

MARC 21 and FRBR

Trond Aalberg

Norwegian University of Science and Technology (NTNU),

Department of Computer and Information Science

*MARC 21 - Experiences, Challenges and Visions
The German National Library in Frankfurt/Main, 14-15 May 2007*

Content

- Intro
- Some words about MARC and FRBR
- FRBR in the context of MARC 21
 - Some examples (walk-through)
 - Possibilities, problems and challenges
- Concluding remarks

MARC formats –

from an outsiders point of view

- A family of formats based on the ISO 2709 standard for information exchange
- The basic data-structure is in fact very simple!
- Variable fields identified by a three-character numeric tag:
 - Control fields of single data element or a series of fixed-length data elements
 - Data fields for complex values using subfields and indicators
- Each data field describes an aspect of the bibliographic item
- Simple in nature but complex in use!
 - large number of data field tags (MARC 21 ~ 200)
 - large number of subfields and indicators (MARC 21 ~ 2000)
 - large number of rules

The dual nature of MARC formats

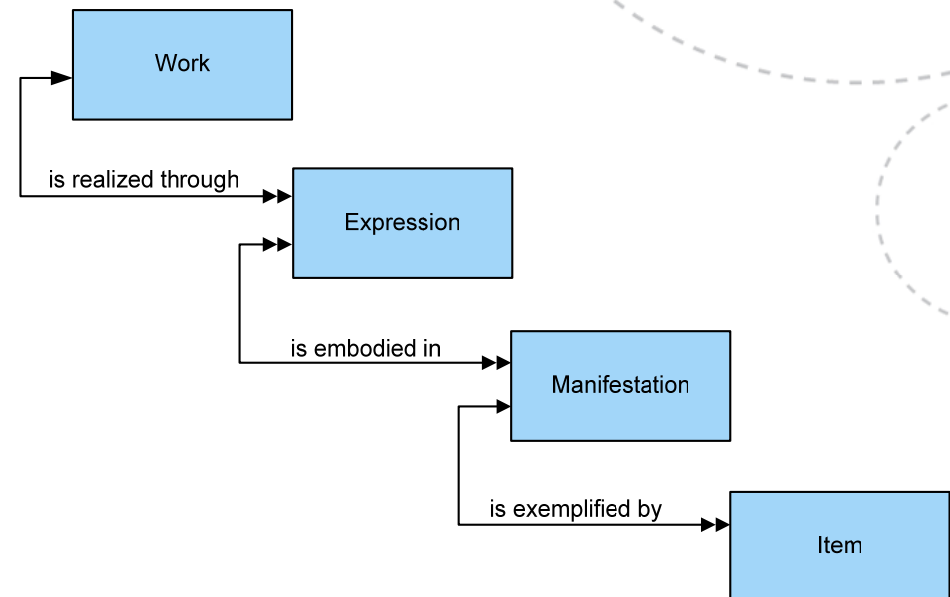
- MARC formats are representation and exchange formats
 - serves interoperability as a common logical data model
 - supports exchange as a standardized record format
- May be stored in different ways, but
 - structure influences how you structure and organize information!
 - would bibliographic information look different without MARC?
- Both a carrier and a requirement

Towards a single MARC format?

- MARC 21
 - Trend in changing from national formats to MARC 21 as exchange format
- UNIMARC
 - Different from MARC 21, basically in the use of tag numbers, but in other features as well
 - Used in some countries, but did not really succeed as an international standard
- But still:
 - Many national or vendor-specific formats in use
 - Often a level of translating and adapting even when using MARC 21 or UNIMARC

IFLA's Functional Requirements for Bibliographic Records - FRBR

- *Aims to establish a precisely stated and commonly shared understanding of what it is that the bibliographic record should provide information about.*
- Defined by the use of an entity-relationship model
- FRBR is a **conceptual model**
 - Not a specific metadata schema or data model



The FRBR model

– the short version

- Group 1
 - Work, Expression, Manifestation, Item
- Group 2
 - Person, Corporate body
- Group 3
 - Object, Event, Concept, Place
- Entities are connected by **relationships**
 - *A work is created by a person*
- Entities are described by **attributes**
 - A person has a *name*

FRBR and MARC?

