

FRBR



ELAG 2003 Workshop 5 report

By Maja Zumer and Ole Husby

Participants

Paula Goossens

Gunilla Jonsson

Vinod Chachra

Antoinette Lemaire

Hildur Gunnlaugsdottir

Iris Marthaler

Thom Hickey

Dan Matei

Liv Holm

Gordana Popovic

Ales Horak

Regina Varniene

Ole Husby

Paul Weston

Andrea Johannsdottir

Maja Zumer

With apologies for the poor character set

Introductory remarks

- FRBR is a conceptual model
- FRBR will move us out of the card context thinking. Practical outcome may be more modest. FRBR has given us a new language.
- We are moving to XXXX (another system).
Why not move to the 4-level model.

Presentations and reports

- Ole Husby on BIBSYS experiences
- Maja Zumer on INWECA
- Thom Hickey on OCLC research
- Maja Zumer on IFLA FRBR working group
- Vinod Chachra on VTLS development
(music database)

Topics discussed

- FRBR'izing
- Identifiers
- User interface (presentation of search results)

Topics not discussed

- Moving from a conceptual model to a data model
- Linking services
- Prototyping and testbeds
- Improving the FRBR model itself (although we quite often were tempted, or identified certain needs ...)

FRBR'izing

- is the art of identifying the group 1 entities
- match them against the rest of the database
- cluster them
- identify the links needed for navigation

There may be several motives for doing this

Common experiences concerning many library catalogues

- it is not easy to FRBR'ize them
- and it is unlikely that it can be fully automated
- Skilled cataloguers know more about the state of the database than the most sophisticated algorithm!

The BIBSYS experience

The analysis shows that even though the information in the MARC records holds attributes relevant for identifying the work, expression and manifestation entities, the accuracy and formal syntax are too simple to be properly handled by programs.

Hegna & Murtomaa

Typical problems

- Lack of role qualifier in 700
- Shared search index for 100 and 700
- Lack of language codes
- Lack of or inconsistent authority control facilities
- Inconsistent registration of multiple works sharing one manifestation

Other common problems

- Heterogeneity concerning cataloguing standards (primary records, photocatalogued records, converted records)
- Original / uniform title not present
- Translator difficult to identify
- Original language not present
- Work title authorities lacking

OCLC Research on FRBR & WorldCat

- Subsets
 - By library, region
 - Example/problem sets
 - Shakespeare, the Bible
 - Humphry Clinker
 - 1,000 random works
 - By genre
 - Dissertations
 - Fiction
- Whole file, 49 million bibliographic records
- PICA NCC file of 3.5 million records

OCLC Approach

- Concentrating on work-level
 - Problems with expression-level clusters
- Efficient, maintainable, understandable
- Few, if any, false matches with correct cataloging
 - Err on the side of missed matches
 - Some accommodation of frequent variants
- Compare with manually clustered

OCLC Algorithm

- A key is generated for each record
- Extract author, title
 - Look up in NACO authority file
 - Added entry information as needed
- Form a key from bibliographic record
 - Author, title, added entry information
 - These can be sorted, compared

OCLC experiences

- Many (17%) records do not have
 - Author main-entry
 - Uniform title
- In general these can not be matched
 - Look at added entries
 - Information at the expression and manifestation levels
 - Handled separately
 - 514,000 clusters involving 1,758,000 records

OCLC Results

- Manual estimate: 1.5 manifestations/work in WorldCat
- Algorithm: ~1.25
- 25,000 clusters have >20 records
- 415,000 clusters have >4 records
- 30% records, 50% holdings in a cluster

Some conclusions

- Cataloguing rules and practice not tailored for FRBR
- Central data entities are still meant for human eyes, not for automated processing
 - meaning too much unstructured text
- Data quality, accuracy and integrity not good enough

Identifiers

- We have several identifiers for manifestations
- But lack them on the work / expression levels
- Or they are
 - Not aligned with the FRBR model
 - Not well enough defined or inconsistently used
 - For network resources only

Identifier requirements

- Persistent
- Globally unique
- Well managed
- Resolvable
- Exchangeable
- (Syntax requirements ???)

Identifiers discussed

- **ISTC (text)**
 - OCLC involvement in ISTC service prototyping
 - Not yet an accepted standard
- **ISWC (music)**
 - ???
- **DOI**
 - cost money
 - uncertainty about FRBR alignment

More identifiers

- URN:NBN
 - Uncertainty about FRBR alignment
 - Inconsistent practice
 - But promising
- URN:XXX
 - Do we need new NIDs within URN?
 - Or a new type of URI?

