

The NewsReader Vision

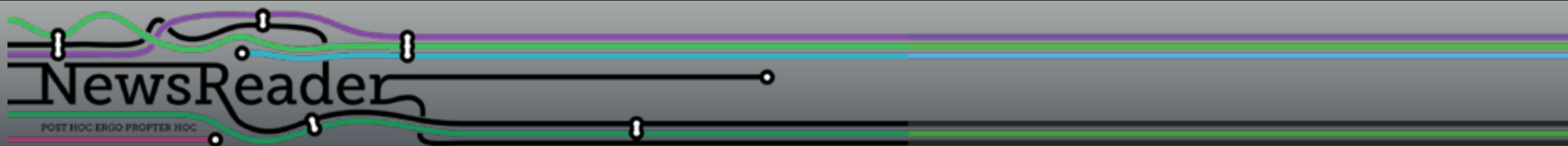
EUROPEAN DATA FORUM,
NOV-2015, LUXEMBOURG

Piek Vossen

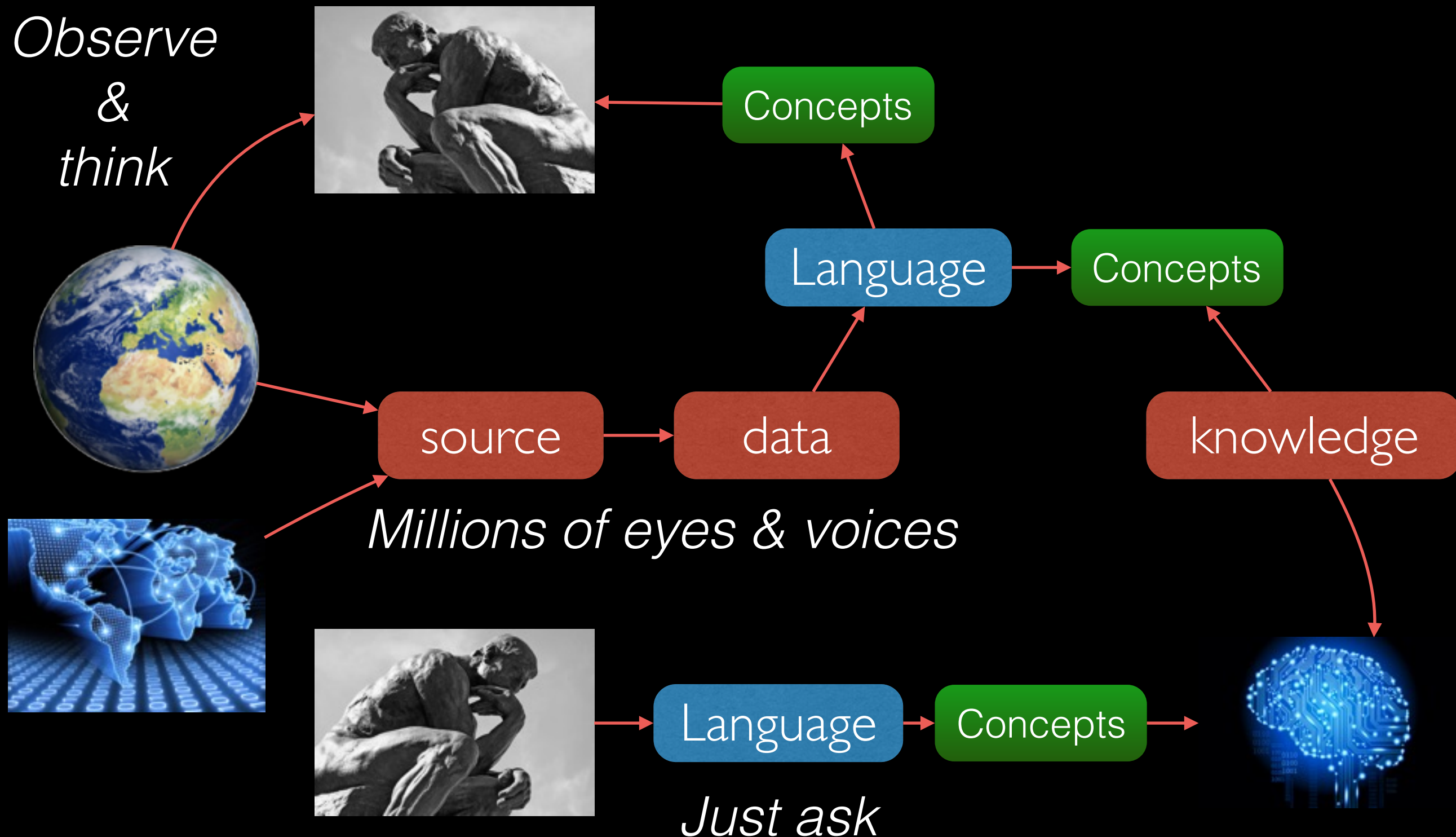
ICT 316404

FP7-ICT-2011-8

www.newsreader-project.eu



Data in a changing world



To learn about the changing world ...we read the news...

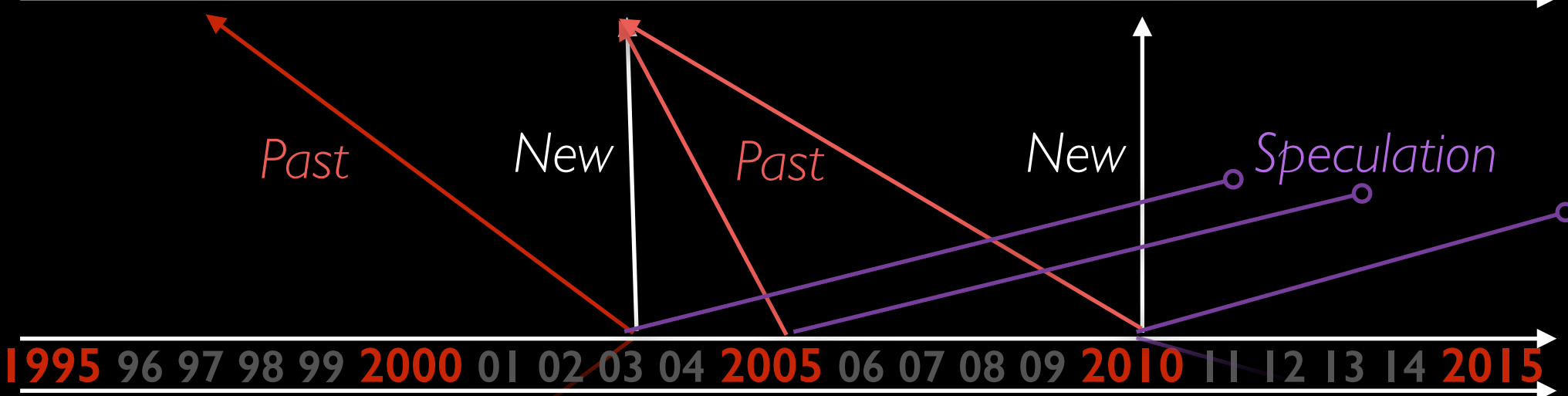
- But can we handle the news?
- Information broker LexisNexis:
 - 1.5 million news articles on a single working day
 - 30,000 different sources



VOLUME OF CHANGE



1995 96 97 98 99 2000 01 02 03 04 2005 06 07 08 09 2010 11 12 13 14 2015



What if
computers could
read the news?



NewsReader (ict316404)

- ***Reading Machine*** processes massive streams of news from many different sources in 4 languages (English, Dutch, Spanish and Italian):
 - Recording the changes in the world as they are told in the media over long periods of time → ***history-recorder***.
 - ***What*** happened, ***where*** and ***when***, ***who*** was involved.
 - From Text to RDF.
 - Who made what statement, where do sources agree and disagree: ***provenance***!

