

# RT-203, Meshing Capability and Threat-based Science & Technology Resource Allocation

**Sponsor: DASD(SE)**

By

**Dr. Carlo Lipizzi**

**10<sup>th</sup> Annual SERC Sponsor Research Review**

**November 8, 2018**

**FHI 360 CONFERENCE CENTER**

**1825 Connecticut Avenue NW, 8th Floor**

**Washington, DC 20009**

**[www.sercuarc.org](http://www.sercuarc.org)**

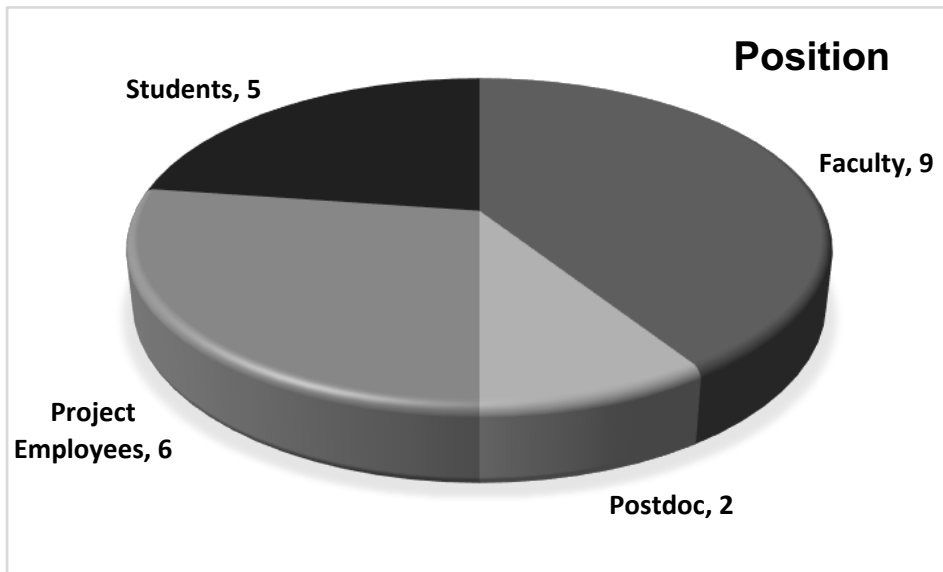
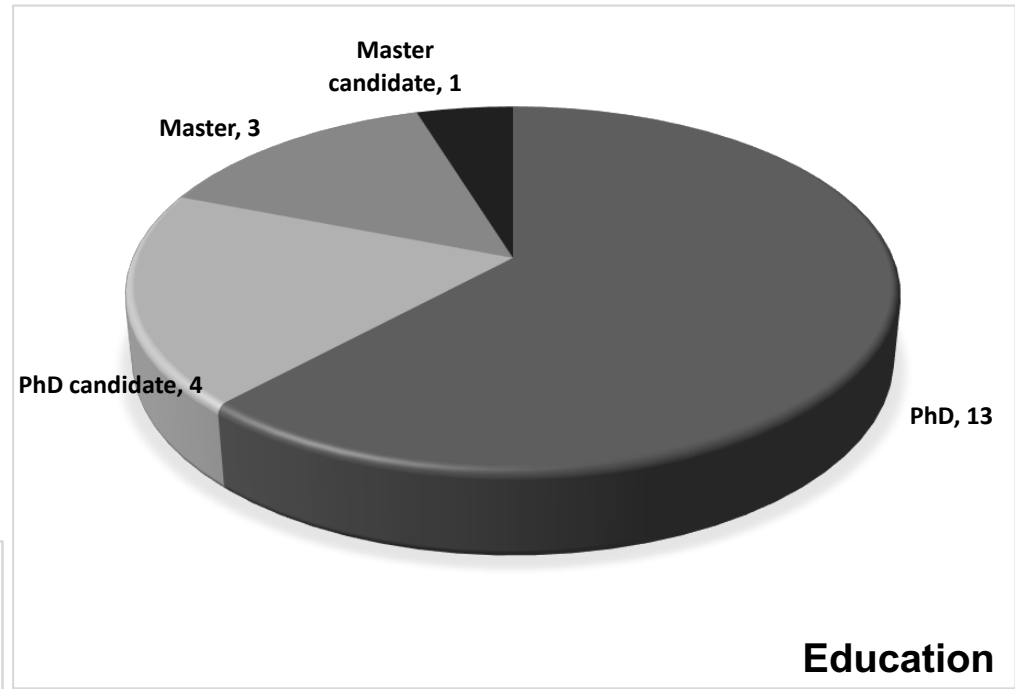


