

Encontro Nacional sobre Preservação Digital
Torre do Tombo
2022-11-03

Metodologia para a elaboração de Planos de Preservação Digital

Contexto

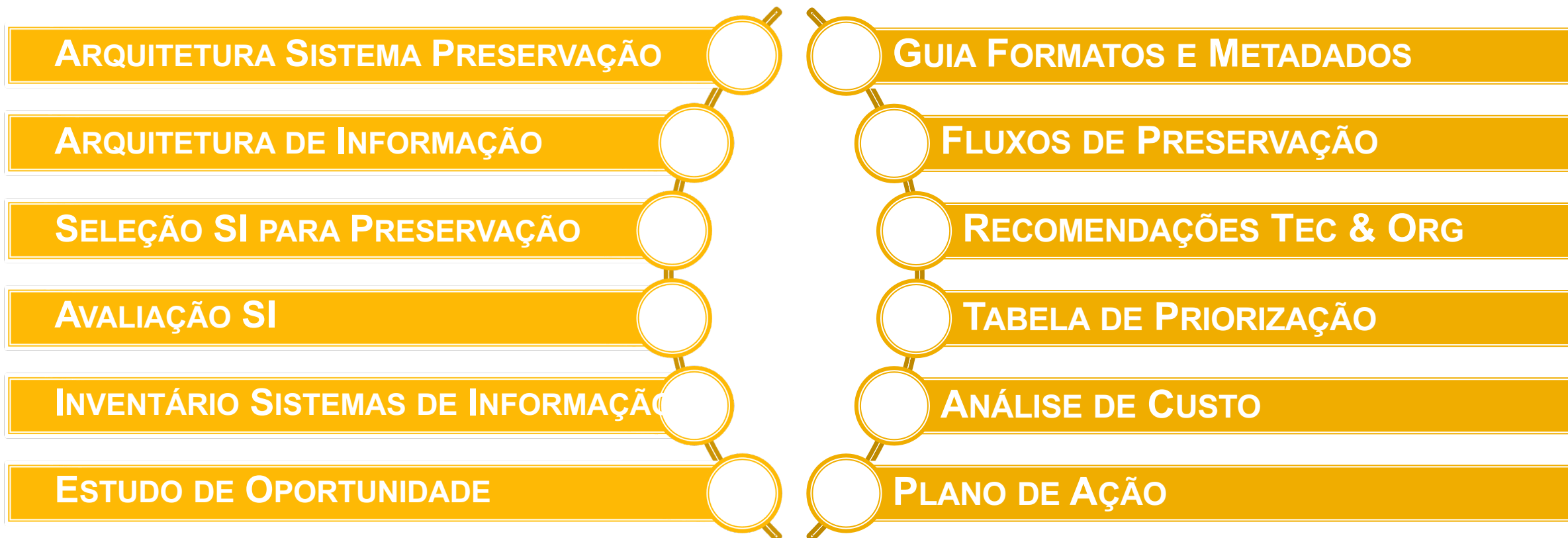
Informação digital

- Riscos da informação em suporte digital
 - Fragilidade de suportes (hardware)
 - Obsolescência de formatos dos objetos digitais
 - Obsolescência de tecnologia
 - Ciclos de vida da informação muito curtos (por exemplo, sítios Internet)
- Estratégias de preservação
 - Preservação da tecnologia
 - Emulação da tecnologia
 - Migração/conversão de formatos
- Solução: ações deliberadas de preservação
 - Política
 - Repositório
 - Equipa

Projeto de Preservação Digital

- Elaboração de um Plano de Preservação Digital (PPD)
 - Avaliação dos sistemas de informação (SI) existentes
 - Desenvolvimento do PPD
- Implementação da Preservação Digital

Passos para um PPD



Andamento do projeto

Reunião 0

Apresentação dos intervenientes.
Apresentação da organização.
Estudo prévio.

Reunião 2

Cenários de preservação.
Aprofundamento da análise dos sistemas a preservar.

Reunião 4

Workflows de preservação - revisão individual de cada sistema.

Reunião 6

Quadro de Priorização.
Recomendações técnicas e organizacionais.
Análise de custos.

mês 1

mês 2

mês 3

mês 4

mês 5

mês 6

Reunião 1

Análise orgânica e funcional.
Inventário dos sistemas de informação.
Análise e avaliação de cada sistema.

Reunião 3

Análise exhaustiva dos sistemas ao nível de conteúdos a preservar, responsabilidades, periodicidade.

Reunião 5

Guia de formatos e esquemas de metainformação.

