

VIRTUAL RESEARCH ROOM PROJECT

i.e. Methodology of Big Data for Medical Data
Sets



Hungarian Hospital Association XXXI. Congress

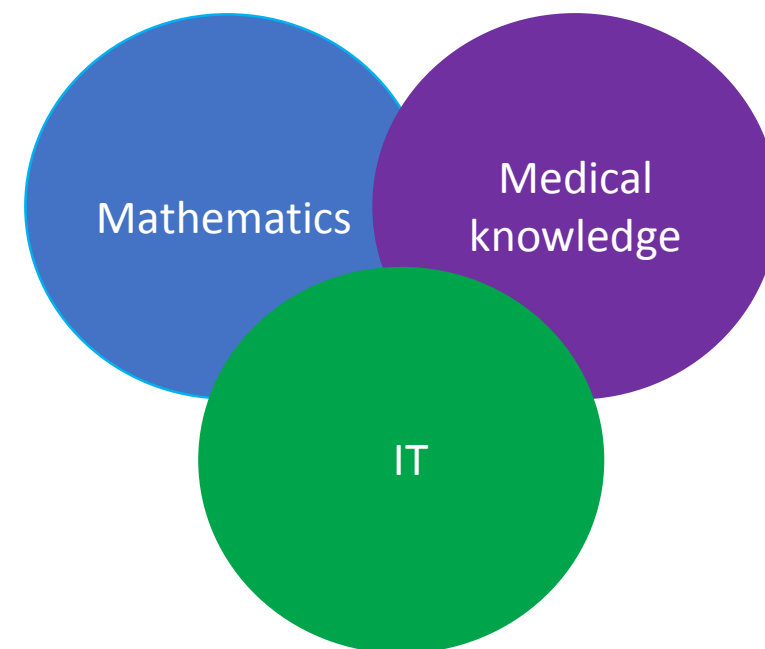
Eger, 11. April 2019

Gyula Király lead researcher, Hospitaly Ltd.



Project ID:	GINOP-2.1.1-15-2016-00898 R&D project
Duration of the project:	1 February 2018 - 31 May 2019
Subject of the project:	<p>„Protocol for secure management of medical professionals and patient life history data, creating an innovative patient management methodology through the development of an artificial intelligence-based prototype medical research room ”</p> <p>i.e.</p> <p><i>„ Designing a research room with the use of Big Data technology with the use of depersonalized health data from healthcare institutions to create, support or reject hypotheses.”</i></p>
IT background:	Server capacity - 440 seeds, 5 Tbyte RAM, 10 nodes, shared storage

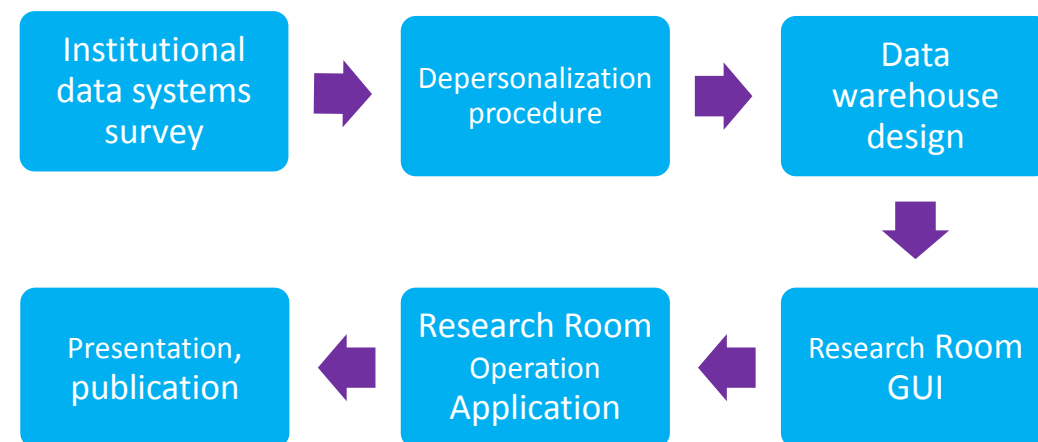
- The research task was based on the principal that an innovative solution on Big Data technology can only be recognized by a team of researchers and professionals who are able to communicate with each other on a greater scale.
- **This triple unit was composed of representatives of the medical, mathematical and IT fields.**
- To achieve a successful project, additional active participants of legal, economic and project management professionals are required.



