

Agenda

09:00 - 09:45 Project overview and overall status 2nd progress

09:45 - 12:15 Presentations (2.5 hours)

- WP02 System architecture (30 min) & WP03 Benchmarking (20 min)

10:45 - 11:00 BREAK

- WP04 Event detection (35 min) & WP05 Event modeling (35 min)

12:15 - 13:00 LUNCH

13:00 - 15:30 Presentations (2.5 hours)

- WP06 Knowledge store (30 min) & WP07 Decision Support Suite (30 min)

- WP08 User evaluation results (30 min) & WP09 Exploitation (25 min)

15:30 - 15:45 BREAK

15:45 – 16:00 Management, usage of resources, dissemination (15 min)

16:00 - 16:30 PO & Reviewers meet

16:30 - 17:00 Preliminary comments and discussion

17:00 - 17:30 Innovation questionnaire

Overview

- Objectives and Implementation
- Achievements 2nd year
- Response to recommendations
- Prospects 3rd year

ON A WORKING DAY

Friday
10:15
February 6

Tuesday
09:34
February 10

1 MILLION ARTICLES
FROM 30,000 SOURCES



1 ARTICLE



text

text

text

100 EVENT MENTIONS PER ARTICLE

naf

naf

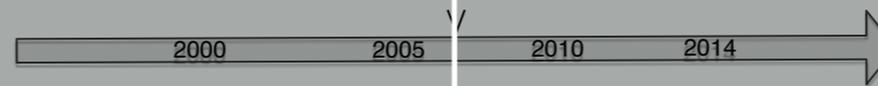
naf

5 EVENT INSTANCE PER ARTICLE

sem

sem

Knowledge Store



PLACES

EVENTS

ENTITIES



<http://www.telegraph.co.uk/finance/newsbysector/industry/engineering/10125280/Porsche-family-buys-back-10pc-stake-from-Qatar.html>

17 Jun 2013

Porsche family buys back 10pc stake from Qatar

Descendants of the German car pioneer Ferdinand Porsche have bought back a 10pc stake in the company that bears the family name from Qatar Holding, the investment arm of the Gulf State's sovereign wealth fund.

<http://www.telegraph.co.uk/finance/newsbysector/industry/engineering/10125280/Porsche-family-buys-back-10pc-stake-from-Qatar.html>

17 Jun 2013

Porsche family buys back 10pc stake from Qatar

Descendants of the German car pioneer Ferdinand Porsche have bought back a 10pc stake in the company that bears the family name from Qatar Holding, the investment arm of the Gulf State's sovereign wealth fund.

MENTIONS

INSTANCES

WHO:

http://dbpedia.org/page/Category:Porsche_family

<http://dbpedia.org/page/Category:Qatarholding>

<http://dbpedia.org/page/Category:PorscheSE>

<http://www.telegraph.co.uk/finance/newsbysector/industry/engineering/10125280/Porsche-family-buys-back-10pc-stake-from-Qatar.html>

17 Jun 2013

Porsche family **buys back** 10pc stake from Qatar

Descendants of the German car pioneer Ferdinand Porsche have **bought back** a 10pc stake in the company that bears the family name from Qatar Holding, the investment arm of the Gulf State's sovereign wealth fund.

MENTIONS

INSTANCES

WHAT:
buy_back

<http://english.alarabiya.net/en/business/banking-and-finance/2013/06/17/Qatar-Holding-sells-10-stake-in-Porsche-to-family-shareholders.html>

Monday, 17 June 2013

Qatar Holding sells 10% stake in Porsche to founding families

Qatar Holding, the investment arm of the Gulf state's sovereign wealth fund, has sold its 10 percent stake in Porsche SE to the luxury carmaker's family shareholders, four years after it first invested in the firm.

