

## Some facts about the „Alephino“ database

- To map bibliographic data to relational databases (RDBMS) is not possible with respect to international or national standards such as MARC21, Unimarc or MAB2. That's particularly with regard to:
  - Record structures as well as record and field lengths are generally variable. You'll hardly find a record that's equal to another with respect to its record structure, its fields and/or their repetitions.
  - The maximum number of fields per record exceeds the limitations of the respective RDBMS (Oracle = 1000).
  - Fields have sub-structures (subfields, indicators) and hence are not atomic - in contrast to the general rules of normalization.
  - The normalization of data is an option and must not necessarily be applied to each record.
  - The standards mentioned above work with a multitude of field groups that can potentially be repeated. Those repetitions make also a contradiction to canonical normalization rules.
  - The abilities of standard database systems, when it comes to text retrieval, are not sufficient for the needs of bibliographic data. The process of splitting text contents into words under consideration of „stop words“, partial numeric indexing, special phrase indexing etc. must allow a much greater variety.

Many well known professional library software solutions, as long as they're based on relational standard databases, do not make use of their relational record structures but instead bibliographic records are stored in an unstructured way using field types „CLOB“ or „BLOB“.

This – by the way – also applies to our product „Aleph 500“.

Hence any access to bibliographic data needs mediation by the – supplier specific – application logic. So the assumption to always be on the safe side when using standard databases is questionable with regard to the safety of investment.

- In contrast the database „Alephino“
  - Uses the native data schema of the respective bibliographic standard internally. This way the data exchange is simple and reliable as long as the data do meet standards.
  - Complies with all requirements on text retrieval.
  - Provides a very good performance on retrieval of bibliographic data as well as combination of data in terms of „views“ or „joins“.
  - Consists of a single file from point of view of the respective operating system. All data tables regardless whether bibliographic or administrative data are included in terms of a virtual file system and hence an Alephino database can be used under Windows as well as Linux and Solaris OS without need to convert anything, simply by copying the database file.
  - Includes an integrated save concept based on redundancy in storage. With that we make sure that Alephino databases can be recovered anytime without loss of data. The respective save and recovery procedures are handled via web-services.
  - Indexing and linking between records happens automatically along with the storage of data. Hence there's no need for supplementary processes. Data are searchable without any delay.
  - A disadvantage compared to relational databases is that data retrieval always requires indexing. Data fields not included in an index list cannot be searched by default.
  - All tables resp. record types are dealt with equally in terms of the database. There's no different handling of bibliographic records such as titles, authorities, subjects and thesaurus and administrative data like patrons, vendors, orders, loans etc.

- Some more facts in numbers:
  - max. records per table =  $2^{24} - 1$  (16.777.215 )
  - max. record length = 30.000 characters (due to GUI limitations)
  - max. field length = 2.000 characters (due to GUI limitations)
  - internal character set = UTF-8
  - max. number of fields / record =  $35^3 * 200$  repetitions \* 35 subfields

**Alephino 5.0 – Data model (simplified)**

