

CERTIFICAÇÃO DE REPOSITÓRIOS DIGITAIS

Miguel Ferreira
Diretor técnico e da Qualidade
mferreira@keep.pt

Seminário (r)evolução da Informação
Pública: preservar, certificar e acessibilizar

7 de novembro de 2011

O que é um repositório digital?

Porquê certificar?

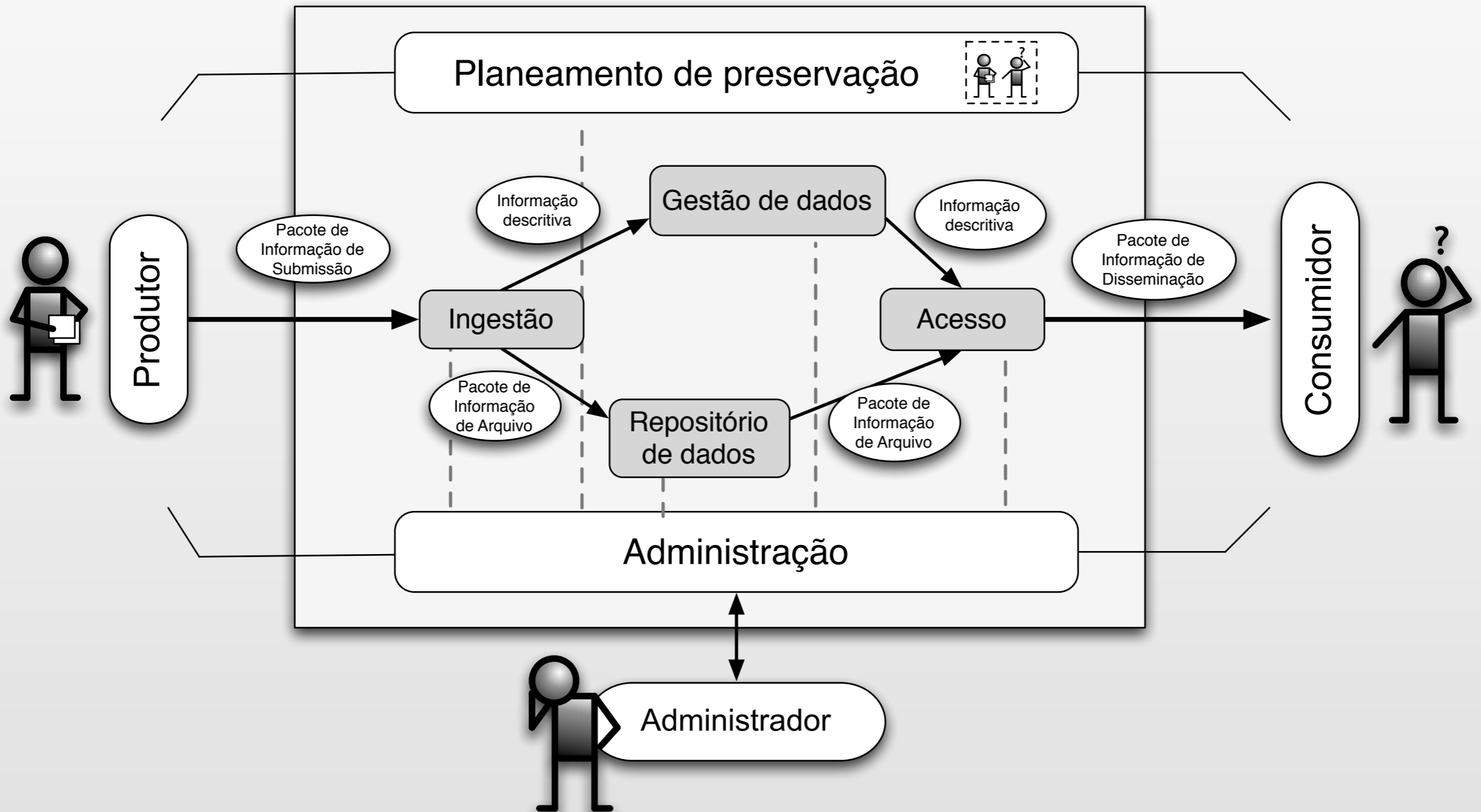
Ferramentas para a certificação de repositórios digitais

Conclusões

O que é um repositório digital?

