

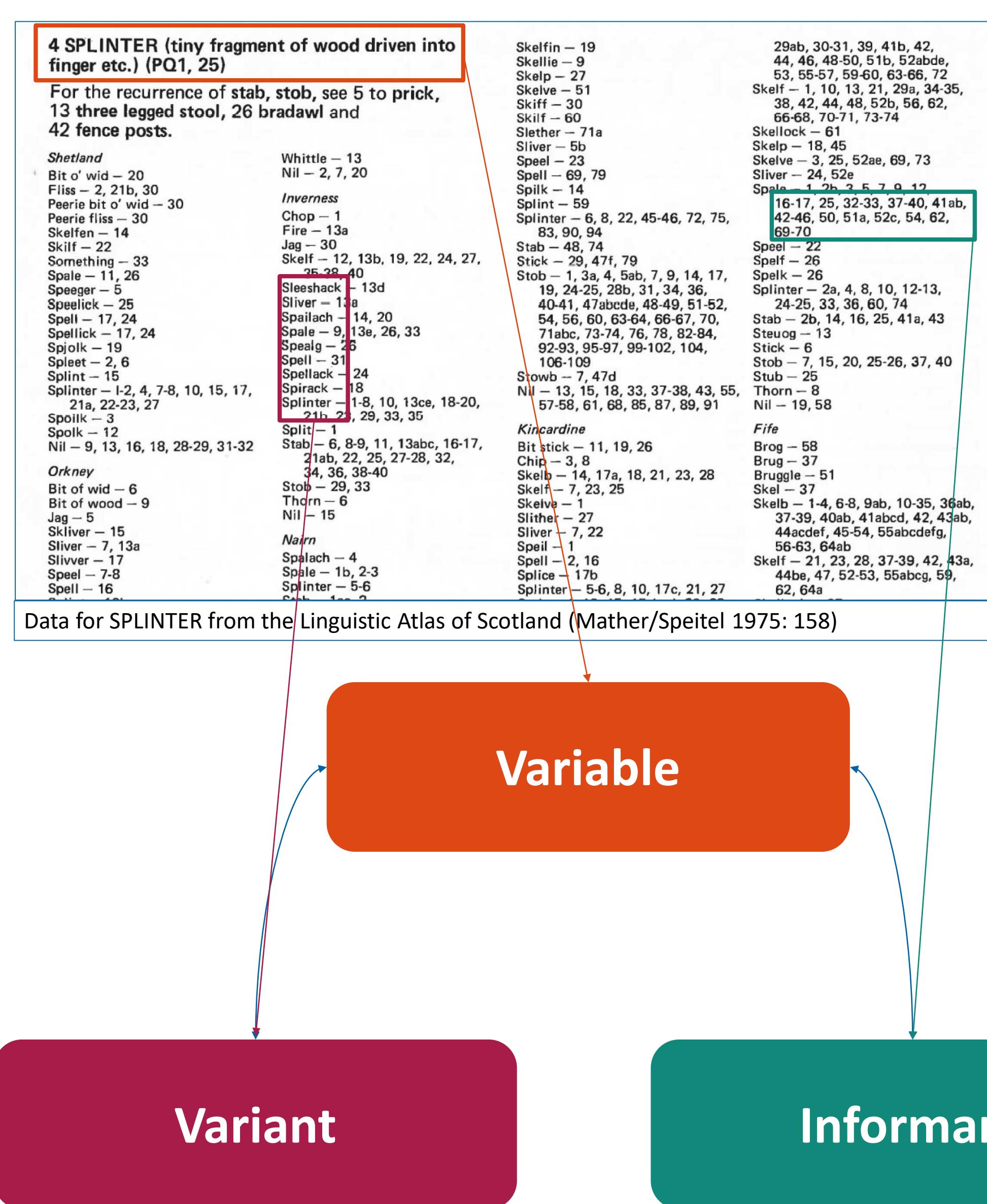
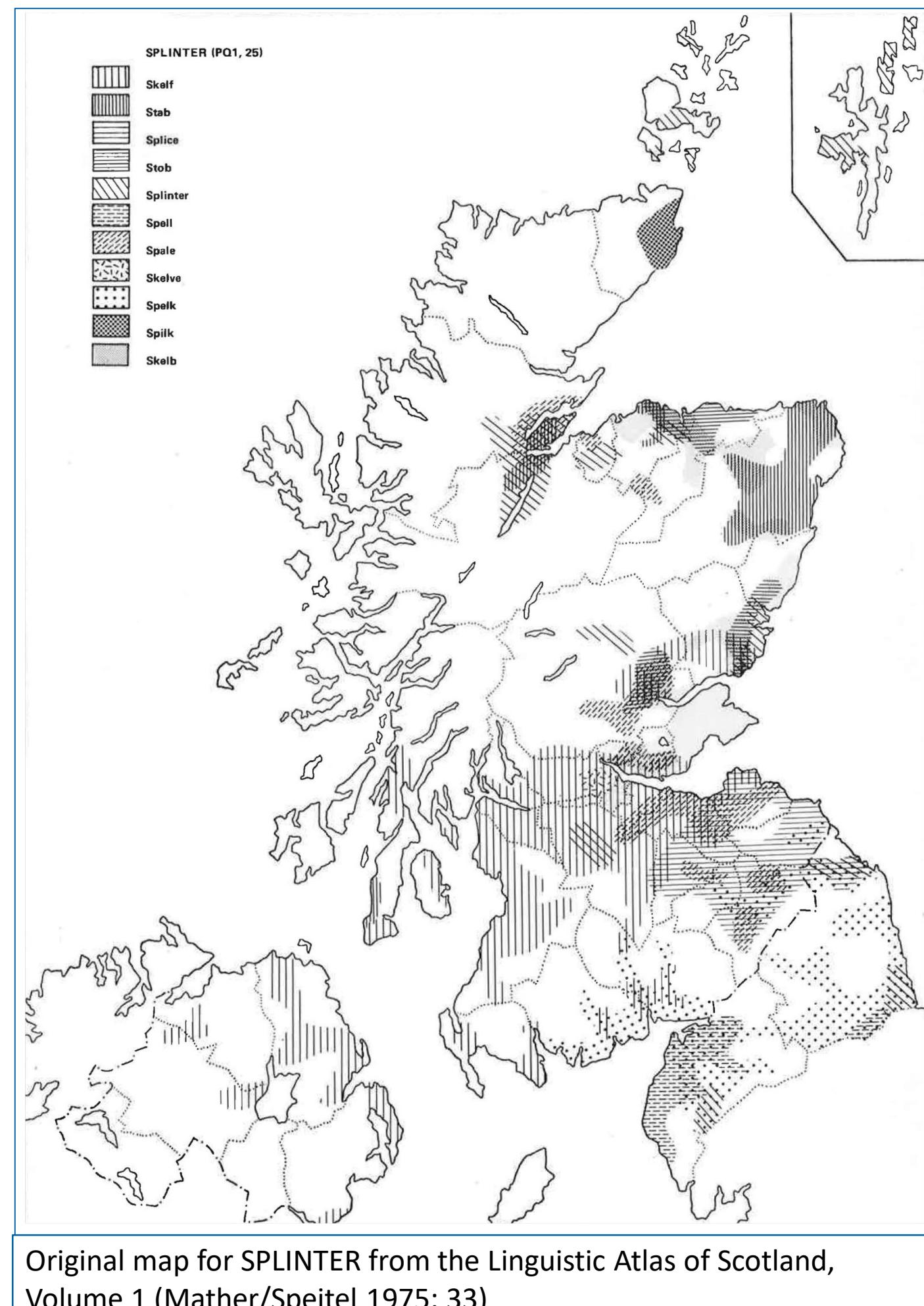
How Digitizing Language Data can Lead to New Analyses