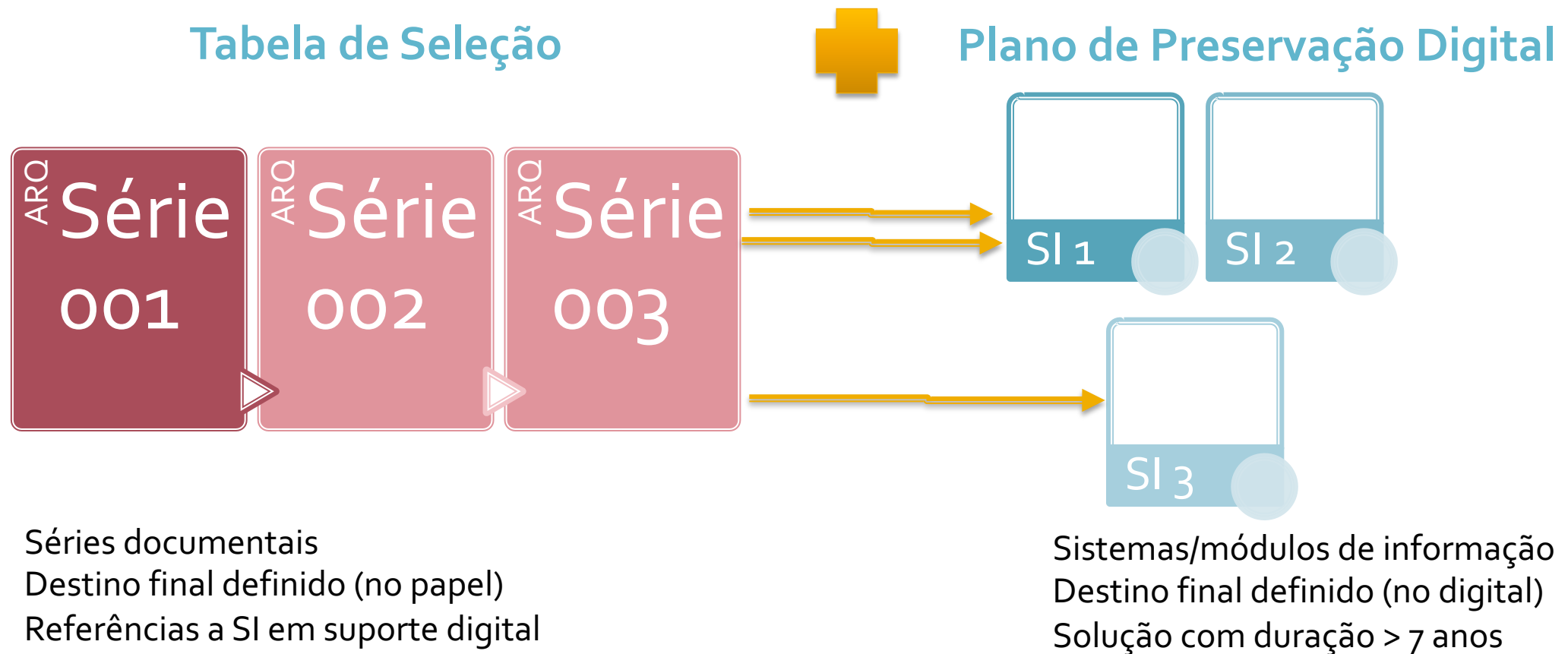
Reunião 7

Apresentação do PPD.

Estudo prévio

- Maturidade digital da organização
 - Grau de informatização
 - Competência técnica
- Viabilidade de constituição de equipa de preservação
- Existência de recursos financeiros
- Compromisso da gestão de topo

Documentos orientadores



Principais referências

- Direção-Geral do Livro, dos Arquivos e das Bibliotecas (DGLAB): Recomendações para a produção de Planos de Preservação Digital
- ISO 14721:2012 Modelo OAIS – Open Archive Information System
- ISO 16363:2012 Audit and certification of trustworthy digital repositories
- InterPARES (International Research on Permanent Authentic Records in Electronic Systems)
- Library of Congress Recommended Formats Statement
- Digital Preservation Coalition (DPC)
- Digital Curation Centre (DCC)

Delimitação do âmbito do PPD

Inventário dos Sistemas de Informação

Ref.	Designação	Unidade orgânica	Dimensão (MB)	Estado
Slo1	Sistema de gestão de projetos	Divisão de Produção	2300	Ativo
Slo2	Sistema de aquisições	Serviço de Contabilidade	600	Ativo
Slo3	Sistema financeiro	Serviço de Contabilidade	500	Ativo
Slo4	Sistema de recursos humanos	Serviço de Recursos Humanos	1000	Ativo
Slo5	Sistema de gestão documental	Serviço de Informática	3000	Semiativo
Slo6	Website antigo	Serviço de Comunicação	250	Inativo
...				

Análise da cada SI

Clique nos balões abaixo para passar à secção pretendida:

Secção 1: Identificação

Secção 5: Segurança

Secção 2: Função

Secção 6: Auditoria

**Secção 3: Descrição e
Avaliação Arquivística**

**Secção 7:
Recomendações**

**Secção 4: Avaliação
Técnica**

Secção 8: Conclusões

Grelha de análise - início

Plano de Preservação Digital – Grelha de Análise

1. Identificação

a) Nome do sistema	
b) Nome do módulo	
c) Tipo de sistema	
d) Âmbito do sistema	
e) Análise pelo Encarregado de Proteção de Dados	
f) Referência no PPD	
g) Nomes alternativos	
h) Referências alternativas	
i) Gestor do sistema	
j) Proprietário do sistema	
k) Gestor dos dados	
l) Proprietário dos dados	

