



FRBRización: usar los campos de relación de UNIMARC para identificar Obras

**Manolis Peponakis,
Michalis Sfakakis and
Sarantos Kapidakis**

Laboratorio de Bibliotecas Digitales y Publicaciones
Electrónicas
Departamento de Biblioteconomía y Archivística
Ionian University
Kerkira (Corfu), Greece

*Trad.: Lourdes Alonso Viana
Biblioteca Nacional de España*

Meeting:

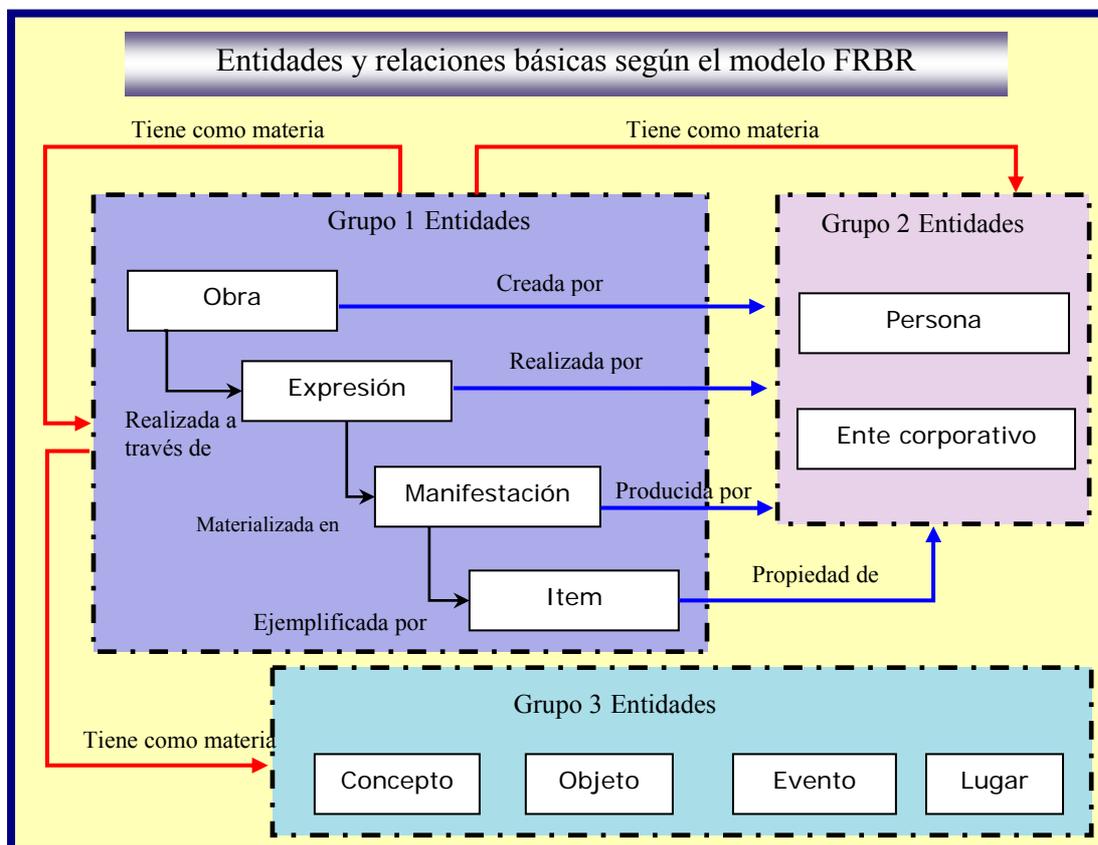
187 — Advancing UNIMARC: alignment and innovation — IFLA UNIMARC Programme (UNIMARC)

Resumen:

El objetivo principal de este estudio es fusionar las prácticas de FRBRización de MARC 21 con la semántica del formato UNIMARC y subrayar algunas diferencias entre ellos en el contexto de la FRBRización. El principal interés es examinar la posibilidad de usar los campos de relación de UNIMARC para identificar las entidades obra de los Requisitos Funcionales para Registros Bibliográficos (FRBR). En nuestro enfoque, sugerimos que todos los registros relacionados a través de los campos 45X pueden pertenecer a la misma obra que el registro que contiene dichos campos. Para probar este planteamiento hemos usado registros de autores clásicos griegos del Catálogo Colectivo de Bibliotecas Académicas Helénicas

