

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?

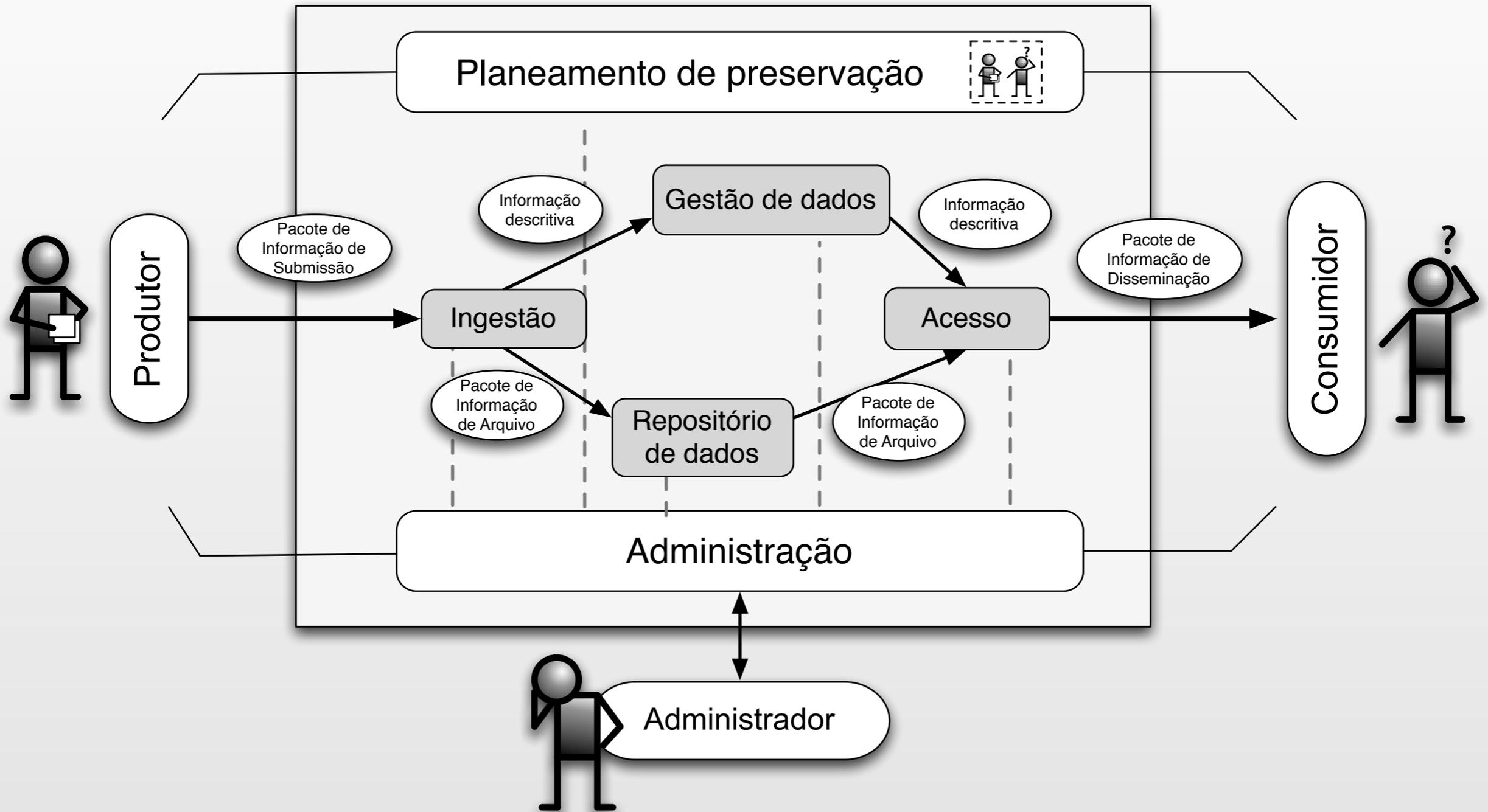
Documentos de apoio

