

A criação de uma rede de preservação comum de património digital



Continuidade Digital

Aspetos a abordar

0. Diagnóstico prévio
1. Antecedentes, grupo de trabalho, objetivos, ponto de partida
2. Análise transversal para as CdP: convergência ou divergência?
3. Modelo de custos
4. Governança
5. Ponto de situação atual
6. Perspetivas futuras

(dados recolhidos 2014-2015)



Continuidade Digital

0. Diagnóstico prévio

- Omnipresença das TIC
- Produção administrativa, científica, artística – mediação tecnológica
- Muitos objetos patrimoniais são nado digitais
- Obsolescência do sistema intermediário
- Ausência de ferramentas para o substituir



Necessidade de assegurar preservação



Continuidade Digital

0. Diagnóstico prévio

Necessidade de assegurar preservação digital

- problema premente e transversal
- ausência de mercado de tecnologias direcionadas à preservação digital
- constrangimentos orçamentais das entidades públicas e privadas dificultam a abordagem individualizada





Continuidade Digital

0. Diagnóstico prévio

Preservação digital envolve:

- produtores, detentores e utilizadores da informação



Preservação digital exige:

- estratégia concertada
- recursos financeiros elevados
- conhecimentos especializados

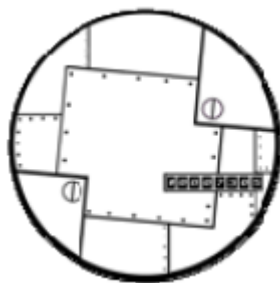




Continuidade Digital

0. Diagnóstico prévio

DGLAB - RODA:



Repositório de Objetos Digitais Autênticos

2006-2007

articulação com a política de gestão de documentos da Administração Pública

Continuidade Digital

1. Antecedentes:

Objetivos

Conclusões

Continuidade

Disponíveis em:

<http://1seminariopreservacaopatrimoniodigital.dglab.gov.pt/>

1º SEMINÁRIO de PRESERVAÇÃO COMUM de PATRIMÓNIO DIGITAL

19 e 20 de Setembro de 2013





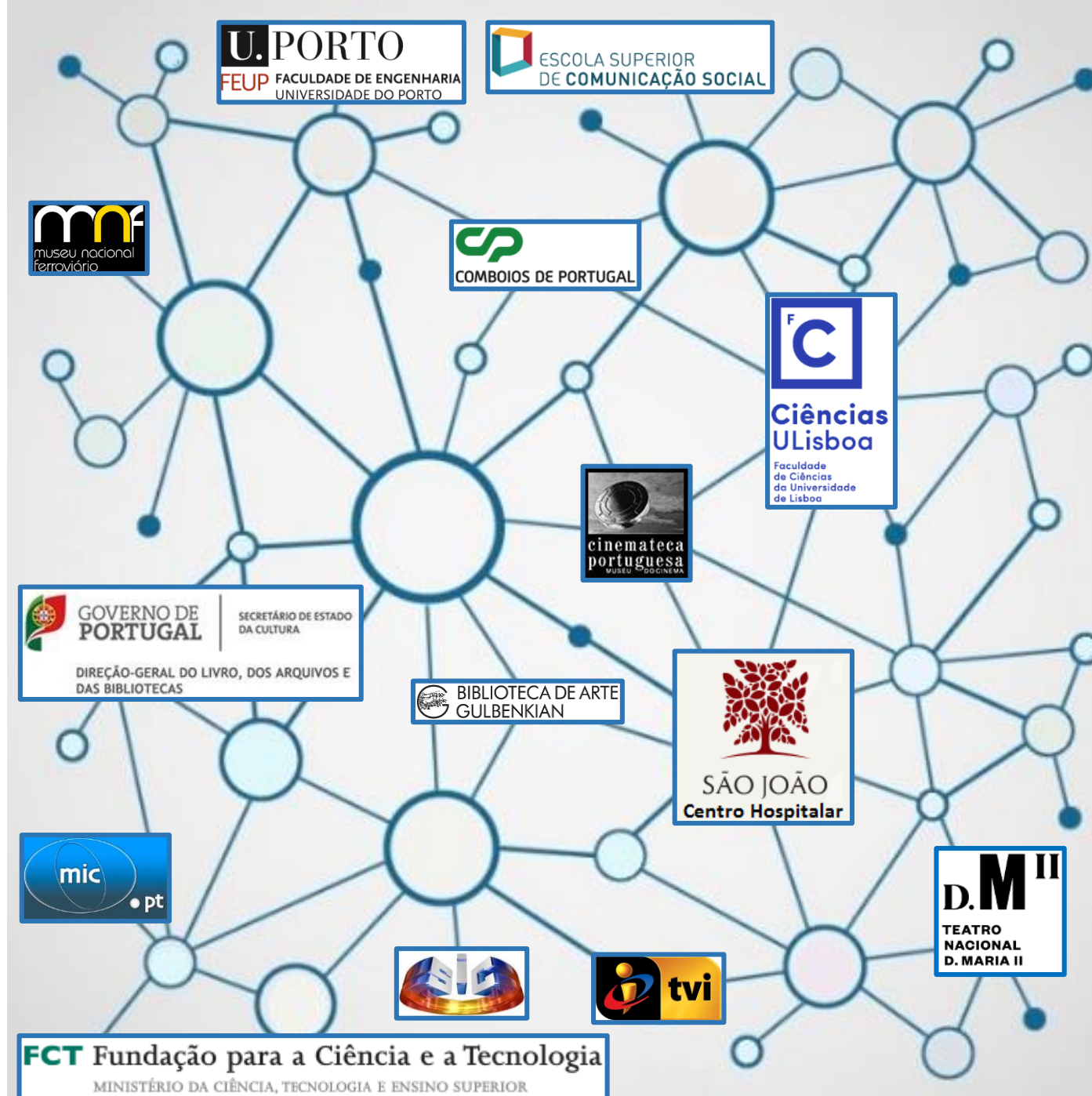
Continuidade Digital

1. Grupo de trabalho:

Iniciativa: DGLAB

Base:
participantes no 1.º seminário

Características:
transversalidade (diversidade
de CdP)