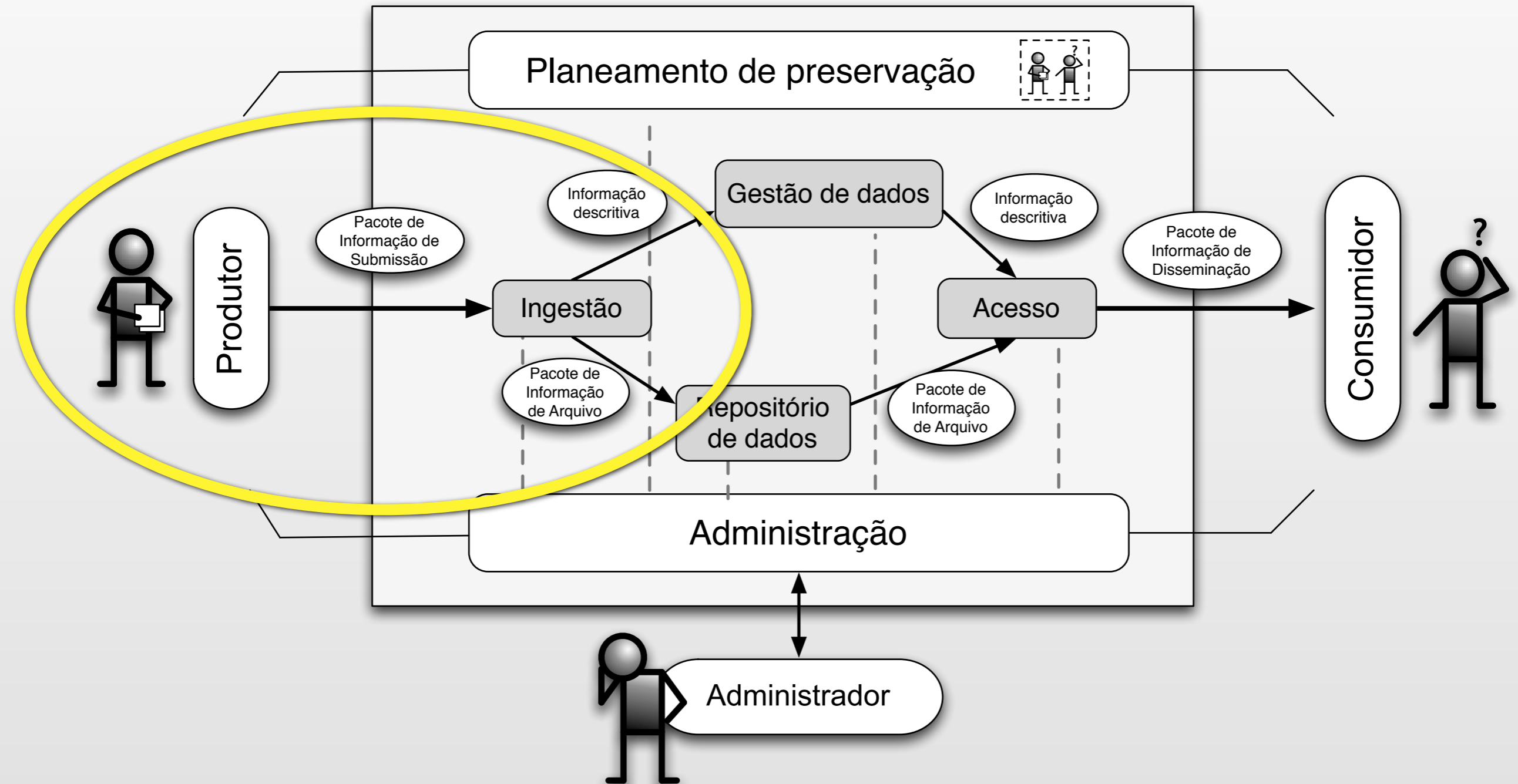
Open Archival Information System - OAIS

ISO 14721:2003



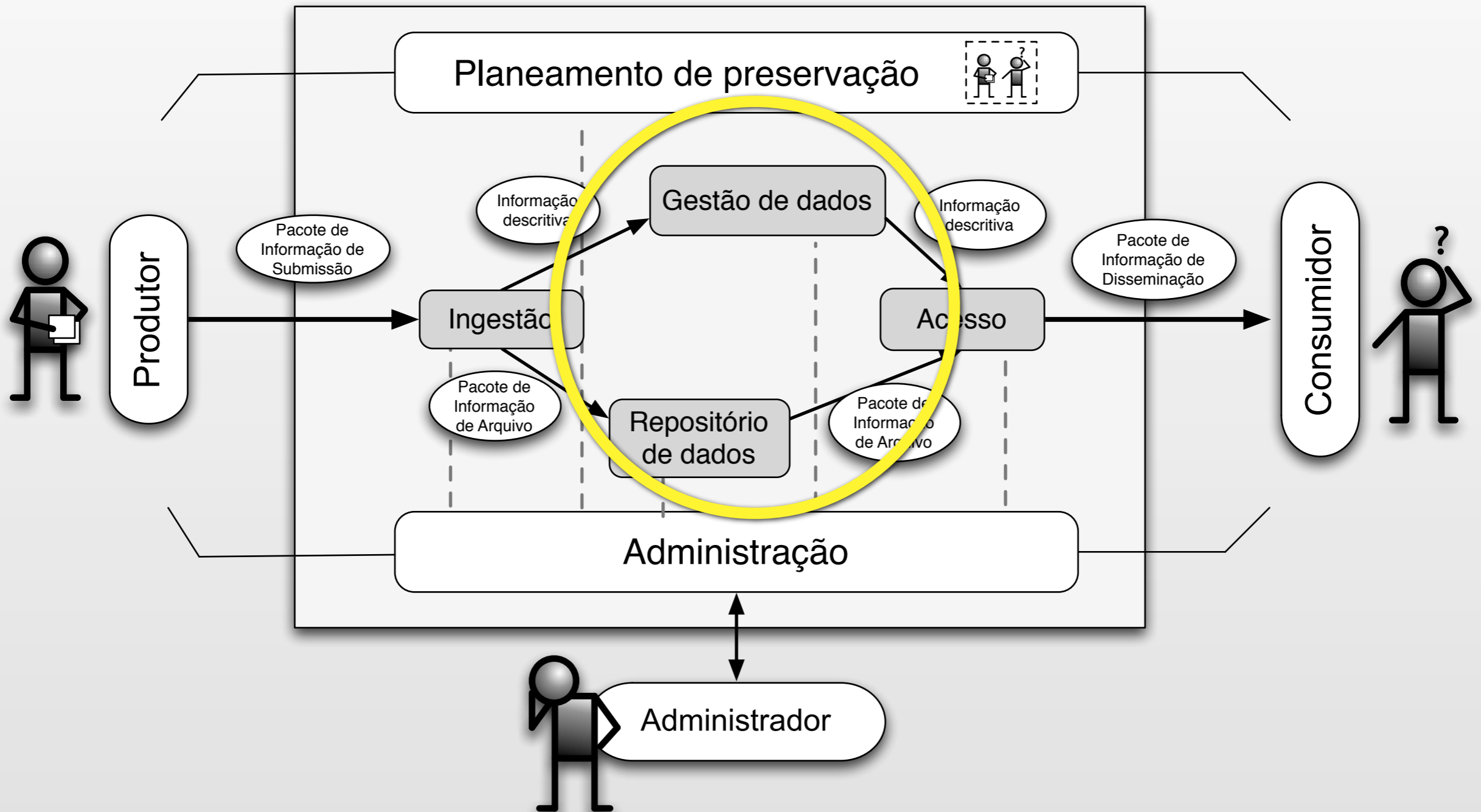
Open Archival Information System - OAIS

ISO 14721:2003



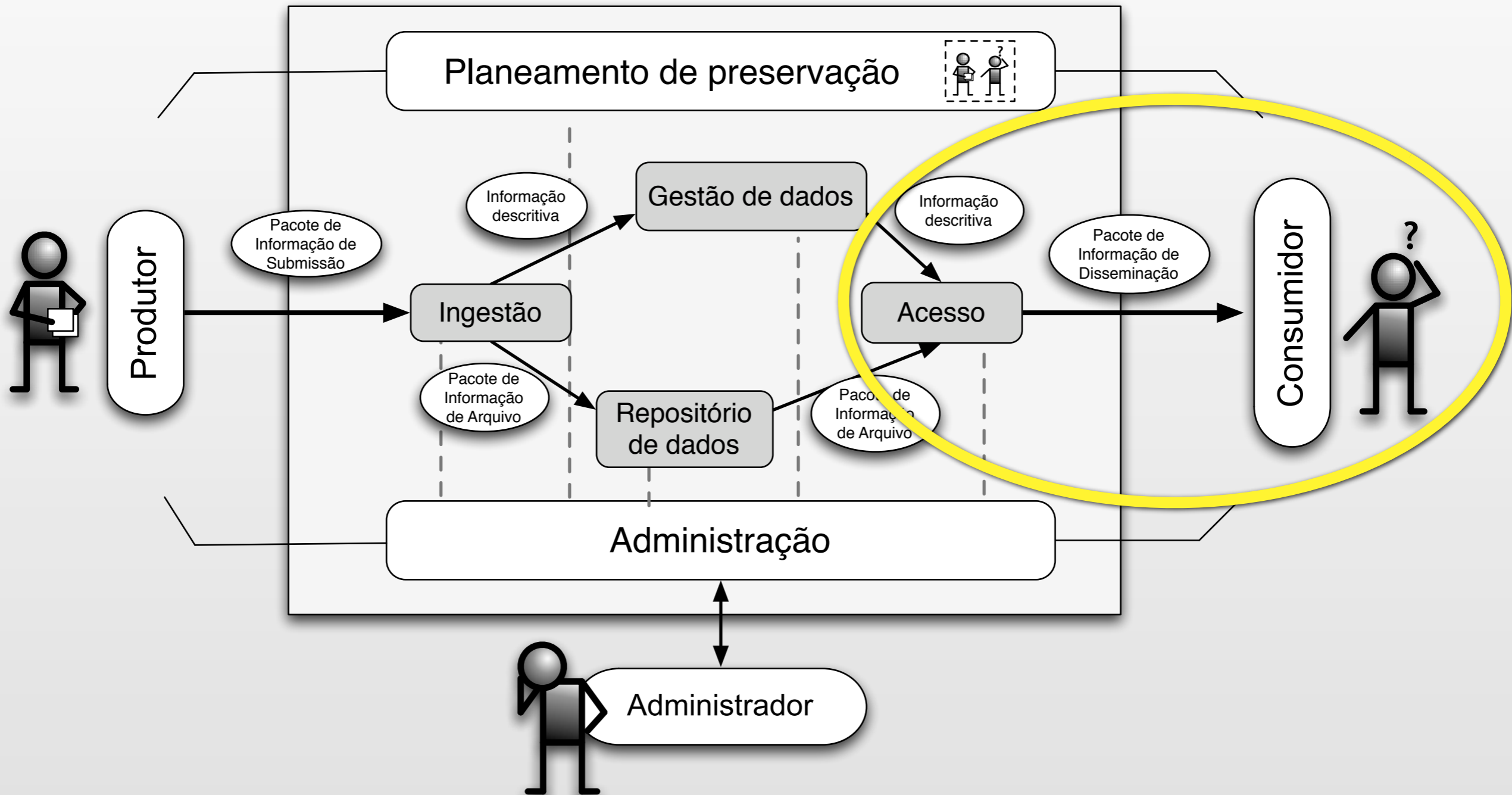
Open Archival Information System - OAIS

ISO 14721:2003



