



Fundação para a Computação Científica Nacional
Foundation for National Scientific Computing

rARC - Preservação da Web através de replicação distribuída em larga escala

André Nogueira



ARQUIVO DA WEB
PORTUGUESA



União Europeia
FEDER

- A Web é a maior fonte de informação construída
 - Jornais, livros, documentação técnica, blogs
 - Informação publicada exclusivamente na Web
- A informação na Web é efémera
 - Após 1 ano, apenas cerca de 20% de um conjunto de endereços da Web ainda apontam para um conteúdo válido

1996

ALTAVISTA

Technology

View Multimedia From Our Vantage Point

[Click here for advertising information - reach millions every month!](#)

Search and Display the Results

Search with Digital's Alta Vista [[Advanced Search](#)] [[Add URL](#)]



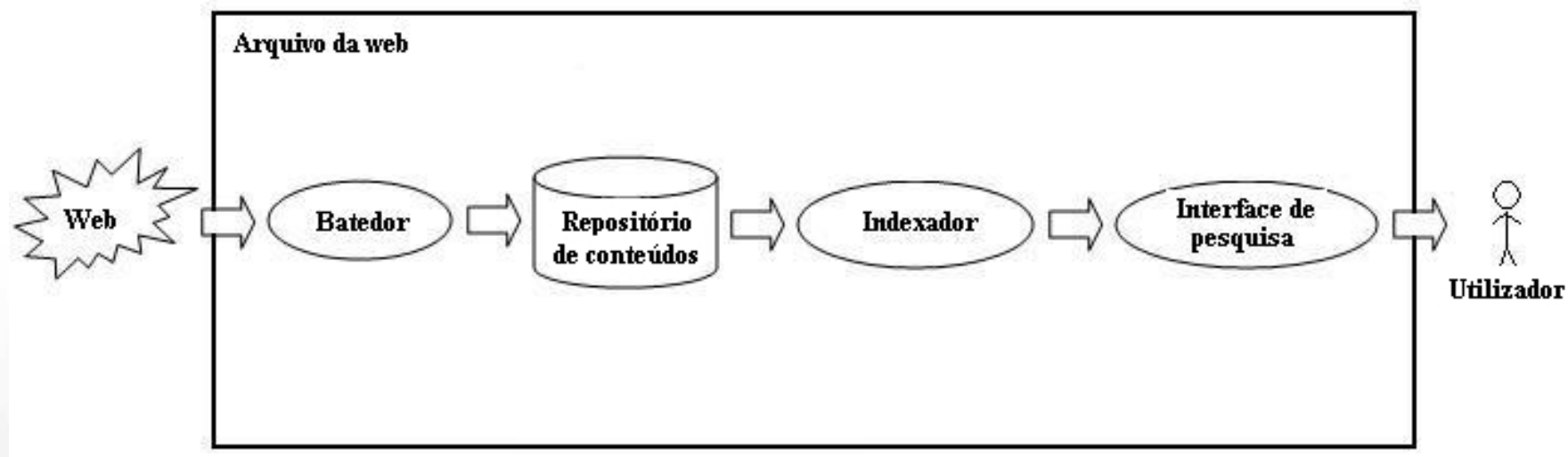
[Download free demo versions of Alta Vista Technology software](#)

ALTAVISTA

[[Creative](#)] [[Search](#)] [[Humor](#)] [[Email](#)]

- Preservação iniciada pelos arquivos da Web
- Primeira iniciativa: Internet Archive
 - Primeira recolha em 1996







fccn

entre 1 / Janeiro / 1996

e 1 / Julho / 2009

Pesquisar

Resultados 1 - 10 de cerca de 163.471

[Portal FCCN - 28 March, 2002](#)

Portal **FCCN** **FCCN** RCTS RCTS2 DNS GigaPix Projectos Localização Software **FCCN** Prod. Informat ... sequência do sucesso das iniciativas anteriores, a **FCCN** e a Universidade de Coimbra, com o impres ... SELF-SERVICE ©2001 **FCCN** - Fundação para a Computação Científica Nacional Todos os direitos ...
<http://www.fccn.pt/> - [outras versões](#)

[Portal FCCN - 20 December, 2001](#)