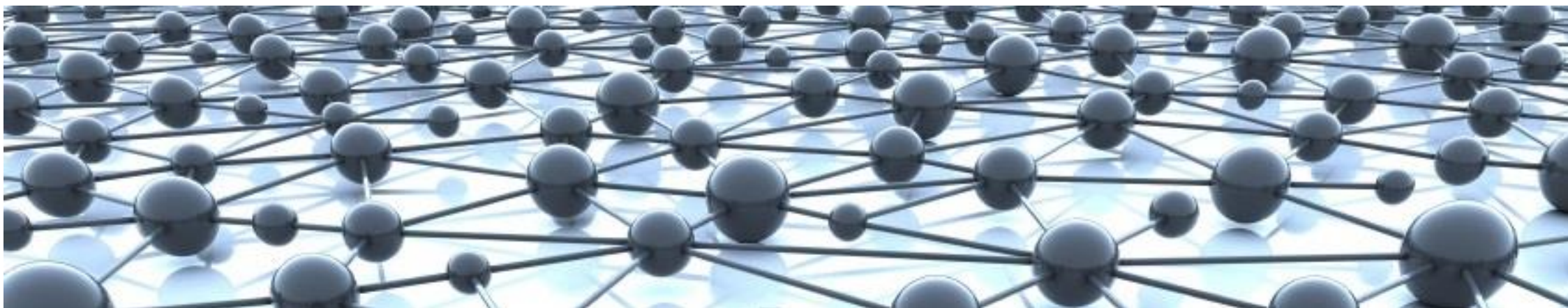


Continuidade Digital

1. Objetivo:

analisar a viabilidade de uma rede comum para preservar património digital

- estruturas comuns
- cooperação
- partilha: recursos, custos, conhecimento, responsabilidade





Continuidade Digital

1. Ponto de partida

vontade efetiva de preservar património digital

preservar património digital = preservar objetos físicos (ficheiros) que veiculam informação digital

homogeneidade da natureza da informação em ambiente digital:
(codificação binária)



domínio de implementação

=

limite máximo de convergência dos atores



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

0. Metodologia

I Quadro regulador

II Formatos utilizados

III Valores

IV Práticas





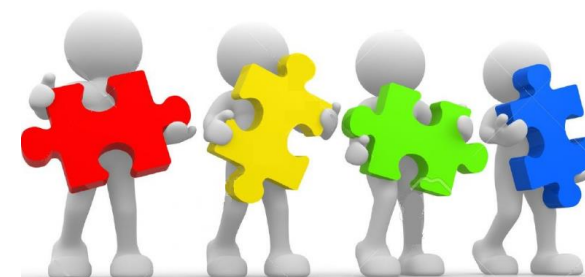
Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

0. Metodologia:

Levantamento e recolção de dados, posteriormente analisados

- debate participado com os membros do GT
- levantamento documental e de fontes de autoridade
- inquéritos: 1.º fechado, exclusivamente elementos do GT; 2.º dirigido a uma comunidade alargada





Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

I Quadro regulador - Legislação

Transversal, nacional e europeia:

- Lei dos Direitos de Autor e Conexos



- Lei do Depósito Legal

- Licenças Públicas *Creative Commons sobre dados espaciais*





Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

I Quadro regulador - Legislação

Multidomínio:

Atividade jornalística - Imprensa, Rádio e Televisão





Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

I Quadro regulador - Legislação

Referências explícitas a preservação de património/digital:

- Art.º 11.º, Lei de Bases do Património Cultural - “dever de preservação, defesa e valorização do Património Cultural”
- Cap. VII, Lei da Televisão, n.º 1, 2 e 3 do art.º 92.º - “conservação do Património Televisivo”
- Cap. VII, Lei da Rádio, art.º 83.º - “conservação do Património Radiofónico - registos de interesse público”
- Portaria 247/2000 - Regulamento de conservação arquivística para os Hospitais e Serviços de Saúde. Processo clínico é de conservação permanente

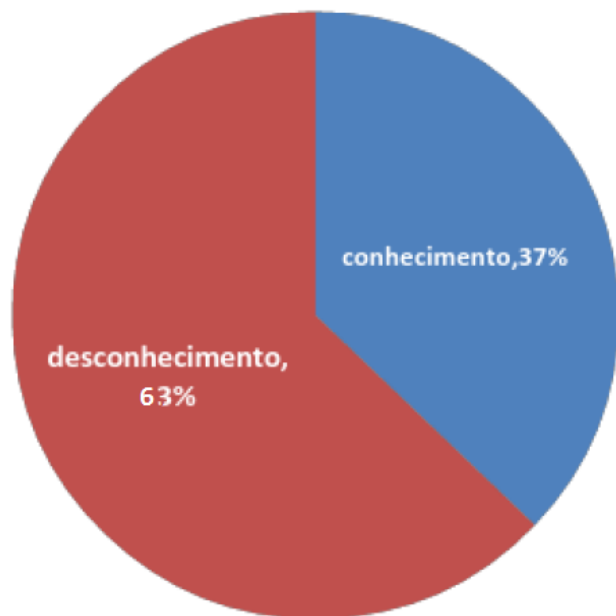


Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

I Quadro regulador - Legislação

Acentuado desconhecimento das CdP sobre regulação do património digital:



mas: fontes de autoridade mais utilizadas: critérios de avaliação de ODs - diplomas legais e regulamentos

Exceções: Museus (3.ª fonte mais utilizada)
Outras (2.ª fonte mais utilizada)

2. análise transversal para as CdP: convergência ou divergência?