FRBR

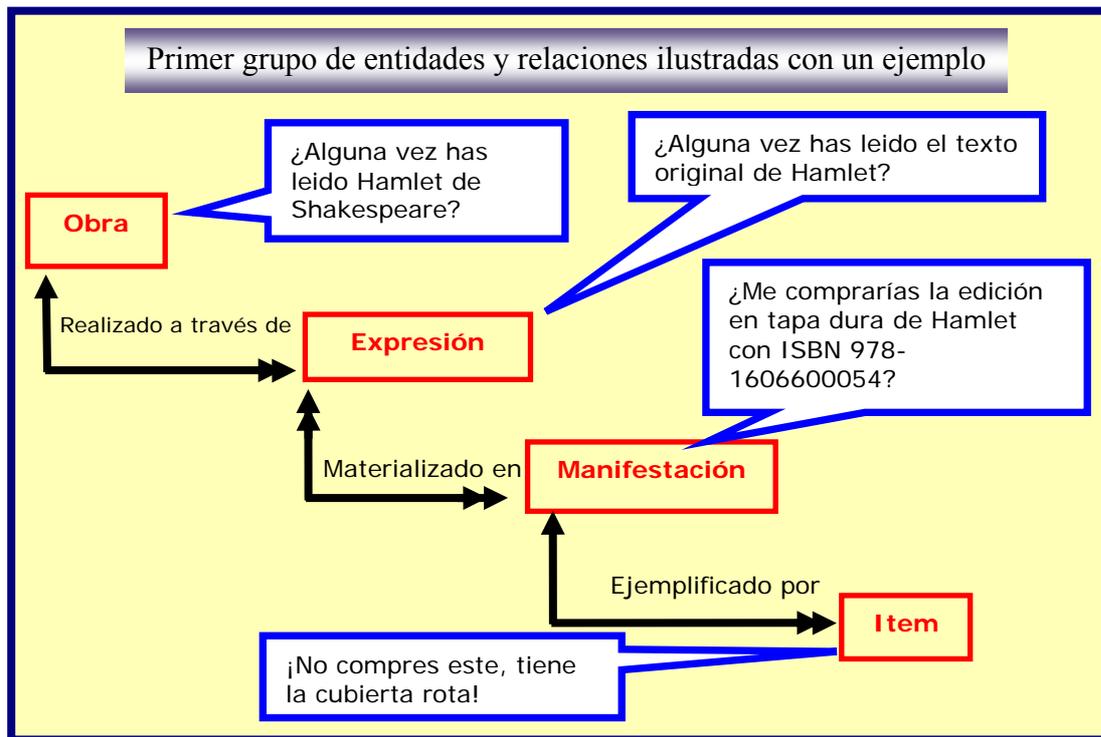
El modelo de Requisitos Funcionales para Registros Bibliográficos (FRBR) es conceptual, basado en las relaciones entre entidades y ha sido desarrollado por la IFLA. El Cuadro 1 es una representación gráfica de las relaciones básicas entre entidades.



Cuadro 1: Entidades y relaciones básicas (basado en la representación gráfica de Manguinhas et al., 2010)

FRBR es “un modelo de relaciones entre entidades como visión generalizada del universo bibliográfico, con intención de ser independiente de cualquier código o implementación de catalogación” (Tillet, 2004). No es ni un esquema de metadatos ni unas reglas de catalogación. Las normas para el Acceso y Descripción de Recursos (RDA), sucesoras de las AACR, son un código de catalogación que pone en práctica las FRBR.

Dado que el objetivo de esta ponencia es el primer grupo de entidades, el cuadro número 2 trata de dar un ejemplo sencillo de su significado.



Cuadro 2: Primer grupo de entidades y relaciones ilustrado con un ejemplo (Peponakis et al, 2010)

