

Digitising the longest-running terrestrial insect database in the world

Gijs M. Gerrits & Roel van Klink



Background

In 1959 a field study was started to monitor abundances of ground beetles (*Carabidae*) by means of standardised pitfall trapping (see Fig. 1). The project still runs today with close to **1 million specimens** collected. **FAIR publication** of historical data on invertebrates is of high importance. The data can serve as baseline to study spatiotemporal population and community dynamics of insects and other groups.



Figure 1. Pitfalls were placed in various habitat types near Wijster, in the northern part of the Netherlands.



Figure 2. Standardising and checking a dataset that has been collected over the course of decades, involves a lot of detective work.

Objectives

- Collate and digitise the first part of the raw data
- Create reference dataset with **error rate <0.1%**
- Publish as **Darwin Core archive** with **Humboldt extension** on GBIF
- Explore the possibilities of automation with the use of AI algorithms

Workflow

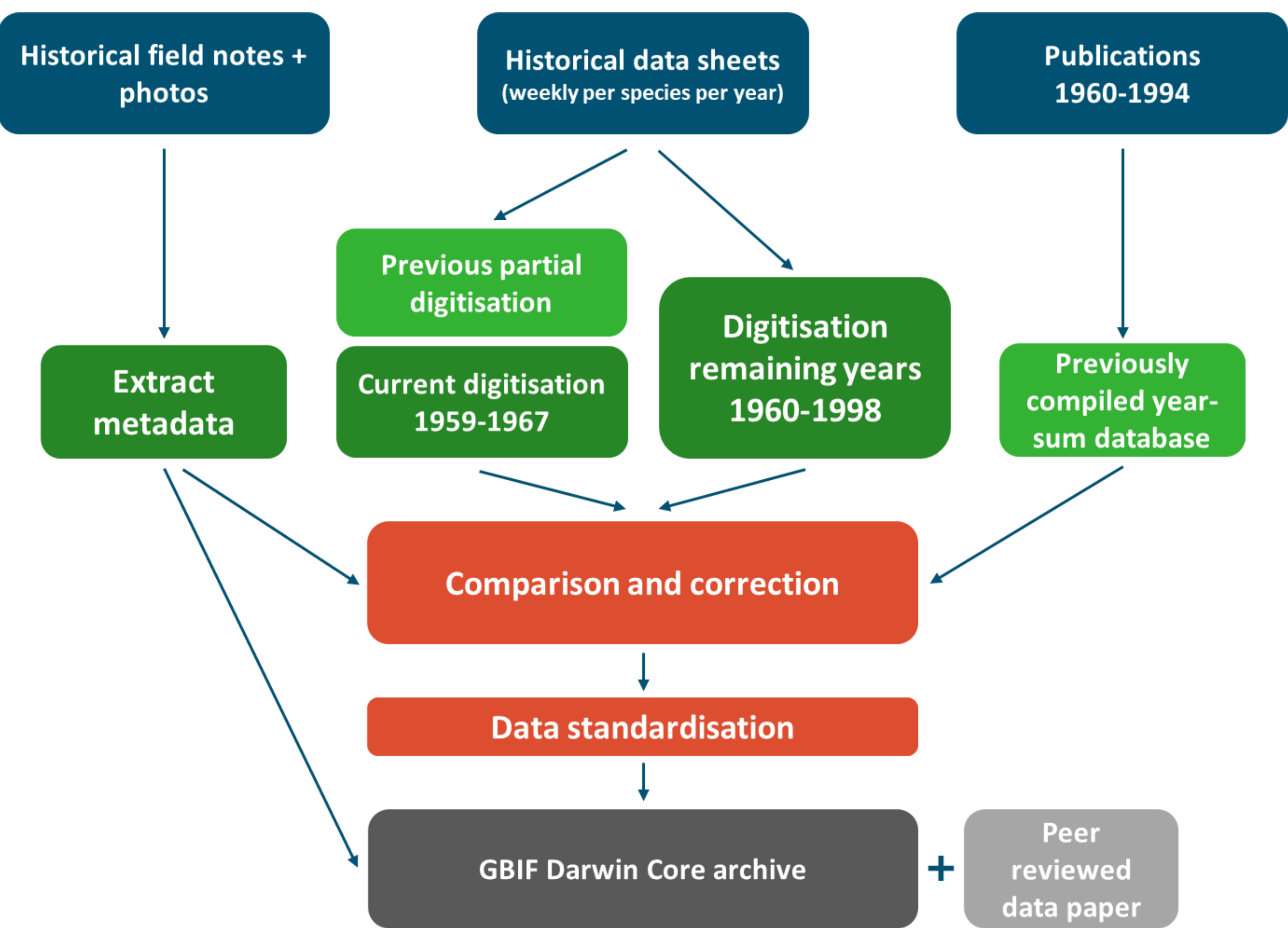


Figure 3. Schematic representation of the workflow; available raw data sources provide the input for collation, digitisation, correction and integration of all the data, which are published on GBIF and accompanied by a datapaper.

