

# OAIS

## Open Archival Information System: uma brevíssima introdução



**Luis Fernando Sayão**  
**Comissão Nacional de Energia Nuclear**  
**Centro de Informações Nucleares**  
[lsayao@cnen.gov.br](mailto:lsayao@cnen.gov.br)





# ... e há uma preocupação mundial em torno da questão

GIORDANO  
BRUNO  
*Notano.*

De l'infinito universo  
*et Mondi.*

*Ad illustrissimum signor de  
Mauuiscera.*



Stampato in Venetia.  
Anno. M. D. LXXVIII.



United Nations Educational,  
Scientific and Cultural organization

“... entre a profusão de atividades no domínio do universo digital, nós estamos criando uma nova forma de herança que é parte integral da herança da humanidade.”

A Conferência Geral da Unesco (2001) reconheceu que existe uma herança digital de valor permanente sob risco de se perder, se nenhuma ação urgente e apropriada não for tomada



## :: O risco de uma amnésia digital é real...

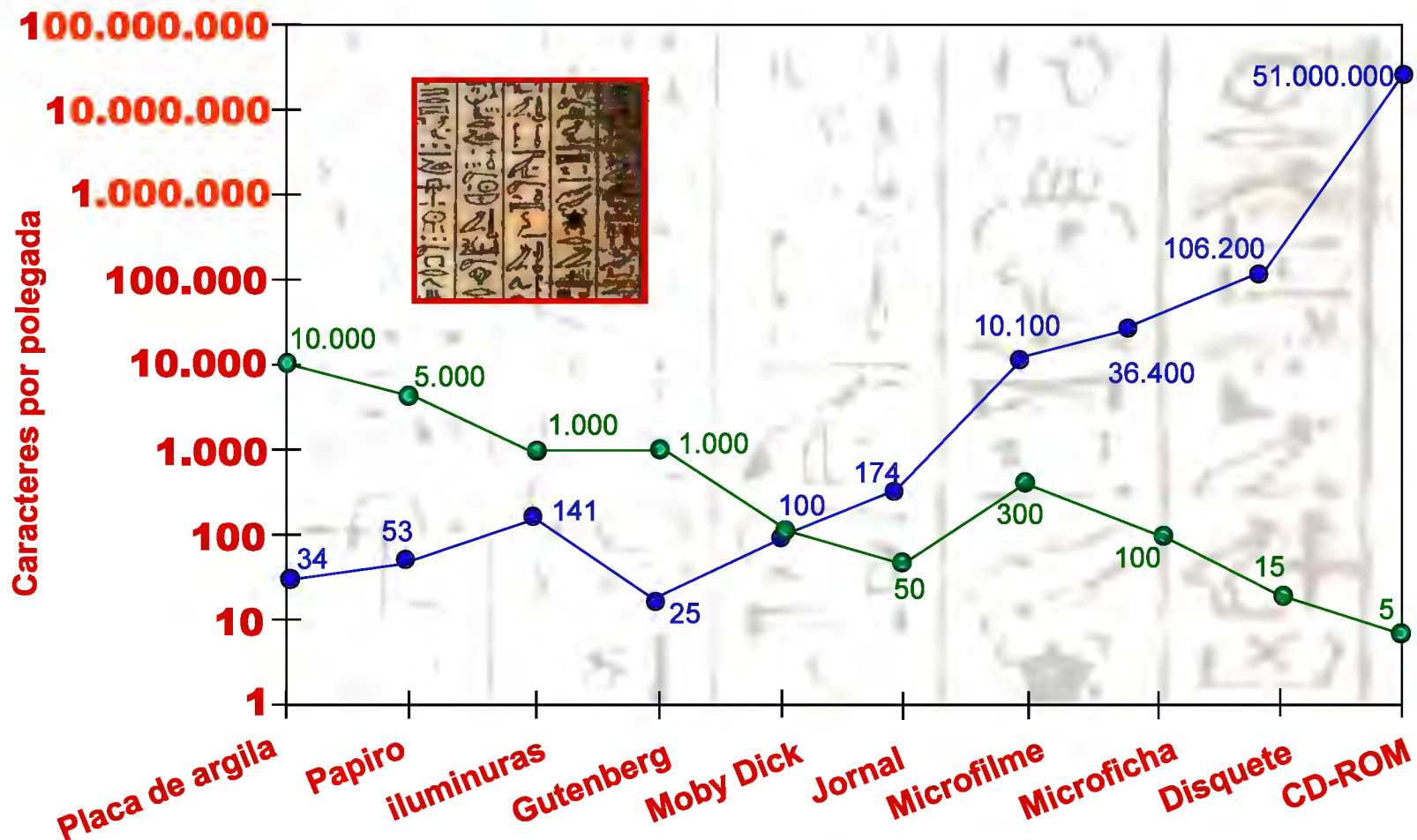


001001100010010  
011011000010010  
110011000110010  
001001011000111  
011110001001011  
000011001110110  
111001101100011  
001001100010010  
011011000010010  
110011000110010  
001001011000111

O dilema que se instala é que a tecnologia digital coloca a humanidade sob o risco de uma amnésia digital, ao mesmo tempo em que abre oportunidades extraordinárias em todos os campos