<http://europe.autonews.com/>

prov-attributedTo

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

prov-o:attributedTo

The Porsche-Piech family bought 10 percent stake in Porsche Automobil Holding

Porsche family buys back 10pc stake from Qatar

fn:Commerce_money_transfer

type

dbp:Porsche

fn:Buyer

Event buy

fn:Seller

dbp:QatarHolding

fn:Goods

sem:hasTime

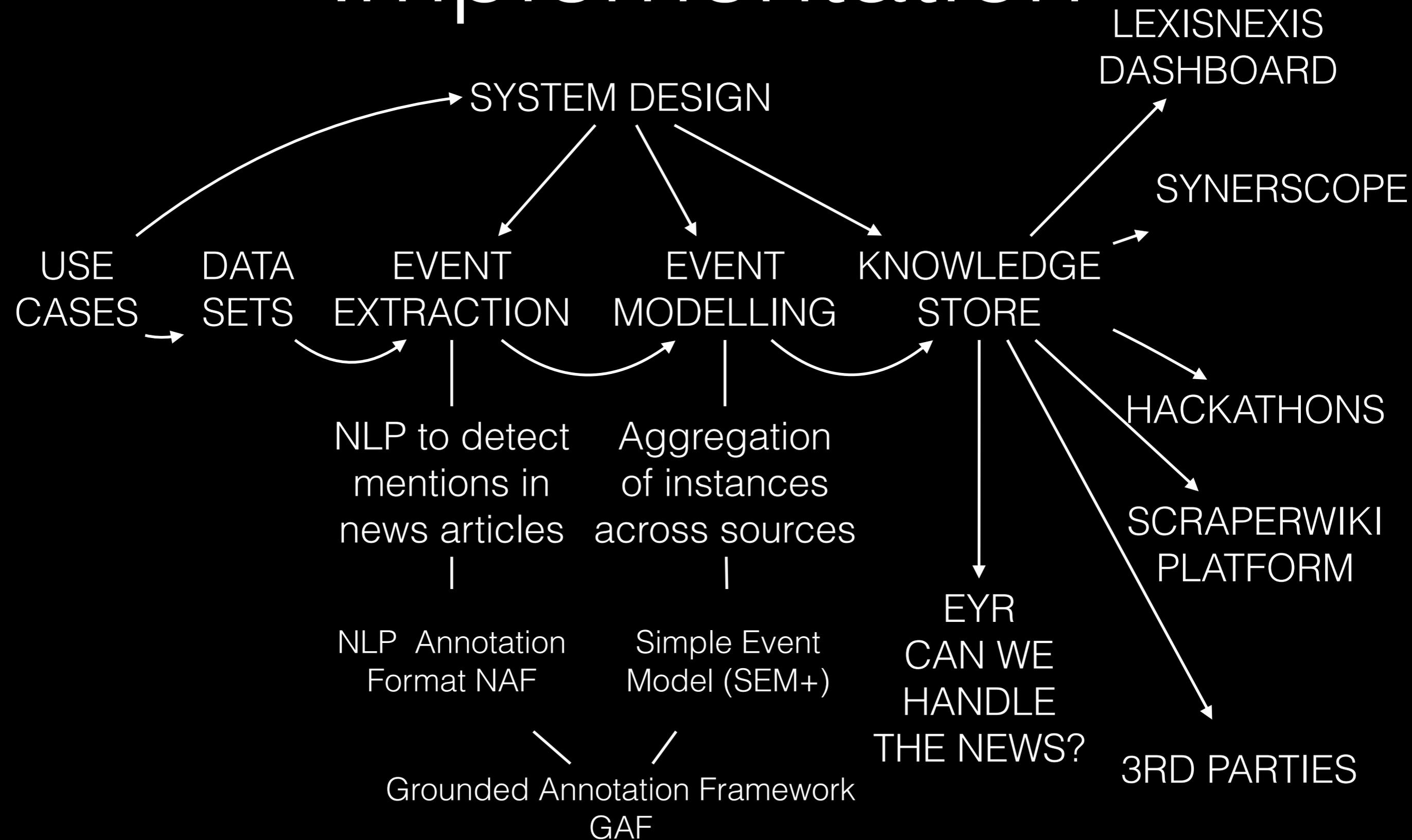
fn:Money

Entity
10% stake in
PorscheAutomobilHolding

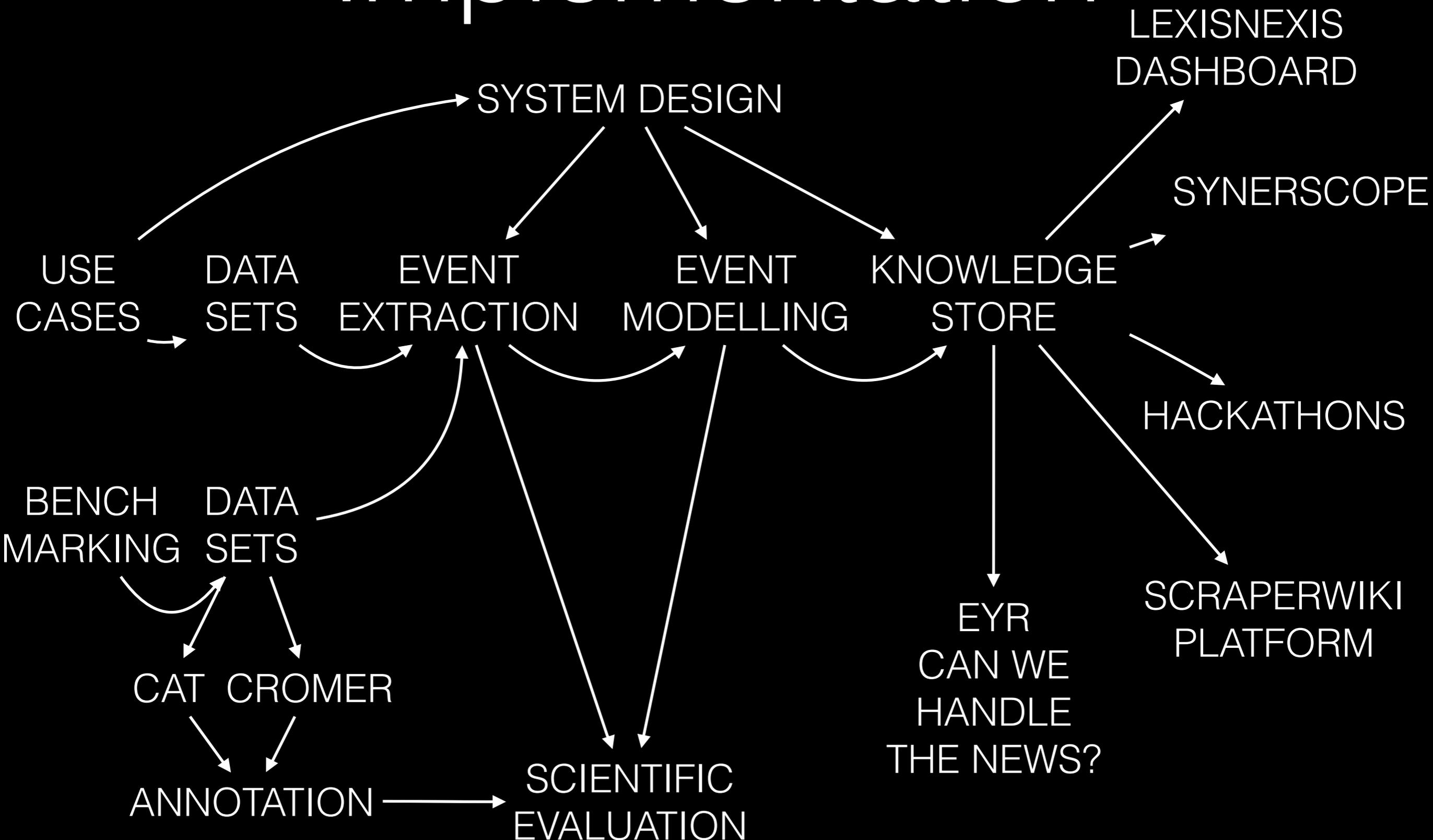
Monday,
17 June 2013

?

