# 67th IFLA Council and General Conference
# August 16-25, 2001

**Code Number:** 005-188(WS)-E
**Division Number:** IV
**Professional Group:** Division Bibliographic Control
**Joint Meeting with:** UBCIM
**Meeting Number:** 188
**Simultaneous Interpretation:** -


## Mapping UNIMARC data to the RLG/CERL Hand Press Books database


**Joe Altimus**
The Research Libraries Group
Mountain View, USA



### *Abstract:*

*The experience of the Research Libraries Group, Inc. (RLG) and Consortium of European Research Libraries (CERL) with round trip conversion of UNIMARC data to a MARC 21-based format shows that there are considerable difficulties in such a conversion. However, various techniques can be used to deal quite successfully with the incompatibilities of those two formats.*


Since 1997 the Research Libraries Group, Inc. (RLG) has provided its users with a database of bibliographic records for hand press books (HPB) published up to 1850. The database is produced through an agreement with the Consortium of European Research Libraries (CERL). The CERL Executive Office oversees the process of getting records for hand press materials from its members and sending them to RLG for processing and loading to the Hand Press Books database.

Until very recently, RLG's databases have been comprised solely of RLIN MARC elements, which are basically the same as MARC 21 (formerly USMARC), with a few additions. However, most of the records that CERL supplies to RLG for the HPB database use UNIMARC format. Ideally RLG's HPB database would have been based on a set of MARC elements that could accommodate UNIMARC easily, as well as the other types of MARC format data that CERL sends to RLG (IBERMARC, UKMARC, etc.). However, the costs associated with creating such a database were too high for RLG

and CERL. Instead all data that CERL contributes to the HPB database must be translated to RLIN MARC format.

There is also a flow of records out of the HPB database. CERL has recently developed UNIMARC export capabilities from HPB through work with RLG to specify an RLIN MARC to UNIMARC conversion, and through work with Crossnet, Inc., to develop a Z39.50-based conversion application. Thus CERL and RLG have experience in round trip mapping of UNIMARC data to and from RLIN MARC.

From the start CERL was aware of the problems associated with converting UNIMARC data to a MARC 21-based format. CERL's objective was to be able to retrieve an UNIMARC format record from the HPB database with no loss of the data that had been in the UNIMARC record sent to RLG. CERL and RLG quickly realized the difficulty of achieving this goal because of the incompatibilities between UNIMARC and RLIN MARC. In a few cases, one format defines an element that the other format lacks (e.g., the UNIMARC 503 has no RLIN MARC equivalent). Often, however, the two formats both define the same element, but the details differ (indicator values, subfields, coded values).

CERL and RLG use several techniques to cope with incompatibilities. UNIMARC defines a couple of dozen extended Latin script characters that are not in RLG's character set definitions. CERL defined surrogates for those UNIMARC characters (e.g., \IJ\\ stands for the Capital Letter IJ character in ISO 5426). The few UNIMARC fields with no RLIN MARC equivalent, such as the UNIMARC 503, are converted to RLIN MARC 886. This completely preserves the source data.

As just mentioned, many UNIMARC fields are similar to RLIN MARC ones, but not enough to allow complete conversion. RLG also uses the RLIN MARC 886 for any UNIMARC field of that type, but in addition, it is also converted to the most appropriate RLIN MARC field(s). For example, the UNIMARC 303 is translated to the RLIN MARC 500 as well as the 886. This allows the RLIN HPB record to retain the source UNIMARC data in a functional way, because the RLIN MARC 886 is not treated by RLG as an indexable or displayable element. This "double conversion" technique can be used to cope with any UNIMARC/RLIN MARC incompatibility when a significant number of UNIMARC subfields or coded values for a particular UNIMARC field lack RLIN equivalents.

The disadvantage of the double conversion technique is that redundant data occurs in an HPB record. The RLIN MARC 886 data permits complete reversibility of the UNIMARC source data supplied to RLG. However, that same data, or some portion, is also present in the HPB record in some other RLIN MARC field. It would be extremely difficult to write an export program with the necessary rules to automatically remove the redundant data. At this time, the record recipient must manually remove redundant data.

RLG and CERL's experience with round trip conversion of UNIMARC data to a MARC 21-based format shows that although there are considerable difficulties in such a conversion, various techniques can be used to deal quite successfully with the incompatibilities of those two formats.