

# Acesso persistente a itens de informação: Proposta de sustentabilidade para a solução adotada pelo INPE

*Gerald Jean Francis Banon*

<gerald.banon@gmail.com>

URL do documento original:

<http://urlib.net/J8LNKB5R7W/3NKH24H>

Esta apresentação está licenciada com base numa licença 3.0 CC BY-NC-ND



I Seminário Temático da Rede RBP de 2017  
São José dos Campos, 25 de maio de 2017

## Conteúdo

**URL persistente**

**Alguns identificadores globais**

**Sistema de resolução de IBI**

**Proposta de manutenção de IBI**

**Referências**

## ***URL persistente***

(1/13)

Segundo Richard WHITT<sup>†</sup> (2017):

“Digital dark age”

The **integrity of academic scholarship** relies in some measure on the supporting references cited in documents. In the online context, there has been uncovered the growing prevalence of so-called “**reference rot**,” a combination of link rot (broken URLs) and content drift (links leading to changed or vanished information).

“Era da escuridão digital”

A **integridade de trabalhos acadêmicos** depende, de certa maneira, das referências de apoio citadas em documentos. No contexto on-line, tem sido descoberta a crescente prevalência das chamadas “**referências corrompidas**”, uma combinação de vínculos corrompidos (URLs quebradas) e derivação de conteúdo (vínculos que levam a informações alteradas ou desaparecidas).

<sup>†</sup> Corporate Director for Strategic Initiatives at Google

# *URL persistente*

(2/13)

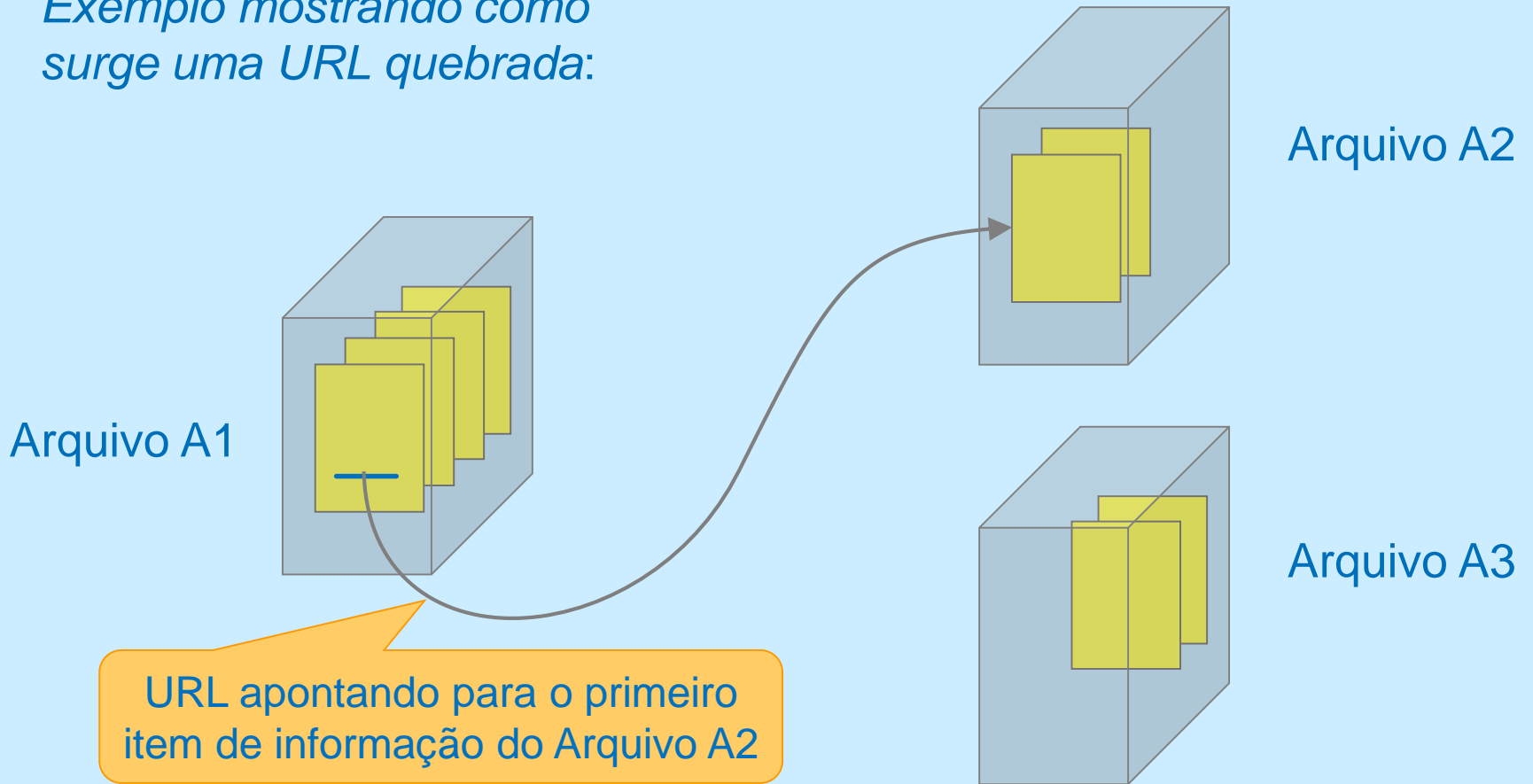
*Exemplo mostrando como surge uma URL quebrada:*



# URL persistente

(3/13)