<http://english.alarabiya.net>

<http://www.telegraph.co.uk>

CERTAIN;TRUE;NON-FUTURE

Qatar Holding sells 10% stake in Porsche to founding families

Porsche family buys back 10pc stake from Qatar

fn:Commerce_money_transfer

type

dbp:Porsche

fn:Buyer

**Event₁₂
buy/sell**

fn:Seller

dbp:QatarHolding

fn:Goods

sem:hasTime

fn:Money

**Entity₂₃
10% stake**

2013-06-17

?

ON A WORKING DAY

Friday
10:15
February 6

Tuesday
09:34
February 10

1 MILLION ARTICLES
FROM 30,000 SOURCES

1 ARTICLE

100 EVENT MENTIONS PER ARTICLE

5 EVENT INSTANCE PER ARTICLE



PLACES

Knowledge
Store

EVENTS



ENTITIES

1995 96 97 98 99 2000 01 02 03 04 2005 06 07 08 09 2010 11 12 13 14 2015

Table 8: Mentions to instance reduction for 2.2M English news articles on the automotive industry from 2003 till 2015

Types	Mentions	DBpedia	nonDBpedia	Total
Events	420,010,878	-	90,866,364	90,866,364
Person	16,821,830	169,852	1,356,699	1,526,551
Organisation	23,841,719	54,236	1,939,175	1,993,411
Location	11,839,365	93,985	265,342	359,327



All mentions

20% events

30% changes in the world

10% entities

70% talks about the world

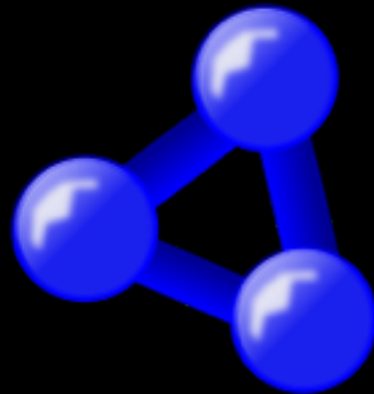
Triple haystack problem



- Entity or class of entities
- Significant negative implications
- Biography of what happened before?

