.C.: President Barack Obama will make a historic visit to Cuba "in coming weeks," a US official said Wednesday, a powerful sign of the thaw in relations between the former Cold War foes. "We can confirm that tomorrow the administration will announce the president's travel to Latin.....

[More articles...](#)

Seoul's spy service says North Korea is preparing attacks North Korean leader Kim Jong Un has ordered preparations for launching attacks on South Korea, Seoul's spy agency disclosed, as worries about the North...
Articles : 100 | Last update : Feb 18, 2016 11:21:00 AM | Start : Feb 18, 2016 5:33:00 AM | Sources : 69 | Peak : 3 | Current rank : 3

South: North Korea planning terror attack
cnn Thursday, February 18, 2016 11:21:00 AM CET | info [other]

Tools

Thursday, February 18, 2016
11:43:00 AM CET

[RSS](#) | [KML](#) | [MAP](#)

[Facebook](#)

[subscribe](#) | [manage](#)

[info](#)



Languages -

Select top stories in other languages.

- | | | | | | |
|----|----|----|----|----|----|
| ar | bg | cs | da | de | el |
| en | es | et | fi | fr | hr |
| hu | it | lt | lv | mt | nl |
| pl | pt | ro | ru | sk | sl |
| sv | sw | tr | zh | | |

Show additional languages

Interface: en - English

Legend +

Country Watch

The country most in the news at the moment.



United States of America (US)

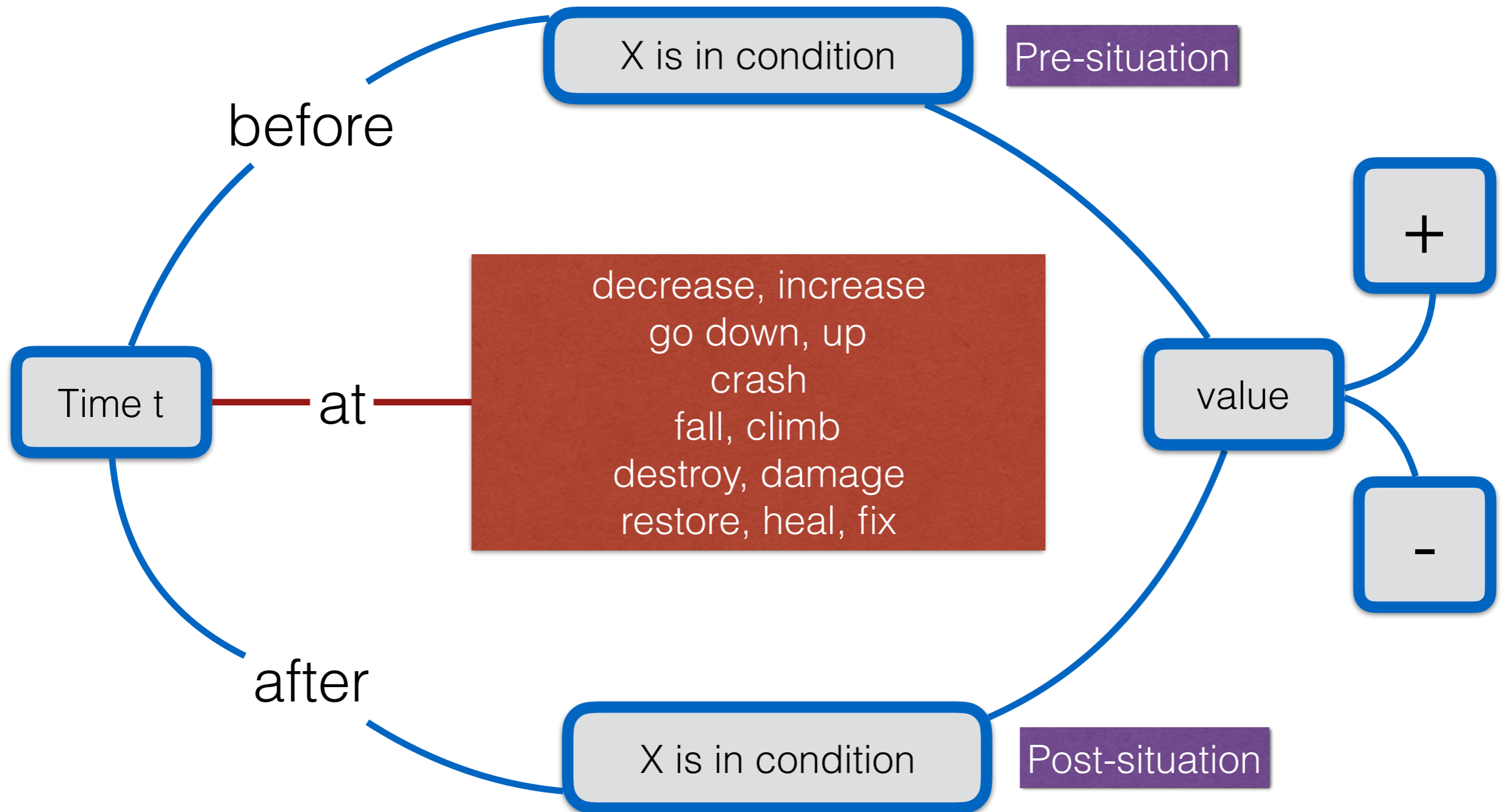
[News about United States of America](#)