- The team is composed of doctors with vast clinical experience, academic research mathematicians and IT professionals, who have developed a routine in building national health systems. They work together in different working groups.
- Data warehouse experts are reliable for creating and operating warehouse data.
- The project emphasized the importance of data security and protection, which is provided by the working group of legal compliance.
- A healthcare economist manages the development of the business model.

PROJECT ORGANIZATION

Vertical Working Group



Horizontal Working Group



➤ Clinical area

- supporting clinical hypotheses which exist and are internationally proven but not confirmed in Hungary
- supporting clinical hypotheses which are not confirmed on a large number of samples
- testing clinical hypotheses which have not been supported

➤ Public funding area

- investing the real value of DRG calculation
- examining impacts of quality care versus financial limits
- examining impacts on the practice of public funding versus the increase of waiting lists

➤ Quality control area

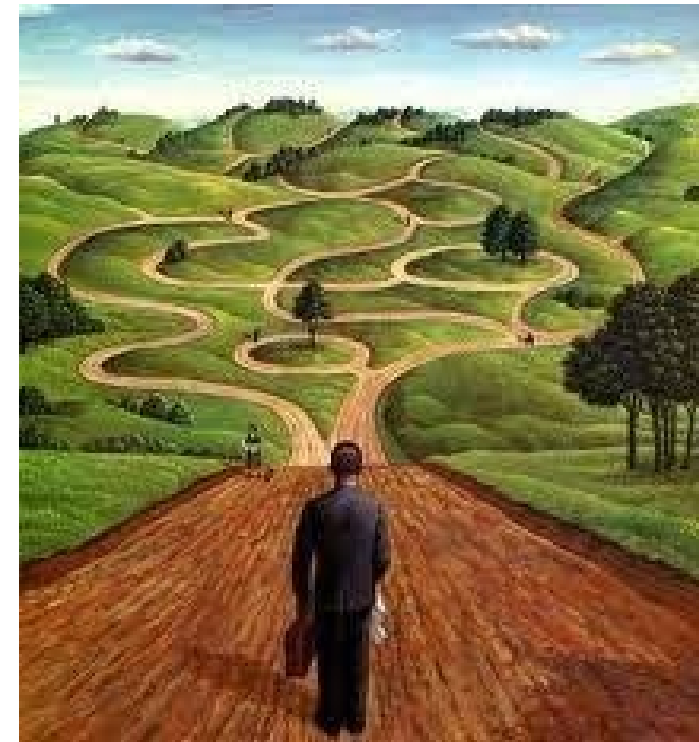
- examining inside protocols and routines
- controlling the scale of economies
- testing professional competences
- controlling the effectiveness of therapies, interventions and tests
- checking, confirming, rejecting professional protocols

➤ Public health area

- investigating lifestyle-related diseases
- examining efficacy of therapies influenced by lifestyle
- controlling risk indicators of the population

