

# Curso de Preservação Digital

## Sistema LOCKSS

Arthur Heleno Lima R de Souza

*Engenheiro de Computação*



Ministério da  
**Ciência, Tecnologia  
e Inovação**



# OBJETIVO DO CURSO

Capacitar profissionais no entendimento da ferramenta de Preservação Digital Distribuída **LOCKSS**, solução tecnológica utilizada na **Rede Cariniana**.

Compreender as **funcionalidades** mais importantes, **procedimentos** e o **gerenciamento** da ferramenta.



# SUMÁRIO

## ➤ Sistema LOCKSS

- Conceitos
- Características técnicas;

## ➤ Caixa LOCKSS;

## ➤ Props Server;

## ➤ Rede Privada LOCKSS (PLN)

- Conceitos
- Arquitetura distribuída
- Como criar uma PLN

## ➤ Processos de Preservação Digital do LOCKSS

- Ingestão
- Teste de Integridade
- Acesso
- Gerência



# CARACTERÍSTICAS

- ▶ Lot Of Copies Keep Stuff Safe(**LOCKSS**);
- ▶ Desenvolvido pela **Stanford University**;
- ▶ Arquitetura distribuída;
- ▶ Segue o modelo **OAIS**.

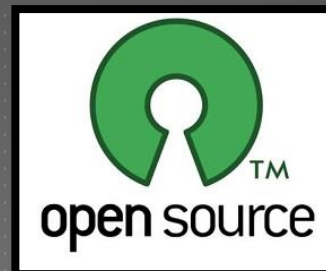


# SISTEMA LOCKSS



## Requisitos mínimos do Sistema LOCKSS

Processamento	Intel x86 (ou compatível), 1 GHZ, 1 núcleo
Memória	1 GB
Espaço em disco	4 TeraByte+ (Cariniana)

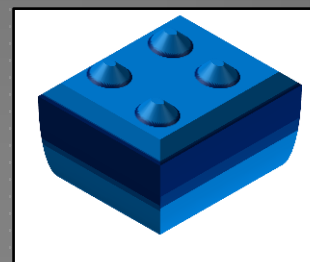


**Código Aberto**

Manutenção de baixo custo (contribuições da comunidade)

Customizações

Interoperabilidade



**Integração**

Linguagem Java

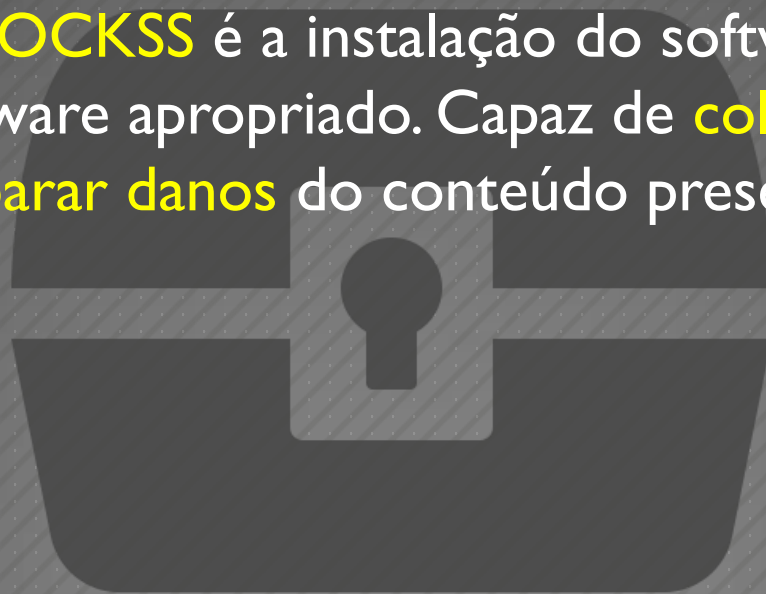
Modular

Extensão



# CAIXA LOCKSS

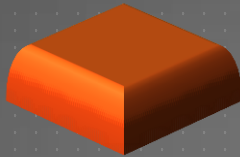
Uma **caixa LOCKSS** é a instalação do software LOCKSS em um hardware apropriado. Capaz de **coletar, preservar** e **reparar danos** do conteúdo preservado.



# CAIXA LOCKSS



=



Daemon

+



Interface

+

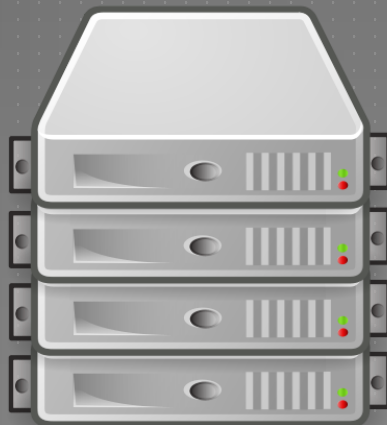


Conteúdo



# PROPS SERVER

É um servidor web de **apoio** às caixas LOCKSS.  
Fornece dados nos formatos **.XML** e **.JAVA**,  
contendo informações sobre os nós da rede, dados  
sobre a preservação e atualizações.





# PROPS SERVER

## Base de dados de Metadados

- [METS] Metadata Encoding and Transmission Standard

## Base de dados de Títulos

- [XML]
- Definições dos objetos preserváveis
- Lista dos endereços IP

## Props Server

## Repositório de Plugins

- [JAVA -XML]
- Plugins

## Gerência do cache

- ...

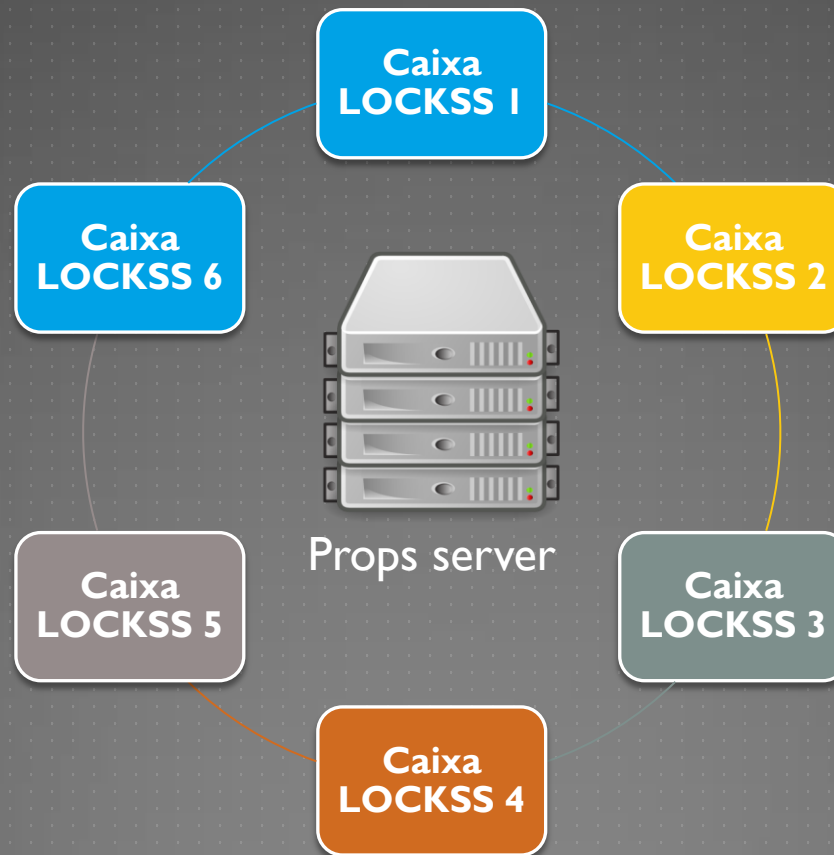


# PRIVATE LOCKSS NETWORK (PLN)

- ▶ Rede Privada LOCKSS para preservação do patrimônio cultural;
- ▶ Uma rede de comunicação entre caixas LOCKSS com capacidade de buscar dados no servidor de suporte (props server)
- ▶ Estudos apontam que cada rede deve ter no mínimo 6 nós (parceiros);
- ▶ Geograficamente dispersas;
- ▶ Comunica-se via P2P, utilizando a rede mundial de computadores (internet).



