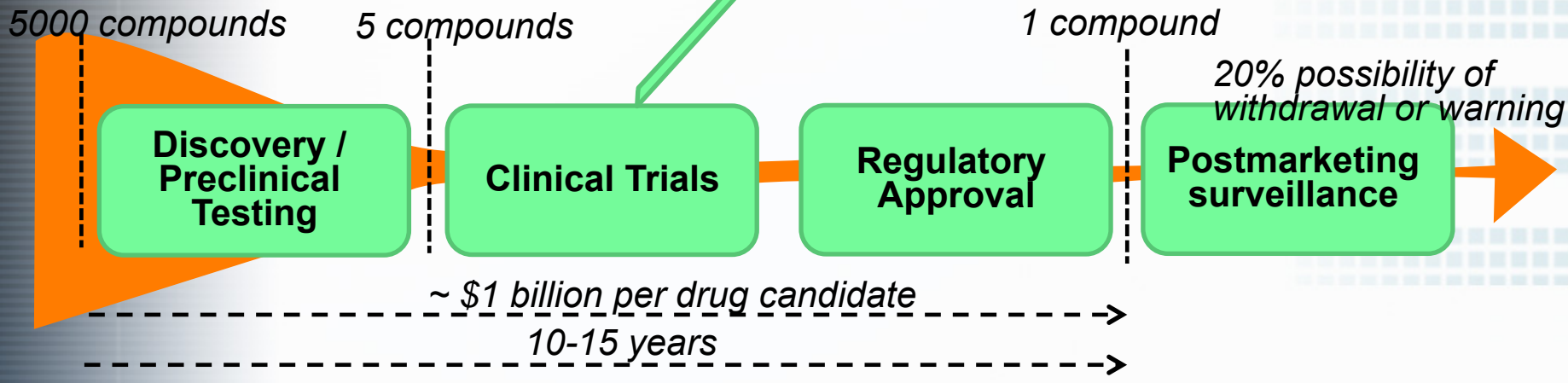
Data-driven, semantic-enriched, social-boosted Clinical Research and Healthcare

Prof. Theodora Varvarigou

ICCS/NTUA

PONTE Landscape

Patient Recruitment:
- **30%** - **40%** of clinical trial costs
- **60%** - **80%** of trials do not meet their temporal endpoint because of recruitment issues
- **30%** of trial sites fail to recruit even a single participant
- only **15%** of clinical trials conclude on schedule



- $53 billion worth of drugs fell off patent in 2012
- $44 billion sales at risk in 2015 due to patent expiration

Pharma focuses efforts on Drug Repositioning (40% of R&D resources)