# ::: Densidade de informação X longevidade das mídias

**A informação no formato digital é mais frágil do que os fragmentos de papiro encontrados nas tumbas dos faraós**



Baseado em Paul Conway – Preservação no universo digital

## **:: A vulnerabilidade das mídias digitais...**



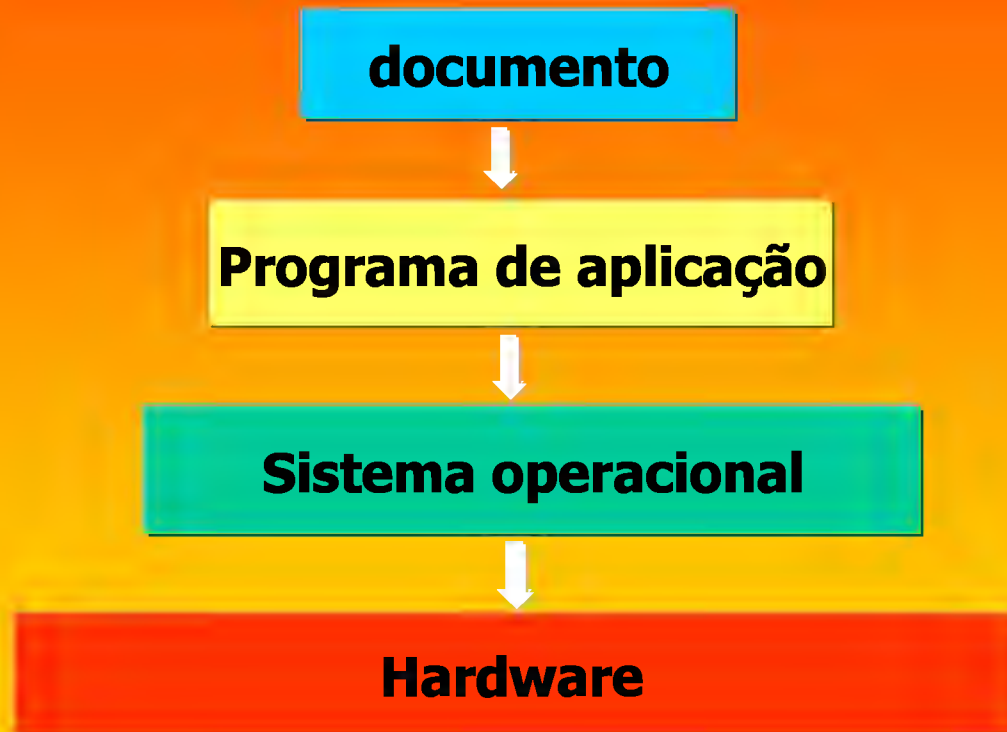
- :: degradação dos bits – efeitos físicos, químicos, magnéticos etc.**
- :: obsolescência da mídias**
- :: obsolescência de formatos**
- :: dependência a softwares incompatíveis ou obsoletos**
- :: dependência a ambientes de softwares obsoletos – S.O. I/O drivers, etc.**
- :: dependência a hardwares obsoletos**

# ... causada pelas suas inúmeras dependências

[gerencias e tecnológicas]



001001100010010  
011011000010010  
110011000110010  
001001011000111  
011110001001011  
000011001110110  
111001101100011  
001001100010010  
011011000010010  
110011000110010  
001001011000111





...e a apesar da nossa incapacidade preservar e controlar a informação digital



“ A nossa compulsão em produzir informações digitais é infinitamente maior do que a nossa capacidade de garantir acesso a ela, e não há em contrapartida estratégias consolidadas para garantir o acesso de longo prazo as informações de valor ”

## **...mesmo porque tudo mudou na forma de registrar informação**



### **REGISTROS INVISÍVEIS**

**Pela primeira vez em 3.500 anos de atividade de gestão de documentos, produzimos registros que não existem para o olho humano. Completamente diferentes das placas de argila da Babilônia, dos papiros egípcios, dos pergaminhos romanos, do papel moderno e mesmo do microfilme**

### **GESTÃO**

**Pela primeira vez, profissionais das mais diversas áreas, sem capacitação para a gestão de estoques informacionais, são responsáveis por criá-los e armazená-los**

### **DESMATERIALIZAÇÃO dos SUPORTES**

**E o mais importante, pela primeira vez não estamos, produzindo, gerenciando e guardando artefatos físicos, mas tentando entender e preservar padrões virtuais que dão a informação digital seu conteúdo, estrutura, contexto e assim o seu significado, que são completamente controlados por software**