Olhando o TRAC em mais detalhe

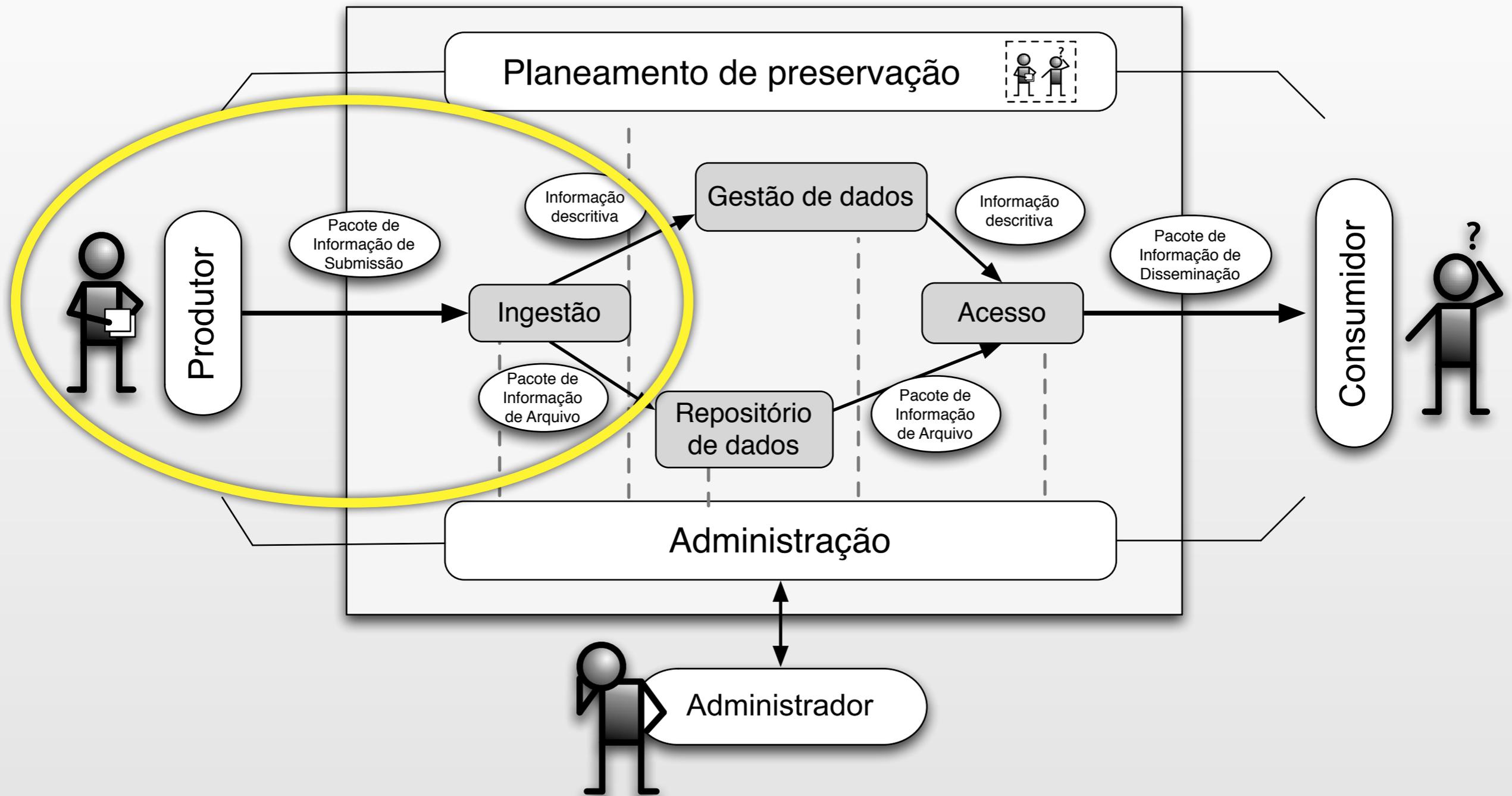
Conclusões

O que é um repositório digital?