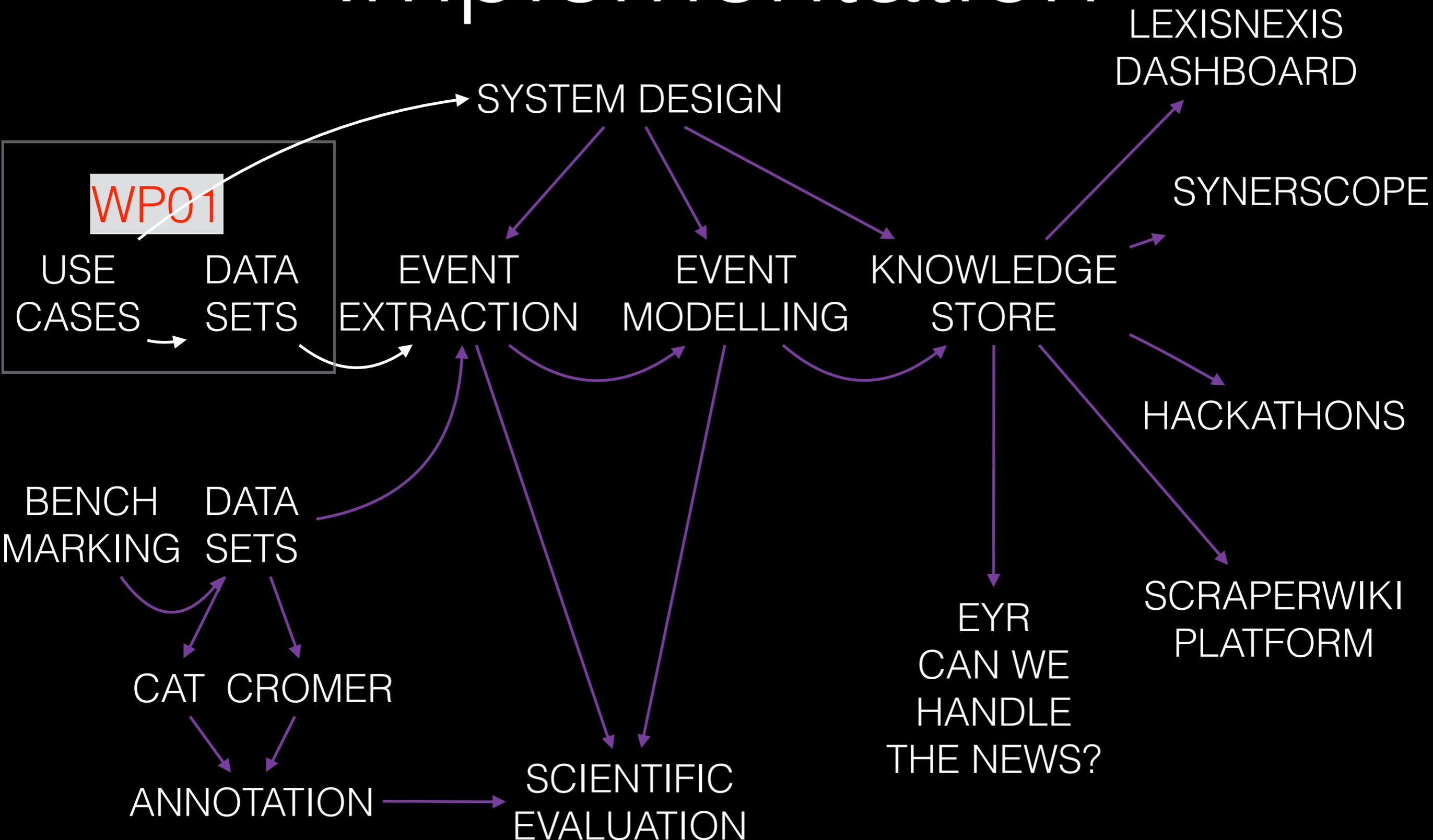
Implementation



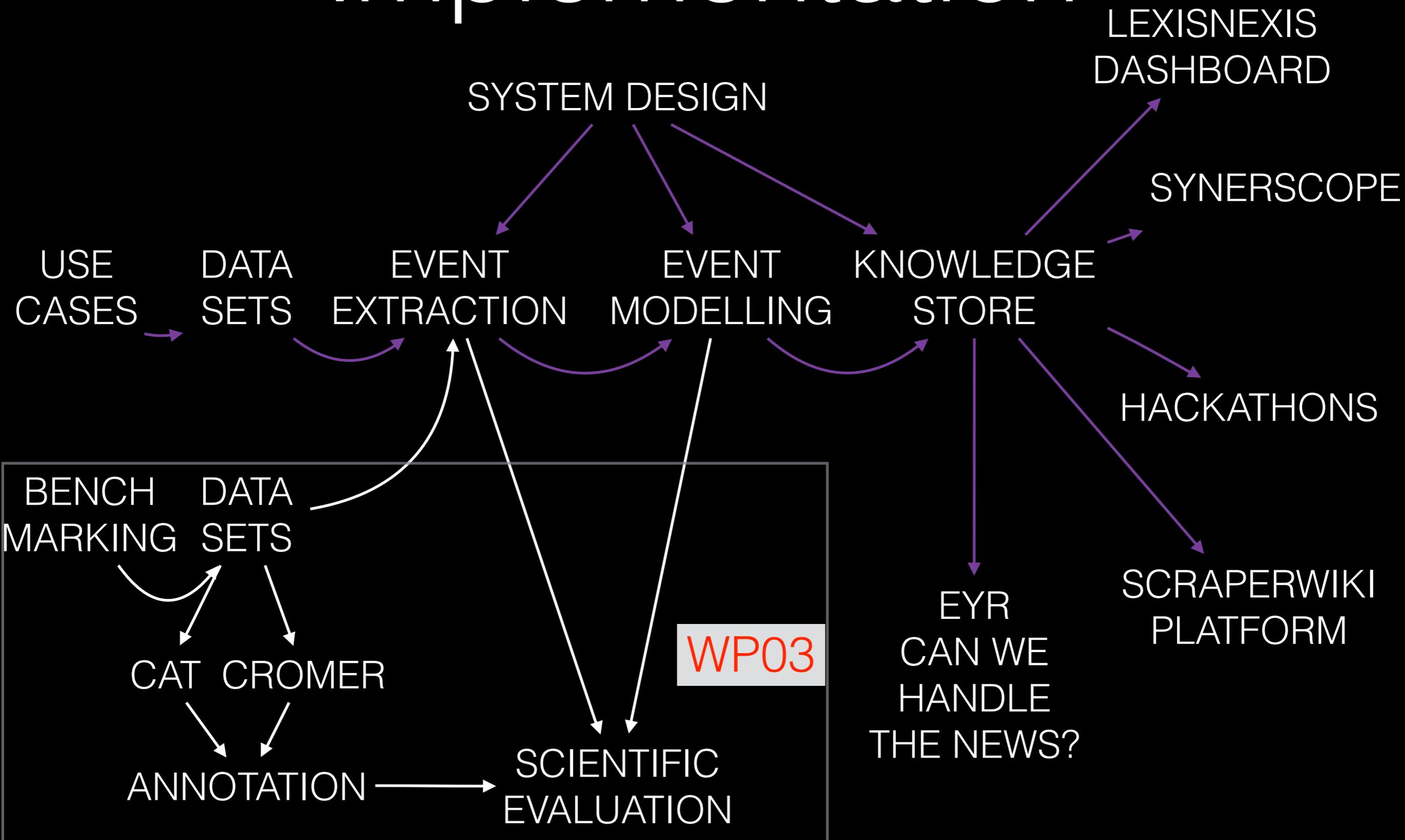
Implementation



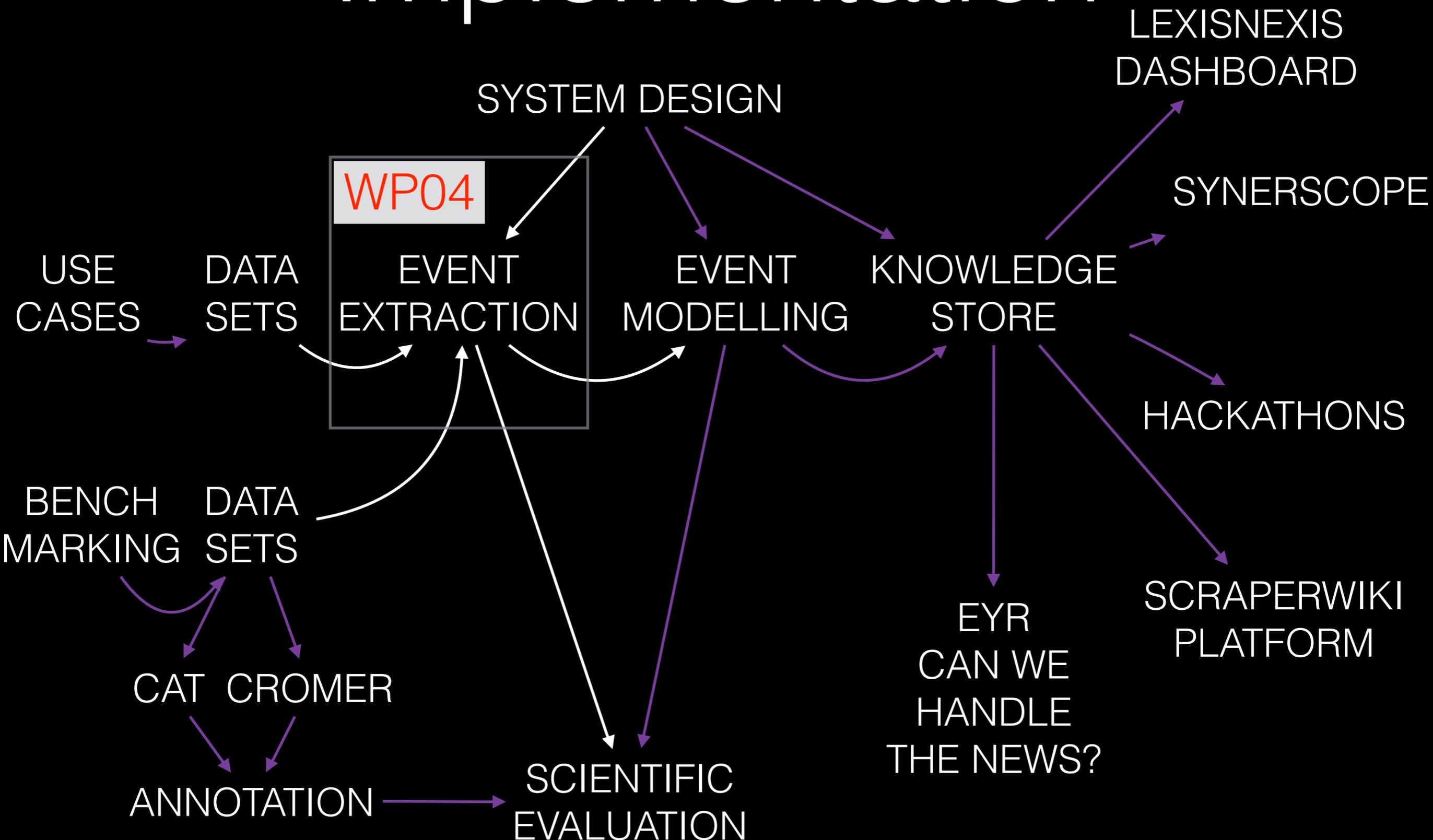
Implementation



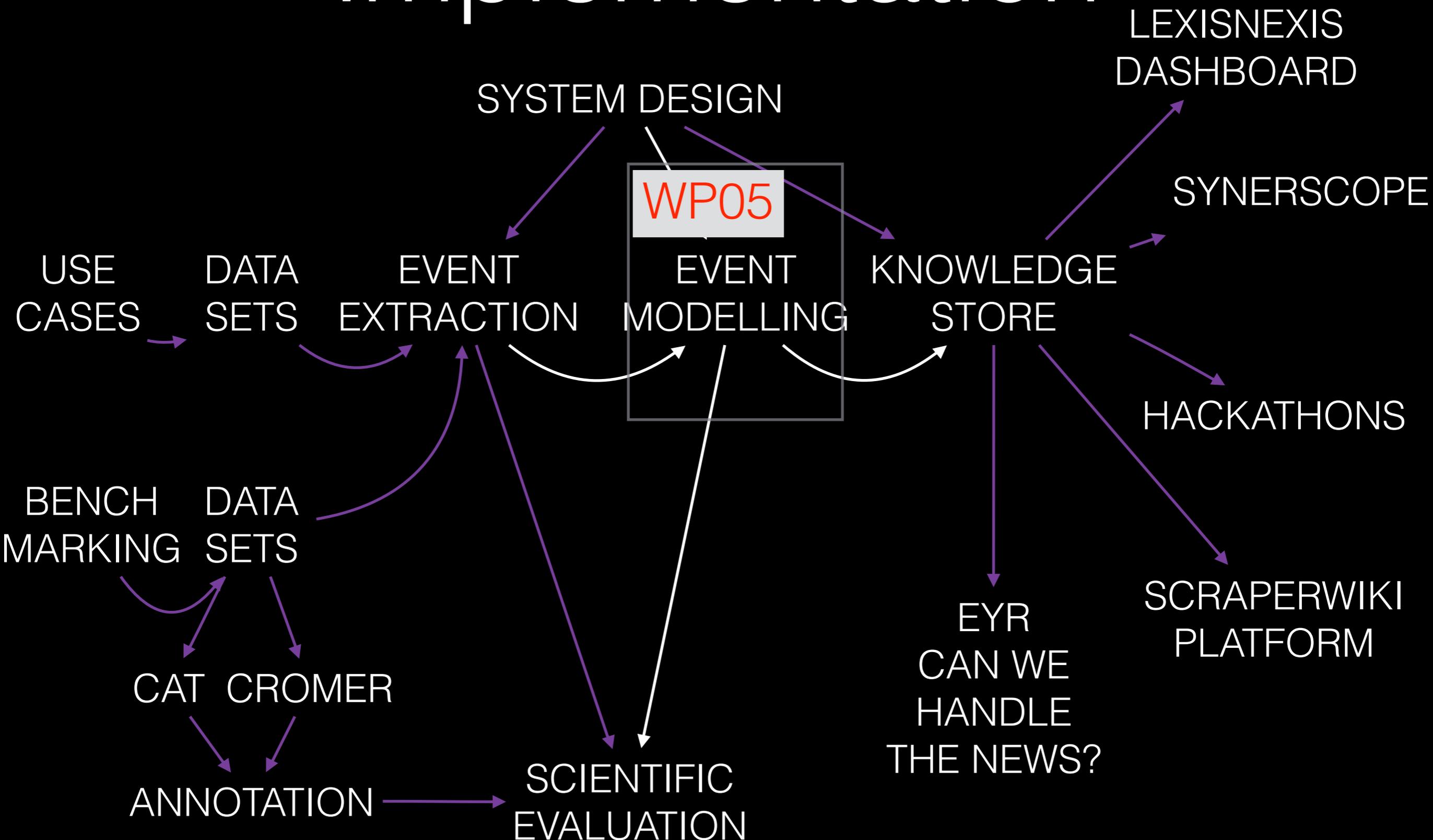
Implementation



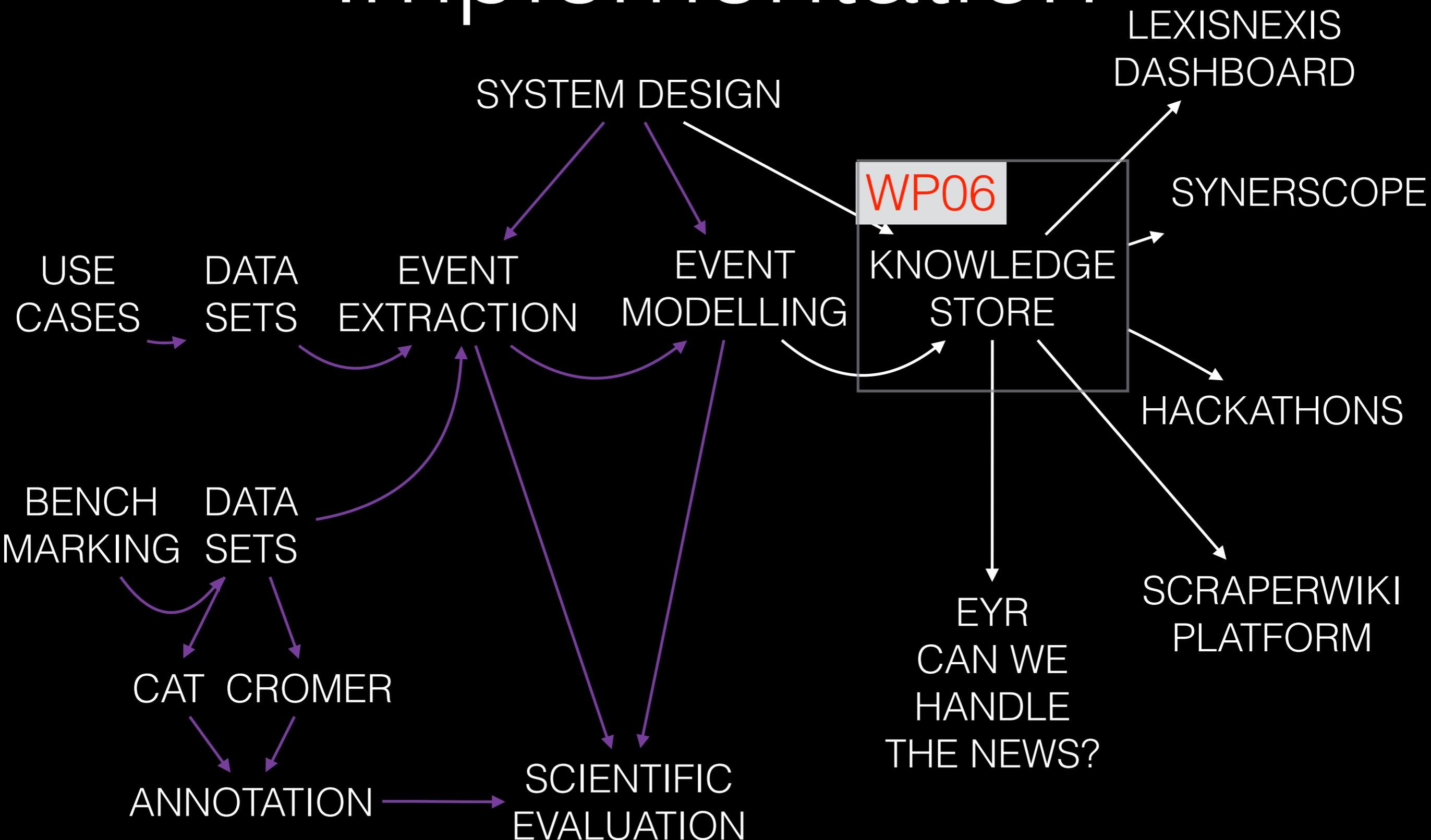