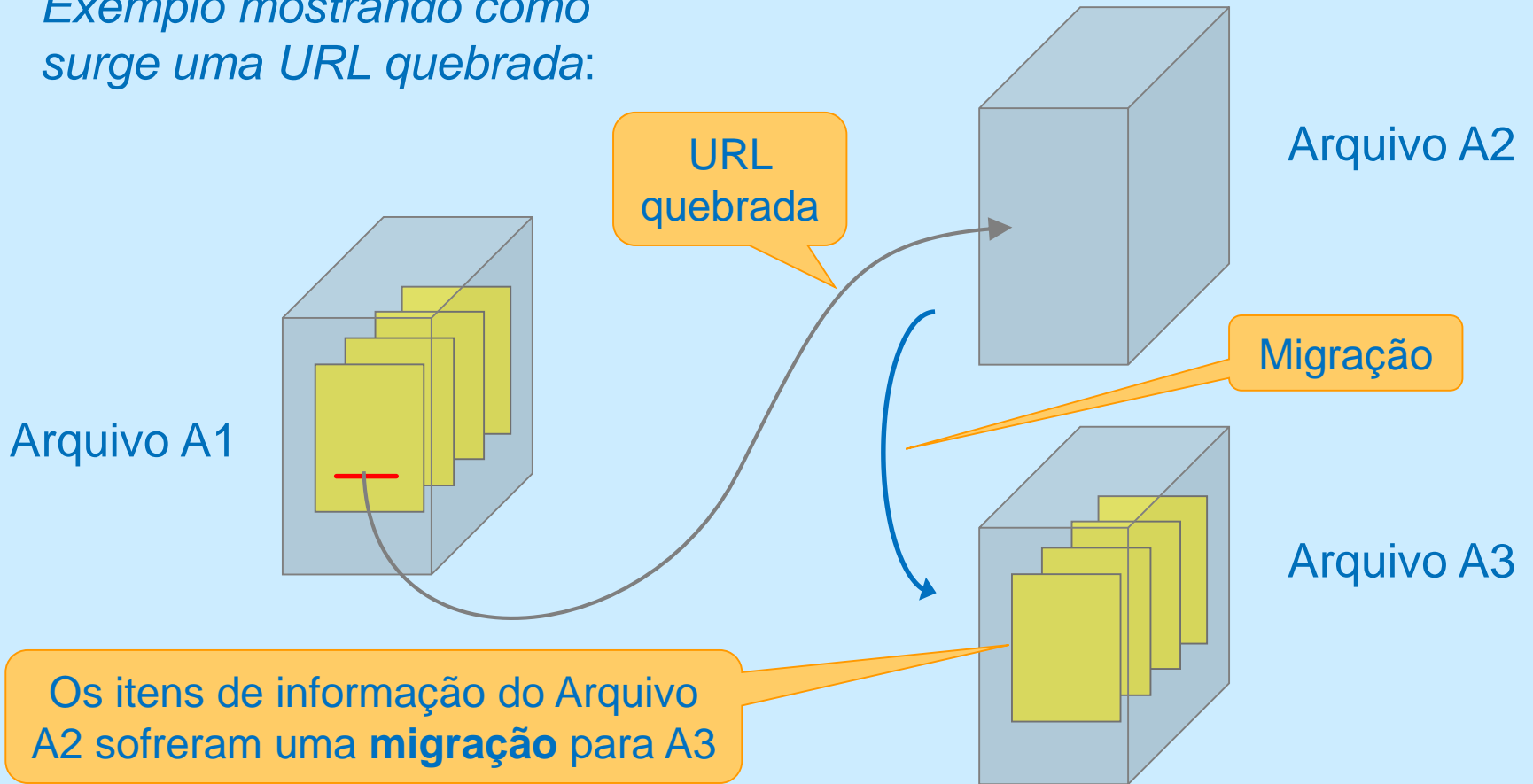
*Exemplo mostrando como surge uma URL quebrada:*



# URL persistente

(4/13)

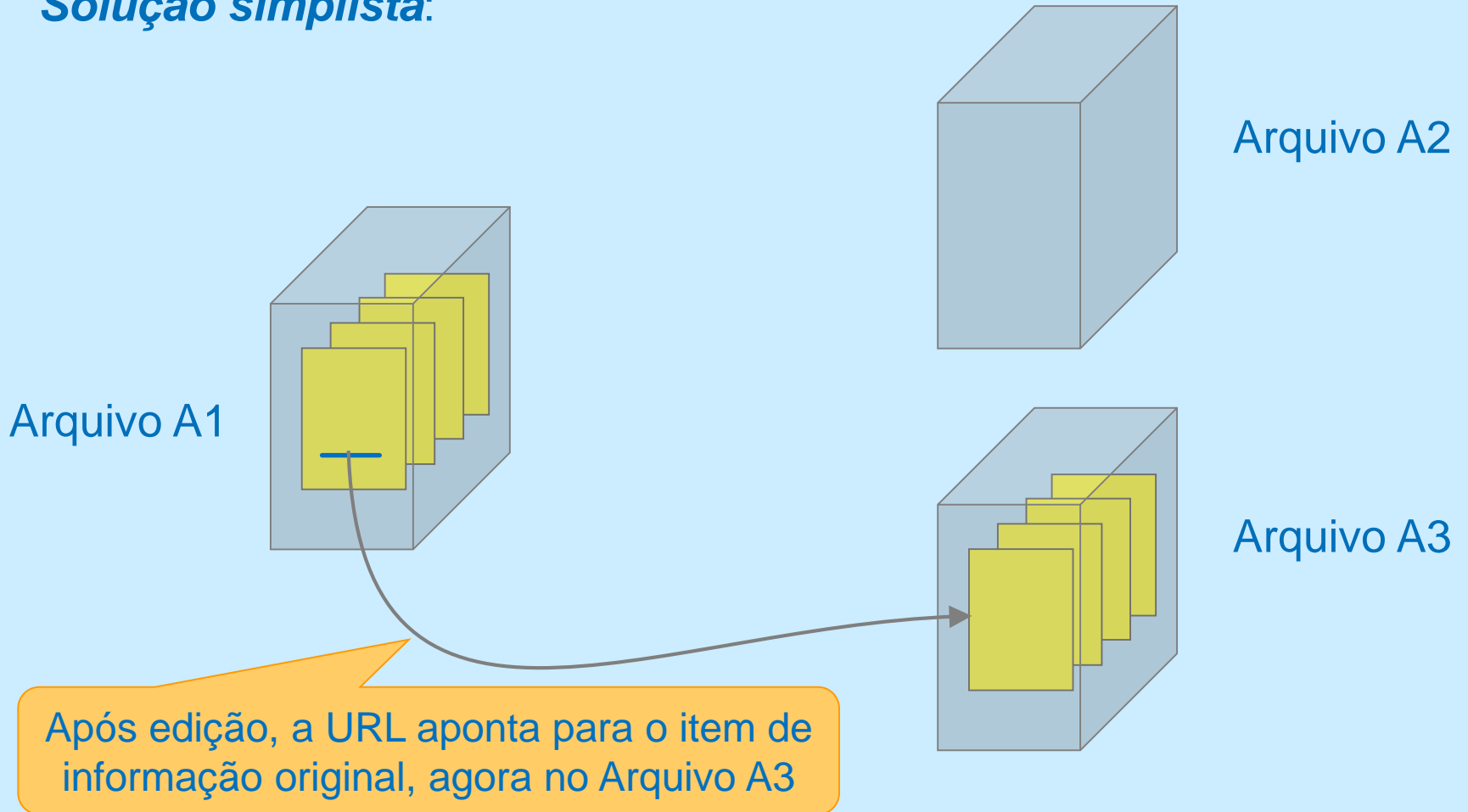
Exemplo mostrando como surge uma URL quebrada:



# *URL persistente*

(5/13)

*Solução simplista:*



## ***URL persistente***

(6/13)

Por razões de negócio, sociais e técnicas **as migrações são inevitáveis** e necessárias para assegurar a preservação digital.

