3.2 Project Overview

If OCR4all is opened in the browser, the user is automatically taken to the "Project Overview" start page:

- " **Settings** ": The "Settings" option is used to select the work that is to be edited in the following. To do this, the desired work is selected from a drop-down list under "Project Selection", which was previously `ocr4all/data/Werktitel` created as a folder under (see 1.2). In addition, "Gray" is selected under "Project image type".
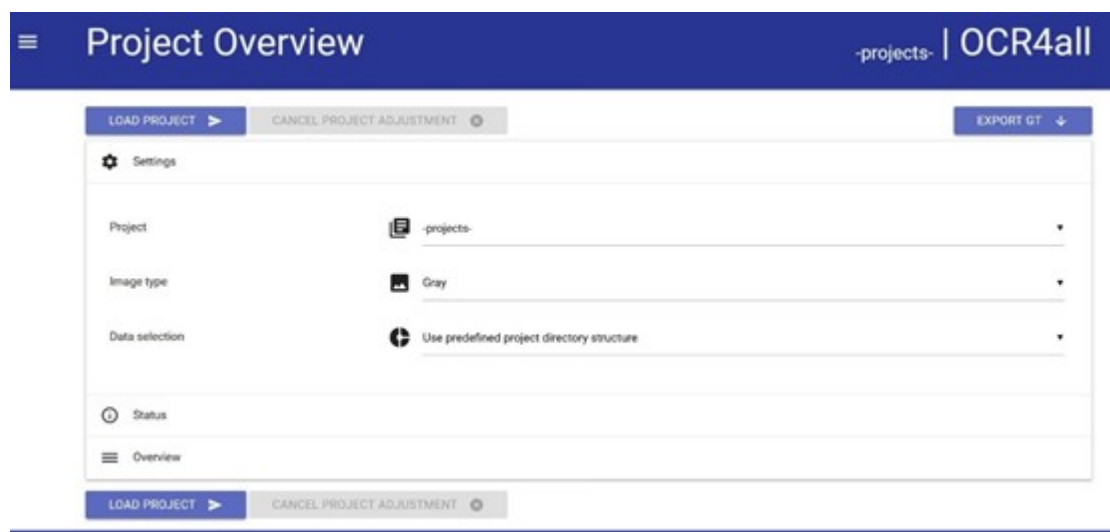

Fig. 2: Settings in the Project Overview.

- After the work has been selected, the work is loaded using the "LOAD PROJECT" button! Since OCR4all requires specific, fixed file names and formats (0001.png etc.), the LOAD function can be followed by a conversion request. This can be done directly in OCR4all.

- It is irrelevant whether the 'input' folder contains a PDF or individual images. If possible, however, individual images are usually preferable.
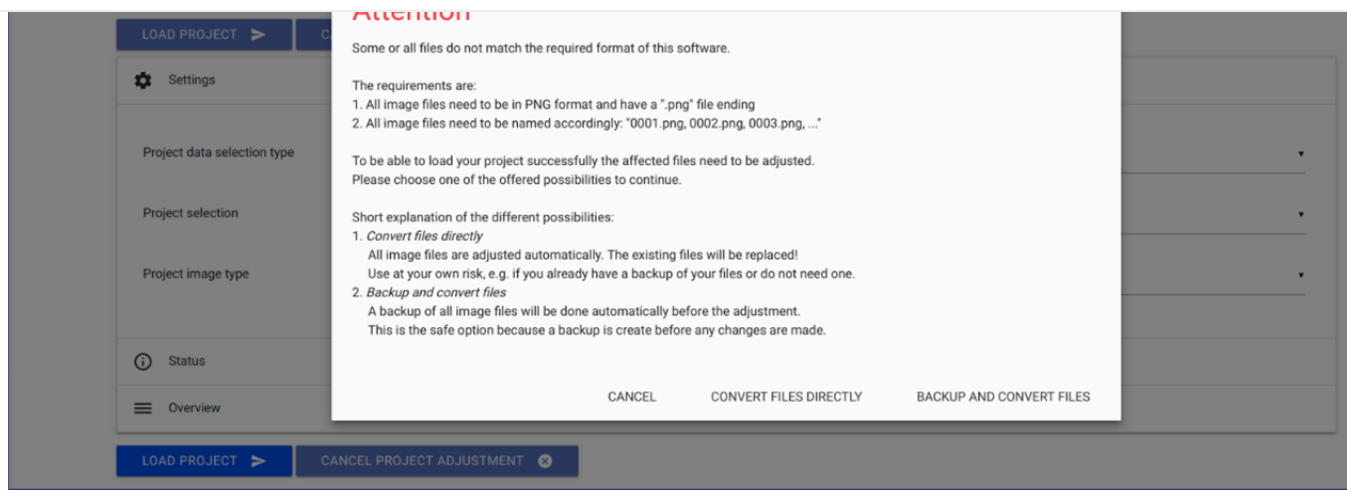
Fig. 3: Possible data conversion. (Here a PDF in the 'input' folder)

- In OCR4all, all data generated during the workflow and for its functioning is saved in a single PageXML file for each scanned page and is no longer kept in the form of many individual files. If project data from earlier versions are still available, it is now possible to automatically convert the project to the structure required by OCR4all. Fig. 4: Possible conversion of a legacy project

- **Under the " Overview " option** , the current processing status of the loaded work can now be viewed at any time, arranged in a table. Each page of this work is given its own line as a so-called "Page Identifier", the following columns form the typical and suggested OCR4all workflow from left to right. If certain work steps of the overall workflow have been carried out, the corresponding pages are marked as processed (check mark) in the "Overview". The display itself can be changed or particularly comprehensive works can be searched through using the options "Show... entries" and "Search".

Page Identifier	Preprocessing	Noise Removal	Segmentation	Region Extraction	Line Segmentation	Recognition
0001	✓	✗	✓	✗	✗	✗
0002	✓	✗	✓	✗	✗	✗
0003	✓	✗	✓	✗	✗	✗
0004	✓	✗	✓	✗	✗	✗
0005	✓	✗	✓	✗	✗	✗
0006	✗	✗	✗	✗	✗	✗
0007	✗	✗	✗	✗	✗	✗
0008	✗	✗	✗	✗	✗	✗
0009	✗	✗	✗	✗	✗	✗
0010	✗	✗	✗	✗	✗	✗

Showing 1 to 10 of 13 entries

Previous 1 2 Next

- By clicking on individual "**page identifiers**", the processing states of individual scanned pages and the data already available for them can be viewed throughout the entire work process. The "Images" column and the associated options "Original", "Binary", "Gray" and "Noise Removal" are used for this purpose.
- With the "Export GT" button (top right) all data generated during the project can be exported and packed as a zip folder within 'data'.