Portal **FCCN** **FCCN** RCTS Contactos Serviços Projectos Directório Documentos Ligação das EB1 Software Infraestrutura Agenda RCTS2 DNS GigaPix Projectos Localização Software **FCCN** Prod. Informat ... reinserção social; etc. Com base em protocolos estabelecidos com o MCT, a **FCCN** tem dotado estas ...
<http://www.fccn.pt/RCTS/escolas/> - [outras versões](#)

[http://www.fccn.pt/ - 13 October, 1996](#)

Nacional (**FCCN**) Av. Brasil, 101 1799 Lisboa Codex Tel. +351 1 8481906 Fax. +351 1 8472167 email. info@**fccn** ... : Wednesday, 10-Jan-96 18:13:25 CUT Internet URL- <http://www.fccn.pt:80/> ...
<http://www.fccn.pt/> - [outras versões](#)

[Speedmeter FCCN - 19 May, 2007](#)

Speedmeter **FCCN** CONDIÇÕES DE UTILIZAÇÃO O speed meter é uma funcionalidade desenvolvida internamente pela **FCCN** que permite efectuar medições de débito. Este sistema é disponibilizado "as is" não se garantindo a fiabilidade total dos resultados apresentados, pelo que, em caso algum, a **FCCN** será ...
<http://speedmeter.fccn.pt/> - [outras versões](#)

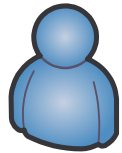
[Speedmeter FCCN - 30 April, 2006](#)

Speedmeter **FCCN** Descrição Técnica A ferramenta Speedmeter aqui disponibilizada pela **FCCN** permite ... um computador e o servidor do speedmeter instalado na sala técnica da **FCCN**. Esta ferramenta baseada ... débito

- Armazenamento centralizado
 - Incêndio ou inundação das instalações
- Cópias de segurança → Aumento dos custos
- Recolha contínua → Aumento do espaço de armazenamento
- Resultados de uma recolha do Arquivo da Web Portuguesa
 - Sítios Web visitados: 450 mil
 - Volume de informação: 3 TBytes
- Volume de informação de todas as recolhas: 10 TBytes

- 4,5 milhões de portugueses usam a Internet
- Actualmente, um PC tem 100 GBytes de capacidade de disco
- Se 4,5 milhões de utilizadores disponibilizarem 10% (10 GBytes) de espaço de disco → Total disponível seria 45 PBytes

Replicador de ficheiros ARC (rARC)