# :: Mas como preservar?

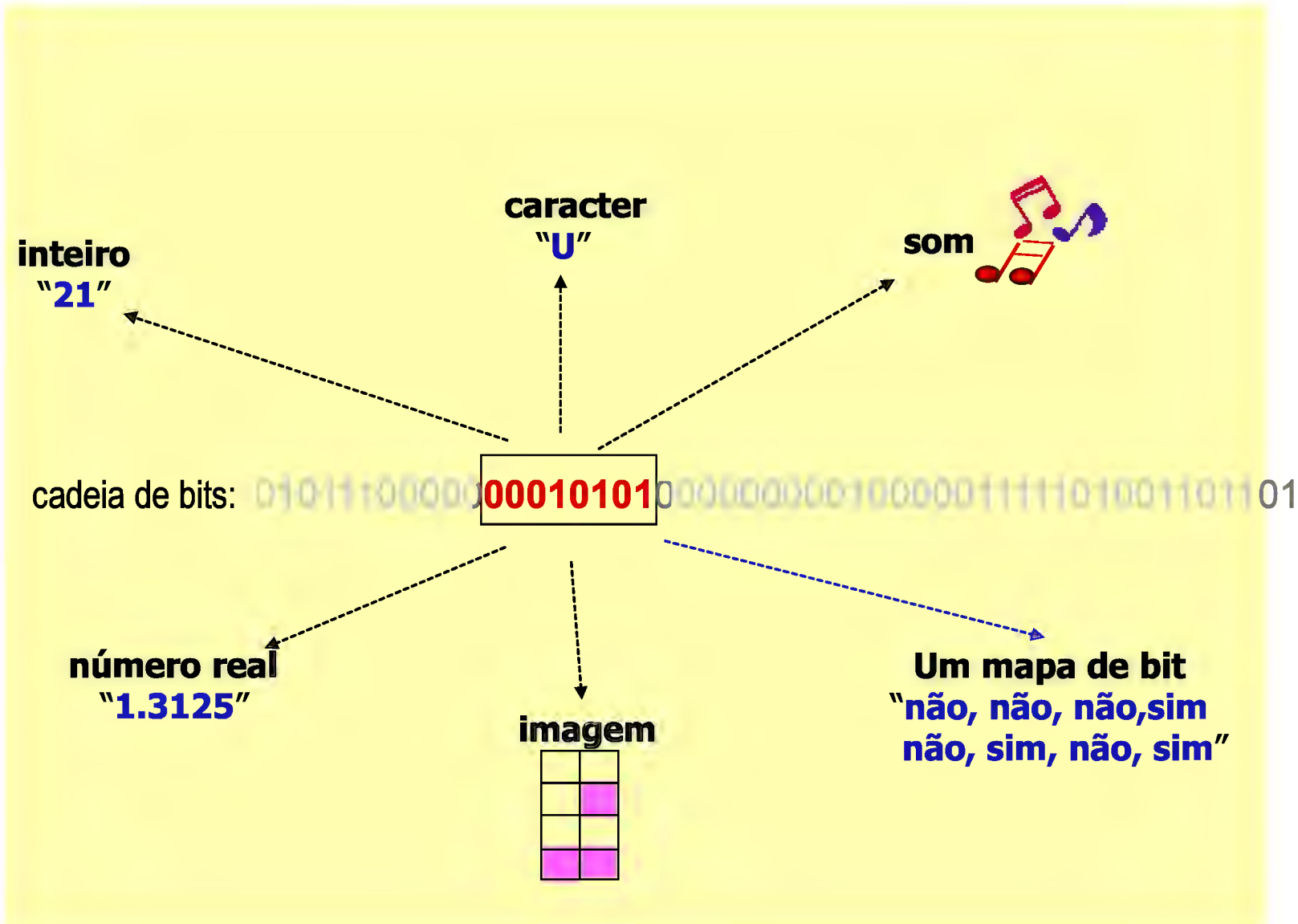


001001100010010  
011011000010010  
110011000110010  
001001011000111  
011110001001011  
000011001110110  
111001101100011  
001001100010010  
011011000010010  
110011000110010  
001001011000111  
011110001001011  
001001  
0110110

**A natureza digital de uma informação é dada pela forma como ela é armazenada: seqüências de 0's e 1's, isto é, cadeias de bits (bitstream)**

**Podemos, portanto, concluir que a forma mais básica de preservar a integridade de um objeto informacional é preservar a configuração de bits que o define como um objeto único**

**:: Um byte numa cadeia de bits pode representar qualquer coisa**  
[e deve ter o seu significado interpretado]



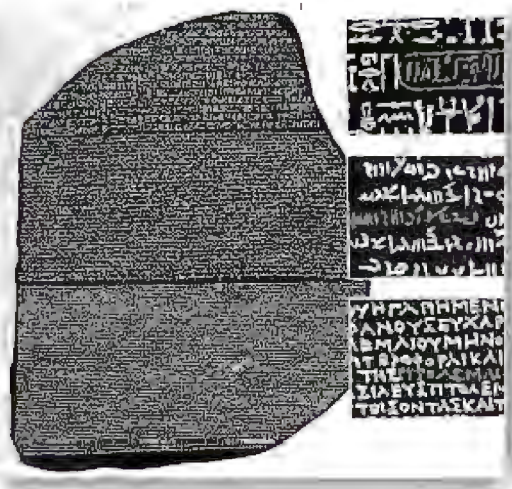
## :: Salvar os bits é necessário mas não suficiente

“ Recuperada uma cadeia de bits ela deve ter o seu significado interpretado, pois a maioria dos arquivos contém informações cujo significado só pode ser explicitado pelo software que as criou. ”

