

UNIMARC: Estudo da sua utilização na PORBASE

*Maria Margarida Lopes, Nuno Freire,
Dulce Fontes*

Biblioteca Nacional de Portugal
Campo Grande, 83, 1741-081 Lisboa
E-mail: mmlopes@bn.pt, nuno.freire@bn.pt,
dfontes@bn.pt

Hugo Manguinhas, José Borbinha

INESC-ID
Rua Alves Redol 9, Apartado 13069, 1000-029 Lisboa
E-mail: hugo.manguinhas@ist.utl.pt, jlb@ist.utl.pt

RESUMO

O UNIMARC é o formato normalizado utilizado na PORBASE para troca de dados bibliográficos com as bibliotecas cooperantes, sendo também utilizado pelos sistemas de gestão documental de grande parte das bibliotecas em Portugal. É uma norma complexa, da família de formatos MARC – MACHine-Readable Cataloging, que define um elevado número de elementos tanto para uso generalizado como para utilização específica. Este artigo apresenta uma análise da frequência da utilização dos elementos do UNIMARC, campos e subcampos, levada a cabo sobre a totalidade dos registos da PORBASE. Apresenta-se também uma comparação com os resultados de um estudo semelhante publicado para o formato MARC21 numa amostra de registos da OCLC, em que se detectam semelhanças relevantes, o que pensamos valida assim ambos os estudos. Espera-se que as conclusões desta análise sejam importantes para a manutenção e evolução das Regras Portuguesas de Catalogação, das recomendações e opções internas da PORBASE, e ainda para as actividades relacionadas da IFLA, especialmente do programa ICABS (IFLA-CDNL Alliance for Bibliographic Standards), em que a BNP (Biblioteca Nacional de Portugal) tem um papel relevante como coordenadora das actividades UNIMARC.

PALAVRAS-CHAVE: UNIMARC, PORBASE, metadados, catalogação

INTRODUÇÃO

O formato MARC – MACHine-Readable Cataloging foi concebido e desenvolvido em 1965 pela Biblioteca do Congresso, com a finalidade de criar um formato legível por computador de armazenamento de informação bibliográfica. Entretanto, foram surgindo variantes nacionais do formato equivalentes a diferentes práticas de catalogação, dificultando assim a troca internacional de registos bibliográficos [1].

O UNIMARC – The Universal MARC Format, foi criado pela IFLA em 1977 com o objectivo de facilitar a troca internacional de registos, resolvendo assim o problema da incompatibilidade entre os diversos formatos MARC nacionais¹. Vários países que não chegaram a desenvolver formatos próprios adoptaram o UNIMARC como formato nacional.

Um desses casos foi a BNP – Biblioteca Nacional de

Portugal. A PORBASE, constituída como o Catálogo Colectivo das Bibliotecas Portuguesas, adoptou o UNIMARC como formato normalizado desde a criação de rede em 1986². É uma base de dados cooperativa, resultante da contribuição de mais de 160 catálogos das diversas bibliotecas e centros de documentação que nela participam e constitui, por isso, um universo de estudo privilegiado pela sua abrangência, reflectindo as práticas de catalogação a um nível nacional e, simultaneamente, as políticas de tratamento bibliográfico desenvolvidas pela BNP.

O formato tem sofrido uma evolução constante com a adição de novos campos e a actualização dos já existentes, com a adição de novos subcampos, como se resume na Tabela 1. Tal tem acontecido por várias motivações: para acompanhar quer a evolução das ISBD, para acompanhar o avanço da tecnologia, e ainda para permitir a criação de novos pontos de acesso para melhor recuperação da informação [2].

Análise recentes levadas a cabo pelo Texas Center for Digital Knowledge, University of North Texas, no qual foi analisada uma amostra de 400.000 registos MARC21 fornecida pela OCLC [3], revelaram que mais de 50% dos campos ou subcampos não são utilizados em qualquer registo, e que apenas 36 campos ou subcampos concentram 80% de toda a utilização do formato [4]. Foi a publicação desse estudo e contactos posteriores com os respectivos autores que nos motivaram para a realização de um estudo semelhante para o formato UNIMARC Bibliográfico na PORBASE.