SEMANTIC QUERIES

Events



Entities



eso:FinancialTransaction
61,405

eso:IntentionalEvent
326,524

dbp:Company
217,456

dbp:Person
111,375

eso:Buying
5,568

eso:LegalEvent
19,772

dbp:Bank
300

dbp:CEO
527

investment

hearing

dbp:Fortis

dbp:Sberbank

dbp:Lippens

dbp:Kerimov



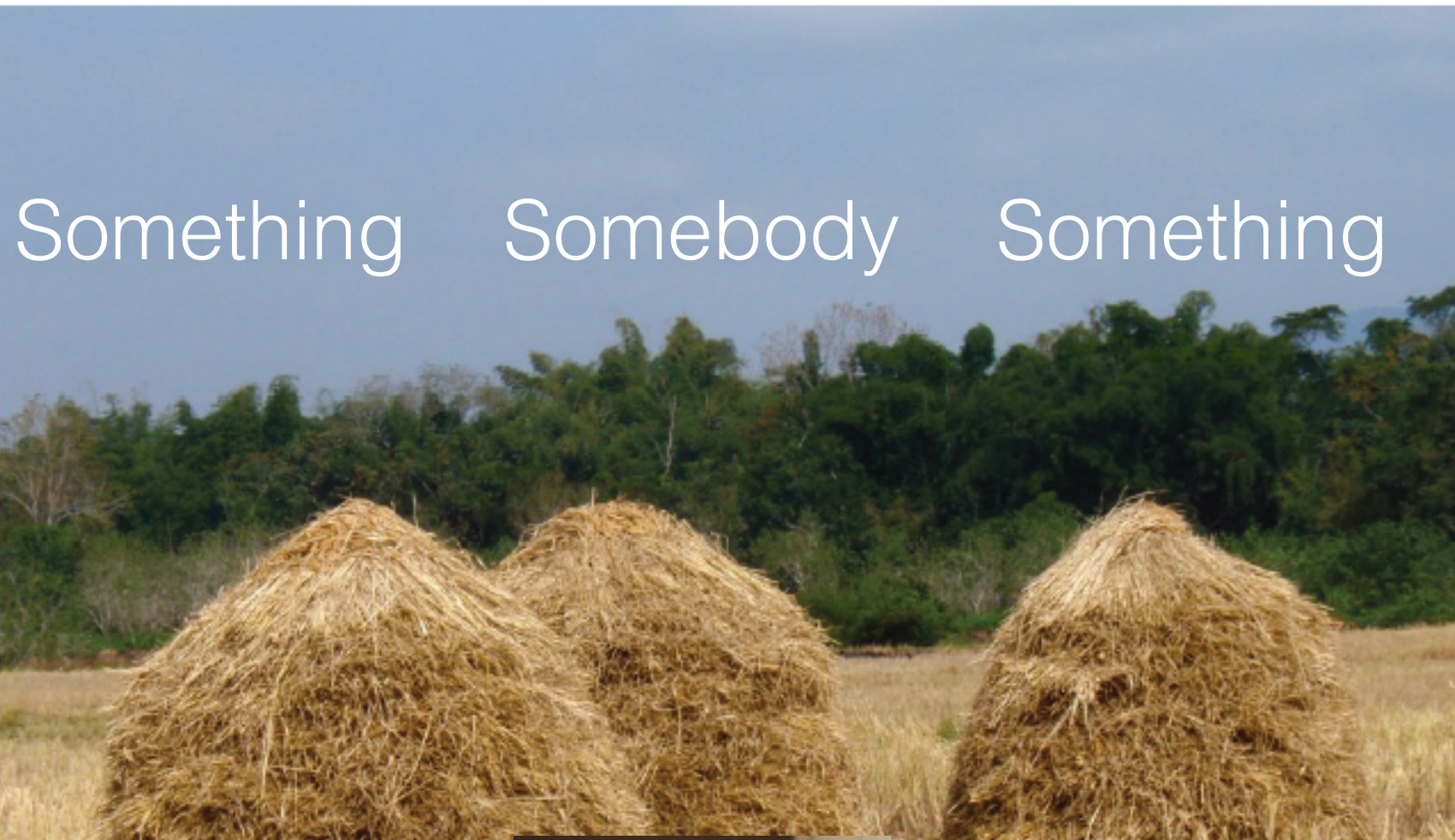
1995 96 97 98 99 2000

01 02 03

2005 06 07 08 09 2010 11



Triple haystack problem



- Entity or class of entities
- Significant negative implications
- Biography of what happened before?



Sulyman Kerimov

Maurice Lippens

Parliamentary hearing

Triple haystack results

- Fortis CEO Maurice Lippens [H1] refused to appear for the Dutch Parliamentary Enquiry [H2], convinced Sulyeman Kerimov to invest 100M euros [H3]
- Porsche Wiedeking [H1] sued by shareholders [H2] has led the takeover quest on VW [H3]
- Renault [H1] stopped production in Iran [H2], after American alliance and pressure to boycott [H3]
- <http://www.newsreader-project.eu/results/demos/synerscope/>



newsreader-project.eu

Come hack with NewsReader!

25 November 2015
Amsterdam Public Library

At the hackathon you can get your hands dirty with the NewsReader system. We will make available the semantic database that the NewsReader system extracted from 10 million news articles about the global automotive industry 2003 - 2015.

You decide what applications you build on top of this database.

Not a developer but full of ideas on what to do with this database? No worries, there will be NewsReader team members to guide you through and developers to team up with.

