sdsbrowser

Making old data collected in the SDS stone artefact documentation system available and enjoyable

---

Clemens Schmid
@nevromeCS
github.com/nevrome
clemens@nevrome.de

Christoph Rinne
github.com/chrinne
crinne@ufg.uni-kiel.de

A project by the
Johanna Mestorf Academy
jma.uni-kiel.de

CAA 2019
R as an archaeological tool:
current state and directions (vol. II)
04/2019

TOC

- Data acquisition and sharing in archaeology
- The SDS System
- Making SDS data accessible:
  R package + Shiny Webapp
- Lessons for the future

Data acquisition and sharing

Bad hidden data

Problems!

- Archaeology often produces **Bad Data** (pretty shitty Big Data)
  - **Volume** – low end of the spectrum
  - **Variety** – a lot! and a significant lack of metadata
  - (Velocity)
  - **Veracity** – highly different data quality
- Lack of data (and code) **sharing**
- Lack of **long term storage** strategies

Solutions?

- Positive pressure
  - Funding agencies and Journals require data sharing now
  - Open Science and Reproducible Research Movement
- Tools and enabling features
  - Data Storage Providers (e.g. *Archaeology Data Service*)
  - Data Journals (e.g. *Journal of Open Archaeology Data*)
  - Data Licensing models
  - new **Software Frameworks** that simplify good data sharing practices

The SDS System

SDS – Systematic and Digital Documentation of Stone Artefacts

A. Drafehn/M. Bradtmöller/D. Mischka, SDS – Systematische und digitale Erfassung von Steinartefakten (Arbeitsstand SDS 8.05). Journal Of Neolithic Archaeology 10, 2008. doi:10.12766/jna.2008.25

> *The following contribution offers a **recording system for stone artefacts**. This **compilation of existing systems** from the Upper Paleolithic to the Neolithic should serve as a starting point for **systematic and quantitative analysis** with a **uniform coding** and standardized listing system of the conventional attributes recorded at lithic inventories. Therewith a valuable basis for **comparative analysis and digital exchange** of the data should be guaranteed.*

Features:

- Description of **individual** artefacts
- ~**200** variables (*color*, *weight*, *preservation*, *measurements*) with >**1300** predefined values/categories
- Table layout with separate **form sheets**
- Digit based **value encoding**

SDS – A form sheet

Form sheet 1																
Variable 1				Variable 2				Variable 3					Variable 4			
site				year				feature number					excavation number			
digits: 4				digits: 4				digits: 5					digits: 5			
L	A	0	5	2	0	0	9	1	0	8	0	2	1	0	0	...
L	A	0	5	2	0	0	9	1	0	8	0	2	1	0	0	...
L	A	0	5	2	0	0	9	1	0	8	0	2	1	0	0	...
L	A	0	5	2	0	0	9	1	0	8	0	2	1	0	0	...
L	A	0	5	2	0	0	9	1	0	8	0	2	1	0	0	...
L	A	0	5	2	0	0	9	1	0	8	0	2	1	0	0	...
L	A	0	5	2	0	0	9	1	0	8	0	2	1	0	0	...
L	A	0	5	2	0	0	9	1	0	7	0	1	1	0	0	...
L	A	0	5	2	0	0	9	1	0	7	0	1	1	0	0	...
L	A	0	5	2	0	0	9	1	0	7	0	1	1	0	0	...
L	A	0	5	2	0	0	9	1	0	7	0	1	1	0	0	...
L	A	0	5	2	0	0	9	1	0	7	0	1	1	0	0	...
L	A	0	5	2	0	0	9	1	0	7	0	1	1	0	0	...
L	A	0	5	2	0	0	9	1	0	7	0	1	1	0	0	...
L	A	0	5	2	0	0	9	1	0	7	0	1	1	0	0	...
L	A	0	5	2	0	0	9	1	0	7	0	1	1	0	0	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