# REDE PRIVADA LOCKSS



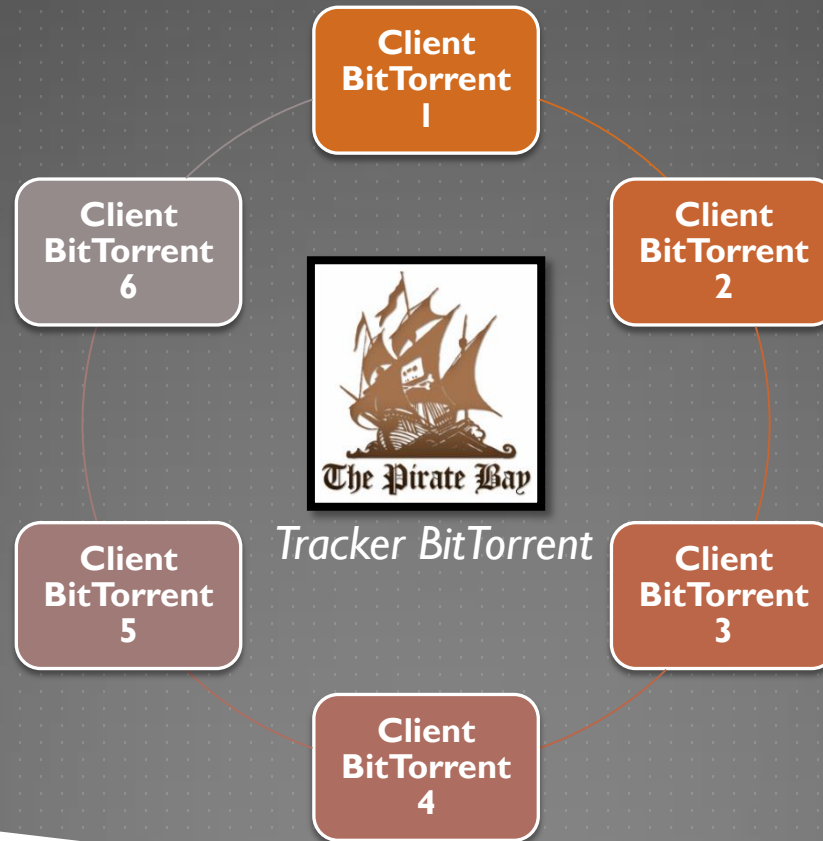
???

Mas desta forma, o sistema não se caracterizaria como um sistema **centralizado**?

○ **Props server** age apenas como um **repositório de informações** para que não precise modificar esses dados em cada caixa manualmente. Não compete a ele nenhum tipo de **processamento de dados**.



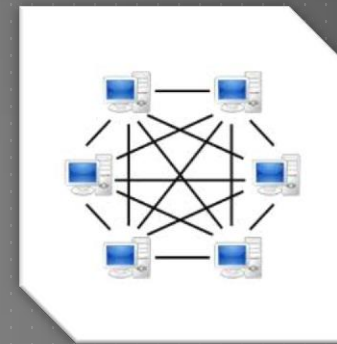
# ANALOGIA: SISTEMAS DISTRIBUÍDOS [BITTORRENT]



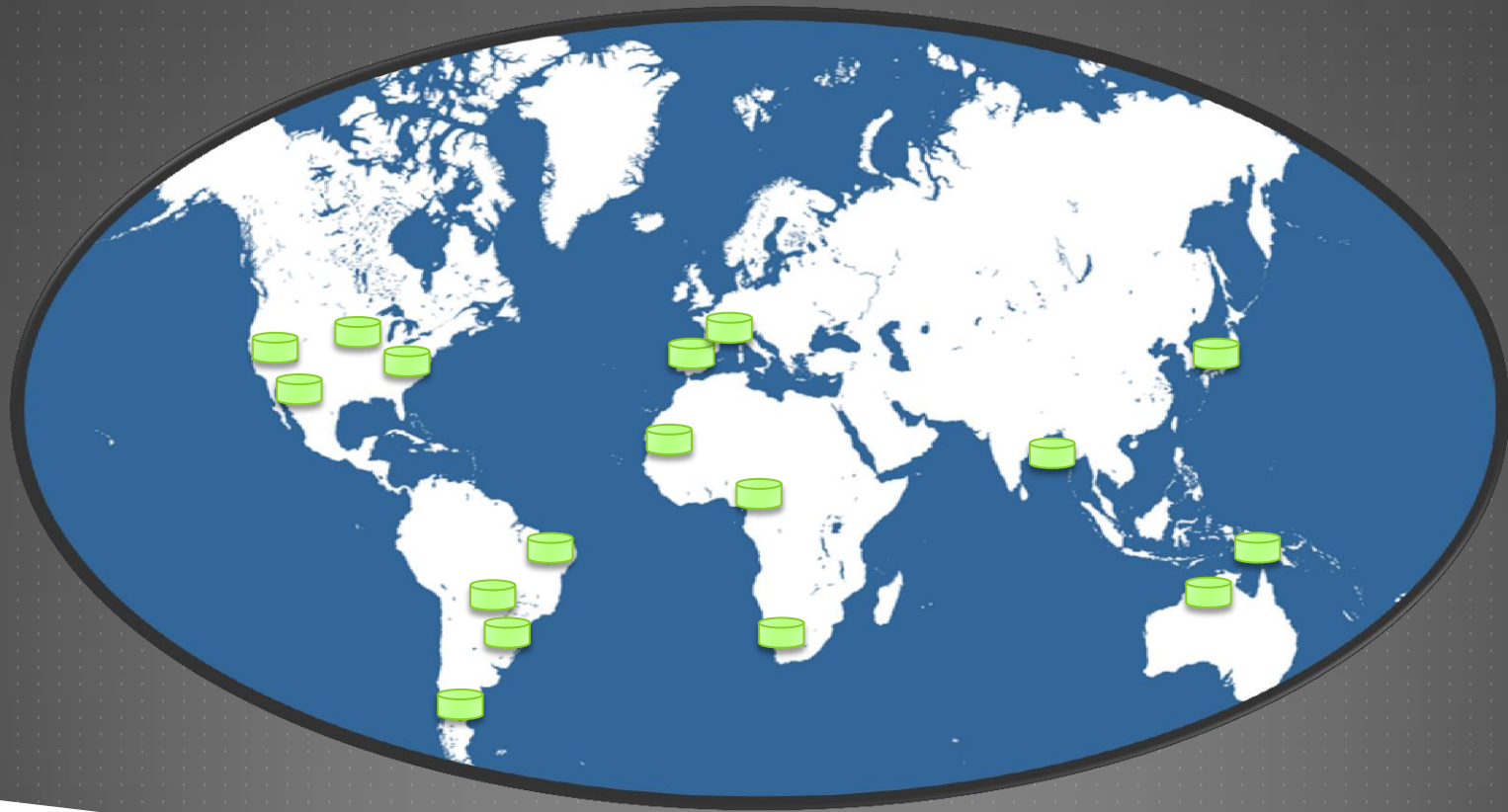
# P2P

**Peer-to-peer** é uma arquitetura de redes de computadores onde cada um dos pontos ou nós da rede funciona tanto como cliente quanto como servidor, permitindo compartilhamentos de serviços e dados **sem a necessidade de um servidor central de processamento**.