Blocos	Campos por bloco		
	2ª ed. (1980)	2ª ed. 4º update (2002)	2ª ed. 5º update (2005)
0--	9	18	20
1--	20	23	24
2--	9	10	10
3--	30	34	34
4--	26	36	40
5--	16	18	18
6--	17	20	20
7--	9	11	11
8--	2	6	6
Total	138	176	183

Tabela 1: Evolução do formato UNIMARC Bibliográfico

¹ <http://www.unimarc.info>

² <http://www.porbase.org>

Um esquema de metadados como o UNIMARC tem como preocupação essencial identificar correctamente a informação, de modo a que a sua recuperação seja a mais correcta possível e que os sistemas informáticos indexem correctamente, e sem ruído, toda a informação disponível. Apresenta para isso uma estrutura coerente de etiquetas, indicadores e códigos de subcampos, agrupando a informação em blocos lógicos que correspondem aos critérios de identificação e de pesquisa bibliográficos.

No estudo aqui apresentado pretendemos realizar uma análise da utilização dessa estrutura, verificando a frequência com que os campos e os subcampos definidos pelo formato UNIMARC ocorrem em toda a PORBASE. Ou seja, pretendemos perceber com que grau de diversidade se usa a riqueza dos elementos disponíveis nos processos de catalogação.

Este estudo pretende fornecer ainda pistas da utilização do UNIMARC Bibliográfico que possam ser úteis na avaliação dos custos/benefícios associados à evolução do formato, tal como feito também para o MARC21 [5]. É-nos assim possível também identificar, ainda que empiricamente, os elementos nucleares da descrição bibliográfica, como se propõe no final. O propósito é colaborar desta forma para os esforços de normalização e a redefinição das políticas de cooperação na PORBASE.

METODOLOGIA

Os resultados aqui apresentados reflectem o estado da PORBASE no dia 15 de Janeiro de 2007, dizendo respeito apenas aos registos bibliográficos. Foram considerados como universo deste estudo, apenas os registos assinalados como concluídos, excluindo desta forma os registos ainda em fase de processamento.

O total de registos bibliográficos identificados para análise foi assim de 1.225.357. Todos esses registos, assim como a sua história desde o dia 1 de Outubro de 2005, estão preservados no sistema REPOX instalado na BNP.

A versão do UNIMARC Bibliográfico que serviu de base à nossa análise estatística foi a 4ª revisão da 2ª edição (2002) [6], por considerarmos que é a versão que mais se aproxima da que está efectivamente implementada no sistema de gestão bibliográfica HORIZON, que suporta a PORBASE. Nesta versão, o formato define 176 campos e 1.194 subcampos, conforme ilustrado na Tabela 2. Os campos do bloco 9--, por serem de definição e uso local, foram ignorados para a análise.

Bloco	Campos	Subcampos
0-- Bloco de Identificação	18	57
1-- Bloco de Informação Codificada	23	70
2-- Bloco de Informação Descritiva	10	45
3-- Bloco de Notas	34	76
4-- Bloco de Entradas Relacionadas	36	612
5-- Bloco de Títulos Relacionados	18	91
6-- Bloco de Assuntos	20	125
7-- Bloco de Responsabilidade Intelectual	11	81
8-- Bloco de Uso Internacional	6	37
Total	176	1194

Tabela 2: Campos e subcampos existentes no UNIMARC bibliográfico (4ª revisão da 2ª edição)

Para cada um dos elementos (campo e subcampo) foram identificados os registos bibliográficos em que os mesmos ocorrem, assim como a sua frequência (uma vez que existem elementos que se podem repetir).

Para o efeito, os registos bibliográficos foram ainda agrupados por quatro tipos, de acordo com os seguintes critérios, definidos a partir da etiqueta de registo:

- Monografias: tipo de registo (posição do carácter 6) 'a' e nível bibliográfico (posição do carácter 7) 'm';
- Publicações periódicas: tipo de registo (posição do carácter 6) 'a' e nível bibliográfico (posição do carácter 7) 's';
- Documentos electrónicos: tipo de registo (posição do carácter 6) 'l' e nível bibliográfico (posição do carácter 7) qualquer um;
- Outros – restantes conjunções

ANÁLISE

O total de registos bibliográficos da PORBASE analisados foi de 1.225.357, correspondentes à ocorrência de um total de 33.757.510 elementos descritivos (campos e subcampos), distribuídos como se descreve na Tabela 3.

Tipo de documento	Número de registo na PORBASE	Ocorrências de campos ou subcampos
Monografias	1.062.275	28.986.847
Publicações Periódicas	68.342	1.719.040
Documentos Electrónicos	822	30.282
Outros	93.918	3.021.341
Total	1.225.357	33.757.510

Tabela 3: Caracterização dos registos bibliográficos na PORBASE segundo o tipo de documento.