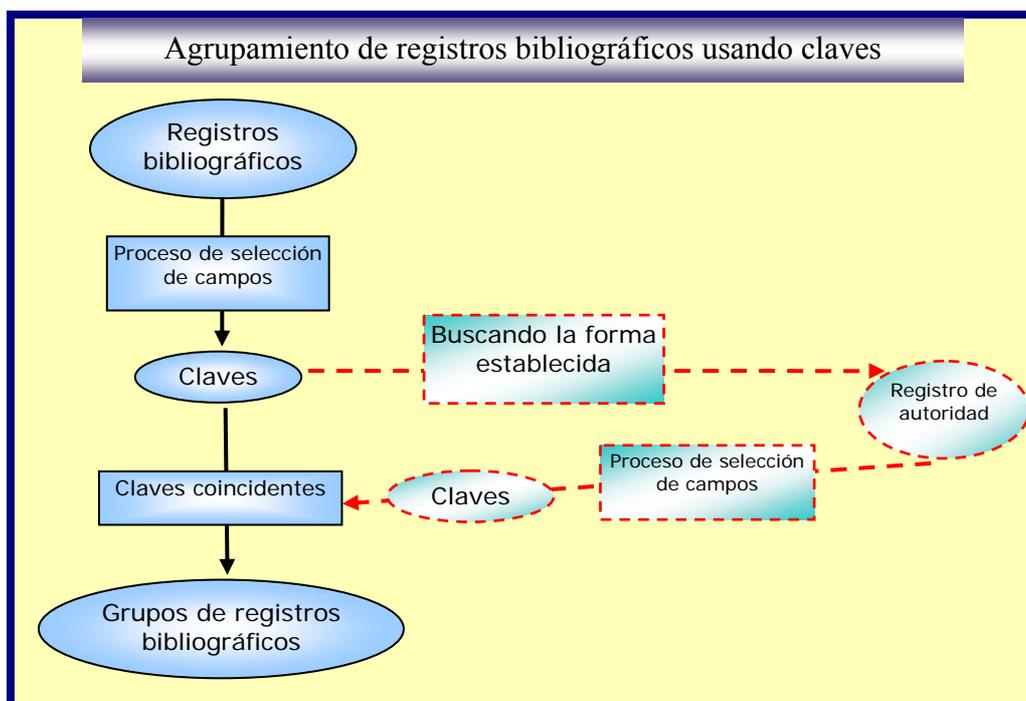
FRBRización

Es ampliamente aceptado que los catálogos tradicionales han alcanzado sus límites y, como ha sugerido Yee (2005: p.77), es esencial proceder a “un uso más inteligente de nuestros millones de registros bibliográficos, de autoridades, y de localización existentes en MARC 21 para mejorar el diseño del sistema y para FRBRizar los índices y presentaciones de nuestros OPAC”. Por esto, las bibliotecas tienen que desarrollar herramientas que sean efectivas en colecciones heterogéneas y esquemas de metadatos (Naun, 2010: p.333). Los FRBR ofrecen una percepción contemporánea de los datos bibliográficos, pero, como señala Rajapatirana (2005), “recatalogar no es una opción”. El principal desafío para las bibliotecas es, por tanto, el uso de los registros bibliográficos existentes para proporcionar servicios de valor añadido. Esta decisión lleva a inventar métodos que permitan la reconstrucción de los datos existentes en nuevos formatos.

La FRBRización es el proceso de búsqueda y síntesis de las entidades FRBR usando registros previamente catalogados-codificados en otros esquemas. Babeu (Babeu, 2008: p. 17) plantea minuciosamente que los términos *catálogo FRBR*, *sistema FRBRizado*, *implementación FRBR* son términos intercambiables que describen el proceso, pero sin tener un significado claro. La misma Babeu prefiere el término “Catálogo inspirado en FRBR” para describir el proceso de FRBRización en el contexto del proyecto Perseus.

El punto de partida de cualquier intento de FRBRización es la identificación de los registros bibliográficos que representan una *obra* y, después, la identificación dentro de ese grupo de las potenciales *expresiones* y *manifestaciones*. La identificación de las *obras* es el paso más crítico porque implica a toda la base de datos y marca los pasos siguientes. Usando los registros bibliográficos individualmente y comparándolos para que el agrupamiento tenga éxito se producen varias “claves”. La misma clave significa la misma *obra*¹. Tanto la bibliografía destacada (Aalberg 2006, Freire et al 2007, LC FRBR display tool) como la definición de *obra* de los FRBR coinciden en que hay tres elementos esenciales que una clave debe incorporar: el autor de la *obra*, el título de la *obra* y el tipo de material en el que la *obra* puede ser expresada (por ejemplo, en película o en texto).

La generación de entidades FRBR del grupo 1 está basada en claves autor-título. Hay dos métodos para generar claves. En el primero, los datos que constituyen las claves se toman directamente de los registros bibliográficos. En el segundo caso, se usa un registro de autoridad como intermediario. Una representación gráfica del proceso puede verse en el cuadro 3, donde la línea discontinua designa la mediación de un registro de autoridad para la generación de claves².



Cuadro 3: Proceso de agrupamiento de registros bibliográficos usando claves (usando o no un registro de autoridad) (Peponakis et al, 2010)

Los beneficios de este segundo enfoque son obvios ya que ofrece la posibilidad de obtener información extra derivada del registro de autoridad. Por lo tanto, es posible buscar y encontrar diferentes representaciones lingüísticas de la misma entidad, como en el caso de “Aristophanes”, que aparece también como “Aristofanis” y “Aristophanis”.

¹ Esta frase es discutible puesto que, de hecho, claves diferentes no implican necesariamente una *obra* diferente. La diferencia entre las claves puede cuantificarse usando una variedad de indicadores de similitud donde se establece un límite por encima del cual dos registros puedan considerarse como pertenecientes a la misma *obra*.

² El algoritmo OCLC incluye el registro de autoridad, pero no así la herramienta LC (<http://www.loc.gov/marc/marc-functional-analysis/tool.html>).