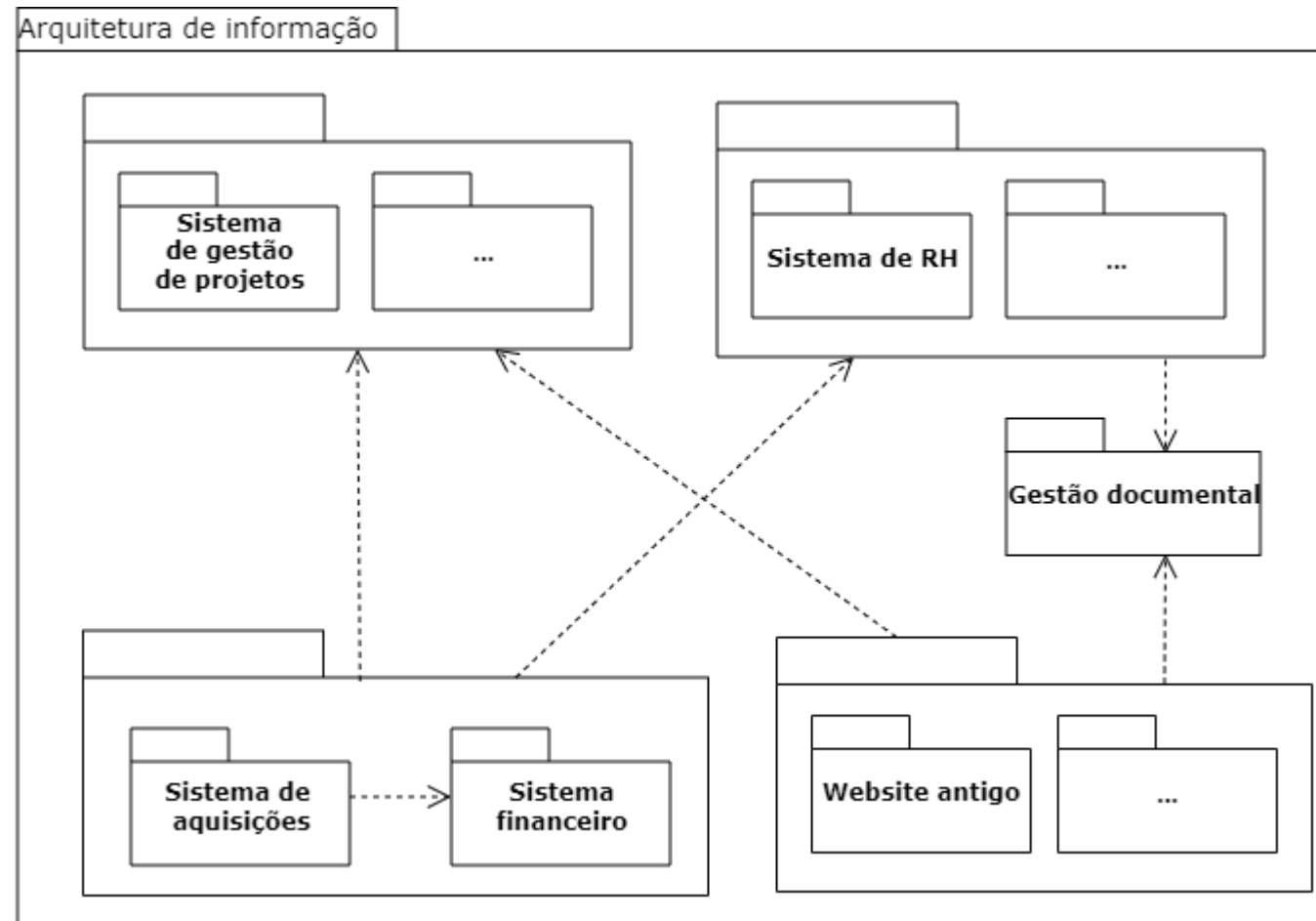
Grelha de análise - conclusão

Plano de Preservação Digital – Grelha de Análise