- **Title:** Meshing Capability and Threat-based Science & Technology Resource Allocation
- This research is focused on providing a computational model to support the planning cycle injecting relevant threat-based intelligence and operational scenarios into the more traditional capabilities-based planning
- This approach will better inform the technical communities charged with developing future weapons systems and has been piloted in late 2016 at the U.S. Army Armaments Research and Development Engineering Center (ARDEC) in the armament-systems domain

- **Replicate the process developed at ARDEC in 2016** to validate this notional computational architecture
- **Enhance the visualization and analytic capability** to allow rapid, high fidelity decision making
- **Introduce additional parameters and variables** to refine the decision making framework. Real-world scenarios will be modeled to project evolving threats, doctrine, partner force interoperability, and other operational environmental conditions (political, military, socio-economic, information, infrastructure, physical environment)
- Deliver the results with an agile approach, **developing prototypes/proofs of concepts with increasing capabilities, using a partially automatic learning approach**
- Project phases:
  - Phase I (FY 17): awarded
  - Phase II (FY 18): proposed
  - Phase III (FY 19): proposed

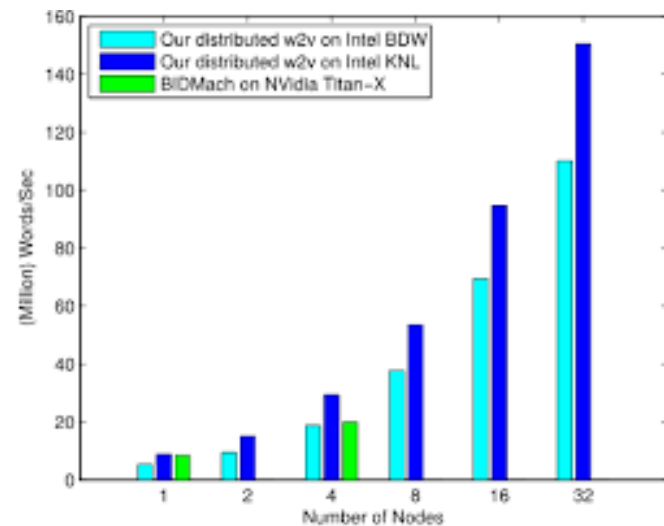
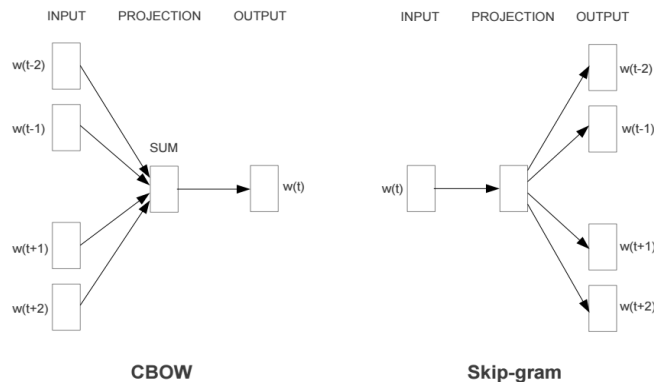
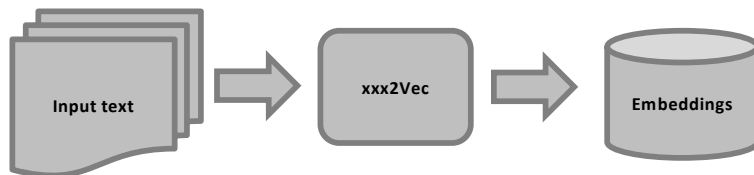
- Project Kickoff – Define collaborators’ Scope of Work (end of April ‘18)
- Monthly status report (mid of each month from June)
- Interim Process Report/Mid-year review (11/15/18)
- Final Technical Report (4/15/19)