Partes clave: UNIMARC y MARC 21

Para la construcción de las claves hay dos elementos comunes en todos los casos, esto es, título y autor. Puede conseguirse una mayor especificación usando el tipo de registro. El algoritmo de OCLC no incluye esta última información y crea “conjuntos de *obras* FRBR” en lugar de *obras*, mientras que la herramienta de visualización de LC FRBR sí contiene el tipo de registro. En nuestro enfoque, también incluimos el tipo de registro. Nosotros, por tanto, construimos las claves usando tres partes. La primera parte es el autor, la segunda es el título y la tercera el tipo de registro. Para todas ellas hay diferencias en la semántica de UNIMARC y de MARC 21.

La diferencia crucial entre MARC 21 y UNIMARC está relacionada con si existe o no una entrada principal. En el contexto de MARC 21 la entrada principal es obligatoria. Por el contrario, en UNIMARC es opcional.

Clave de autor

Si existe el campo de autor principal (campos 700,710, 720), seleccionamos este campo. Si no existe, seleccionamos un campo según el siguiente orden (Sfakakis and Kapidakis, 2009):

- El primer campo de nombre de autor personal, por ejemplo etiqueta 701, sin subcampo \$4 o donde el \$4 tenga valor igual a “070” (es decir, código de relación para autor)
- El primer campo de entidad o reunión, por ejemplo etiqueta 711, sin subcampo \$4 o donde el \$4 tenga valor igual a “070”
- El primer campo de nombre de familia, por ejemplo etiqueta 721, sin subcampo \$4 o donde el \$4 tenga valor igual a “070”

Otra norma que mejora la deducción mediante prueba es la mención de responsabilidad (\$f en el campo 245 del MARC 21) que después selecciona las coincidencias de las formas establecidas del nombre de los campos mencionados anteriormente.

Clave de título

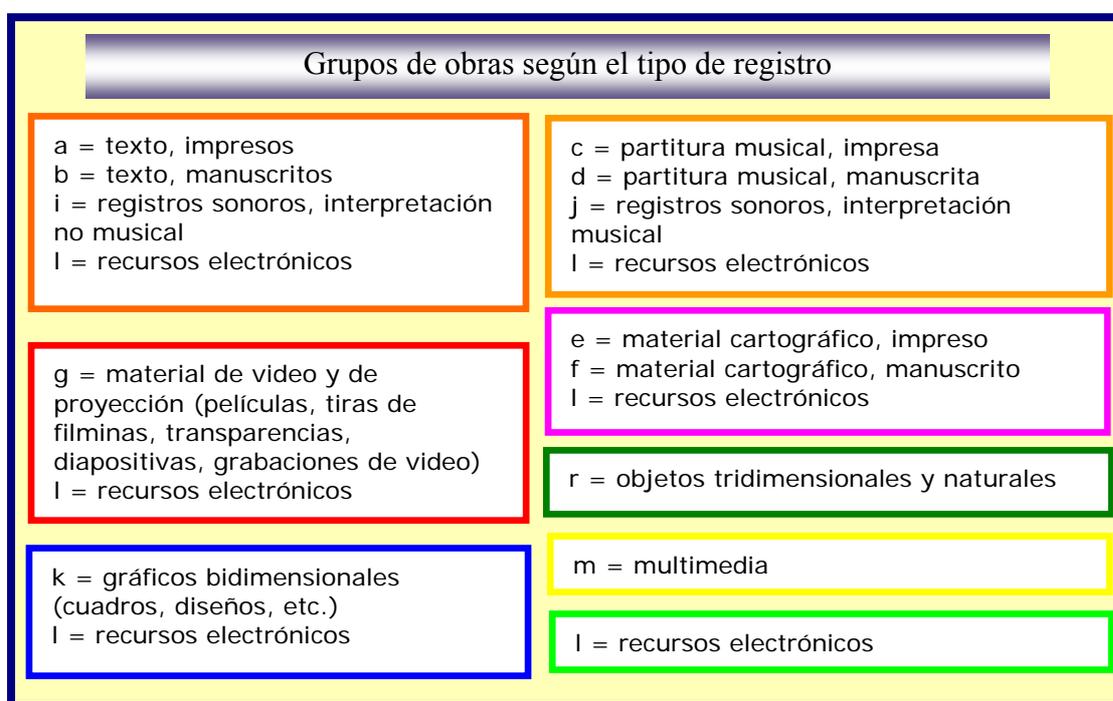
El algoritmo OCLC (Hickey and O’Neill, 2005) define el orden de selección siguiente en los campos de título:

- Título uniforme (entrada principal) (MARC21 130 => UNIMARC 500, Indicador 2 valor 1)
- Título uniforme (no es entrada principal) (MARC21 240 => UNIMARC 500, Indicador 2 valor 0)
- Traducción del título proporcionada por el centro catalogador (MARC21 242 => UNIMARC 541)
- Título principal (UNIMARC 200 => MARC21 245)
- Variantes del título (MARC21 246 => UNIMARC 517)
- Título anterior (MARC21 247 => UNIMARC 520)

Según estas definiciones, los campos de relación de UNIMARC 45X se refieren a registros considerados como diferentes *expresiones y manifestaciones* de la misma *obra*, tales como otras ediciones, traducciones y copias. La lista anterior no incluye campos de relación. Así pues, la identificación y, consecuentemente, la recuperación de los registros relacionados es una cuestión importante durante el proceso.