A solução simplista não é definitiva e é **inviável** na prática.

A solução definitiva passa pelo uso de **URL persistentes**.



## ***URL persistente***

(7/13)

Item de informação

Uma URL é uma referência usada para recuperar um recurso localizado na *Web*.

A forma geral (simplificada) de uma URL é:

**esquema** : // **domínio** / **caminho** / **recurso**

Para se manter **persistente** cada elemento formando uma URL deve se manter para sempre **inalterado**.

Usualmente uma URL persistente possui a forma simplificada:

**esquema** : // **domínio** / **identificador**

Identificador de item de informação

# URL persistente

(8/13)

Exemplos de URL:

URL com referência explícita à localização atual do item de informação na *Web*

`http://mtc-m16c.sid.inpe.br/col/sid.inpe.br/mtc-m18/2012/07.12.18.08/doc/CCSDS%20650.0-M-2.pdf`

`http://urlib.net/8JMKD3MGP8W/3C9EP6P`

Prefixo

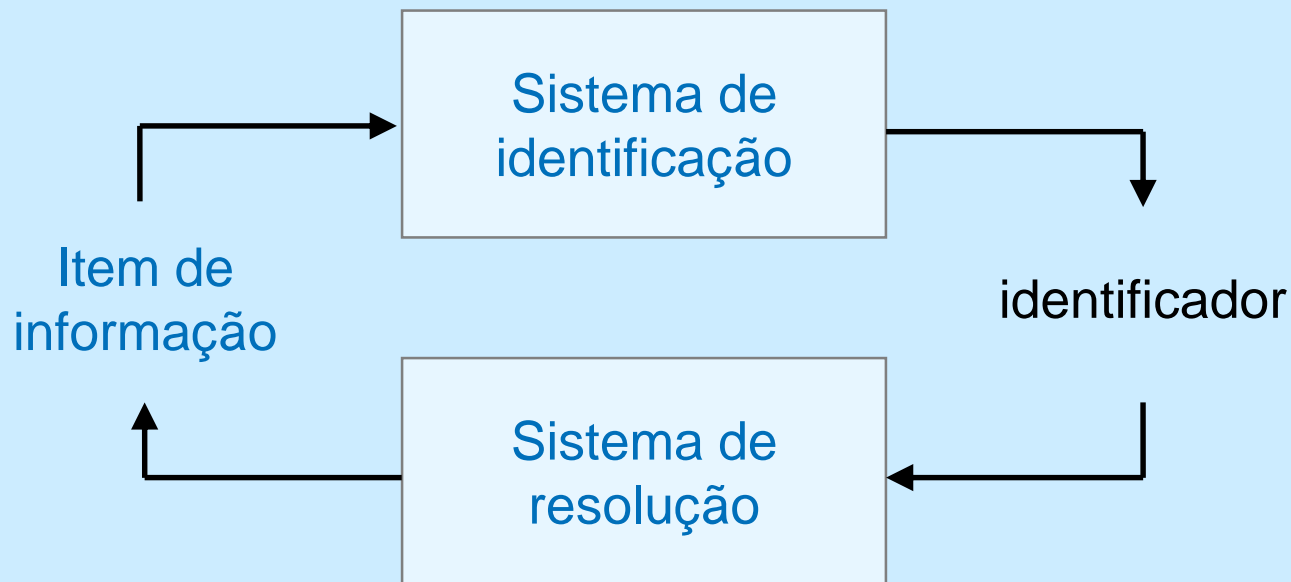
Sufixo

URL persistente

## *URL persistente*

(9/13)

A construção de URL persistente passa pelo uso de **um sistema de identificação** que associa a cada item de informação um **identificador único** (exclusivo) e de **um sistema de resolução** capaz de retornar o item de informação com base no seu identificador.



## ***URL persistente***

(10/13)

Segundo a recomendação CCSDS 652.0-M-1 (sobre Auditoria e Certificação de Repositórios Digitais Confiáveis), um Arquivo **confiável** deve identificar todos seus itens de informação.

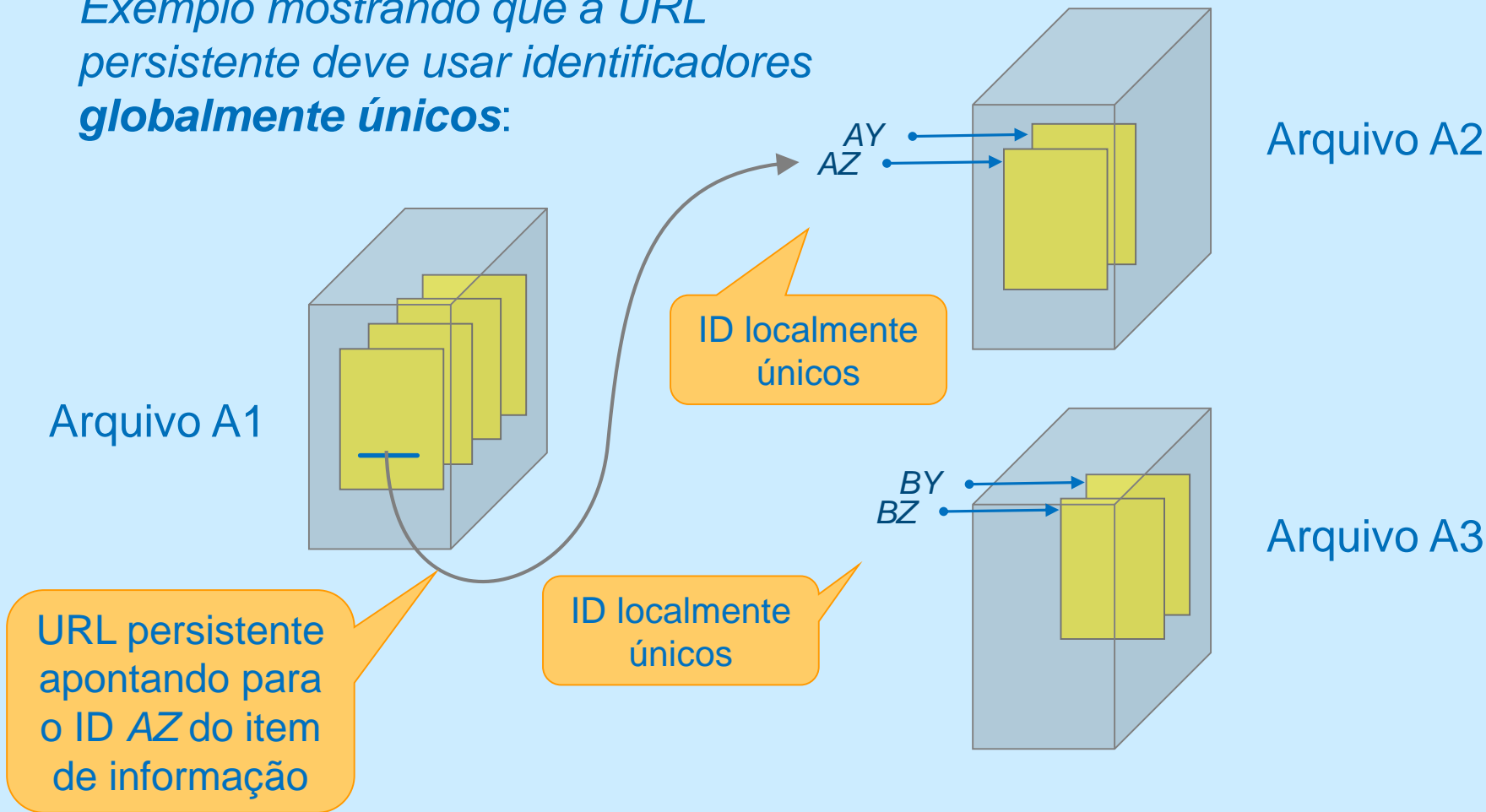
No entanto, para tornar as URL persistentes, a identificação dos itens de informação **deve ser globalmente único** (exclusivo).