Bloco	Campos por bloco	Campos não utilizados		Campos com utilização inferior a 1%		Campos com utilização entre 1% e 50%		Campos com utilização superior a 50%	
0--	18	22%	4	44%	8	22%	4	11%	2
1--	23	13%	3	61%	14	13%	3	13%	3
2--	10	0%	0	40%	4	30%	3	30%	3
3--	34	6%	2	65%	22	29%	10	0%	0
4--	36	6%	2	94%	34	0%	0	0%	0
5--	18	0%	0	78%	14	22%	4	0%	0
6--	20	35%	7	50%	10	10%	2	5%	1
7--	11	18%	2	36%	4	36%	4	9%	1
8--	6	33%	2	50%	3	0%	0	17%	1
Total	176		22		113		30		11

Tabela 4: Utilização verificada dos campos nos registos da PORBASE, por bloco.

Em todos os blocos do UNIMARC, o grau mais frequente de utilização dos campos situa-se abaixo de 1%.

Esta baixa frequência de utilização é especialmente significativa no caso dos blocos 4-- e 6--. No caso do bloco 4-- todos os campos, sem excepção, são usados em menos de 1% dos registos; no caso do bloco 6--, apenas 15% dos campos são utilizados em mais de 1% dos registos.

No entanto, é necessário ter em conta que, nestes dois casos, estes valores são inferiores àquilo que é efectivamente a prática de catalogação nacional. No caso do bloco 4-- tal acontece porque não são criados na PORBASE registos de analíticos, nem integrados os registos de analíticos provenientes dos cooperantes³. No caso do bloco 6--, tal ocorre porque no momento da integração dos registos na PORBASE o conteúdo destes campos é eliminado dos registos⁴.

Os blocos 1--, 2-- e 8-- são aqueles que apresentam valores mais significativos no intervalo acima dos 50% de utilização: no caso do bloco 1--, 13% dos campos são usados em mais metade dos registos; para o bloco 2-- há 30% de campos com este grau de utilização; para o bloco 8--, este valor situa-se nos 17%.

Quer o bloco 1--, quer o 8--, têm diversos campos passíveis de preenchimento ou correcção automática pelos sistemas, que julgamos poder contribuir largamente para este resultado. Parece-nos por isso importante salientar também o bloco 7--, em que 45% dos campos são usados em mais de 1% dos registos,

³ Esta decisão deve-se ao facto de, tratando-se de um bloco de entradas relacionadas, a sua consistência ser muito difícil de gerir numa base cooperativa onde se têm de integrar registos gerados em sistemas diferentes.

⁴ Esta prática aplica-se a todo o bloco, com excepção do campo 675, mantendo-se apenas os casos em que a linguagem de indexação utilizada está identificada como SIPOR (Siporbase) ou CLIP, e decorre do facto de se considerar que, nos restantes casos, os cooperantes utilizam linguagens de indexação diferentes e não consolidadas entre si (o objectivo da iniciativa CLIP).

reflectindo, este assim, uma prática efectiva de catalogação.

Ocorrência	Todos os Tipos	Monog.	Public. Períod.	Doc. Elect.	Outros
0 Utilizações	22	49	60	118	52
[00,0%-00,5%[103	89	75	10	66
[00,5%-01,0%[10	6	8	6	13
[01,0%-05,0%[14	7	7	13	12
[05,0%-10,0%[8	5	4	4	13
[10,0%-25,0%[8	8	3	5	7
[25,0%-50,0%[0	1	7	8	2
[50,0%-75,0%[2	2	4	4	1
[75,0%-100%[9	9	8	8	10
Total	176	176	176	176	176

Tabela 5: Frequência de utilização dos 176 campos UNIMARC.

Refinando estes resultados, relativamente à frequência de campos preenchidos, os registos bibliográficos analisados apresentam o cenário descrito na Tabela 5.

A primeira conclusão é que 22 campos (12,5%) não possuem nunca qualquer ocorrência, independentemente do tipo de material descrito nos registos. Esses campos são:

- 014, 016, 072, 073
- 117, 131, 140
- 332, 334
- 437, 462
- 608, 615, 616, 626, 660, 670, 680
- 716, 730
- 850, 886

No entanto, se considerarmos o número de campos que foram utilizados em menos de 1% dos registos, este é de 135 (76,7% de todos os campos). Para as monografias o número é ainda maior, de 144 campos, sendo para as publicações periódicas de 143. Isto é, grande parte da descrição de monografias é feita recorrendo-se apenas a 32 campos, e de publicações periódicas a 33 campos.