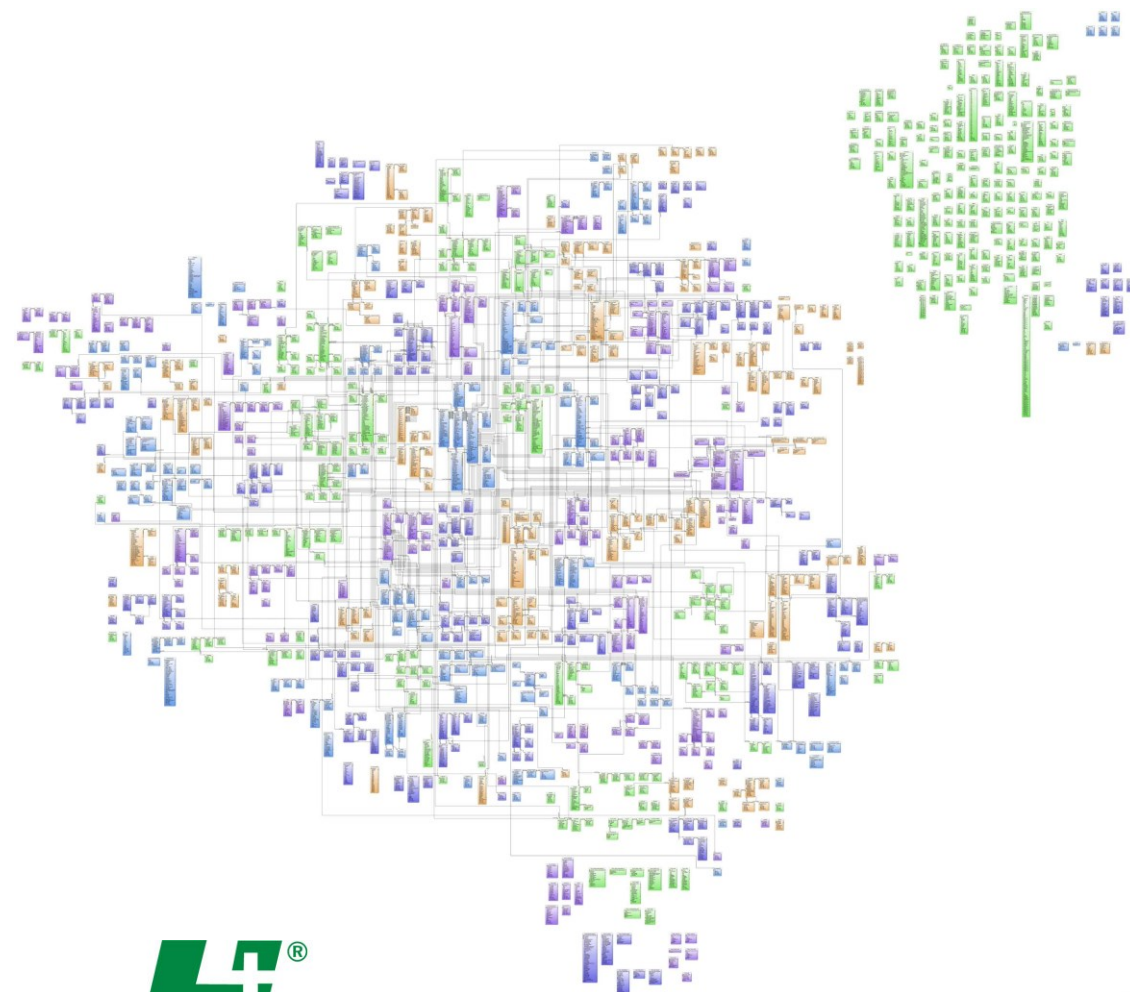
➤ Other area

- inspecting the distortion effect of informatics solution and targeted data collection
- analyzing development of data validation



Defining the scope of the health datasets

- The goal of the project is to implement a virtual research room where quantity and quality data meet the criteria for accepting or rejecting the hypotheses. Furthermore, appropriate technical infrastructure is needed from the medical industry.
- At first we assessed the database structure of a single, randomly selected institution's medical informatics system. After that we selected such data (codes, dates, numeric value, other classified data) for which we have the right tools for clear interpretation.



H⁺OSPITALY

medMátrix (HIS)

Replacing personal data with demographic data

- The social security number with its associated personal data are separated within the institute on the appropriate infrastructure.
- In the meantime, case ID's provide certain demographic characteristics which are required for research. These are supplemented directly or by data transformation.
- The research extends to involve the possibility of relevant anamnesis data.