Implementation



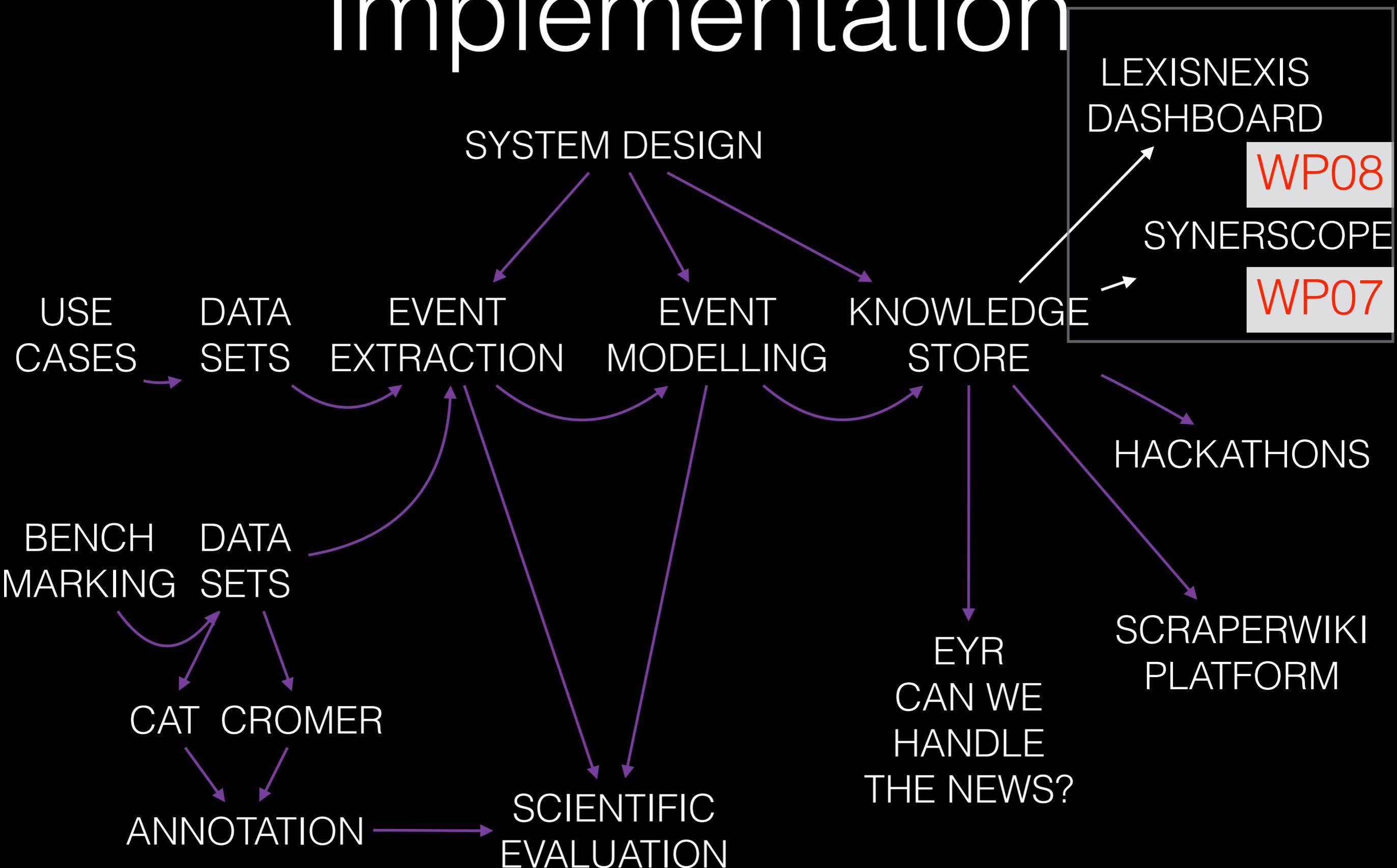
Implementation



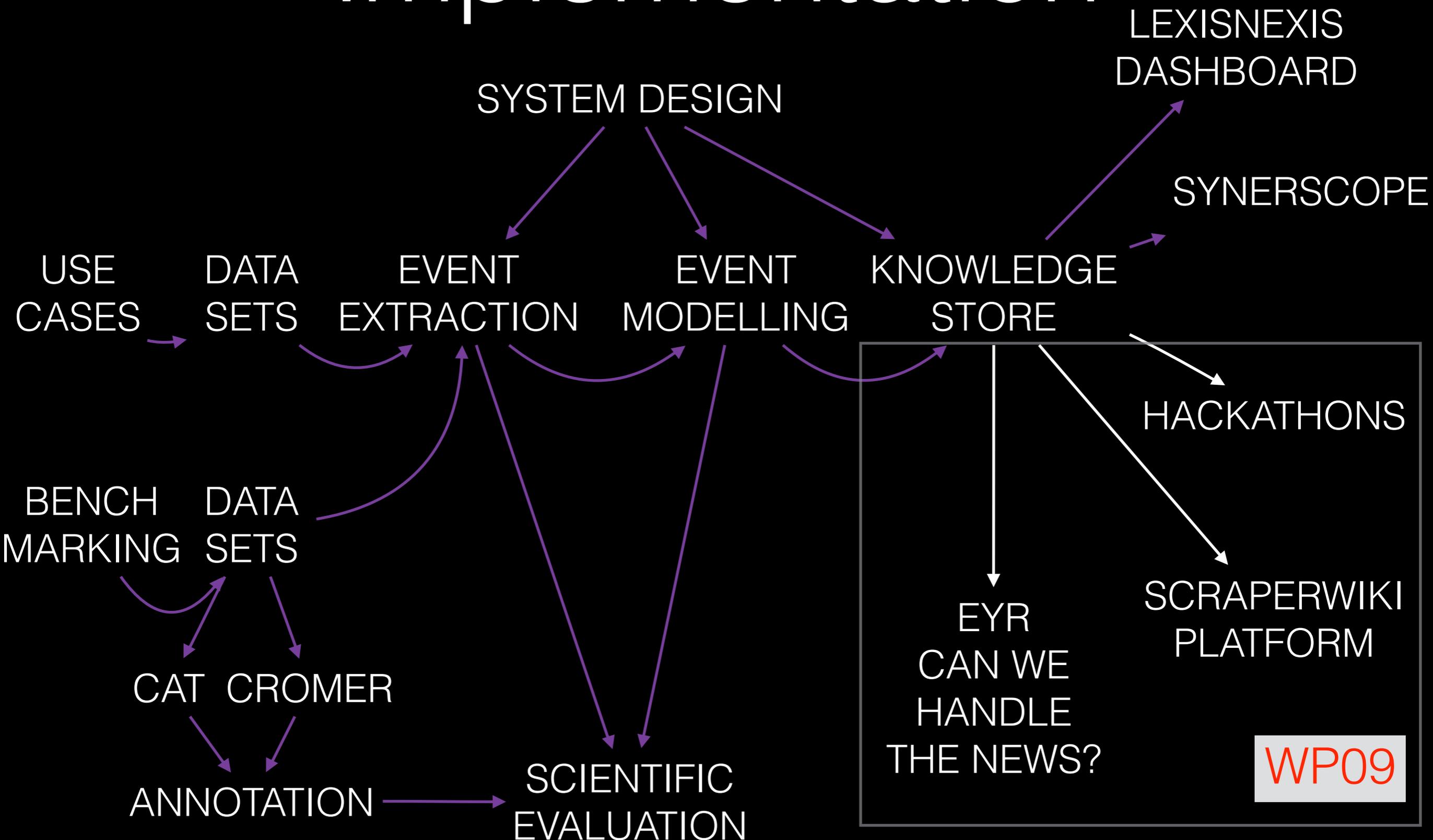
Implementation



Implementation



Implementation



Overall schedule

- 3 cycles: specification - development and evaluation: year-1, year-2 and year-3
 - End-user evaluations: month 13, 25, 36
 - Benchmark evaluations, hackathons and shared task
- Each year: increase in volume, complexity and quality

NewsReader Project overview Results SECOND year

Over half a billion triples attested

<https://vimeo.com/120574800>

Results of the 2nd year

- 2nd cycle: user feedback → design and specification → implementation → data and text processing → evaluation.
- Volume, scalability and robustness
- Multilingual and cross-lingual benchmarking
- Glass-box system that will drive future research and development