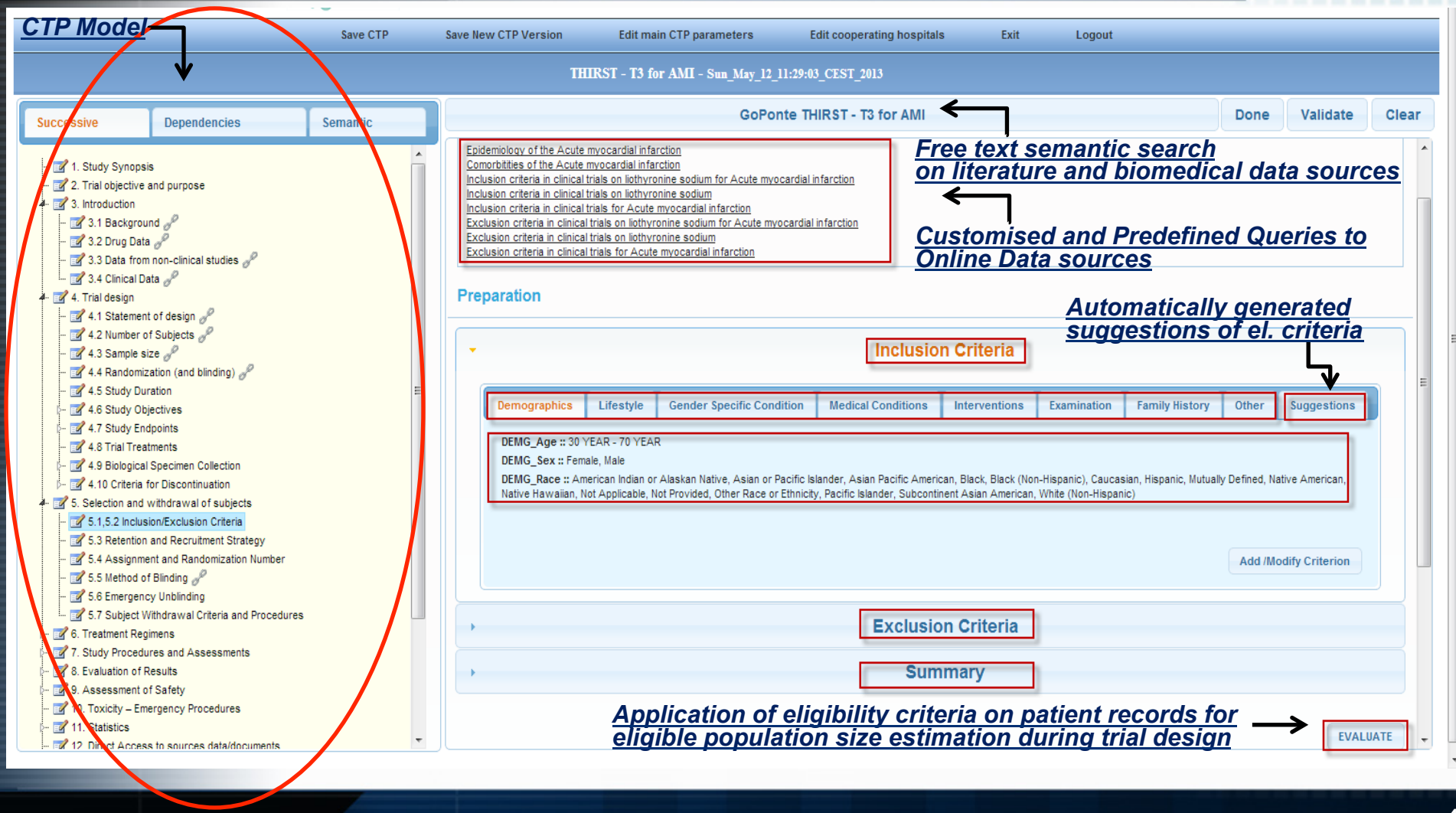
PONTE in a nutshell

PONTE aimed at providing a novel platform facilitating:

- the **generation** and **evaluation** of the **test of hypothesis** in the biomedical domain,
- the **design** of a drug repositioning **clinical trial** and offering
- **automatic pre-screening of potential study participants**

Towards this direction the platform exploits and extends *Semantic Web* technologies in order to offer *advanced decision support functionalities* to all the above aspects and to achievement *semantic interoperability* between clinical trials and healthcare patient records for patient recruitment purposes.

PONTE Interface



More at http://www.ponte-project.eu/

PONTE Data Infrastructure

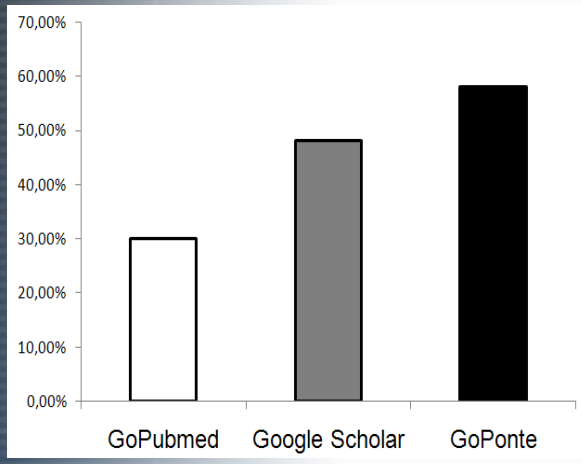
- Linked Data (incl. DrugBank, Diseasome, LinkedCT, KEGG)
- PubMed and ClinicalTrials.gov
- The Web
- Patient Records at Healthcare

GoPONTE

- Is the PONTE semantic search engine
- Uses Yahoo Boss! Search API for fetching web results and the MeSH, GO and UniProt ontologies for enriching the query and for indexing and annotating the results
- It searches across abstracts, semi-structured documents, web pages.

