

A Media Suite for Journalists (Nirmal Roy, Ioannis Petros Samiotis, David Zhang, Soroush Rasti, Nikhil Bhuwakar, Mari Wigham, Philo van Kemenade, Erwin Verbruggen, Roeland Ordelman)

During ICT with Industry 2020, participants of the Sound and Vision use case explored strategies for finding relevant data collections that are not yet known to a professional user, such as a researcher or journalist, who is seeking information on a certain topic. To describe this process, we coined the term 'collection recommendation': given a journalist's need for information (query), which collections are most likely to help in answering it. Starting from the FAIR principles, the team developed a conceptual model for the stages in a collection recommendation system. The point of departure was our observation of a tendency for journalists to involve others in their quest for information, referred to as engaged journalism, open-source journalism, or citizen intelligence. This source of information was incorporated in our system using a crowdsourcing/niche-sourcing model. This model interacts with a multimedia collection miner and a collection analyser that extract useful information describing the collections. The analyser could use entity extraction techniques, potentially making it possible to interconnect data sources (linked journalism). Ultimately, the recommender engine in the last stages of the pipeline uses both crowdsourced and extracted information to recommend certain collections to the journalist. To produce a good recommendation, the recommender may use various techniques to improve the system's understanding of the information need of a user (e.g., relevance feedback, user profiling) and to perform relevant collection retrieval (e.g. diversification algorithms). On the basis of this collection recommendation model, the team listed a number of very interesting research questions on various levels. We are very positive about the prospect of using the model and questions to define the next research and development steps in the emerging field of data research and data journalism.