[News sources from United States of America](#)

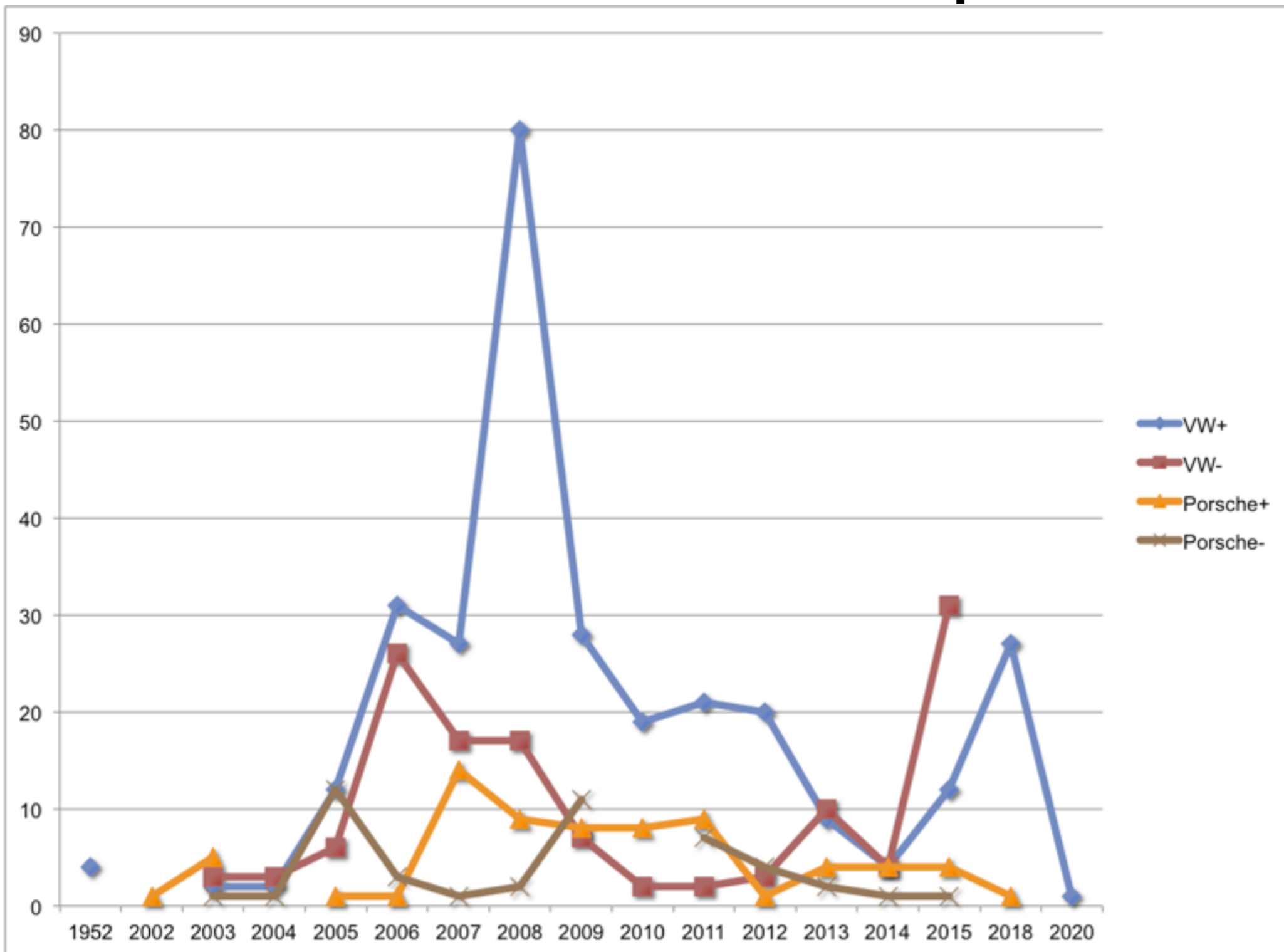


Triggers for impact of the
event on the participants

