



Continuidade digital: relatório de projeto - passo 1

Normas, legislação, terminologia, formatos

Relatório relativo ao passo 1 do cronograma aprovado do projeto “Continuidade Digital”

Ficha técnica MIP:

Título: Continuidade digital: relatório de projeto: passo 1

Id: dt_004

Classificação: (MEF) 900.10

Autores: Ana Rodrigues, Francisco Barbedo, Lucília Runa, Mário Sant'Ana

Colaboradores: Ana Franqueira, António Vieira, Cristina Ribeiro, Filipe Montargil, Gabriel David, Jakub Szczypa, João Machado, Jorge Souto, José Manuel Costa, Luís Corujo, Miguel Azguime, Paula Guimarães, Paulo Leitão, Rui Machado, Teresa Borges, Teresa Gião

Descritores: 1. Normas. 2. Formatos. 3. Legislação, 4. Terminologia. 5. Representação de informação

Data/Hora: 2014-10-07

Formato de dados: Texto (docx)

Estatuto de utilização: Acesso reservado à equipa de projeto

Relação: versão - 1.0



DGLAB, 2014

<http://creativecommons.org/licenses/by-sa/4.0/>

Conteúdo

1 Introdução	4
2 Legislação.....	5
3 Terminologias	6
4 Normas	10
5 Formatos	19
6. Conclusões	31

1 Introdução

O projeto “continuidade digital” tem como objetivo efetuar uma análise coletiva e envolvente que permita determinar a viabilidade de construir uma rede para a preservação comum de património digital.

Por rede comum entende-se uma estrutura de oportunidades, aberta e transversal. Aberta a todas as instituições públicas ou privadas que detenham património digital e que possuam a vontade de empreender ações concretas e comprometidas no sentido de o preservar prolongadamente.

Transversal porque pretende abranger os mais diversificados e variados domínios culturais e comunidades de prática (CdP).

O cronograma aprovado contém um conjunto de passos os quais, por sua vez, são constituídos por atividades que coordenadamente analisam tópicos específicos os quais, na sua totalidade, facultarão uma perspetiva global e detalhada sobre os diferentes objetos patrimoniais, bem como as soluções de preservação adequadas ou simplesmente possíveis.

A metodologia base para todos os passos e atividades consiste em procurar determinar pontos de aproximação e também de afastamento entre as diferentes comunidades e domínios culturais representados no grupo de trabalho (GT). Pretende-se identificar possibilidades e/ou dificuldades na construção de estruturas comuns para preservar os objetos digitais inseridos em diferentes domínios culturais.

Nesse sentido o presente relatório apresenta a metodologia de análise seguida para cada atividade bem como as conclusões obtidas no que se refere ao **passo um**, onde foram reunidas, comparadas e analisadas legislação, terminologias, normas de representação e formatos reportados pelos membros do GT.

Pontos significativos do presente relatório:

1. Há alguns textos legislativos que preconizam a obrigatoriedade de conservar objetos que a legislação especificamente enquadra. Ver ponto 2.4
2. Duas classes de estruturas vocabulares foram identificadas: orientadas à atividade e orientadas à representação da atividade. As estruturas que contêm maior número de termos e conceitos são as que se inserem na categoria 2, orientada à representação informativa do objeto. Ver ponto 3.5
3. Existem normas partilhadas por diferentes CdP. Ver ponto 4.4, quadro e gráfico 4.2
4. A maior parte das normas analisadas centram-se, quanto ao objeto, num domínio patrimonial específico. Ver ponto 4.5.1
5. A maior parte das normas analisadas contempla informação relacionada com o contexto de produção dos objetos representados, embora esta informação seja sobretudo valorizada pelos arquivos, logo seguidos dos museus. Ver ponto 4.5.5

6. A utilização da representação multinível não é exclusiva de normas de arquivo. Ver ponto 4.5.6
7. As normas categoriais estruturam a informação em zonas ou grupos e, dentro destes, em elementos ou categorias de informação. As normas combinatórias recorrem a meta dados, esquema de meta dados e linguagens de definição de descrição. Ver ponto 4.5.7
8. O objetivo das normas é comum, centrando-se na recuperação troca e partilha de informação. Ver ponto 4.5.8
9. Há normas categoriais (mais clássicas e temporalmente anteriores) e combinatórias (mais recentes). As primeiras apresentam um elevado grau de equivalência. As segundas, pela sua aproximação relacional, oferecem uma flexibilidade de descrição compatível com vários domínios patrimoniais. Ver ponto 4.5.9
10. Experiências práticas de descrição “cruzada” permitiram perceber a possibilidade de descrever um objeto proveniente de um domínio patrimonial com uma norma aplicada noutro domínio. Ver ponto 4.5.10
11. Foram comunicados um total de 54 formatos utilizados. Ver ponto 5.5
12. Nenhum formato é usado por todas as CdP. Ver ponto 5.7
13. A categoria com mais formatos usados é imagens em movimento (23). Ver ponto 5.15
14. A categoria com menos formatos usados é datasets (4). Ver ponto 5.16

2 Legislação

Metodologia

2.1 No contexto da análise legislativa procedeu-se, para cada CdP, ao reconhecimento e identificação dos diplomas reguladores da respetiva atividade - regimes jurídicos, Lei de Bases, Códigos deontológicos -, assim como de aspetos específicos - com especial interesse para a preservação digital, tendo os mesmos sido agrupados em torno dessas duas categorias.

2.2 Nesse sentido, foi identificada a legislação reguladora e específica - tanto a nível nacional como europeu - das seguintes CdP:

- Património Cultural;
- Arquivos;
- Bibliotecas;
- Informação clínica;
- Jornalismo;
- Televisão;
- Rádio;

- Cinema;
- Fotografia;
- Música;
- Multimédia e Entretenimento.

2.3 Para efeitos de descrição e citação bibliográfica, procedeu-se à normalização da referenciação dos diplomas, considerando, no âmbito do presente projeto, os seguintes elementos:

- N.º de referência;
- Designação;
- Descrição do conteúdo;
- Observações.

Conclusões

2.4 Do levantamento efetuado, constatou-se a existência de diplomas:

- Com interesse para todas as CdP (Lei dos Direitos de Autor e Conexos; Lei do Depósito Legal);
- Multidomínio, aplicáveis à atividade jornalística, exercida através da Imprensa, Rádio e Televisão, e Património Cultural;
- Com referências explícitas a preservação de património/digital, designadamente:
 - Art.º 11.º da Lei de Bases do Património Cultural sobre “dever de preservação, defesa e valorização do Património Cultural”;
 - Cap. VII da Lei da Televisão, n.º 1, 2 e 3 do art.º 92.º sobre “conservação do Património Televisivo”;
 - Cap. VII da Lei da Rádio, art.º 83.º sobre ” conservação do Património Radiofónico - registos de interesse público”;
 - Portaria 247/2000 - Regulamento de conservação arquivística para os Hospitais e Serviços de Saúde. O processo clínico é de conservação permanente.

A lista da legislação consultada está disponível no anexo 1.

3 Terminologias

Metodologia

3.1 A análise de estruturas vocabulares foi precedida pela identificação das mesmas, tentando incluir todos ou pelo menos o máximo de domínios culturais representados. A pesquisa realizou-se através da internet pelos meios e recursos habituais. Em alguns casos não foi possível aceder aos documentos propriamente ditos por se tratar de normas que implicam pagamento para obter acesso. Noutros casos foram encontradas referências a estruturas vocabulares a que não foi possível aceder em linha ao respetivo conteúdo.

3.2 De uma forma geral foi possível encontrar terminologias relativas a diferentes CdP. No entanto a origem de produção e a fidedignidade das mesmas é variável. Por exemplo: no domínio cultural musical não foram encontradas estruturas de carácter normativo à exceção de uma ISBD que ainda se encontra em draft, elaborada na Biblioteca do Congresso (n.º21 do quadro). Encontram-se, no entanto, inúmeras estruturas terminológicas compulsadas por diferentes autores: pessoas individuais, empresas comerciais (por exemplo a Naxos <http://www.naxos.com/>), blogs, etc.

3.3 O anexo 1 representa as estruturas vocabulares identificadas e analisadas.

3.4 Metodologicamente havia que encontrar a forma mais eficaz de comparar termos entre as estruturas seleccionadas de forma a descortinar semelhanças e equivalências. Considerando a elevada variabilidade de palavras e termos utilizados procuraram-se aproximações semânticas de conceitos e não especificamente de palavras.

Neste cenário a clara diferenciação em duas classes distintas e a existência de formas de apresentação ocultas que impossibilitam a comparação.

As restrições à visualização total das estruturas vocabulares determinaram em parte a impossibilidade de comparar termo a termo. Mas mesmo que isso fosse possível não seria a aproximação mais lógica pelas razões já referidas da existência de duas classes de estruturas e da polissemia que habitualmente se verifica.

Optou-se por partir de um vocabulário nuclear estabelecido no documento estratégico do projeto e procurar esses termos nas estruturas vocabulares identificadas. A sua presença ou não, ou a frequência de termos presentes, poderia determinar o peso que um determinado conceito, independentemente do termo pelo qual é representado, se reveste para uma determinada CdP.

	arquivos	e/arquitet	teatro	filme	udio visual	filmografia	fotografia	multidomini	saúde	
termos base	DTA	AAT (Getty)	Theatre glossary	national film and sound archive	Unesco- audiovisual	GLOSSARY OF FILMOGRAPHIC TERMS	CLIP	UNESCO ciencias sociais	MESH (medical subject headings)	terms mais comuns por domínio
Acesso	1	1	0	1	1	0	0	1	1	6
Autenticidade	1	1	0	0	1	0	0	0	0	3
Avaliação	1	1	0	1	1	0	0	1	0	5
Comunidades de Prática	0	0	0	0	0	0	0	0	0	0
Custódia	1	0	0	0	1	0	0	0	0	2
Identificação de património digital	0	0	0	1	0	0	0	0	0	1
Infoestrutura para a preservação digital	0	0	0	0	0	0	0	0	0	0
Modelo de negócio	0	0	0	0	0	0	0	0	0	0
Património digital	0	1	0	1	1	0	0	1	0	4
Preservação digital -	1	1	0	1	0	0	1	0	0	4
Certificação e segurança de repositórios	0	1	0	0	0	0	0	0	1	2
Direitos de autor	1	1	0	0	1	1	0	1	1	6
Perspetiva de usabilidade	1	1	0	0	1	0	0	0	0	3
Responsabilidade partilhada	0		0	0	0	0	0	0	0	0
digital art	0	1	0	0	0	0	0	1	0	2
N de termos comuns por domínio	7	9	0	5	7	1	1	5	3	

Quadro 3.1- presença de termos por domínio

1 = presente

0 = ausente

As células assinaladas a verde representam a existência do termo ou conceito. As células brancas representam a inexistência do termo ou conceito na estrutura do domínio observado.