- Why is this interesting?
 - Bibliographic catalogues are based on MARC formats
 - Any major change in the world of bibliographic information has to consider this “legacy” information
 - MARC may be “old-fashioned” but will be around for many more years
- Important questions:
 - Is MARC 21 (or other MARC formats) able to express FRBR?
 - What is needed to make the FRBR model more explicit in MARC records?
 - Can we improve the format or improve the use of the format?
- An evolutionary approach for realizing FRBR is more likely to succeed than a revolutionary one

The MARC record

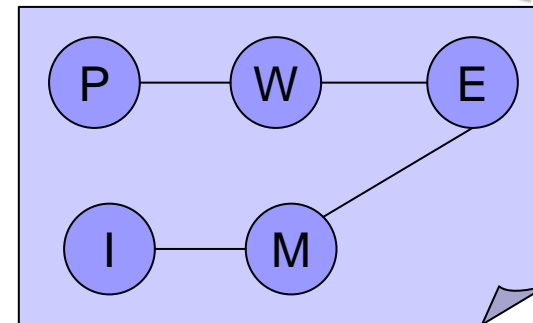
- A MARC record is a self-contained unit of information
 - A unit of information that can be exchanged and reused by others
 - Usually no dependencies to other records
- Includes the information that is needed to
 - Find, identify, select, obtain (FRBR user tasks) **manifestations**
- In the context of FRBR the bibliographic record is basically a **Manifestation surrogate**
 - But contains information that describes the many aspects of a publication (including other FRBR entities)

A simple example

- A single person that is the author of a single book

Person (1)

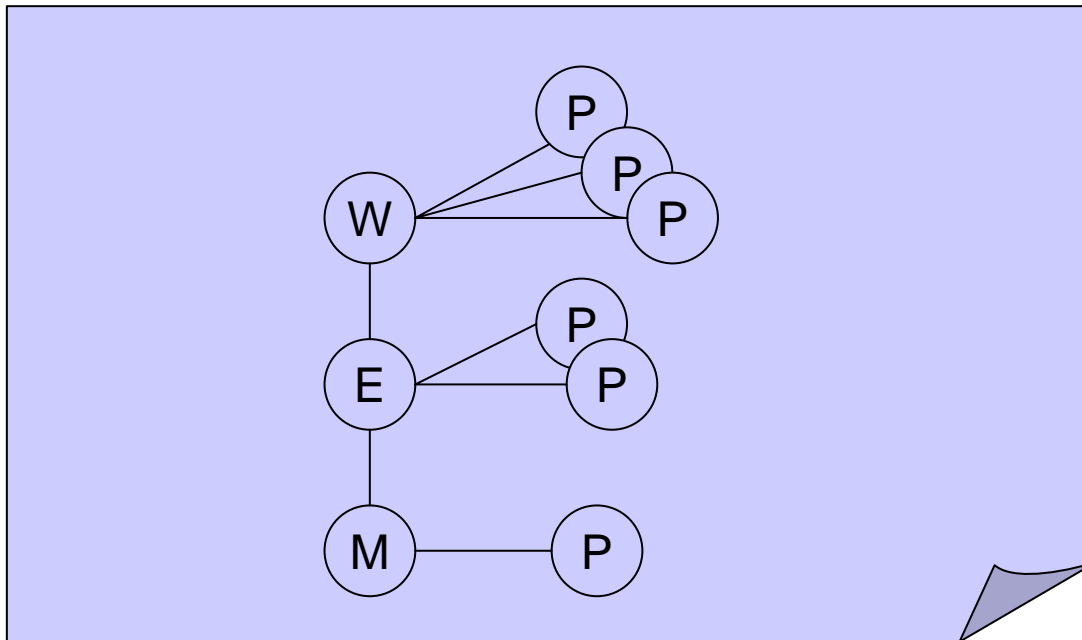
- *has created* Work (1)
- *is realized through* Expression (1)
- *is embodied in* Manifestation (1)
- *is exemplified by* Item (1+)



- A MARC record is perfectly able to capture this scenario and many existing records already express this simple scenario
 - implicit or explicit