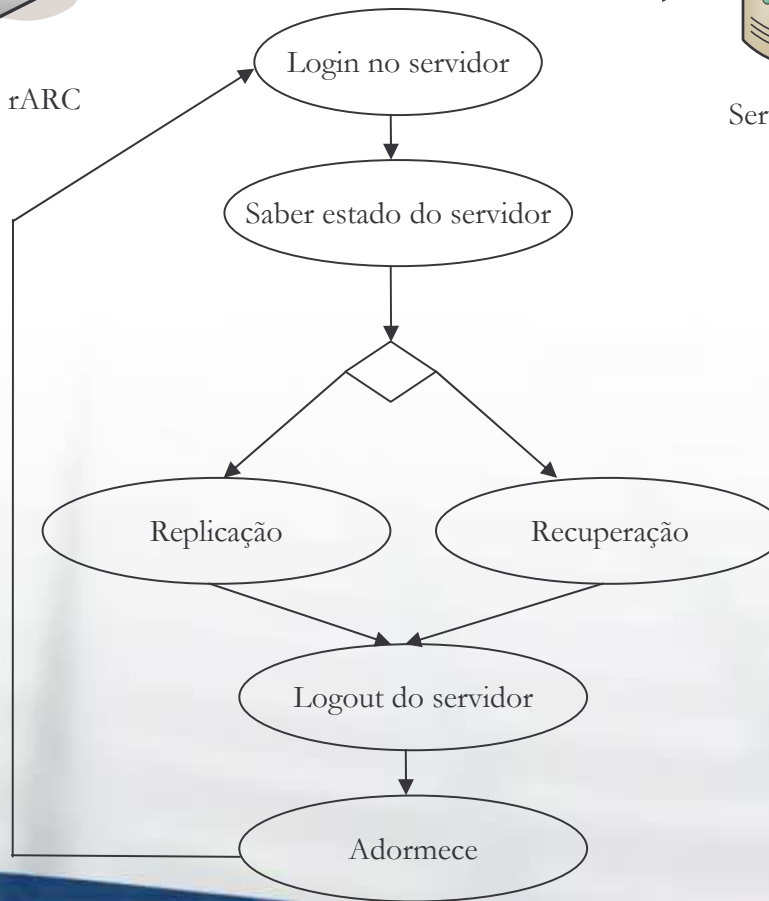
Utilizador

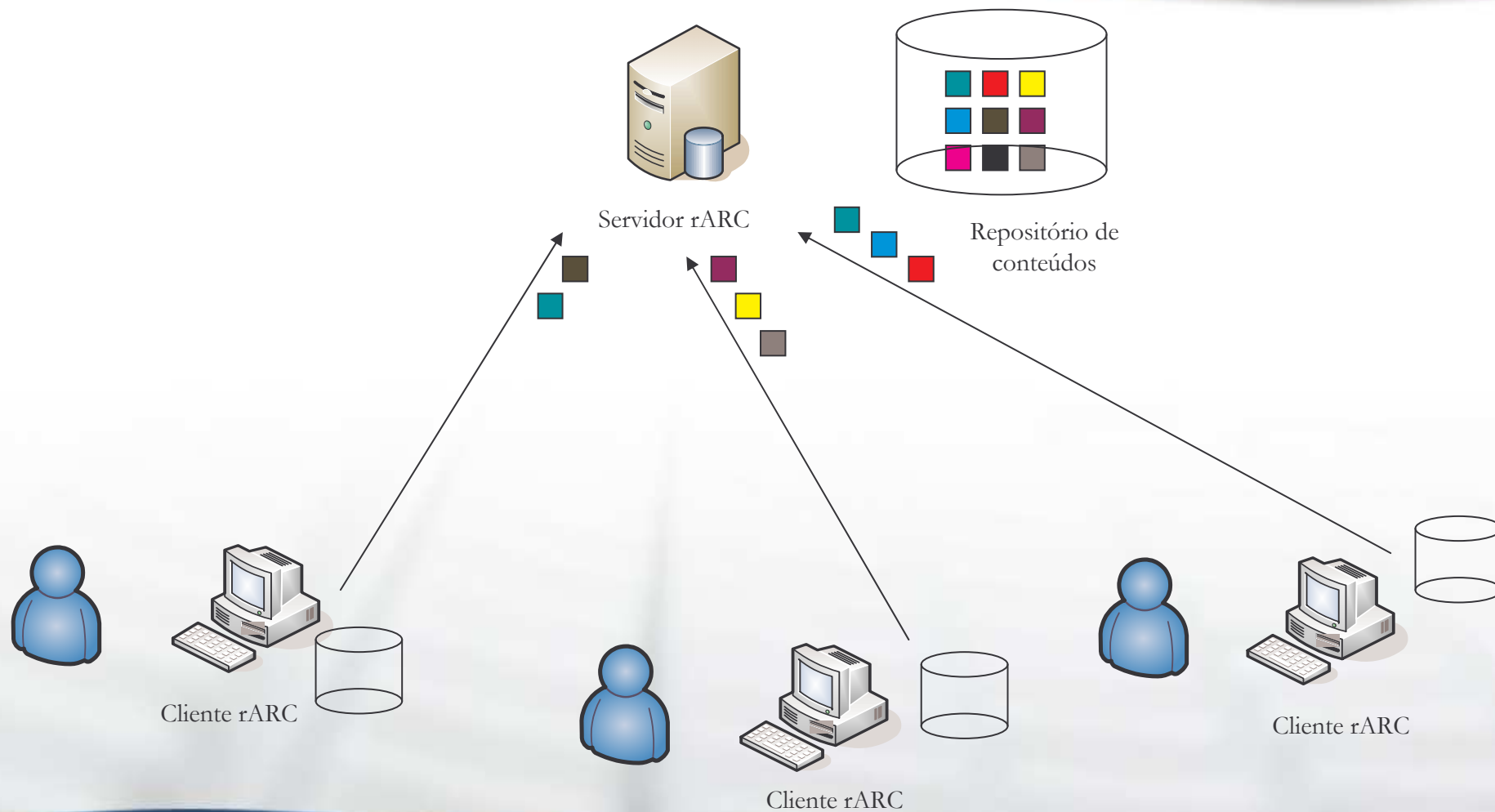


Cliente rARC

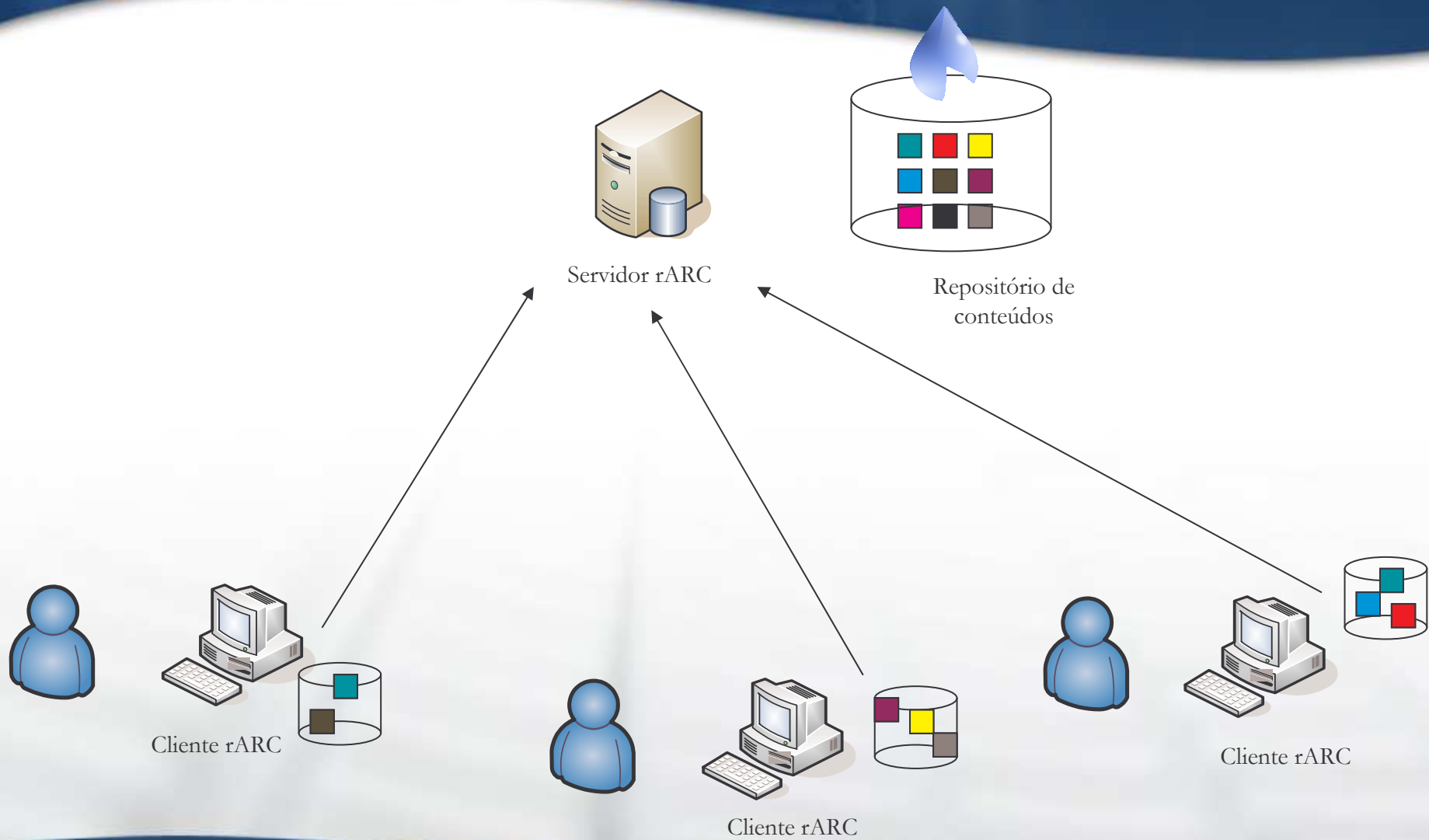