I Quadro regulador - Terminologias

Ponto de partida da análise:

- vocabulário nuclear estabelecido no documento estratégico do projeto
- localização dos termos nas estruturas vocabulares identificadas
- presença, frequência, ou ausência = peso do conceito no âmbito da CdP





Continuidade Digital

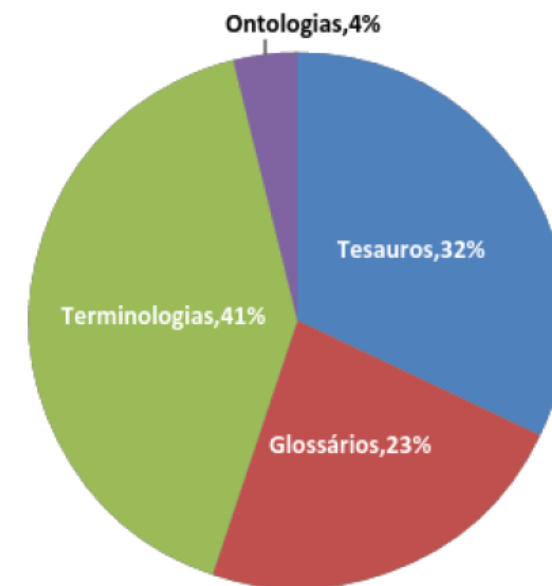
2. análise transversal para as CdP: convergência ou divergência?

I Quadro regulador - Terminologias

instrumentos de indexação mais utilizados:

2 grandes classes de estruturas vocabulares:

- apoio à execução de atividades: termos muito específicos e praticamente exclusivos da CdP a que se destinam
- apoio à descrição ou representação de atividades: maior aproximação e similitude de termos entre CdP





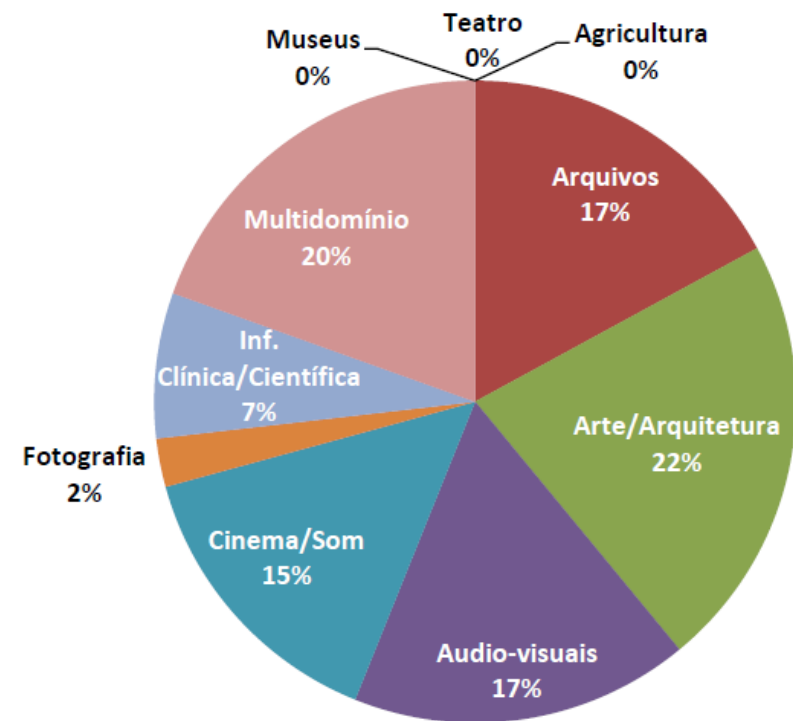
Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

I Quadro regulador – Terminologias

termos mais usados pelo conjunto das CdP:

- Acesso / Direitos de autor
- Avaliação / Preservação digital
- Património digital
- Autenticidade / Perspetiva de usabilidade
- Custódia / Certificação e segurança do repositório /
/ Arte digital
- Identificação de património digital



número de termos encontrados por domínio



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

I Quadro regulador - Normas

identificadas ca. 105

- normas, regras, guias, orientações, recomendações, etc.
- diferentes graus de pertinência
- internacionais / nacionais
- gerais / específicas
- 3 grandes classes: Arquivos, Bibliotecas e Museus