More info and registration at:



newsreader-project.eu

NewsReader Final Workshop

24 November 2015
Amsterdam Public Library

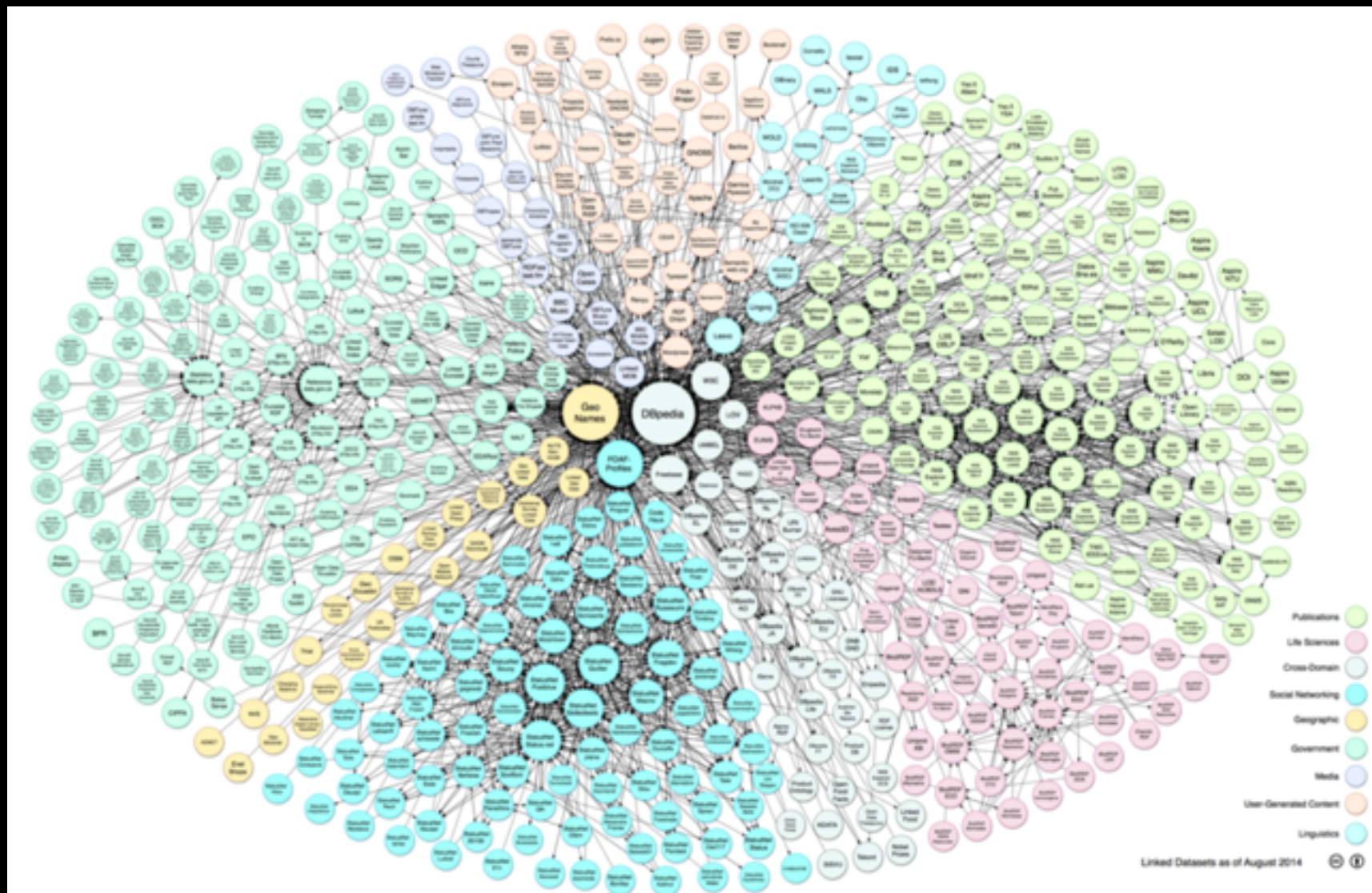
The workshop will bring together start-ups, established companies, and researchers and developers to present and discuss the Newsreader project, the technological domains it draws from and future applications of these technologies.

