



knowledge on the health, safety and environmental impact of nanoparticles

# Creation of a critical and commented database on the health, safety and environmental impact of nanoparticles

## Challenges and objectives

Oded Maimon, Abel Browarnik, Rafi Korenstein, Francois Rossi, Hildo Krop, Pieter van Broekhuizen, Hanno Wittig, Anna Maria Lemor

Find more information on

[www.nhecd-fp7.eu](http://www.nhecd-fp7.eu)

[maimon@eng.tau.ac.il](mailto:maimon@eng.tau.ac.il)

[abel@eng.tau.ac.il](mailto:abel@eng.tau.ac.il)



NHECD has received funding from the European Commission's Seventh Framework Programme [FP7/2007-2013] under *grant agreement* n° 218639. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the given information.

# NHECD - Goal

The goal of **NHECD** is to build a ***free access***, ***robust*** and ***sustainable*** system including a ***knowledge repository*** on the **impact of nanoparticles on health, safety and the environment**

# NHECD - rational

- Most NanoTox knowledge repositories focus mainly on *metadata*.
- Repositories manually operated, therefore
  - only a limited amount of data is processed
  - taxonomy and ontology used to guide document classification and information extraction processesunsystematic

# NHECD

Funded by The European Commission FP7, **NHECD** is intended at converting the unstructured body of knowledge produced by the different groups of users (such as researchers and regulators) into a database of scientific papers and reviews (e.g., whitepapers), augmented by layers of information extracted from the above papers.

# NHECD - components

- A robust content management system (Documentum) as its backbone, to hold unstructured data (e.g., scientific papers and other relevant publications)
- A mechanism for automatically updating the documents repository, to ensure the creation of a large, updated and developing collection of published NanoTox data
- A mechanism to harmonize the repository, to hold it compatible at the metadata level with existing databases.

# NHECD - internals

**NHECD** will include:

- development of a systematic *domain model* of concepts and terms (i.e., a wide set of domain taxonomies) to support the classification of papers.
- development of the information extraction process. Particular, *domain-specific* zoning and text mining algorithms will be applied to reach the defined goals.

# NHECD - audience

The unique features of this database will allow different user groups – *academics, industry, public institutions* and the *general public* – to easily access, locate and retrieve information relevant to their needs

