

SCEUR – Serviço Centralizado de Estatísticas de Utilização de Repositórios

Construindo um Serviço de Estatísticas
(Hélder Silva, José Carlos Ramalho e Miguel Ferreira)

Estatísticas

If you want to inspire confidence, give plenty of statistics. It does not matter that they should be accurate, or even intelligible, as long as there is enough of them.

Lewis Carroll, escritor e matemático Inglês (século XIX)

Agenda

- Contexto
- Motivação
- Estatísticas no repositório
- Estatísticas no consórcio
- Serviço Centralizado de Estatísticas
- Resultados
- Conclusões
- Trabalho futuro

Contexto

Dissertação de **mestrado** (Hélder Silva)

Projeto **RCAAP**

Agências de fomento



Motivação

Aumento da quantidade de repositórios

Portugal: 39 *

Brasil: 60 *

Espanha: 77 *

Aparecimento de consórcios

Portugal: RCAAP

Brasil: OASIS

Espanha: RECOLECTA

Necessidade de estatísticas globais

Métricas baseadas em utilização

Alternativa às métricas baseadas em citações

Imediatamente disponíveis

Mais abrangentes

* Segundo o sítio OpenDOAR (2011-11-11)

(<http://www.opendoar.org/>)

Estatísticas no repositório

Já **disponibilizadas** nas diferentes plataformas de *software*

Downloads e **visualização** de metadados

Têm algumas **limitações**

- Trabalhos colaborativos

- Podia ser mais ricas

Estatísticas no consórcio

Geração de estatísticas é **trabalhosa**

Que estatísticas são **importantes**?

Têm **limitações**

Comparação entre um ou mais repositórios é **manual**

Estatísticas podem estar **acesso restrito**

Não há garantias da sua **qualidade**

Questões pertinentes a nível central

Normalização

Escalabilidade

Serviço Centralizado de Estatísticas

Objetivos

Construir **arquitetura** para recolher e processar dados de utilização de repositórios