PONTE results

- 5 new ontologies (incl. Eligibility Criteria, CTP, Hypothesis, Patient Record, Domain Ontologys)
- 40 services (incl. information retrieval, automatic generation of research questions and eligible population size)
- GoPONTE, a semantic search engine with on-the-fly results' annotation
- 1 integrated platform offering the PONTE services
- The **THIRST study** on Thyroid Hormone Replacement therapy in patients with ST-Elevation Myocardial Infarction



Summary of the ratio of the relevant results retrieved compared to the full list of results retrieved

- **Zero effort** population estimation **during** trial design at the recruitment sites / Faster go/no-go decisions about sites
- Average *CTP preparation time* **in half** through direct literature linking, population size estimations and decision support
- Over **50%** of eligibility criteria of completed drug repositioning clinical trials were suggested by the platform
- More than **80%** of the CTP parameters were semantically linked with literature for direct semantic searches
- **84%** of eligibility criteria can be expressed through the EI. Criteria Ontology in a machine processable way

PONTE

Tagaris, A., Andronikou, V., Chondrogiannis, E., Tsatsaronis, G., Schroeder, M., Varvarigou, T. and Koutsouris, D. (2014), Exploiting Ontology based search and EHR Interoperability to facilitate Clinical Trial Design, In Dionysios-Dimitrios Koutsouris, Athina A. Lazakidou, Concepts and Trends in Healthcare Information Systems, Annals of Information Systems Volume 16, 2014, pp 21-4

Tagaris, A., Andronikou, V., Karanastasis, E., Chondrogiannis, E., Tsirmpas, H., Varvarigou, T. and Koutsouris (2014) PAT: An intelligent authoring tool for facilitating clinical trial design, 25th European Medical Informatics Conference - MIE2014.

Matskanis, N., Andronikou, V., Massonet, P. Mourtzoukos, K., Roumier, J (2012) A Linked Data Approach for Querying Heterogeneous Sources, KEOD 2012 in IC3K (International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management).

Andronikou, V., Xefteris, S., Varvarigou, T. (2012) A Novel, Algorithm metadata-aware Architecture for Biometric Systems, IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications, 2012.

Tsatsaronis, G., Mourtzoukos, K., Andronikou, V., Tagaris, T., Varlamis, I., Schroeder, M., Varvarigou, T., Koutsouris, D., Matskanis, N. (2012) PONTE: A Context-Aware Approach for Automated Clinical Trial Protocol Design, 38th International Conference on Very Large Databases, VLDB-2012.

Tagaris, A., Chondrogiannis, E., Andronikou, V., Tsatsaronis, G., Mourtzoukos, K., Roumier, J., Matskanis, N., Schroeder, M., Massonet, P., Varvarigou, T. (2012) Semantic Interoperability between Clinical Research and Healthcare: the PONTE approach, Semantic Interoperability in Medical Informatics (SIMI2012) Workshop co-located with the ESWC2012.

Roumier, J., Liebman, M., Tsatsaronis, G., Matskanis, N., Tagaris, A., Mourtzoukos, K., Chondrogiannis, E., Andronikou, V. , Massonet, P., Schroeder, M. (2012) Semantically-assisted Hypothesis Validation in Clinical Research, e-Challenges e2012 Conference.

Chondrogiannis, E., Andronikou, V., Mourtzoukos, K., Tagaris, T., Varvarigou, T. (2012) A novel Query Rewriting Mechanism for Semantically interlinking Clinical Research with Electronic Health Records, International Conference on Web Intelligence, Mining and Semantics, WIMS'12.

Chondrogiannis, T., Matskanis, N., Roumier, J., Andronikou, V., Massonet, Ph., (2011). Enabling semantic interlinking of medical data sources and EHRs for clinical research purposes, eChallenges e-2011 Conference.

Andronikou, V., Karanastasis, E., Chondrogiannis, T., Tserpes, K., Varvarigou, T. (2010). Semantically-enabled Intelligent Patient Recruitment in Clinical Trials. 5th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3GPCIC 2010), November 4-6, 2010, Fukuoka Institute of Technology, Fukuoka, Japan, IEEE Conference Publishing Services.

OpenScienceLink Landscape

No universal well-structured repositories of scientific and research data for **experimentation** and **benchmarking** of pertinent research works in a given thematic area

Poor linking of research with **data journals** and open access **datasets**

Waste on resources due to research duplication and limited access to data

Peer Review Processes remain **fragmented**, **lengthy**, **biased** and in several cases **weak** and **inefficient**

Reviewers are still equipped with only few tools and need to perform **time-consuming**, **incomplete** searches across **global literature**

Current **evaluation metrics and systems** do *not* fully reflect the actual **quality**, **novelty** and **impact** of the published work

The **dynamics** of the field, the research work and the researcher are *not* taken into consideration

Limited results of **science** in dealing with **great challenges**, such as poverty, climate change, unemployment, social exclusion, which aim at a healthy and productive population

The OpenScienceLink in a nutshell

OpenScienceLink exploits the open access trend and resources in the scientific world coupled with recent advances in the **social analytics** and **Semantic Web** for facilitating addressing the above problems, while also enabling a range of **new business models** for several stakeholders in the **scientific publishing** and **academic value chains**.