Servidor rARC

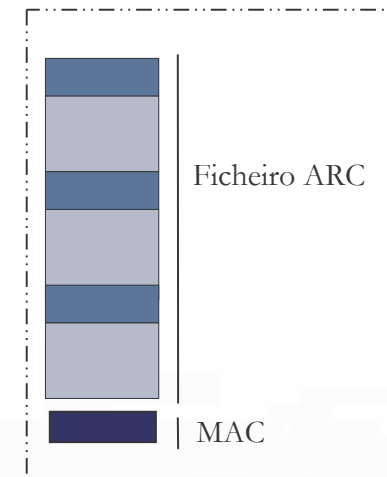




Processo de recuperação

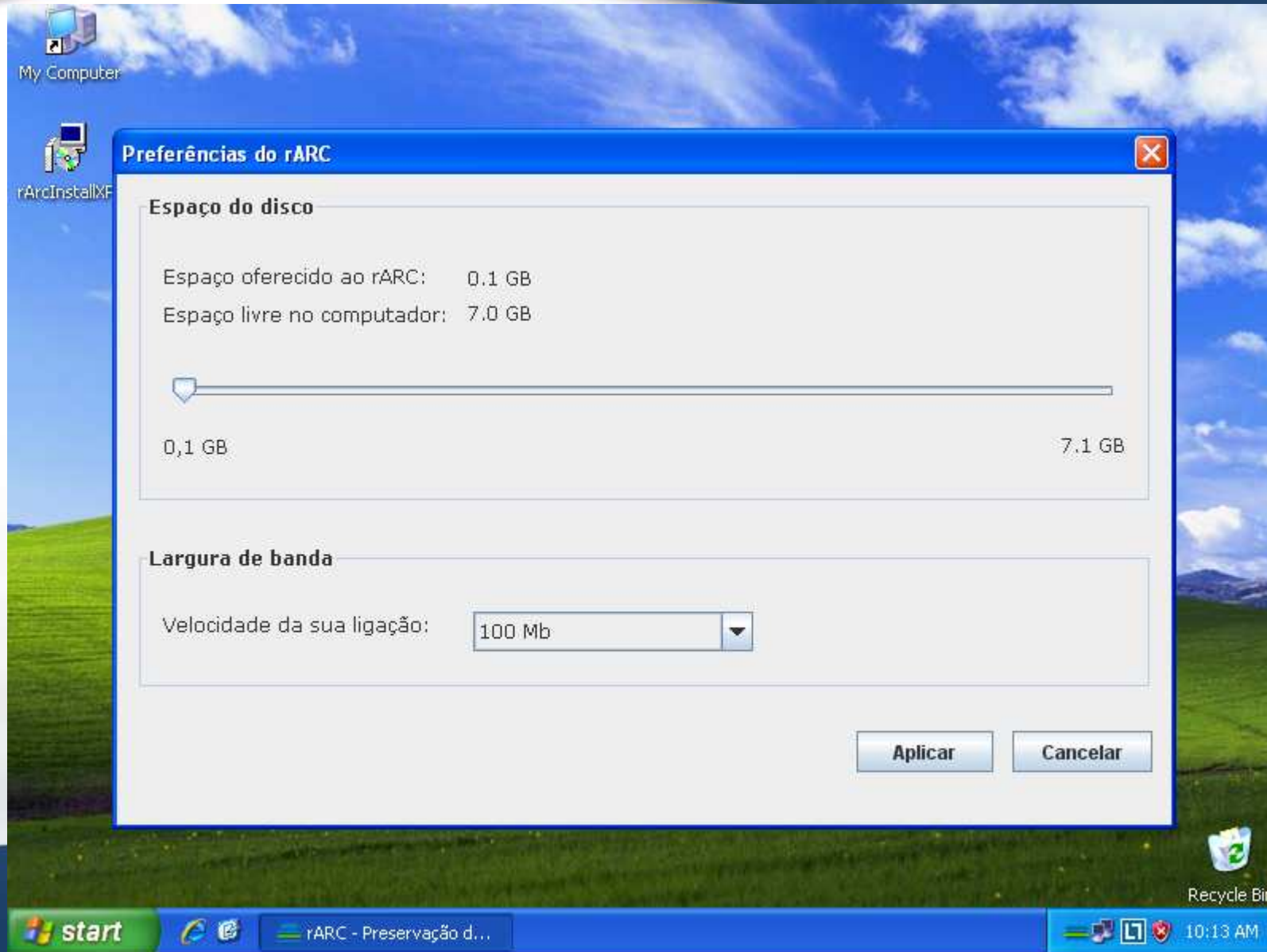


- Uma cápsula é constituída por:
 - Ficheiro ARC
 - Informação segurança (MAC)
- Permite garantir a confidencialidade e integridade dos conteúdos



Cifrado

Como se tornar um contribuidor?