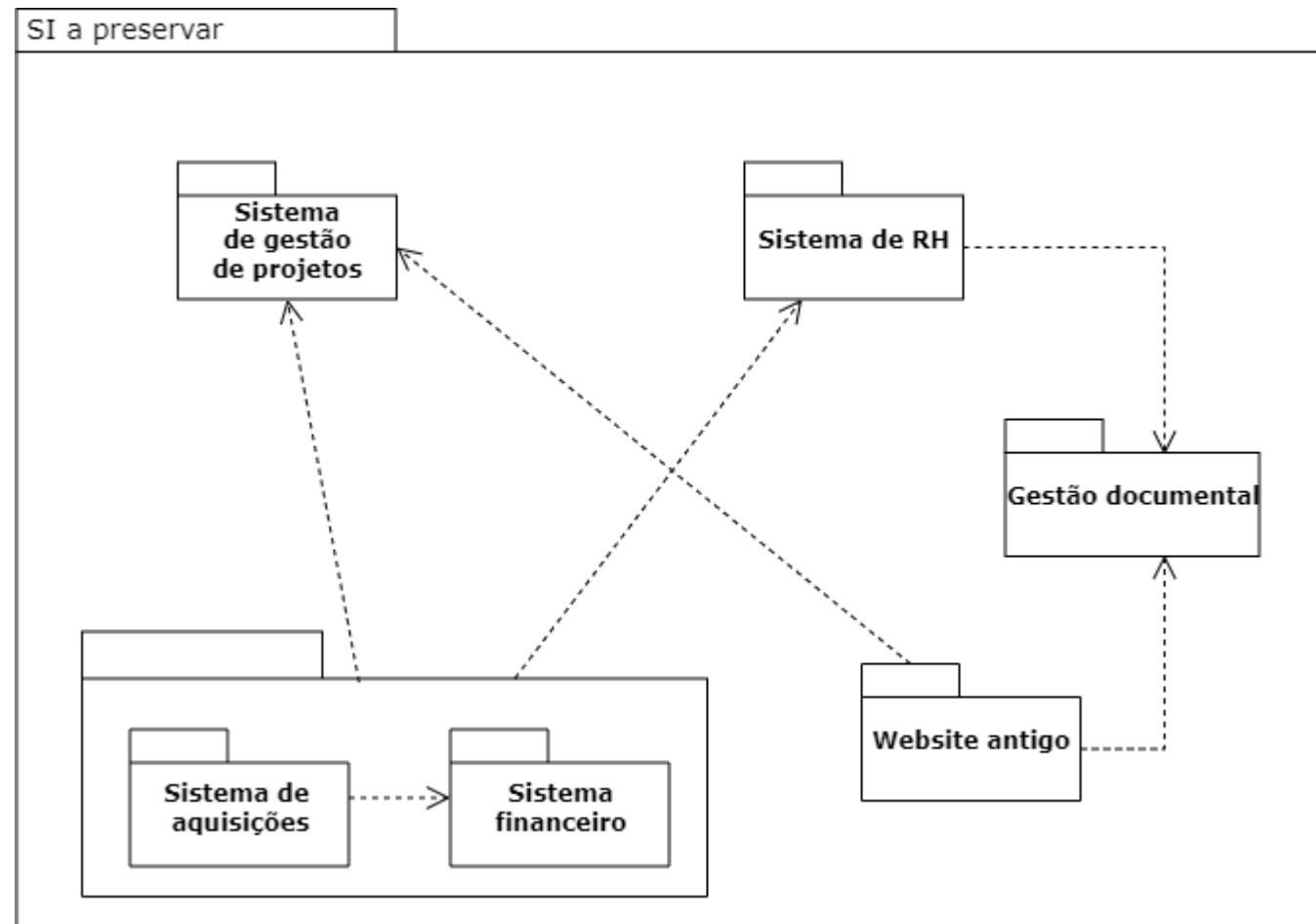
8. Conclusões

a) Decisão	
b) Condição	
c) Período de efetividade	
d) Data	
e) Entidade responsável	

