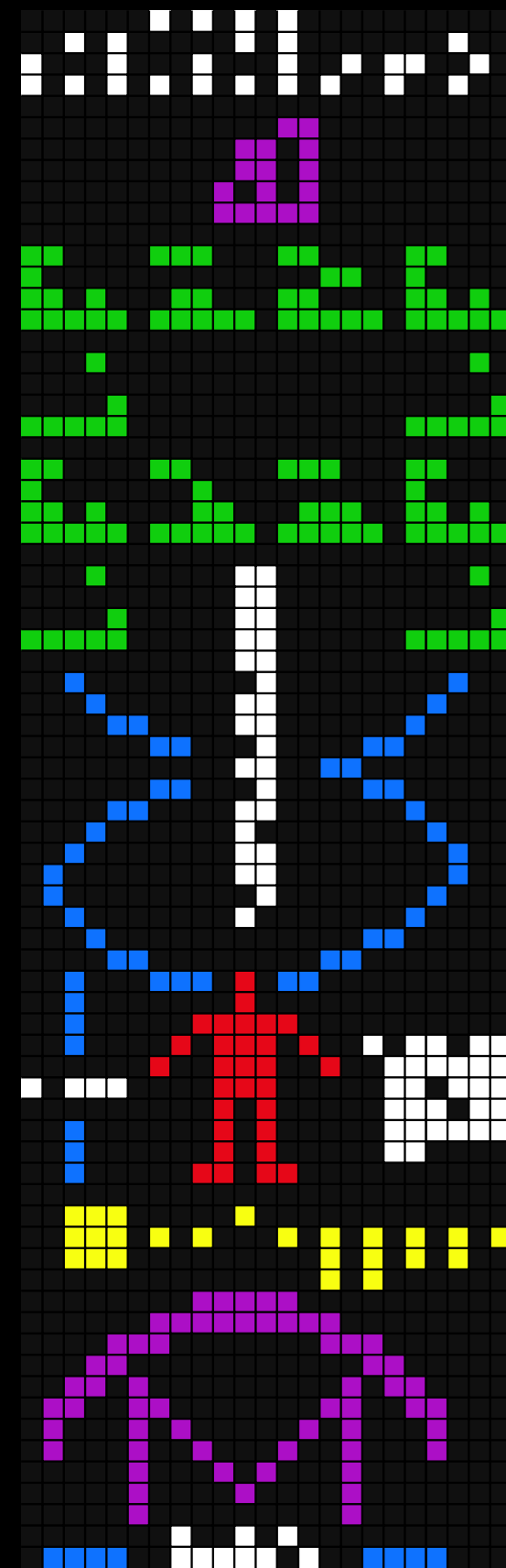


Many methods, technologies, standards, and languages exist to structure and describe data. The aim of this thesis is to find common features in these methods to determine how data is actually structured and described. Existing studies are limited to notions of data as recorded observations and facts, or they require given structures to build on, such as the concept of a record or the concept of a schema. These presumed concepts have been deconstructed in this thesis from a semiotic point of view. This was done by analysing data as signs, communicated in form of digital documents. The study was conducted by a phenomenological research method. Conceptual properties of data structuring and description were first collected and experienced critically. Examples of such properties include encodings, identifiers, formats, schemas, and models. The analysis resulted in six prototypes to categorize data methods by their primary purpose. The study further revealed five basic paradigms that deeply shape how data is structured and described in practice. The third result consists of a pattern language of data structuring. The patterns show problems and solutions which occur over and over again in data, independent from particular technologies. Twenty general patterns were identified and described, each with its benefits, consequences, pitfalls, and relations to other patterns. The results can help to better understand data and its actual forms, both for consumption and creation of data. Particular domains of application include data archaeology and data literacy.

## Describing Data Patterns

Jakob Voß



Jakob Voß

# Describing Data Patterns

A general deconstruction  
of metadata standards