Disponibilizar serviços associados a esses dados

- Geração de estatísticas variadas

- Geração de gráficos em tempo-real

- Subscrição de estatísticas por correio-eletrónico

- Permitir embeber gráficos em sítios externos

Serviço Centralizado de Estatísticas

Decisões

Seguir **recomendações** internacionais

Eventos a **publicar**

Consulta de metadados

Downloads

Depósitos

Publicar também **metadados descritivos**

dc.author

dc.type

dc.rights

dc.format

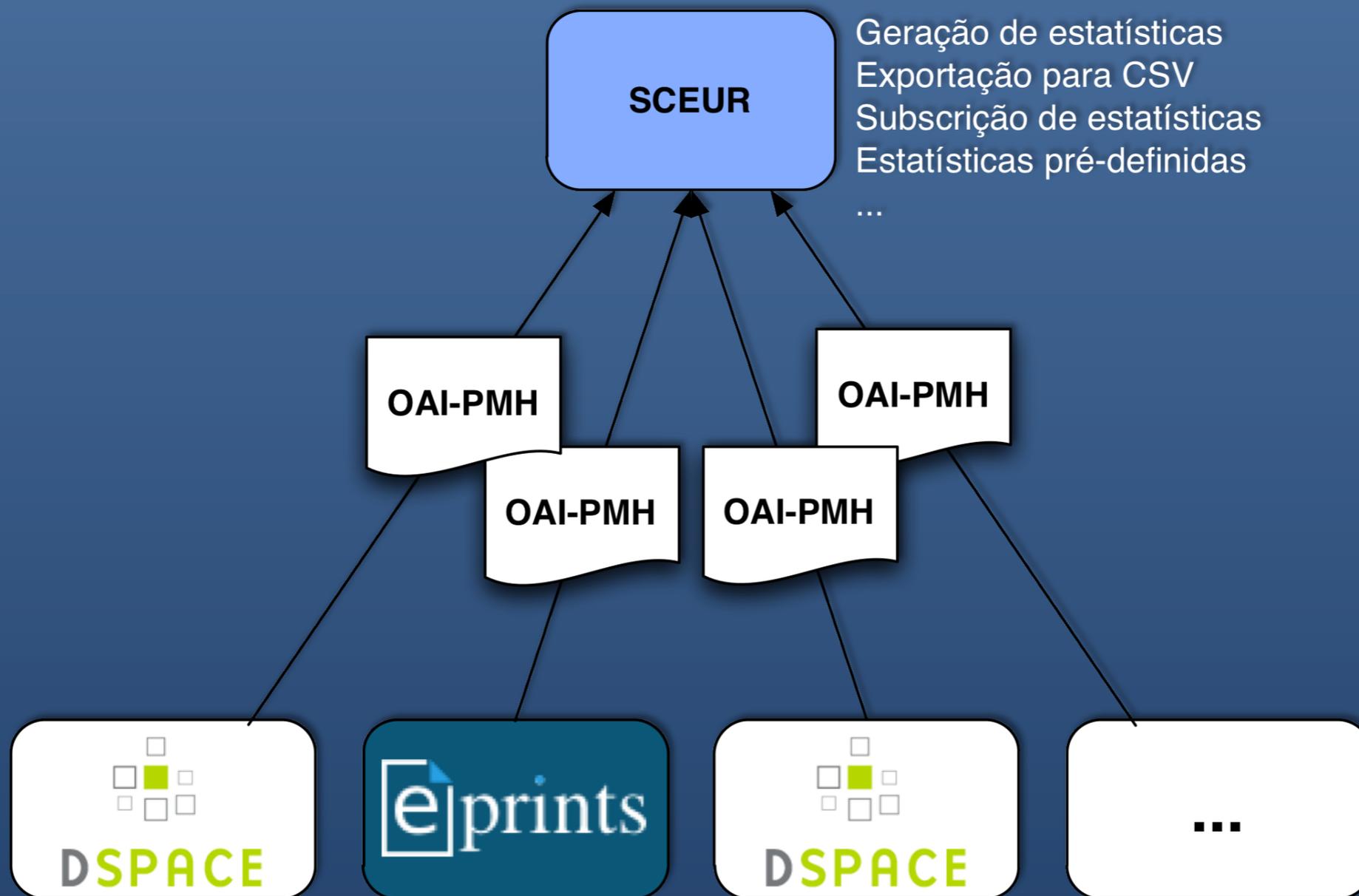
dc.language

Usar OAI-PMH como protocolo de **interoperabilidade**

Desenvolver **add-ons** para os repositórios

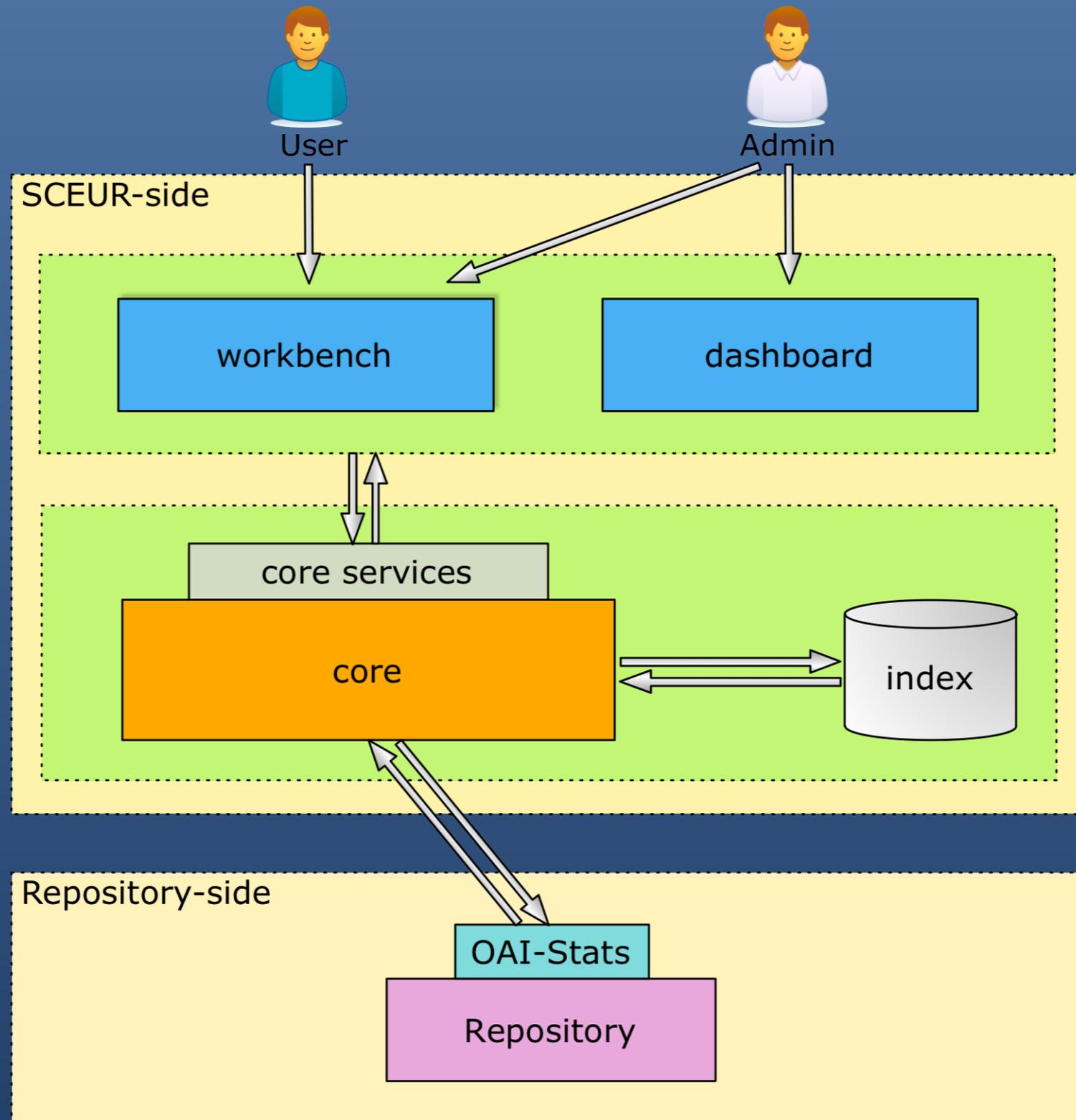
Serviço Centralizado de Estatísticas

Arquitetura



Serviço Centralizado de Estatísticas

Módulos



Serviço Centralizado de Estatísticas

Módulos - OAI-Stats

Add-on para o DSpace

Partilha dados de utilização via **OAI-PMH**

Partilha também de **metadados** descritivos

Encapsula dados de utilização dando resposta a **3 questões:**

O quê?

Quem?

Quando?