User interfaces

- Focus on searching for bibliographic entities (group 1)
- What is a query?
- How to display a result set?
- What is a result set?

What is a query?

- Did not discuss that really
- but searching for **author, title, subject**, or **known item** will have (quite) different sets of requirements
 - perhaps more different than we are used to
- and users and context differ

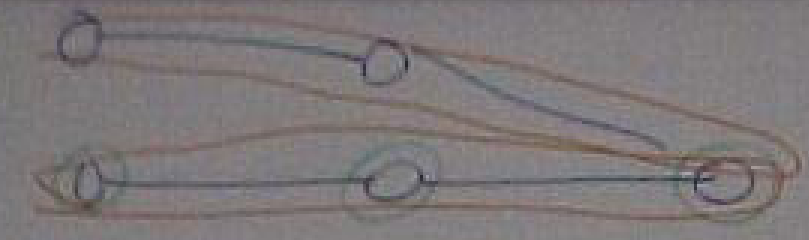
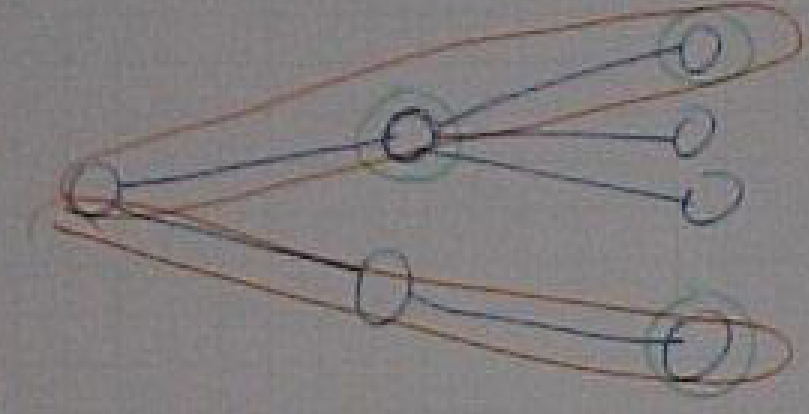
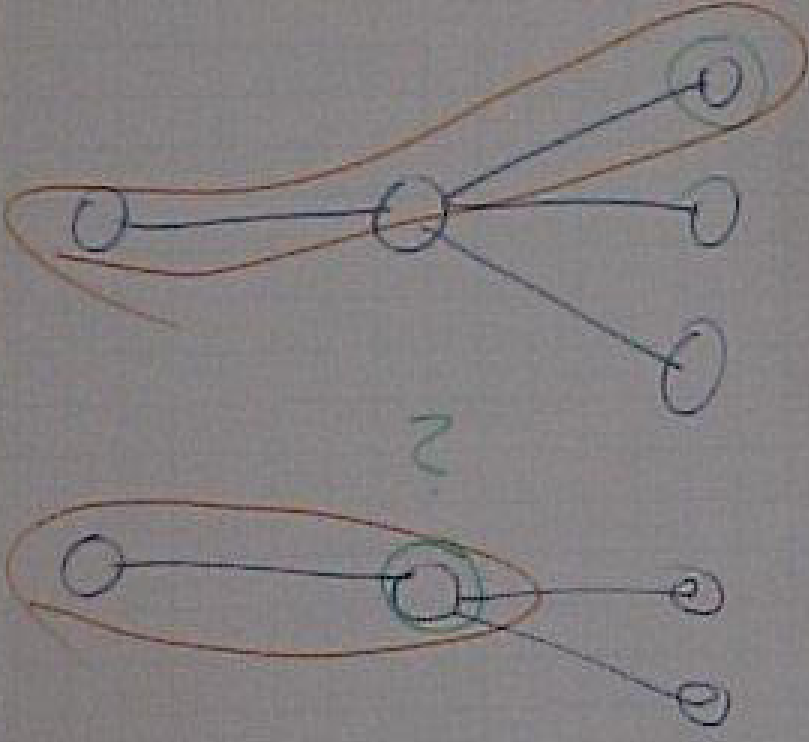
How to display a result set

- Consensus that expandable lists are in many contexts a good method
- Problems seem to arise often on the expression level
- How should an expression be named in a hit list?
- How to (if) group or cluster different expressions in the same language

What is a result set?

- We do not know yet
- But we think there are needs to
 - display them
 - store them
 - manipulate them (intersect, refine ...)
 - exchange them
- And what is the **hit count** ???

but anyhow, this is what they look like:



TOO MANY BRAINS ~~SMILING~~ HITS
SEE YOU NEXT YEAR!