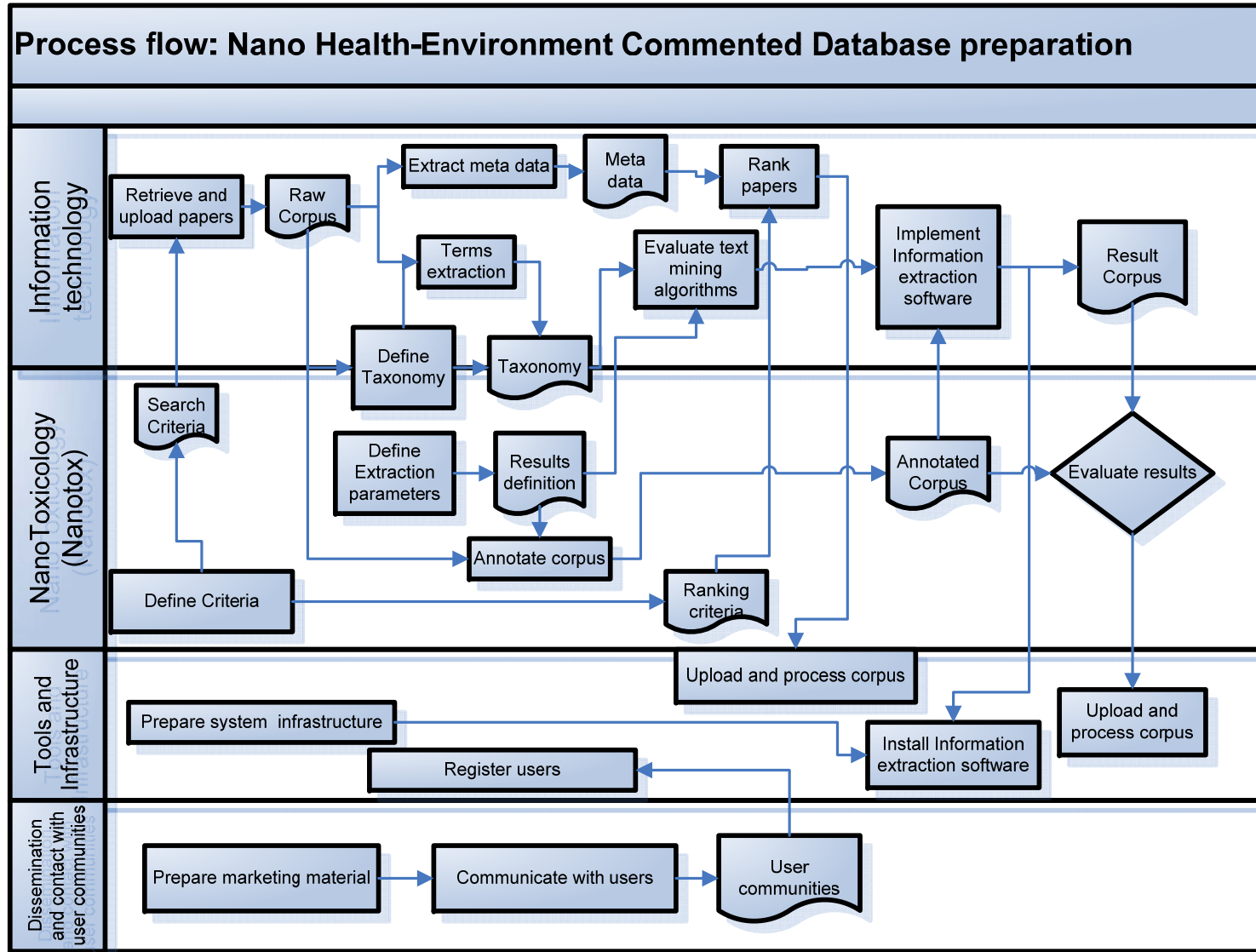
# NHECD - results

The creation of the *NHECD* knowledge repository

- Will enrich public understanding of the impact of nanoparticles on health and the environments
- will support a safe and responsible development and use of nanotechnology
- will represent a useful instrument for the implementation of relevant regulatory measures and legislation