- **Total number** \_\_\_\_\_ **22**
  - "Permanent" members \_\_\_\_\_ 18
    - $\geq 50\%$  of their time \_\_\_\_\_ 9
  - "Temporary" members \_\_\_\_\_ 4



- Data/Text driven approach
- Using a proxy-domain to source the data
- Systems are developed as agile growing prototypes with modular components. Most of the components are developed separately for a better reuse

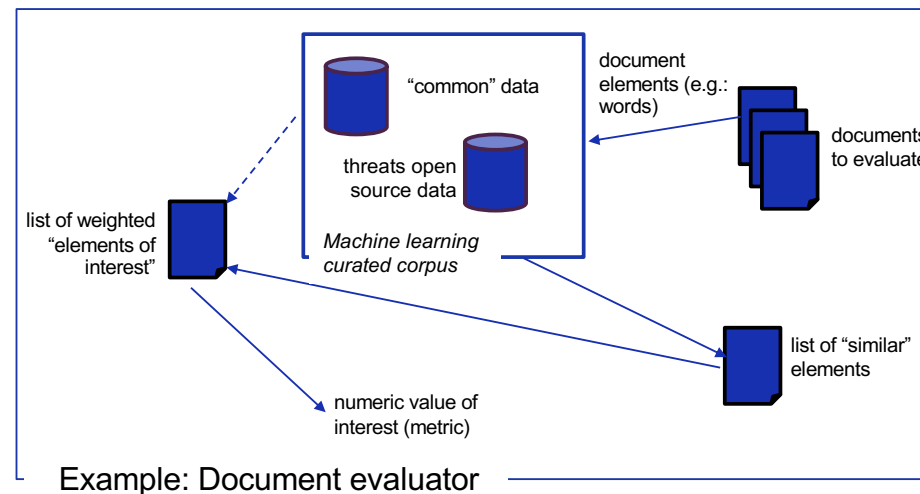
- All our models are based on data/text, with human validation
- We collect the data we use. This is text related to a specific “proxy” domain, selected to be more easily associated – for content and complexity – to the final target domain
- We use a combination of traditional Natural Language Processing (mainly for the preprocessing) and embeddings, that are feature vectors for conversational elements in that text (such as words), calculated via Python libraries based on neural networks, such as Word2Vec
- From the embeddings we extract specific metrics we use for risk evaluation and for visualization
- We use dedicated GPUs to create the embeddings from the corpora, driving down the time to obtain the embeddings from days to hours