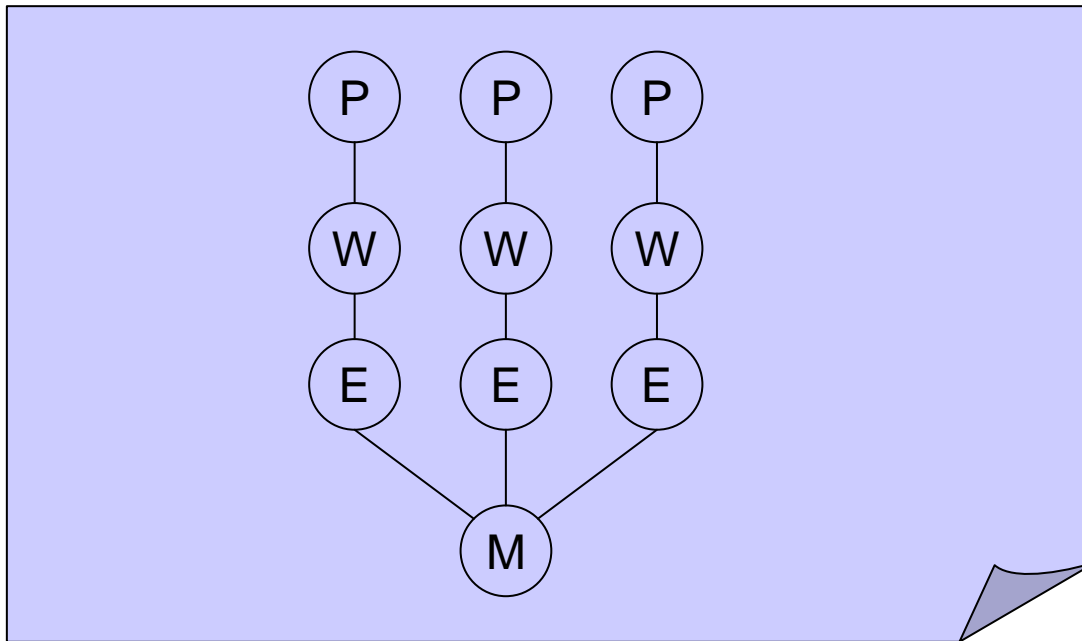
But what about ?

- Many occurrences of group 2 entities



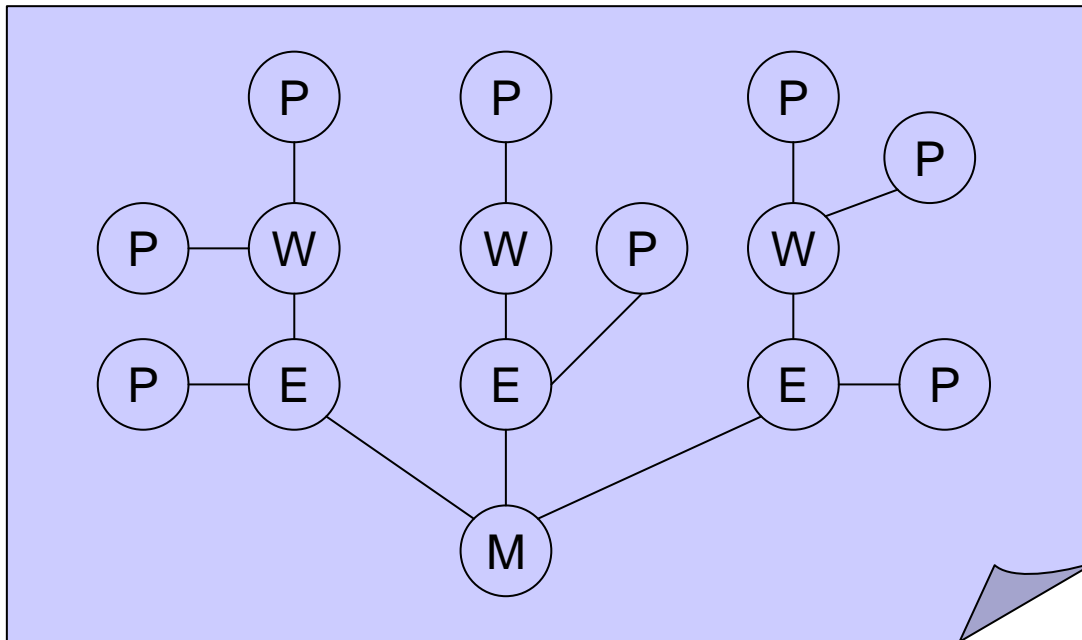
But what about ?