Arquitetura de informação

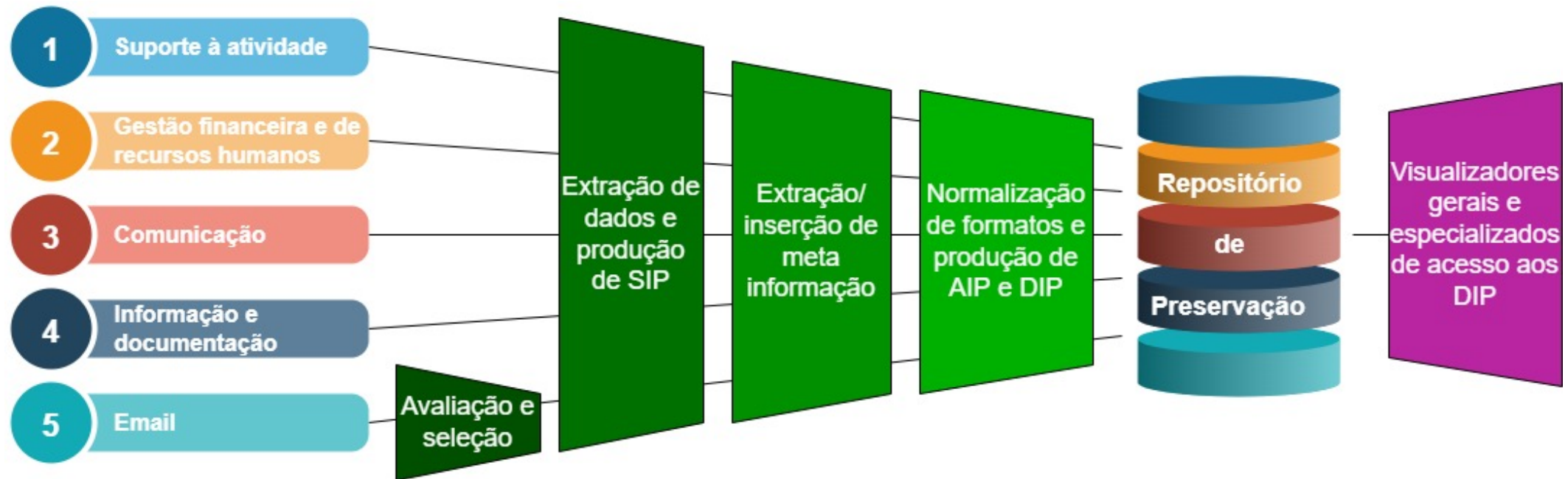


SI a preservar



Estruturação do projeto

Fluxo geral de preservação



Comparação de repositórios

	RODA 4	Archivematica 1.13.2	Preservica Cloud Edition
Open source / comercial	Open source	Open source	Comercial
OAIS	Sim	Sim	Sim
PREMIS	Sim	Sim	Não (mas os metadados podem ser traduzidos para este esquema)
METS	Sim	Sim	Não (mas os metadados podem ser traduzidos para este esquema)
EAD	Sim	Não (mas o AtoM sim)	Sim
DC	Sim	Sim	Sim
Formatos suportados	Múltiplos formatos	Múltiplos formatos (FPR)	Múltiplos formatos (PRONOM)
Crawling de websites	Não (mas suporta formato WARC)	Não (mas suporta formato WARC)	Sim (WARC)

Aspetos organizacionais

- Conselho da Preservação Digital
- Equipa de Preservação
- Políticas
 - Assinaturas digitais
 - Regras para o uso de correio eletrónico
 - Regras para o uso de pastas partilhadas
 - Articulação com o Sistema de Gestão Documental

Formatos preferenciais

Tipo de dados	Formato preferencial	Formato aceitável
Texto	PDF/A (ISO 19005-2:2010)	PDF, DOCX, ODT
Folha de cálculo	PDF/A	PDF, XLSX, ODS
Apresentação	PDF/A	PDF, PPSX
Imagem	TIFF (*.tif)	Photoshop (*.psd, *.psb)
		JPEG2000 Part 2 (*.jpf, *.jpx)
	JPEG2000 (sem perdas)(*.jp2)	Proprietary Camera Raw formats (*.nef, *.crw)
		BMP (*.bmp)
	PNG (*.png)	GIF (*.gif)
		Digital Negative DNG (*.dng)
JPEG/JFIF (*.jpg)		

Formato preferencial	Formato aceitável	Tipo de dados
MXF	MPEG-4 (*.mp4)	Vídeo
Motion JPEG 2000 (ISO/IEC 15444-4:2004/Cor 1:2009)	MKV	
WAV ou WAVE (*.wav) – com metadados embutidos	MP3	Áudio
BWF (*.bwf)		
AIFF (*.aif, *.aiff)	AAC	
DXF (*.dxf)	DWG (*.dwg)	Vetores
SIARD 2 (*.siard)	SQL Data Definition Language	Bases de dados
Comma Separated Values (*.csv)	Microsoft Access (*.accdb; *.mdb)	
Tab Delimited Text (*.txt)	OpenDocument Database File (*.odb)	
XML com esquema (ASCII ou Unicode)	DBF (*.dbf)	

Perfis de metadados

- Dublin Core – Coleções planas
- EAD Encoded Archival Description – Coleções com níveis
- METS – Metadata Encoding & Transmission Standard
- PREMIS - Dicionário de Dados para Preservação de Metadados