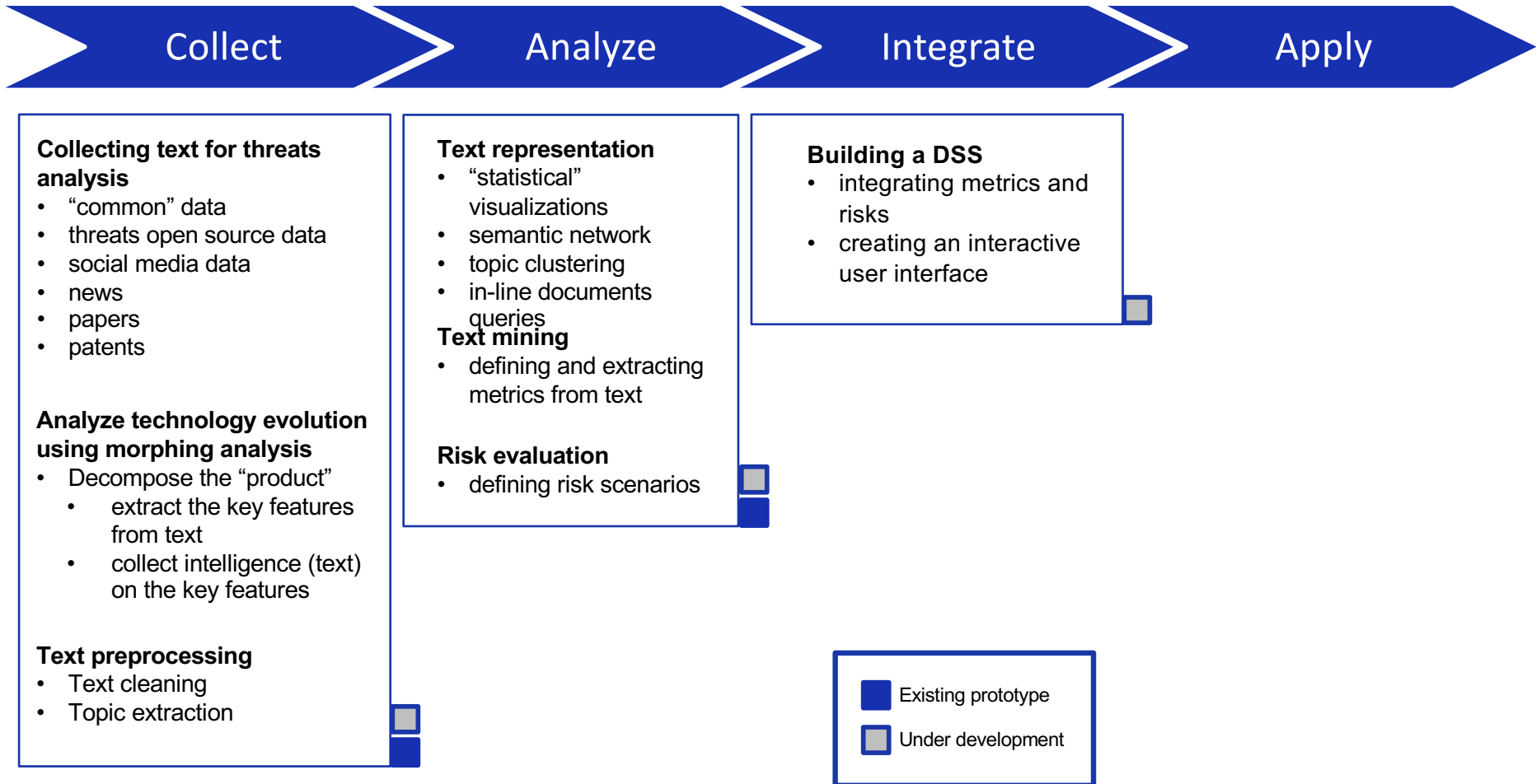
- We are developing 2 Systems and several Components
- Systems are:
  - **Risk panel – Planning Support System.** Primary outcome from this system will be an interactive panel that can be used for all the what-is analyses, with a future layer of Machine Learning trained by the user interactions and suggesting the “best” scenarios
  - **Technology Monitoring System.** This will scan for elements outside the defined domain, scanning a broader set of sources looking for those technologies that are emerging. It will provide the Sponsor with a way to be prepared for future technologies that may have an impact on their activities