Serviço Centralizado de Estatísticas

Módulos - OAI-Stats

O quê?

```
- <ctx:referent>
  - <ctx:identifier>
    http://comum.rcaap.pt/bitstream/123456789/12/1/getPDF.pdf
  </ctx:identifier>
  <ctx:identifier>http://comum.rcaap.pt/handle/123456789/12</ctx:identifier>
  - <ctx:metadata-by-val>
    <ctx:format>http://dublincore.org/documents/dcmi-terms/</ctx:format>
    - <ctx:metadata>
      <dcterms:type>article</dcterms:type>
      <dcterms:rights>open access</dcterms:rights>
      <dcterms:format>application/pdf</dcterms:format>
      <dcterms:language>eng</dcterms:language>
      <dcterms:author>Afonso, Joao</dcterms:author>
      <dcterms:author>Veiga, Pedro</dcterms:author>
    </ctx:metadata>
  </ctx:metadata-by-val>
</ctx:referent>
+ <ctx:requester></ctx:requester>
- <ctx:service-type>
  - <ctx:metadata-by-val>
    <ctx:format>http://dublincore.org/documents/dcmi-terms/</ctx:format>
  - <ctx:metadata>
    <dcterms:format>info:eu-repo/semantics/objectFile</dcterms:format>
  </ctx:metadata>
</ctx:metadata-by-val>
</ctx:service-type>
```