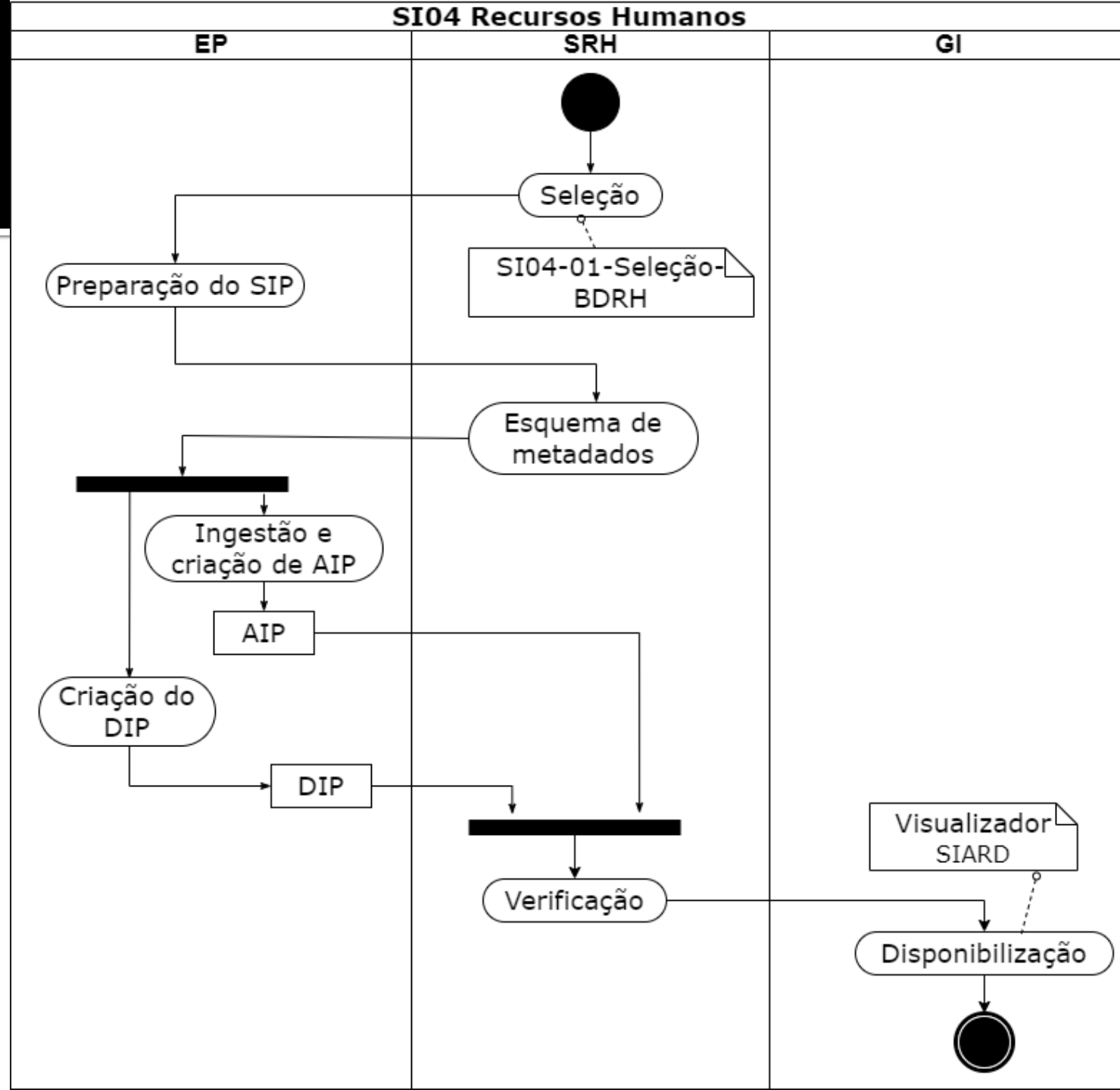
Preparação detalhada

Procedimento de preservação

Ficha de Análise			
Procedimentos de Preservação			
1. Designação do sistema	Sistema de Recursos Humanos		
2. Identificação do sistema	SI04		
3. Área responsável	Serviço de Recursos Humanos		
4. Volume de crescimento/ano	50 MB		
5. Periodicidade	Inicial e a seguir anual		
6. Atores envolvidos	SRH		
7. Workflow de Preservação	Intervenientes	Sistemas ou aplicações	Resultado da atividade
<i>Inventário</i>	N/A		
<i>Seleção</i>	Inventário e seleção simultâneos	SI04-01-SeleçãoRH	Tabelas a preservar
<i>Preparação do SIP</i>	EP	SIARD Suite	SIP
<i>Esquema de Metadados</i>	SRH	DC, METS e PREMIS	Metadados inseridos
<i>Ingestão e criação do AIP</i>	EP	Software de preservação	AIP
<i>Preparação do DIP</i>	EP	Software de preservação	DIP
<i>Verificação</i>	SRH	Amostragem	Declaração de verificação
<i>Disponibilização</i>	GI	VisualizadorSIARD	Consulta de acordo com os requisitos do RGPD

Procedimento de preservação

Nota: visualizador independente do sistema de produção, de forma a garantir continuidade mesmo após desaparecimento do sistema de produção



Declaração de verificação

Identificação do sistema verificado: S1xx

Designação do sistema verificado: XXXXXXXXXXXXXXXXXXXX

Etapas do workflow de preservação verificadas:

Etapa	Verificada	Com erros
Inventário		
Seleção		
Preparação do SIP		
<u>Esquema de Metadados</u>		
Ingestão e criação do AIP		
Preparação do DIP		

Método de verificação aplicado: *(exaustivo, amostragem, outros)*

Resultado da verificação:

Aceite

Rejeitado

Responsável pela verificação:

Data da verificação:

Recomendações técnicas sistemas operacionais

1. Codificação de caracteres: UTF-8
2. Uniformizar formatos de ficheiros
3. Recolher metadados na criação dos documentos
4. Timestamp nas alterações facilita preservação incremental
5. Sinalizar objetos digitais já ingeridos no repositório
6. Melhorar a exportação de dados de aplicações fechadas

Recomendações técnicas sistemas distribuídos

1. Regras para a estruturação das pastas e subpastas
 - atribuição de nomes a pastas e ficheiros
 - identificação e atribuição de responsabilidade pela gestão destas pastas
2. Critérios de seleção dos conteúdos a incorporar
 - prazos de retenção dos ficheiros nestas pastas
3. Inventário dos conteúdos das pastas e seleção
4. Regras e responsabilidade na gestão do email
5. Inventário das mensagens nos emails institucionais