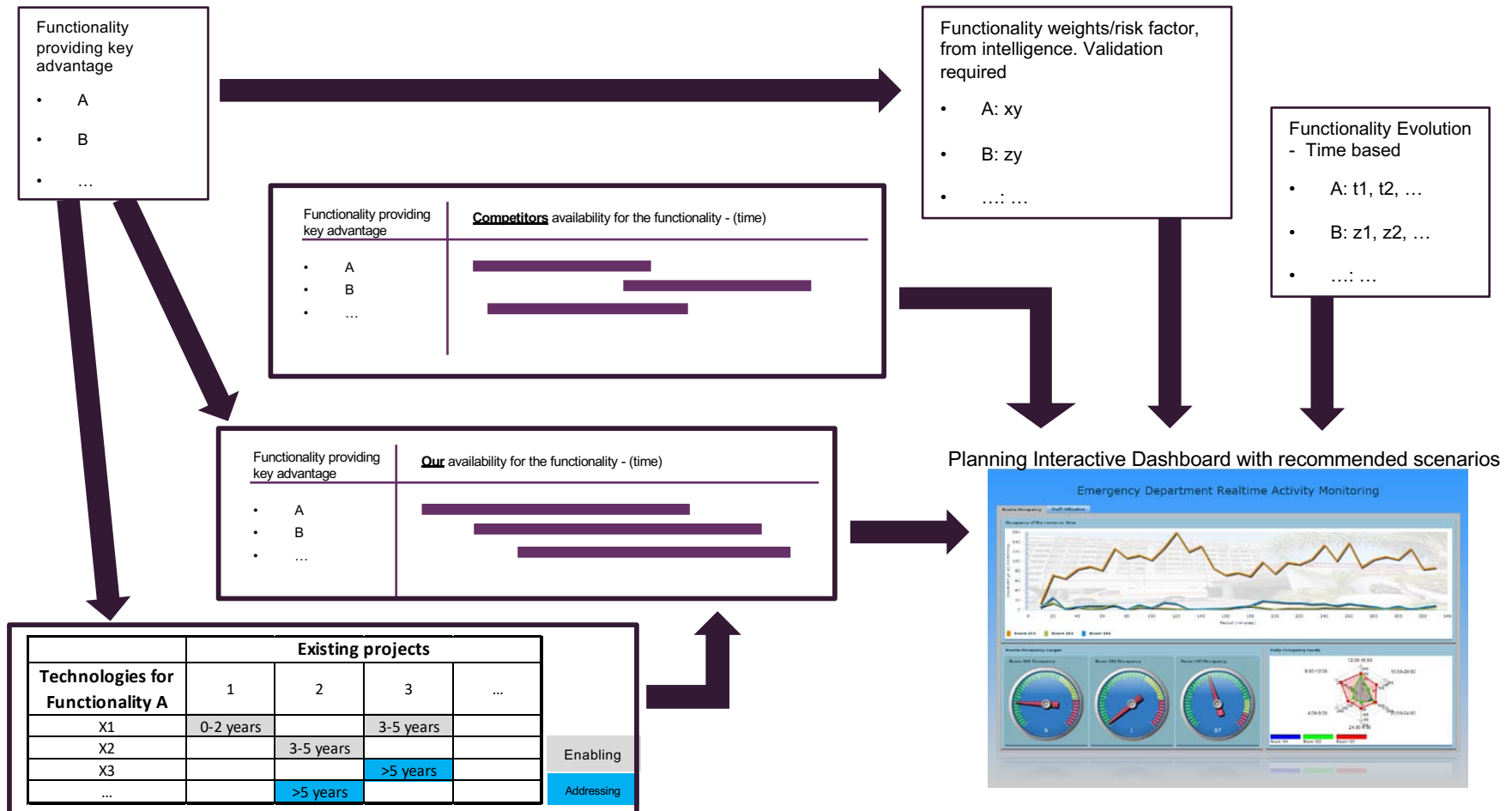
- **Basic Components**
  - Text Cleaning, Topic extraction, Chunking/named entity recognition, News scanner, Word2Vec embeddings matrix, Word/chunk semantic similarity, Opinion polarity evaluator
- **Complex Components**
  - Weighted “critical topic”- list generator, Document evaluator, Social validator, Social Media collection and analysis
- As 10/20/18 our components cover about 20 tasks, with about 1000 lines of Python code