- Many works/expressions in one publication



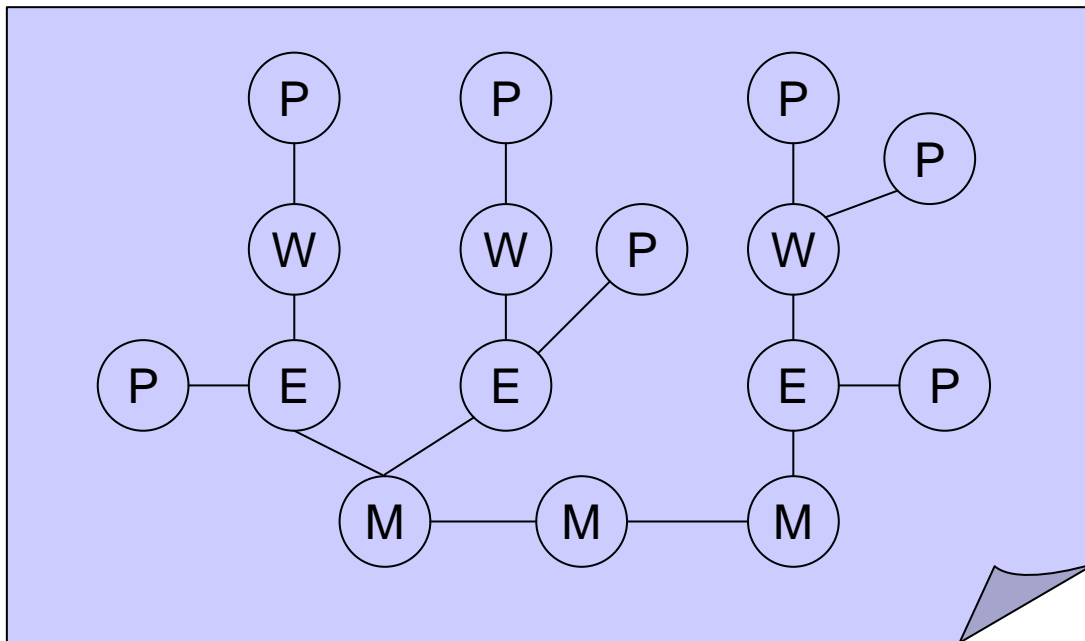
But what about ?

- Many works and many group 2 entities



But what about ?

- Multivolume publications with parts



Requirements for FRBR in bibliographic information

- Two fundamental requirements:
 - Entities must have well-defined identities
 - By the use of **consistent descriptive information** or by the use of **identifiers**
- Relationships must be well-defined
 - By semantics: the precise meaning of the relationship
 - By targets: the “to” and “from” of the relationship
- Attributes are important but less significant if the first two requirements are met
 - Except the ones that are needed for identification

Identifying works

- The notion of a work is inherent in any intellectual product
 - Properties required to identify a work:
 - Creator(s)
 - The title of the work
 - Date and form
 - (and sometimes other attributes)
- As a general rule any manifestation will embody at least one expression that is a realization of a work
- Is the work-level already sufficiently described in records?
- What are the fields that uniquely describe the work?