SDS – Hierarchy of form sheets

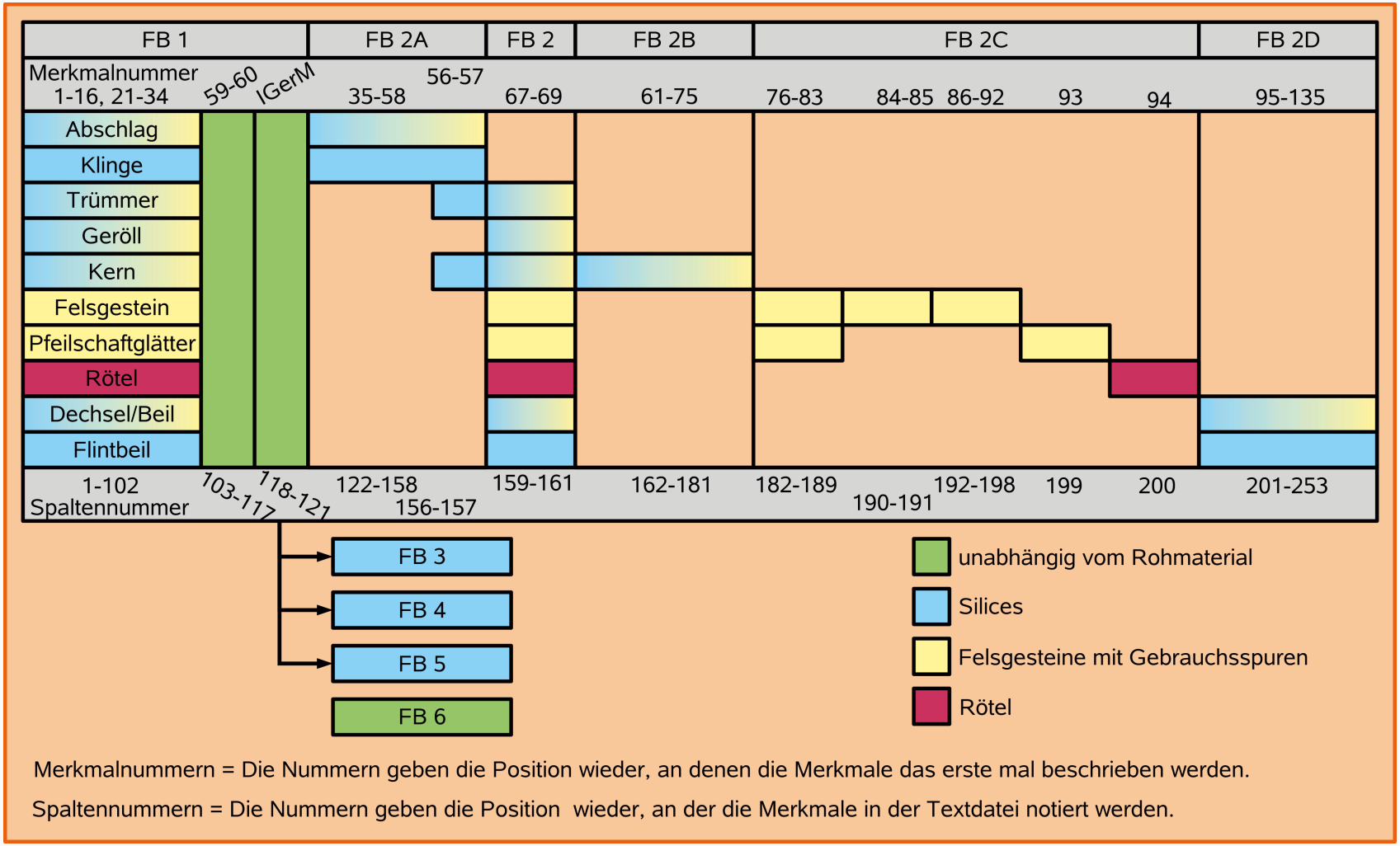

Figure 1: SDS Form sheet organization. FB = Formblatt = Form sheet. Drafehn/Bradtmöller/Mischka 2008, 3.

SDS – The actual implementation

10 years of use on 20+ sites by dozens of researchers and students with no common supervision: Cultural Evolution!

Practical solutions (aka *Hacks*):

- Elimination of the **digit separation**
- **Ignoring** most variables but **adding** new variables and values
- Creation of a **multi artefact description form sheet** with only ~20 variables

Consequences:

- Inconsistent and **incongruous file structures**
- **Undocumented variables** and **values** that only appear in individual datasets
- Absolute **domination** of the **multi description system**

Making SDS data accessible

General steps

1. Data collection:

- Reconstruct the **SDS user history**
- **Hunt for SDS data** on hard drives and forgotten network storages

2. Cleaning:

- **Repair or document hacks** to establish a modern SDS layout
- **Merge form sheet tables**

$$\underbrace{1 + 2}_{\text{Preparation}} + \underbrace{3 + 4 + 5}_{\text{R + Shiny}}$$

sdsanalysis – Backend

3. Data sharing:

- **Host SDS datasets**
- Provide functions to **download them to R**

```
sdsanalysis::get_available_datasets()
sdsanalysis::get_single_artefact_data("Kuesterberg") -> Kuesterberg
```

4. Increasing human readability:

- Provide functions to **decode the alphanumerical value coding scheme** (*hashes*)

```
Kuesterberg %>% sdsanalysis::lookup_everything()
```

R Package **sdsanalysis**:
https://github.com/Johanna-Mestorf-Academy/sdsanalysis

sdsbrowser – Webapp

5. Data presentation:

- Make SDS datasets directly explorable and downloadable **online**

R Shiny Webapp **sdsbrowser**:
https://github.com/Johanna-Mestorf-Academy/sdsbrowser
hosted at: **http://sds.ufg.uni-kiel.de**

Lessons for the future

Lessons

Modesty:

- Users will ignore/modify/slaughter overly complicated databases and documentation systems. **Keep it simple**

Supervision:

- Data collection requires supervision and **technical guidelines**

Long term strategies:

- Long term data sharing must be discussed **at the very beginning**, because people and data get lost quickly

Scripting Languages:

- Scripting Languages and Web Frameworks like R and Shiny provide **simple tools** to make **complicated data** more **visible and accessible**

sdsbrowser

Making old data collected in the SDS stone artefact documentation system available and enjoyable

---

Clemens Schmid
@nevromeCS
github.com/nevrome
clemens@nevrome.de

Christoph Rinne
github.com/chrinne
crinne@ufg.uni-kiel.de

A project by the
Johanna Mestorf Academy
jma.uni-kiel.de

CAA 2019
R as an archaeological tool:
current state and directions (vol. II)
04/2019