Como motivar os contribuidores?



Os 20 maiores contribuidores do rARC

A seguinte tabela apresenta os 20 contribuidores que disponibilizam mais espaço.

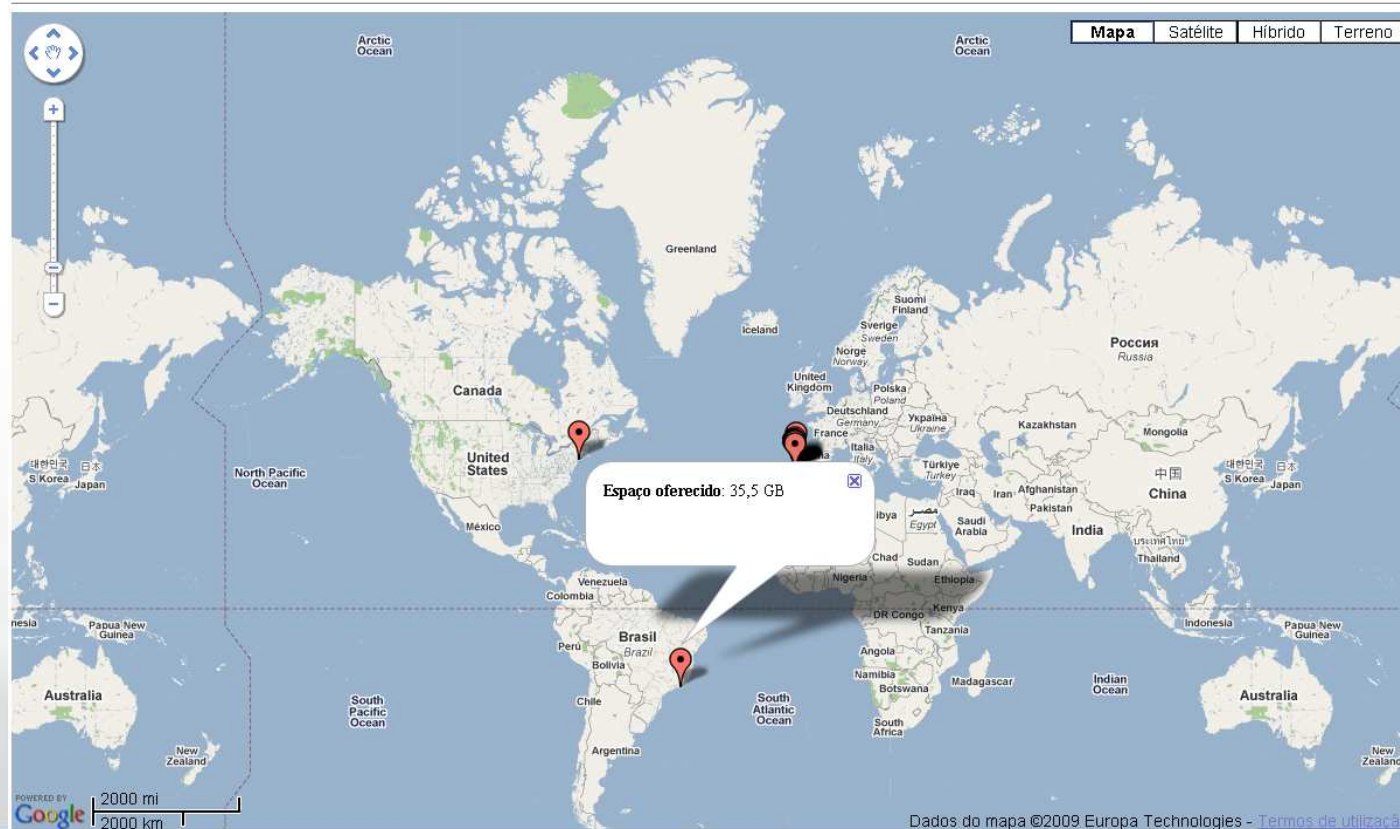
[Lista completa](#)