Tipo de identificador de registro

Como ya hemos mencionado, existen diferencias en la semántica de MARC 21 y UNIMARC en relación al identificador de registro que define el “tipo de registro”. De acuerdo con las directrices de UNIMARC para recursos electrónicos, hay una opción al catalogar material digitalizado (un mapa, por ejemplo) de usar el valor de identificador de registro de recurso electrónico (en lugar de mapa impreso). Basándonos en esto, hemos usado el valor “l = recursos electrónicos” en varios grupos como puede verse más abajo. Por otro lado, en MARC 21 está claramente definido que “las clases de recursos electrónicos se codifican según su aspecto más significativo (por ejemplo, texto, gráfico, material cartográfico, sonido, música, imagen en movimiento)”. Para que todo se aglutine, es decir, todo esté bajo la misma obra, sugerimos la siguiente agrupación para registros con diferentes identificadores: como se muestra en el cuadro 4, los registros con diferentes valores para tipo de registro pueden pertenecer o no a la misma obra (véanse ejemplos con los registros 4, 6 y 7 en el cuadro 5)



Cuadro 4: Sugerencia de agrupación basada en identificador de registro

Un ejemplo

Según las FRBR los tres registros siguientes pertenecen a la misma obra que consiste en dos *expresiones* y tres *manifestaciones*. Basándonos en lo expuesto, la clave que reunirá todos los registros siguientes bajo la misma *obra* será “autor = HOMERO – título = LA ILIADA – tipo de recurso = TEXTO”

Registro 1 - Libro	
Título / Autor	The Iliad / Homer ; translated by E.V. Rieu
Publicación	Harmondsworth : Penguin Books , 1954
Descripción física	xxv, 466 p., 20 cm.
Título uniforme	La Iliada
Autor	Homero
Traductor	Rieu, Emile Victor, 1887-1972
Idioma del texto	Inglés
Registro 2 - Libro	
Título / Autor	The Iliad / Translated by E. V. Rieu
Publicación	Baltimore : Penguin Books , [1964, c1950]
Descripción física	469 p., 18 cm.
Título uniforme	La Iliada
Autor	Homero
Traductor	Rieu, Emile Victor, 1887-1972
Idioma del texto	Inglés
Record 3 - Book	
Título / Autor	Ομήρου Ιλιάδα / μετάφραση Ν. Καζαντζάκη, Ι. Θ.Κακριδή
Publicación	Αθήνα : Εστία, [1997]
Descripción física	401 σ., 22 εκ.
Título uniforme	La Iliada
Autor	Homero
Traductor	Καζαντζάκης, Νίκος ; Κακριδής, Ιωάννης Θ.
Idioma del texto	Griego moderno

Tabla 1: Los tres registros constituyen una *obra*, dos *expresiones* y tres *manifestaciones*³