Creators of works

- Creators are in many cases essential for the identification of a work
 - 100, 110, 111
 - 700, 710, 711
- with the appropriate relator term/code for "is creator of"

The title of works

- Often equivalent to Uniform titles
- Do all records have a uniform title entry?
 - No, most records only have title statements
 - If not, is the title statement equivalent to work title?
 - In some cases the title statement can be used as work title, but in many cases not
- Not all uniform titles can be used as work titles:
 - 240 \$a Symphonies, \$n no. 5, op. 67, \$r C minor. \$p Allegro con brio. \$k Selections; \$o arr.
 - 240 \$a Sinfoniat \$b Beethoven \$e nro 5 \$j op67 \$r c-molli \$u 0005 \$v 0067
 - 240 \$a Symfoni \$n nr 5 \$n op. 67 \$r c-moll, "Ödessymfonin"

Identifying works based on 245 title

- Problem: Lack of uniform title when title statement is significantly different from original title – such as translations.
- May result in a large number of errors
 - Different title statements on different editions
 - Erroneous or inconsistent representation of title statement

Examples

- The same work and the same title in 245

100 \$a Ballard, J. G., \$d 1930-
245 \$a **Cocaine nights** / \$c J.G. Ballard.
260 \$a London : \$b Flamingo, \$c 1996
300 \$a 328 p. ; \$c 23 cm.

100 \$a Ballard, J. G., \$d 1930-
245 \$a **Cocaine nights** / \$c J.G. Ballard.
250 \$a 1st Counterpoint ed.
260 \$a Washington, D.C. : \$b Counterpoint, \$c 1998.
300 \$a 328 p. ; \$c 23 cm.

- The same work

100 \$a Burgess, Anthony, \$d 1917-1993.
245 \$a **Ernest Hemingway and his world** / \$c Anthony Burgess.
260 \$a [London] : \$b Thames and Hudson, \$c c1978.
300 \$a 128 p. : \$b ill. ; \$c 24 cm.

100 \$a Burgess, Anthony, \$d 1917-1993.
245 \$a **Ernest Hemingway** / \$c Anthony Burgess.
260 \$a New York : \$b Thames and Hudson, \$c 1999.
300 \$a 128 p. : \$b ill. ; \$c 24 cm.

Work or not

Plays / Ibsen, Henrik

A Doll's House; The Wild Duck, The Lady From the Sea

The Pillars of Society, A Play In Four Acts / Ibsen, Henrik

The Richard Mansfield Acting Version of Peer Gynt By

When We Dead Awaken: A Dramatic Epilogue In Three

The Collected Works of Henrik Ibsen / Ibsen, Henrik

The Works of Henrik Ibsen / Ibsen, Henrik

The Lady From the Sea / Ibsen, Henrik

The Late Plays of Henrik Ibsen / Ibsen, Henrik

Four Plays By Ibsen: An Enemy of the People, A Doll's House, The Master Builder, Peer Gynt / Ibsen, Henrik

... Brand / Ibsen, Henrik

Ibsen's Prose Dramas. -- / Ibsen, Henrik

Plays: Five / Ibsen, Henrik

Selections / Ibsen, Henrik

Ghosts; The Warriors At Helgeland; An Enemy of the People / Ibsen, Henrik

Emperor and Galilean: A World-history Drama. A Doll's House. Ghosts / Ibsen, Henrik

The Lady From the Sea / Ibsen, Henrik

Dukkehjem / Ibsen, Henrik

Peer Gynt / Ibsen, Henrik

Hedda Gabler / Ibsen, Henrik

Byggmester Solness / Ibsen, Henrik

Gengangere / Ibsen, Henrik

John Gabriel Borkman / Ibsen, Henrik

Vildanden / Ibsen, Henrik

Lille Eyolf / Ibsen, Henrik

Folkefiende / Ibsen, Henrik

Unges Forbund / Ibsen, Henrik

Rosmersholm / Ibsen, Henrik

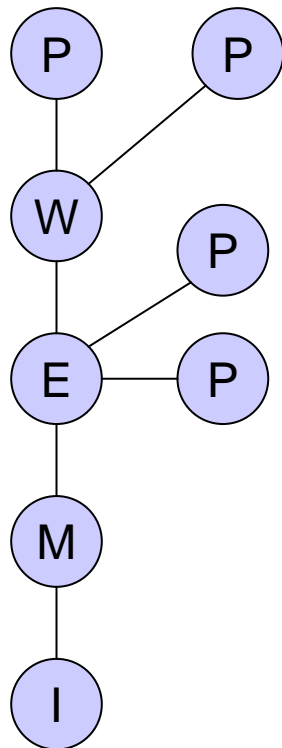
Identifying expressions

- Any manifestation will embody at least one expression
 - An expression is always a realization of only one work
 - If there is a work identified there is always an expression in between the work and the manifestation
- Properties required to identify an expression
 - The work
 - The persons or corporate body (*realized by*)
 - Dates
 - Language, form, and more (and sometimes other attributes)
- Can be grouped according to characteristics
 - But attributes for identification of distinct entities are difficult; such as translator, date of translation
 - Somewhat easier for performances (music, theatre, etc)

