COMUNICAÇÕES

O "COMMON COMMUNICATION FORMAT" DESENVOLVIDO PELA UNESCO*

Alan Hopkinson
Consultor da UNESCO. British Library Bibliographic Services Division, Londres.

1 - INTRODUÇÃO

"UNESCO COMMON COMMUNICATION FORMAT - CCF"¹ é um formato de intercâmbio desenvolvido para ser usado por instituições da comunidade de informação que pretendem intercambiar, mutuamente, registros bibliográficos. Particularmente, é destinado àquelas instituições que possuem registros de monografias e seriados — materiais típicos de bibliotecas, assim como registros de artigos de periódicos, de contribuições em anais de congressos e outras partes de documentos físicos, que constituem unidades intelectualmente independentes, sendo estes os tipos de registros criados pelos serviços de indexação e resumos.

Um formato de intercâmbio é formado por três componentes:

- a) Regras para a organização de dados em computadores para serem intercambiados (incluindo regras para o tamanho do meio físico de armazenamento);
- b) Códigos para a identificação dos diferentes elementos de dados no registro (por exemplo: autor, título, escala de mapas, data de início do periódico);

*Tradução do original em inglês por Meireluce da Silva Ferreira e Maria Carmen Romcy de Carvalho, técnicas de informação do IBICT.

RESUMO

Descreve o "Common Communication Format — CCF", um formato de intercâmbio desenvolvido pela UNESCO - PGI (Programa Geral de Informação), para ser usado por instituições que pretendem intercambiar registros bibliográficos. São feitas comparações entre o CCF e outros formatos de registros bibliográficos, sendo evidenciada a necessidade do CCF agir como ponte de ligação entre as comunidades de informação atuando como um formato de intercâmbio sobre o qual se basearia um formato interno para bases de dados de sistemas de bibliotecas e de serviços secundários.

Descritores: Registros Bibliográficos; Formatos Internacionais de Intercâmbio; Common Communication Format/UNESCO - PGI.

c) Regras para a formulação dos diferentes elementos de dados (muito relacionado ao item b). Os elementos de dados identificados separadamente pelos códigos no formato de intercâmbio têm que ser definidos, não somente em termos de conteúdo, mas também de forma se os registros forem adequados para uso por outra instituição.

O intercâmbio efetivo entre instituições somente pode ser efetuado, se os registros das mesmas estão em conformidade com os itens a, b e c.

No primeiro caso, existe um formato padrão para o intercâmbio de dados em fita magnética que foi estabelecido pela ISO — International Organization for Standardization, a ISO 2709². A aceitação universal deste padrão tem ajudado muitíssimo à comunidade de informação. O padrão é aceito para intercâmbio de dados em fita magnética, e está também sendo usado para o formato bibliográfico em dados enviados "online" por linha telefônica.

O segundo componente relaciona os "tags", indicadores e códigos de sub-campos, em suma, códigos que definem os diferentes elementos de dados no registro. Não existe padrão universalmente aceito para isto. Uma razão que os fazem variar entre as diferentes implementações da ISO 2709 é que as instituições têm diferentes exigências a respeito dos dados que elas desejam intercambiar. A maioria dos

O "Common Communication Format" desenvolvido pela Unesco
Alan Hopkinson

países, por exemplo, tem achado necessário desenvolver seus próprios formatos MARC nacionais e existem também formatos desenvolvidos ou especificamente planejados para serviços de indexação e resumo. Todos eles envolvem diferentes esquemas de "tags" e outros identificadores.

O último componente, a forma e conteúdo dos elementos de dados, variam de acordo com as regras de catalogação usadas, como também variam de acordo com o modo pelos quais os diferentes elementos de dados, prescritos pelas regras, são divididos e separadamente identificados.

O "Common Communication Format" implementa a ISO 2709 na íntegra. Ele possui seu próprio sistema de "tags" e outros identificadores para a identificação de elementos de dados, mas não assume qualquer código de catalogação em particular. Este ponto será elaborado depois.