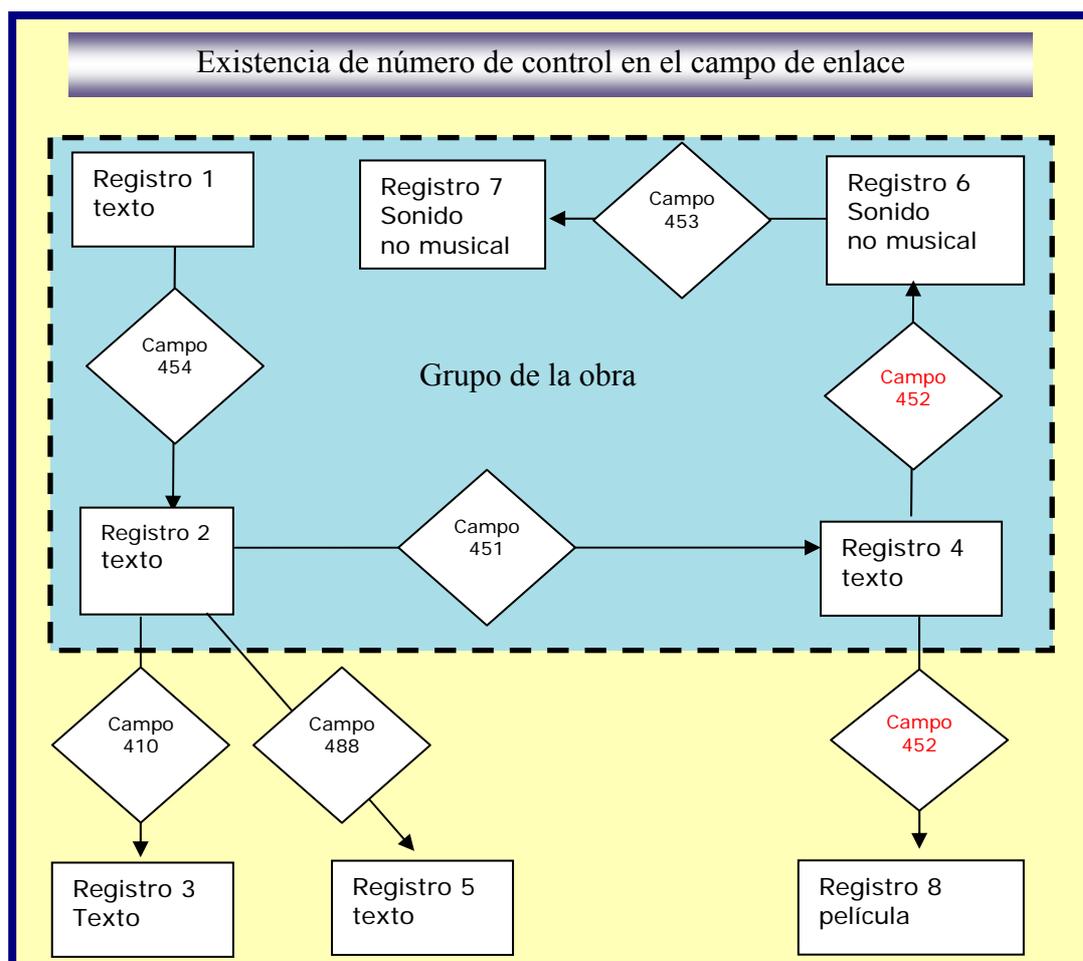
Construyendo campos de relación

Teniendo en cuenta el hecho de que UNIMARC permite tanto la existencia como la ausencia de un número de control del registro hacia el que enlazar, hemos tratado con cada opción de forma separada. Principalmente, la existencia (o no) del número de control está relacionada con la técnica de enlace empleada. Normalmente, en los casos de campos integrados, el número de control existe mientras que no existe en los casos de técnica estándar.

³ La *obra* es el texto de La Iliada de Homero, la primera *expresión* es la traducción inglesa de Rieu (registros 1 y 2) y la segunda *expresión* es la traducción al griego moderno de Kazantzakis y Kakridis (registro 3). Cada registro representa una *manifestación* diferente.

Registros UNIMARC con campos relacionados que integran el campo 001

En el caso de un número de control existente de un registro relacionado, se considera que todos los registros relacionados con los campos 45X pertenecen a la misma *obra* si el identificador de registro lo permite, sin importar el resultado de la implementación clave. En el caso de grupos de identificadores de registros diferentes, constituyen *obras* diferentes, pero, aún así, relacionadas. Por ejemplo, en el cuadro 5 el registro 4 enlaza con los registros 6 y 8, pero sólo los registros 4 y 6 pertenecen a la misma obra. El registro 8 tiene un tipo de registro diferente (para los grupos de tipos de registro véase el cuadro 4).



Cuadro 5: Campos 001 integrados. El fondo azul claro (dentro de la línea discontinua) es la *obra*.

Registros UNIMARC con campos de relación que no integran el campo 001

En este caso, los datos de los campos de relación pueden usarse para generar claves. Hemos observado que en la técnica estándar, la información del campo 45X es más formal que en el 200\$a. En realidad, no existe una descripción de una *manifestación* específica, sino más bien un título más formal (cercano al título uniforme). Así pues, incluso en el caso de los campos 451, 452, 455 y 456 es más efectivo usar estos campos que el 200\$a.

Para definir el orden de selección de los campos de relación hemos considerado el campo de idioma 101 (especialmente en el caso de 453 “traducido como” y 454 “traducción de”). Si el indicador era 1 “ítem como una traducción de la obra original o una obra intermedia”, el campo 454 “traducción de” se situaba justo debajo del título uniforme. Si el indicador era 0 “ítem en el idioma original de la obra”, no usábamos el campo 453.

Evaluación del resultado de añadir campos de relación para generar claves

Para fortalecer la hipótesis de que relacionar campos puede usarse para incrementar la efectividad de la recuperación hicimos un experimento. Usamos como muestra registros del *Catálogo Colectivo de Bibliotecas Académicas Helénicas*. Este catálogo es una base de datos UNIMARC con más 3.500.000 registros de 54 bibliotecas. Algunas de las características principales de esta base de datos son los datos multilingües, la ausencia de registros de autoridad comunes y las diferentes políticas de catalogación de los colaboradores.