*CCSDS: Consultative Committee for Space Data Systems*  
(Comitê Consultivo para Sistemas Espaciais de Dados)

# URL persistente

(11/13)

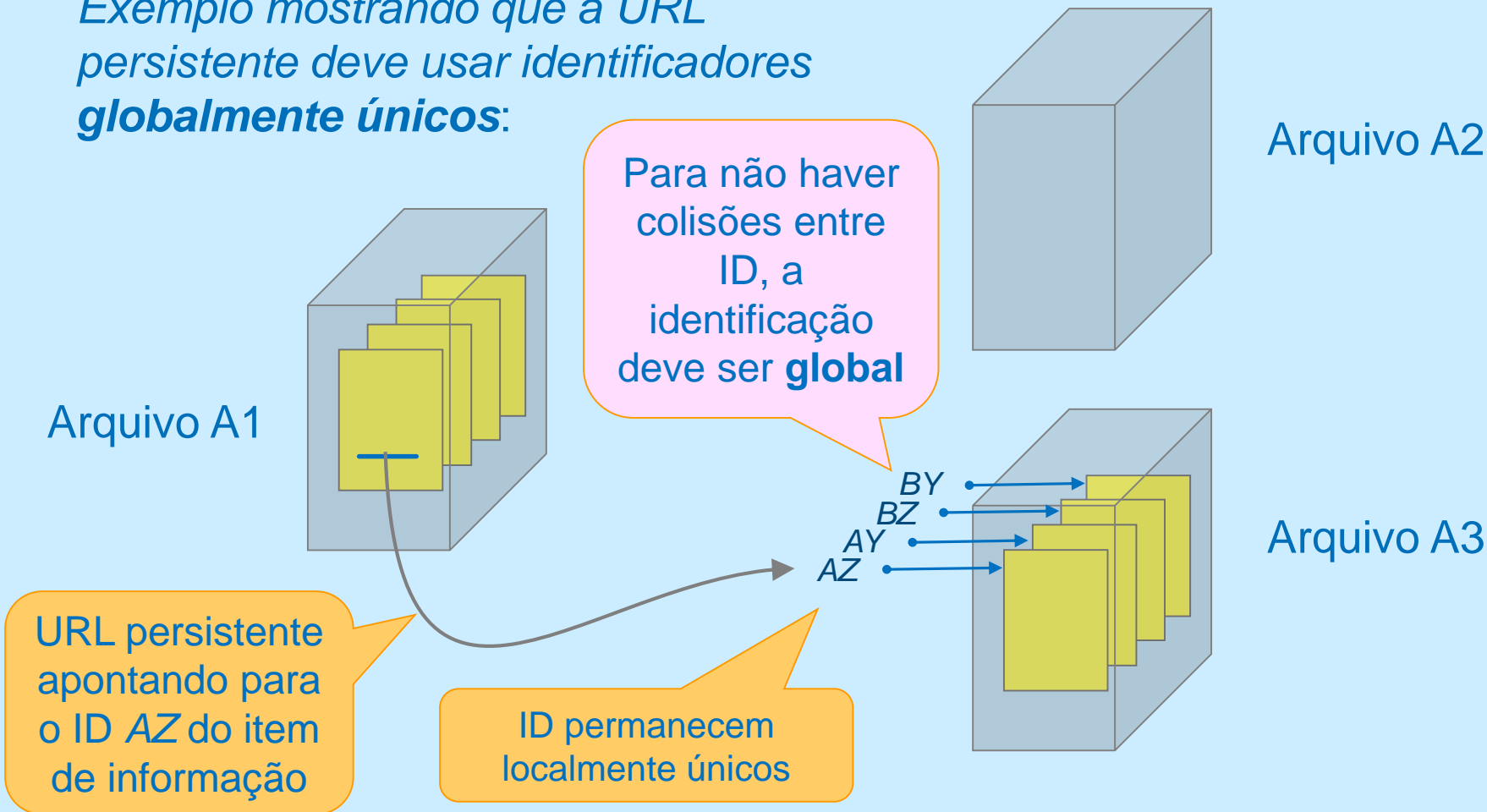
Exemplo mostrando que a URL persistente deve usar identificadores **globalmente únicos**:



# URL persistente

(12/13)

Exemplo mostrando que a URL persistente deve usar identificadores **globalmente únicos**:



## URL persistente

(13/13)

Exemplos de 2 URL em referências bibliográficas de um trabalho acadêmico de 2015 disponível na **Plataforma SciELO**:

<https://dx.doi.org/10.1590/1809-4422ASOC675V1812015en>

URL quebrada

IEA - INSTITUTO DE ECONOMIA AGRÍCOLA. Banco de dados. São Paulo: IEA, 2010. Disponível em: <[http://ciagri.iea.sp.gov.br/bancoiea/indices\\_new.aspx](http://ciagri.iea.sp.gov.br/bancoiea/indices_new.aspx)>.