Event & Situation Ontology (ESO) implications

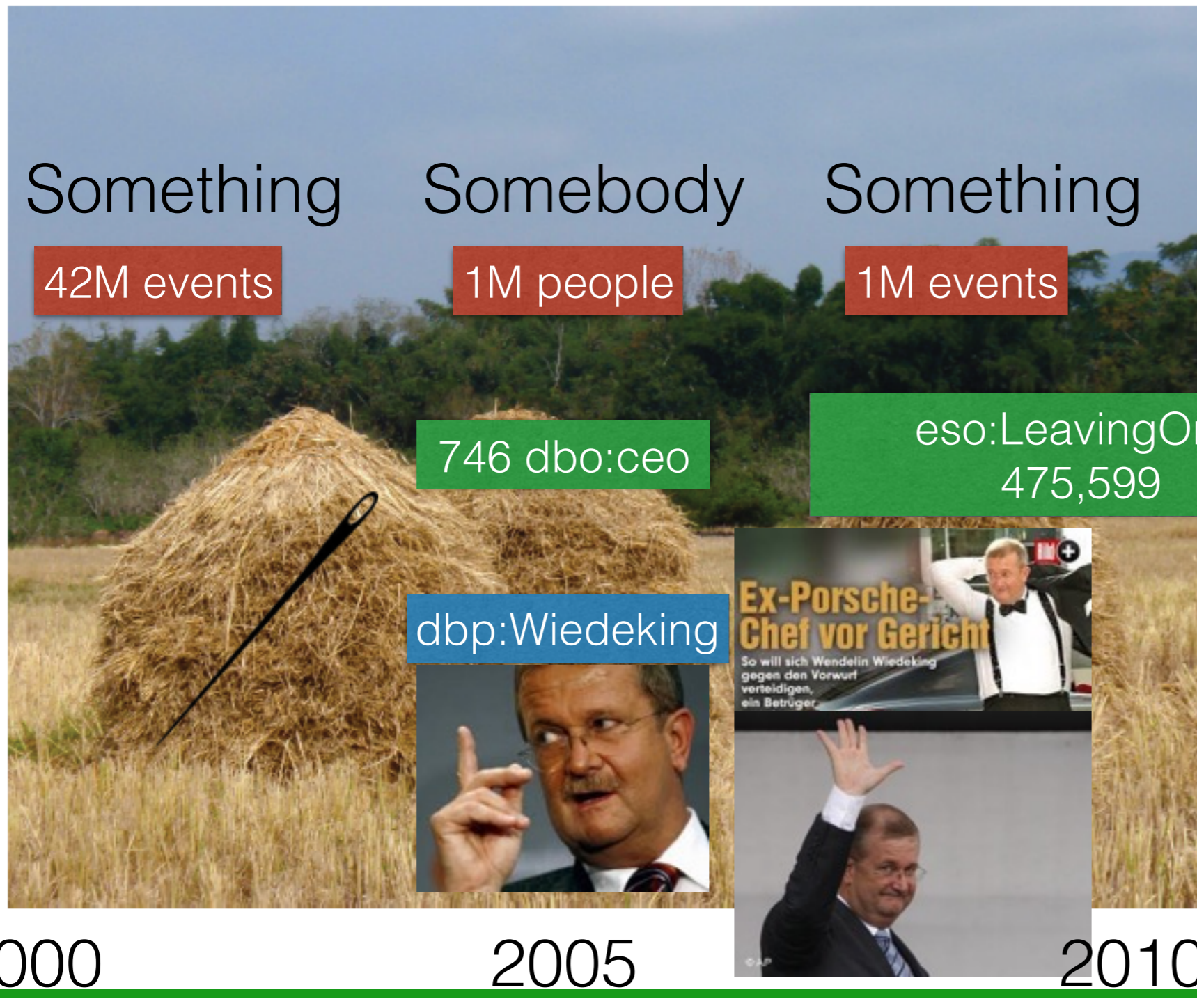


ESO impact



- Nr. *negative* or *positive* **impact** events for VW and Porsche

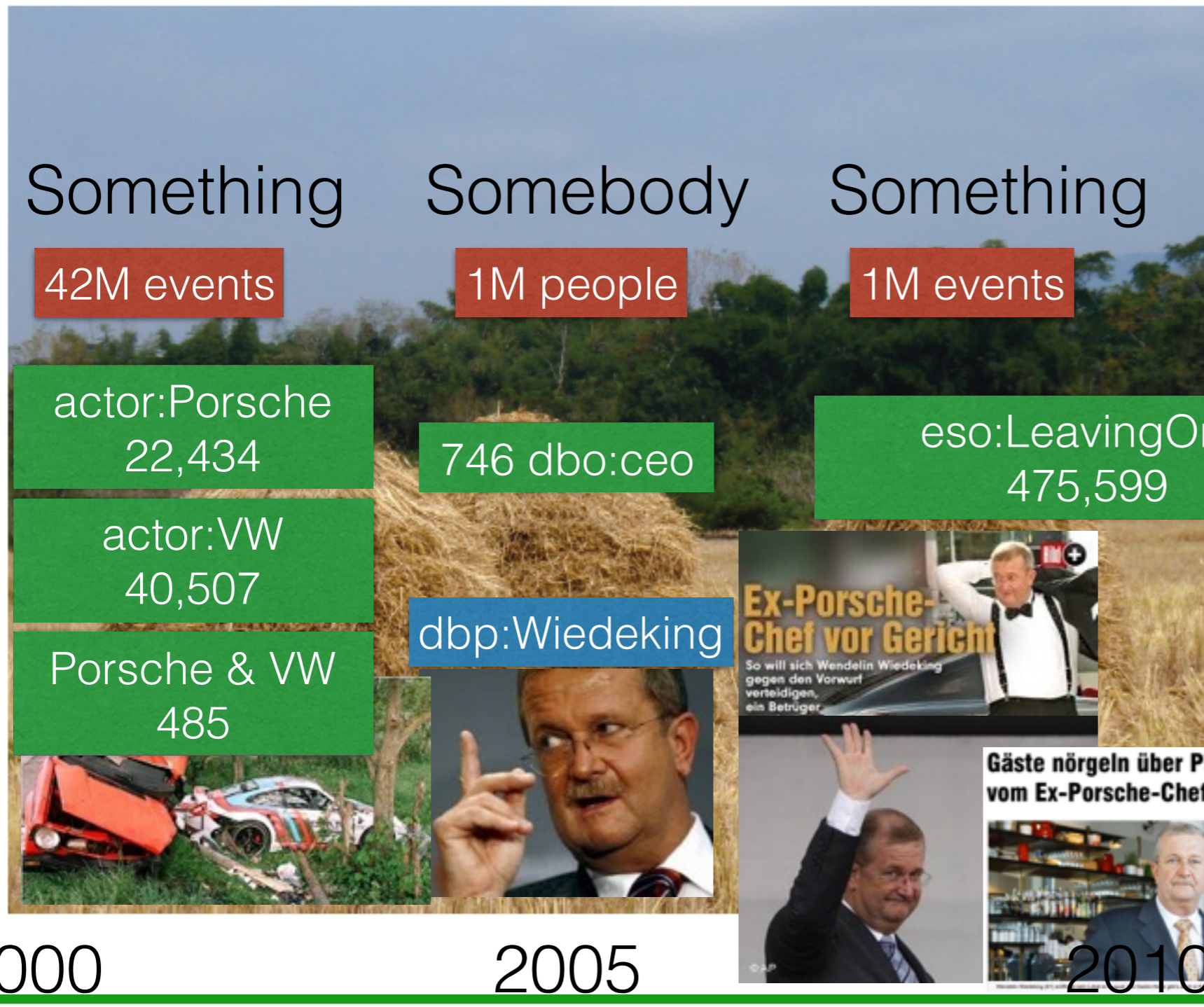
Triple haystack method “Finding hidden events”