Seleccionamos *obras* de autores clásicos griegos porque dichas obras tienen tanto *expresiones* como *manifestaciones* y constituyen un área ideal para comprobar la efectividad de los algoritmos de FRBRización. Para evitar resultados controvertidos, excluimos manualmente aquellos registros que representaban fragmentos de *obras* u *obras* encuadernadas juntas en un único volumen.

Dado que la política de uso de campos de relación del *Catálogo Colectivo de Bibliotecas Académicas Helénicas* no implica el 001, aplicamos el método únicamente donde los campos de relación se utilizan para construir claves. Primero, empleamos el algoritmo OCLC ligeramente modificado⁴ para verificar la efectividad del procedimiento de FRBRización con nuestro conjunto de datos. La observación más importante fue la baja recuperación, el algoritmo recogía sólo un número limitado de registros. Hasta donde pudo precisarse, el algoritmo parecía funcionar correctamente.

La muestra consistía en 307 registros pertenecientes a 12 obras. Así pues, el éxito total sería un resultado de 12 claves generadas. El título más significativo para la identificación de la *obra* es el campo de título uniforme. Desafortunadamente, como podemos observar en el gráfico 1, sólo la mitad de los registros tenían dicho campo. Concretamente, los 307 registros (100%) tenían el campo 200; 168 registros (54,7%) tenían el campo 500; 5 (1,6%) el campo 510; 42 (13,6%) el 517; 9 (2,9%) tenían el 540; y 19 registros (6,1%) tenían campos 45X. Sólo 3 registros (0,97%) tenían tanto título uniforme como campos 45X.

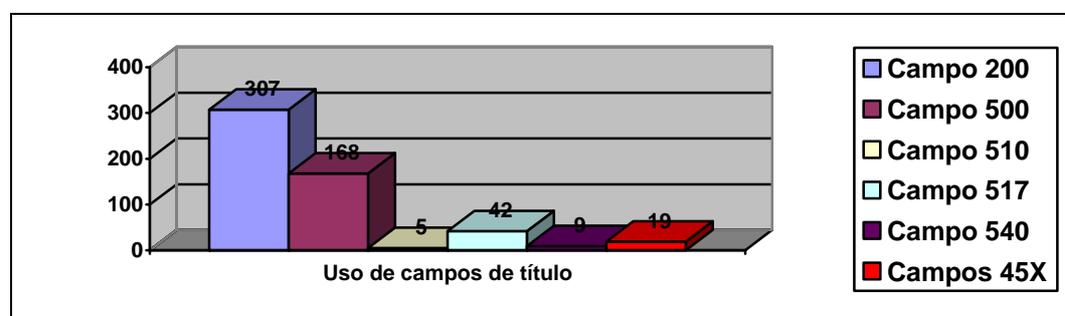


Gráfico 1: Distribución de campos de título

⁴ No hemos usado registros de autoridad y nuestros metadatos estaban en UNIMARC, en lugar de en MARC 21.

Para evaluar la efectividad de nuestro algoritmo, realizamos un agrupamiento mediante enlace único en dos conjuntos de claves de *obra* extraídos de nuestros registros de prueba. El primer grupo consistía en claves generadas sin usar campos de enlace relación (basado en OCLC), mientras que el segundo usaba campos de relación como los definidos en el anterior apartado. Al agrupar, obtuvimos 85 grupos del primer conjunto y 78 del segundo.

El uso de campos de relación mejora la efectividad de la síntesis de la *obra* alrededor de un 9%. Comparando los números de grupos resultantes no obtenemos una indicación precisa de la efectividad del proceso en general. En nuestro caso, en aquellos en los que se comprobó el contenido de los conjuntos y sólo alguno contenía registros similares, vimos que la diferencia entre los dos enfoques de obtener grupos adicionales es 0,9. Es más, la mejora se confirmó también con el uso de medidas para evaluar la agrupación y el promedio de agrupamiento tales como el *índice RAND corregido* y *silhouette width information*. El *índice RAND* mide el porcentaje de decisiones correctas (coincidencias de claves correctas), mientras que el *RAND corregido* aumenta la sensibilidad de dicha medida. *Silhouette width* evalúa cuán satisfactoriamente está siendo agrupada una clave, es decir, si está situada en el grupo correcto. Concretamente, los valores para el *índice RAND corregido* y el promedio de *silhouette width information* eran 0,56 y 0,81 respectivamente en el primer conjunto de claves de *obra*; mientras que los valores de agrupación en el segundo, usando campos de relación, mejoraron a 0,61 y 0,83 respectivamente. El *índice RAND* se acerca más a nuestra estimación, mientras que la existencia de muchos grupos de ejemplares únicos afecta a la alta mejora de la *silhouette width information*.