Conclusões

3.5 Foi possível identificar duas grandes classes de estruturas vocabulares:

- a. A primeira delas consiste nas estruturas desenvolvidas para apoiar diretamente a execução de uma determinada atividade. Estas estruturas têm termos muito específicos e praticamente exclusivos da CdP a que se destinam. É o caso, por ex., da GEMET Thesaurus (n.º 18 do quadro), relativo a ciências do ambiente. Neste tipo de normas o vocabulário é muito “tipado” e orientado para aspetos técnicos ou tecnológicos constantes da atividade a que se referem.
- b. Uma segunda é composta por estruturas desenvolvidas para apoiar a descrição ou representação de uma determinada atividade. Nesse sentido existe uma maior aproximação de termos patentes nestas estruturas, sendo possível encontrar uma maior similitude de termos. Este tipo de normas é normalmente produzido por organizações especializadas na custódia de objetos patrimoniais que funcionalmente têm de proceder à referência e representação descritiva desses mesmos objetos. Nestes casos existe uma semântica subjacente comum, que se traduz na existência de alguns termos e conceitos similares.

Aparentemente a diferenciação das formas de representação do mesmo conceito aproxima-se da prática instalada na CdP a que se refere cada estrutura terminológica.

3.6 As formas de apresentação destas estruturas são essencialmente duas: uma delas é uma representação sistemática em que o conjunto de termos é apresentado sob a forma de lista alfabética; a outra é uma representação estruturada sob a forma de tesouro, em que são estabelecidas relações entre os termos bem como os restantes mecanismos específicos deste tipo de ontologia (termos preferenciais, por ex.).

3.7 Sob o ponto de vista de visualização encontramos situações diferentes:

- O acesso aos diferentes termos é feito através de pesquisa sem possibilidade de ser efetuada navegação nos termos (browsing) ou efetuar o descarregamento do conjunto total dos termos constantes da estrutura.
- O acesso permite a visualização direta dos termos ainda que estes possam estar organizados por grupos alfabéticos.
- Noutros casos constatou-se uma apresentação híbrida, em que são disponibilizadas as duas possibilidades combinadas de visualização.

Uma análise de n.º de termos por domínio e termos mais frequentes dentro de um domínio encontra-se representada nos seguintes gráficos:

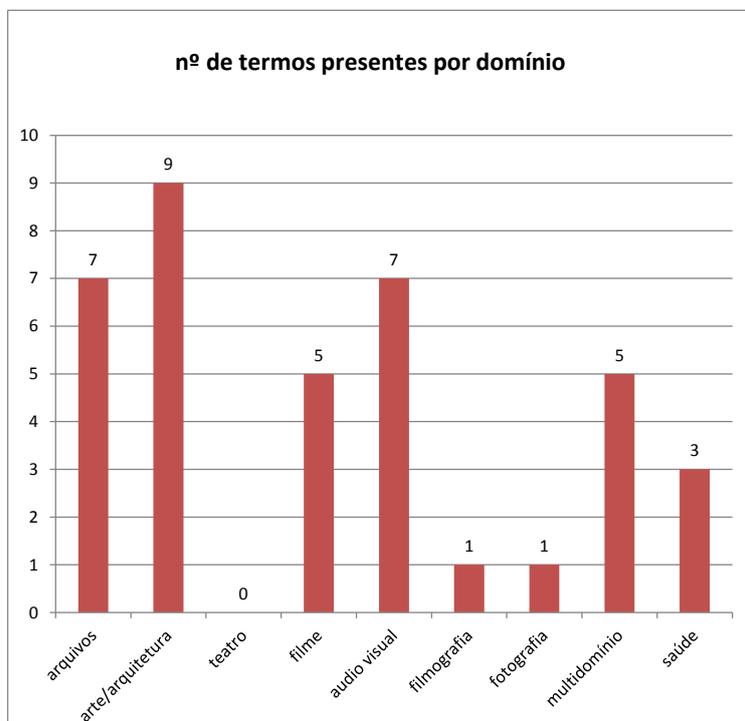


Gráfico 3.1

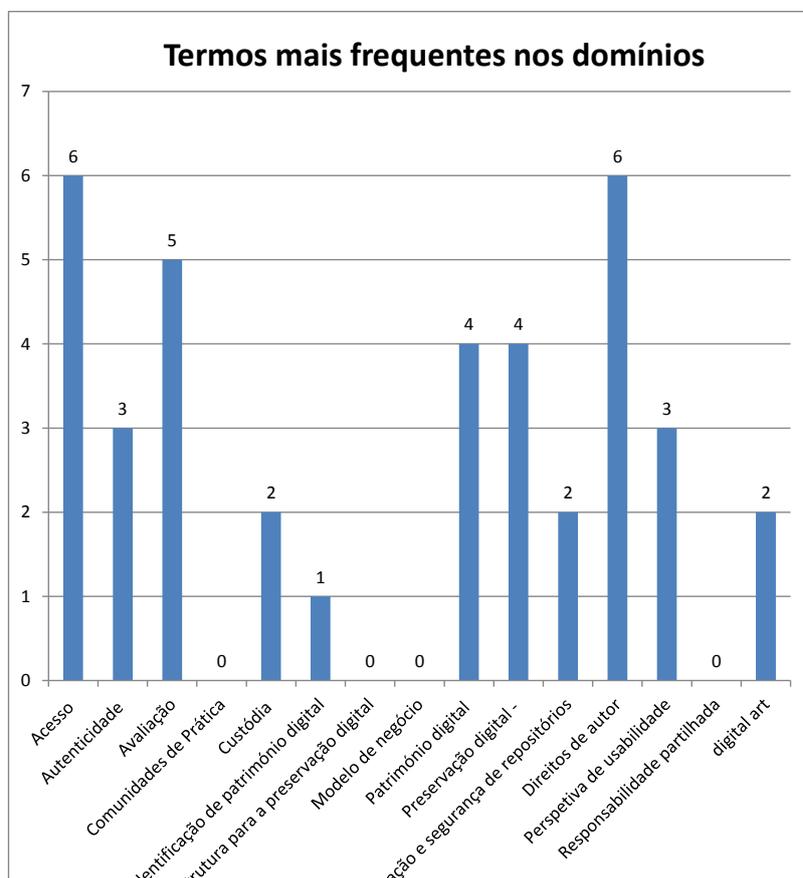


Gráfico 3.2

3.8. A média de n.º de termos presentes nas estruturas observadas é de 4,2, havendo várias que não apresentam qualquer dos termos base utilizados.

Constata-se que o maior n.º de termos presentes se concentra nas estruturas vocabulares pertencentes à classe dois, ou seja, orientada para a representação do objeto. As comunidades de prática de arquivos, áudio visual e museus contêm a maior concentração de termos.

Outras estruturas vocacionadas para a execução não contêm qualquer termo ou contêm-nos de forma limitada. Sirva de exemplo o vocabulário de termos filmográficos.

4 Normas

4.1 O **objetivo geral** definido para a tarefa 3 compreendia a identificação das normas utilizadas pelas diferentes CdP representadas no GT para criar informação sobre os respetivos objetos patrimoniais. São elas: jornalismo, televisão, cinema, museus, arquivos, bibliotecas, informação clínica, dados científicos, música, jogos eletrónicos.

Metodologia

4.2 Para o efeito procedeu-se a um primeiro levantamento geral das normas disponíveis para cada uma delas. O levantamento não foi exaustivo. Pretendeu-se, no entanto, que fosse representativo, pelo que foram consideradas:

- normas, regras, guias, orientações, recomendações, etc.
 - Internacionais, emanadas de organismos como: International Council on Archives (ICA), International Federation of Television Archives (IFTA), International Council of Museums (ICM), International Federation of Library Associations and institutions (IFLA), Music Library Association (MLA), Online Audiovisual Catalogers (OLAC), International Organization for Standardization (ISO), EUROPEANA, Archives Portal Europe Network of Excellence (APEX), etc.
 - Nacionais, elaboradas por entidades como: Instituto Português da Qualidade (IPQ), Arquivos, bibliotecas e museus nacionais, coordenadores das políticas adotadas nos diferentes setores - Instituto dos Arquivos Nacionais/Torre do Tombo-Direção-Geral de Arquivos-Direção-Geral do Livro, dos Arquivos e das Bibliotecas; Biblioteca Nacional de Portugal; Instituto Português de Museus-Instituto dos Museus e da Conservação-Direção-Geral do Património Cultural; Library of Congress; National Archives of Australia; Association of Canadian Archivists, etc.
 - Gerais e específicas.

A lista de normas identificadas corresponde ao anexo no separador *normas*.

4.3 Encontra-se estruturada em duas grandes áreas:

a/ a primeira remete para a identificação e caracterização das normas, contemplando a seguinte informação:

- CdP à qual se reportam;
- universo de aplicação (internacionais ou nacionais);
- objetivo (descrição, codificação, gestão, etc.);
- objeto (documentação, recursos bibliográficos, recursos musicais, peças / objetos, imagens em movimento, etc.);
- sigla ou designação genérica por que são conhecidas;
- referência bibliográfica (tão completa quanto possível);
- observações.

Para o conjunto das normas identificadas foram definidos dois níveis de pertinência:

- maior pertinência (assinalada a branco), que corresponde a normas utilizadas para a representação dos objetos das diferentes CdP;
- menor pertinência (assinalada a cinzento), que corresponde a normas de codificação da descrição, de gestão, ou a normas muito específicas.

b/ a segunda corresponde à indicação das que são utilizadas por cada uma das CdP.

De uma primeira análise, e considerando quer a sobreposição de objetos entre as normas dos diferentes domínios, quer a abordagem de base utilizada em cada uma delas, destaca-se, desde logo, a possibilidade de se considerar três grandes classes de CdP: arquivos, bibliotecas e museus -, integrando-se as demais nestas três grandes categorias. Veja-se, a título meramente exemplificativo, o caso do cinema, que utiliza como normas *The FIAF cataloguing rules for film archives* e a *Descrição bibliográfica internacional normalizada (ISBD): edição consolidada*.

