Data Quality Assurance

Quality control of the TCS AH episode data is carried out according to the diagram shown in figure below. It includes data processing steps and control points realized with a help of an on-line tool based on Redmine (www.redmine.org). The usage of the tool allows assignment and management of personnel responsible for the data integration, as well as tracking the progress of that integration. The uploaded data are prepared, checked and validated to ensure that they conform to commonly used standards and formats. Any inconsistencies in the data structure are identified by the quality control team. Once the data and metadata are approved, they are made available on the IS-EPOS Platform test portal and made available for the final quality and integrity check, as well as for an additional check done by the data owner. After all the checks, the data is published on the IS-EPOS Platform.

blocked URL

Quality Control Workflow of the AH Episode Access Service.
Orlecka-Sikora, B., Lasocki, S., Kocot, J. et al. (2020) An open data infrastructure for the study of anthropogenic hazards linked to georesource exploitation., Sci Data 7, 89, doi: 10.1038/s41597-020-0429-3.

Detailed steps of the data publication and Quality Control process:

- **Step 1.** Transfer of episode data to the local data center - raw data is uploaded to proper 'buffer' subdirectory by the data provider. Each data provider has access to a dedicated buffer directory. 'Administrator' for this episode and his 'Control group' in EAC are assigned.
- **Checkpoint 1.** Administrator sets episode as a new task in task management system. The Control group roles are distributed and the workflow Observer is appointed.
- **Step 2.** Data standardization and validation of formats. Data is verified, converted (if needed) and homogenized by people assigned by Administrator.
- **Checkpoint 2.** The completeness and quality of prepared data is checked
- **Step 3.** Metadata preparation according to TCS AH metadata scheme. All data files are described with sets of metadata prepared according to the rules.
- **Checkpoint 3.** The metadata is checked
- **Step 4.** Contextual analysis of data quality followed by appearance of episode in TCS AH maintenance area. Correct data is placed in final directories according to episode data structure. Now data is visible to all users who have permission to see this episode. Episode can be published to the test instance of IS-EPOS Platform if needed.
- **Checkpoint 4.** Administrator checks metadata sets and accepts episode as correct
- **Step 5.** Episode publication at TCS AH website.