More at http://opensciencelink.eu/

OpenScienceLink Output

- **One integrated platform** incorporating **Semantic Web** and **social analytics** technologies for the provision of **5** (five) **pilot services**:

  #1: Data journals development based on the OpenScienceLink model for scientific datasets

  #2: Novel open, semantically-assisted **peer review process**

  #3: Research Trends Detection and Analysis

  #4: Dynamic **researchers' collaboration** based on non-declared, semantically-inferred relationships

  #5: Scientific field-aware, productivity- and impact-oriented enhanced **research evaluation services**

Current OpenScienceLink Results (1/2)

- **OpenScore**:
  - A new evaluation metric which uniquely incorporates scientific output aspects such as work volume, thematic breadth, career timeline and linking with scientific community
  - High correlation with existing evaluation metrics
- **Automatic Abbreviations expansion mechanism:**
  - It detects abbreviations along with their meaning regardless of whether their long form is provided in the document or not
  - 95% success rate in detection of true meaning of abbreviations in online biomedical documents
- (***ongoing***) **Trends detection:**
  - Temporal analysis of:
    - Social networks activity for detection of potentially new topics
    - Biomedical concepts across literature with particular focus on the ones of low but rising occurence

Current OpenScienceLink Results (2/2)

- **Collaborations suggestion service:**
  - Provision of recommendations which are relevant to the expert's topic/domain and are not part of his existing collaborations
  - Correctness of implicitly identified relationships among researchers: >= 60%
- **Biomedical Data Journal:**
  - An open access journal aiming to facilitate the presentation, validation, use, and re-use of datasets, with focus on publishing biomedical datasets that can serve as a source for simulation and computational modelling of diseases and biological processes
  - Implements the OpenScienceLink model for datasets and allows the publisher to exploit the platform trends detection and analysis services
  - Publisher: PROCON

OpenScienceLink

Karanastasis, E., Andronikou, V., Chondrogiannis, E., Tsatsaronis, G., Eisinger, D., Petrova, A. The OpenScienceLink architecture for novel services exploiting open access data in the biomedical domain, 18th Panhellenic Conference on Informatics, PCI2014, 02-04 October 2014.

Chondrogiannis, E., Andronikou, V., Karanastasis, E., Varvarigou, T.: An intelligent ontology alignment tool dealing with complicated mismatches. In: Proceedings of the 7th International Workshop on Semantic Web Applications and Tools for Life Sciences, Berlin, Germany (2014)

Chondrogiannis, E., Andronikou, V., Karanastasis, E., Varvarigou, T.: A Novel Framework for User-Friendly Ontology-Mediated Access to Relational Databases, World Academy of Science, Engineering and Technology (WASET) International Journal of Computer, Electrical, Automation, Control and Information Engineering. Vol. 9(3) (2015).

Chondrogiannis, E., Andronikou, V., Karanastasis, E., Varvarigou, T.: Meaning Inference of Abbreviations Appearing in Clinical Studies. In: Proceedings of the 2015 Symposium on Languages, Applications and Technologies (SLATE-15), Madrid, Spain (2015).

Chondrogiannis, E., Andronikou, V., Karanastasis, E., Varvarigou, T.: An Advanced Query and Result Rewriting Mechanism for Information Retrieval Purposes from RDF datasources. Accepted for publication in the International Conference on Knowledge Engineering and Semantic Web (KESW-2015), Moscow, Russia (2015). Paper will be published by Springer in CCIS 518 proceedings.