Serviço Centralizado de Estatísticas

Módulos - OAI-Stats

Quem?

```
- <ctx:requester>  
  <ctx:identifier>data:0a28242b48dea9247c25c1438ad4a571</ctx:identifier>  
  <ctx:identifier>data:193.136.44.0</ctx:identifier>  
- <ctx:metadata-by-val>  
  <ctx:format>http://dublincore.org/documents/dcmi-terms/</ctx:format>  
- <ctx:metadata>  
  <dcterms:spatial>PT</dcterms:spatial>  
  </ctx:metadata>  
</ctx:metadata-by-val>  
</ctx:requester>
```

Serviço Centralizado de Estatísticas

Módulos - OAI-Stats

Quando?

```
- <ctx:context-object identifier="oai:comum.rcaap.pt:download_64"
timestamp="2009-11-02T16:59:02Z">
+ <ctx:referent></ctx:referent>
+ <ctx:requester></ctx:requester>
+ <ctx:service-type></ctx:service-type>
+ <ctx:resolver></ctx:resolver>
</ctx:context-object>
</ctx:context-objects>
```

Serviço Centralizado de Estatísticas

Módulos - workbench



SCEUR workbench

Serviço Centralizado de Estatísticas de Utilização de Repositórios

1) Tipo de eventos

Selecionar o tipo de evento para produzir a estatística.

- Consultas de metadados
- Downloads de Documentos
- Consultas de metadados + Downloads
- Depósitos

2) Repositórios

3) Período de tempo

4) Tipo de estatística

5) Personalização do gráfico

Gerar gráfico

Serviço Centralizado de Estatísticas

Módulos - workbench



SCEUR workbench

Serviço Centralizado de Estatísticas de Utilização de Repositórios

1) Tipo de eventos

2) Repositórios

Escolha um ou mais repositórios (para fazer comparações), sendo que a última opção considera todos os repositórios disponíveis.

- ARCA - IGC Repository (arca)
- Biblioteca Digital do IPB (bibliotecadigital)
- Repositório Comum (comum)
- Repositório Digital da Universidade da Madeira (digituma)
- IC-online (iconline)
- Repositório do ISCTE-IUL (iscte)
- Repositório Aberto (raberto)
- Repositório do Centro Hospitalar de Lisboa Central, EPE (rchl)
- Repositório Científico do Centro Hospitalar do Porto (rchporto)
- Repositório da Universidade de Lisboa (repul)
- Repositório do Hospital Prof. Doutor Fernando Fonseca (rhff)
- RIHUC (rihuc)

Serviço Centralizado de Estatísticas

Módulos - workbench



SCEUR workbench

Serviço Centralizado de Estatísticas de Utilização de Repositórios

1) Tipo de eventos

2) Repositórios

3) Período de tempo

Configure o período de tempo para a estatística que pretende obter. Existem duas grandes opções: datas específicas e datas pré-definidas.

Datas específicas - pode definir só a data inicial (significando todos os eventos desde essa data), só a data final (significando todos os eventos até essa data), as duas datas ou nenhuma (significando todos os eventos disponíveis no sistema).

Datas pré-definidas - pode usar uma das datas pré-definidas para usar as subscrições por e-mail no caso de querer, por exemplo as estatísticas do mês anterior (valor que se refere sempre ao mês anterior contrariamente às datas específicas que são "estáticas").

Datas específicas

Data inicial

Formato: AAAA ou AAAA-MM ou AAAA-MM-DD

Data final

Datas pré-definidas

Datas pré-definidas

- ✓ Este ano (por meses)
- Este ano (acumulado)
- Este mês (por dias)
- Este mês (acumulado)
- Último mês (por dias)
- Último mês (acumulado)

4) Tipo de estatística

Serviço Centralizado de Estatísticas

Módulos - workbench

1) Tipo de eventos

2) Repositórios

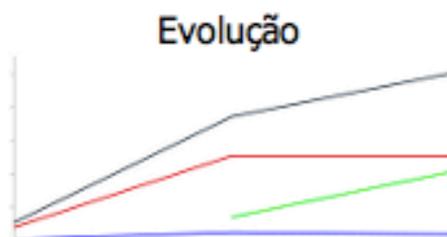
3) Período de tempo

4) Tipo de estatística

Selecione qual o tipo de estatística que deseja obter. Existem duas grandes opções: estatísticas de evolução ou estatísticas de ranking.