Added entries

- Is used for adding more access points not provided by other fields
- Is used to deal with multiple names and titles associated to an item
- Or to add information about constituent parts – analytical entries
- MARC 21: 7XX
 - A small number of data fields used for a number of purposes
 - Meaning and structure is managed by the use of indicators + relator codes and/or terms

Additional persons (or corporate bodies)



- Added entries can be used to associate more persons with the entities
 - Added entry fields in MARC21 (7XX)
- Relator codes are used to express what kind of entity the person is associated to
 - And the semantics of the relationship
- The applicability of this is depending on how ambiguous the relator codes are
- Without relator code the added entry is without meaning and it is impossible to know the kind and target of the relationship
- Descriptions may exist but are hard to interpret automatically

Author example

Two authors

100 \$a Sjöwall, Maj, \$d 1935-
 245 \$a Brandbilen som försvann. \$b Roman om ett brott. \$c [Av] Maj Sjöwall och Per Wahlöö.
 260 \$a Stockholm, \$b Norstedt, \$c 1969.
 300 \$a 249, (1) p. \$c 23 cm.
 700 \$a Wahlöö, Per, \$d 1926-1975. \$e joint author.

100 \$a Sjöwall, Maj, \$d 1935-
 240 \$a Brandbilen som försvann. \$l Á dönsku
 245 \$a Brandbilen som forsvandt / \$c Maj Sjöwall og Per Wahlöö ; på dansk ved Grete Juel Jørgensen.
 260 \$a [S.l.] : \$b Superpocket, \$c 2002.
 300 \$a 275 s.
 440 \$a Roman om en forbrydelse ; \$v 5
 700 \$a Wahlöö, Per, \$d 1926-1975
 700 \$a Jørgensen, Grete Juel

Three authors?

Managing complex information

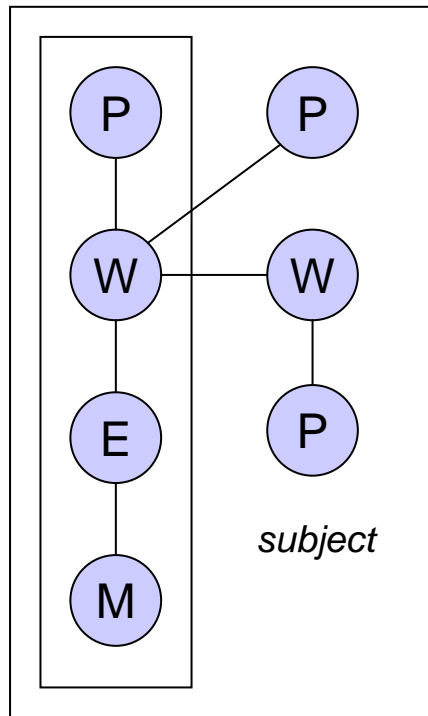
- Sometimes there is a need to organize the fields by more than tags and indicators
- MARC 21 : **\$8** - FIELD LINK AND SEQUENCE NUMBER
- E.g. associating added entry fields that pertain to the same constituent item

700 1_ \$82\c\$84\c\$aDi Giuseppe, Enrico,\$d1938-\$4prf
 700 12\$81\c\$aSiegmeister, Elie\$d1909-\$tFrom my window;\$oarr.
 700 12\$82\c\$aMozart, Wolfgang Amadeus,\$d1756-1791.\$tDon Giovanni\$pMio tesoro.
 700 12\$83\c\$aFlotow, Friedrich von,\$d1812-1883.\$tMartha.\$pAch! So fromm, ach! so traut.\$lItalian
 700 12\$84\c\$aPuccini, Giacomo,\$d1858-1924.\$tTurandot.\$pNessun dorma.
 700 12\$85\c\$aRespighi, O

