The widespread use of patient Electronic Health Records in hospitals generates large volumes of data offering exciting opportunities for novel discoveries in medicine. Indeed, we consider these records for a secondary use that is to constitute patient cohorts useful to experiment complex biomedical hypotheses. For instance, these medical records are used to monitor the safety of drugs and to alert about drugs that are prescribed but should be withdrawn from the market.

Beside these patient records, data describing the current state of human knowledge about diseases, genetics and drug mechanisms has been recently published. This “state-of-the-art” data can be used in conjunction with medical records to help guiding knowledge discovery. However, medical records are encoded in a hospital-specific manner, making them hard to interconnect with state-of-the-art data. Our research project is to integrate medical records data from the Stanford Hospital (available in a database named INTREPID) with the state-of-the-art biological knowledge, using an appropriate informatics toolbox: the semantic web. Once combined, these two datasets will be used to gain new insight on why distinct populations react differently to drugs and, when possible, to explain the biological mechanisms that may explain these differences.