Conclusiones y trabajo futuro

Primero y más destacado tiene que quedar constancia de que en ocasiones no hemos trabajado con *obras*, sino con conjuntos de *obras*. Aunque similares a los conjuntos de OCLC, en nuestro caso con una diferencia significativa: se distinguen más explícitamente si se incluye el identificador de registro. Además, como se señala en las FRBR “el concepto de lo que constituye una *obra* y dónde descansa la línea de separación entre una *obra* y otra puede de hecho verse diferente entre una cultura u otra. Consecuentemente, las convenciones bibliográficas establecidas por varias culturas o grupos nacionales pueden diferir según el criterio usado para determinar las fronteras entre una y otra *obra*” [FRBR (p. 16)].

Los resultados revelan un ratio bajo de recuperación incluso añadiendo campos de relación para generar de claves. La razón principal para este pobre resultado es la ausencia de títulos uniformes (campo 500) en combinación con la gran diversidad de títulos principales existentes (campo 200). Entre los 307 registros, había 141 títulos principales únicos (campo 200) mientras que, como muestra el gráfico 1, había 550 campos de título en total. Si usamos sólo un campo de cada registro, se obvian aparentemente 243 títulos relevantes, siendo esta una cantidad casi igual a los datos totales usados realmente. Centrándonos en un incremento significativo del ratio de recuperación, vamos a usar también estos datos, esto es, los campos de título ignorados previamente para identificar las *obras*, de modo que en lugar de seleccionar sólo un campo de título, compararemos todos los campos entre sí.

Referencias bibliográficas

- Aalberg, T. (2006). A Tool for Converting from MARC to FRBR. In: *ECDL 2006, Alicante, Spain, 17-22 September 2006*. Gonzalo, J. et al. (eds.) Berlin, Heidelberg: Springer, pp. 453–456. Disponible en <http://www.springerlink.com/content/5356711834963732/fulltext.pdf>. [último acceso 29/05/2011].
- Babeu, A. (2008). Building a "FRBR-Inspired" Catalog: The Perseus Digital Library Experience. [Internet] Perseus Digital Library. Disponible en <http://www.perseus.tufts.edu/~ababeu/PerseusFRBRExperiment.pdf>. [Último acceso 29/05/2011].
- Freire, N., Borbinha, J. and Calado, P. (2007). Identification of FRBR Works Within Bibliographic Databases: An Experiment with UNIMARC and Duplicate Detection Techniques. In: *ICADL 2007, Hanoi, Vietnam, 10-13 December 2007*. Berlin, Heidelberg: Springer, pp. 267–276. Disponible en <http://www.springerlink.com/content/d06r28v440n1x420/>.
- Hickey, T.B. and O'Neill, E.T. (2005). FRBRizing OCLC's WorldCat. *Cataloging & Classification Quarterly*. 39 (3/4), pp. 239-251.
- IFLA (1998). Functional Requirements for Bibliographic Records. Disponible en <http://www.ifla.org/VII/s13/frbr/frbr.pdf>. [Último acceso 29/05/2011].
- LC FRBR Display Tool (The Library of Congress' Network Development and MARC Standards Office) <http://www.loc.gov/marc/marc-functional-analysis/tool.html>.
- Manguinhas, H., N. Freire, and J. Borbinha. "FRBRization of MARC records in multiple catalogs." In *Proceedings of the ACM International Conference on Digital Libraries*, 225-234, 2010
- Naun, C.C. (2010) "Next generation OPACs: A cataloging viewpoint." *Cataloging and Classification Quarterly* 48 (4), pp. 330-342.
- Peponakis, M.; Sfakakis, M.; Kapidakis, S. (2010) "FRBRization: Seeking for the "key" to Works' Identification" (written in Greek). En *Proceedings of the 19th Hellenic Conference of Academic Libraries*. Disponible en http://library.panteion.gr/19libconf/conference_en.php [Último acceso 29/05/2011]
- Rajapatirana, B. and Missingham, R. (2005). The Australian National Bibliographic Database and the Functional Requirements for the Bibliographic Database (FRBR). *The Australian Library Journal*. 54 (1), pp. 31-42. Disponible en <http://www.alia.org.au/publishing/alj/54.1/full.text/rajapatirana.missingham.html>. [Último acceso 29/05/2011]
- Sfakakis, M. and Kapidakis, S. (2009). Eliminating query failures in a work-centric library meta-search environment. *Library Hi Tech*. 27 (2), pp. 286-307

Tillett, B. (2004). What is FRBR? A conceptual model for the bibliographic universe. [Internet]. Disponible en <http://alia.org.au/publishing/alj/54.1/full.text/tillett.html> . [Último acceso 29/05/2011].

Yee, M.M. (2005). FRBRization: a Method for Turning Online Public Finding Lists into Online Public Catalogs. *Information Technology and Libraries*. 24 (3), pp. 77-95. Post-impresion disponible en <http://repositories.cdlib.org/postprints/715/>. [Último acceso 29/05/2011]