2 - HISTÓRIA E DESENVOLVIMENTO

O "Common Communication Format — CGF" foi desenvolvido por um grupo de especialistas, convidados pela UNESCO e que trabalham no contexto da UNESCO-PGI (Programa Geral de Informação). Em 1978, realizou-se o Simpósio Internacional sobre Formatos de Intercâmbio Bibliográfico³ patrocinado pela UNESCO. A UNESCO estava interessada neste assunto, porque havia recebido muitos pedidos de instituições de todo o mundo para consultorias no estabelecimento de bibliografias regionais e nacionais. Os peritos encontravam dificuldades quando recomendavam que formato um sistema deveria usar, embora os formatos fossem divididos, a grosso modo, em duas categorias: — aquele usado pelas bibliotecas nacionais, a família MARC da qual o CALCO é um membro —, e aquele usado pelos serviços de indexação e resumo, tais como o formato CEPAL que, embora mais diverso, constitui um exemplo do Manual de Referência (Reference Manual) do UNISIST.⁴ Este formato foi desenvolvido juntamente pela UNESCO e pelo ICSU-AB (International Council of Scientific Unions Abstracting Board) devido a necessidade sentida por padrões e guias para serviços secundários, considerando a automação de suas bases de dados. O Simpósio Internacional dedicou-se ao problema da existência de duas categorias de formatos que refletiam as duas diferentes comunidades de informação: bibliotecas e serviços secundários. Estava claro que no futuro, a nível de padronização de registros bibliográficos, deveria haver mais cooperação para que as necessidades do usuário final fossem satisfeitas, pois, pesquisadores encontram

referências de artigos em bases de dados de serviços de resumo e indexação, mas têm que recorrer à biblioteca para encontrar o periódico ou a monografia citados.

O Simpósio resolveu tentar quebrar as barreiras entre as duas comunidades de informação. Uma maneira de contribuir para isto foi o planejamento de um formato, que não seria direcionado especificamente a qualquer comunidade, mas que pudesse ser utilizado por qualquer instituição, que estivesse fornecendo registros a ambas.

O Grupo "Ad-Hoc" para Estabelecimento de um Formato de Comunicação foi reunido. Primeiramente, os membros solicitaram que a UNESCO encomendasse um diretório de elementos de dados, que considerasse os elementos de dados dos vários formatos de intercâmbio internacional, incluindo o UNIMARC⁵, o USSR/USA "Common Communication Format", o MEKOF⁶, o formato de ISDS⁷ e o Reference Manual. Um consultor preparou um índice KWOC para os elementos de dados dos formatos e uma definição abreviada para cada elemento de dados. Isto constitui a base para as discussões em torno da definição de um conjunto obrigatório, básico, de elementos de dados. Elementos de dados opcionais foram então acrescentados, tornando possível a inclusão de registros bibliográficos completos no formato. Entretanto, ficou claro que para o formato ser aceito não seria possível prescrever exatamente a forma e conteúdo de cada elemento de dados, embora fosse possível ser preciso em alguns elementos de dados, tais como, aqueles cuja forma e conteúdo tivessem sido prescritos por outros padrões, isto é, os números-padrões (ISBN e ISSN) e título-chave.

Suficiente consenso não pode ser obtido no caso de outros elementos de dados. Isto nunca poderia ser alcançado sem as regras de catalogação internacionalmente aceitas, as quais são ainda algo para o futuro. Assim, quando o "Common Communication Format - CCF" foi publicado pela UNESCO em meados de 1984, a representação de muitos elementos de dados foi especificada como sendo, "de acordo com as práticas da instituição que prepara o registro", o que significa "de acordo com as regras de catalogação próprias de cada instituição". Assim, uma instituição que recebe um registro no CCF necessitaria saber as regras usadas pela instituição - fonte, a fim de determinar se vale a pena utilizar os registros. Isto não significa dizer que os registros criados de acordo com regras diferentes não seriam totalmente aproveitadas por uma instituição

receptora. Estando em "Formato Comum de Comunicação" e tendo os mesmos identificadores, os registros podem ser facilmente interpretados por programas da instituição receptora. Eles podem ser usados para se saber o que uma organização tem em termos de acervo, supondo que os registros originam-se de uma biblioteca; ou podem ser usados para produção de Sumários Correntes, por exemplo, se os registros são originários de serviço de informação secundária. Eles podem ser somados a base de dados "online", embora a ausência de regras comuns de entrada (pontos de acesso) possam tornar a recuperação da informação um pouco mais difícil e menos eficiente. Os problemas surgem somente quando registros de diferentes fontes preparados de acordo com diferentes códigos de catalogação são inseridos em uma base de dados com vistas a produzir um produto impresso, tal como um catálogo. Os registros criados sob diferentes regras não podem ser facilmente arquivados junto com um produto impresso. Existe, no momento, o "International Standard Bibliographic Descriptions" que assegura uniformidade na área descritiva do registro, mas ainda não existem padrões universais para a forma de cabeçalhos.