Estatísticas de evolução nas seguintes dimensões: por evento, por evento e nome do autor, por evento e handle do documento.

Estatísticas de ranking sobre as seguintes características: handle, autor, dc.type, dc.rights, dc.language, dc.format (no caso dos downloads e depósitos) e por país que originou o evento.



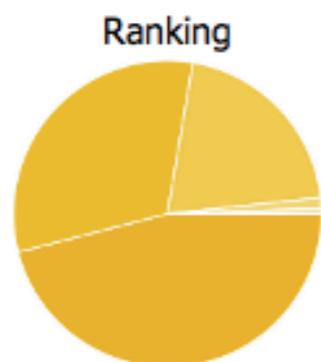
- Total de eventos ao longo do tempo
- Total de eventos ao longo do tempo por autor

Nome do autor, e.g., António Moreira ou Moreira, António

- Total de eventos ao longo do tempo por documento

Handle do documento, e.g., <http://hdl.handle.net/1822/5820>

Pode fornecer vários handles, separados por ponto e vírgula, e as suas estatísticas serão somadas



- Top 10 documentos por handle
- Top 10 autores
- Top 10 documentos por dc.type
- Top 10 documentos por dc.rights
- Top 10 documentos por dc.language
- Top 10 documentos por dc.format
- Top 10 países

Serviço Centralizado de Estatísticas

Módulos - workbench

1) Tipo de eventos

2) Repositórios

3) Período de tempo

4) Tipo de estatística

5) Personalização do gráfico

Personalize o gráfico gerado, alterando a cor, adicionando um título, adicionando o fundo em grelha, entre outros.

Título do gráfico

Cor do gráfico

Esconder etiquetas (labels)?

Fundo em grelha?

Aplicar gradiente ao fundo do gráfico?

