

e-SmartBIMS. Novel components to enable smart interoperability and analytics in Biobanks

Consorcio: Noray Bioinformatics; Ibermatica; VizzDAT; Karolinska Institutet

Tecnología: Tecnologías del Lenguaje

Descripción general:

SmartBIMS develops novel software products that will enable interoperability and exploitation of the Biobank data. The main components are: an automation tool to codify medical texts to SNOMED-CT, an Application Program Interface (API) to facilitate Biobank Big Data management and sharing and mobile APPs to allow the standardized data collection and visualization. All of them will be integrated in a commercial platform called SmartBIMS that will give answer to current technological market needs.

In SmartBIMS project, the following new technological products will be generated:

- 1) An innovative Kit of automatic coding of medical texts to SNOMED-CT. Both a tool and a methodology will be developed to support the automated coding of texts written in natural language to SNOMED-CT. It will include a coding system that, based on NLP and machine learning techniques, automates the coding of texts in SNOMED-CT. It will also allow the correction of texts in real time to guide the introduction of the correct terminology to encode the information of the system. A Validation Panel will show the quality and viability of the tagged information, so that Biobanks can validate, modify, correct and influence the new future coding of the system. These functionalities will be implemented in a web service to be easily integrated into other systems.
- 2) An analysis API adapted to the facilities of the Biobanks that will provide them with Big Data capabilities to exploit all the data (clinical, research) associated with different types of biological samples. It will also allow providing new analysis services to customers and academic departments.

3) A mobile APP suite that will make easy the experimental & clinical data collection and its visualization in a standardized way.

4) A novel web product called SmartBIMS (Smart - Biobanks Information Management System) that will communicate with the previous components to enable automatic coding in SNOMED-CT thus allowing the harmonization of data and interoperability with other Biobanks, health systems and researchers, and will provide standardized data analytics capabilities. It will also allow closing the research cycle in such a way that all the data generated by a research study can be associated in a consistent way with the original samples, allowing the reuse of data and knowledge. The Biobank may offer this new information in its portfolio to researchers, who will avoid duplication of efforts, thus reducing costs and time.

Ibermatica will develop the SNOMED-CD KIT, which will allow the automatic coding of medical texts to SNOMED-CT.

- Creation of the knowledge base and correction of errors
- Correct introduction of texts in natural language for clinical environments
- Analysis and coding of texts from biobanks to SNOMED-CT

Programa: Eurostars (CIIP-20202004)

Duración: 27 meses (2022 – 2022)

Presupuesto global proyecto: 807.876,00 €

Presupuesto Grupo Ayesa: 166.625,00 €