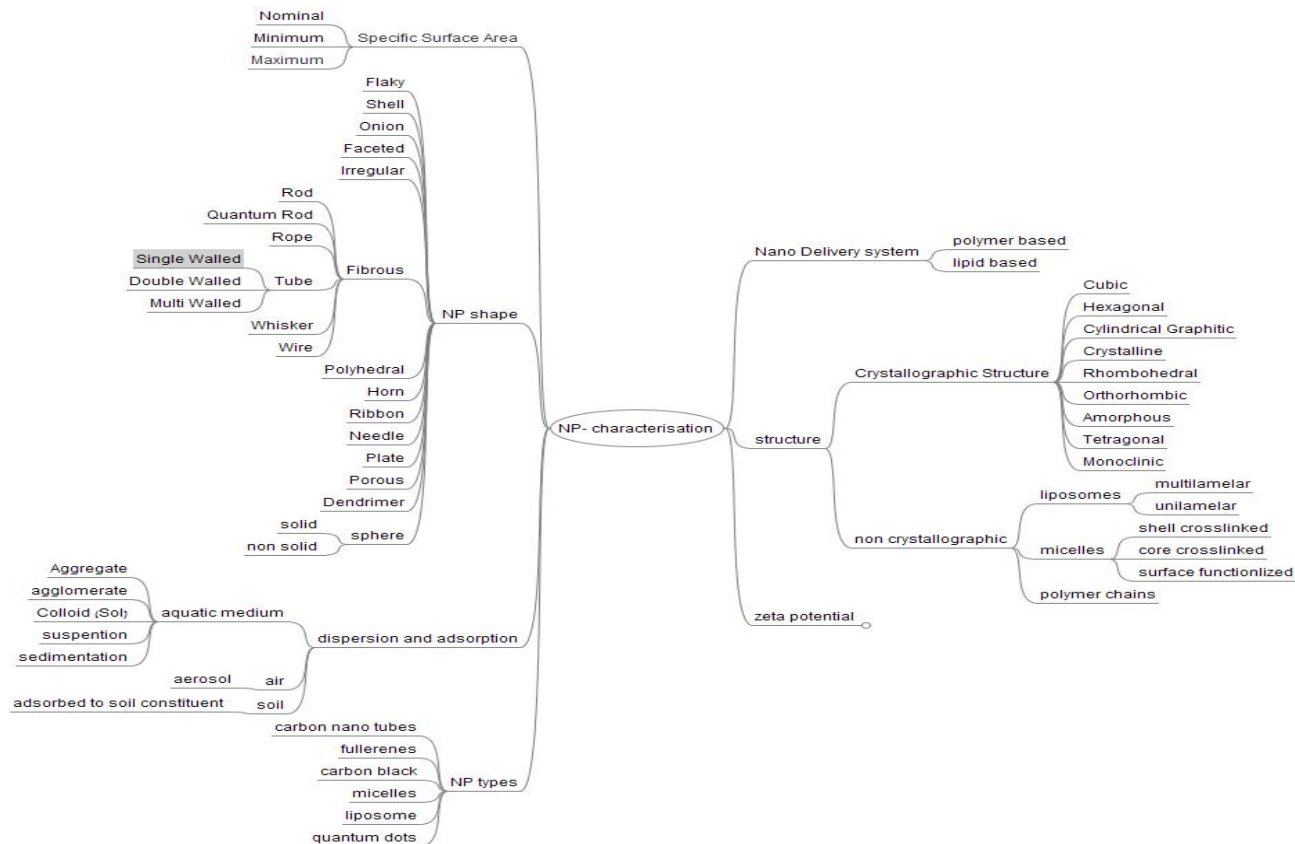


# NHECD – the model



# NHECD - implementation

- **Taxonomies:** essential to the NHECD process, which uses taxonomies to classify papers, both for information extraction and later on for repository navigation

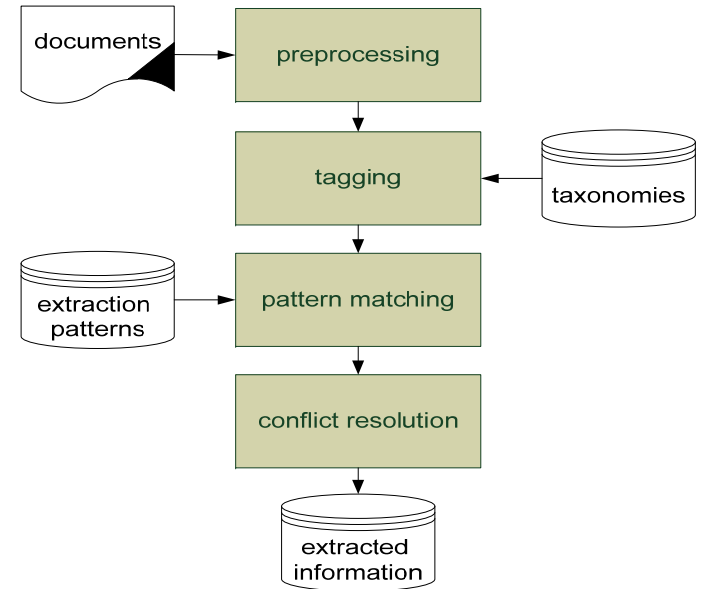


# *NHECD* - implementation

**Crawling:** The process of automatically obtaining scientific papers and data about the paper (e.g., author, publication date, name of the journal, keywords, abstract, etc.) by visiting scientific paper repositories available on the web (whether restricted to subscribers or available to everyone) and searching by keywords on the paper's text

# *NHECD* - implementation

**Information extraction:**  
methods and algorithms used to enable users to ask specific questions about attributes and receive answers and a link to the paper along with a pointer within the document where the information exists, and allow, in the future, data mining on the extracted information patterns