## Functional View



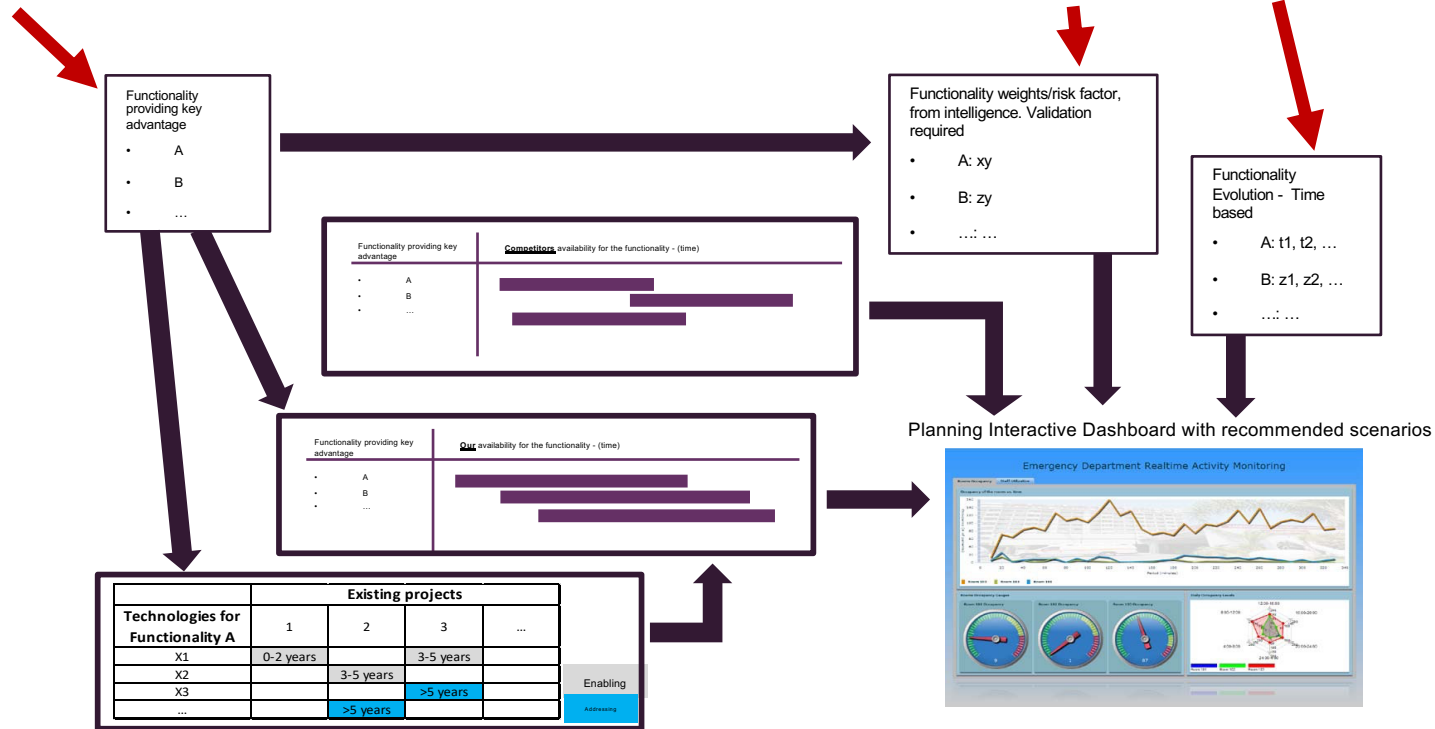
## Overall architecture



## Details/components

Extracted from “domain-specific” dataset\*

Calculated using existing components

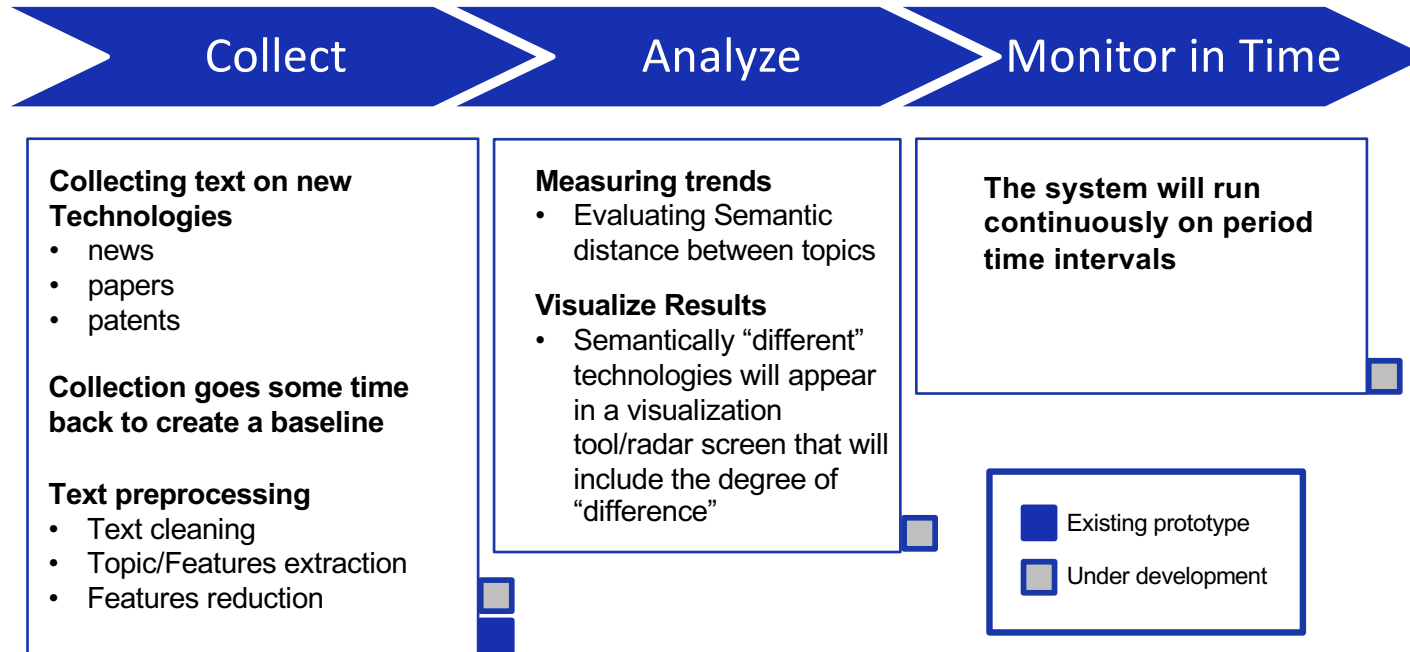


Technologies for Functionality A	Existing projects			
	1	2	3	...
X1	0-2 years		3-5 years	
X2		3-5 years		
X3			>5 years	
...		>5 years		

- Technologies extracted via embeddings on full dataset
- Projects extracted via embeddings on “domain-specific” dataset\*
- Timeline extracted via embeddings on “domain-specific” dataset\*