Quanto aos campos mais utilizados, são 11 (6,3% do

total) os campos preenchidos em mais de 50% dos registos (independentemente do tipo de material). São os seguintes:

- 001, 005
- 100, 101, 102
- 200, 210, 215
- 675
- 700
- 801

O conjunto acima referido é constituído pelos campos que representam a estrutura mais básica de um registo bibliográfico (autor, título, dados de publicação, descrição física e classificação), ou a campos obrigatórios que são preenchidos automaticamente pelo sistema, como é o caso dos campos do bloco de identificação.

No caso do campo 801, um campo obrigatório do UNIMARC, a elevada taxa de utilização não corresponde, nem ao efectivo preenchimento do campo pelo catalogador, nem ao preenchimento automático pelo sistema, mas ao resultado de trabalhos de correção sistemática que têm sido feitos recentemente sobre a PORBASE, pois trata-se de um campo invariavelmente ignorado.

No caso do campo 100, a sua utilização é de praticamente 100%, pois na maior parte dos sistemas o seu preenchimento é automático em algumas posições. No entanto, vale a pena referir que a experiência com os registos dos cooperantes da PORBASE mostra que em cerca de metade dos registos esse preenchimento não é devidamente completado.

Esta análise global pode ainda ser complementada por uma abordagem mais detalhada por tipo de material. Por serem os únicos grupos de material com casos relevantes em quantidade e perfeitamente identificados, os dados seguintes irão focar-se nos registos de monografias e de publicações periódicas.

REGISTOS DE MONOGRAFIAS E DE PERIÓDICOS

No universo do estudo, as monografias representam 1.062.275 do total dos registos bibliográficos e as publicações periódicas 68.342.

Na Tabela 6 podemos ver que são apenas 20 os campos preenchidos em pelo menos 10% desses registos (os campos aparecem ordenados pela sua percentagem de ocorrência).

Na Tabela 7 podemos verificar que neste caso são 22 os campos preenchidos em pelo menos 10% desses registos (os campos aparecem ordenados pela sua percentagem de ocorrência).

Campo	Designação	Registos	%
001	Identificador do Registo	1.062.275	100,00%
100	Dados Gerais de Processamento	1.062.243	100,00%
200	Tít. e Menção de Responsab.	1.062.148	99,99%
101	Língua de publicação	1.060.525	99,84%
210	Publicação, Distribuição, etc.	1.059.481	99,74%
215	Descrição Física	1.053.987	99,22%
102	País de Publicação	1.051.054	98,94%
801	Fonte de origem	966.143	90,95%
005	Identificador da Versão	931.623	87,70%
700	Autor (resp. principal)	785.887	73,98%
675	CDU	726.549	68,40%
225	Colecção	273.382	25,74%
010	ISBN	263.599	24,81%
106	Mat. Textual - Atributos físicos	262.558	24,72%
702	Autor (resp. secundária)	259.914	24,47%
105	Dados Codificados: Material textual - Monografias	225.060	21,19%
205	Menção da Edição	168.927	15,90%
021	Número do Depósito Legal	138.403	13,03%
710	Colectividade-autor (resp. principal)	119.751	11,27%
701	Autor (co- resp. principal)	112.963	10,63%

Tabela 6: Campos preenchidos em mais de 10% dos registos de monografias.

Campo	Designação	Registos	%
001	Identificador do Registo	68.342	100,00%
100	Dados Gerais de Processamento	68.341	100,00%
200	Tít.e Menção de Responsabilidade	68.333	99,99%
210	Publicação, Distribuição, etc.	68.115	99,67%
101	Língua de publicação	68.083	99,62%
102	País de Publicação	67.793	99,20%
005	Identificador da Versão	65.807	96,29%
801	Fonte de origem	64.341	94,15%
215	Descrição Física	44.800	65,55%
712	Colectividade-autor (resp. secundária)	41.486	60,70%
207	Publicações em Série - Numeração	38.804	56,78%
675	CDU	38.571	56,44%
110	Dados Codificados: Publicações em Série	32.915	48,16%
326	Nota de Periodicidade (Publ. Série)	26.481	38,75%
106	Material Textual - Atributos Físicos	25.125	36,76%
620	Acesso pelo local	24.093	35,25%
303	Notas Gerais - Informação Descritiva	22.649	33,14%
702	Autor (resp. secundária)	19.913	29,14%
606	Nome comum usado como assunto	18.681	27,33%
011	ISSN	15.939	23,32%
530	Título chave (Publicações em série)	12.889	18,86%
802	Centro ISDS	8.243	12,06%

Tabela 7: Campos preenchidos em mais de 10% dos registos de periódicos.