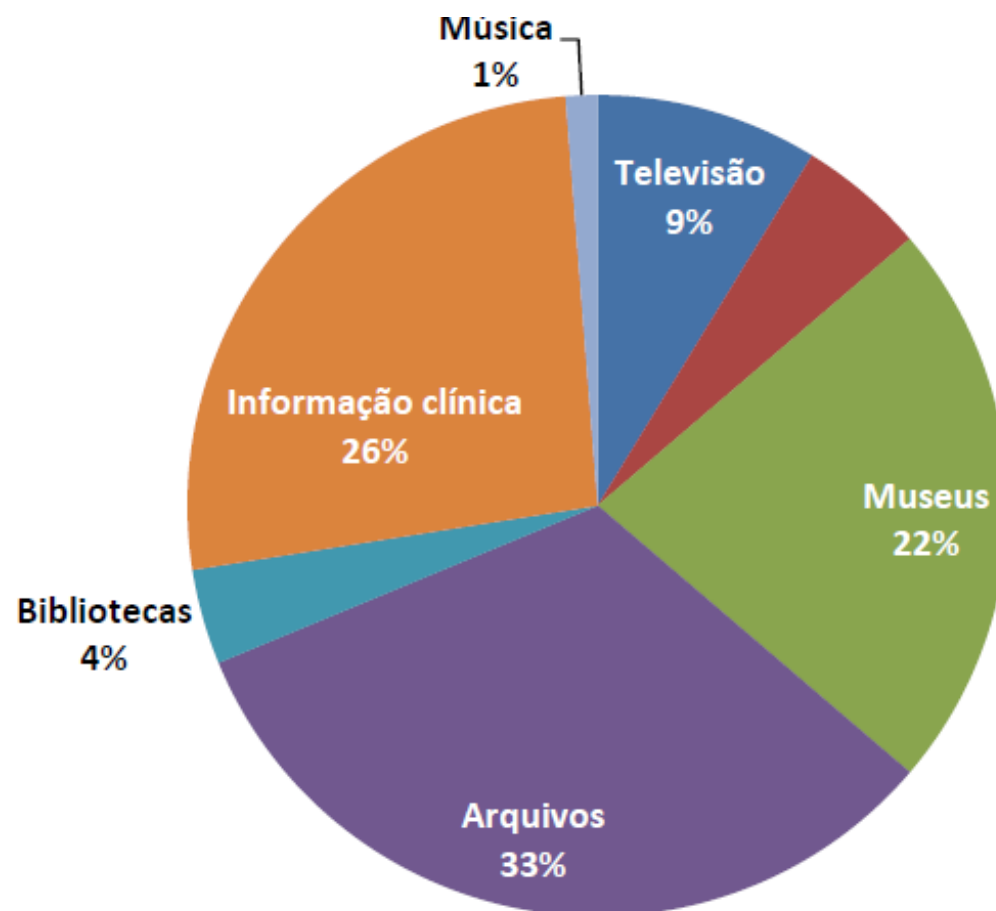


Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

I Quadro regulador - Normas

repartição pelas CdP representadas no GT:





Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

I Quadro regulador - Normas

Normas partilhadas	CdP
ISAD (G)	Arquivos, Televisão
ISBD	Bibliotecas, Cinema
EDM	Arquivos, Museus
DC	Arquivos, Música

Análise comparativa:

- ISDIAH/EAG
- ISAD(G)/EAD2
- ISAAR(CPF)/EAC (CPF)
- ODA 3
- DC; MPEG 7
- ISBD; CIDOC
- CIDOC-CRM
- Normas de Inventário
- EBU-TECH 3293



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

I Quadro regulador – Normas – Conclusões:

- Maioria centra-se, quanto ao objeto, num domínio patrimonial específico
- Heterogeneidade de práticas descritivas
- Perspetiva vertical do património
- **Exceções:** CIDOC-CRM e Dublin Core (assumem-se como multidomínios)



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

I Quadro regulador – Normas – Conclusões:

- Contexto de produção – relevância não exclusiva para arquivos (ex. museus, bibliotecas)
- Representação multinível não é exclusiva dos arquivos (ex. museus, bibliotecas, informação clínica)
- Objetivos comuns
- Experiências de descrição cruzadas - possibilidade de descrever objetos de um domínio patrimonial com uma norma aplicada noutra (ex. informação clínica)
- Mapeamentos - atestam possibilidade de mapeamento para uma estrutura única



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

I Quadro regulador - Normas - Conclusões

Estruturação da informação

a) **Categoriais:** Arquivos, Informação clínica, Bibliotecas e Museus

- Centradas no objeto
- Zonas ou grupos + elementos ou categorias de informação
- Bastante equivalentes (quanto aos elementos de informação definidos)

b) **Combinatórias** (Dublin Core, a CIDOC-CRM, a EBU-TECH 3293 (baseada na Dublin Core) e a MPEG 7), metadados, esquema de metadados e linguagem de definição de descrição



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

I Quadro regulador – Normas – Conclusões:

Registo de informação

Arquivos, Informação clínica e Museus definem forma como a informação deve ser registada nos elementos ou categorias de informação - regras e convenções específicas



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

II Formatos

Total de 68 formatos utilizados pelas diferentes CdP

- Arquivos: 57 formatos
- Bibliotecas: 53 formatos
- Museus: 46 formatos
- “Outros”: 50 formatos

54 formatos utilizados por mais do que uma CdP

38 formatos utilizados por todas as CdP

Categorias previstas pelo sítio web da Biblioteca do Congresso:

Imagens fixas

Som

Texto

Genérico

Imagens em movimento;

Datasets

Arquivo WEB

Geoespacial



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

II Formatos - Conclusões

Não apresentam especial singularidade

Maior especificidade: Comunidade médica e na Televisão

Unanimidade da utilização da metalinguagem XML

Heterogeneidade:

- necessidade de controlo
- necessidade de acompanhar a sua evolução de mercado
- necessidade de identificar formatos aceitáveis visando a preservação continuada





Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

III Valores: autenticidade

1/ relevância:

- Elementos internos: **Identidade**, Identificação, **Integridade**, **Conteúdo**, Contexto, Usabilidade e **Estrutura**
- Elementos externos: **Tecnológico**, **Procedente**, **Documental**, Jurídico, **Administrativo** e Procedimental



Mais prevalecente; mais valorado

Menos prevalecente; menos valorado



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

III Valores: autenticidade

2/ admissibilidade de modificações e tipo de modificações (sem perda de autenticidade):

as CdP reconhecem: **não é possível preservar sem alterar**

as CdP defendem: alterações criteriosas

- **rejeitadas: alterações de conteúdo**

- **aceites: alterações de estrutura, informação contextual**

alterações documentadas e controladas



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

IV Práticas

procedimentos das CdP em aspectos específicos de preservação de informação digital:

- a. avaliação da informação digital
- a. programas de preservação digital
- b. sistemas de armazenamento
- c. sistemas TIC específicos para a gestão e preservação digital
- e. custódia de ODs



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

IV Práticas

procedimentos das CdP em aspectos específicos de preservação de informação digital:

a. avaliação da informação digital

- considerada pelas CdP uma atividade necessária
- a maioria utiliza sempre critérios de avaliação
- CdP que mais utiliza critérios de avaliação: Arquivos
- CdP que menos os utilizam: Bibliotecas e Museus



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

IV Práticas

a. avaliação da informação - fontes utilizadas

- maioritariamente: referenciais nacionais (sobretudo leis e regulamentos, seguidos de documentos técnicos)
- menos utilizados: documentos técnicos internacionais
- critérios e práticas não formalizadas
- CdP que mais utilizam fontes de autoridade: **Arquivos e Bibliotecas** (sobretudo normas internacionais e nacionais)



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

IV Práticas

procedimentos das CdP em aspetos específicos de preservação de informação digital:

b. programas de preservação digital

- maioria dos que existem não estão formalizados
- ações de preservação esporádicas (a mais frequente é o backup)