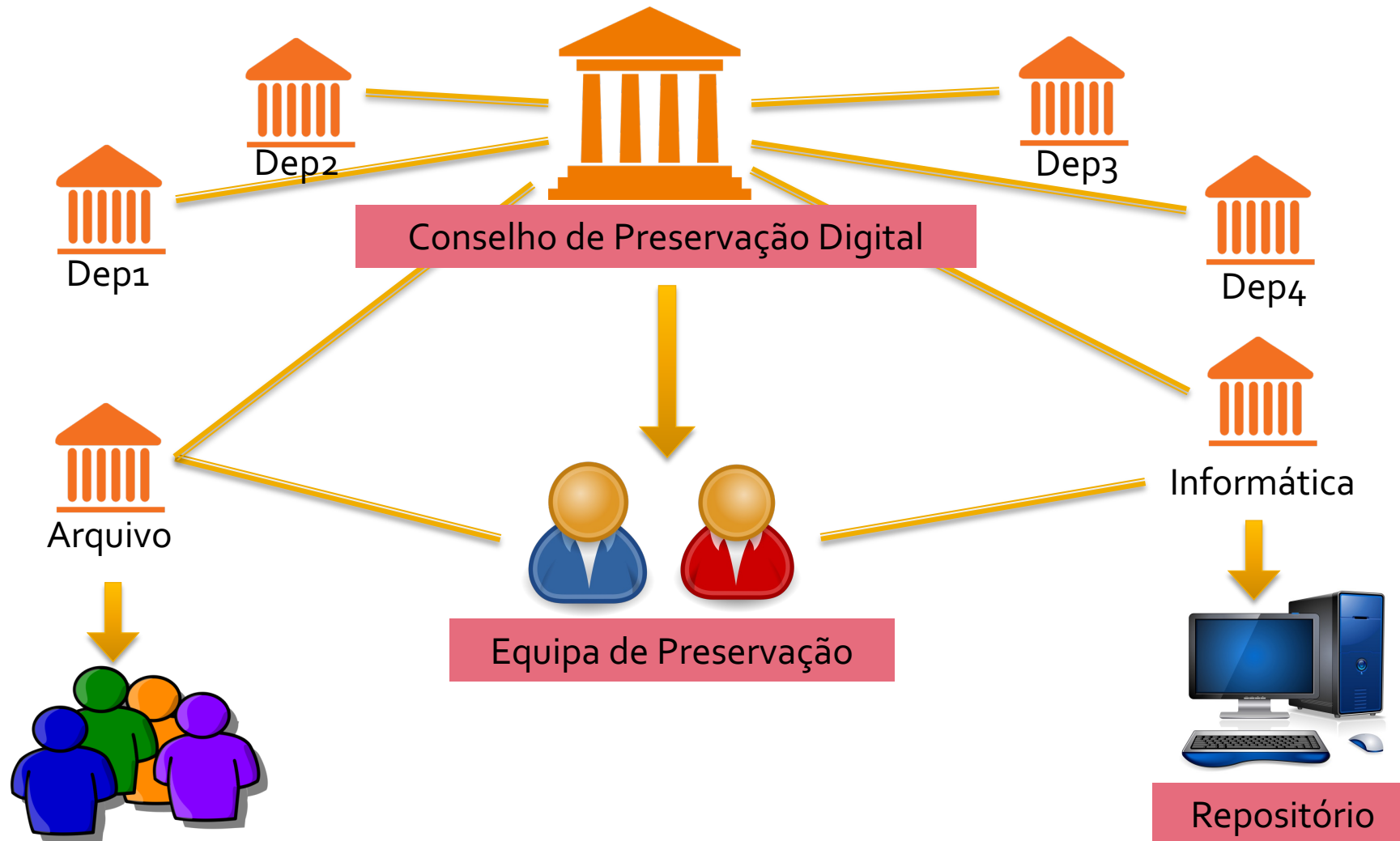
Recomendações sobre o repositório

1. Repositório de Preservação deve possuir as seguintes funcionalidades
 1. Elaboração de relatórios adequados, exportação de dados para folhas de cálculo
 2. Verificar a integridade dos dados incorporados
 3. Registrar (*logs*) das verificações de integridade realizadas
 4. Registrar todos os acessos ao sistema
 5. Registrar todas as ações realizadas sobre ficheiros e por quem, incluindo eliminações
 6. Suportar a diversidade de tipos de conteúdos existentes atualmente
 7. Gerar identificadores únicos e persistentes para todos os pacotes AIP
 8. Suportar controlo de acesso diversificado.
2. Substituir assinaturas digitais por confiança no Repositório
3. Desenvolver módulos de preparação de pacotes de submissão
 - Em particular para a fase incremental
4. Idem para a fase de visualização
 - Em particular para sistemas complexos e de interesse alargado

Recomendações organizacionais

1. Equipa de Preservação
 - Técnico de Arquivo (metadados)
 - Técnico da Informática (operações de preservação, migrações, normalização)
2. Preservação digital integrada desde o início
 - Processos de aquisição ou desenvolvimento de software
 - Definição de cadernos de requisitos e avaliação das aplicações.
3. Pedidos de acesso à informação
 - Responsabilidade do Arquivo
 - Apoio da Equipa de Preservação, como segunda linha

Gestão e operacionalização



Recomendações finais

- Regulamento Geral de Proteção de Dados (RGPD)
 - Assegurar conformidade dos dados ingeridos no repositório de preservação
- Tabela de Seleção de Documentos
 - Promover a revisão de acordo com as conclusões do PPD
 - Ambos os documentos articulados
 - Cobrir todos os conteúdos informacionais, independentemente do suporte

Plano de ação

Priorização dos SI

Ref.	Designação	Unidade orgânica	Dimensão (MB)	Estado	Relevância	Prioridade
Slo1	Sistema de gestão de projetos	Divisão de Produção	2300	Ativo	Alta	2
Slo2	Sistema de aquisições	Serviço de Contabilidade	600	Ativo	Média	3
Slo3	Sistema financeiro	Serviço de Contabilidade	500	Ativo	Média	3
Slo4	Sistema de recursos humanos	Serviço de Recursos Humanos	1000	Ativo	Alta	1
Slo5	Sistema de gestão documental	Serviço de Informática	3000	Semiativo	Média	2
Slo6	Website antigo	Serviço de Comunicação	250	Inativo	Baixa	1
...						