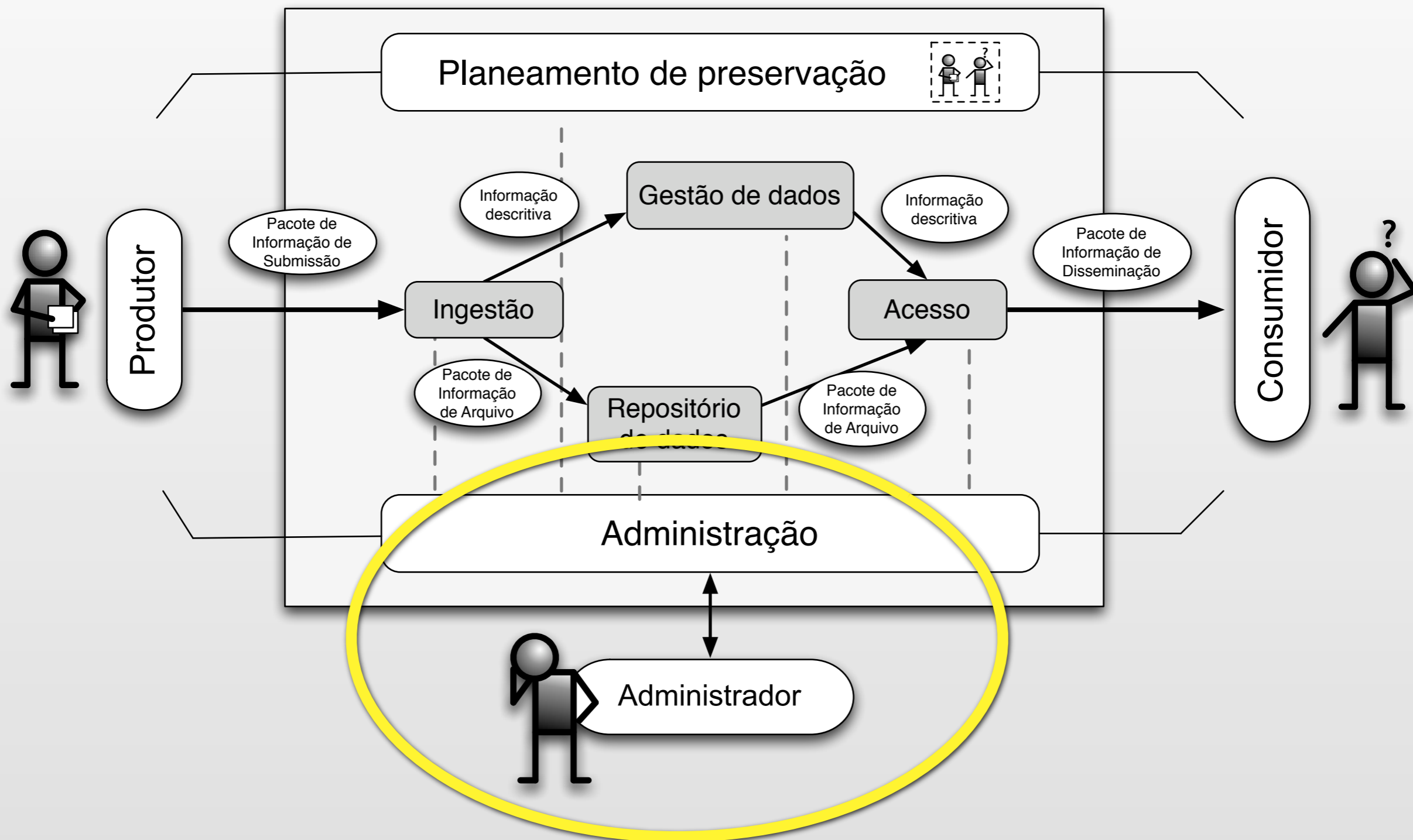
Open Archival Information System - OAIS

ISO 14721:2003



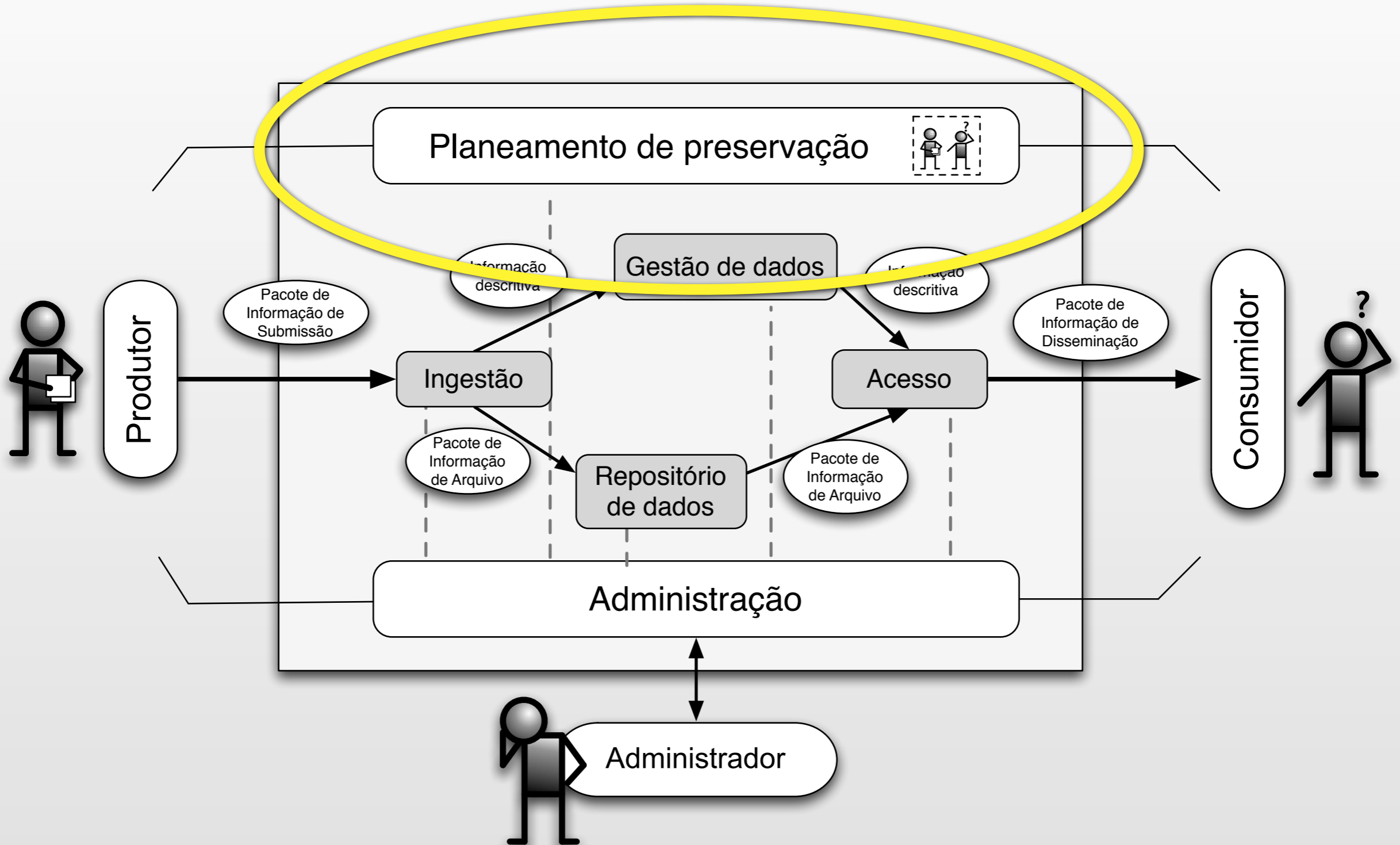
Open Archival Information System - OAIS

ISO 14721:2003



Open Archival Information System - OAIS

ISO 14721:2003



Porquê certificar um repositório digital?

Estabelecer um **clima de confiança** em torno do repositório e da informação que nele se encontra custodiada

- ▶ **Produtores, Consumidores**, Gestores, Operadores, Financiadores, etc.

Capacidade de **demonstrar de forma objetiva** quanto confiável é o nosso repositório

Postura de **transparência** perante todos os intervenientes

Demonstrando que todos os processos e procedimentos são seguidos

- ▶ Demonstrar **com base em evidências**

Quantos de vocês **confiam** nos sistemas de informação da
Segurança Social?

Serão estes sistemas capazes de **preservar** os vossos registos durante 20, 30 ou 40 anos?

Já imaginaram quantos **governos, políticas e sistemas de governação** estes registos terão de atravessar?

Já imaginaram quantos **sistemas** de informação, **tecnologias**,
arquiteturas de **hardware**, etc., os vossos registos irão
encontrar?

A derradeira questão é

Quantas pessoas terá a Segurança Social a **pensar** nestas questões?