Sources and hierarchy of errors

(A)

Pitfall catches

Field/Lab notes

Year sheet

Digitisation

Data standardisation

Pterostichus nigrita

n = 373

(B)

Error type	Severity	Prevention/correction
species mix-up	very bad	check with year-sum database
wrong year	very bad	enter data per year, check with year sums in year-sum database
wrong pitfall	bad	check with row and column totals
date	less problematic	check with row and column totals
n individuals	less problematic	check with row and column totals
type-o in species name	fixable	check standard list of taxon names

Figure 4. In every step of processing the data, errors can and do occur; fig. 3a on the left, shows the steps from pitfall catch to the publication of the number of carabid species *P. nigrita* caught in 1961. Fig. 3b on the right shows how errors are not equal in severity and necessitate different approaches in preventing, or correcting them.

Year-sheet

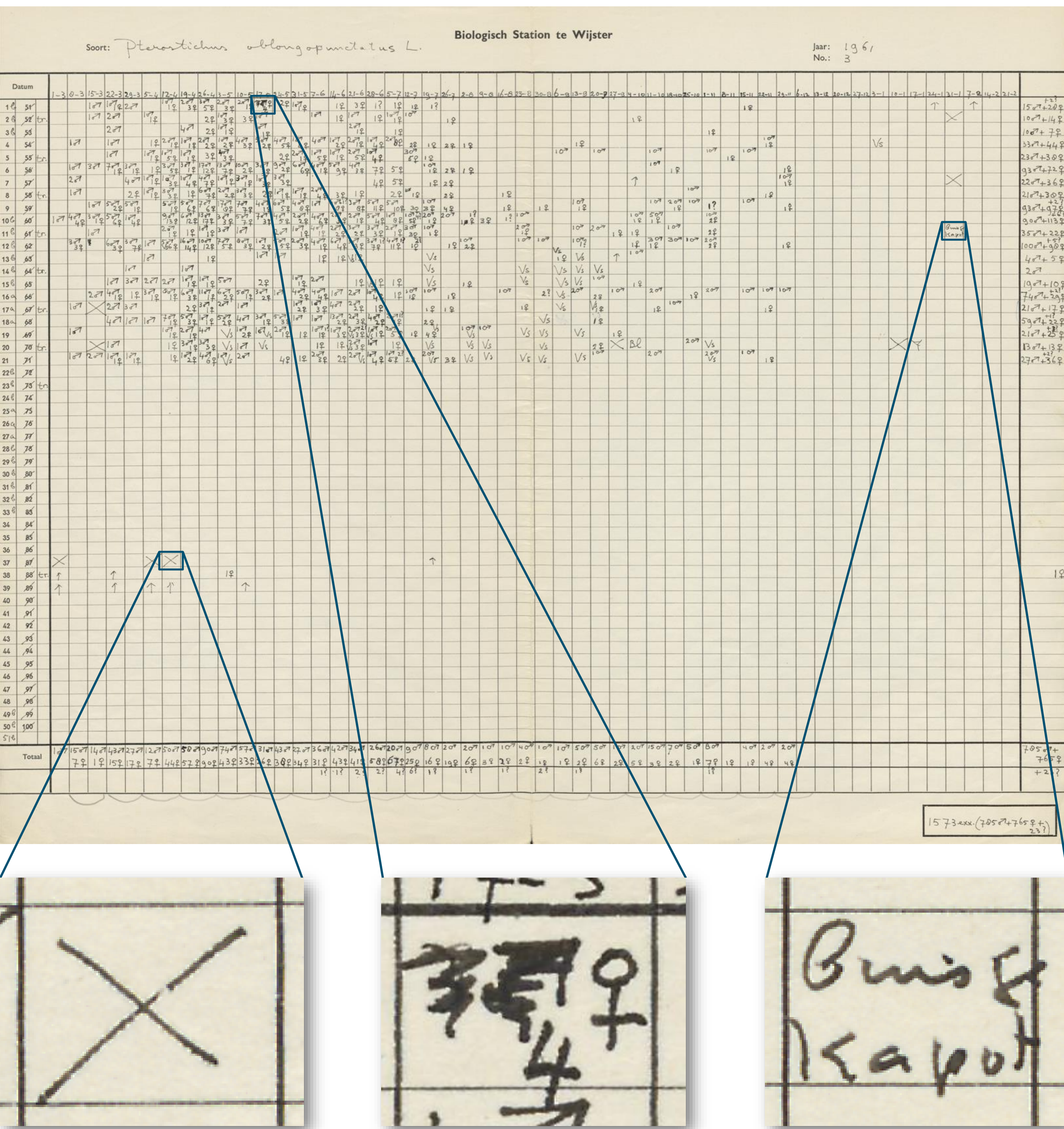


Figure 5. Example of a year-sheet for ground beetle species *Pterostichus oblongopunctatus*, with pitfalls in rows and sample dates in columns. The cut-outs of cells show entries that may provide challenges to automating the digitisation process.

Acknowledgements

We wish to thank Rikjan Vermeulen, Alje Woldering, Willem Jongbloed and Bjorn de Roos (WBBS) for continuing the time series to the present day. We thank Henk de Vries (WBBS), Karien Lahaise and Menno Hooft (Naturalis) for access to the historical records, and Hans Turin (Loopkeverstichting) for the year-sum database. We thank Lia Hemerik (WUR) for useful comments on the poster. This work is financed by NLBIF. The earlier digitisation effort was financed by the Prins Bernhard Cultuur Fonds