Considerando, no entanto, o objetivo específico desta tarefa: proceder à análise comparativa das normas ou orientações utilizadas, mapear os respetivos elementos constitutivos, no sentido de encontrar eventuais correspondências, e avaliar possibilidades de os mapear para uma estrutura comum, optou-se pela individualização.

Para uma das normas identificadas, não foi possível aceder ao respetivo texto propriamente dito, por se tratar de um documento que implica pagamento para obter acesso (MPEG 7).

4.4 Passando de imediato para a caracterização do universo das normas identificadas, verifica-se que se distribuem pelas CdP da seguinte forma:

Normas identificadas	#
Jornalismo	---
Televisão	7
Cinema	4
Museus	15
Arquivos	22
Bibliotecas	2
Informação clínica	21
Dados científicos	---
Música	1
Jogos eletrónicos	---
Total	72

Quadro 4.1

Estes dados encontram melhor expressão no gráfico que se segue:

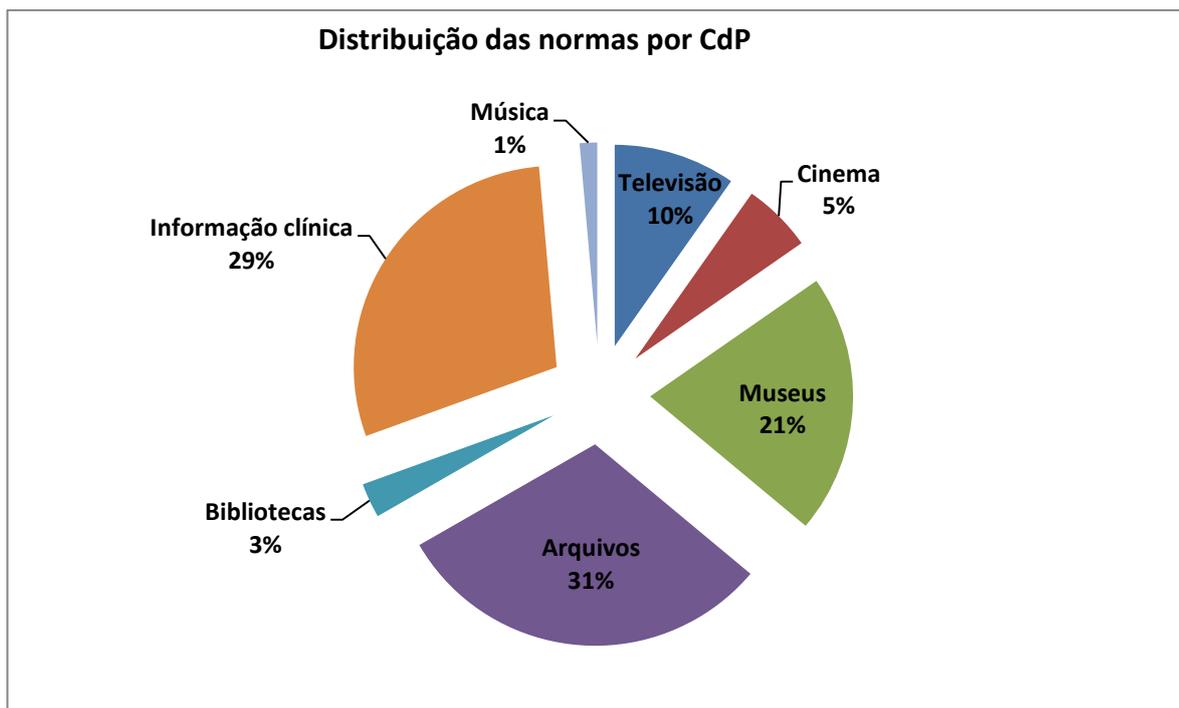


Gráfico 4.1

Do conjunto de normas identificadas:

- quatro são partilhadas por duas CdP, segundo a distribuição esquematizada no quadro:

Normas partilhadas	CdP
ISAD (G)	televisão, arquivos
ISBD	cinema, bibliotecas
EDM	museus, arquivos
DC	arquivos, música
Total	8

Quadro 4.2

Estes dados encontram melhor expressão no gráfico que se segue:

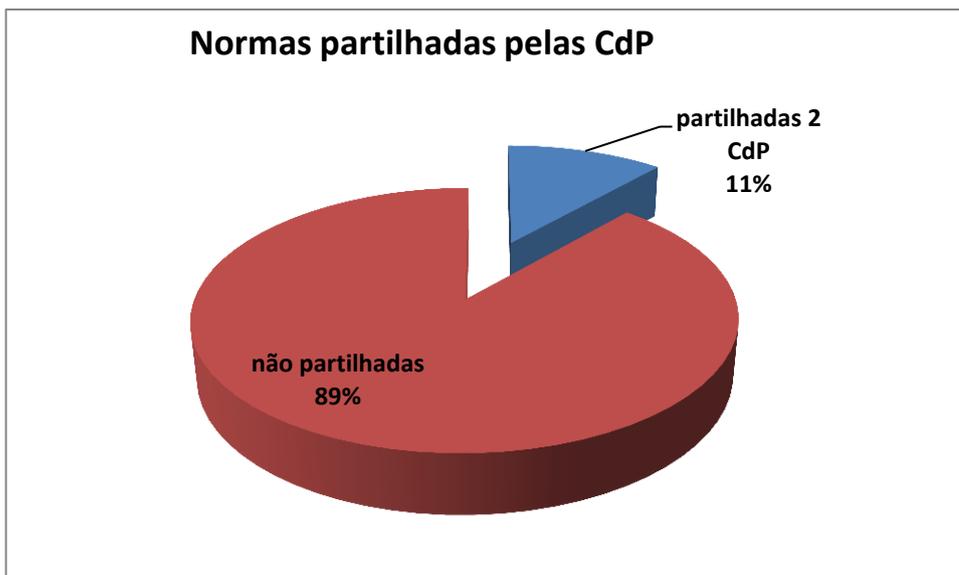
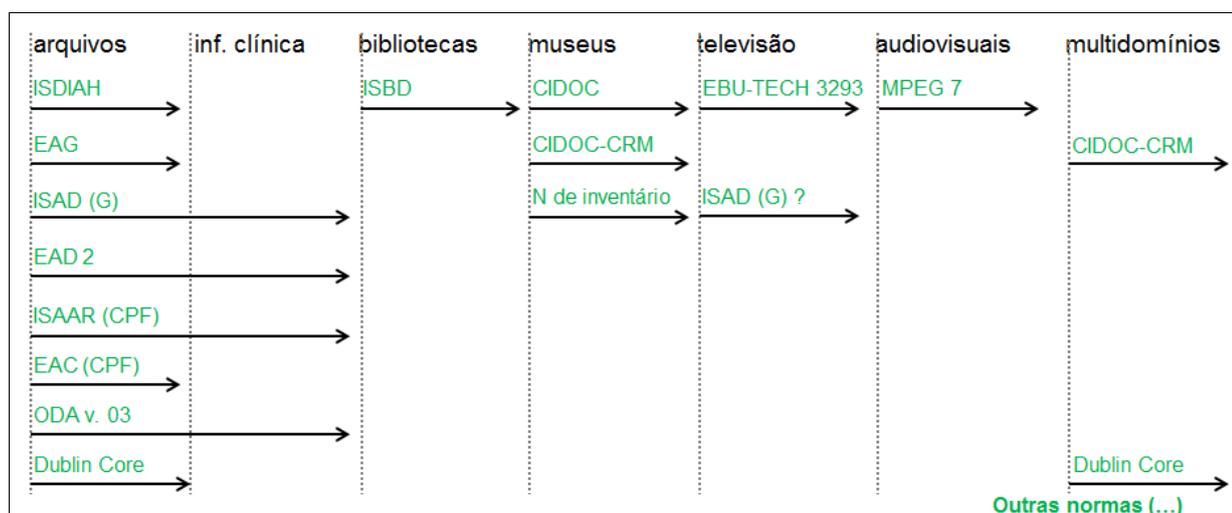


Gráfico 4.2

Do conjunto de normas identificadas, foram selecionadas, para análise comparativa, as que constam do quadro que se segue:

Normas selecionadas para análise comparativa



Quadro 4.3

Em relação a esta seleção há a mencionar que embora as normas de codificação da descrição utilizadas pelos arquivos tenham sido consideradas e assinaladas no anexo que corresponde ao separador *normas* como de menor pertinência, apresentam, para a análise, a vantagem de estruturar a informação de forma mais atomizada do que as suas congéneres, o que pode facilitar a abordagem comparativa entre os diferentes domínios patrimoniais.

A análise comparativa efetuada encontra-se documentada de forma mais detalhada no anexo que corresponde ao separador *normas_map*.

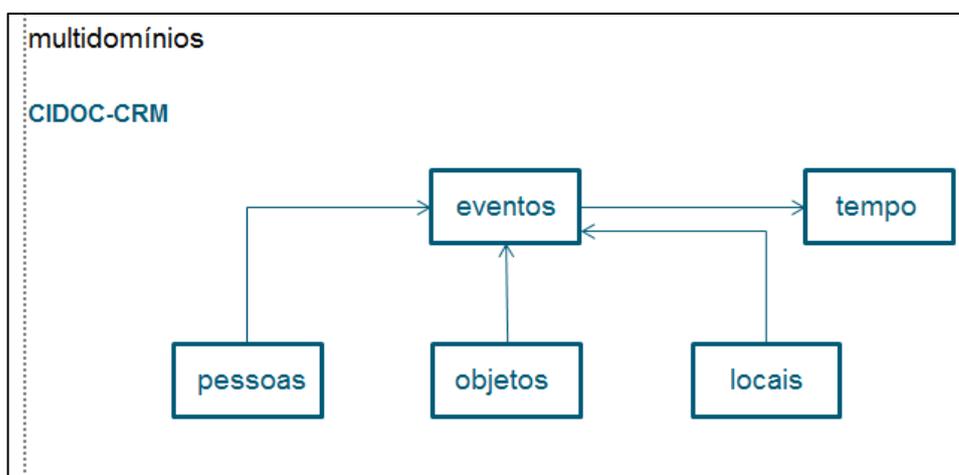
Conclusões

4.5 As conclusões a que foi possível chegar são as seguintes:

4.5.1 A maior parte das normas analisadas centram-se, quanto ao objeto, num domínio patrimonial específico. Constituem exceção a CIDOC-CRM e a Dublin Core, vocacionadas para uma aplicação transversal, multidomínio. Como pode constatar-se através do quadro 4.3, estão a ser utilizadas pelos arquivos e pelos museus, respetivamente.