Segundo **Tanenbaum & Steen**, uma das arquiteturas para a **construção de sistemas distribuídos** é a arquitetura **P2P** e sua principal característica é a **descentralização**.



# ARQUITETURA DISTRIBUÍDA – CÓPIAS GEOGRAFICAMENTE DISPERSAS



# CRIANDO UMA REDE PRIVADA LOCKSS

**Passo 1:** Instalar uma caixa LOCKSS.





# CRIANDO UMA REDE PRIVADA LOCKSS

**Passo 2** : Configurar a caixa e apontar para um Props Server



```
Configuration URL: [http://props.lockss.org:8001/daemon/lockss.xml]
```



# CRIANDO UMA REDE PRIVADA LOCKSS

**Passo 3** : Configurar acesso e rede

P2P: 9729 [LCAP]

Interface: 8081

Proxy: 8080



# CRIANDO UMA REDE PRIVADA LOCKSS

## Passo 4 : Gerencie as caixas LOCKSS via uma interface



<http://caixalockss.instituicao.br/8081>



- *Gerenciamento e configuração*
- *Intuitivo*
- *Acesso remoto e seguro*



# PROCESSOS



## Ingestão

- Conteúdo
- Metadados



## Armazenamento

- Autenticidade
- Integridade



## Gerência

- Interface Gráfica
- Integração

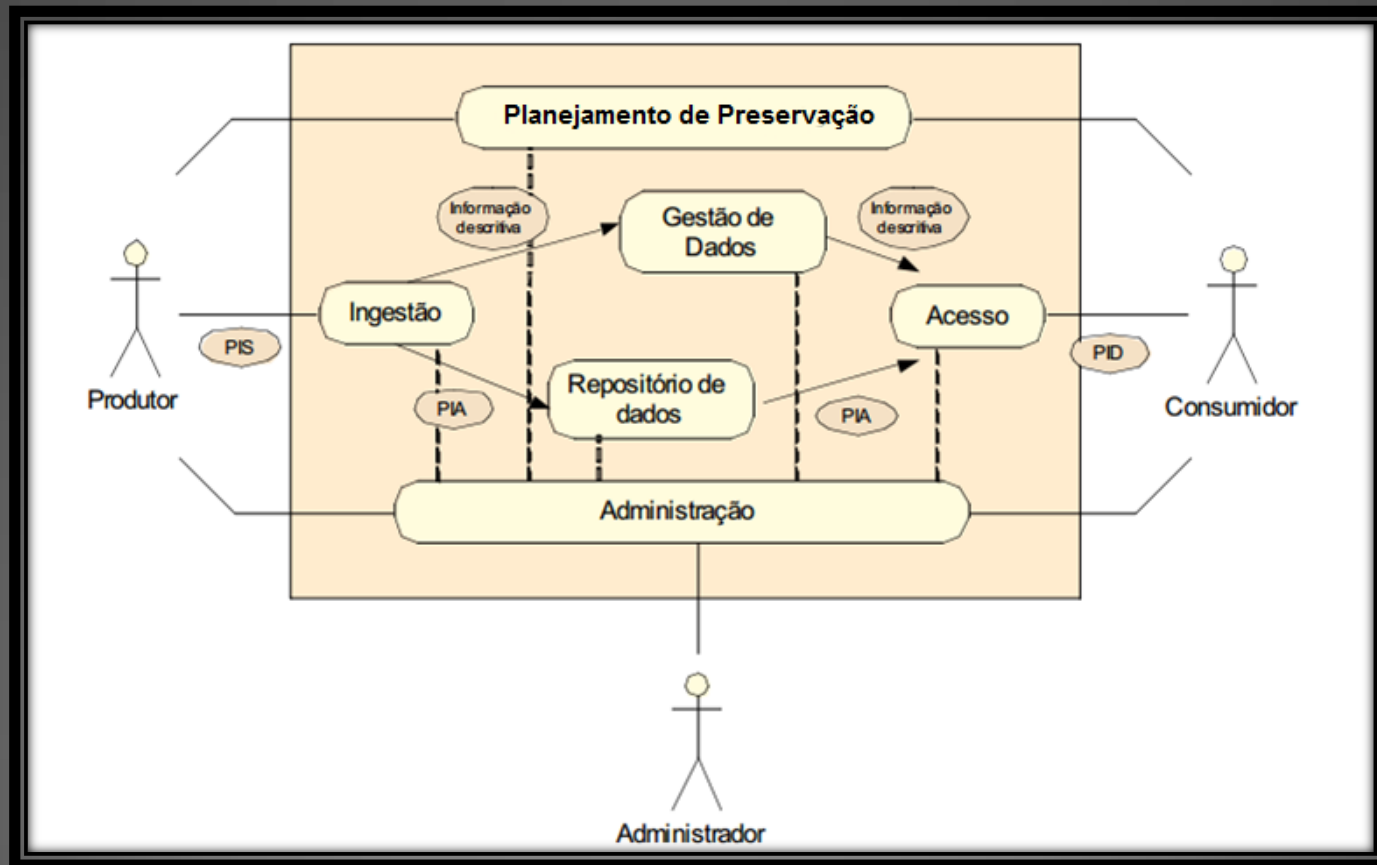


## Acesso

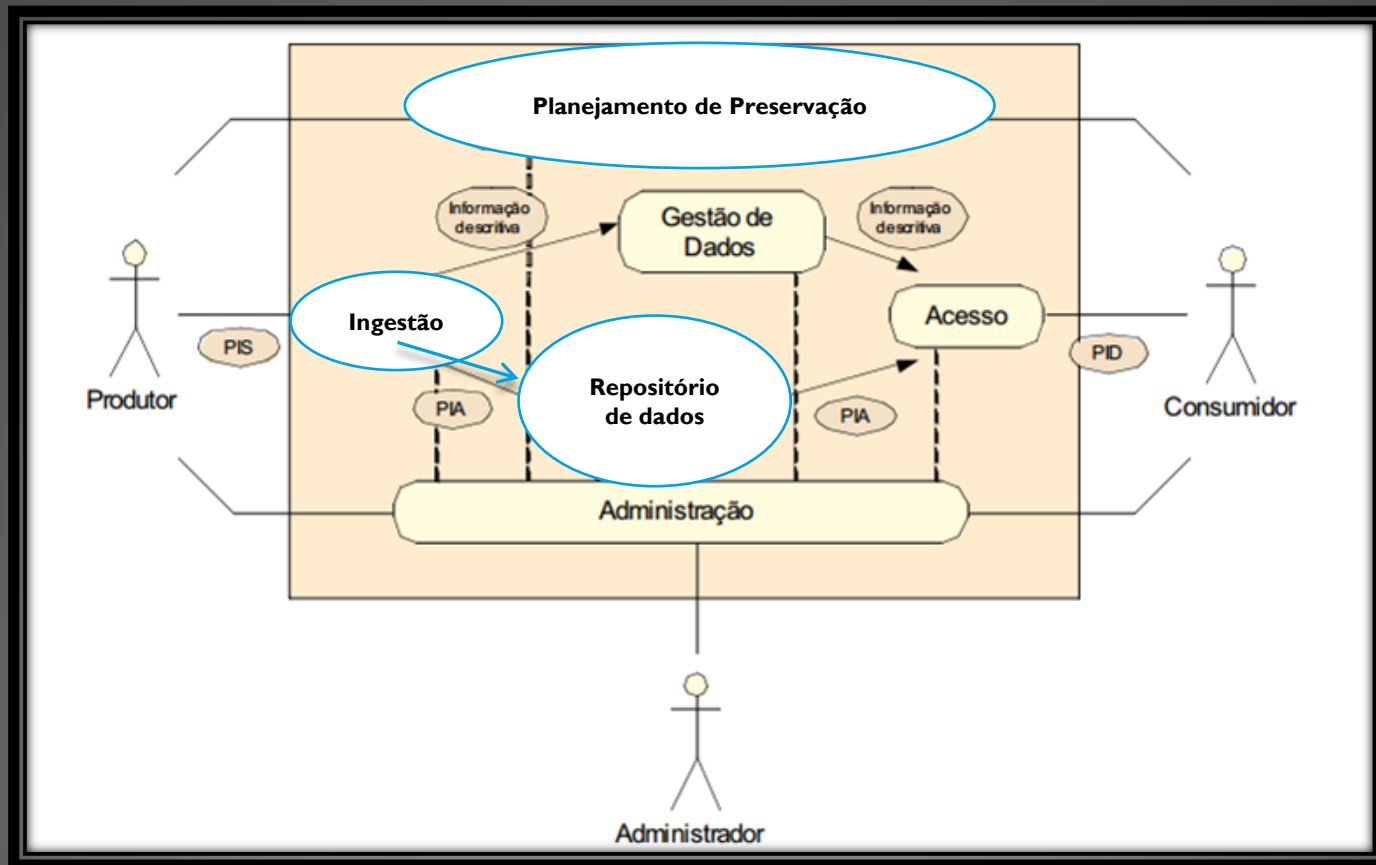
- Proxy
- Migração