Results of the 2nd year

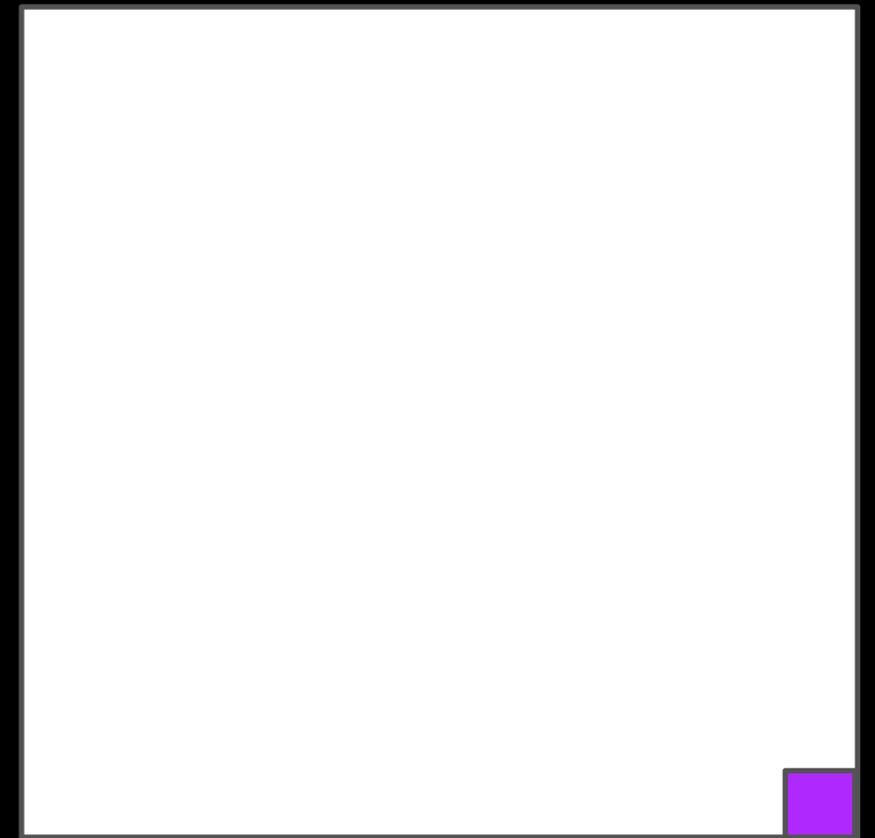
- Volume, complexity and richness:
 - From 100K articles to 1.5M articles
 - From English to *Spanish, Dutch, Italian* and *cross-lingual*
 - From 47M triples to 535M, including 2M inferred and 96M DBpedia
- (improved) STORM and (new) HADOOP architectures for parallel processing —> WP02
- Easily deployable Virtual Machines (VMs) for English, Spanish, Dutch, Italian —> WP02, WP04
- KnowledgeStore handles massive data and requests (over 100K during Amsterdam hackathon) —> WP06



LN daily output



Typical usage, monthly



**Car use case,
10 years data**



Wikinews,
since 2005

Techcrunch



= 40,000 articles
= 80 reams of A4



LN daily output



Typical usage, monthly

1,3M articles
100K articles/year
95GB compressed
processed in 1 week

10M obj
1M TRiG
226GB
500M triples
1 week

**Car use case,
10 years data,
6M articles**



Worldcup:
212K articles

Techcrunch

= 40,000 articles
= 80 reams of A4



Wikinews,
since 2005



Criminal networks
160K articles

Results of the 2nd year

- Multi-lingual, cross-lingual, cross-document, cross-topic benchmarking (WP03, WP04, WP05):
 - 120 Wikinews articles translated and annotated using CAT and CROMER
 - 982 ECB+ articles for cross-document and cross-topic coreference
 - benchmark results for English surpass State-of-the-Art on various data sets
 - other languages approaching English State-of-the-Art

Results of the 2nd year

- Interoperable© semantic pipelines across languages!!!!
- Resulting in a unified RDF representation of unstructured textual data in different language

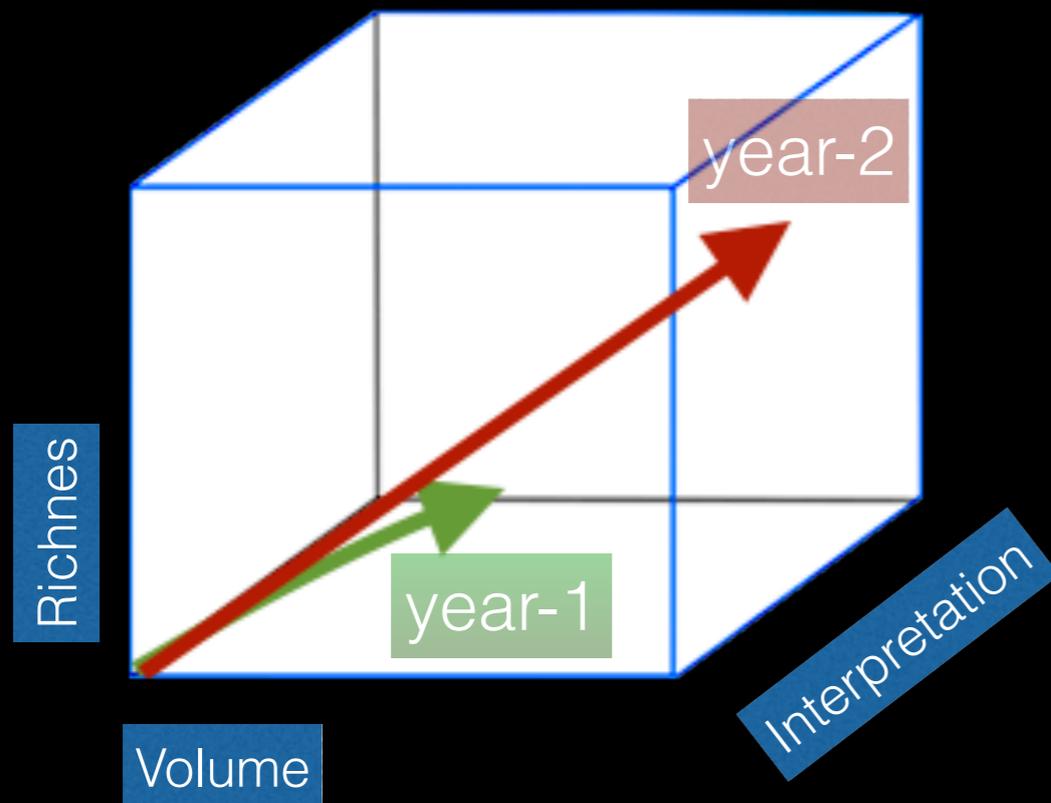
Results of the 2nd year

- 2nd cycle: user feedback → design and specification → implementation → data and text processing → evaluation.
- Knowledge crystallization and reasoning:
 - ESO and reasoner to obtain implications of events
 - Domain modeling to learn about entity interpretations and their roles
- Perspective model: sources can disagree with themselves
- Microstories: capture the core of the story through bridging

Results of the 2nd year

- 2nd cycle: user feedback → design and specification → implementation → data and text processing → evaluation.
- 3 Hackathons and 3 user evaluations: valuable feedback on all levels, e.g. quality, quantity, richness, API, interfacing
- 3rd party take-up: NGOs, governments, banks, security
- SemEval2015 timelines: organization and participation
- SynerScope mention-base and event-base views

Progress



- **Richness:**

- new NAF layers
- more SEM relations
- perspective layer

- **Volume**

- 20 times more, 20 times faster

- **Interpretation:**

- alternatives in layers
- crystallization: ESO + domain
- inferred properties
- timelines & microstories

Recommendations

- [Recommendation 1] Highlight the technical innovations and progress made by individual components of the system, as compared to (1) the state-of-the art, and (2) the starting point
 - Benchmarking on standard data sets & Wikinews
 - Generic modules and adapted modules
 - Deliverables indicate project results and progress
 - D9.1 Draft Exploitation plan defines background/foreground
 - D4.1.2 Event detection, version 2 table with status

Pipelines and modules for Dutch, Spanish and Italian. Table 15 presents the current state of all the available modules.