Ferramentas para a certificação de repositórios digitais

TRAC

Trustworthy Repositories Audit and Certification

Publicado em 2007 pela

- ▶ RLG (Research Library Group)
- ▶ NARA (National Archives and Records Administration)

Enumera um conjunto de **requisitos** que vão desde a gestão organizacional às infraestruturas de suporte e **que visam** assegurar a **confiança em torno de um repositório**

Os **objetivos** do TRAC são:

- ▶ Fornecer uma ferramenta que permita **auditar, avaliar,** e potencialmente **certificar** repositórios digitais
- ▶ Estabelecer a **documentação necessária** para realizar uma auditoria
- ▶ **Delinear o processo de certificação**
- ▶ Estabelecer metodologias apropriadas para **determinar a robustez e a sustentabilidade de um repositório** digital

O documento apresenta-se sob a forma de uma **checklist** contendo todos os requisitos que um repositório digital deve possuir para que possa ser confiável

A aplicação do **TRAC** **potencia a confiança** junto dos utilizadores do repositório pois

- ▶ Estabelece um **clima de transparência** relativamente aos processos implementados
- ▶ **Auxilia na realização de auditorias internas e externas**

O TRAC encontra-se em vias de se tornar uma norma ISO

- ▶ ISO/DIS 16363 - Audit and certification of **trustworthy digital repositories = TRAC**
- ▶ ISO/DIS 16919 - **Requirements for bodies providing audit and certification of candidate trustworthy digital repositories**

A data prevista para a **publicação** da norma é **Junho de 2012**

DRAMBORA

Digital Repository Audit Method Based on Risk Assessment

Desenvolvido pelo **Digital Curation Centre (DCC)** e pelo **DigitalPreservationEurope (DPE)**

Nasce da **experiência acumulada** após realizar um conjunto de **auditorias** ao longo do ano de 2006

- ▶ Tendo por base uma **versão preliminar do TRAC**

O DRAMBORA é um **documento** e uma **ferramenta interativa** que sistematiza um **processo de auto-avaliação**

- ▶ Foca-se mais em aspetos ligados à **gestão estratégica** e à organização, e **menos em aspetos técnicos** relacionados com o repositório e respectiva plataforma tecnológica
- ▶ Abordagem top-down
- ▶ <http://www.repositoryaudit.eu/>

O DRAMBORA **convida os administradores** de repositórios digitais a

- ▶ Elaborar um **perfil organizacional**, documentando a sua política de depósito, objetivos, responsabilidades, atividades e material custodiado;
- ▶ **Identificar e avaliar os riscos** que poderão impedir a prossecução da sua missão e que ameaçam a salvaguarda dos seus materiais;
- ▶ **Gerir eficazmente os riscos**, mitigando a sua probabilidade de ocorrência;
- ▶ Estabelecer **planos de contingência** eficazes para minimizar os efeitos provocados por riscos que não puderam ser evitados
- ▶ **Relatar os resultados** do processo de auto-avaliação

DSA

Data Seal of Approval



Estabelecido em 2009

“**Selo de garantia**” emitido por um grupo de especialistas em preservação digital

- ▶ i.e. O conselho do Data Seal of Approval

Atesta que um repositório é capaz de **preservar dados científicos** para futura referência e processamento

- ▶ sem que isso acarrete elevados custos ou investimentos para as entidades que os custodiam

Conjunto de boas-práticas que se pretendem que sejam seguidas por organizações responsáveis pela preservação de dados científicos

O DSA é composto por 16 requisitos

- ▶ 3 dizem respeito ao **processo de ingestão**
- ▶ 10 ao repositório e seus **processos internos**
- ▶ 3 dizem respeito ao **processo de disseminação**

O processo de obtenção do “selo” não requer a visita de auditores externos

- ▶ Todo o processo é **conduzido em-linha**
- ▶ O auditado apenas tem de **apresentar evidências de que cumpre os 16 requisitos** definidos
- ▶ Alguns requisitos, mas não todos, podem ser cumpridos em regime de **outsourcing**

Exemplos de requisitos

- ▶ O **produtor deposita os seus dados** no repositório **com informação suficiente para que outros possam aferir a sua qualidade científica** e de forma compatível com as normas éticas exigidas pela disciplina em questão?
- ▶ O repositório **assegura a integridade dos dados** e dos metadados sob sua custódia?
- ▶ O **consumidor respeita os níveis de acesso** à informação definidos pelo repositório?