The afternoon will feature invited talks, demos, a panel discussion and a networking reception.

Confirmed speakers:
Prof. dr. Frank van Harmelen (Vrije Universiteit Amsterdam)
Bernardo Magnini (FBK, Trento)

More info and registration at:





World
of
Data



Real
World

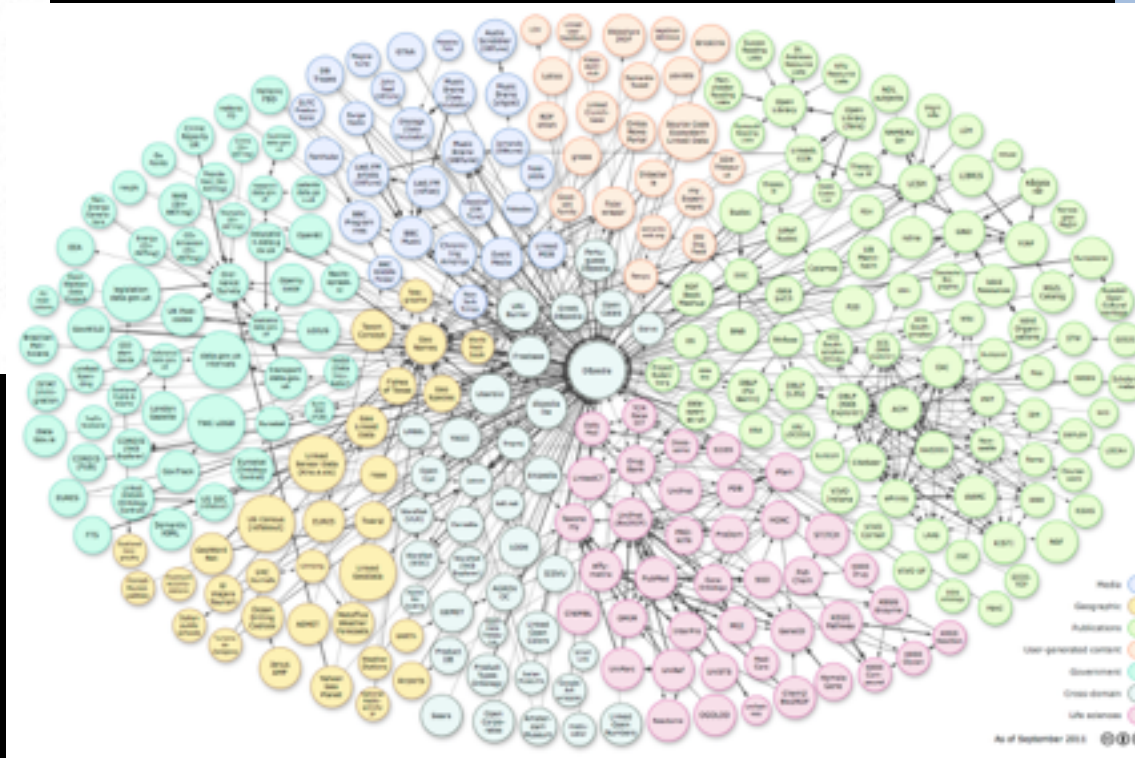
Social web

Web of things

Classical web



Semantic web



cosm.com: last 1,000 reporting sensors





Demonstration event



Protests in Cairo's Tahrir Square were brought by citizen journalism. Photograph: Mohammed Abed/ AFP/Getty Images



LOD helps NLP



NLP helps LOD