Name	2nd year	Version	1st year	Language(s)	Third-party software	Resources
ixa-pipe-tok	No update	1.5.0	✓	All		
ixa-pipe-pos	Update	1.3.0	✓	EN, ES	API Apache OpenNLP project	WSJ treebank, Ancora

NewsReader: ICT-316404

February 20, 2015

fbk-causalrel	New	1.0.0		EN		Tagger, YamCha Explicit Discourse Connectives Tagger, YamCha mallet	2014 data TimeBank
vua-factuality	No update	1.0	✓	EN			FactBank
opinion-miner	No update	2.0	✓	EN, NL			OpENER models

NewsReader: ICT-316404

February 20, 2015

Event Detection, version 2

81/99

vua-alpino	Update	1.0		NL	Alpino parser	
vua-heideltime	New	1.0		NL	HeidelTime, TreeTager	
vua-wsd	Update	1.2		NL		DutchSemCor
vua-ontotagging	Update	1.0		NL		Predicate Matrix
vua-srl	New	1.0		NL	TimBL	SoNaR SRL data
vua-framenet-classifier	New	1.0		NL		Predicate Matrix
ixa-heideltime	✓	1.0.0		ES	HeidelTime	
fbk-tokenpro	no update	2.1	✓	EN, IT		
fbk-morphopro	no update	1.3.2-1	✓	EN, IT		word form list
fbk-tagpro	no update	1.5.0	✓	EN, IT	Yamcha, TinySVM	ELRA tagset
fbk-lemmapro	no update	2.0	✓	EN, IT		
fbk-entitypro	no update	1.4.3	✓	EN, IT	Yamcha, TinySVM	ICAB corpus
fbk-chunkpro	no update	2.0	✓	EN, IT	Yamcha, TinySVM	
fbk-depparsepro	no update	1.0	✓	IT	MaltParser	Turin University Treebank

Event Detection, version 2

82/99

fbk-eventpro	new	1.0		IT		Snowball Italian stemmer, Multi-WordNet domains, derIvaTario lexicon, Yamcha, TinySVM	EVENTI-Evalita 2014 data
fbk-factpro	new	1.0		IT		derIvaTario lexicon, Yamcha, TinySVM	Fact-Ita Bank
text-classification	New	1.0.0		All		JEX	
Predicate Matrix	Update	1.1.0	✓	EN, ES			PropBank, NomBank, Ancora, VerbNet, FrameNet, SemLink, WordNet

Table 15: EN, IT, NL and SP modules.

pipelines and modules for Dutch, Spanish and Italian. Table 15 presents the current state of all the available modules.

Name	2nd year	Version 1st	Language(s)	Third-party software	Resources
ixa-pipe-tok	No update	1.5.0	✓	All	
ixa-pipe-pos	Update	1.3.0	✓	EN, ES	
vua-alpino	Update	1.0	NL	Alpino parser	
vua-heideltime	New	1.0	NL	HeidelTime, TreeTager	
vua-wsd	Update	1.2	NL		DutchSemCor
vua-ontotagging	Update	1.0	NL		Predicate Matrix
vua-srl	New	1.0	NL	TimBL	SoNaR SRL data
vua-framenet-classifier	New	1.0	NL		Predicate Matrix
ixa-heideltime	✓	1.0.0	ES	HeidelTime	
fbk-tokenpro	no update	2.1	✓	EN, IT	
fbk-morphopro	no update	1.3.2-1	✓	EN, IT	word form list
fbk-tagpro	no update	1.5.0	✓	EN, IT	Yamcha, TinySVM, ELRA tagset
fbk-lemmapro	no update	2.0	✓	EN, IT	
fbk-entitypro	no update	1.4.3	✓	EN, IT	Yamcha, TinySVM, ICAB corpus
fbk-chunkpro	no update	2.0	✓	EN, IT	Yamcha, TinySVM
fbk-depparsepro	no update	1.0	✓	IT	MaltParser, Turin University Treebank

fbk-causalrel	New	1.0.0	EN		Tagger, YamCha	2014 data
					Explicit discourse connectives tagger, YamCha, allet	TimeBank
						FactBank
						OpenNER models
						February 20, 2015
fbk-eventpro	new	1.0	IT		Snowball Italian stemmer, Multi-WordNet domains, derIvaTario lexicon, Yamcha, TinySVM	EVENTI-Evalita 2014 data
fbk-factpro	new	1.0	IT		derIvaTario lexicon, Yamcha, TinySVM	Fact-Ita Bank
text-classification	New	1.0.0	All		JEX	
Predicate Matrix	Update	1.1.0	✓	EN, ES		PropBank, NomBank, Ancora, VerbNet, FrameNet, SemLink, WordNet

Table 15: EN, IT, NL and SP modules.

Recommendations

- [Recommendation 2] State the way toward the final end-user requirements (the global end-user decision making tool) on every aspect of the system (eg, quality of semantic processing, quality and size of training data, ergonomics of the end-user interface, etc.).
 - Hackathons & end-user evaluations as feedback on all system aspects —> crystallization instead of training data, choosing from alternative interpretations for entities, events, roles and concepts
 - 3rd party collaborations: Dutch House of Representatives, Spanish Ministry of Industry, Dutch National bank, Bank of England, Dutch intelligence/policy bodies and NGOs

Recommendations

- [Recommendation 3] Considering the large volume of data expected for each use case scenario and the continuum upload of documents in the KnowledgeStore the consortium should evaluate an alternative architecture for the proposed local DSTS database and import process.
 - Only selected data is loaded
 - Selection scenarios clarified during hackathons and user interactions
 - Implementation in 3rd year
 - Further experiments planned within streaming architecture

Prospect 3rd year

- Further benchmarking and improvement of the NLP modules on the basis of the annotated data, also for Dutch, Spanish, Italian and cross-lingual
- Full integration of modules with the KnowledgeStore as one system: black-box production variant
- Setup of streaming experiments to handle incoming news and integration with a given KnowledgeStore database

Prospect 3rd year

- Implementation of perspective modules and microstories analysis
- Event-coreference based on topic clustering and microstories
- Storylines based on microstory integration

Prospect 3rd year

- Further development of ESO and the reasoner following the crystallization methodology
- Cross-lingual semantic processing and integration
- Domain adaptation and background modeling

Prospect 3rd year

- Setup experiments to demonstrate the perspective layer of the data within the SynerScope tool
- Exploration of other use cases
- Final end-user evaluations
- A final hackathon and final workshop: European DataForum, Luxembourg, Nov. 16-17 2015
- Exploitation in new projects and by industrial partners

Crystallization