4.5.2 As normas analisadas para arquivos, informação clínica, bibliotecas e museus, com exceção da CIDOC-CRM, contemplam separadamente a descrição e a codificação das descrições. As normas utilizadas pela Televisão (com exceção da ISAD (G)), pelos Audiovisuais e as transversais (multidomínio), contemplam, na mesma norma, as duas funcionalidades.

4.5.3 As normas analisadas centram-se no objeto. Constitui exceção o CIDOC-CRM, centrada no evento e orientada para o objeto, como pode constatar-se através da análise do quadro que se segue:



Quadro 4.4

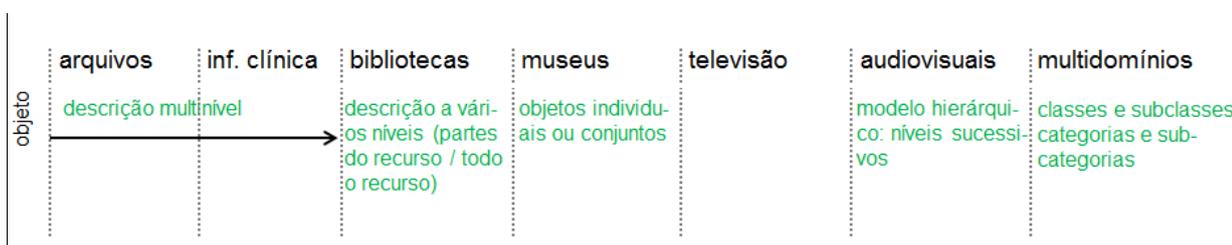
4.5.4 Ainda quanto ao objeto, e socorrendo-nos da terminologia utilizada nas próprias normas, pode ser sistematizado, em cada caso, da seguinte forma:

	arquivos	inf. clínica	bibliotecas	museus	televisão	audiovisuais	multidomínios
objeto	documentação contexto: produtores, detentores ...		recursos publicados	objetos e espécimes informação contextual: histórica, geográfica, teórica	recursos áudio e vídeo	imagens fixas, gráficos, modelos 3D, áudio, vídeo	recursos heterogéneos (património cultu- ral) informação contextual: histórica, geográfica, teórica

Quadro 4.5

4.5.5. Há a salientar, relativamente aos arquivos, o peso assumido pelo contexto de produção, que se materializa na existência de normas específicas para a descrição dos produtores / detentores e das funções. Este aspeto assume também relevância, embora não equivalente, nas normas utilizadas pelos museus, de entre as quais o CIDOC-CRM se assume como multidomínios. Nas demais normas esta informação é contemplada, mas com muito menor peso e incidência.

4.5.6 Há igualmente a mencionar, para os arquivos e informação clínica, o peso assumido pela descrição multinível, cujas definição e regras são expressamente enunciadas na ISAD (G) e nas ODA. O modelo hierárquico é igualmente mencionado no domínio dos audiovisuais. Para as bibliotecas e os museus, como pode verificar-se através da análise do quadro 6, prevê-se a descrição dos respetivos objetos como um todo, ou das várias partes que o constituem, dos objetos individuais ou dos conjuntos em que se integram.



Quadro 4.6

4.5.7 Quanto à estruturação da informação, arquivos, informação clínica, bibliotecas e museus, recorrem a zonas ou grupos e, dentro destes, a elementos ou categorias de informação.

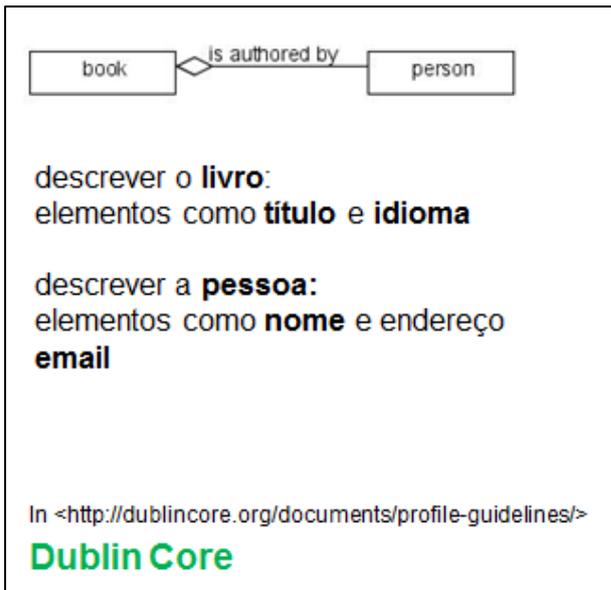
Arquivos, informação clínica e museus definem a forma como a informação deve ser registada dentro dos elementos ou categorias de informação, indicando regras e convenções específicas para o efeito.

Arquivos, informação clínica e museus procedem ao levantamento e definição da terminologia a utilizar, tendo em vista a normalização e a interoperabilidade.

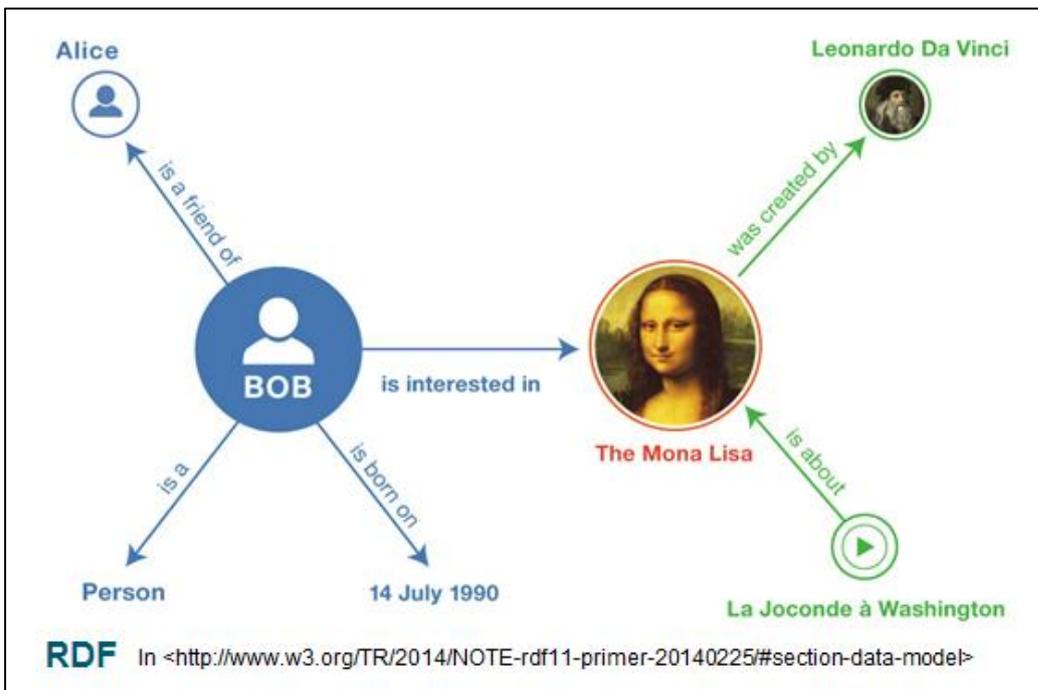
Arquivos, informação clínica e bibliotecas definem a sequência a utilizar no registo da informação, bem como a pontuação a utilizar para a sua separação.

Normas como as ISDIAH, ISAD (G), ISAAR (CPF), ODA, ISBD, CIDOC e Normas de inventário, podem ser classificadas, tendo em conta as suas características, atrás enunciadas, como categoriais.

De entre as normas analisadas existe um outro grupo, composto pelas normas combinatórias como a Dublin Core, a CIDOC-CRM, a EBU-TECH 3293 (ela própria baseada na Dublin Core) e a MPEG 7, constituídas por meta dados, esquema de meta dados e linguagem de definição de descrição. O modelo utilizado, que pode ser representado de forma mais ou menos complexa, pode ser esquematizado nos quadros que se seguem:



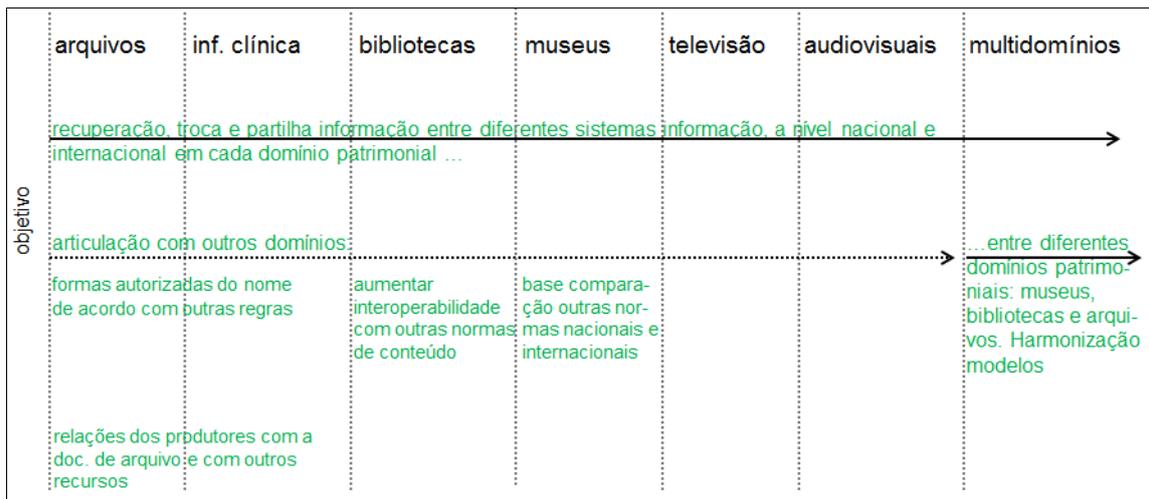
Quadro 4.7



Quadro 4.8

Estes modelos foram desenvolvidos para responder a situações em que a informação, que se encontra na Web, é proveniente de diferentes sistemas de informação e processada por diferentes aplicações. Um dos objetivos é o de permitir a partilha de toda essa informação, que se reporta a diferentes domínios patrimoniais, sem que haja perda do seu significado.