URL com referência explícita à localização do item de informação

No INPE, **todos** os 24688 itens de informação originais são acessíveis por URL persistentes

INPE - INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. Uso de imagens de satélite de sensoriamento remoto para mapear a área cultivada com cana-de-açúcar no estado de São Paulo - Safra 2009/10. 2010. Disponível em: <<http://urlib.net/sid.inpe.br/mtc-m19@80/2010/02.09.18.47>>.

URL persistente

# Alguns identificadores globais

(1/3)

Entidades sem fins lucrativos

Nome	Data	Prefixo	Sufixo	Manutenção
PURL	1995	Identificador próprio	Espaço de nomes local	OCLC (Zepheira)
		Depende da criação de centralidades		
Handle (DOI)	1995 (2000)	Identificador próprio	Espaço de nomes local	CNRI (DOI Foundation)
		Depende da criação de uma centralidade		
IBI	1995	Endereço Internet	Data de criação do IBI	AMI (a ser criada)
		Aproveita a centralidade da Internet		

Em uso no INPE

OCLC: *Online Computer Library Center*

CNRI: *Corporation for National Research Initiatives*

AMI: *Associação para a Manutenção de IBI*



## *Alguns identificadores globais*

(2/3)

Nome	Taxa para o usuário (Arquivo)
PURL	0
Handle	US\$ 50 por ano e por prefixo
DOI	US\$ 1 por DOI (CrossRef )
IBI	0

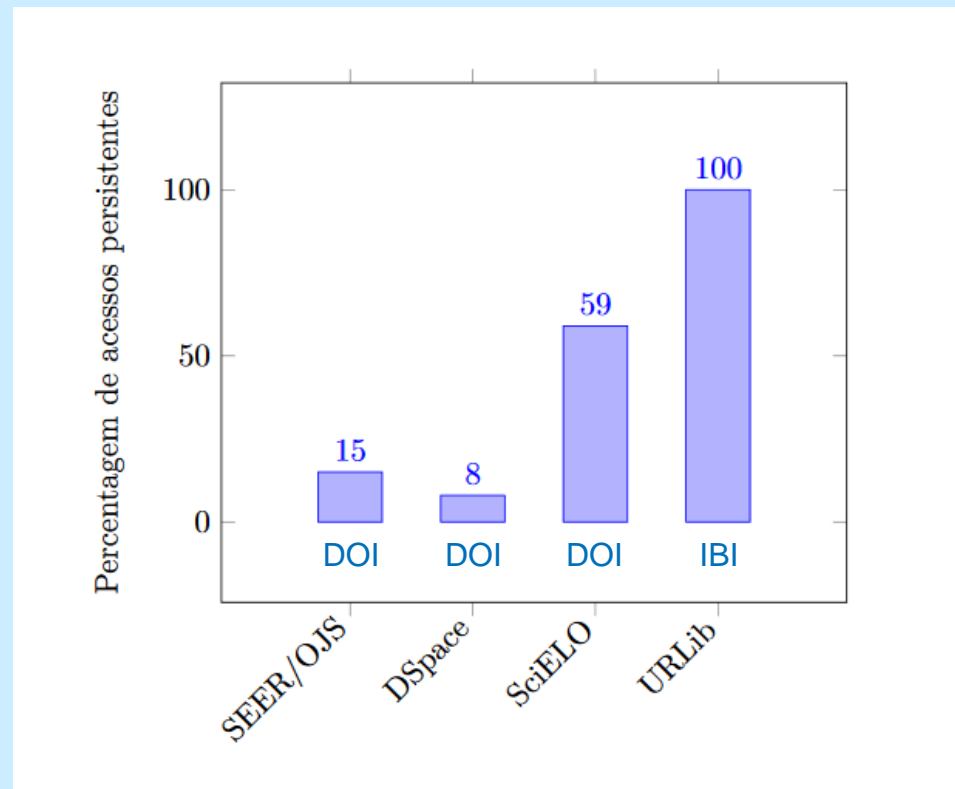
(novembro de 2015)

A Rede SciELO hospede 573.525 artigos (maio 2017)

## *Alguns identificadores globais*

(3/3)

*Uso de acesso persistente por plataformas no Brasil:*

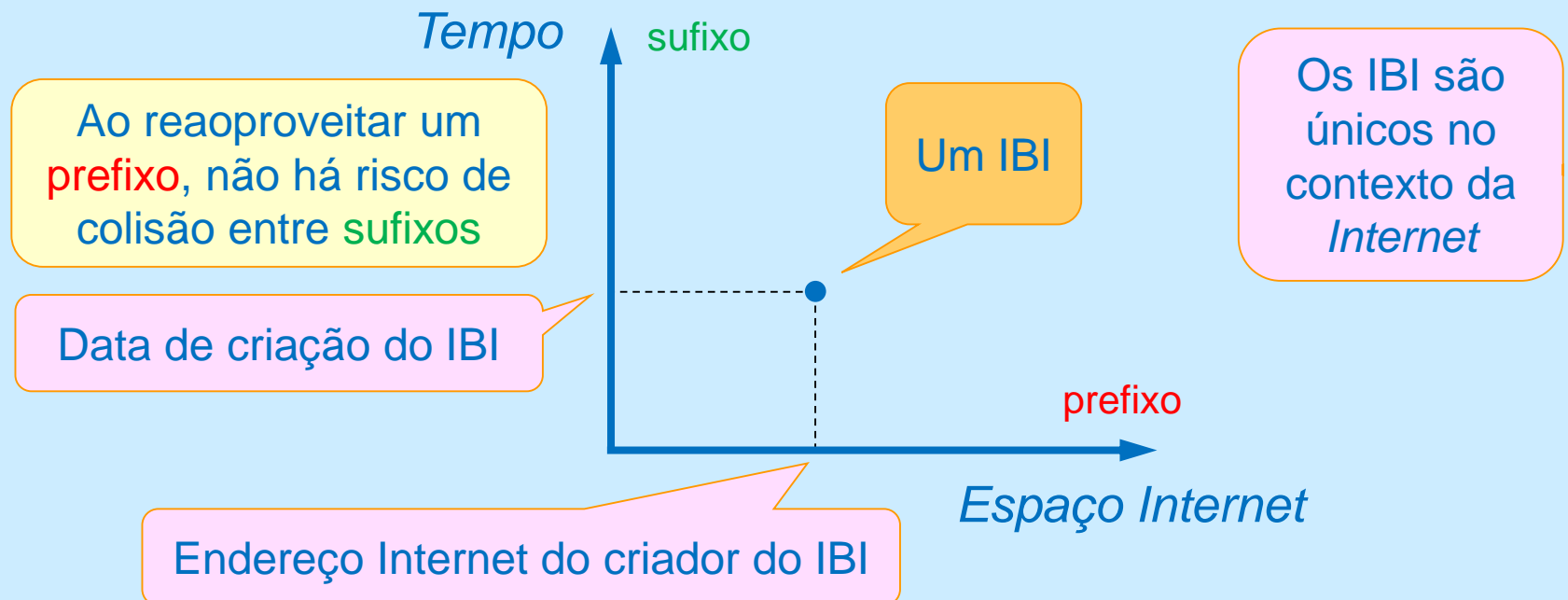


# Sistema de resolução de IBI

(1/5)

*Definição de IBI (Identificador com base na Internet):*

Um IBI é um ponto no espaço-tempo, onde o espaço é o conjunto de todos os possíveis endereços *Internet*.