740 \$aUna casa di bambola\$wcasa di bambola
 740 \$aSpettri
 740 \$aL'anitra selvatica\$w'anitra selvatica
 740 \$aEt dukkehjem\$wdukkehjem
 740 \$aGengangere
 740 \$aVildanden

*Readable and
searchable, but no
structure*

Works and persons as subject entries



- MARC 21
 - 600/610/611 fields for person/corporate/meeting names
 - 630 for uniform titles
- Subjects are distinct entries in a record
- In FRBR subject relationships are always from works
- Are we able to tell whether the relationship is to the person, work, or both

Example

The subject entry is correct, but does the name entry and uniform title reflect creator and work?

100 \$a Beethoven, Ludwig van, \$d 1770-1827.
240 \$a Selections
245 \$a Beethoven for dummies \$h [sound recording].
260 \$a New York : \$b EMI, \$c p1996.
300 \$a 1 sound disc : \$b digital, stereo. ; \$c 4 3/4 in.
440 \$a Classics for dummies
500 \$a The 1st and 3rd works for orchestra; the 2nd for violin and orchestra; the 4th for piano; the 5th for piano and orchestra; the 6th for SATB solos, SATB chorus, and orchestra.
546 \$a The 6th work sung in German.
600 \$a Beethoven, Ludwig van, \$d 1770-1827.

Aggregations

- Whole/part relationships may exist between all group 1 entities
- Can be of different types depending of the role of the part in the overall composition
- A range of techniques in use to express different types of “something being part of something”
 - Series
 - Analytical entries
 - Record Linking
 - Linking entry fields
 - Part-names in title fields

Series

- Some series are works

100 1_ \$a Tolkien, J. R. R. \$q (John Ronald Reuel), \$d 1892-1973.
 245 14 \$a The two towers / \$c J.R.R. Tolkien ; illustrated by Alan Lee.
 490 1_ \$a 490 1_ \$a The lord of the rings ; \$v pt. 2
 800 1_ \$a Tolkien, J. R. R. \$q (John Ronald Reuel), \$d 1892-1973. \$t Lord of the rings (2002) ; \$v pt. 2.

The title in the series entry title in one record, may be the main entry work in another record

100 1_ \$a Tolkien, J. R. R. \$q (John Ronald Reuel), \$d 1892-1973.
 245 14 \$a The lord of the rings / \$c by J.R.R. Tolkien.
 250 __ \$a [50th anniversary 1 vol. ed.]
 260 __ \$a Boston : \$b Houghton Mifflin, \$c [2005]

240 10 \$a Lord of the rings
 245 10 \$a Hringadróttinssaga / \$c eftir J.R.R. Tolkien ; Þorsteinn Thorarensen íslenskaði ; ljóðapýðingar Geir Kristjánsson.

- But not all series entries are relevant as works

800 1_ \$a Bach, Johann Christian, \$d 1735-1782. \$t Works. \$f 1984 ; \$v v. 7.

Analytical entries

- Is solved differently by different agencies (or formats)
- Added entries or by listing in notes?

100 1 \$a Tolkien, J. R. R. \$q (John Ronald Reuel), \$d 1892-1973
 245 14 \$a The lord of the rings ; \$b The fellowship of the ring ; The two towers ;
 The return of the king / \$c by J.R.R. Tolkien
 740 4 \$a The fellowship of the ring
 740 4 \$a The two towers
 740 4 \$a The return of the king

100 1_ \$a Tolkien, J. R. R. \$q (John Ronald Reuel), \$d 1892-1973.
 245 14 \$a The lord of the rings / \$c by J.R.R. Tolkien.
 505 0_ \$a The fellowship of the ring ---The two towers ---The return of the king.

Both solutions can be machine-interpreted, but the use of formatted notes adds a new level of complexity

Part-names in title fields

- The use of part names and part numbers in title fields indicates the presence of a part of a whole
- Such as the “parts” of the Bible

130 0_ \$a Bible. \$p N.T. \$l Scots.
245 10 \$a The New Testament in Scots / \$c translated by William Laughton Lorimer.