Os **resultados da auditoria são publicados** em-linha no sítio Web do DSA

- ▶ <http://www.datasealofapproval.org/>

1. The data producer deposits the research data in a data repository with sufficient information for others to assess the scientific and scholarly quality of the research data and compliance with disciplinary and ethical norms.

Minimum Required Statement of Compliance:

3. We are in the implementation phase.

This guideline cannot be outsourced.

Applicant Entry

Statement of Compliance:

4. This guideline has been fully implemented for the needs of our repository.

Evidence:

The types of dataset that the ADS accept into the archive are described in our Collections Policy. See in particular sections 2.3.1 to 2.3.2.3 which discuss how we assess the intellectual content of any potential dataset plus its reuse value, whether it is adequately documented and its suitability for digital preservation.

Collections Policy: <http://archaeologydataservice.ac.uk/advice/collectionsPolicy>

More specific advice and guidance for data producers on metadata and documentation that should be associated with an archive can be found in our Guidelines for Depositors and also our Guides to Good Practice Series which is in the process of being reworked and updated.

Guidelines for Depositors: <http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors>

Guides to Good Practice: <http://guides.archaeologydataservice.ac.uk/>

Where data is submitted for archive that does not have appropriate contextual information, this should be highlighted at the ingest stage and the data producer will be contacted and asked to supply more information.

This process is documented in the ADS Ingest Procedures Manual

Ingest Procedures Manual available from: <http://archaeologydataservice.ac.uk/advice/preservation>

Olhando o TRAC com mais detalhe

A - Infraestrutura organizacional

- ▶ A 1.1 - O repositório possui uma **missão definida que reflete o seu compromisso para com a preservação** e a manutenção do acesso continuado à informação digital custodiada.
- ▶ A 4.1 - O repositório possui **planos de negócio** a curto e longo-prazo que garantam a **sustentabilidade financeira do repositório ao longo do tempo**.

B - Gestão de objetos digitais

- ▶ B 1.1 - O repositório identifica claramente as **propriedades significativas** que é capaz de preservar para os objetos digitais sob a sua custódia.
- ▶ B 4.1 - O repositório implementa **estratégias de preservação** devidamente documentadas

C - Tecnologias, infraestrutura e segurança

- ▶ C 2.1 - O repositório possui o **hardware apropriado para os serviços** que presta à sua comunidade de interesse e todos os procedimentos necessários para avaliar a sua necessidade de substituição.
- ▶ C 3.1 - O repositório **monitoriza sistematicamente** fatores como **dados**, sistemas, pessoal, **planta física do edifício**, **segurança**, ambiente externo ao repositório, etc.

Conclusões

O recurso a **ferramentas como o TRAC** permite identificar

- ▶ Pontos **fortes**
- ▶ Pontos **fracos**
- ▶ Formas de **mitigar riscos** e potenciais pontos de falha

A **publicação dos resultados** das auditorias (internas ou externas) e restante documentação, **confere maior confiabilidade** ao repositório aos olhos de terceiros

- ▶ Princípio da transparência

Não há atualmente **entidades certificadoras** de repositórios

- ▶ O DSA é apenas um primeiro passo neste sentido
- ▶ Quando o TRAC se tornar uma norma ISO, talvez isso mude de figura
- ▶ No entanto, nada nos impede de realizar auditorias internas ou até externas

Aspetos a ter em conta para quem gere um repositório

- ▶ Possuir **procedimentos documentados** sobre todas as atividades realizadas em torno do repositório
- ▶ Possuir registos e **evidências de todas as atividades realizadas**
- ▶ **Identificar potenciais riscos** e delinear planos de contingência caso esses riscos não possam ser evitados
- ▶ Monitorizar o meio ambiente e **definir planos de preservação** digital (i.e. ser pro-ativo e não reativo)



Miguel Ferreira
Diretor técnico e da Qualidade
mferreira@keep.pt

KEEP SOLUTIONS
Rua Rosalvo de Almeida, nº 5
4710-429 Braga, Portugal

www.keep.pt



Sistemas de gestão de informação digital

Miguel Ferreira
Diretor técnico e da Qualidade
mferreira@keep.pt

KEEP SOLUTIONS
Rua Rosalvo de Almeida, nº 5
4710-429 Braga, Portugal

www.keep.pt