

# PUBLISH FIRST

A novel, rapid digitisation/publishing workflow for herbaria low on staff time and technical resources

26 - 28 SEPT 2022: GBIF  CESP WORKSHOP



STEPS - 1: SCAN

2: Upload



(See: <https://min.io>)



A python app listens to create events on the bucket



First 10k images p/m: €0

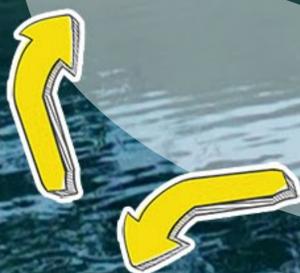


Google Vision API

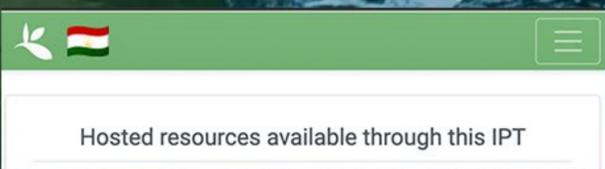
parses handwritten and printed text in image - some mistakes but generally good, even with handwritten cyrillic



Google Translate API for the Russian translation (original and translation are stored verbatim in dwc:dynamicProperties)



A simple parser splits out obvious label information (scientific name, altitude, etc). This is output into a dwc stored on the bucket



IPT set up to publish automatically from source file on bucket