ANÁLISE POR SUBCAMPOS

Procedendo a uma análise semelhante à apresentada acima, mas ao universo dos subcampos, obtemos a Tabela 8, Tabela 9 e Tabela 10.

Ocorrência	Todos os tipos	Monografias	Publicações Periódicas
0 Utilizações	590	753	809
[00,0%-00,5%[487	343	303
[00,5%-01,0%[29	20	18
[01,0%-05,0%[31	27	20
[05,0%-10,0%[19	9	6
[10,0%-25,0%[15	18	5
[25,0%-50,0%[3	4	13
[50,0%-75,0%[7	7	10
[75,0%-100%[13	13	10
Total	1194	1194	1194

Tabela 8: Frequência de utilização dos 1194 subcampos UNIMARC.

A primeira conclusão é que praticamente metade dos subcampos disponíveis no UNIMARC nunca foi utilizada na PORBASE (590 subcampos, ou seja, 49% do total). Além disso, 41% dos subcampos têm uma utilização inferior a 0,5%. Ou seja, apenas cerca de 10% dos subcampos são utilizados em mais de 0,5% dos registos, independentemente do tipo de material.

Analisando especificamente as monografias, o número de subcampos nunca utilizados é de 753, ou seja, 63%. Com uma utilização inferior a 0,5% dos registos encontramos 92% dos subcampos. Isto significa que a descrição de monografias é feita utilizando praticamente apenas 98 dos 1194 subcampos disponíveis.

No caso das publicações periódicas, o número de subcampos nunca utilizados é ainda mais significativo – 809, correspondentes a 68%; e com uma utilização inferior a 0,5% dos registos encontramos 93% dos subcampos. Neste caso, a catalogação de publicações periódicas é feita utilizando apenas 82 dos 1194 subcampos disponíveis.

Quanto aos subcampos mais utilizados, independentemente do tipo de material, encontramos os seguintes, com utilização em mais de 50% dos registos:

- 100 \$a, 101 \$a, 102 \$a
- 200 \$a \$f
- 210 \$a \$c \$d
- 215 \$a \$d
- 675 \$3 \$a \$v \$z
- 700 \$3 \$a \$b
- 801 \$a \$b \$g

Subcampos	Designação	% de utilização
100 \$a	Dados gerais de processamento	100,00%
200 \$a	Título próprio	99,98%
101 \$a	Língua do texto	99,82%
210 \$a	Lugar de edição	99,63%
210 \$d	Data de publicação	99,53%
210 \$c	Nome do editor	99,51%
102 \$a	País de publicação	98,93%
215 \$a	Indicação específica da natureza do doc.	98,26%
801 \$a	País	90,95%
801 \$b	Agência	90,95%
801 \$g	Regras de catalogação	82,69%
200 \$f	Primeira menção de responsabilidade	77,33%
215 \$d	Dimensões	76,64%
700 \$a	Autor – Palavra de ordem	73,98%
700 \$3	Número de reg. de entrada de autoridade	72,86%
700 \$b	Outra parte do nome	71,58%
675 \$a	Notação CDU	68,43%
675 \$3	Número de reg. de classificação	68,13%
675 \$z	Língua da edição	63,64%
675 \$v	Edição	63,64%
200 \$e	Informação de outro título	32,33%
700 \$f	Autor - Datas	29,62%
215 \$c	Outras indicações físicas	29,18%
225 \$a	Título próprio da coleção	25,73%
106 \$a	Dados cod. mat. textual - Forma física	24,78%
702 \$a	Autor sec. – Palavra de ordem	24,49%
702 \$3	Número de reg. de entrada de autoridade	24,48%
702 \$b	Outra parte do nome	24,19%
702 \$4	Código de função	24,17%
010 \$a	ISBN	23,54%
105 \$a	Dados codificados dos livros	21,25%
225 \$v	Colecção - Volume	16,24%
200 \$g	Outras menções de resp.	16,18%
205 \$a	Menção de edição	15,91%
021 \$b	Depósito Legal - Número	13,05%
021 \$a	Código de país	13,02%
710 \$a	Colectividade-autor – Palavra de ordem	11,27%
710 \$3	Número de reg. de entrada de autoridade	11,26%
701 \$a	Co-autor – Palavra de ordem	10,64%
701 \$3	Número de reg. de entrada de autoridade	10,64%
701 \$b	Outra parte do nome	10,53%
701 \$4	Código de função	10,31%

Tabela 9: Subcampos preenchidos em mais de 10% dos registos de monografias da PORBASE.