Mapping the (Digital) Linguistic Atlas of Scotland

Markus Pluschkovits & John Kirk

The Linguistic Atlas of Scotland

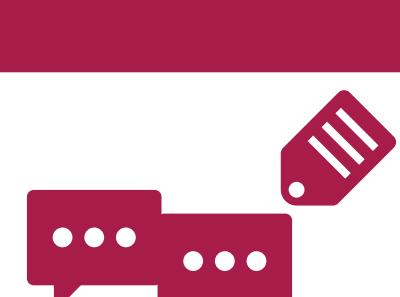
The Linguistic Atlas of Scotland was an atlas project carried out by Mather & Speitel, with the first paper questionnaires being sent out in the 1950s until the eventual publication of the individual volumes throughout the 1970s and 1980s. We are currently in the process of digitizing the lexical section of the Atlas, which spans over 200,000 responses to 153 lexical variables from over 1700 localities from the 33 pre-1975 counties of Scotland, seven counties in Ulster, and the two northernmost counties of England. These data allow for a re-analysis of a great many lexical, semantic and geographical presentations.



The Digitization Process

- OCR was employed to import list of informants into the database
 - From informants, relevant localities were created
 - Utilized Nominatim to find coordinates for localities
 - Localities which could not be found automatically were corrected by student assistants, using a dedicated interface which allowed for customized Nominatim queries (to account for spelling variation/errors in OCR). This user interface also allowed for older ordnance maps as base maps, from which coordinates could be selected (for those localities which no longer exist under the same name)
 - Afterwards, response data was imported into the database using OCR
 - We employed customized OCR to identify the relevant entities during the OCR process (Variable, Variant and Informant) by utilizing the formatting and page layout of the printed material
 - All OCR data was then manually checked by student assistants
 - Any piece of data therefore consists of a variant to a variable produced by an informant (who is connected to a locality)
 - New lexemizations (i.e. the reanalysis of the material) are conducted by associating variants with variant groups (the ‘lexemes’)
 - Therefore possible to map original material as well as reanalysed material
 - Tech Stack:
 - PostgreSQL database

Variants



- Variants are saved to the variable to make disambiguation among polysemous variants possible
 - Multiple variants are grouped together in variant groups, a process we call ‘lexemization’.
 - Lexemization is the basis for reanalysis of the material (see e.g. Kirk/Hessel 2022)

The Digital Linguistic Atlas of Scotland

The Digital Linguistic Atlas of Scotland switched from area maps to more accurate point maps, which also allow to group the data by counties for a quick overview. Additionally, different base maps allow for correlation between the distribution of variants and topographical features. Every point on the map corresponds to a locality, which consists of one or more informants. Clicking on a point indicates the name of the locality, as well as the number of informants.