:: essa interpretação é fortemente dependente de formatos, códigos e estruturas – que não podem ser explicitamente representadas na cadeia de bits

:: portanto, para se **compreender um documento** é necessário conhecer o **significado** do seu conteúdo na linguagem para o qual ele foi construído

:: no caso de um documento digital essa linguagem é um **programa de computador**



:: **salvar a cadeia de caracteres sem salvar o seu interpretador é como salvar o hieróglifos sem salvar a Pedra de Rosetta**

# :: Por que é tão difícil preservar documentos digitais?

“ A preservação digital envolve não somente a retenção do objeto informacional em si, mas também do seu significado. É necessário, portanto, que as técnicas de preservação sejam capazes de compreender e recriar a forma original ou a função do objeto de forma que seja assegurada sua autenticidade e acessibilidade ”

:: não podemos simplesmente salvar documentos digitais como documentos físicos

0110010100100010001  
0001110100110101101  
1100100010100100111  
0110010100100010001  
0001110100110101101  
1100100010100100111  
0110010100100010001  
0001110100110101101  
1100100010100100111

1100100010100100111  
01100001110100110101101  
001111100100010100100111  
100010110010100100010001  
01100001110100110101101  
001111100100010100100111



## :: Mas o que devemos preservar?



O problema primordial da preservação é o conteúdo, ou seja, a substância intelectual contida nos objetos informacionais.

Quando um livro é preservado no seu formato original, todos os aspectos do livro são preservados, seu layout, seu formato, seu conteúdo. Todos os elementos estão inextricavelmente unidos.

... e o processamento ocorre na cabeça do leitor

Quando um documento é preservado, suas assinaturas, carimbos, autenticação, também são preservados



## :: Mas o que devemos preservar?



Mas no mundo digital a noção de conteúdo é complexa e estratificada

Objetos digitais são facilmente decompostos em elementos

No mundo digital, a primeira tarefa é identificar os aspectos múltiplos que devem ser preservados

**A preservação digital deve garantir que o maior número possível desses aspectos devam persistir no tempo.**

# no mundo digital a noção de conteúdo é complexa e estratificada

## Fixar os limites do objeto

um livro é um livro, mas qual é o limite de um objeto digital?

## Preservar a cadeias de bits

é a presença física, formada por séries de “0”s e “1”s

## Preservar o conteúdo

capacidade de acessar o conteúdo no seu nível mais baixo tal como um texto ASCII

## Preservar a apresentação

forma, layout, fontes, tamanho, margens, colunas, cores, paginação... em muitos casos está especificado separadamente – folha de estilo

## Preservar as funcionalidades

componentes multimídia, conteúdo dinâmico, interoperabilidade, busca

## Preservar a autenticidade

verificar se o documento é o que diz ser, se as transformações preservaram a sua forma original; se sofreu mudanças não autorizadas

## Localizar e referenciar ao longo do tempo

nomear e referenciar um objeto digital de forma a distingui-lo de outras versões, cópias e edições

## Preservar a proveniência

origem e cadeia de custódia confirmam a autenticidade e a integridade

## Preservar o contexto

objetos digitais são também definidos por suas dependências a hardware, software, modo de distribuição e links com outros objetos



# Afinal o que é preservação digital?

## :: Preservação Física

Foco nas mídias e na sua renovação

## :: Preservação Lógica

Foco nos formatos e dependência de hardware e software que mantenham legíveis e interpretáveis a cadeia de bits

## :: Preservação intelectual

Foco no conteúdo intelectual e na sua autenticidade e integridade

[é necessário ainda preservar do aparato – na forma de metadados – necessários para localizar, recuperar e interpretar a informação digital; assim como o monitoramento e a instrumentalização da comunidade alvo, para que ela possa compreender plenamente a informação no momento do seu acesso]









001100011100011010100010110001011101001

## O que é o Modelo de Referência OAIS?



**O Modelo Referencial OAIS – Open Archival Information System – é uma estrutura conceitual que disciplina e orienta um sistema de arquivo dedicado a preservar e manter o acesso a informação digital por longo prazo**

001100011100011010100010110001011101001

## Como é definido um OAIS segundo a Norma ISO



Um OAIS consiste numa organização de pessoas e sistemas que aceitaram a responsabilidade de preservar informação e torná-la disponível para uma **Comunidade Alvo**



## Quais são os objetivos de um OAIS?

**O objetivo do modelo de referência é aumentar o grau de consciência e compreensão dos conceitos relevantes para o arquivamento de objetos digitais, especialmente entre instituições não arquivísticas.**

:: ampliar a consciência e a compreensão dos conceitos arquivísticos relevantes necessários para a preservação e o acesso de longo prazo

:: prover conceitos necessários para organizações não arquivísticas participarem efetivamente no processo de preservação

:: permitir comparações: arquitetura, operação, estratégias e técnicas de preservação