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

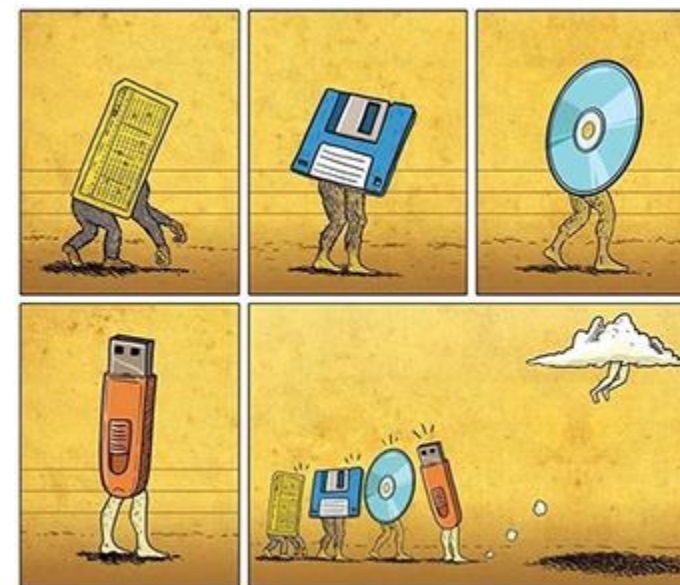
IV Práticas

procedimentos das CdP em aspetos específicos de preservação de informação digital:

c. sistemas de armazenamento = **ponto crítico**

A maior parte das CdP utiliza:

- servidores
- sistemas de ficheiros
- dispositivos externos como DVDs





Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

IV Práticas

procedimentos das CdP em aspectos específicos de preservação de informação digital:

d. sistemas TIC específicos para a gestão e preservação digital

- CdP remetem para a existência de um sistema
- Geralmente não dedicado à preservação e gestão de informação digital

Se forem consideradas as exigências das normas OAIS, a percentagem deve rondar os 0%



Continuidade Digital

2. análise transversal para as CdP: convergência ou divergência?

IV Práticas

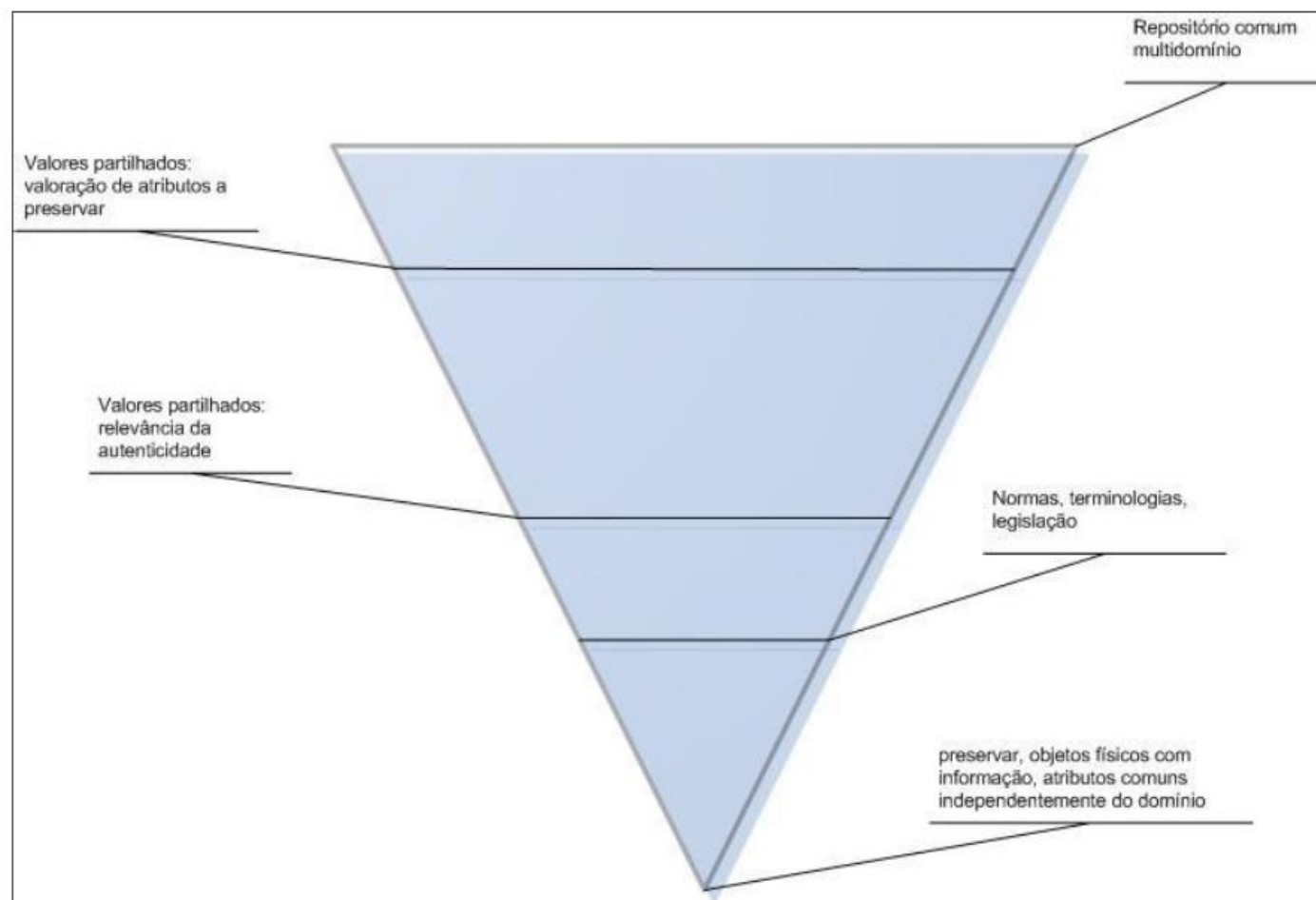
e. Custódia de Ods

- Maioria das CdP reporta custódia de ODs patrimoniais (nado digitais e resultantes de transferência de suporte)
- Percentagem com algum significado refere o contrário (% sensivelmente igual à que refere não realizar avaliação)
- Dimensão do universo custodiado: c. 2,35 PB (arquivos das Televisões)
- Expetativa de crescimento prospetivo: c. de 20 TB por ano de nova informação