MAJOR MILESTONES OF THE PROJECT

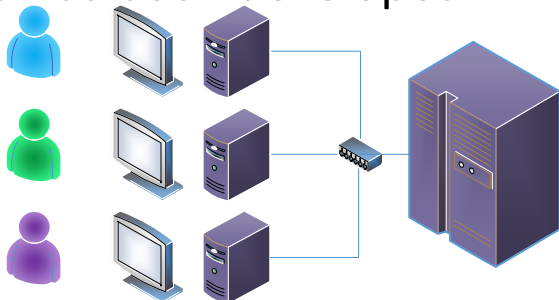
Replacing social security number with pathway ID

- During the research conducted in an institute, a safe, independent and irreversible patient pathway ID process and methodology was developed.
- Even within the institution, the social security number is transformed into a patient pathway ID using the closed, multi-factor cryptographic procedure defined in the methodol



Research data warehouse creation

- The collected depersonalized data, suitable for identifying a patient pathway is moved to the data warehouse.
- Data markets of clinical needs has been completed. The necessary environment for running mathematical functions and procedures have been prepared.
- The methodology for continuous data download has been developed.



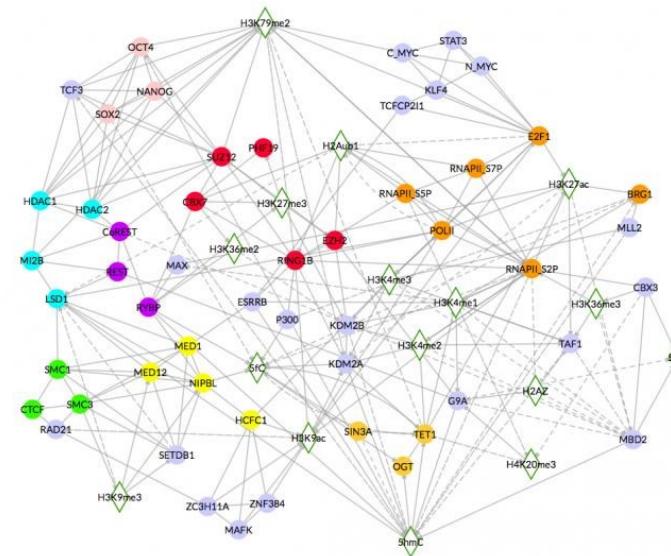
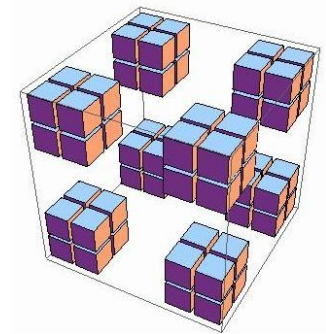
MAJOR MILESTONES OF THE PROJECT

Develop of operating model and publication rules

- institutional data asset register
 - settlement of used data sets accessed through research algorithm
 - possibility of contin conn
 - country and language independence
 - closed research administration
- Scientific findings are either individual or institutional competence, however publications are common results.



- The results and methodology allow to test large number of data sets for medical hypotheses. Furthermore, suspected but not yet examined relationships are supported by modern statistics and datamining which include: support vector machine, principal component analysis (PCA) and other machine learning algorithms.
- Based on our results, more reliable prediction in medical practice, reduced number of negative tests and faster diagnoses is possible.