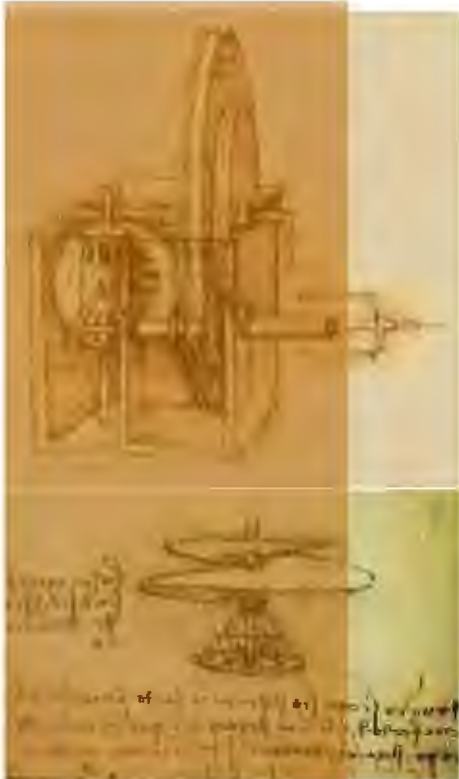
:: estabelecer fundamentos que possam ser expandidos para informações que não estão no formato digital

:: ampliar o consenso em torno dos elementos e processos voltados para a preservação de longo prazo e acesso a informação digital, promover a expansão do mercado e o apoio dos fornecedores

:: orientar a identificação e produção de padrões relacionados ao OAIS



# Aplicabilidade



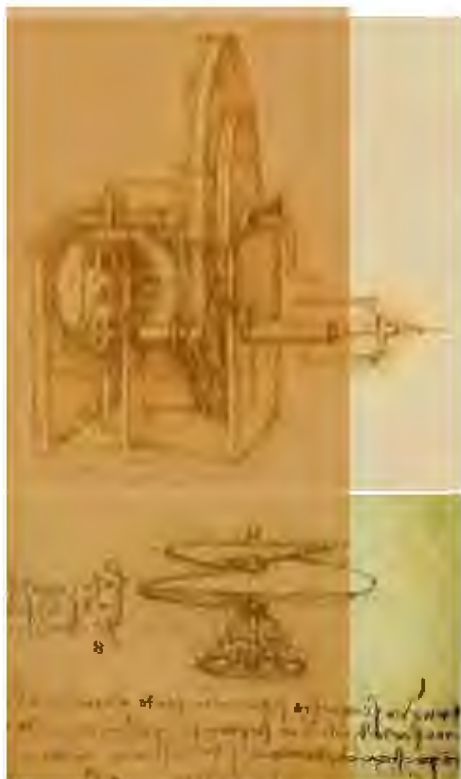
**:: O modelo OAIS pode ser aplicado a qualquer arquivo – físico ou digital -, mas ele é especificamente dirigido para organizações que têm a responsabilidade de tornar a informação disponível por longo prazo**

**:: O Modelo é também de interesse das organizações e pessoas que criam informações que podem necessitar de preservação de longo prazo, bem como das organizações que adquirem tais informações**

**:: O modelo de referência não especifica um projeto ou qualquer tipo de implementação**

0110010100100010001  
0001110100110101101  
1100100010100100111  
0110010100100010001  
0001110100110101101  
1100100010100100111  
0110010100100010001  
0001110100110101101  
1100100010100100111

# Antecedentes



0110010100100010001  
0001110100110101101  
1100100010100100111  
0110010100100010001  
0001110100110101101  
1100100010100100111  
0110010100100010001  
0001110100110101101  
1100100010100100111

**Desenvolvimento coordenado pelo CCSDS – Consultative Committee for Space Data System - ligado à NASA por encomenda da ISO**

**:: Participação ampla em escala mundial**

[do Brasil: INPE e CTA]

**:: Um produto de 10 anos de trabalho**

**.. Red Book – maio de 1999**

**.. Blue Book – janeiro de 2002**

[<http://www.ccsds.org/documents/650x0b1.pdf>]

**.. Aprovação da ISO 1471 – fevereiro de 2003**

**.. longo prazo?**

**.. aberto?**

**(Modelo de Referência OAIS é reconhecido como a mais importante documento conceitual voltado para a preservação digital)**

# Ambiente OAIS

O Ambiente OAIS é derivado da interação de quatro entidades

:: **PRODUTORES**

Pessoas ou sistemas clientes que fornecem a informação a ser preservada

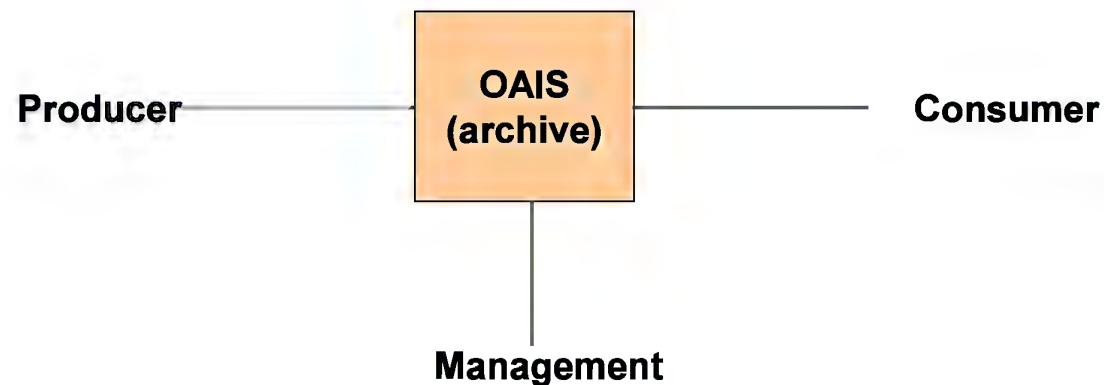
:: **ADMINISTRAÇÃO**

Entidade responsável pelo estabelecimento de políticas mais gerais do arquivo.