Subcampos	Designação	% de utilização
100 \$a	Dados gerais de processamento	100,00%
200 \$a	Título próprio	99,98%
101 \$a	Língua do texto	99,59%
210 \$d	Data de publicação	99,56%
102 \$a	País de publicação	99,17%
210 \$c	Nome do editor	98,27%
801 \$a	País	94,16%
801 \$b	Agência	94,15%
801 \$g	Regras de catalogação	94,04%
210 \$d	Data de publicação	87,53%
200 \$f	Primeira menção de responsabilidade	71,06%
215 \$d	Dimensões	64,58%
712 \$3	Número de reg. de entrada de autoridade	60,70%
712 \$a	Colect.-autor sec. – Palavra de ordem	60,66%
207 \$a	Numeração	56,77%
675 \$a	Notação CDU	56,44%
675 \$3	Número de reg. de classificação	56,43%
712 \$4	Código de função	54,03%
675 \$v	Edição	52,97%
675 \$z	Língua da edição	52,89%
110 \$a	Dados cod. Publicações em série	48,17%
326 \$a	Periodicidade	38,74%
106 \$a	Dados cod. mat. textual - forma física	36,78%
620 \$d	Acesso pelo local - Cidade	34,78%
303 \$a	Nota informação descritiva	33,14%
702 \$a	Autor sec. – Palavra de ordem	29,14%
702 \$3	Número de reg. de entrada de autoridade	29,13%
702 \$b	Outra parte do nome	28,91%
702 \$4	Código de função	28,82%
200 \$e	Informação de outro título	28,46%
606 \$a	Assunto – Elemento de entrada	27,33%
606 \$3	Número de reg. de entrada de autoridade	27,33%
606 \$2	Código do sistema	25,04%
11 \$a	ISSN	21,90%
712 \$b	Colect.-autor sec. – Subdivisão	20,44%
530 \$a	Título chave	18,87%
200 \$g	Outras menções de resp.	12,68%
802 \$a	ISDS	12,06%

Tabela 10: Subcampos preenchidos em mais de 10% dos registos de periódicos da PORBASE.

O “UNIMARC ESSENCIAL”

Uma análise conjunta da Tabela 9 e da Tabela 10 permite-nos identificar um conjunto muito restrito de campos e subcampos, preenchidos em mais de 50% dos registos, que constituem o UNIMARC que é utilizado de facto pelas bibliotecas cooperantes da PORBASE. Identificamos assim 20 subcampos, quer no caso das monografias, quer no das publicações periódicas, a que podemos chamar de “UNIMARC essencial”.

Verifica-se que este conjunto de elementos corresponde, de uma forma muito aproximada, ao nível mínimo de

preenchimento de registos que foi definido em 2004 para integração dos registos das bibliotecas cooperantes na PORBASE [7]. Podemos prová-lo com a Tabela 11 e a Tabela 12, para os casos das monografias e das publicações periódicas (assinalando-se a negrito as diferenças entre o “UNIMARC essencial” e o nível mínimo de preenchimento da PORBASE).

O “UNIMARC essencial” para as monografias	Nível mínimo da PORBASE
100\$a	100 \$a
101\$a	101 \$a
102\$a	102 \$a
200\$a \$f	200 \$a
210\$a \$c \$d	210 \$a \$c \$d
215\$a \$d	215 \$a
675\$3 \$a \$v \$z	675 \$a \$v \$z
700\$3 \$a \$b	7-- (um campo do bloco)
801\$a \$b \$g	801 \$a \$b \$g
	966 \$l \$s

Tabela 11: Comparação do “UNIMARC essencial” e nível mínimo da PORBASE para as monografias.

O “UNIMARC essencial” para as publicações periódicas	Nível mínimo da PORBASE
100 \$a	100 \$a
101 \$a	101 \$a
102 \$a	102 \$a
200 \$a \$f	200 \$a
207 \$a	207 \$a ou 303 \$a
210 \$a \$c \$d	210 \$a \$c \$d
215 \$d	215 \$d
675 \$3 \$a \$v \$z	675 \$a \$v \$z
712 \$3 \$4 \$a	7-- (um campo do bloco)
801\$a \$b \$g	801 \$a \$b \$g
	966 \$l \$s

Tabela 12: Comparação do “UNIMARC essencial” e nível mínimo da PORBASE para as publicações periódicas.

