

Research Assistant Vacancies in the Land Economics Group:

OCR & Historical Data Digitization

Project: HARVEST – Historical Agricultural Records as Virtualized Evidence on Structural Transformation

Location: University of Bonn

Duration: Up to 12 months (flexible)

Working hours: 8 –16 hours/week (depending on position)

Number of positions: 2–3

1. Project Overview

HARVEST is a research project aimed at constructing a comprehensive, long-run dataset of agricultural productivity by integrating modern statistical sources (e.g., FAOSTAT) with newly digitized historical records dating back to 1900. A central component of the project is the large-scale digitization of historical agricultural documents, including statistical yearbooks, colonial reports, and archival publications. These materials are processed using modern document image analysis and OCR pipelines, followed by structured validation and harmonization.

2. Objectives of the Position

The Research Assistants will support the digitization and data construction pipeline, contributing to:

- Extraction of structured data from historical documents
- Implementation and monitoring of OCR workflows
- Validation and cleaning of extracted datasets
- Construction of harmonized and reproducible data outputs

3. Key Responsibilities

Digitization & OCR:

- Pre-processing scanned documents (deskewing, denoising, contrast enhancement)
- Running OCR pipelines on historical tables and reports
- Supporting layout detection and table extraction workflows

Data Processing & Validation:

- Cleaning OCR outputs (numeric alignment, formatting, error correction)
- Identifying anomalies and inconsistencies in extracted data
- Supporting validation using internal consistency checks

Data Harmonization:

- Assisting in unit conversion (e.g. acres/hectares, tons/pounds)

- Supporting construction of commodity and geographic concordances
- Organizing structured datasets and metadata

General Research Support:

- Managing datasets and documentation
- Assisting with reproducibility (scripts, version control)
- Supporting preparation of research outputs (tables, figures)

4. Required Qualifications

- Currently enrolled as a Bachelor's or Master's student in Computer Science, Data Science, Economics, Statistics, or related fields at the University of Bonn
- Basic programming skills (preferably Python)
- Strong attention to detail and data accuracy
- Ability to work independently and manage tasks

5. Preferred Qualifications

- Experience with OCR tools
- Experience with document image processing or computer vision
- Experience with data cleaning and manipulation
- Familiarity with version control (Git)
- Interest in economic history, development economics, or data construction

6. Learning Opportunities

This position offers hands-on experience with:

- OCR and document digitization pipelines
- Large-scale data construction and harmonization
- Reproducible research workflows
- Applications of AI in economic research

7. Supervision and Work Structure

- Direct supervision by the Principal Investigator
- Structured tasks with clear milestones
- Flexible working arrangements
- Integration into a collaborative research workflow

8. Compensation

Compensation follows standard student assistant rates at the University of Bonn.

9. Application Process

Please submit:

- Short CV
- Brief statement of interest (max. 1 page), including relevant skills, experience, and availability

Applications should be sent out to: a.schioppa@ilr.uni-bonn.de