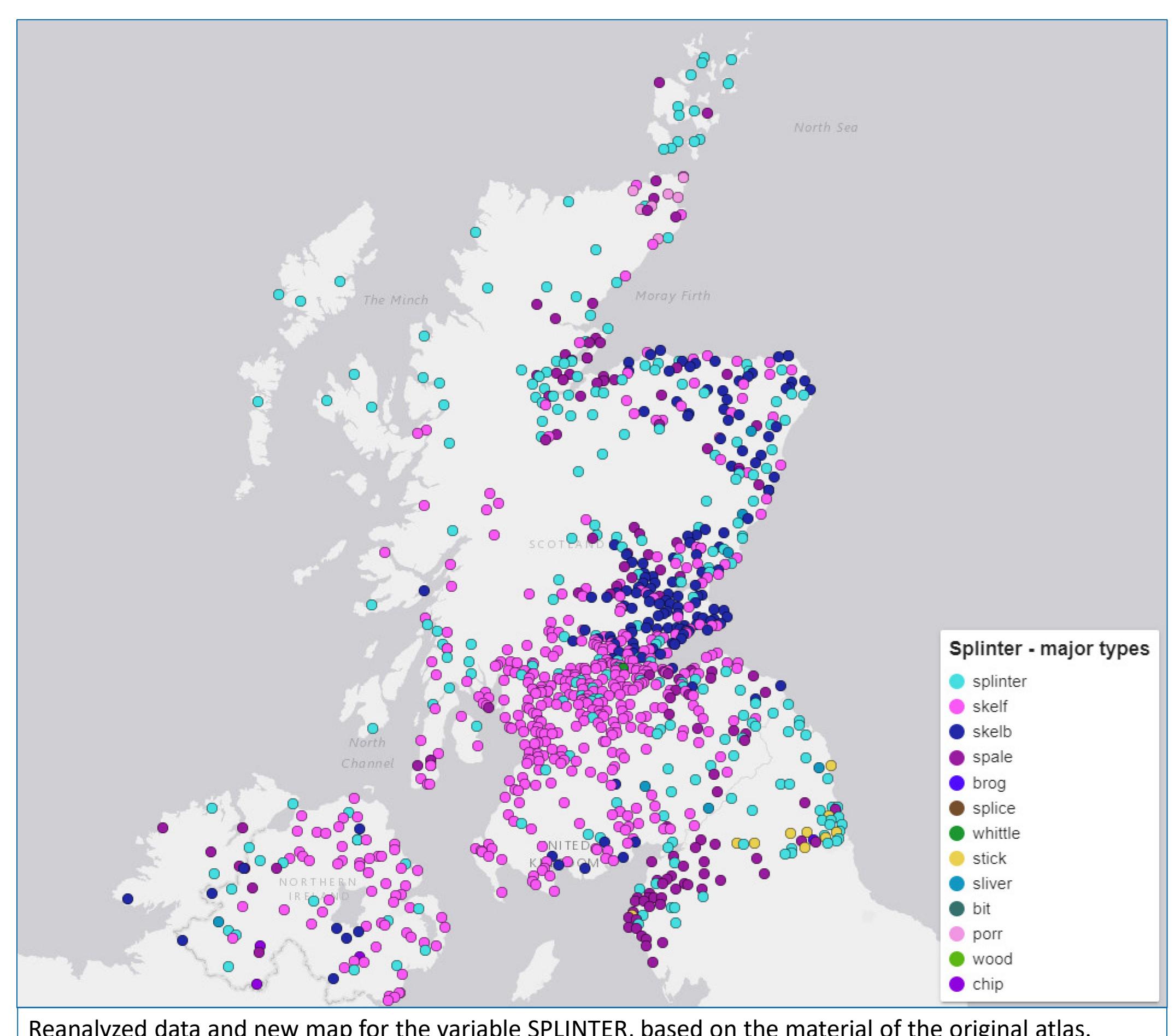
Informants



- Informants were imported separately
 - Informants are indicated by number in the original data
 - Numbers repeat for each county – not uniquely identifiable by number alone, but by county and number
 - Informants are used to connect variants to a locality

Trust us!

Try it out yourself and check out our maps at
[www.lasdb.dioe.at!](http://www.lasdb.dioe.at/)



Digital Literacy Assessment Toolkit

Digital Linguistic Atlas of Scotland

Centre for Translation Studies, in Cooperation with the
Special Research Programme German in Austria –
Variation, Contact, Reception (FWF-FGG)

John Kirk (University of Vienna): john.kirk@univie.ac.at
Markus Pluschkovits (University of Vienna, Austrian Centre for Digital Humanities and Cultural Heritage): markus.pluschkovits@uni.ac.at



BRAND

- References:**

DLAS (2022): *The Digital Lexical Atlas of Scotland*. <<https://lasdb.dioe.at/>> [04.11.2022].

Hessel, Christian / Kirk, John (2020): “Digitising Collections of Historical Linguistic Data: The Example of the Linguistic Atlas of Scotland”, in: *Journal of Data Mining and Digital Humanities. Special Issue Special Issue on Visualisations in Historical Linguistics*: 1–17.

Github (2023): GitHub-Repository of the Digital Linguistic Atlas of Scotland. <<https://github.com/german-in-austria/lasDB>> [26.06.2023].

Kirk, John / Hessel, Christian (2020): “Towards a Digital Lexical Atlas of Scotland”, in: *Scottish Language* 39: 1–39.

Kirk, John & Pluschkovits, Markus (forthcoming.): “Introducing the Digital Lexical Atlas of Scotland”, in: *The languages of Scotland and Ulster in a global context, past and present: Proceedings of FRLSU13, München, 28-30 October 2021 (= Publications of the Forum for Research on the Languages of Scotland and Ulster 6)*.

Mather, I. Y. / Speitel, H.-H. (1975): *The Linguistic Atlas of Scotland. Scots Section. Volume 1*. London: Croom Helm.