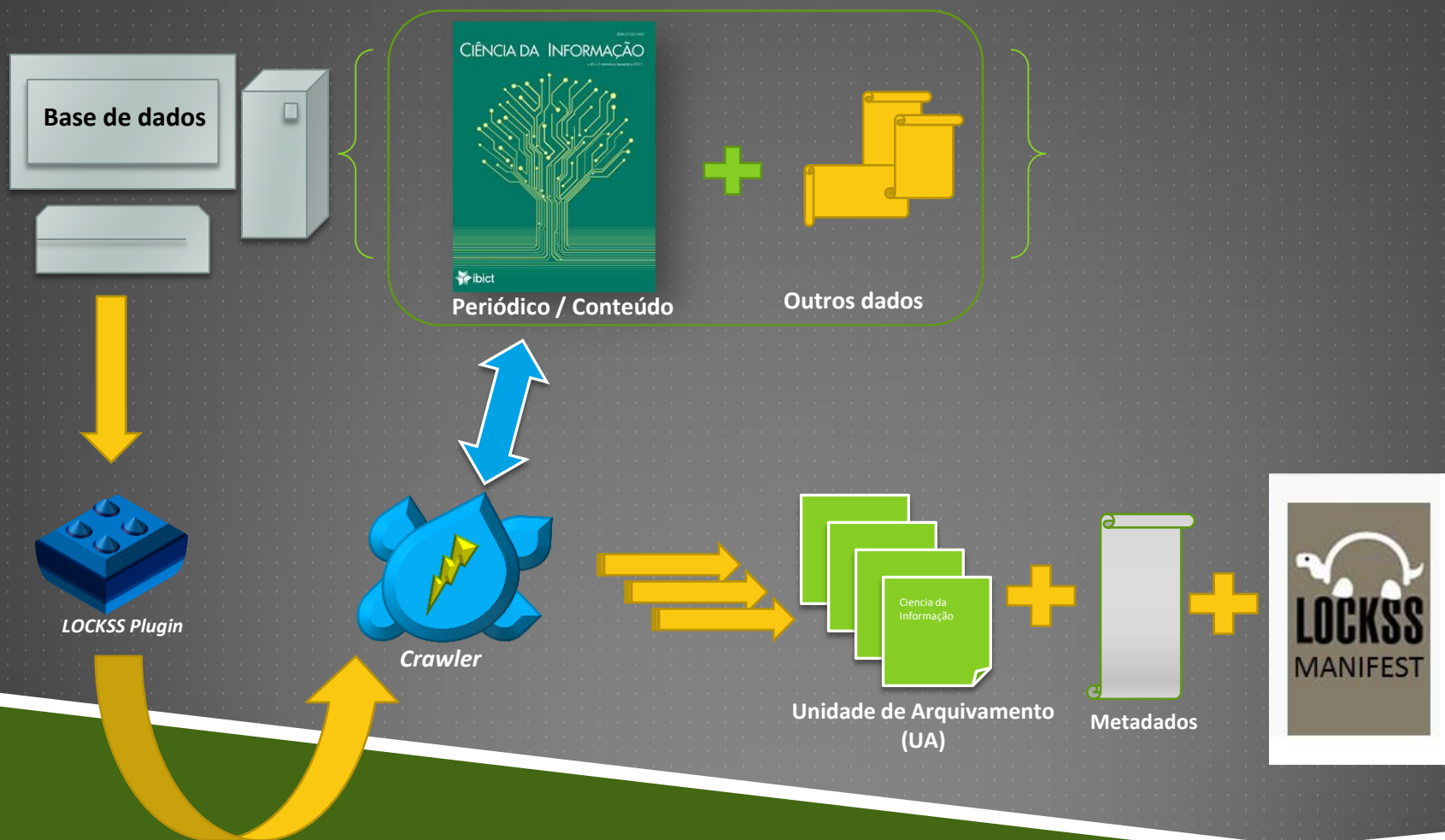
# DIAGRAMA OAIS



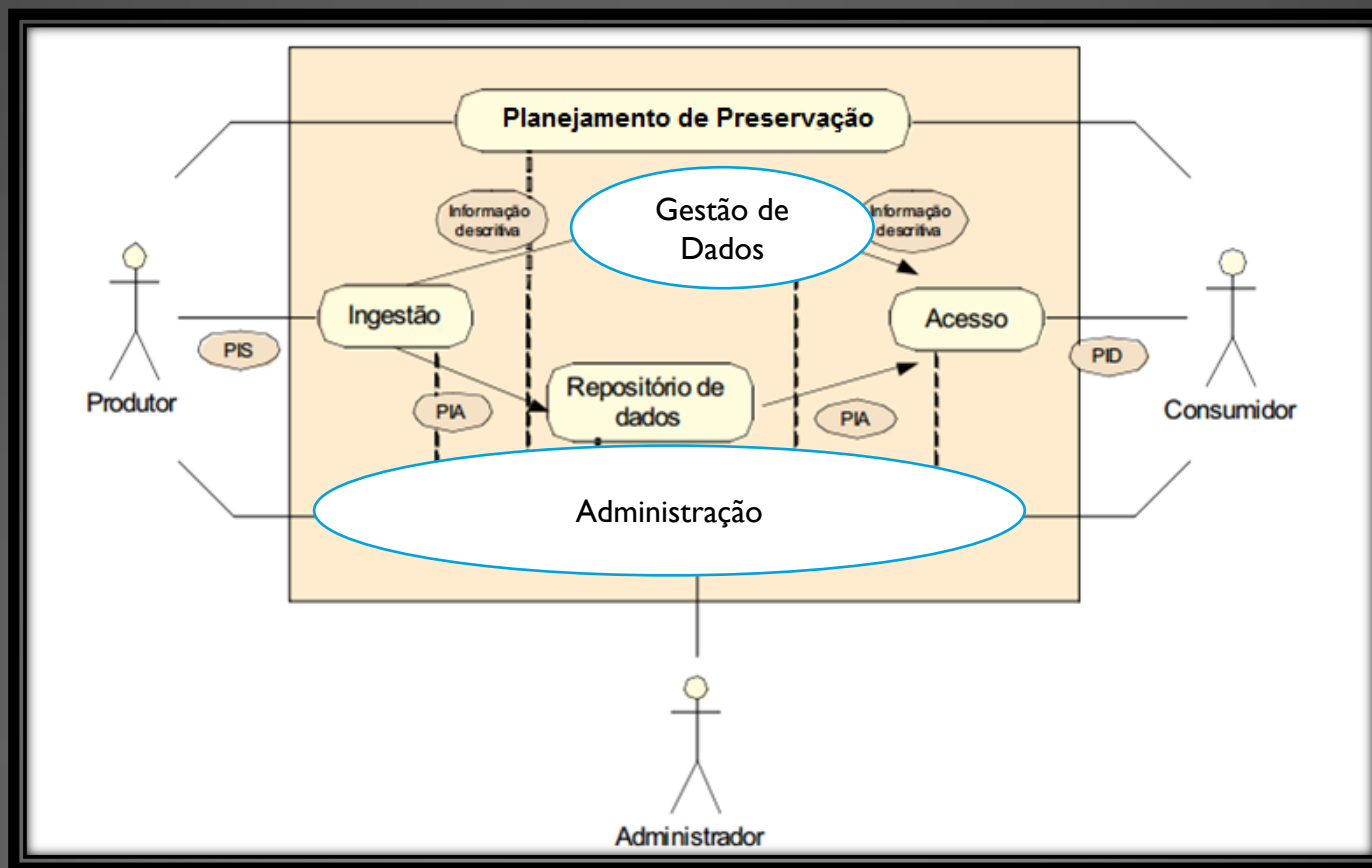
# HARVEST & INGEST



# INGESTÃO (PERIÓDICOS SEER/OJS)



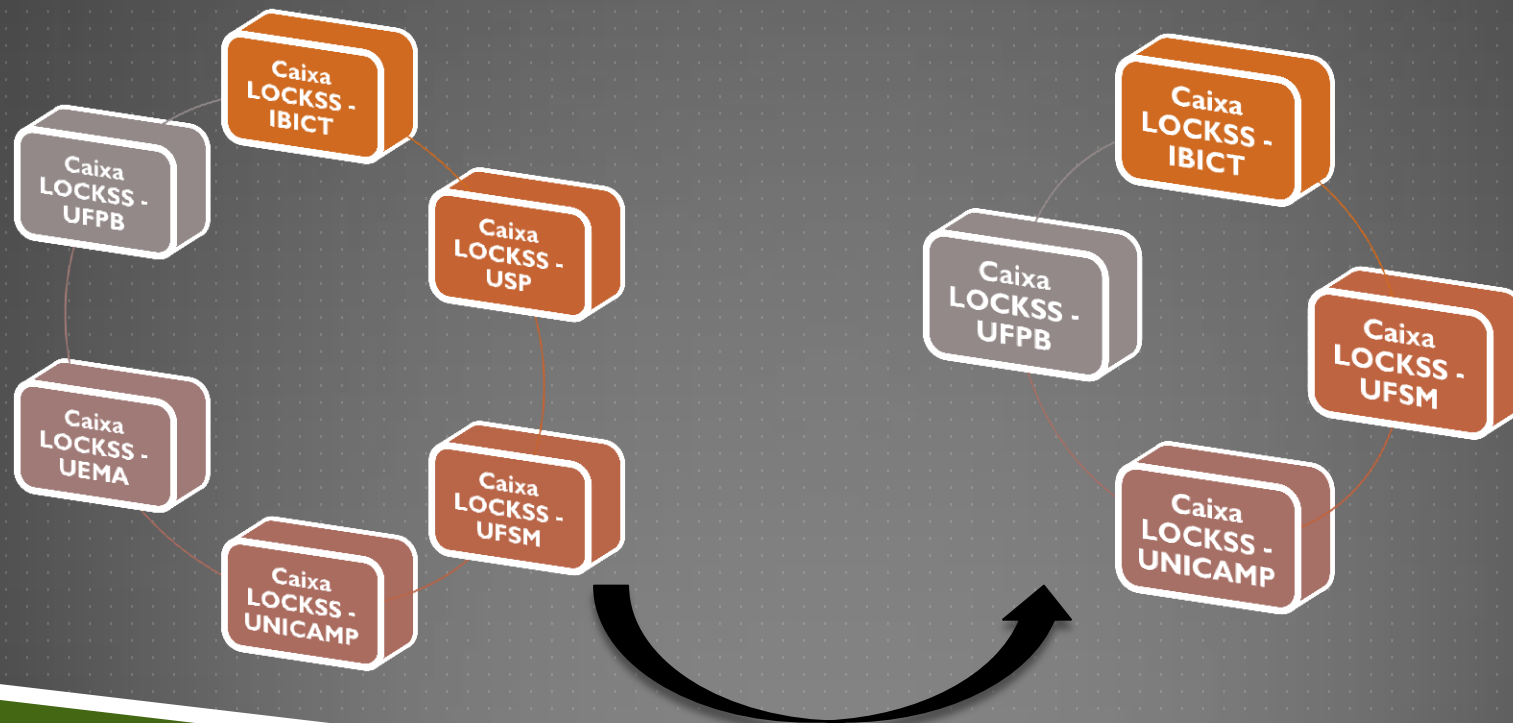
# TESTE DE INTEGRIDADE





# PRESERVAÇÃO – TESTE DE INTEGRIDADE

## Votes

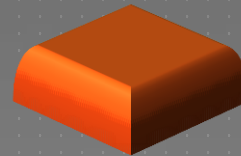


# PRESERVAÇÃO – TESTE DE INTEGRIDADE

## Identificação



Conteúdo preservado



**DAEMON**



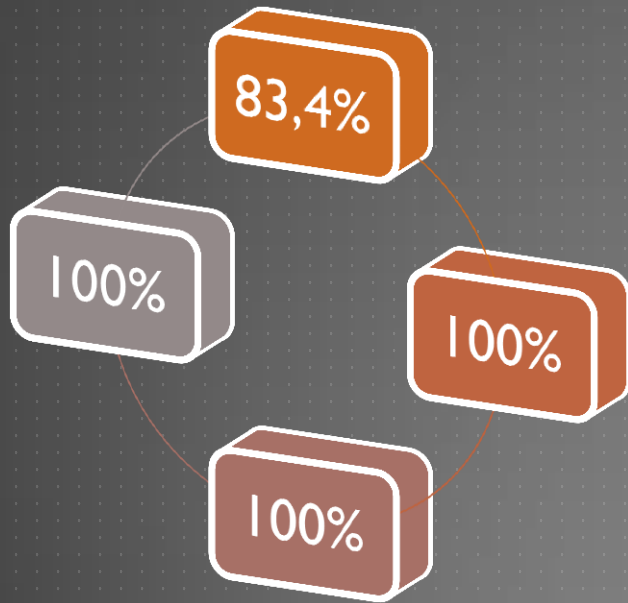
```
8D77GFDGH77SD4  
4SDF66SDG9HJ7SD  
56FS6DF4SD4G7H8  
8SDF66DF6DF7S5D  
F5SDF4SD4F79DF8
```