1. Choose an entity or class of entities
2. Ask for Events with significant impact
3. What did you miss from the past that helps explaining the impact event?

Triple haystack method

“Finding hidden events”



1. Choose an entity or class of entities

2. Ask for Events with significant impact

3. What did you miss from the past that helps explaining the impact event?

Car wars

Ferdinand Piëch

Wendeling Wiedeking



and I want to run

I run

and I want to run

I run



Volkswagen



PORSCHE



Triple haystack method “Finding hidden events”



Something

42M events

Somebody

1M people

Something

1M events



2000



2005



Parliamentary hearing

1. Choose an entity or class of entities
2. Ask for Events with significant impact
3. What did you miss from the past that helps explaining the impact event?

Triple haystack hidden event results

- Porsche Wiedeking [H1] sued by shareholders [H2] has led the takeover quest on VW [H3]
- Fortis CEO Maurice Lippens [H1] refused to appear for the Dutch Parliamentary Enquiry [H2], convinced Sulyeman Kerimov to invest 100M euros [H3]
- Peugeot [H1] stopped production in Iran [H2], after American alliance (GM) and pressure to boycott [H3]

NewsReader intelligence

- ☑ association (needles in haystacks)
- ☑ correlation (only with help of data scientist)
- ☐ no legal liability
- ☐ no scientific proof
- ☑ help to find a story
- ☑ tell a story

Response to review year 2

- Recommendation 1: Assess whether a project/service exploitation mode is considered relevant and, if yes, how it could be implemented.
- Exploitation plan describes the measures we took to continue exploitation as project/service
- Currently done: Dasym; and planned: FT, Commerz Bank, Spanish Ministry, etc...
- Details presented later today

Response to review year 2

- Recommendation 2: Assess the questions raised before about the quality of the system (the quality of the semantic tagging): Is the quality enough for commercial usage? How much effort is required to adapt the NLP pipelines to a new domain?
 - benchmark data and results for 4 languages
 - high-level tasks: event coreference, timelines, triple evaluation
 - domain adaptation
 - Details presented later today

Response to review year 2

- Recommendation 3: Synchronize work required for WP8 and final review meeting date, in a way to provide valuable feedback and achieved deliverable on "Evaluation for Decision Making" for the date of review.
- Details presented later today

Response to review year 2

- Suggestion 1: Build a demo scenario, available through a video, demonstrating the power of NewsReader system for discovering "hidden" information.
 - Videos and triple haystack method
- Suggestion 2: Continue to highlight innovations, strengths (and possible weaknesses) of individual specific technical components.
 - 120 peer reviewed publications, 166 presentations, interviews & talks for specialist and broader public

NewsReader II

- Beyond the current exploitation: more research!
- More complex data structures: event coreference, timelines, storylines, perspectives, IDAP
- Beyond the sentence, beyond the document, beyond the language, beyond text!
- Multimodal knowledge-to-data processing: FET proposal DECIPHER
 - Beyond the reading machine: machines that think, perceive and read
 - Joint inferencing and deep-learning approaches to data maintaining alternatives