Data Warehouse Preparation:

- ~ 600,000 patients, ~ 17 million cases, ~ 44 million relevant laboratory test results

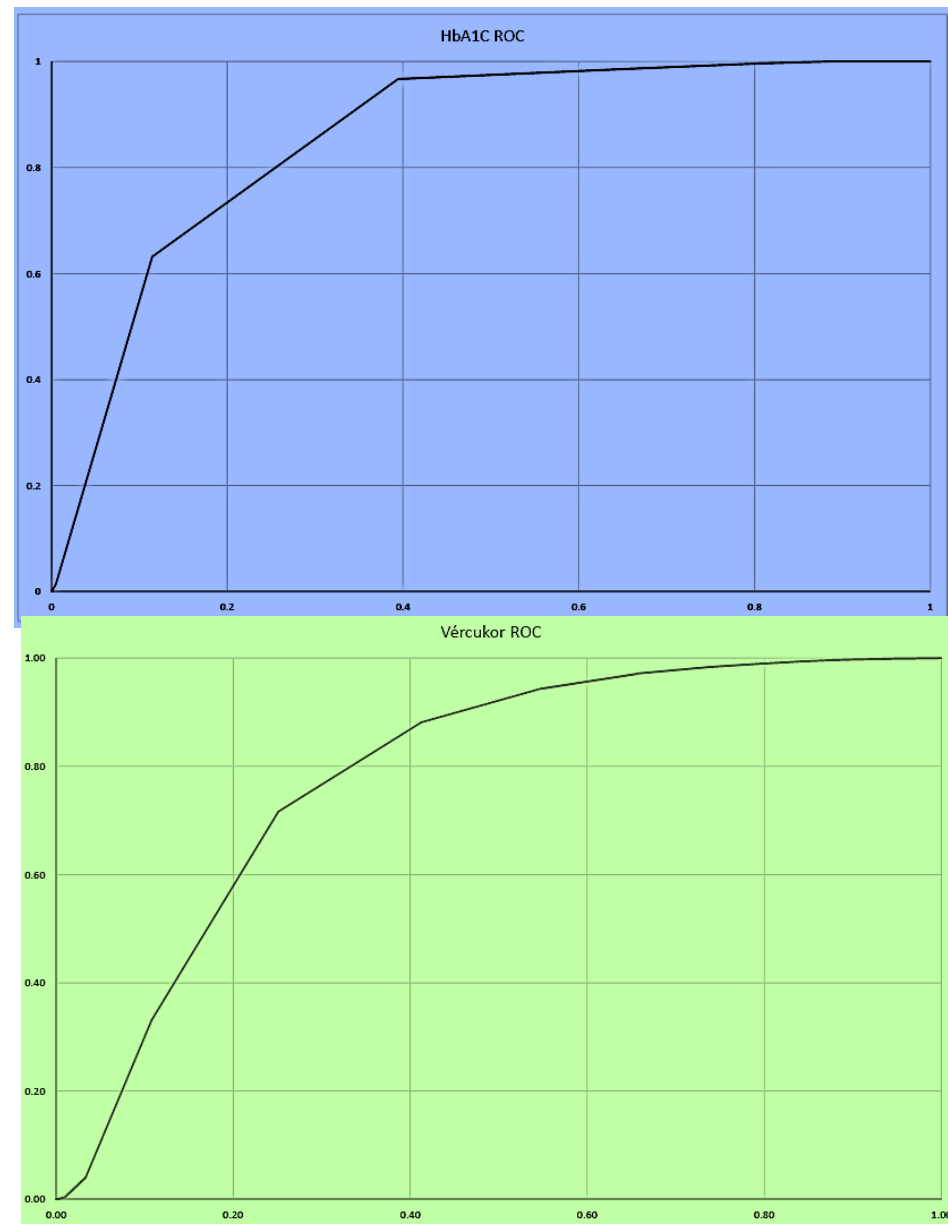
Mathematical suggestion:

- Based on HbA1C data set, it is easier to separate a diabetic from a non-diabetic person.
- AUC is 0.86, while blood sugar is 0.79

Mathematical question and **Clinical answer**:

- Is it possible that the relatively low AUC value caused some treated diabetic patients to have a normal blood sugar level? – **YES**
- Is it possible that some patients who are not diagnosed with diabetes are actually diabetic? – **YES**

A SIMPLE RESEARCH EXAMPLE



THANK YOU FOR YOUR ATTENTION!