3 - O "COMMON COMMUNICATION FORMAT" E OUTROS FORMATOS

Por que necessitamos de outro formato de intercâmbio, quando já temos o UNIMARC e o Reference Manual, sem mencionar os formatos nacionais de intercâmbio? Esta é uma crítica frequentemente direcionada ao "Common Communication Format". Entretanto, ambos os formatos atendem bem apenas à metade da comunidade de informação.

O UNIMARC é idealmente destinado a servir as bibliotecas nacionais com seus próprios formatos nacionais e que desejam intercambiar registros entre si. Se não fosse pelo UNIMARC, o intercâmbio entre bibliotecas nacionais teria que ser realizado, em cada caso, em bases bilaterais, com programas de conversão sendo escritos entre cada par de formatos, entre os quais se dá o intercâmbio. Uma alternativa seria que todos usassem um formato nacional, por exemplo o MARC americano, mas se este método fosse o escolhido haveria um clamor internacional todas as

*Seis (seis) volumes do International Standard Bibliographic Descriptions foram publicados pelo Escritório Internacional da IFLA para o Controle Bibliográfico Universal (UBC) relativos a publicações monográficas, seriadas, materiais não-livros, materiais cartográficos, obras raras e partituras; há também um ISBD geral, um roteiro que orienta a estrutura dos outros ISBD's.

vezes que o formato nacional escolhido fosse mudado, já que isto significaria ajustes em todos os demais programas. O UNIMARC é suficientemente próximo à maioria deles, se não a todos os formatos MARC, justamente para evitar conversões complicadas entre ele e os formatos MARC nacionais, e ele só poderá ser mudado através de acordo internacional. A principal desvantagem do UNIMARC é que ele é direcionado a materiais de biblioteca e não há um modo fácil, por exemplo, de relacionar o registro de um artigo ao registro do periódico que o contém. Não há problemas de tipo analítico, tratados pelo AACR — Anglo American Cataloguing Rules, que consiste geralmente em uma entrada (ponto de acesso) para um trabalho já publicado independentemente, mas que agora se encontra contido em um volume juntamente com outros trabalhos. Entretanto qualquer item mais complexo torna-se um pouco incontrolável.

O Reference Manual refere-se especificamente a este problema. A solução que ele apresenta era boa quando as fitas não constituíam, apenas o único meio para intercâmbio, mas eram também muito usadas no processamento de dados textuais, para armazenamento de dados, em situações onde os discos são hoje usados. Cada nível bibliográfico, que se constitui num item que se deseja recuperar, tem o seu próprio registro, e dentro do mesmo registro estão todos os detalhes, tais como, monografia, coleção ou seriado, onde pode ser encontrado. Assim, existirá um registro abreviado ou seriado em cada registro de cada artigo naquele seriado. Os elementos de dados no registro global são identificados por "tags" que indicam o nível bibliográfico daquele elemento de dados. Assim, o título de monografia tem "tag" diferente do título do seriado, o mesmo se aplica para autor, detalhes de coleção, etc. Em uma base de dados organizada deste modo, haveria um alto índice de repetição. Em uma fita de intercâmbio, isto não importa tanto; na verdade saber que todo registro está completo, torna o processo mais fácil.