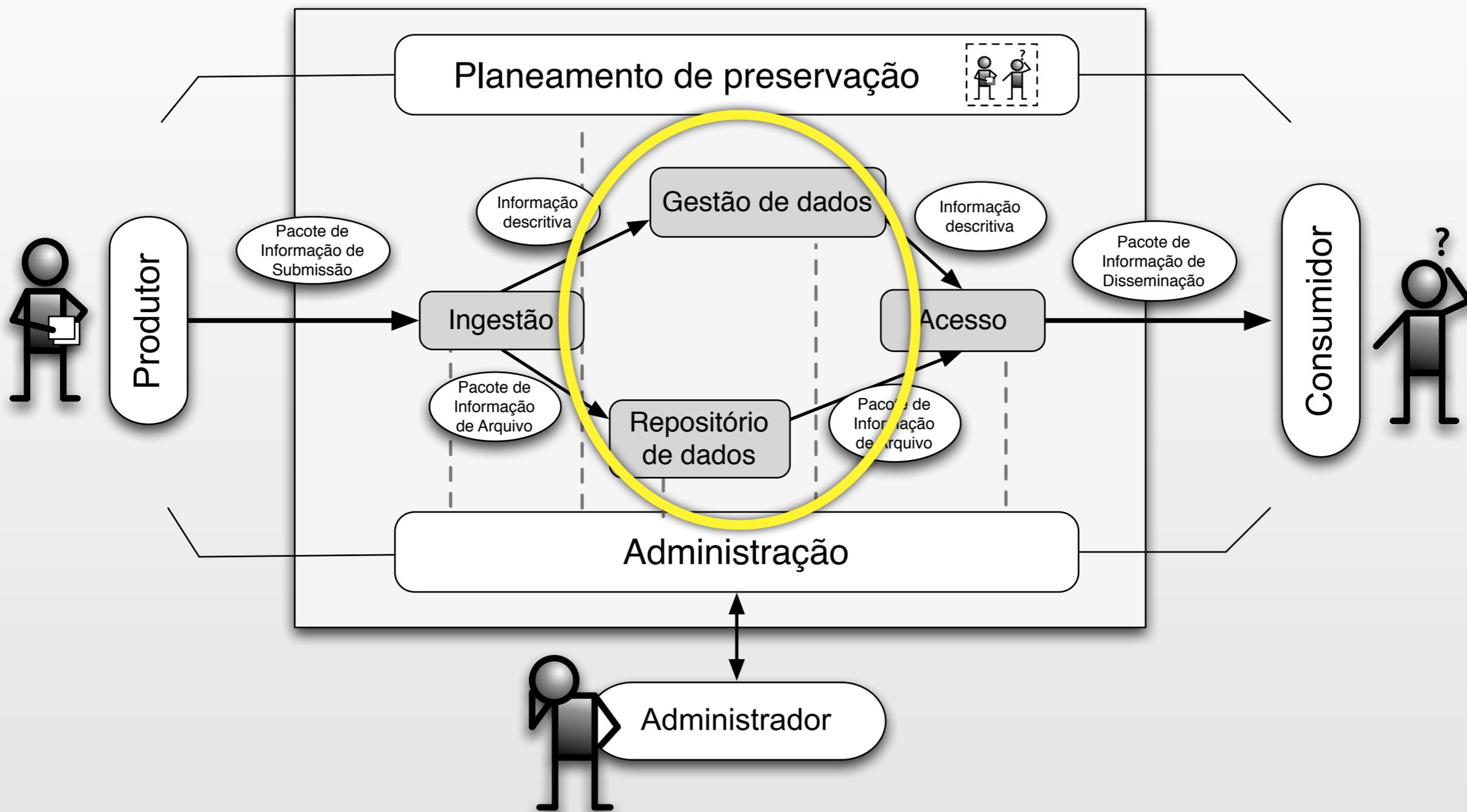
OPEN ARCHIVAL INFORMATION SYSTEM



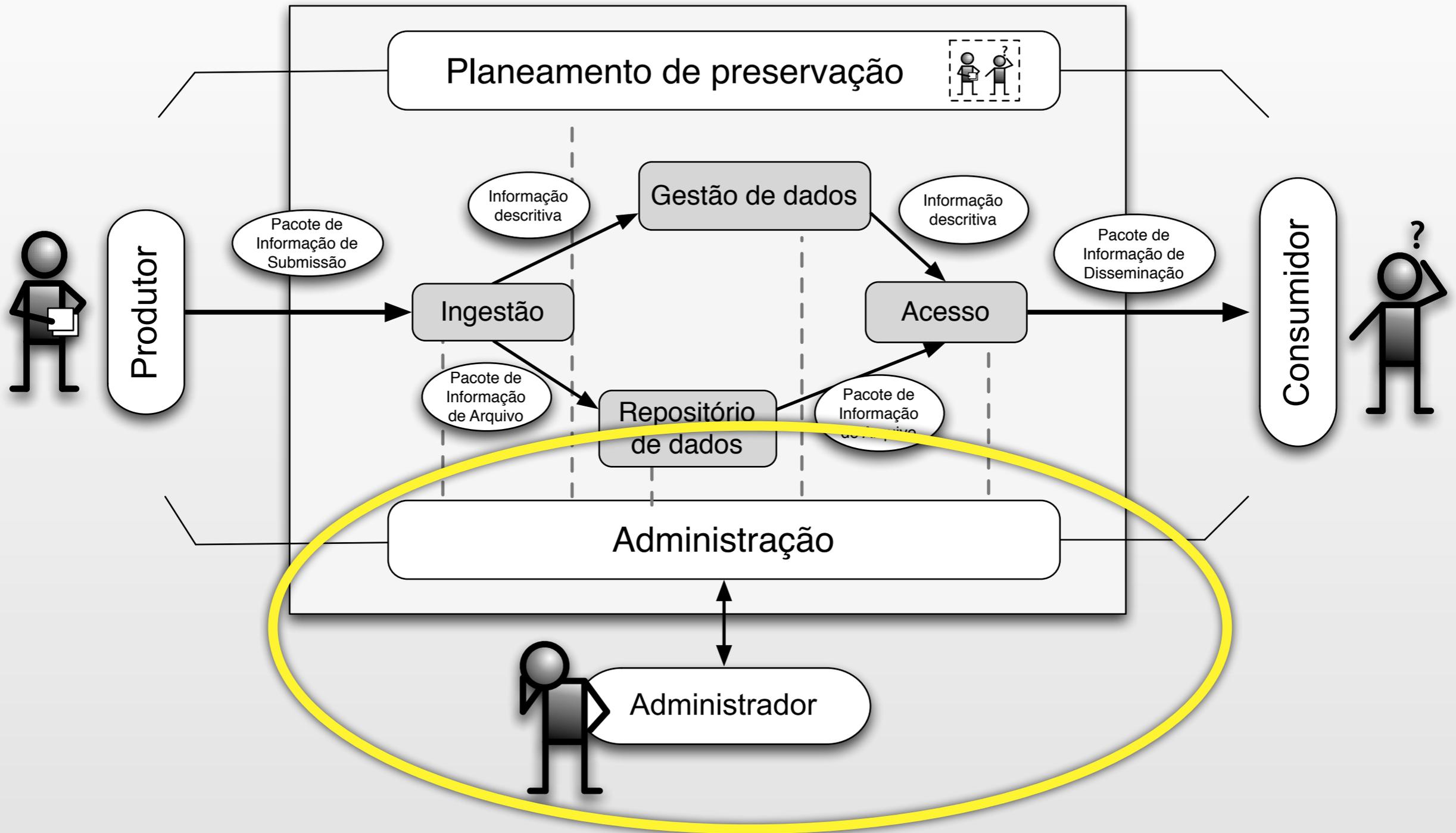
OPEN ARCHIVAL INFORMATION SYSTEM



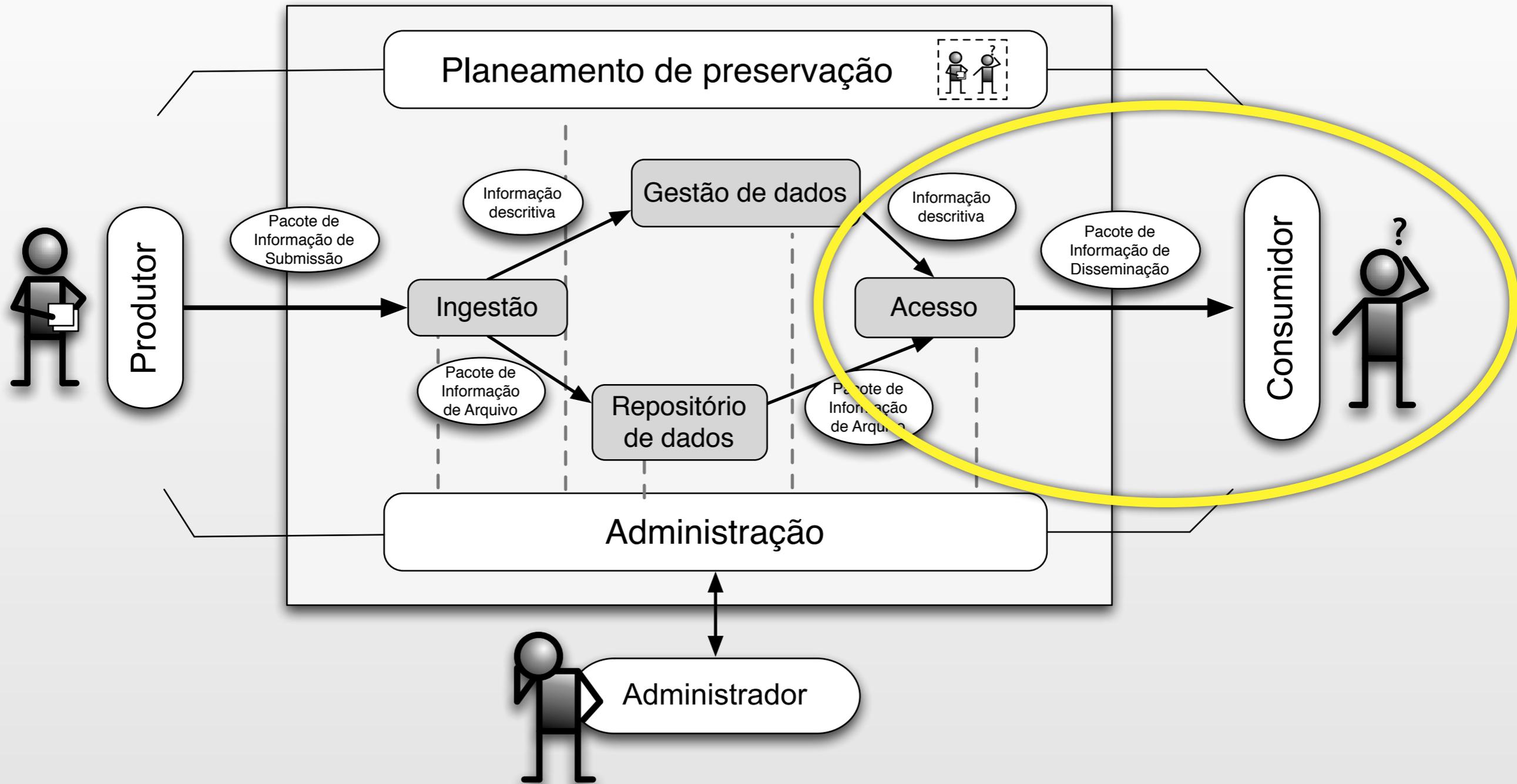
OPEN ARCHIVAL INFORMATION SYSTEM



