# Quality assurance, data curation and data preparation in Artificial Intelligence systems applied to medical imaging

*PhD Candidate:* Giuseppe Roberto BONFITTO

*Email:* g.bonfitto@studenti.unibs.it

**[XXXVIII Cycle]**

*Tutor:* Prof. Alberto SIGNORONI
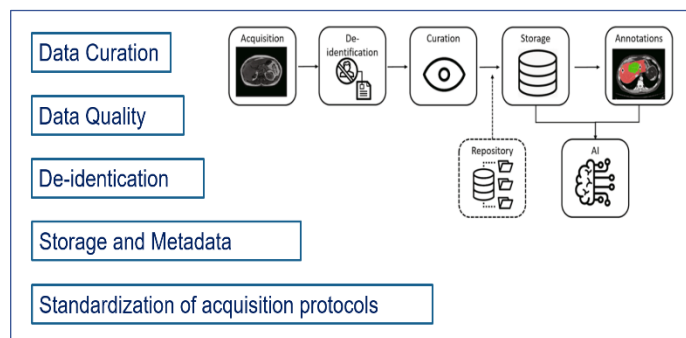
## Background

Data preparation and curation are critical to the success of artificial intelligence systems. To this purpose, data curation and data preparation lines should cover several key steps, including image acquisition in clinical centers, image de-identification to remove personal information and protect patient privacy to control the quality of image information, image storage, and management, and finally image annotation. Another important aspect concerns the size of the data sets. Medical datasets are typically smaller, in the hundreds or thousands. Therefore, the high quality of the same data set is necessary to be contrasted with the small size of the data sets. There are multiple definite challenges that could negatively affect the quality of a dataset. These challenges are bad data collection practices, missing or incomplete values, non-standardized inconvenient storage solutions, security, and privacy.

## Objectives

Provide guidelines for the care of radiological imaging data, from their acquisition and data curation to retention up to the definition of a standard for quality assurance for data, an essential element for their use in artificial intelligence systems.

## Methodologies

This project will develop on the study of data preparation and treatment issues for AI systems in medical imaging, on the data preparation evaluation scales, on the characteristics of the data sets already used, and finally on the imaging modalities, the anatomical districts, and pathologies mainly addressed by Artificial Intelligence software.



Data Curation

Data Quality

De-identication

Storage and Metadata

Standardization of acquisition protocols

## Expected Results and Impact

The expected result is to focus attention on the issue of Data Curation in the healthcare sector, a founding element of all datasets and AI software. It will be demonstrated that a correct standardization of data preparation will allow for greater effectiveness and validity of artificial intelligence systems in medical imaging.