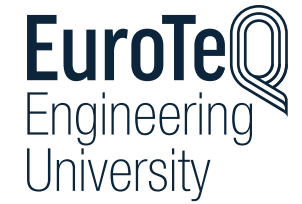# Challenge: Automatic data creation from written and audio sources I

Introduction

General description of the challenge of automating certain data collection recording using new application and programming.

Who is behind this initiative?

Ororo, a.s., with the collaboration of Charles University, Prague

Ororo company has been focused since 2011 on developing cutting-edge information systems that are tailored to large service and service companies. The emphasis is on efficient management, process optimization and innovation in the water sector. Ororo specializes in linking operational data using Internet of Things (IoT) with the operation and status of equipment or infrastructure. The company believes that linking these data with other information and operational events is key to effective management and decision-making.

What is the challenge?

Automatic creation of metadata from audio recordings of conversations in minority languages

Interested student teams are asked to engage in the above area by

- Aiming to develop a system for fully automated creation of descriptive records of individual audio recordings that will contain technical information (e.g. length, list and track of automatically recognized sections of the recording, date recording) and factual information (e.g. recognized conversation topics, keywords, locations mentioned, breakdown into sections).
- Optionally, the system could recognize the dialect used. The result will be research audio data with accurate metadata and ready to be stored in research data repositories. It will be possible to work with recordings more efficiently, to study a given language and language group. It is a great contribution to the promotion of research and preservation of minority languages**.**

Expectations and relevant considerations for the challenge / theme

Operation of a system that will provide a metadata service for audio files, which are not processed due to their size. These are recordings with a large amount of data that are not easily accessible to the users of the database. Linguistic and cultural value of diverse linguistic minorities.

Contact: jan.pokorny@ororo.cz