4.5.8 O objetivo das normas analisadas é comum, como pode aliás constatar-se através da análise do quadro que se segue:



Quadro 4.9

Para a maior parte delas a preocupação centra-se fundamentalmente na recuperação, troca e partilha de informação entre diferentes sistemas de informação, centrados num domínio patrimonial específico. Para as multidomínio, essa preocupação é assumida de forma transversal. Todas elas, no entanto, assumem a necessidade de articulação com outros domínios: no caso dos arquivos, essa preocupação traduz-se na total autonomização da descrição dos produtores / detentores, que para além de facilitar a partilha da informação, permite o estabelecimento de relações com a descrição da documentação por eles produzida / detida, ainda que dispersa por diferentes entidades detentoras, mas também com outros tipos de recursos, nomeadamente biblioteconómicos, museológicos, etc.

Ainda no domínio dos arquivos, na norma relativa à descrição dos produtores, a mesma preocupação traduz-se na existência do elemento de informação “Formas autorizadas do nome de acordo com outras regras”, que a ser utilizado facilita a partilha de registos de autoridade entre diferentes domínios, nomeadamente bibliotecas e museus.

No caso das bibliotecas, a ISBD assume, de forma expressa, a preocupação de aumentar a interoperabilidade com outras normas de conteúdo. No caso dos museus, na CIDOC e nas Normas de inventário, é assumida a preocupação de fornecer uma base de comparação com outras normas nacionais e internacionais. Por sua vez a CIDOC-CRM assume uma vocação de harmonização de diferentes modelos de descrição.

4.5.9 Centrando-nos numa análise mais detalhada dos elementos de informação / classes e das respetivas correspondências entre as diferentes normas, com base no mapeamento que corresponde ao anexo no separador *normas_map*, pode concluir-se o seguinte:

- As normas analisadas, e centrando-nos primeiro nas categoriais, pese embora a especificidade que decorre dos diferentes objetos para os quais foram elaboradas, são bastante equivalentes.
- Verifica-se que os elementos ou as categorias de informação podem estar preparados para o registo mais ou menos atomizado dos diferentes tipos de informação. Este aspeto - maior ou menor número de elementos para o registo da informação - nem sempre parece estar relacionado com o grau de

relevância atribuído à informação a registar. Veja-se, a título exemplificativo, o caso do *Título*: as normas arquivísticas preveem a obrigatoriedade do registo de um *Título*, pelo que disponibilizam um único elemento de informação, sendo os restantes, caso existam, remetidos para o elemento de informação *Notas*. Bibliotecas, museus e televisão preveem diferentes elementos de informação para registar esta informação. De igual forma está prevista, nos arquivos, a indicação do *tipo de Título* registado, sem que exista um elemento de informação específico para esse fim, ao contrário do que se verifica para os museus, que preveem, para além do elemento *Title*, o elemento *Title type*.

Um outro exemplo é o caso, nos arquivos, do elemento *Fonte imediata de aquisição ou transferência*, cujo objetivo é o de identificar a fonte imediata de aquisição ou transferência da unidade de descrição. Corresponde, nos museus, aos elementos de informação *Acquisition method*, *Acquisition date*, *Acquisition source* (Acquisition information group) e *Current owner*, *Depositor*, *Entry date*, *Entry number*, *Entry reason* (Object entry information group).

- Algumas das especificidades detetadas prendem-se com o âmbito das normas. A título meramente exemplificativo, veja-se o caso das bibliotecas, que contemplam elementos de informação específicos para *Dados matemáticos (recursos cartográficos)* e *Apresentação musical (música notada)*. Este tipo de informação é igualmente relevante no domínio dos arquivos. O que se verifica é que tanto as normas internacionais do ICA como as normas portuguesas sobre elas desenvolvidas são gerais, não contemplando documentos especiais. Em caso de necessidade este tipo de informação é registado em elementos de carácter mais genérico.
- Existem outras especificidades nas normas analisadas que não decorrem do respetivo objeto. Sirvam de exemplo os elementos destinados ao registo da informação relativa à atual localização dos objetos e espécimes, nos museus, que não tem paralelo nas demais normas analisadas. Embora seja igualmente considerada relevante no domínio dos arquivos, é objeto de registo autonomizado. Verifica-se exatamente o mesmo tipo de situação para as intervenções de conservação e restauro.
- Por outro lado, há a mencionar a existência de elementos de descrição vocacionados para informação equivalente mas que, pelas diferentes características dos objetos a que se reportam, assumem especificidade. Sirvam de exemplo *Fonte imediata de aquisição ou transferência* e *Acquisition method*, respetivamente no domínio dos arquivos e dos museus, já anteriormente mencionados, por contraposição a *Modalidades de aquisição*, nas bibliotecas, que se destina a acomodar informação relativa ao preço ou a outras modalidades de aquisição, se o recurso não estiver para venda.
- Existem, no entanto, elementos especialmente vocacionados para determinados tipos de objetos. É o caso, nas bibliotecas, dos que constituem a *Zona de Edição* ou a *Zona de publicação, produção, distribuição, etc.*, esta última com correspondência também nas normas utilizadas pela Televisão.
- Quanto às normas combinatórias, a utilização da Dublin Core encontra-se testada no âmbito de vários portais nacionais e internacionais. Quanto à

CIDOC-CRM, para além do mapeamento disponibilizado no anexo que corresponde ao separador *normas_map*, está prevista a realização de testes práticos, no sentido de determinar da sua maior ou menor adequação à representação dos objetos das diferentes CdP consideradas.

4.5.10 Relativamente à informação clínica, foram identificadas várias normas específicas, tendo sido incluída, no anexo que corresponde ao separador *normas*, a referenciada como de utilização prática: OpenEHR Foundation - *Open European Health Record*. No sentido de perceber até que ponto, e dada a especificidade apontada para a informação clínica e, mais precisamente, para os processos clínicos, existe ou não correspondência entre ela e as outras normas identificadas, e para apurar a possibilidade de mapeamento para uma estrutura comum, foram analisados casos práticos - processos clínicos concretos, devidamente enquadrados no respetivo contexto de produção. Da análise efetuada foi possível concluir que, apesar da sua complexidade, as ODA respondem às necessidades de descrição, quer no que respeita aos níveis de descrição necessários, quer no que se reporta aos elementos de informação para acomodar a informação necessária.

5 Formatos

5.1 O objetivo desta atividade consistia em conhecer os formatos de arquivo digital (formatos) usados, de forma genérica, pelas diversas CdP.

Metodologia

5.2 Foram considerados todos os formatos passíveis de utilização, independentemente do fim para que foram criados (acesso ou preservação), assim como uma amostra, o mais representativa possível, do universo das CdP detentoras de património digital.

INQUÉRITO:

5.3 Com o objetivo de serem referenciados pelas diferentes CdP foi elaborada, num ficheiro excel, uma lista de formatos, aberta (em que as CdP podiam acrescentar os formatos que entendessem) e organizada por categorias (não definidas de forma rígida mas apenas no sentido de permitir maior facilidade conceptual). Essa lista resultou do exposto no sítio web *The Digital Formats Web site*

[<http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml>] da Biblioteca do Congresso norte-americano e de contributos, *ad hoc*, de colaboradores do projeto *Continuidade digital*. Concretamente foi listado um total de 176 formatos estruturado por 6 categorias (Imagens fixas, som, texto, genéricos, imagens em movimento e datasets) das 8 previstas (Geoespacial e Arquivos web não foram contempladas) no *The Digital Formats Web site*. Nesta lista incluíram-se as versões dos formatos apenas no sentido de ajudar à identificação

dos mesmos, tendo-se sempre optado por trabalhar qualquer conclusão, em função dos grandes grupos de formatos.

AMOSTRA

5.4 O inquérito foi dirigido às nove CdP, com representação no GT do projeto *Continuidade digital*:

- Arquivos;
- Bibliotecas
- Ciência/Saúde;
- Cinema;
- Jogos eletrónicos;
- Multimédia
- Museus;
- Televisão
- Música

Duas CdP não representadas no GT foram ainda consideradas, embora as suas respostas incidissem apenas nos formatos específicos da sua área de atuação:

- Fotografia (imagem fixa);
- Rádio (som, genéricos)

RESULTADOS (alguns aspetos mais salientes)

5.5. TOTAL DE FORMATOS UTILIZADOS

Um total de 54¹ formatos é utilizados pelo conjunto das CdP inquiridas. Neste ponto importa realçar que as duas CdP não presentes no GT (Fotografia e Rádio) responderam essencialmente quanto aos formatos relativos à sua área de especialidade (imagem fixa e som²). Acrescente-se ainda que se optou por assinalar apenas por uma vez os formatos que são enquadrados em várias categorias³;

Lista de formatos ordenada alfabeticamente:

AAF	DOC	JPEG 2000 Encodings	RealMedia
AC-3	DOCX	JPEG 2000 File Formats	RIFF
ASF (som: genérico; imag. mov.)	DPX	MP2	TIFF

¹ O formato TXT foi acrescentado, pela CdP Música, à lista referida no sítio web *The Digital Formats Web sit* (ver p. 19)

² A CdP também respondeu ainda relativamente à categoria genéricos

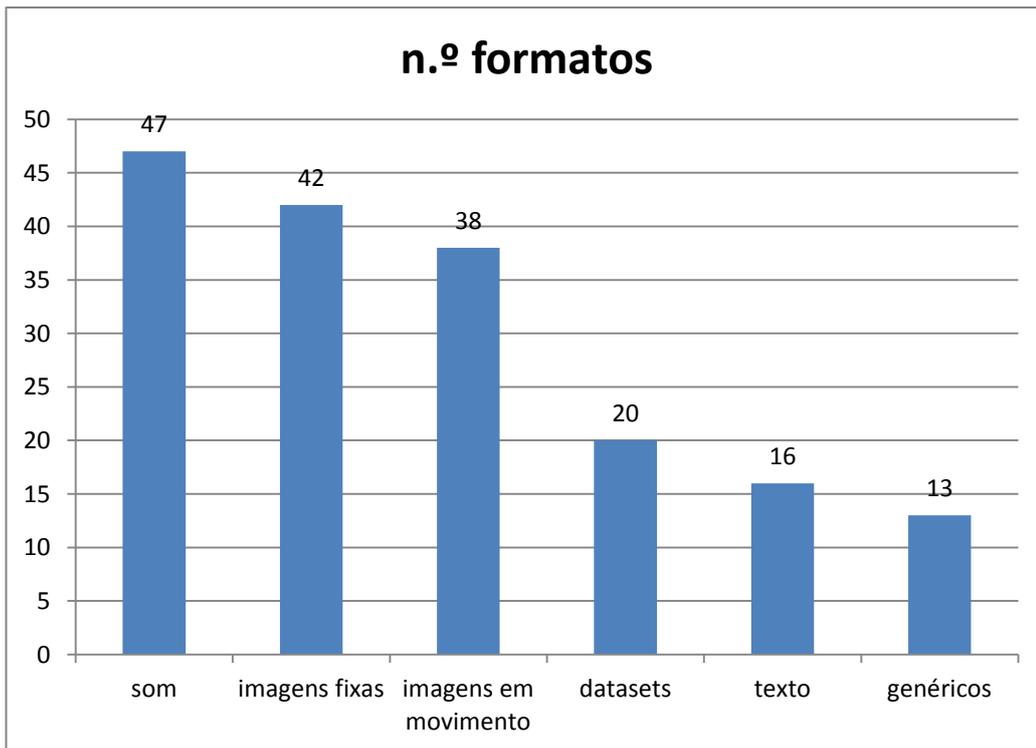
³ Por exemplo o formato ASF inclui-se em 3 categorias (som: genérico; imag. mov.), o formato XML em 2 (texto; genérico)...