:: **CONSUMIDORES**

Pessoas ou sistemas clientes que interagem com os serviços do OAIS para recuperar e adquirir informações preservadas de interesse. Uma classe especial de Consumidores é a Comunidade Alvo (Designated Community). [ Comunidade Alvo é o conjunto de Consumidores capazes de compreender a informação preservada]

:: **ARQUIVO OAIS**



# Ambiente OAIS - exemplos

O Ambiente OAIS é derivado da interação de quatro entidades

:: ARQUIVO: Planetary Data System

ADMINISTRAÇÃO: National Aeronautics Data Administration – NASA

PRODUTORES: NASA Flight Project

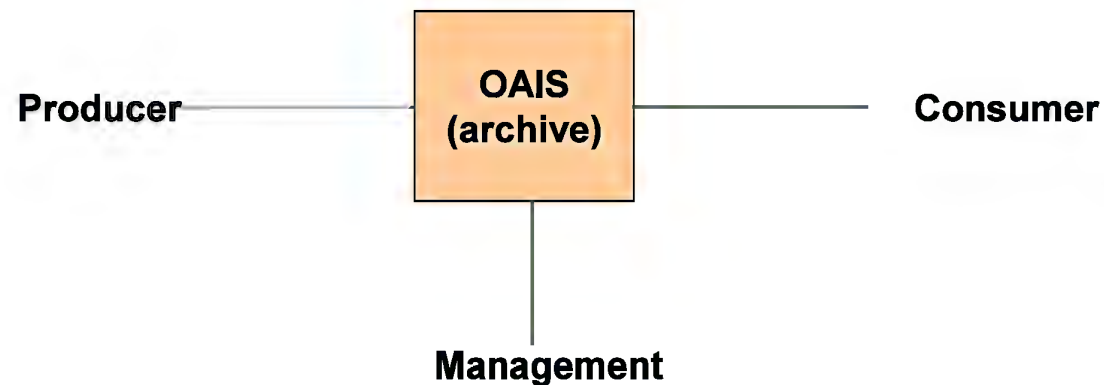
COMUNIDADE ALVO: Comunidade Científica Planetária

:: ARQUIVO: Electronic and Special Media Records Administration

ADMINISTRAÇÃO: National Archives and Records Administration - NARA

PRODUTORES: U.S. government agencies

COMUNIDADE ALVO: Público em geral



# Responsabilidades de um OAIS



- :: Negociar a aceitação apropriadas de informações vindas de Produtores de Informação**  
[critérios: tipo, assunto, fonte, originalidade, singularidade, etc. | mídias, formatos ...  
| conformação com padrões e políticas | informações descritivas | informações para preservação]
- :: Obter controle suficiente das informações para garantir a preservação de longo prazo**  
[propriedade do objeto físico X propriedade intelectual | acordo de transferência | copyright | autoridade para modificar]
- :: Determinar, por si mesmo ou em conjunto com outros parceiros, que comunidades devem se tornar Comunidades Alvo e, portanto, capacitadas a entenderem a informação fornecida**
- :: Assegurar que a informação a ser preservada seja compreensível para a Comunidade Alvo de forma independente, ou seja sem auxílio dos especialistas que produziram a informação**
- :: Seguir políticas e procedimentos documentados que garantam que a informação esteja preservada contra todas as contingências, e que possibilite que a informação seja disseminada como cópia autenticada do original ou rastreável até o original**
- :: Tornar a informação preservada disponível para a Comunidade Alvo**  
[visões | buscas | documentação | acesso X preservação | restrição de acesso distribuição]

# Modelo de Informação do OAIS

## <DEFINIÇÃO DE INFORMAÇÃO>

:: Uma definição clara de informação é fundamental para que um OAIS possa preservá-la

:: **Informação:** qualquer forma de conhecimento passível de intercâmbio, esta informação é expressa (representada) por meio de algum tipo de dado [exemplo: um string de bits acompanhados por uma descrição | um livro expresso por caracteres observáveis combinados com o conhecimento da linguagem | dicionário e gramática para interpretação para compreensão (informação de representação)]

:: **Pode existir de duas formas (OBJETO DE DADOS):**

- .. Objeto Físico [ exemplo: rocha lunar]
- .. Objeto Digital [ seqüência de bits]

## <BASE DE CONHECIMENTO>

:: **A Base de Conhecimento:** uma pessoa ou sistema possui um conjunto de informações que permite que compreendam a informação recebida