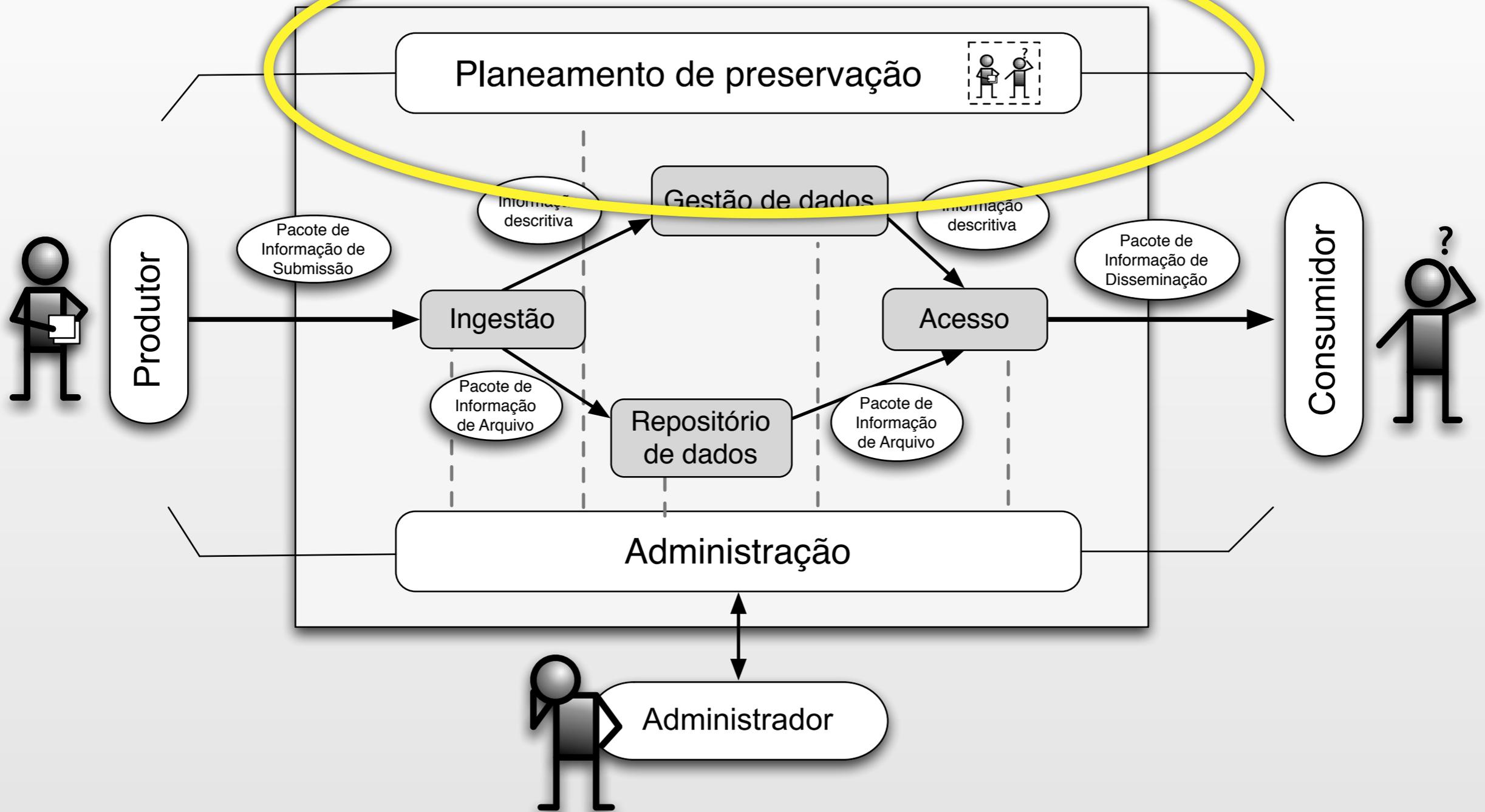
OPEN ARCHIVAL INFORMATION SYSTEM



OPEN ARCHIVAL INFORMATION SYSTEM



OPEN ARCHIVAL INFORMATION SYSTEM



Porquê certificar um repositório digital?

Estabelecer confiança em torno do repositório

- ▶ Produtores, Consumidores, Gestores, Operadores, Financiadores, Etc.

Demonstrar de forma objetiva quanto confiável é o nosso repositório

Estabelecendo critérios, requisitos mensuráveis ou capazes de serem aferidos 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 quantas tecnologias, sistemas de informação, arquiteturas de hardware, etc., os vossos registos irão encontrar?