Por outro lado, o CCF tem um mecanismo que leva em consideração a necessidade de ligação de registros. Ele pode ligar registros em níveis bibliográficos diferentes, os quais quando colocados juntos constituem o registro de um item bibliográfico, ou pode ligar registros com diferentes tipos de relações, tais como, o título do periódico e seu título anterior e vice-versa, ou um trabalho e suas traduções. O mecanismo de ligação foi planejado, para ser flexível para ser compatível com o complexo mecanismo do UNIMARC e a estrutura plana*de

"registro plano = registro sem sub-registros

registro de Reference Manual, que atribui um "tag" diferente para um elemento dado, dependendo se está a um nível analítico, ou monográfico, seriado ou coleção. A conversão entre o CCF e o MEKOF foi também considerada; embora a estrutura do registro do MEKOF seja do tipo plana, ele possui um "tag" de 5 dígitos utilizando um dígito para indicar, de forma consistente, o nível bibliográfico do elemento de dados. Entretanto, a flexibilidade com a qual o CCF é dotado significa que os usuários devem ser cautelosos: devem certificar-se de que eles próprios usam o mecanismo de modo consistente, ou então, que instituições receptoras dos registros tenham construído dentro de seus programas, as instruções corretas que decodifiquem todos os mecanismos particulares usados. Além disso, os registros são divididos em segmentos que apontam para outros registros. Portanto, é necessário assegurar que todas as pequenas partes de um registro sejam intercambiadas juntas. Assim, procedimentos de organização (housekeeping) têm que ser desenvolvidos para um intercâmbio entre as partes, antes que o intercâmbio possa ocorrer, e no momento não há diretrizes para isto.

Conseqüentemente, há ainda espaço para a utilização de todos os três formatos internacionais de intercâmbio já mencionados.

Também não há razão para que o AGRIS — Agricultura! Information System e o INIS - International Nuclear Information System desistam de seus próprios formatos que foram planejados para sistemas particulares.

4-O FUTURO

O objetivo principal de todos os formatos de intercâmbio é facilitar o livre fluxo de informação neste caso, referências bibliográficas. Para isto ser alcançado, provavelmente mais importante seria os computadores estarem aptos a conversarem entre si, o que seria alcançado por especialistas fora da biblioteca e da comunidade de informação. Depois disto, o que é necessário são formas internacionalmente aceitas de cabeçalhos ou "authority files". Ao mesmo tempo, mais esforço deveria ser feito em conversões entre formatos de intercâmbio existentes (e aqui os formatos nacionais estão incluídos) e torná-los prontamente disponíveis. Assim o "Common Communication Format" poderia ser realmente uma ponte efetiva entre a comunidade da biblioteca e a comunidade de informação e certamente ganharia popularidade como um formato sobre o qual se apoiaria um formato interno para bases de dados de sistemas de bibliotecas e de serviços secundários.

REFERÊNCIAS BIBLIOGRÁFICAS

- 1 CCF: the Common Communication Format Paris, Unesco, 1984.
- 2 INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. Documentation: format for bibliographic information interchange on magnetic tape. (2nd. ed.) Geneva, ISO, 1981 (ISO 2709- 1981).
- 3 INTERNATIONAL SYMPOSIUM ON BIBLIOGRAPHIC EXCHANGE FORMATS. Towards a common bibliographic exchange format?: proceedings. Budapest, OMKDK, 1978.
- 4 MARTIN, M. D. Reference manual for machine-readable bibliographic descriptions. Paris, Unesco, 1974 (SC. 74/WS/20). 2nd ed., edited by H. Dierickx and A. Hopkinson. Paris, Unesco, 1981 - (PGI/81/WS/22).
- 5 UNIMARC: Universal MARC Format London, IFLA International Office for UBC, 1977.
- 6 International Interchange Format - MEKOF-2: specification of data elements, Moscow, ICSTI, 1977 (Official English Translation).
- 7 ISDS manual. Paris: Unesco, 1984.
- 8 CCF, op. cit. p. 19-24.
- 9 LONG, Anthony. UK MARC and US/MARC: a brief history and comparison. In: Journal of Documentari, 40 (1): 1-12, 1984.
- 10 HOPKINSON, Alan. International access to bibliographic data: MARC and MARC-related activities. In: Journal of Documentation, 40 (1): 13-24, 1984.

ABSTRACT

It describes the "Common Communication Format — CCF", which has been developed by UNESCO's PGI (General Information Programme) as an exchange format to be used by institutions for the mutual exchange of bibliographic records. Comparisons between CCF and other formats of bibliographic records are made, and the need is established for CCF to become a link between the information communities — an exchange format giving internal support for databases in systems for libraries and secondary services.