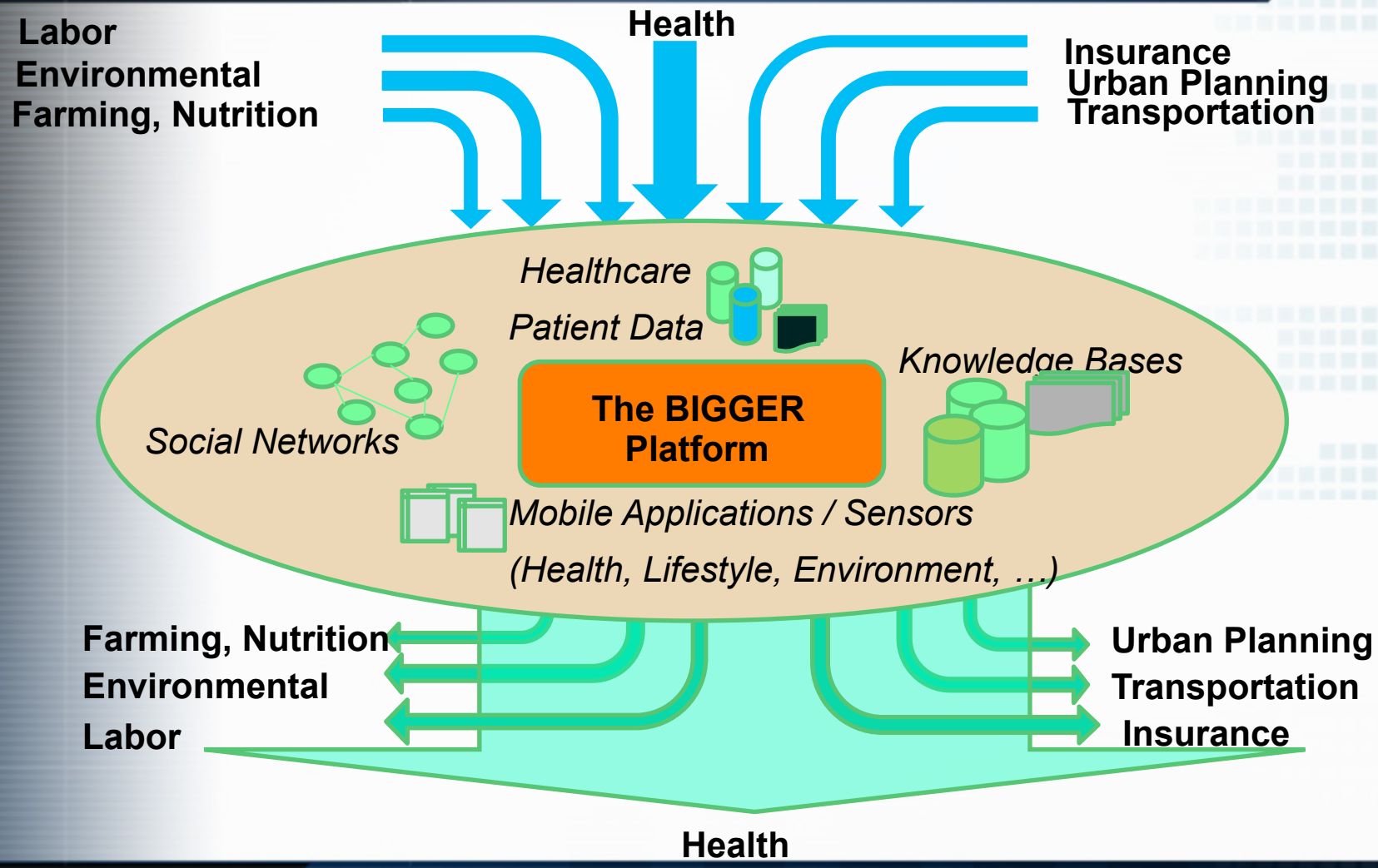
Chondrogiannis, E., Karanastasis, E., Andronikou, V., Varvarigou, T.: Building a Repository for Inferring the Meaning of Abbreviations used in Clinical Studies. Accepted for publication in the 2nd International Conference on Artificial Intelligence(ICOAI 2015), Amsterdam, Netherlands (2015). Paper will be published in Journal of Computers (JCP, ISSN: 1796-203X)

