

An R Framework to Handle Time Series and Its Meta Information

Matthias Bannert

KOF Swiss Economic Institute, ETH Zurich
Contact author: bannert@kof.ethz.ch

Keywords: meta data, time series, object relational mapping, postgresSQL, localization

Meta information is inevitable to find, select, share and re-trace suitable data for research projects. The steadily growing amount of data and the plethora of globally available providers have emphasized the need of proper meta information. While the ability to reproduce statistical analysis has been a major topic in the *R* community in recent years, handling meta information in *R* has not been addressed comprehensively. Thus structured meta information is often not available in context (i.e. an *R* session) when running an analysis in *R*.

This talk suggests a framework that uses object-relational mapping to handle and store time series data alongside comprehensive meta information. A new *R* package called **miro** (Meta Information for R Objects) to create, update, delete and assign meta information objects to existing *R* objects is introduced as part of the framework. Though **miro** can be used as a standalone package to organize meta information inside *R* it is most useful when combined with the new **pgSQLtmap**. The **pgSQLtmap** package maps data and the meta information objects created by **miro** to a *PostgreSQL* database. Both packages were separated to make **miro** independent of database related dependencies such as **RPostgreSQL** [2]. As opposed to the pioneering **TSdbi** [3] package family which only offers rudimentary support for meta information, the framework presented in this talk has extensive support for meta information including translations.

The **miro** package structures meta information and provides methods to access, display and mutate information. Meta objects created by **miro** reside in a separate environment and leave data itself basically untouched which makes **miro** completely agnostic to users' preferences for a particular time series package. Thus **miro** can be used with S3, S4 and reference class based time series implementations alike. On the database level the suggested framework makes use of *PostgreSQL*'s data type `hstore` [1] which implements storage of key-value pairs in a single field. This allows for a flexible amount of meta information items and a varying number of languages.

The framework presented in this talk can be used to build a time series database that can be queried semantically. Because meta information is made available in context *R* can be used as a flexible user interface to share data with applied practitioners. Context aware meta information can facilitate reproducible research by using the information to label tables and figures generically. Originally designed to manage aggregated time series stemming from longitudinal surveys this talk focuses on time series only but the general idea could be extended to other types of datasets as well. Future versions of the framework could support other databases by using a standard data type such as JSON that is supported by newer *PostgreSQL* and other DBMS.

References

- [1] Bartunov, O. and T. Sigaev. PostgreSQL: Documentation: 9.1: hstore. <http://www.postgresql.org/docs/9.1/static/hstore.html>. Accessed: 2014-03-24.
- [2] Conway, J., D. Eddelbuettel, T. Nishiyama, S. K. Prayaga, and N. Tiffin (2013). *RPostgreSQL: R interface to the PostgreSQL database system*. R package version 0.4.
- [3] Gilbert, P. (2013). *TSdbi: TSdbi (Time Series Database Interface)*. R package version 2013.9-1.