*Hash*

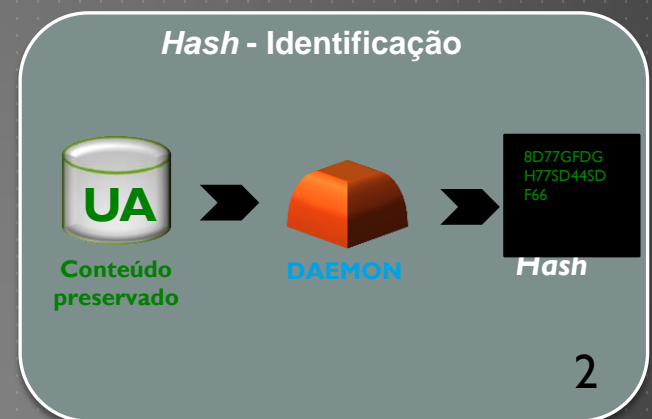
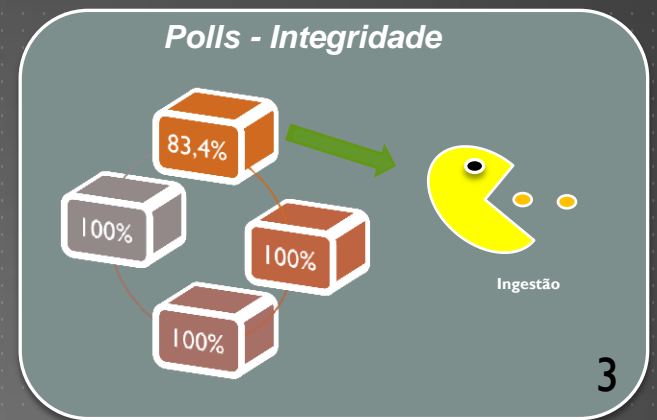
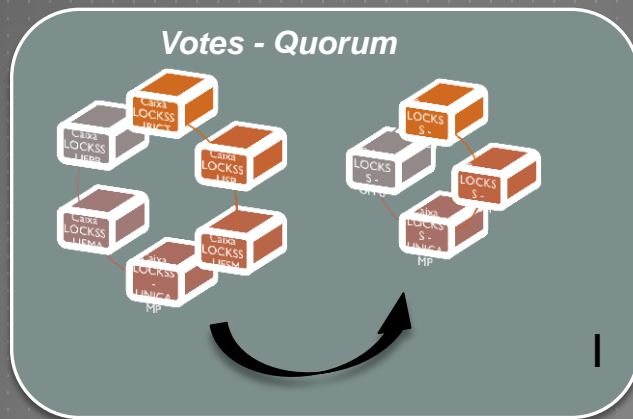
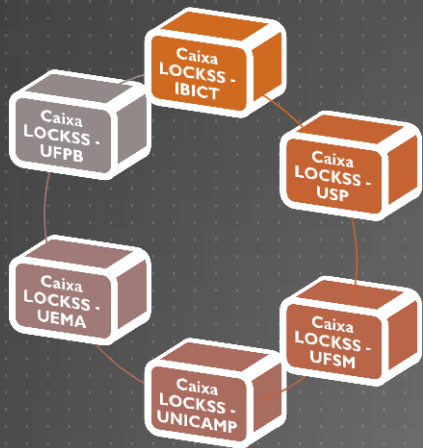


# PRESERVAÇÃO – TESTE DE INTEGRIDADE

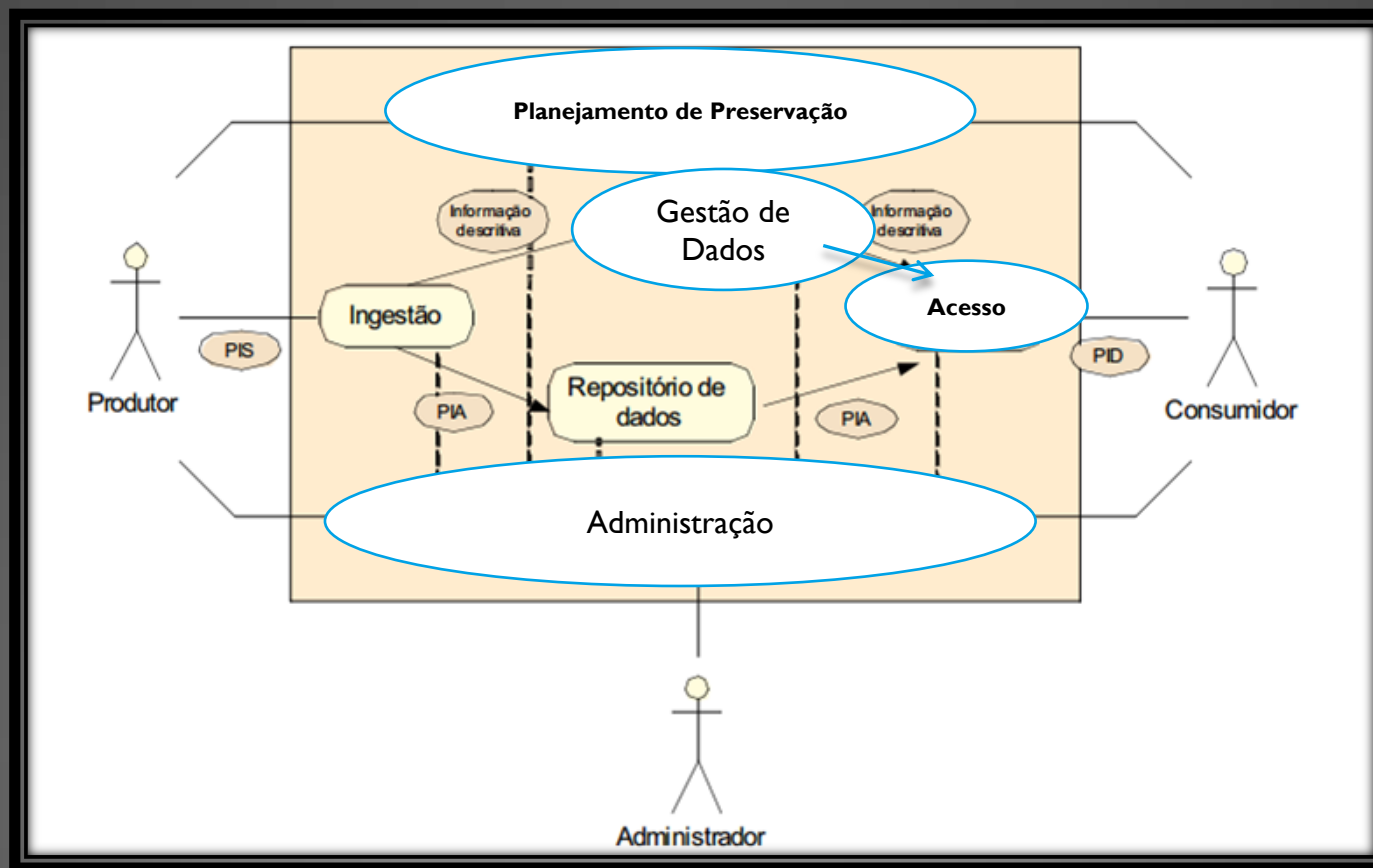
Teste de integridade



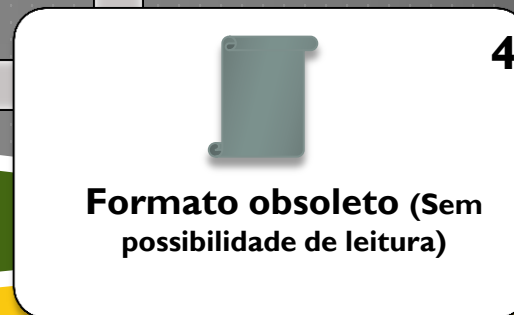
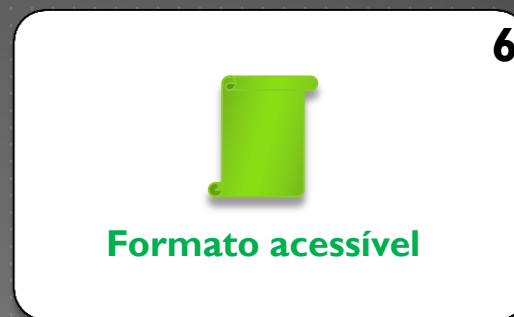
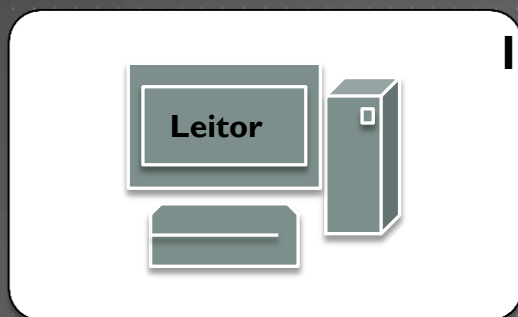
# PRESERVAÇÃO