## <INTERPETAÇÃO DO OBJETO DE DADOS>

:: A interpretação do Objeto de Dados como informação significativa pela Comunidade Alvo é obtida por meio da combinação da **Base de Conhecimento** dessa comunidade e a **Informação de Representação** associado ao **Objeto de Dados**

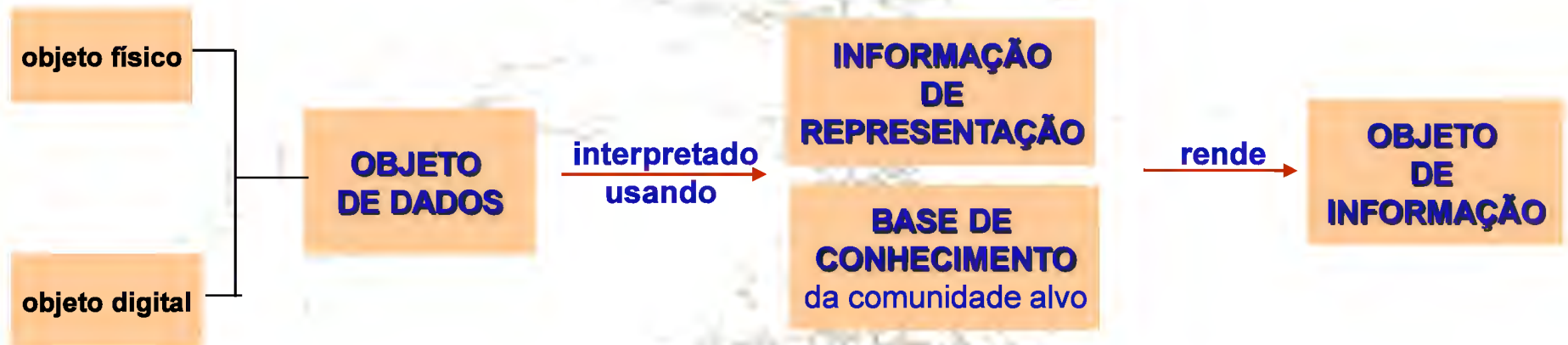


# Modelo de Informação do OAIS

## <INTERPETAÇÃO DO OBJETO DE DADOS>

:: A interpretação do Objeto de Dados como informação significativa pela Comunidade Alvo é obtida por meio da combinação da **Base de Conhecimento** dessa comunidade e a **Informação de Representação** associado ao **Objeto de Dados**

:: Podemos concluir que de modo geral que “Dado interpretado usando sua Informação de Reapresentação produz Informação (**Objeto de Informação**)”



:: Para um Objeto de Informação ser preservado com sucesso é crítico para o OAIS claramente identificar e compreender o **Objeto de Dados** e a sua Informação de Representação

:: Isto é um desafio importantíssimo para a preservação digital !!!









# Modelo de Informação do OAIS

## <DEFINIÇÃO DE PACOTE DE INFORMAÇÃO>

### :: SIP – PACOTE DE SUBMISSÃO DE INFORMAÇÃO

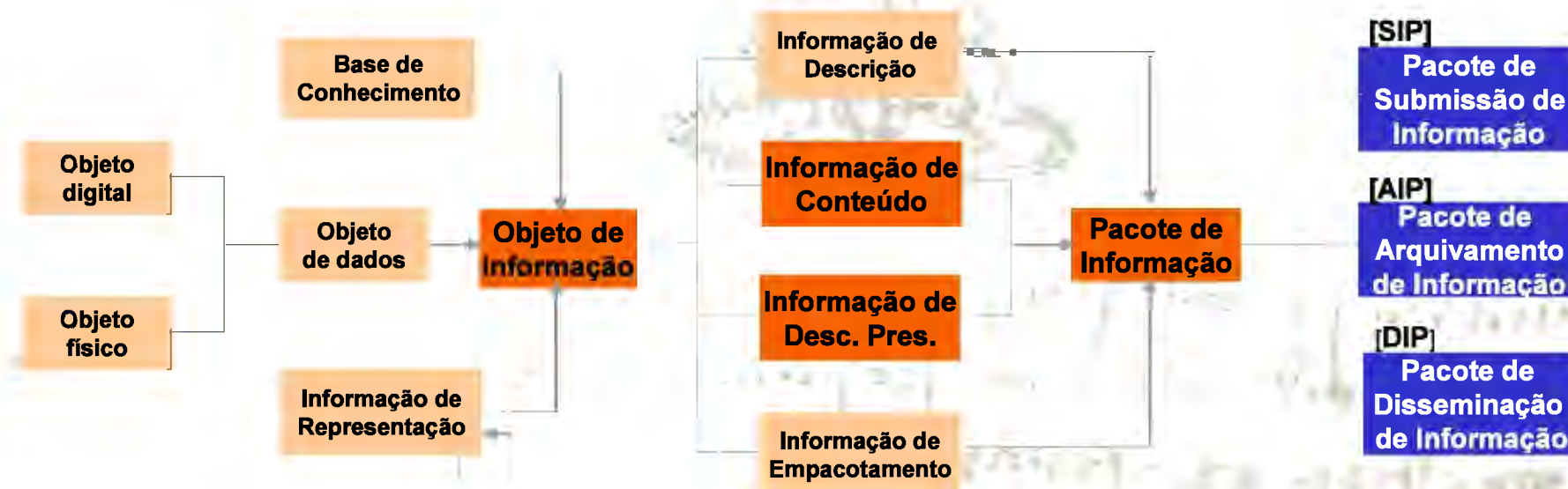
[é o Pacote de Informação que entregue pelo Produtor a um OAIS para construção de um ou mais AIP, Pacote de Arquivamento de Informação]