COMPARAÇÃO COM O MARC21

Comparando as duas grandes conclusões obtidas pelo estudo da utilização do MARC21 nos registos da OCLC [8] com a nossa análise do UNIMARC na PORBASE, somos levados a concluir da existência de grandes semelhanças no que se refere à concentração da utilização do formato num conjunto muito restrito de elementos descritivos.

Dessas conclusões destacamos:

- Subcampos que ocorrem pelo menos uma vez nos registos:
 - MARC21: a percentagem é inferior a 50%.
 - UNIMARC: essa percentagem é de 51%.
- Subcampos concentrando 80% do formato:
 - MARC2: 4% dos subcampos concentram a utilização de 80% do formato; ou seja, os 36

subcampos que apresentam mais ocorrências nos registos totalizam 80% das ocorrências no conjunto de registos analisados.

- o UNIMARC: A Tabela 13 mostra que neste caso essa concentração é maior, pois 80% das ocorrências são de apenas 26 subcampos, ou seja, apenas 2,2% do total. Já ao nível da utilização dos campos, o resultado é menos concentrado: 80% das ocorrências são de 12 campos, ou seja, 6,8% dos campos disponíveis.

Subcampos	Nº ocorrências	% de ocorrências (simples)	% de ocorrências (acumulada)
675 Sv	1.448.797	4,61%	4,61%
675 Sz	1.445.933	4,60%	9,20%
101 Sa	1.246.818	3,96%	13,17%
200 Sa	1.234.483	3,92%	17,09%
210 Sa	1.228.291	3,90%	20,99%
210 Sc	1.226.361	3,90%	24,89%
100 Sa	1.225.285	3,89%	28,79%
102 Sa	1.217.256	3,87%	32,66%
801 Sa	1.213.973	3,86%	36,52%
801 Sb	1.213.969	3,86%	40,37%
210 Sd	1.209.957	3,85%	44,22%
675 Sa	1.164.435	3,70%	47,92%
675 S3	1.162.459	3,70%	51,62%
215 Sa	1.141.366	3,63%	55,25%
801 Sg	1.027.821	3,27%	58,51%
200 Sf	943.481	3,00%	61,51%
215 Sd	943.348	3,00%	63,51%
700 Sa	836.960	2,66%	67,17%
700 S3	824.479	2,62%	69,79%
700 Sb	810.061	2,58%	72,37%
200 Se	457.810	1,46%	73,82%
702 S4	409.486	1,30%	75,12%
702 Sa	408.970	1,30%	76,42%
702 S3	408.867	1,30%	77,72%
702 Sb	401.652	1,28%	79,00%
215 Sc	359.306	1,14%	80,14%

Tabela 13: Ocorrências dos subcampos mais utilizados⁵.

As razões para estas concentrações podem ser múltiplas. Passam certamente pela elevada complexidade de ambos os formatos, que pretendem abarcar a descrição de todos os tipos de recursos de informação. Mas dado que os formatos foram criados e desenvolvidos usando um conjunto diversificado de orientações e regras, é possível que a não utilização, ou fraca utilização de muitos campos e subcampos, se relacione também com aspectos particulares das regras de catalogação associadas aos

⁵ O total de ocorrências considerado para este cálculo foi de 31.458.092, porque foi retirado ao total geral de 33.757.510, apresentado na Tabela 3, o nº de ocorrências dos campos 1 e 5, uma vez que estes não estão estruturados em subcampos.

respectivos elementos descritivos.

Outro factor a considerar é de que o profissional de informação, ao fazer a descrição dos recursos, ainda encara o UNIMARC simplesmente como um “hospedeiro” dos elementos das ISBD e das regras de catalogação, não explorando as potencialidades que o formato oferece ao nível da recuperação da informação pelos sistemas de informação (sendo por isso os campos de dados codificados aqueles que têm uma utilização mais irregular).

CONCLUSÃO

Este trabalho teve como objectivo um estudo geral de utilização do formato UNIMARC nas práticas de catalogação nacionais, lançando pistas que se espera que possam ser úteis. O trabalho poderá ser complementado por estudos mais analíticos, considerando por exemplo todos os tipos de material, uma vez que muitos dos campos existentes no UNIMARC são campos específicos, utilizados apenas em determinados materiais. Outro exemplo de trabalho complementar poderia consistir num estudo mais pormenorizado de modo a identificar e compreender melhor os factores que eventualmente mais contribuem para que a utilização do formato pelo catalogador tenha as características que apresenta. De qualquer forma pensamos que o estudo aqui apresentado já nos permite retirar um conjunto de conclusões significativas.