AUTOCAD	DTS	MP3	TXT
AVI	DV	MPEG-1	Uncompressed video encodings
BigTIFF	FLAC	MPEG-2	WAV
BWF	Flash (SWF, FLA, FLV)	MPEG-4 File Formats with Encoded Bitstreams	WAVE
Cinepak	H.26n ITU-T video encoding standards	MPEG-4 Video Encodings	WM (Windows Media) (som; imag. mov.)
CSV	HDCAM	MXF	XLS
DGW	ID3	PCM	XLS (linux)
DICOM	Indeo	PDF	XML (texto; genérico)
Digital Betacam, AKA Digibeta or D-Beta	ISO_8211	PNG	ZIP
DivX	ISO_image	PPT	
DNG	JPEG	QuickTime (som; imag. mov.)	

5.6 N.º de formatos em função das categorias inventariados:

Contrariamente ao ponto anterior neste respeitou-se cada valência dos formatos em termos de categoria;

categoria	n.º formatos
som	47
imagens fixas	42
imagens em movimento	38
datasets	20
texto	16
genéricos	13



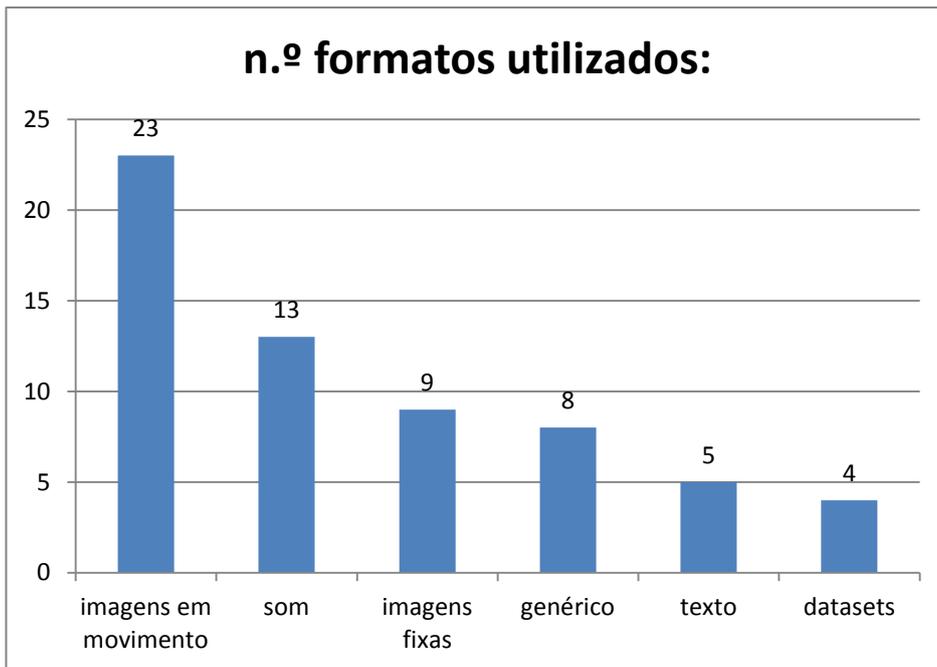
5.7 Formatos utilizados por todas as CdPs

Nenhum formato é utilizado pela totalidade das CdP. Refira-se que, para este ponto, foram consideradas apenas as CdP que responderam exhaustivamente ao inquérito;

5.8 N.º de formatos mais utilizados por categoria:

Imagens em movimento, com 23 ocorrências, foi a categoria que registou maior n.º de formatos utilizados. A categoria *som*, registou 13 formatos, e a categoria *imagens fixas* 9. De mencionar que aqui optou-se por manter o desdobramento dos formatos que são incluídos em mais do que uma categoria;

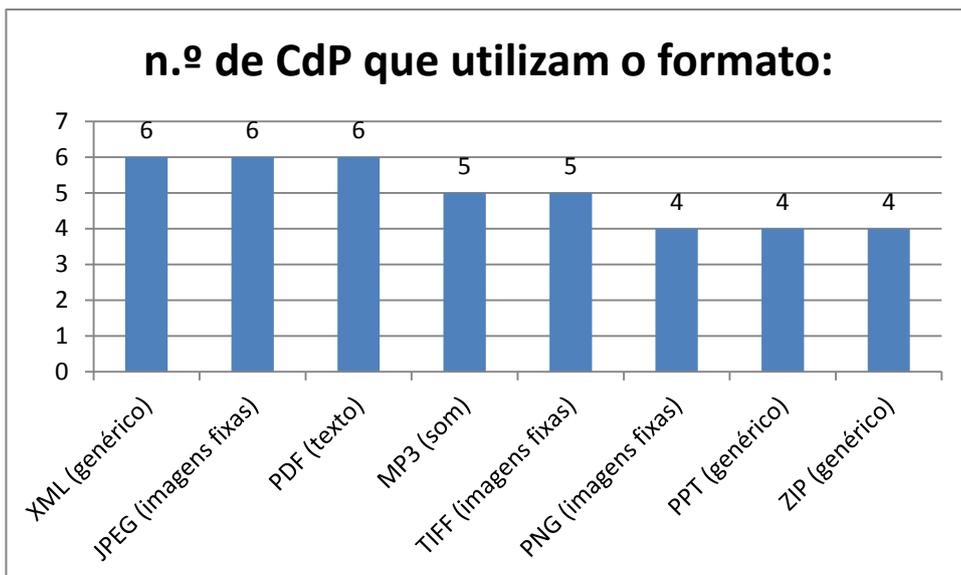
categoria	n.º formatos utilizados:
imagens em movimento	23
som	13
imagens fixas	9
genérico	8
texto	5
datasets	4



5.9 Formatos utilizados, por maior n.º de CdP:

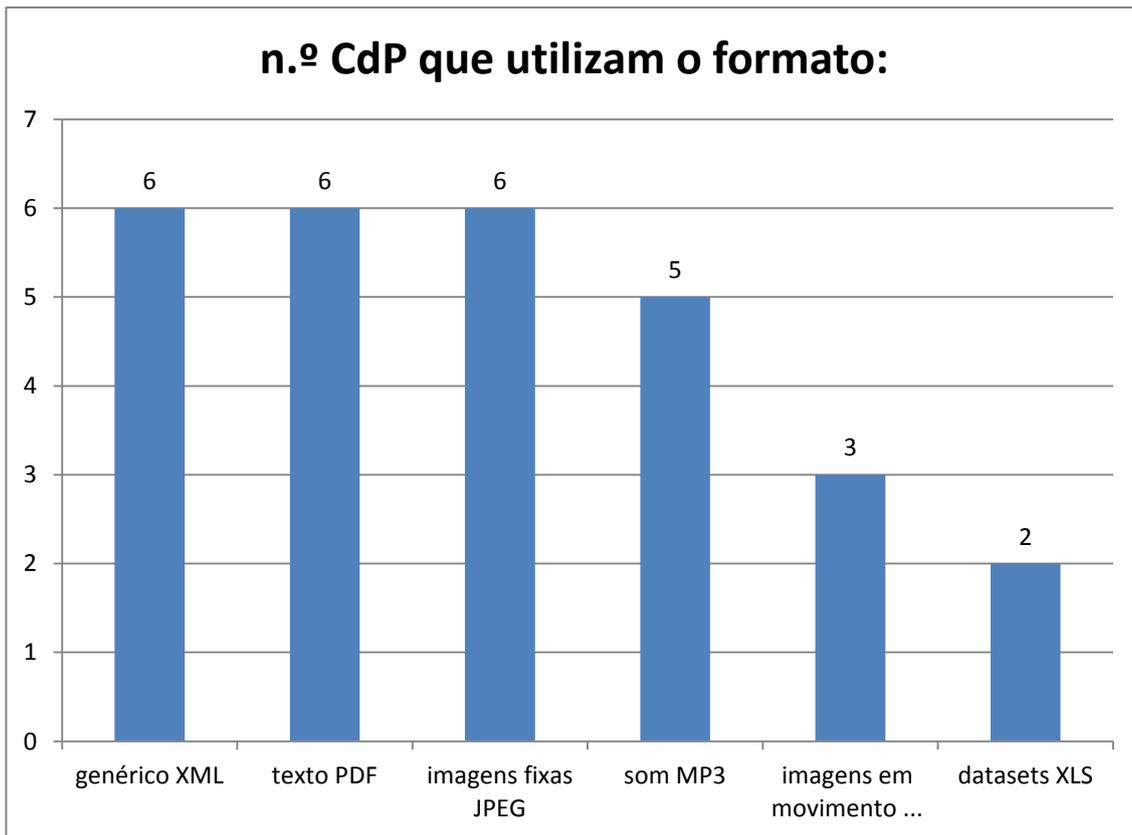
Os formatos XML (categoria genérico), PDF e JPEG utilizados por 6 CdP são os mais usados pelo conjunto das diferentes CdP. Os formatos MP3 e TIFF foram referidos por 5 CdP enquanto os formatos, PNG, PPT e ZIP foram mencionados por 4 CdP. Neste ponto optou-se também por manter o desdobramento dos formatos que são incluídos em mais do que uma categoria;

formato	categoria	n.º de CdP que utilizam o formato:
XML	genérico	6
JPEG	imagens fixas	6
PDF	texto	6
MP3	som	5
TIFF	imagens fixas	5
PNG	imagens fixas	4
PPT	genérico	4
ZIP	genérico	4