### :: AIP- PACOTE DE ARQUIVAMENTO DE INFORMAÇÃO

[ é o pacote de informação que será objeto de preservação]

### :: DIP- PACOTE DE DISSEMINAÇÃO DE INFORMAÇÃO

[é o Pacote de Informação derivado de um ou mais AIP, recebido pelo Consumidor em resposta a uma requisição dirigida ao OAIS]









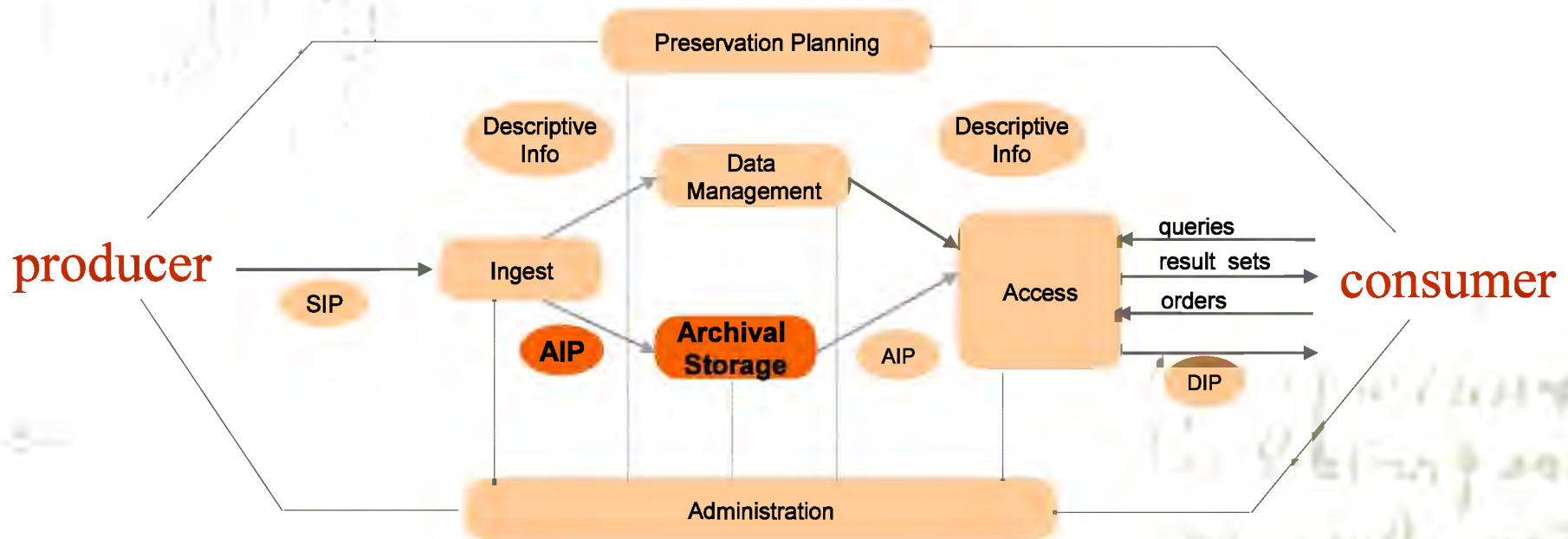


# Modelo de Informação do OAIS

## ARMAZENAMENTO

### Funções

- :: **Recepção dos dados** [recebe um *pedido de armazenamento* de um AIP da Recepção | move o AIP para uma área de armazenamento permanente | seleção do tipo de mídia | preparação dos volumes | transferência física | confirmação de armazenamento para a a Recepção | identificação do armazenamento do AIP]
- :: **Gerencia da hierarquia da área de armazenamento** [posicionamento das AIP's na mídia apropriada]
- :: **Renovação de mídias** [reprodução das AIP's ao longo do tempo]
- :: **Verificação de erros** [garantia estatisticamente aceitável de danos nos AIP's nas transferências internas]
- :: **Recuperação de falhas** [duplicação | armazenamento físico separado]
- :: **Fornecimento de dados** [fornecimento de cópias de AIP's para o Acesso]





# Modelo de Informação do OAIS

## <GERENCIAMENTO DE DADOS>

- :: Serviços e funções para popular, manter e acessar tanto a Informação Descritiva que identifica e documenta os acervos do arquivo quanto os dados administrativos usados para gerenciá-los
- :: Banco de dados que captura e gerencia todos os metadados necessários para operar o sistema
- :: Consultas são dirigidas; visões são geradas; relatórios são emitidos

## <funções>

- :: ADMINISTRAÇÃO DA BASE DE DADOS | ATUALIZAÇÃO | EXECUÇÃO DE BUSCAS