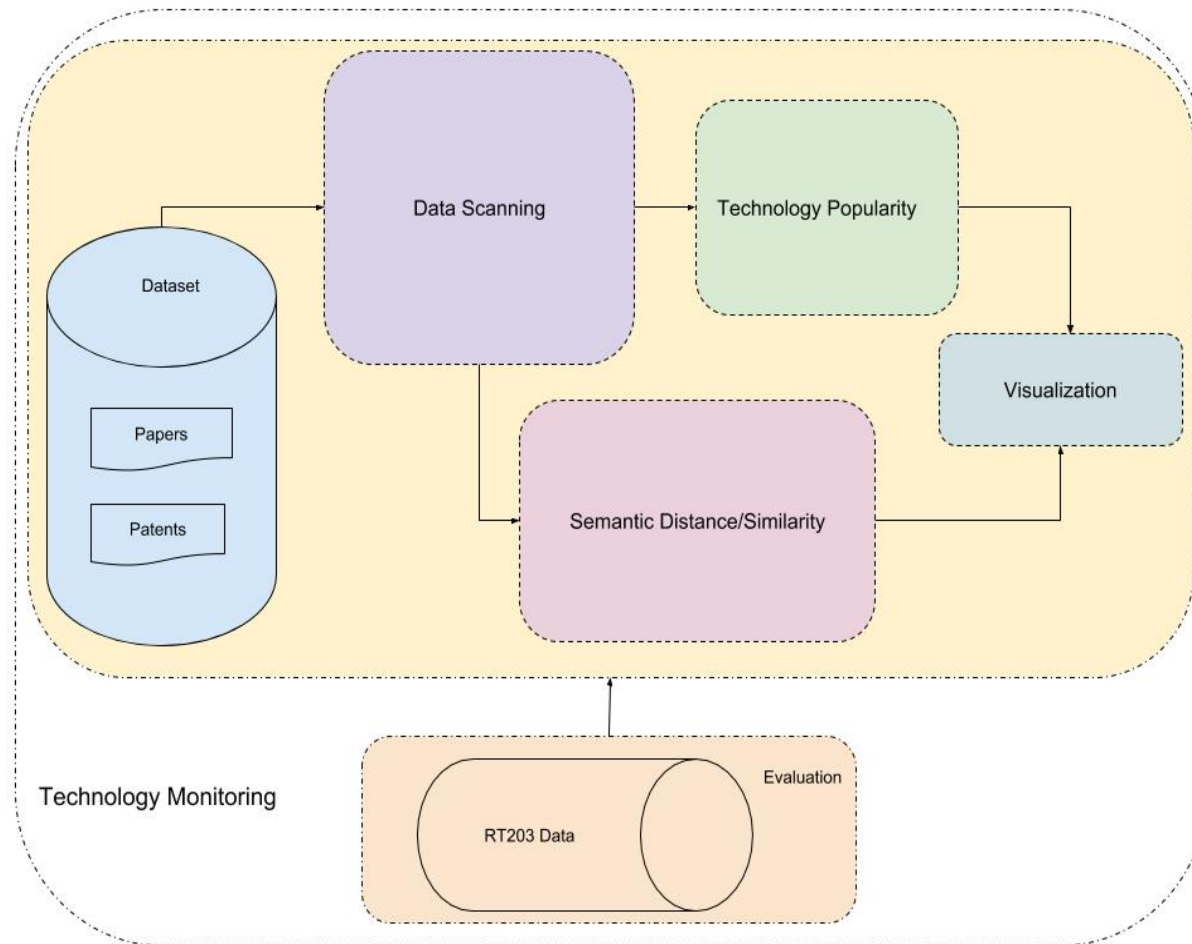
\*: This is a corpus generated by collecting domain-specific text

## Functional View



- This system has a broader but less complex goal, looking for emerging technologies
- In the collection phase, a features reduction step will avoid monitoring technologies too far from our scope

## Overall architecture



- We are working on papers focused on specific aspects of our research
  - *Journal paper* - Text mining in an evolving society: getting insights from text in times of minimally structured conversations. This paper is an evolution of a conference paper Dr. Lipizzi presented at CESUN/Tokyo in June '18 and will contain one of the techniques for text mining used in RT 203
  - *Conference paper* - Product development in industries with long term planning and asynchronous events. This paper is a survey on short term events which directly impact on planned design and development. The aerospace, ship-building and automotive manufacturing industries were used as proxies for this research. This research enabled the RT 203 team to work with relevant, open source material
  - *Journal paper* - Validating online opinions. The proposed method in this paper will provide a way to select users based on the correlation between their comments and a given control metric. This will help us to evaluate the accuracy rate of the information provided by the user. The proposed model is build based on the text mining techniques used in RT 203

- Systems are in an early prototyping phase, with the first working proof of concept ready by Mid of November
- All the basic components are ready to be used
- A first release of the complex components will be available by Mid of November
- The proxy domain has not been selected. The team is working on a “reasonable option”, that is ML/AI in connected environments
- Working on a 2 years project extension with the following goals:
  - 1<sup>st</sup> year of extension will bring the system from the proof of concept state to a working prototype. A preliminary scenario recommendation layer based on reinforced learning will be provided. The prototype will be tuned-up to work on the actual domain
  - 2<sup>nd</sup> year of extension will evolve the prototype and expand the reach of the scenario recommendation system. An handover to the sponsor will be also performed during this stage