5.10 Formato mais utilizado, por categoria:

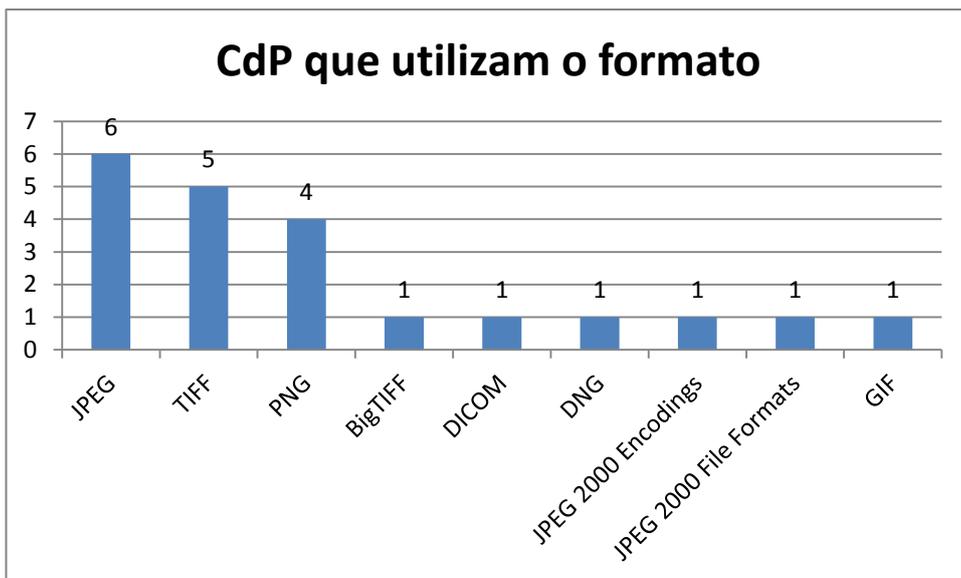
categoria	formato	n.º CdP que utilizam o formato:
genérico	XML	6
texto	PDF	6
imagens fixas	JPEG	6
som	MP3	5
imagens em movimento	AVI + MPEG-4 Video Encodings + Quicktime	3
datasets	XLS	2



5.11 Formatos utilizados por CdP em função da categoria (síntese)

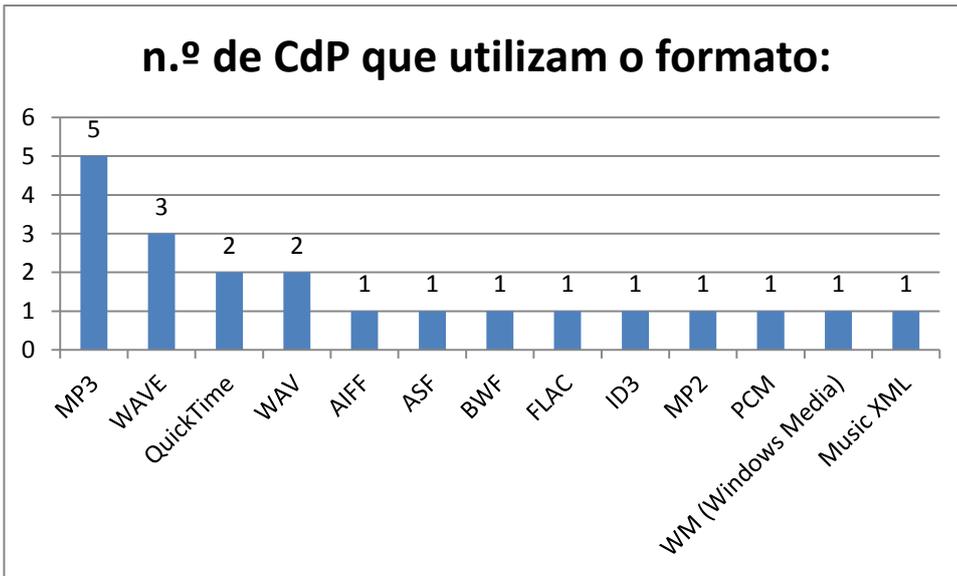
5.11.1 Imagens Fixas (9 formatos referidos):

formatos referidos:	CdP que utilizam o formato	CdPs:
JPEG	6	arquivos; bibliotecas; ciência/saúde; cinema; museus; musica
TIFF	5	arquivos; bibliotecas; ciência/saúde; cinema; musica
PNG	4	arquivos; ciência/saúde; cinema; musica
BigTIFF	1	
DICOM	1	
DNG	1	
JPEG 2000 Encodings	1	
JPEG 2000 File Formats	1	
GIF	1	



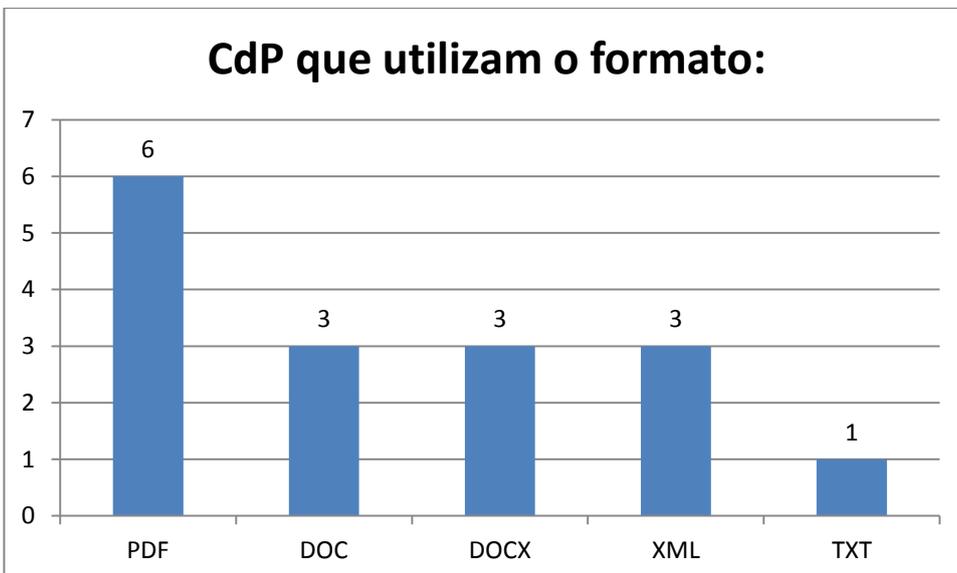
5.11.2 Som (13 formatos referidos):

formatos referidos:	n.º de CdP que utilizam o formato:	CdPs:
MP3	5	arquivos; ciência/saúde; museus; rádio; musica
WAVE	3	museus; rádio; musica
QuickTime	2	museus; musica
WAV	2	museus; rádio
AIFF	1	
ASF	1	
BWF	1	
FLAC	1	
ID3	1	
MP2	1	
PCM	1	
WM (Windows Media)	1	
Music XML	1	



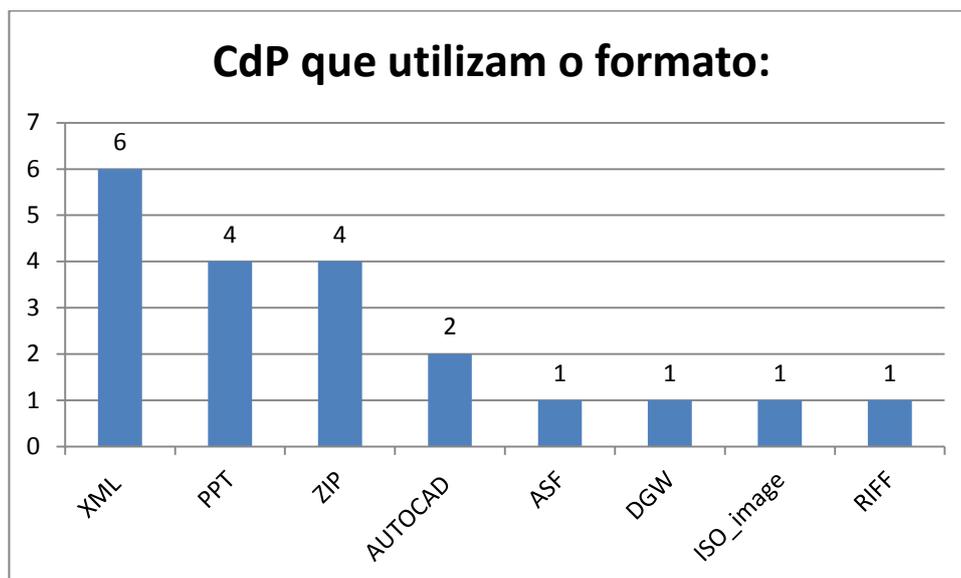
5.11.3 Texto (5 formatos referidos):

formatos referidos:	CdP que utilizam o formato:	CdPs:
PDF	6	arquivos; bibliotecas; ciência/saúde; cinema; museus; musica
DOC	3	arquivos; ciência/saúde; museus;
DOCX	3	arquivos; ciência/saúde; museus;
XML	3	arquivos; ciência/saúde; museus;
TXT	1	musica;