# Sistema de resolução de IBI

(2/5)

Todo item de informação depositado na plataforma *URLib* recebe **dois** identificadores, ambos denominados IBI.

*Exemplo:*

O primeiro serve no armazenamento do item de informação no sistema de arquivos

`sid.inpe.br/mtc-m18/2012/07.12.18.08`

O segundo é curto e opaco

`8JMKD3MGP8W/3C9EP6P`

ABNT NBR 16066:2012

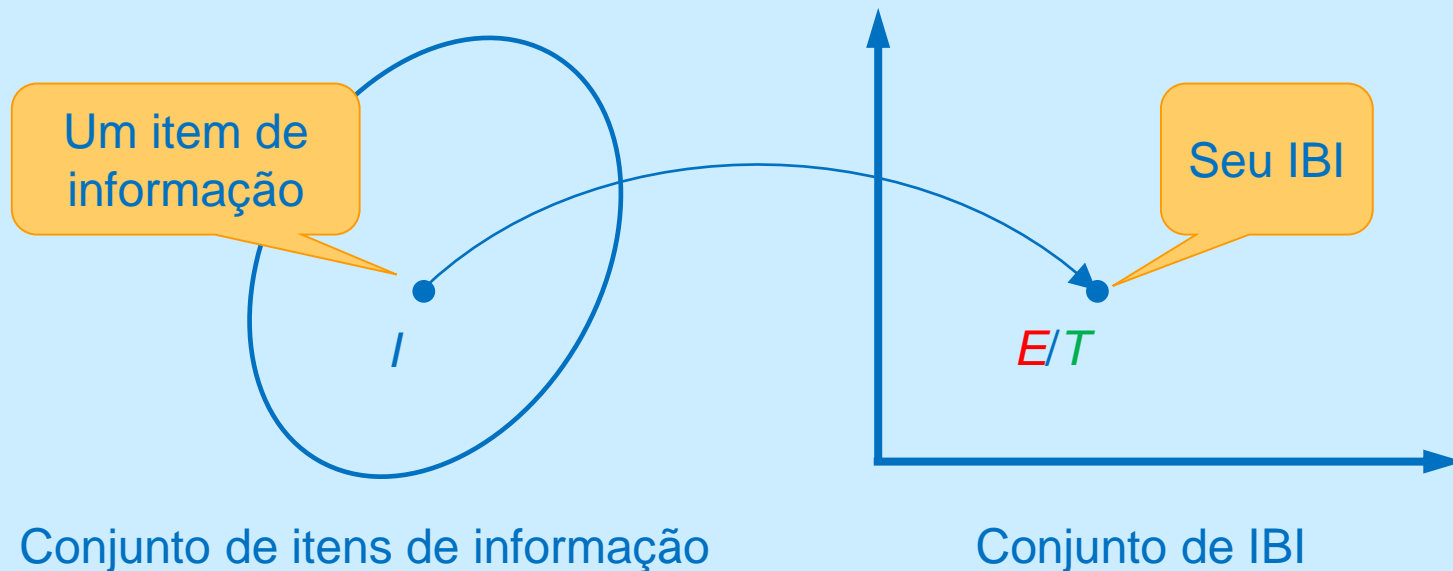
*Sistema para geração de identificador com base na internet (IBI)*

# Sistema de resolução de IBI

(3/5)

*Princípio da identificação por IBI:*

No instante  $T$  do depósito de um item de informação  $I$  num determinado Arquivo de endereço  $E$ , associa-se a  $I$  o IBI  $E/T$ .