130 0_ \$a Bible. \$p N.T. \$p Matthew. \$l Mountain Arapesh. \$f 2000.
245 10 \$a Enyudok iruhin ananin yopinyi barain Matyu nenyem iri.
260 __ \$a [Papua New Guinea : \$b S.I.L.], \$c 2000.

Linking entry fields

- Each linking entry field in a record will contain subfields that is used to identify the item to which the link is being made
 - by data elements
 - record number
- Different field tags represents different link semantics (76X-78X)
 - not a very extensive list, but can use the general purpose 787 + \$g
- Still a question about what entities the link is between
 - The work, expression or manifestation?
 - For some fields the anchors are ambiguous, for others not
 - Mainly used for manifestation linking

FRBR-systems

- FRBRization
 - Is the FRBR model already existing in the data we have?
 - If yes, can we create user interfaces based on this?
 - General interest in using FRBR with existing catalogues
- Many FRBR prototypes and some commercial systems that implement the model
- But no clear agenda for realizing the model in library systems
- General problem: do we know where to start and are we able to change current practise and formats?

Lessons learned (i)

- Mapping tables from MARC to FRBR is only a start
 - Many, many rules are needed for expressing when an entity and/or relationships occurs
- Some tags corresponds to FRBR entities
 - 100, 600, 700 entries are persons
 - 240, 130 indicates the work
 -
- Some relationships are often implicit in the use of tags – others are not
 - 600 person is the subject of a 240 work
 - For added entry persons in 700 we are depending on additional information such as indicators and relator codes

Lessons learned (ii)

- Advanced processing is often needed
 - Text-processing often needed to homogenize values
 - Data must be corrected and sometimes restructured
- Inconsistencies become more visible
 - Errors that nobody ever have noticed before are suddenly eye-catchers
- Requires data of high quality
- High number of rules are needed
 - Cataloguing rules are highly intricate, decoding records too
 - Have to cover current rules and current format
 - And historic versions if not converted
 - Data is sometimes different from what it should be according to the format
- To every rule for interpreting a record there is always an exception

Important issues for expressing FRBR in MARC

- Data elements
 - Cleaning up and solving ambiguity is RDA-issue
 - Representing data is MARC-related, but is it possible to rethink MARC formats when considering legacy data?
- Structural information is the key: indicators, relator codes, field linking and linking entry
 - combined with mapping tables (and mapping rules) from MARC 21 to FRBR
- FRBR requires a high level of data quality
 - consistent use and coding of values
 - more data about works, expressions, relationships,
 - correcting and cleaning up old records
- Keep structure separate from data!

Identifiers

- Things with an identity can be assigned identifiers
- Can we extend the use of identifiers to works, expressions, persons, corporate bodies?
 - URI's are considered (by many) to be important in building the global information infrastructure (e.g. semantic web applications using URI's for all abstract and concrete things)
 - New identifiers such as ISWC, ISAN, ISTC are relevant to FRBR
 - Not a format issue, but rather a question about the cataloguing process and workflow

MARC in XML

- MarcXchange (and MARCXML)
 - is to provide a tool for interchange of MARC records in XML as a supplement to interchange of MARC records in ISO2709.
 - MarcXchange is approved as international standard (ISO 25577)
- Same wine in new bottle?
 - Enables the use of a broad family of XML-tools
 - Supports “roundtripability” from XML back to MARC
 - Solve character code issues
 - Does not really introduce anything new or utilize the potential of XML

Time to leave the ISO 2709?

- The ISO2709 data format has limitations
 - Generic information structure
 - Advanced in terms of the number of different fields that can be defined, but limited in terms of complex structures (not hierarchical)
- No inherent support for entities and relationships
 - although it can be solved with indicators, relator codes and field linking
- Maybe the future is keeping MARC as it is and rethinking bibliographic records in the context of XML and FRBR?

Final remarks.....

- We do not yet have a well developed understanding of the nature of works and expressions
- Should expect many years of discussions and clarification
 - Definitions must be allowed to evolve and mature
 - Into something that easily can be applied
- On the pragmatic side:
 - It is possible to select what is important for the users
 - Advanced FRBR structures only apply to a small part of a catalogue