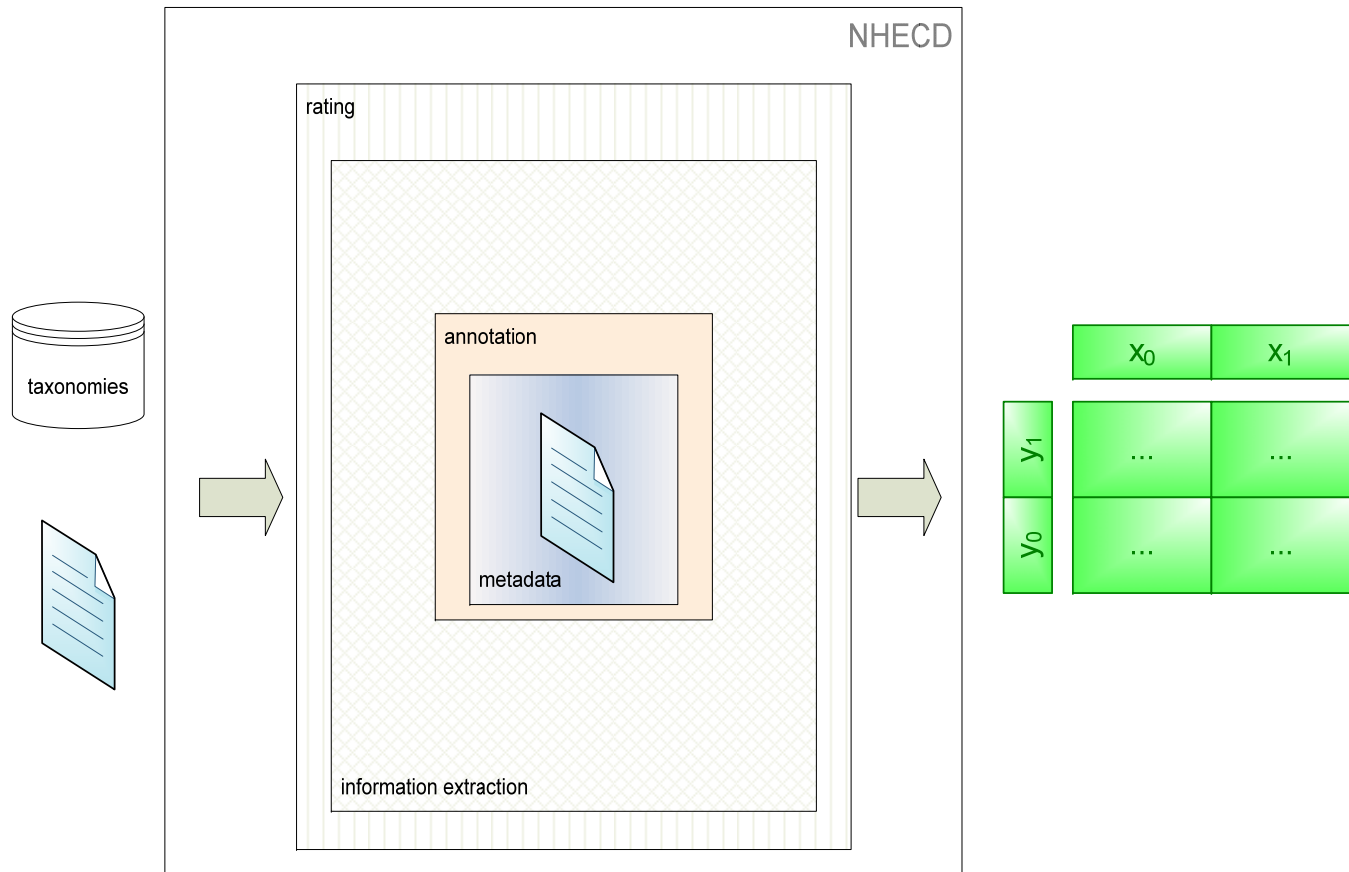
# ***NHECD*** - implementation

## **Rating of scientific papers**

one of NHECD contributions, seen as an additional *comment* layer.

Obtained by processing different sources and parameters

# NHECD - implementation



# ***NHECD*** - implementation

The result of the process consists of:

- A corpus of results, updated on an ongoing, asynchronous basis.
- A *commented* collection of scientific papers. By commented we refer to the added layer of metadata, rating and other information extracted from the document.

# NHECD - Conclusions

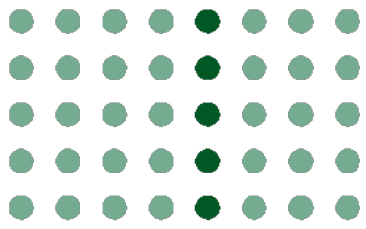
**NHECD** will provide two important *products*:

- An extensive and commented repository of scientific papers and other publications in the Nanotox area, searchable using taxonomies and full text search.
- A set of structured results extracted from the scientific papers populating the NHECD repository. Using these results it will be possible to perform data mining on the results. Data mining will result in validated results and further knowledge discovery. This part of NHECD results is targeted at Nanotox scientists and regulators.



# NHECD - Challenges

- Automatic population
- Information extraction
- Repository up-to-date
- Build/update taxonomies
- Papers rating
- Intelligent retrieval



# NHECD

knowledge on the health, safety and environmental impact of nanoparticles

## Thank you!

**Find more information on**

[www.nhecd-fp7.eu](http://www.nhecd-fp7.eu)

[maimon@eng.tau.ac.il](mailto:maimon@eng.tau.ac.il)

[abel@eng.tau.ac.il](mailto:abel@eng.tau.ac.il)