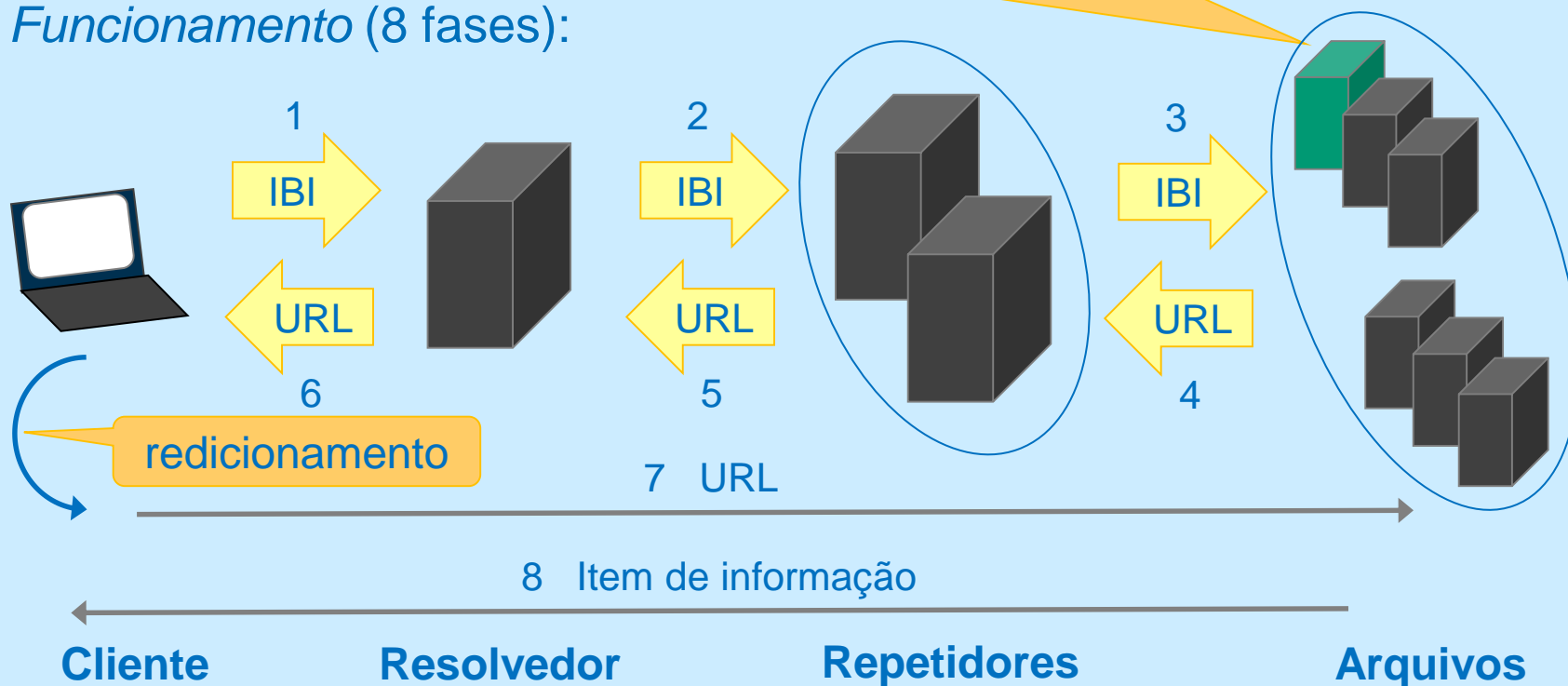


# Sistema de resolução de IBI

(4/5)

O Arquivo que tem o IBI retorna a URL da localização do item de informação (4) e em seguida o próprio item de informação (8)

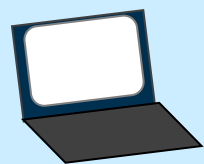
Funcionamento (8 fases):



# Sistema de resolução de IBI

(5/5)

Instalação:



2 atendendo ao INPE  
1 atendendo à SBC

17 Arquivos (13 + 3 + 1)  
(INPE)

Resolverdor **urlib.net**  
(INPE)

3 repetidores  
(INPE)

Cliente

Resolverdor

Repetidores

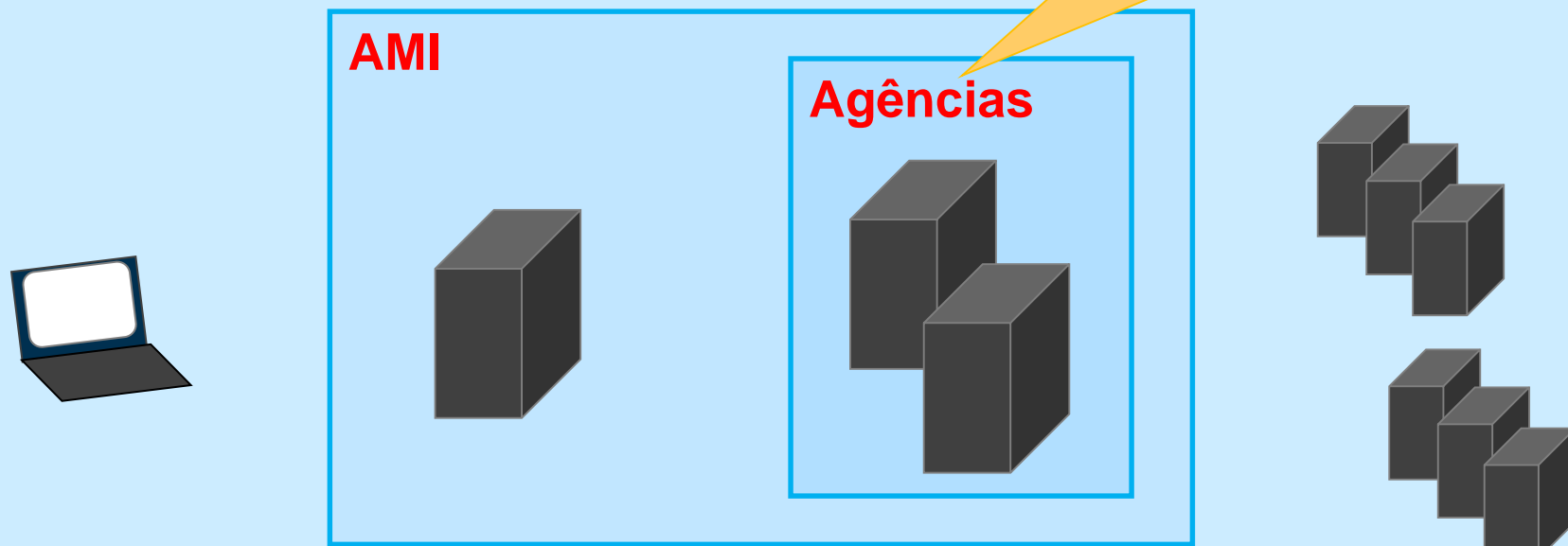
Arquivos

# Proposta de manutenção de IBI

(1/6)

Modelo de sustentabilidade:

As agências de suporte são membros da AMI



AMI: Associação para a Manutenção de IBI (a ser criada)

