Un-working data in the health tech sector: From Data Quality to Data Value
Natalia-Rozalia Avlona (University of Copenhagen)

Over the last few years, the intrusion of the Industry in digital healthcare with AI technologies and data-driven decision making has contributed to the expansion of the "data-market" into the trajectory of health and well-being through the intensive collection of data (Taylor, 2022; Sharon, 2018). In particular, the achievement of quality datasets became an important consideration in the development and deployment of algorithmic systems, since the quality of data feeding these systems seems to define their performance in an "garbage in, garbage out" fashion (Vidge & Derczynski 2020; Kazimzade & Miceli, 2020). This is particularly the case, when the AI powered systems are applied at "high at stakes" domains, such as the healthcare. Whilst data quality is not guaranteed by data quantity, the intensive data collection by the Tech Industry in general seems to correspond to metaphors such as "data is the new oil" (Neelie Croes, 2011) that exemplify a very particular understanding about data and its value. Academic scholarship has scrutinised not only the extractivist material relationships the data-oil and data-assets associations signify but has also showed that what becomes of value is not the data itself but its users (Birch et al 2021; Taffel, 2021; Thatcher, 2016;).

In this paper we identify that there is little research conducted to understand how the achievement of high-quality datasets become datasets of value for the health tech industry and market. We articulate the processes that transform the quality datasets achieved in two European companies in the heath-tech sector, to valuable datasets for a broader diversity of stakeholders. To do so, we study the life-circle of health data from its production towards quality standards to its trade as "valuable" assets. We particularly analyse the findings from the ethnographic research conducted in two European companies developing AI powered platforms systems for patients -to- clinical trials matchmaking and for the diagnosis of lung and pancreatic diseases. Our aim is to manifest how the data practices of experts (designers, computer scientists, managers and engineers) and data guardians in these settings lead to high-quality datasets, and under which processes and logics the high-quality datasets are transformed into valuable datasets outside the context of its production.