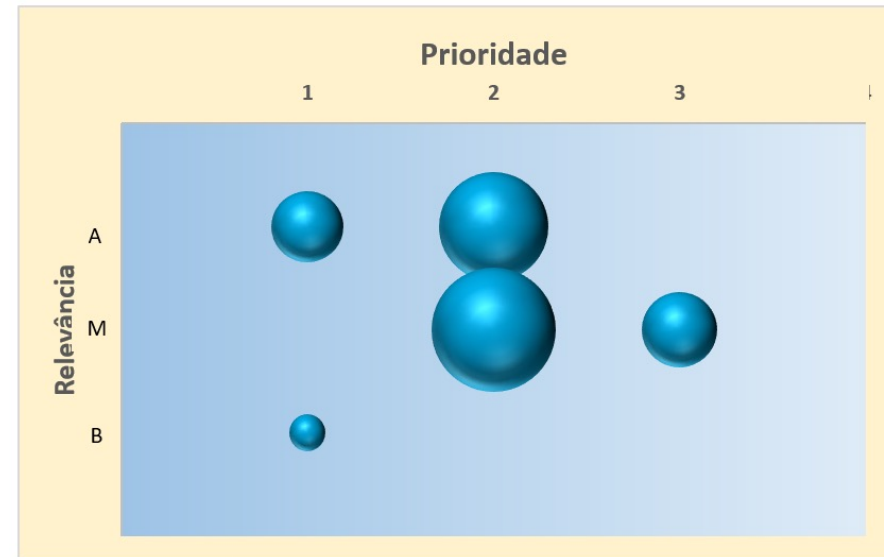
Relevância versus prioridade

Quantidade de SI

Sistemas	Prioridade			
	1ª	2ª	3ª	
Alta	1	1	0	2
Média	0	1	2	3
Baixa	1	0	0	1
	2	2	2	6

Relevância

Volume de dados



Armazenamento estimado (GB)

Prioridade	Dados	Documentos	Total
1	1.000	0.250	1.250
2	2.300	3.000	5.300
3	0.500	0.600	1.100
Total	3.800	3.850	7.650

Crescimento: 2 GB/ano
Total a 3 anos: 13.65 GB

Plano de ação – três fases

PROJETO Plano de Preservação Digital
 Documento Plano de ação v1.0 2022-11-03

FASE 1 - PREPARAÇÃO

FASE 2 - PRESERVAÇÃO INICIAL FASE 3 - PRESERVAÇÃO REGULAR

Fase	Designação	Detalhes	Periodicidade	Mês												1º T				2º T				3º T				4º T			
				1	2	3	4	5	6	7	8	9	10	11	12	Jan	Fev	Mar	Abri	Ma	Jur	Jul	Agc	Set	Out	Nov	Dez				
1	Estrutura organizacional e técnica	Nomear Comissão de Preservação Digital Constituir Equipa de Preservação (EP) Implementar o Repositório de Preservação (RP) Desenvolver módulos especializados Formação da EP no uso do RP		[Gantt chart showing tasks from month 1 to 4]																											
2.1	Preservação - Sistemas prioridade 1	SI06 - Website antigo SI04 - Sistema de Recursos Humanos	Única Anual	[Gantt chart showing tasks from month 5 to 6]												Sem preservação incremental															
2.2	Preservação - Sistemas prioridade 2	SI01 - Sistema de gestão de projetos SI05 - Sistema de gestão documental	Semestral Mensal	[Gantt chart showing tasks from month 7 to 12]												[Gantt chart showing tasks from 1st to 4th trimester]															
2.3	Preservação - Sistemas prioridade 3	SI02 - Sistema de aquisições SI03 - Sistema financeiro	Mensal Anual	[Gantt chart showing tasks from month 7 to 12]												[Gantt chart showing tasks from 1st to 4th trimester]															

3 2 3 3 2 2 3 2 2 2 2 2

Custos estimados

Custos de implementação do PPD. Solução 1

Rubrica	Item	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Total
Hardware	Aquisição						- €
	Manutenção						- €
	Alojamento						- €
Software	Licenças						- €
	Configuração e parametrização						- €
	Desenvolvimento						- €
Implementação e suporte	Formação						- €
	Manutenção e suporte SW						- €
Recursos humanos	Equipa preservação (2 elementos)						- €
	Outros						- €
TOTAL	Custo / ano	- €	- €	- €	- €	- €	- €

Conclusão

- PPD é um documento vivo
 - Deve prever a sua própria revisão periódica
- Os instrumentos de apoio estão disponíveis
 - Grelha de avaliação dos SI
 - Template dos procedimentos de preservação de cada SI