A derradeira questão é

Quantas pessoas terá a Segurança Social a pensar nestas
questões?

Guias de boas-práticas

TRAC

Trustworthy Repositories Audit and Certification

Publicado em 2007 pela

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

Conjunto de requisitos que vão desde a gestão organizacional à infraestruturas de suporte e que visam garantir 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 qual 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

- ▶ Estabelece um clima de transparência relativamente aos processos implementados
- ▶ Permite a realização de auditorias externas demonstrando um elevado nível de confiança quanto à utilização de práticas adequadas na operação do repositório e no tratamento da informação que lhe fora confiada

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

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 de um conjunto de auditorias realizadas pelo DCC 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.

O DRAMBORA convida os administradores de repositórios digitais a

- ▶ Elaborar um perfil organizacional, descrevendo e 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
- ▶ Produzir relatórios sobre o processo de auto-avaliação

DSA

Data Seal of Approval



Estabelecido em 2009

“Selo de garantia” emitido por um grupo de especialistas

- ▶ 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 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 aos produtores
- ▶ 10 ao repositório e seus processos internos
- ▶ 3 dizem respeito ao acesso à informação

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 evidencias de que cumpre os 16 requisitos definidos

Alguns exemplos de requisitos

- ▶ O produtor deposita os seus dados científicos 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

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 - Organizational Infrastructure

A 1 - Governance & organizational viability

- ▶ A 1.1 - Repository has a mission statement that reflects a commitment to the long-term retention of, management of, and access to digital information.

A 2 - Organizational structure & staffing

- ▶ A 2.1 - Repository has identified and established the duties that it needs to perform and has appointed staff with adequate skills and experience to fulfill these duties.

A 3 - Procedural accountability & policy framework

- ▶ A 3.5 - Repository has policies and procedures to ensure that feedback from producers and users is sought and addressed over time.

A 4 - Financial sustainability

- ▶ A 4.1 - Repository has short- and long-term business planning processes in place to sustain the repository over time.

A 5 - Contracts, licenses & liabilities

- ▶ A 5.2 - Repository contracts or deposit agreements must specify and transfer all necessary preservation rights, and those rights transferred must be documented.

B - Digital Object Management

B 1 - Ingest: acquisition of content

- ▶ B 1.1 - repository identifies properties it will preserve for digital objects.

B 2 - Ingest: creation of the archival package

- ▶ B 2.1 - repository has a identifiable, written definition for each AIP or class of information preserved by the repository.

B 3 - Preservation planning

- ▶ B 3.2 - repository has mechanisms in place for monitoring and notification when Representation Information (including formats) approaches obsolescence or is no longer viable.

B 4 - Archival storage & preservation/maintenance of AIPs

- ▶ B 4.1 - Repository employs documented preservation strategies.

B 5 - Information management

- ▶ B 5.1 - Repository articulates minimum metadata requirements to enable the designated community(ies) to discover and identify material of interest.

B 6 - Access management

- ▶ B 6.5 - Repository access management system fully implements access policy.

C - Technologies, Technical Infrastructure, & Security

C 1 - System infrastructure

- ▶ C 1.3 - Repository manages the number and location of copies of all digital objects.

C 2 - Appropriate technologies

- ▶ C 2.1 - Repository has hardware technologies appropriate to the services it provides to its designated community(ies) and has procedures in place to receive and monitor notifications, and evaluates when hardware technology changes are needed.

C 3 - Security

- ▶ C 3.1 - Repository maintains a systematic analysis of such factors as data, systems, personnel, physical plant, and security needs.

Conclusões

O recurso a estes guias de boas-práticas permite identificar pontos fortes, debilidades e formas de mitigar riscos

Publicar os resultados das auditorias internas, assim como os restantes procedimentos de gestão, confere uma maior confiabilidade ao repositório

Não há atualmente entidades certificadoras de repositórios

- ▶ O DSA é apenas um primeiro passo neste sentido
- ▶ Talvez quando o TRAC se tornar uma norma ISO isso mude de figura

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/ou evidências de todas as tarefas 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)



Desenvolvemos o futuro hoje...

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