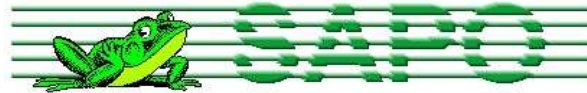
Posição	Nome	Contribuidor desde	Espaço oferecido
1	jonathas	2009-10-01	35,5 GB
2	Andy [TAZ Server]	2009-07-23	30,2 GB
3	Rudolfo Santos	2009-10-17	30,1 GB
4	Nuno Valdez	2009-07-27	30 GB
5	Joao Gomes	2009-09-21	25,5 GB
6	Andy	2009-07-23	24 GB
7	Ana Ventura	2009-10-13	20 GB
8	Rita Sales Lopes	2009-10-14	20 GB
9	Nuno Miguel Dias Mendes Gonçalves	2009-07-20	15 GB
10	Sónia Veloso	2009-10-09	12,7 GB
11	Sandra Santos	2009-10-10	12,1 GB
12	Joana Pena	2009-10-13	10,7 GB
13	CSilva	2009-10-13	10,1 GB
14	CSilva - Work	2009-10-14	10,1 GB
15	Cristina Mota	2009-08-10	10 GB
16	Maria Teresa Costa	2009-10-15	10 GB
17	Sónia Ruivo	2009-07-27	9,6 GB
18	Estela Cruz	2009-10-13	9,2 GB
19	Fernando	2009-10-13	9 GB
20	Eduardo Ribeiro	2009-08-13	7,1 GB

[Lista completa](#)

Mapa com a distribuição geográfica das cópias de segurança

Versão experimental





Servidor de Apontadores Portugueses

- **Novidades**
[Novos Links no SAPO](#), [Congressos](#), [Exposições](#), [Cursos de Formação](#), ...
- **Ensino e Investigação**
[Universidades](#), [Institutos](#), [Escolas](#), [Centros](#), ...
- **Comunicação Social**
[Jornais](#), [Rádios](#), [Televisão](#), [Revistas](#), ...
- **Entretenimento**
[Desportos](#), [Fotografia](#), [Passatempos](#), [Livros](#), [Fora de Casa](#), [Educatonal](#), ...
- **Serviços de Informação**
[Arquivos de Software](#), [Usenet News](#), [Talkers](#), [Archie](#), [Gopher](#), ...
- **Comércio, Indústria e Serviços**
[Serviços](#), [Informática](#), [Turísticas](#), [Financeiras](#), [Transportes](#), ...
- **Páginas Pessoais**
[Páginas pessoais de Portugueses](#), [Lista de Endereços de Correio Electrónico](#)
- **Sociedade e Cultura**
[Museus](#), [Bibliotecas](#), [Eventos](#), [Hospitais](#), [Religião](#), [Tributos](#), [Grupos e Associações](#), [Ministérios e Direcções Gerais](#), ...
- **Regional**
[Terras de Portugal](#), [Regiões de Turismo](#), [Açores](#), [Madeira](#), [Macau](#), [Timor](#), ...
- **Computadores e Internet**
[Documentação](#), [Web Designers](#), [Software](#), [Internet Service Providers](#), [Jogos](#), ...



[Centro de Informática da Universidade de Aveiro](#)

Internet URL- <http://sapo.ua.pt:80/>



- Ferramentas
 - ArgoUML (análise e desenho)
 - Java (desenvolvimento)
 - NetBeans
 - MySQL (desenvolvimento)

- Ferramentas e tecnologias de código aberto
 - Não existe custo de aquisição

- Motivação dos contribuidores
 - Conta do contribuidor
- Arquitectura *Peer-to-Peer*

- Lista de propostas de projectos de colaboração com o Arquivo da Web Portuguesa :
 - Pesquisa de imagens arquivadas
 - Classificação automática de conteúdos Web arquivados
 - Perfis de utilizadores e casos de uso para um Arquivo da Web
 - Criação de uma bancada de teste para resultados de pesquisas temporais
 - ...

- O rARC foi criado porque não existia um sistema que respondesse aos requisitos dos arquivos da Web.
- A comunidade pode contribuir para a preservação dos conteúdos dos arquivos da Web

Obrigado pela atenção.

Questões?

Arquivo da Web Portuguesa: <http://arquivo.pt>

rARC: <http://arquivo.pt/rarc>

Propostas de mestrado: <http://arquivo.pt/propostas>