5.11.4 Genérico (8 formatos referidos):

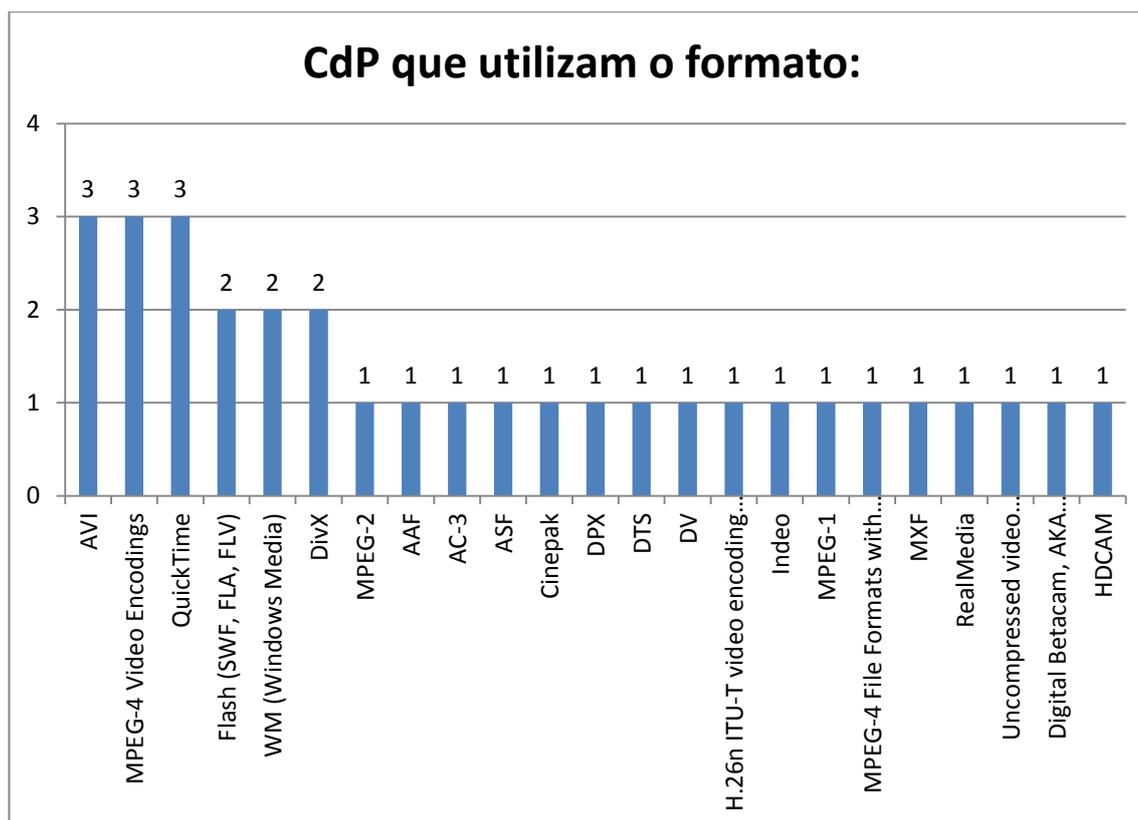
formatos referidos:	CdP que utilizam o formato:	CdPs:
XML	6	arquivos; bibliotecas, ciência/saúde; museus; rádio; tv; música
PPT	4	arquivos; ciência/saúde; museus; tv
ZIP	4	arquivos; ciência/saúde; tv; música
AUTOCAD	2	museus; tv;
ASF	1	
DGW	1	
ISO_image	1	
RIFF	1	



5.11.5 Imagens em movimento (23 formatos referidos):

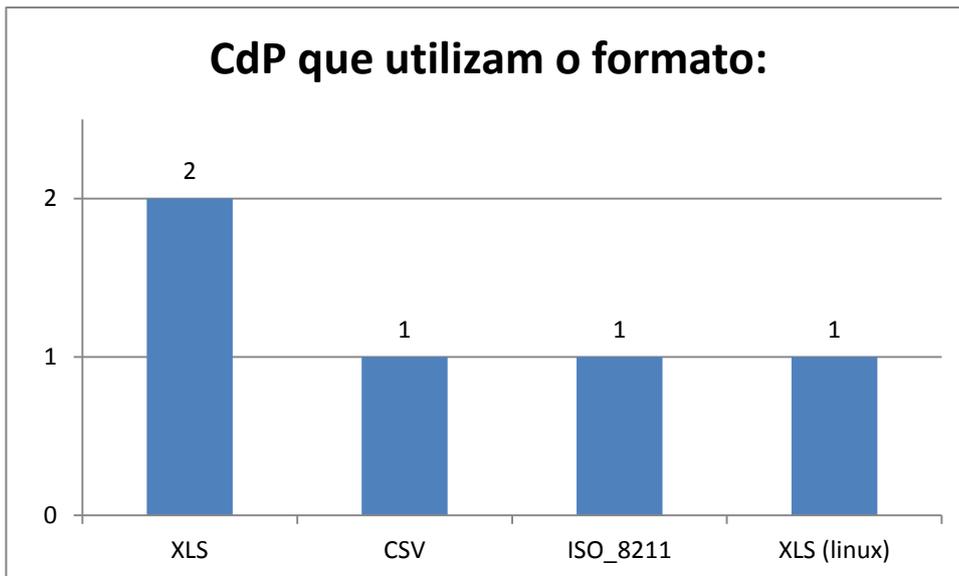
formatos referidos:	CdP que utilizam o formato:	CdPs:
AVI	3	arquivos; museus; tv
MPEG-4 Video Encodings	3	arquivos; cinema; tv
QuickTime	3	arquivos; tv; música
Flash (SWF, FLA, FLV)	2	Museus; música
WM (Windows Media)	2	arquivos; tv
DivX	2	Museus: tv
MPEG-2	1	
AAF	1	
AC-3	1	
ASF	1	
Cinepak	1	
DPX	1	
DTS	1	
DV	1	
H.26n ITU-T video encoding standards	1	
Indeo	1	

MPEG-1	1
MPEG-4 File Formats with Encoded Bitstreams	1
MXF	1
RealMedia	1
Uncompressed video encodings	1
Digital Betacam, AKA Digibeta or D-Beta	1
HDCAM	1



5.11.6 Datasets (4 formatos referidos):

formatos referidos:	CdP que utilizam o formato:	CdPs:
XLS	2	arquivos; museus
CSV	1	
ISO_8211	1	
XLS (linux)	1	



Conclusões

Com base na resposta ao inquérito promovido a uma amostra de CdP detentoras de património digital (Arquivos; Bibliotecas Ciência/Saúde; Cinema; Fotografia; Jogos eletrónicos; Multimédia; Museus; Música, Rádio e Televisão) é possível salientar:

- Um total de 54 formatos utilizados;
- Nenhum formato referenciado por todas as CdP;
- A categoria com maior n.º de formatos é *Imagens em movimento* registando 23 diferentes tipos de formato utilizados, seguindo-se *som* com 13, *imagens fixas* com 9 e *genérico* com 8. As categorias com menos tipos de formato referenciados são *texto* e *datasets* com 5 e 4 respetivamente;
- Os formatos XML (identificado na categoria *genérico*), PDF e JPEG são os mais utilizados pelas diferentes CdP. Seguem-se os formatos MP3 e TIFF referidos por 5 CdP enquanto os formatos PNG, PPT e ZIP foram mencionados por 4 CdP;
- Em cada categoria os formatos mais referenciados são:
 - genérico - XML (6);
 - texto - PDF (6);
 - imagens fixas - JPEG (6);
 - som - MP3 (5);
 - imagens em movimento - AVI/MPEG-4 Video Encodings e QuickTime (4);
 - datasets - XLS (2)

6. Conclusões

A análise comparativa efetuada nos diversos tópicos que constituem este passo, aponta para um conjunto de pontos comuns que importa destacar.

Da legislação analisada a Lei de Bases do Património Cultural define requisitos de preservação aplicáveis transversalmente a informação de valor patrimonial, sendo definido como tal uma significativa diversidade de informação inserível em diversos domínios culturais e produzida por diferentes CdP. Legislação específica como a Lei da Televisão e da Rádio, aponta igualmente para a necessidade de preservar material, embora não seja especificamente referido o suporte digital. A questão dos direitos de autor é extensível a todos os domínios culturais, sendo uma questão estruturante que surge com particular relevância quando aplicada em ambiente digital.

As terminologias analisadas apontam para uma aproximação sempre que o foco das mesmas é a representação do objeto. Divergem no entanto quando estas estruturas são centradas no aspeto executivo e de apoio à atividade centrada no objeto. Esta constatação parece lógica e, a nosso ver, não deve ser particularmente valorizada, uma vez que a preservação de património digital incide claramente sobre o primeiro foco visto que, em princípio pelo menos, os objetos preservados não se encontram já operacionalmente ativos.

Dentro das terminologias orientadas para a representação do objeto há conceitos partilhados, embora possam surgir de forma lexicalmente diferente.

Normativamente é de salientar a existência de elementos básicos comuns, independentemente do domínio cultural em que são aplicados. Alguns testes efetuados permitiram realizar descrições ou catalogações “cruzadas” utilizando uma norma aplicável a objetos provenientes de diferentes domínios culturais. A necessidade de contextualizar o objeto patrimonial, bem como a representação hierárquica, tradicionalmente consideradas como singularidades arquivísticas, manifestaram-se presentes noutras comunidades de prática como os museus e as bibliotecas. As normas combinatórias revelam-se suficientemente poderosas para descrever qualquer objeto independentemente da via de análise adotada (centrada no evento ou no objeto). Na área do jornalismo e da música as observações realizadas permitem concluir haver características passíveis de ser representadas com recurso às normas analisadas. Podemos assim concluir ser possível interoperabilidade entre os vários esquemas descritivos utilizados sem perda de sintaxe ou semântica. A utilização de normas diferentes de acordo com as características do objeto a ser representado, bem como as necessidades específicas do público, não é impeditiva, portanto, de mapear elementos semanticamente idênticos numa estrutura única.

Os formatos referidos pelos membros do grupo de trabalho não apresentam características de especial singularidade, sendo formatos utilizados correntemente. Os que apresentam uma maior especificidade são utilizados na comunidade médica (DICOM) e na televisão, com formatos de tapes digitais normalmente utilizados em contextos de captação profissional de imagens em movimento (Digital Betacam, AKA Digibeta or D-Beta, HDCAM). Destacamos a quase unanimidade declarada da utilização da metalinguagem XML o que é um aspeto relevante se considerarmos que esta metalinguagem fundamenta muitos esquemas normalizados empregues em preservar objetos digitais.

Foram referidos poucos formatos na categoria de “datasets”, situação que provavelmente se alterará com a presença da comunidade de prática científica e ainda com a exploração posterior dos formatos utilizados pela comunidade médica, onde a informação clínica do doente, nas diversas vertentes administrativa, técnica e assistencial é normalmente armazenada em bases de dados. A heterogeneidade identificada dos formatos utilizados permite confirmar a necessidade de um controlo sólido dos mesmos, bem como da respetiva evolução de mercado, e ainda de identificação de formatos aceitáveis sob o ponto de vista de utilização e preservação continuada.