Continuidade Digital

2. análise transversal para as CdP: convergência e divergência



prática de avaliação



Continuidade Digital

3. Modelo de custos

Considerados dois cenários:

- PID empreendida **a título individual**
- PID empreendida no **contexto de uma rede comum**
(plataforma tecnológica comum, custos partilhados)

Determinar:

- benefício financeiro da adesão a uma rede comum com custos partilhados
- número mínimo de aderentes necessários para tornar vantajosa a partilha





Continuidade Digital

3. Modelo de custos

Aspetos a considerar:

- estimativas de crescimento anual
- formatos dos ODs
- sistemas e processos de armazenamento utilizados
- requisitos de rapidez de acesso
- n.º de cópias anualmente solicitadas
- tempos de espera para a sua obtenção



Continuidade Digital

3. Modelo de custos

Metodologia utilizada:

- Orçamento para a contratualização externa dos serviços (empresas especializadas):

Armazenamento em <i>cloud</i>	Acesso imediato	4.462,26 € ano/TB
	Acesso diferido	3.391,99 € ano/TB

custo médio de *storage*: **859,42 €** TB/ano

(média de preços obtidos na *cloud* da *Amazon* e na *Azure* da *Microsoft*)

c. **3.600 €** na primeira modalidade e de **2.500 €** na segunda - acréscimo cobrado por serviços específicos de preservação digital



Continuidade Digital

3. Modelo de custos

- Cálculo das despesas implicadas na PID, diferentes tipos de cenário:

- a título individual e de forma partilhada

recurso a ferramentas previamente identificadas no projeto *4C: Collaboration to Clarify the Costs of Curation* (financiado pela EU)

<http://4cproject.eu/summary-of-cost-models>

Seleção dos modelos considerados mais adequados:



Continuidade Digital

3. Modelo de custos

- Cálculo das despesas implicadas na PID a título individual:

Modelo LIFE		
Objetos de formatos diversificados	período de 10 anos	605.000 euros
Imagens	período de 10 anos	806.000 euros

Modelo MODS		
Vídeo e imagens	período de 10 anos	2.796.901 euros
Todos os formatos (bases de dados, vídeos, imagens, texto, etc. - sem ligação com realidades reportadas pelos membros do GT)	período de 10 anos	13.527.578 euros



Continuidade Digital

3. Modelo de custos

Conclusões da aplicação dos modelos:

- orientadas para calcular custos da PID enquanto atividade
- não orientados para calcular custos de um repositório comum

Elaboração de cenário alternativo:

elaboração de uma matriz para determinar uma unidade de custo base que permita:

- calcular para cada entidade os custos de adesão a uma rede comum



Continuidade Digital

3. Modelo de custos - cenário alternativo

2 tipos de custos diretos:

- fixos – assumem maior peso:
 - recursos humanos
 - manutenção e atualização do repositório
 - manutenção de edifício (espaço físico ocupado, eletricidade, segurança, etc.)
 - outros (computadores para utilização administrativa, material de escritório, etc.)
- variáveis – diretamente proporcionais ao grau de utilização dos serviços:
 - espaço de armazenamento (individualizável, potencialmente evolutivo)
 - complexidade do processamento dos serviços de preservação digital

n.º de ODs (expressa numa unidade binária)	x	esforço de computação (diretamente proporcional ao n.º de Ods, cai exponencialmente de ano para ano)
---	---	--



Continuidade Digital

3. Modelo de custos

Cenário alternativo: questões a considerar

Questão 1: encontrar um rácio equilibrado entre estas duas variáveis:

- dimensão dos objetos expressa numa unidade binária (MB, TB, etc.)
- número de objetos que representam essa dimensão.

Questão 2 determinar as variáveis possíveis e adequadas para estimar o custo de computação, obtendo-se assim o fator a relacionar com o número de ODs



Continuidade Digital

3. Modelo de custos

Conclusões

Fórmula final para obter o custo base poderá assumir a seguinte configuração:

$$\text{Unidade Base Custo} = \text{Custos Fixos} \times \text{Custos Variáveis}$$

$$\text{Custos Variáveis} = (\text{dimensão_espaço} \times \text{n.º_ODs}) (\text{esforço_computação})$$

- caso não seja necessário serviço de armazenamento, retirar a parcela
- considerar sempre a parcela “número de ODs x esforço computação”



Continuidade Digital

3. Modelo de custos

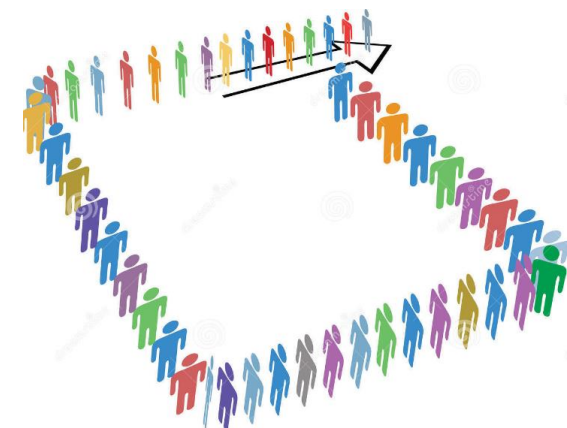
Conclusões

Decrementação de custos em função do aumento de n.º de aderentes:

mais aderentes = menos custos

cada novo aderente representa redução nos custos fixos

custos variáveis dependem do material que cada aderente ingira no repositório



Cenário 1:

		Custos fixos	Custos armazenamento/ processamento
Ano 1	5 aderentes	pagam 20.000 + 50.000	
Ano 2	5 aderentes + 1 novo aderente	pagam 16.557 + 50.000	

Cenário 2: aos custos anteriores acresce joia de inscrição por cada novo aderente



Continuidade Digital

3. Modelo de custos

produto disponibilizado

Simulador de custos de preservação digital (excel)

A utilizar pelas entidades interessadas

Acessível no sítio web do projeto:

<http://1seminariopreservacaopatrimonioidigital.dglab.gov.pt/projeto-continuidade-digital/documentos-de-projeto/>





Continuidade Digital

3. Modelo de custos

Conclusões gerais:

custos a suportar pelos aderentes são eventualmente excessivos

equacionar formas de financiamento da rede que assegurem a sua independência:

- outros serviços pagos: monitorização das tecnologias
- modelo de exploração comercial - cobrança, aos aderentes, de % fixa dos lucros auferidos com venda de cópias de ODs preservados
- publicidade

preservação digital distribuída
(+ investigação, desenvolvimento adicional)



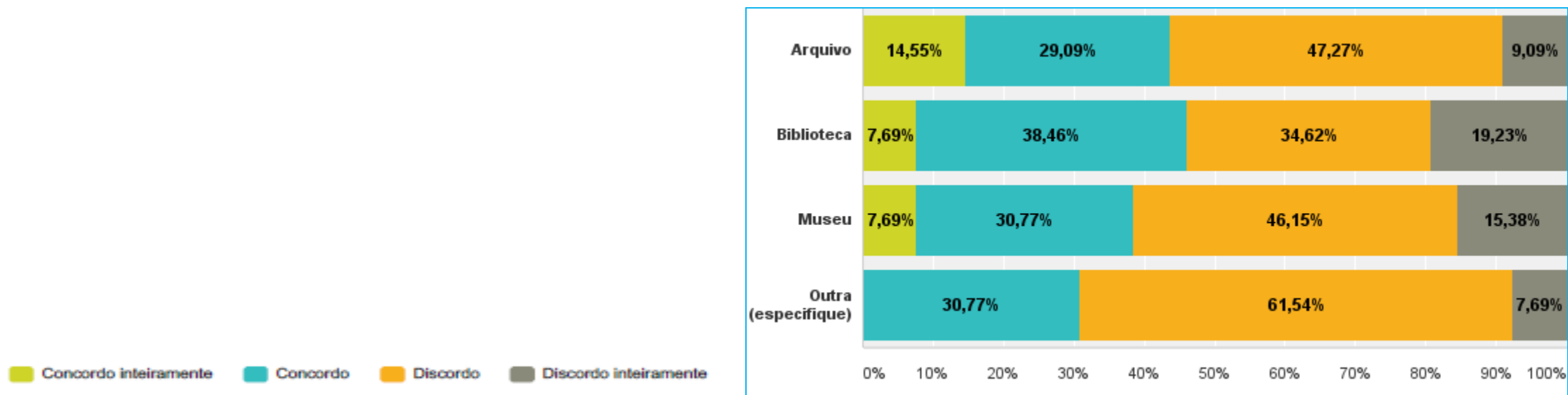


Continuidade Digital

3. Modelo de custos

O que revelam os inquéritos:

relativamente à disponibilidade das instituições para contribuírem financeiramente para um repositório comum, como contrapartida de serviços prestados, a maior parte dos respondentes discorda ou discorda inteiramente:





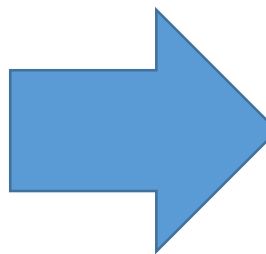
Continuidade Digital

4. Governança

diretamente relacionada com o modelo de custos.

Implica a definição de:

- características
- princípios
- objetivo
- missão
- funções e atividades
- tipo de governação
- órgãos



traduzível de forma efetiva em
Regulamento



Continuidade Digital

4. Governança

Funções a assegurar (comuns a qualquer tipo de organização)

- deliberativa (exercida formalmente por um conjunto de aderentes)
- executiva (núcleo encarregue de executar e gerir a estrutura)
alternativas:
 - várias entidades aderentes;
 - subconjunto de aderentes fixos e subconjunto de aderentes rotativo
 - entidade fixa, definida por lei
 - instituição pública ou privada contratada para o efeito
- tecnológica (manutenção preventiva, corretiva e evolutiva do repositório comum)
- preservação (através de repositório tecnológico comum, valência de receber e preservar ODs)
- investigação (criar conhecimento na área da PID a replicar reutilizar pelos aderentes)





Continuidade Digital

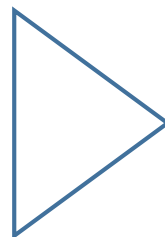
4. Governança

Aderentes

- pessoas coletivas, públicas ou privadas
- ou pessoas singulares

diferentes tipologias

direitos e deveres



em função dos serviços a contratualizar
e do grau de compromisso a assumir



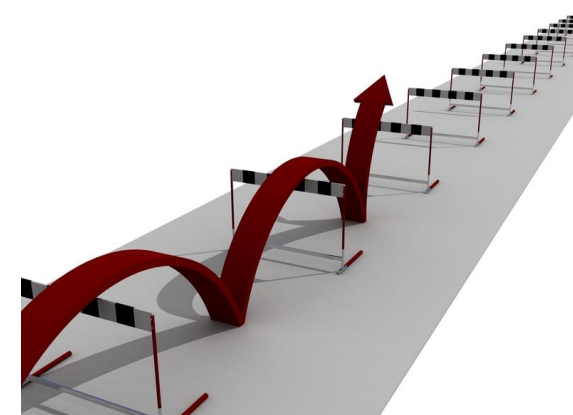


Continuidade Digital

4. Governança

Obstáculos identificados no contexto da AP:

- Natureza jurídica (formalização legal da rede)
- Natureza orçamental
- Constrangimentos de organização e de articulação
(Rede = organização virtual no seio de entidade reais)





Continuidade Digital

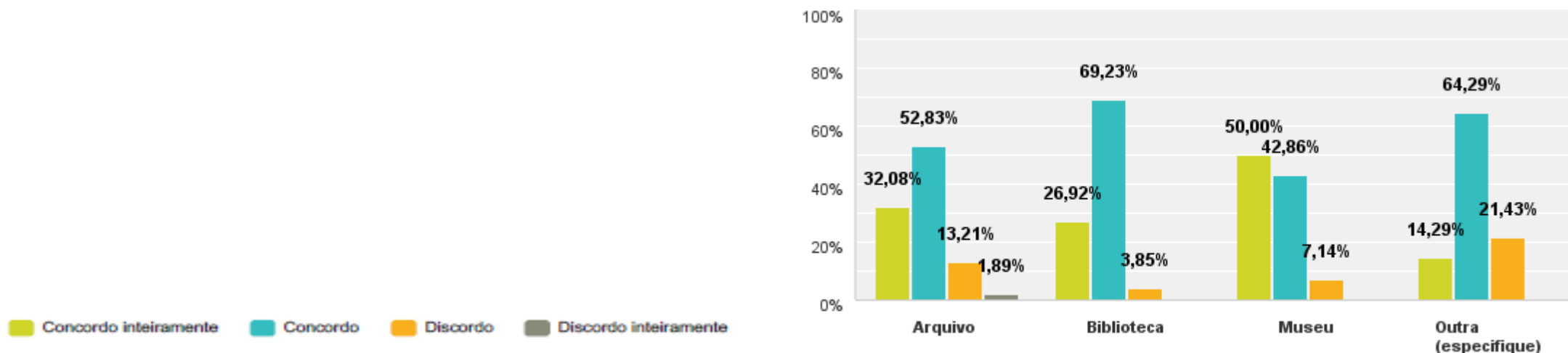
5. Ponto de situação atual

A opinião dos potenciais aderentes a uma rede de preservação comum – **convergências:**

Perceção comum quanto à relevância de preservar património digital

A maioria das CdP

- concorda com um repositório transversal (arquivos, bibliotecas, museus, etc.)





Continuidade Digital



5. Ponto de situação atual

A opinião dos potenciais aderentes a uma rede de preservação comum – **convergências:**

A maioria das CdP

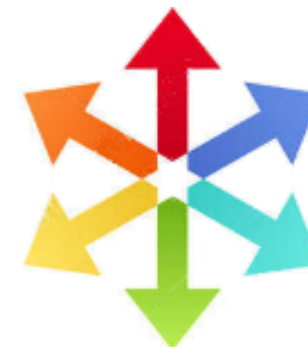
- concorda com uma gestão coletiva de um repositório partilhado
- manifesta disponibilidade para assumir compromisso e responsabilidade num modelo de gestão participativo
- considera financeiramente vantajosa a gestão comum do património digital
- concorda com modelo de exploração comercial para assegurar sustentabilidade financeira (sobretudo Arquivos e Museus)
- optaria por entidade pública para prestação de serviços de preservação digital (sobretudo Arquivos e Bibliotecas)

Nota: algumas CdP, sobretudo, **Museu** e **Outra** admitem confiar ODs a entidades privadas, desde que confiáveis



Continuidade Digital

5. Ponto de situação



A opinião dos potenciais aderentes a uma rede de preservação comum - **divergências**

A maioria das CdP:

- não prevê disponibilidade financeira para contribuir para o financiamento do repositório

Detetada alguma resistência em confiar ODs a uma entidade terceira (sobretudo Bibliotecas e Arquivos)



Continuidade Digital

5. Ponto de situação

O que já existe

DGLAB:



Equipa do GT Continuidade Digital

Visão global e integradora da preservação digital e do acesso ao património

Programa Ciência Aberta - GT Política Nacional de Ciência Aberta

Projetos em curso: ex. Centro Hospitalar São João



Continuidade Digital

5. Ponto de situação

O que há a fazer

- adotar boas práticas:
 - OAIS
 - normalização de formatos
 - normalização de meta informação visando interoperabilidade
- qualidade da plataforma tecnológica e das infraestruturas
- definir custos da preservação digital e incluí-los na programação orçamental
- identificar e delimitar responsabilidades sobre o património digital





Continuidade Digital

6. Perspetivas futuras

Necessidade de investigação adicional:

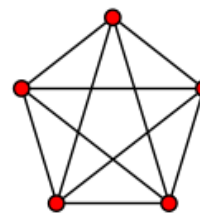
Área tecnológica:

- Modelo de preservação digital distribuída

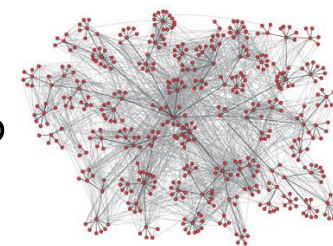
Área social/organizacional:

- Formas de articulação de entidades com culturas organizacionais e orçamentais díspares

- Definição do modelo de rede: mais simples?



mais complexa?



- Análise financeira e económica - identificar possibilidades de sustentação financeira da rede



Continuidade Digital

6. Perspetivas futuras

Objetivos

- Obter o financiamento necessário para prossecução do trabalho (SAMA 2020)
- Disseminar de forma mais participativa os resultados do projeto
- Promover conhecimento na área de preservação digital através de ações formativas e informativas
- Obter uma base alargada para discussão e validação dos resultados do projeto
- Cooptar massa crítica para a constituição efetiva da rede para preservação comum de património digital





Continuidade Digital

mais informação em:

<http://1seminariopreservacaopatrimonioidigital.dglab.gov.pt/projeto-continuidade-digital/>

contactos, dúvidas, esclarecimento, suporte:

DGLAB, inovação e administração eletrónica

dsiae@dglab.gov.pt



Continuidade Digital

dúvidas, questões, comentários, sugestões...