O facto de o UNIMARC oferecer um conjunto muito amplo de campos e subcampos para a descrição bibliográfica tem vantagens do ponto de vista do catalogador e do profissional de informação, pois permite tornar a descrição mais rica e rigorosa e recuperar um elevado número de elementos descritivos. No entanto, do ponto de vista da interoperabilidade e da migração dos dados, esta variedade torna muito complexa a análise, desenho, desenvolvimento e manutenção dos sistemas, com os custos associados. Já do ponto de vista do utilizador final, a riqueza dos metadados disponíveis para a descrição acaba por ser irrelevante, uma vez que a informação existente não é mostrada nas visualizações em ISBD ou semelhantes, que são geralmente apresentadas ao utilizador.

Tendo em conta estes pontos, e dado que a análise feita permite concluir que uma percentagem considerável de campos e subcampos disponíveis no UNIMARC para descrição bibliográfica têm uma utilização praticamente nula, pensamos que esta constatação deve ser tomada em conta na tomada de decisões relativamente à evolução, quer do formato UNIMARC, quer das regras de catalogação que norteiam a sua utilização.

Por último, esperamos também que esta análise possa ser útil na definição das políticas de normalização e no investimento ao nível das ferramentas tecnológicas de apoio à PORBASE.

NOTAS

- [1] IFLA – UNIMARC: an introduction [Em linha]. The Hague : IFLA, 1999. [Consult. 8 Jan. 2007]. Disponível em WWW: <URL:<http://www.ifla.org/VI/3/p1996-1/unimarc.htm>>
- [2] GALVÃO, Rosa Maria – O UNIMARC num contexto de cooperação: evolução do formato. In JORNADAS PORBASE, 9, Lisboa – Novos contextos para a cooperação [Em linha]. Lisboa : BN, 2003. [Consult. 8 Jan. 2007]. Disponível em WWW: <URL:<http://purl.pt/331/1/docs/comunicacao/11manha/unimarcnumcontextodecooperacao.doc>>
- [3] MOEN, William E. ; Miksa, Shawne - MARC content designation utilization [sítio web]. Denton: Texas Centre for Digital Knowledge, 2005. [Consult. 4 Set. 2006]. Disponível em WWW: <URL:<http://www.medu.unt.edu/?p=30>>
- [4] MOEN, William E. ; Miksa, Shawne - Examining Present Practices to Inform Future Metadata Use: An Empirical Analysis of MARC Content Designation Utilization [Em linha]. 2004 [Consult. 5 Nov. 2006]. Disponível em WWW: <URL:<http://www.medu.unt.edu/wp-content/FinalProposalNarrativeReferences1Feb2004.pdf>>
- [5] MOEN, William E. [et al.] - Learning from Artifacts: Metadata Utilization Analysis. In International Conference on Digital Libraries Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries. [Em linha]. Chapel Hill, NC : 2006. p. 270 - 271. [Consult. 4 Set. 2006]. Disponível em WWW: <URL:<http://doi.acm.org/10.1145/1141753.1141813>>
- [6] IFLA - UNIMARC concise bibliographic format (1 Mar 2002) [Em linha] The Hague : IFLA, 2002. [Consult. 5 Nov. 2006]. Disponível em WWW: <URL:<http://www.ifla.org/VI/3/p1996-1/concise2.pdf>>
- [7] BIBLIOTECA NACIONAL. Divisão da PORBASE. Área de Cooperação - Informação Técnica n.º 2/2006 - Nível Mínimo de Preenchimento: campos e subcampos UNIMARC de preenchimento obrigatório e essencial [Em linha]. Lisboa : BN, 2006. [Consult. em 8 Jan. 2007]. Disponível em WWW: <URL:<http://www.porbase.org/referencias/it-2-2006.html>>
- [8] MOEN, William E. ; BENARDINO, Penelope – Assessing metadata utilization: an analysis of MARC content designation use. In DUBLIN CORE CONFERENCE, 2003, Seattle – Supporting Communities of Discourse and Practice – Metadata Research & Applications. [Em linha] Seattle, 2003. [Consult. 5 Nov. 2006] Disponível em WWW: <URL:http://www.siderean.com/dc2003/502_Paper58.pdf>