Cliente

Resolvedor

Repetidores

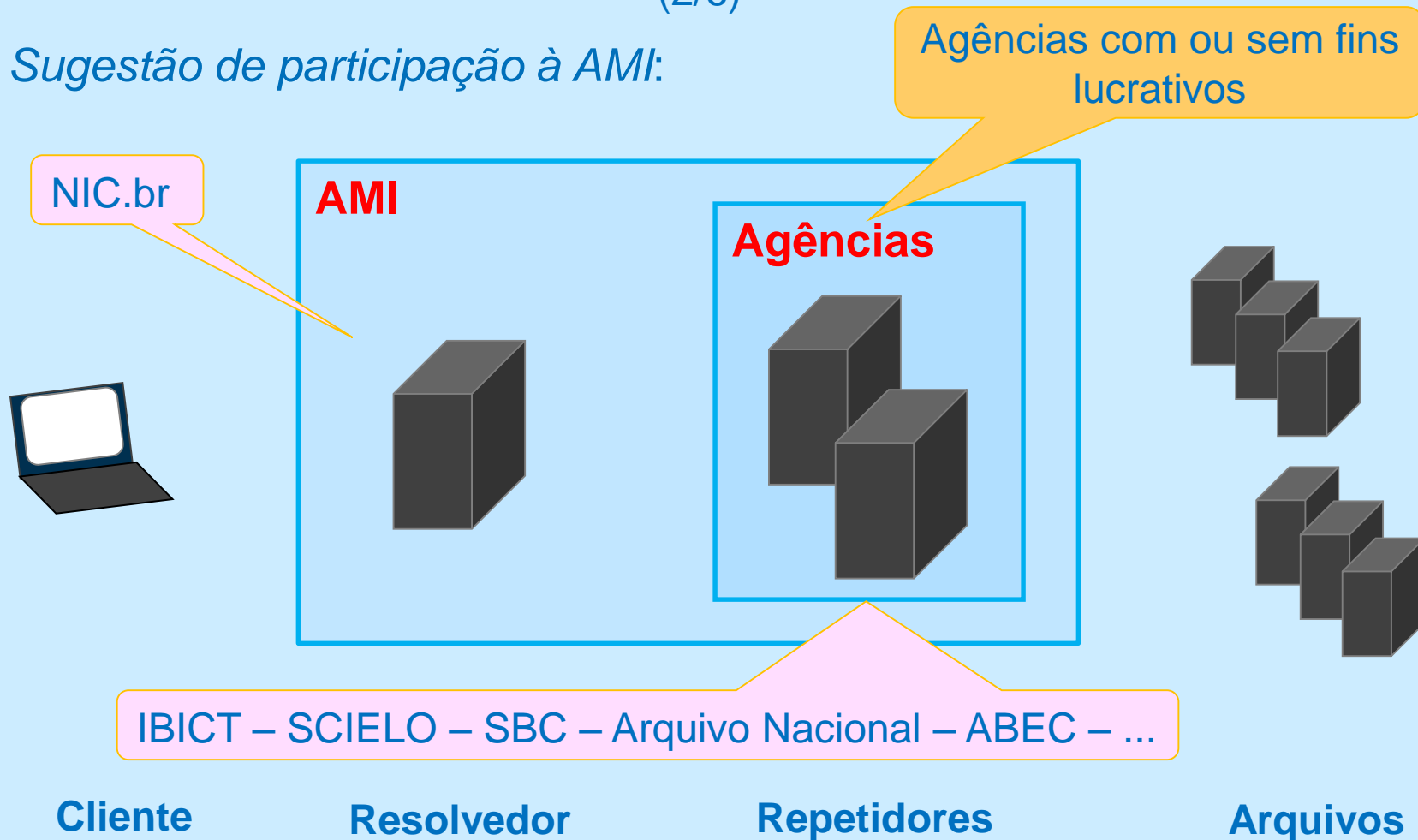
Arquivos



# Proposta de manutenção de IBI

(2/6)

Sugestão de participação à AMI:



## ***Proposta de manutenção de IBI***

(3/6)

*Criação de uma Associação para a Manutenção de IBI (AMI), sem fins lucrativos, com objetivo (entre outros) de:*

Disseminar e promover a geração e o uso correto dos IBI;

Manter as infraestruturas para a resolução dos IBI.

# ***Proposta de manutenção de IBI***

(4/6)

## *Papel das Agências de Suporte:*

Oferecer serviços de suporte aos Arquivos que participem do Sistema de resolução de IBI:

- Receber as solicitações de cadastro de novos Arquivos;
- Fornecer o arquivo .zip personalizado de instalação da plataforma *URLib* ou desenvolver *softwares* compatíveis;
- Treinar os administradores dos Arquivos cadastrados;
- Introduzir o sistema IBI junto à comunidade;

Manter um repetidor do Sistema de resolução de IBI direcionado aos Arquivos cadastrados (opcional);

Atuar como membro da AMI.

## *Proposta de manutenção de IBI*

Segundo Richard WHITT (2017):<sup>(5/6)</sup>

“Digital dark age”

“We’ve GOT to get organized!”<sup>†</sup>

*Hopefully, the foregoing discussion leads to but one inevitable conclusion: **the world must get better organized**, and quickly, to preserve our digital present and future. [...]*

*To be clear, much is happening already. Thousands of dedicated researchers and archivists and academics and government officials and **volunteers are doing what they can**. Institutions large and small are allocating scarce resources to study the intertwined problems and implement changes. And some of the results so far are most impressive. [...]*

*At its root, digital preservation should become **a broad-based social cause**, not limited to any particular set of players.*

<sup>†</sup> The Russians Are Coming, the Russians Are Coming. THE MIRISCH CORP.

## *Proposta de manutenção de IBI*

Segundo Richard WHITT (2017):<sup>(6/6)</sup>

“Era da escuridão digital”

“Temos que nos organizar!”<sup>†</sup>

Esperemos que a discussão anterior leve a uma conclusão inevitável: **o mundo deve se organizar melhor e rapidamente**, para preservar nosso presente e futuro digital. [...]

Para ser sincero, muita coisa já está acontecendo. Milhares de pesquisadores, arquivistas, acadêmicos, funcionários e **voluntários dedicados estão fazendo o que podem**. Instituições grandes e pequenas estão alocando recursos escassos para estudar os problemas interligados e implementar mudanças. E alguns dos resultados desde já são bem impressionantes. [...]

Na sua raiz, a preservação digital deve se tornar **uma causa social de base ampla**, não se limitando a um grupo particular de atores.

<sup>†</sup> The Russians Are Coming, the Russians Are Coming. THE MIRISCH CORP.

Acesso persistente

**Obrigado!**

## **Referências**

(1/2)

BANON, G. J. F. **Auditoria interna sobre o repositório digital da Memória Científica do INPE**. São José dos Campos: INPE, 2010. 60 p. (dpi.inpe.br/banon-pc3@80/2009/11.10.13.03-RPQ). Disponível em: <<http://urlib.net/J8LNKAN8PW/36CT2G2>>.

BANON, G. J. F. **Identificador com base na Internet (IBI): Sistema de identificação**. São José dos Campos: INPE, 2011. 39 p. (iconet.com.br/banon/2009/09.09.22.01-RPQ). Disponível em: <<http://urlib.net/LK47B6W/362SFKH>>.

BANON, G. J. F. **Identificador com base na Internet (IBI): Sistema de resolução**. São José dos Campos: INPE, versão: 2016-01-16. 71 p. Disponível em: <<http://urlib.net/J8LNKB5R7W/3G2EKR5>>.

## ***Referências***

(2/2)

CCSDS. Audit and Certification of Trustworthy Digital Repositories. Magenta Book. Issue 1. September 2011. CCSDS 652.0-M-1. Reston: Consultative Committee for Space Data Systems, 2011. 77 p. Disponível em: <<https://public.ccsds.org/Pubs/652x0m1.pdf>>. Acesso em: 02 abr. 2017.

WHITT, R. S. "Through A Glass, Darkly" Technical, Policy, and Financial Actions to Avert the Coming Digital Dark Ages. Santa Clara High Tech. Law Journal, v. 33, n. 2, p. 117-229, 2017. Disponível em: <<http://digitalcommons.law.scu.edu/chtlj/vol33/iss2/1>>. Acesso em: 02 abr. 2017.