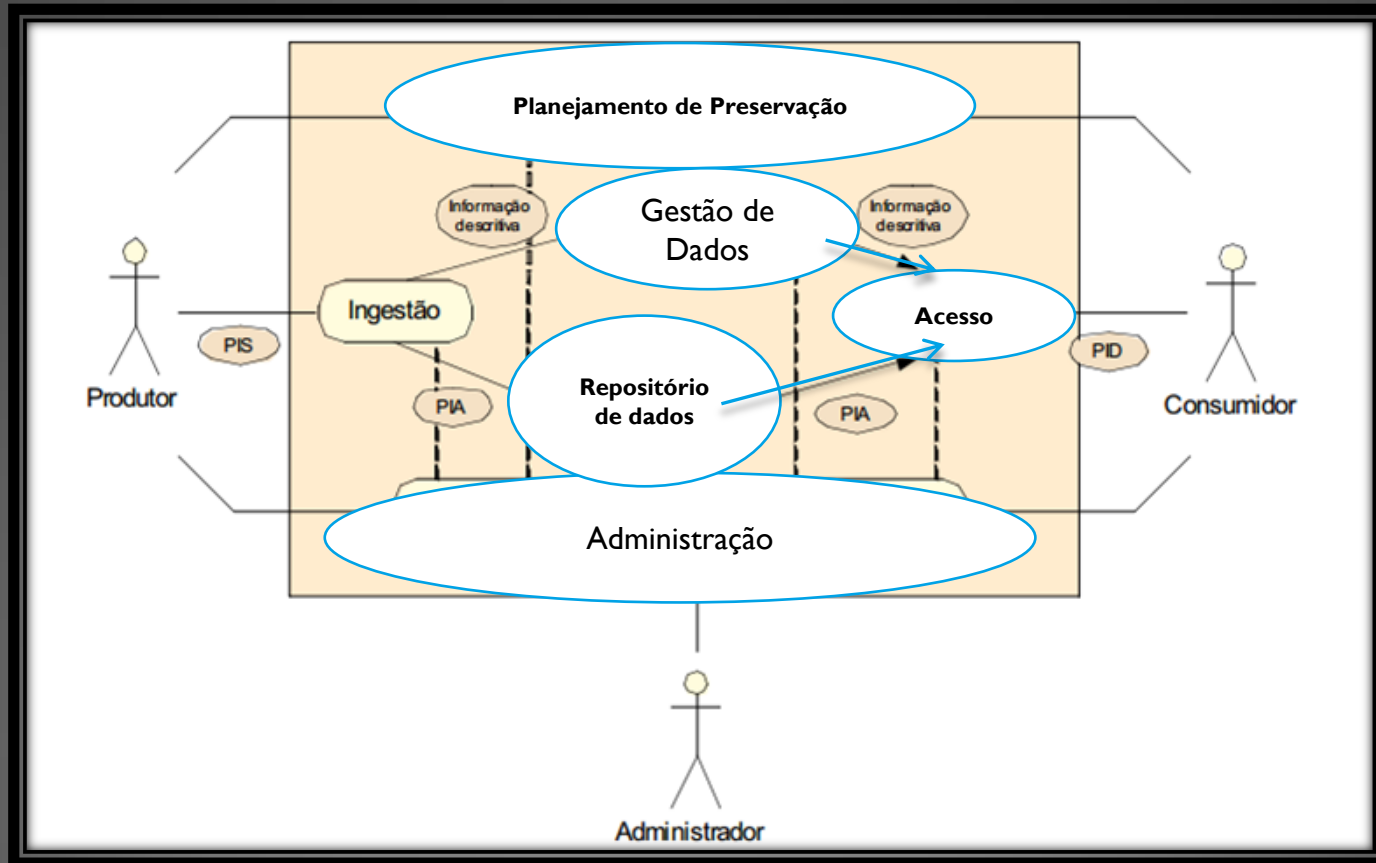
# ACESSO



# ACESSO - “MIGRATION ON ACCESS”



# GERÊNCIA



# GERÊNCIA - INTERFACE GRÁFICA

Welcome to the administration page for LOCKSS box **lockss.ibict.br**.

<a href="#">Journal Configuration</a>	Add or remove titles from this LOCKSS box
<a href="#">Manual Journal Configuration</a>	Manually edit single AU configuration
<a href="#">Admin Access Control</a>	Control access to the administrative UI
<a href="#">Content Access Control</a>	Control access to the preserved content
<a href="#">Content Access Options</a>	Configure the audit proxy and the ICP server
<a href="#">Proxy Info</a>	Info for configuring browsers and proxies to access preserved content on this LOCKSS box
<a href="#">Daemon Status</a>	Status of LOCKSS box contents and operation
<a href="#">Help</a>	Online help, FAQs, credits

[Journal Configuration](#)  
[Admin Access Control](#)  
[Content Access Control](#)  
[Content Access Options](#)  
[Proxy Info](#)  
[Daemon Status](#)  
[Debug Panel](#)  
[Expert Config](#)  
[Logs](#)  
[Thread Dump](#)  
[Contact Us](#)  
[Help](#)

**L O T S   O F   C O P I E S   K E E P   S T U F F   S A F E <sup>TM</sup>**

Daemon 1.59.2 built 16-Jan-13 22:12:44 on build7.lockss.org, Linux RPM 1





# GERÊNCIA – ADICIONANDO TÍTULOS

Add Titles

Add one or more groups of titles

~~Remove Titles~~

Remove selected titles

Deactivate Titles

Deactivate selected titles

Reactivate Titles

Reactivate selected titles

Backup

Backup cache config to a file on your workstation

Restore

Restore cache config from a file on your workstation

Manual Add/Edit

Add, Edit or Delete an individual AU

**L O T S   O F   C O P I E S   K E E P   S T U F F   S A F E <sup>TM</sup>**

Daemon 1.59.2 built 16-Jan-13 22:12:44 on build7.lockss.org, Linux RPM 1



# GERÊNCIA - SEGURANÇA

Allow Access <sup>1</sup>	Deny Access <sup>1</sup>
██████████.0/24	
██████████.0/26	
██████████.254	
██████████2.234	
██████████.184	

Update



# GERÊNCIA – TRANSFERÊNCIA

The screenshot displays the LOCKSS configuration interface, divided into several sections:

- Content Server Options:** Configure how content is collected from publishers and served to users. Includes links for [Content Server Options](#) (Configure content servers and proxies and ICP), [Proxy Client Options](#) (Configure the crawler to access sites through a proxy), and [Export Content](#) (Export preserved content as Zip, WARC, etc.).
- Proxy Client Options:** Configure the LOCKSS crawler to access the net through a proxy server. Includes a checkbox for  Proxy crawls, an input field for HTTP Proxy host (10.0.0.1), and a Port field (80). Buttons include [Update Proxy Client](#) and [Back to Content Access Options](#).
- Content Servers and Proxies:** Manage this box's content servers and proxies.<sup>1</sup> Includes checkboxes for:
  - Enable content server<sup>2</sup> on port 80
  - Enable content proxy<sup>3</sup> on port 8080
  - Enable audit proxy<sup>4</sup> on port 8082
  - Enable ICP server<sup>5</sup> on port [ ]Buttons include [Update Content Servers](#) and [Back to Content Access Options](#).
- Export Options:** Select AU [ ]
  - Export file type<sup>1</sup>: WARC (content only)
  - Compress:
  - Exclude Directories<sup>2</sup>:
  - Translate filenames<sup>3</sup>:  None  Windows  MacOS
  - Export file prefix<sup>4</sup>: [ ]
  - Max export file size (MB)<sup>5</sup>: [ ]
  - Max content file versions<sup>6</sup>: 1
  - [Create Export File\(s\)](#)

LOTS OF COPIES KEEP STUFF SAFE™  
Daemon 1.59.2 built 16-Jan-13 22:12:44 on build7.lockss.org, Linux RPM 1



# GERÊNCIA – TRANSFERÊNCIA



## Serve Content

Serve Content

[Title List](#)

[Help](#)

### Archival Unit

Accounting History Volume 15

Accounting History Volume 16

Al-Mas'q: Islam and the Medieval Mediterranean Volume 23

Asian Journal of Comparative Law Volume 6

Journal of Intellectual Disabilities Volume 13

Medical Humanities Volume 32

Mediterranean Quarterly Volume 19

Mediterranean Quarterly Volume 20

Mediterranean Quarterly Volume 21

### Manifest

[http://ach.sagepub.com/lockss-manifest/vol\\_15\\_manifest.dtl](http://ach.sagepub.com/lockss-manifest/vol_15_manifest.dtl)

[http://ach.sagepub.com/lockss-manifest/vol\\_16\\_manifest.dtl](http://ach.sagepub.com/lockss-manifest/vol_16_manifest.dtl)

<http://www.tandfonline.com/lockss/calm20/23/index.html>

<http://www.bepress.com/asjcl/lockss-volume6.html>

[http://jid.sagepub.com/lockss-manifest/vol\\_13\\_manifest.dtl](http://jid.sagepub.com/lockss-manifest/vol_13_manifest.dtl)

[http://mh.bmj.com/lockss-manifest/vol\\_32\\_manifest.dtl](http://mh.bmj.com/lockss-manifest/vol_32_manifest.dtl)

[http://mq.dukejournals.org/lockss-manifest/vol\\_19\\_manifest.dtl](http://mq.dukejournals.org/lockss-manifest/vol_19_manifest.dtl)

[http://mq.dukejournals.org/lockss-manifest/vol\\_20\\_manifest.dtl](http://mq.dukejournals.org/lockss-manifest/vol_20_manifest.dtl)

[http://mq.dukejournals.org/lockss-manifest/vol\\_21\\_manifest.dtl](http://mq.dukejournals.org/lockss-manifest/vol_21_manifest.dtl)



# TRANSFERÊNCIA (PROXY)

```
// PAC file
// Generated Thu May 02 18:23:04 BRT 2013 by LOCKSS cache lockss.ibict.br

// Generated from /info/ProxyInfo
function FindProxyForURL(url, host) {
  // Educação Volume 2, Educação Volume 4, (14 more)
  if (shExpMatch(url, "http://cascavel.ufsm.br/*"))
  { return "PROXY lockss.ibict.br:8080; DIRECT"; }

  // Perspectivas em Gestão & Conhecimento Volume 1, Perspectivas em Gestão & Conhecimento Volume 2
  if (shExpMatch(url, "http://periodicos.ufpb.br/*"))
  { return "PROXY lockss.ibict.br:8080; DIRECT"; }

  // IBICT PLN Plugin Registry
  if (shExpMatch(url, "http://props.lockss.org:8001/*"))
  { return "PROXY lockss.ibict.br:8080; DIRECT"; }

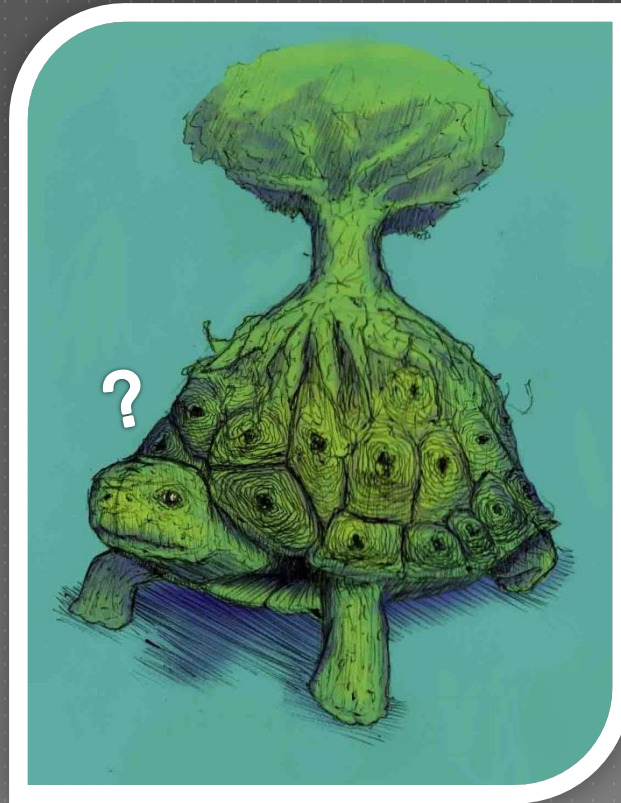
  // Ciência da Informação Volume 1, Ciência da Informação Volume 2, (39 more)
  if (shExpMatch(url, "http://revista.ibict.br/*"))
  { return "PROXY lockss.ibict.br:8080; DIRECT"; }

  // ETD - Educação Temática Digital Volume 1 (1999), ETD - Educação Temática Digital Volume 1-2 (2000), (36 more)
  if (shExpMatch(url, "http://www.fae.unicamp.br/*"))
  { return "PROXY lockss.ibict.br:8080; DIRECT"; }

  return "DIRECT";
}
```



# DÚVIDAS?



# REFERÊNCIAS

K. SKINNER AND M. SCHULTZ - A Guide to Distributed Digital Preservation.,  
Eds. (Atlanta, GA: Educopia Institute, 2010).

LOCKSS TEAM – LOCKSS Support pages. Disponível em: <<http://www.lockss.org/support> > Acesso: 15 nov. 2013.

MARDERO ARELLANO, Miguel Angel. Critérios para a preservação digital da informação científica.  
2008. 354 f. Tese (doutorado) - Universidade de Brasília,  
Departamento de Ciência da Informação, 2008.

Disponível em: <[http://eprints.rclis.org/15412/1/Tese\\_Miguel\\_Angel\\_Mardero\\_Arellano.pdf](http://eprints.rclis.org/15412/1/Tese_Miguel_Angel_Mardero_Arellano.pdf)>. Acesso em: 20 nov. 2013.

REICH, Victoria A. Lots of copies keep stuff safe as a cooperative archiving solution for ejournals.  
Issues in Science and Technology Librarianship, [s.l.], fall 2002.

Disponível em: <<http://www.library.ucsb.edu/istl/02-fall/article1.html>>. Acesso em: 09 out. 2013.

TANENBAUM, A. S., STEEN, M.V. - Distributed Systems: Concepts and Design. 4ª Edição (2005).



**Obrigado!**

The end



Ministério da  
**Ciência, Tecnologia  
e Inovação**