BIGGER: The main idea

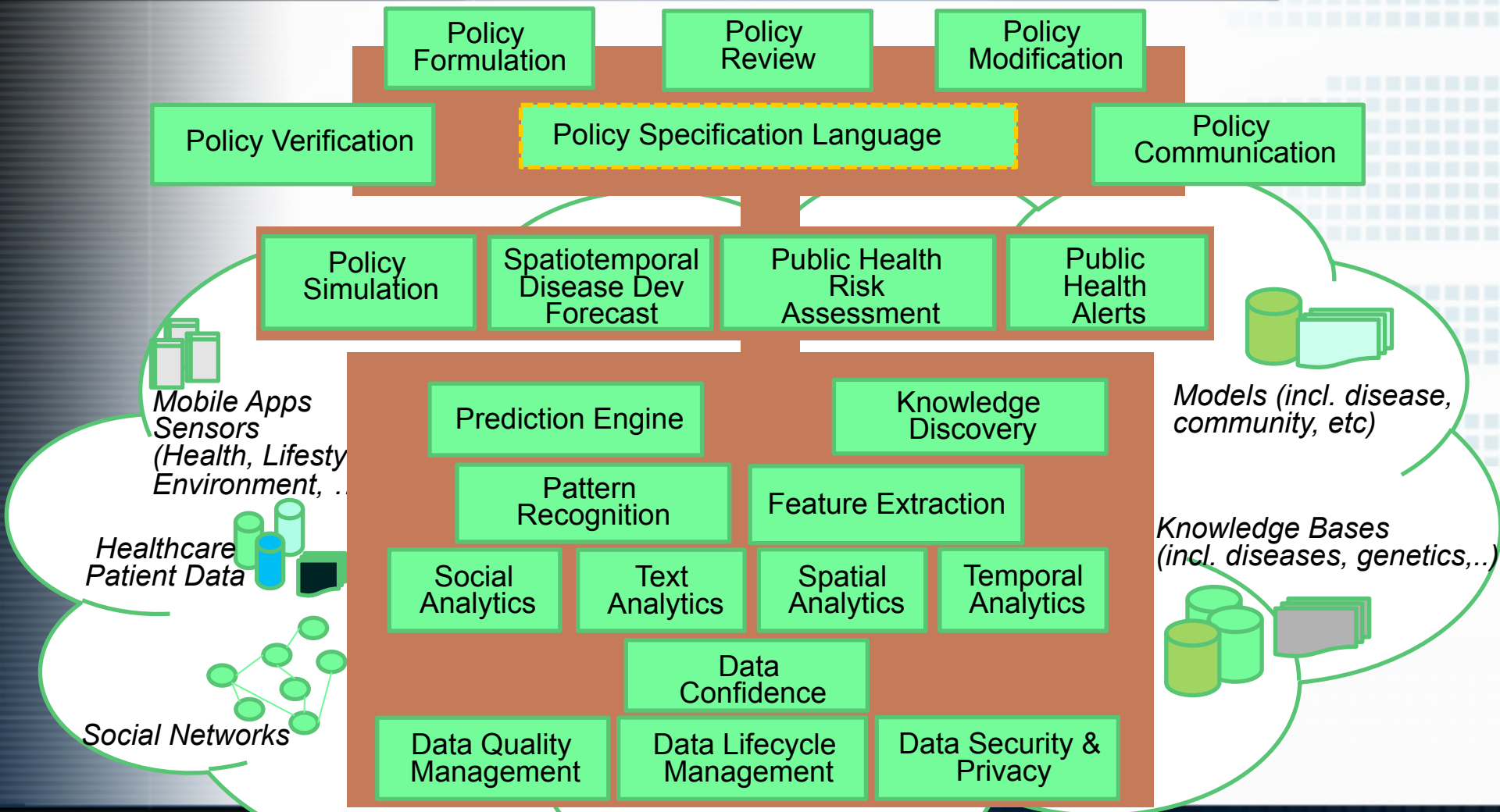
BIGGER aims at exploiting big data technologies applied on a variety of heterogeneous, content- and context-variable data sources for assessing the following question:

How to **incorporate public health issues** in **policies across the public sector** (such as transportation, urban planning, labor, insurance, education) in order to allow for **prompt response** in cases of infectious **diseases outbreaks** and **robust, effective policy making** in cases of **non-communicable diseases**?

The BIGGER vision



The BIGGER platform



BIGGER Pilots

Pilot #1: Short term public health policy making requiring prompt response. Driven by the continuously increasing prevalence and incidence of communicable diseases due to rising numbers of immigrants and the financial crisis, this pilot will focus on *an infectious disease*, and more specifically early identification of its signs and their estimated spatio-temporal spreading.

Pilot #2: Long term public health-driven cross-sectorial policy making on CVD for disease and disease worsening prevention. As preventing cardiovascular disease is not only about better medical treatment, but also about improving access to healthier foods and creating environments that encourage physical activity, this pilot will focus on policy development for CVD prevention and stabilisation.

Potential Collaboration

- To use anonymised, annotated (with contextual information) health and lifestyle datasets (both already collected and real-time generated ones) collected by the applications as part of the BIGGER Data Infrastructure
- To improve content annotation in collected data for the latter to serve the BIGGER purposes
- To launch a new “study” in ResearchKit for supporting the Pilot #2
- To develop a DREAM Challenge for researchers to compete in offering public sector policy models on top of or supplementing the ones to be incorporated in the BIGGER platform