Largura

Altura

Largura*Altura deve ser inferior a 300000

Serviço Centralizado de Estatísticas

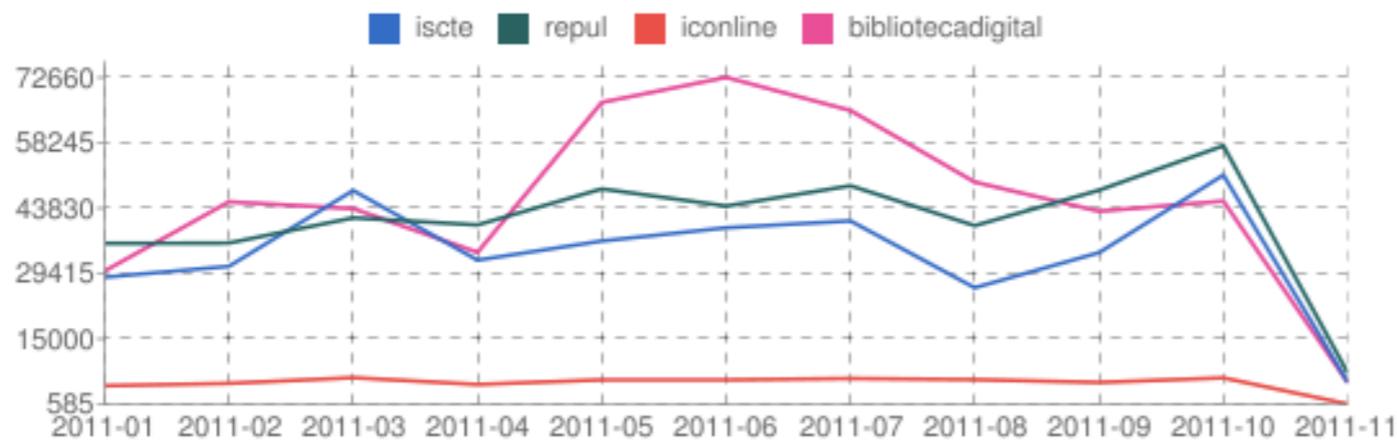
Módulos - workbench

5) Personalização do gráfico

Gerar gráfico

Gráfico gerado

Evolução da consulta de metadados em 2011



Data	bibliotecadigital	iconline	iscte	repul
2011-01	29953	4650	28598	36072
2011-02	45225	5172	30991	36133
2011-03	43731	6429	47726	41690
2011-04	34080	4931	32365	40128
2011-05	67169	5960	36586	48088
2011-06	72693	5945	39500	44293
2011-07	65377	6243	41022	48779
2011-08	49583	6007	26272	39966
2011-09	43151	5380	34027	47791
2011-10	45393	6392	51133	57600
2011-11	5372	616	5464	7480
Total	501727	57725	373684	448020

Opções:

Subscrever

Exportar para CSV

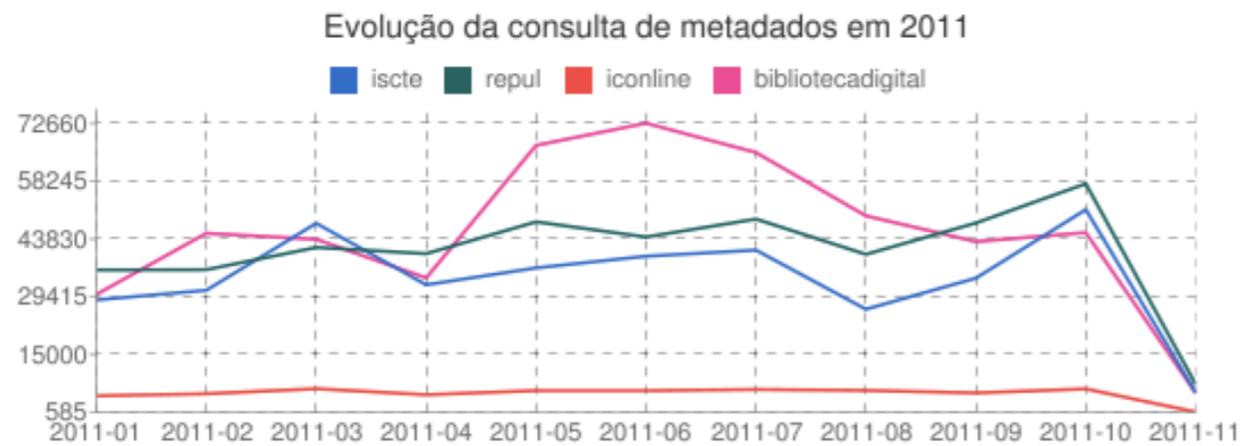
HTML snippet

```
<iframe src="http://sceur.rcaap.pt/dataHandler?c=view&ct=e1&h=&a=&ctr=bibliotecadigital&c_bibliotecadigital=FF1994&ctr=iconline&c_iconline=FF3333&ctr=iscte&c_iscte=3366CC&ctr=repul&c_repul=006666&dt=a&bd=&e d=&cd=thisYear&gl=true&co=EAB300&wi=540&he=200&t=Evolução da consulta de metadados em 2011&l=pt" marginheight="0px" marginwidth="0px" frameborder="0" scrolling="no" height="200" width="540" >&nbsp;&nbsp;&nbsp;</iframe>
```

Serviço Centralizado de Estatísticas

Módulos - workbench

Gráfico a subscrever



Data	bibliotecadigital	iconline	iscte	repul
2011-01	29953	4650	28598	36072
2011-02	45225	5172	30991	36133
2011-03	43731	6429	47726	41690
2011-04	34080	4931	32365	40128
2011-05	67169	5960	36586	48088
2011-06	72693	5945	39500	44293
2011-07	65377	6243	41022	48779
2011-08	49583	6007	26272	39966
2011-09	43151	5380	34027	47791
2011-10	45393	6392	51133	57600
2011-11	5372	616	5464	7480
Total	501727	57725	373684	448020

Dados necessários

E-mail: *

Nome do gráfico (também usado como "Assunto" no e-mail enviado): *

Descrição do gráfico:

Periodicidade: *

Semanal Mensal

* - Campos de preenchimento obrigatório

Subscrever

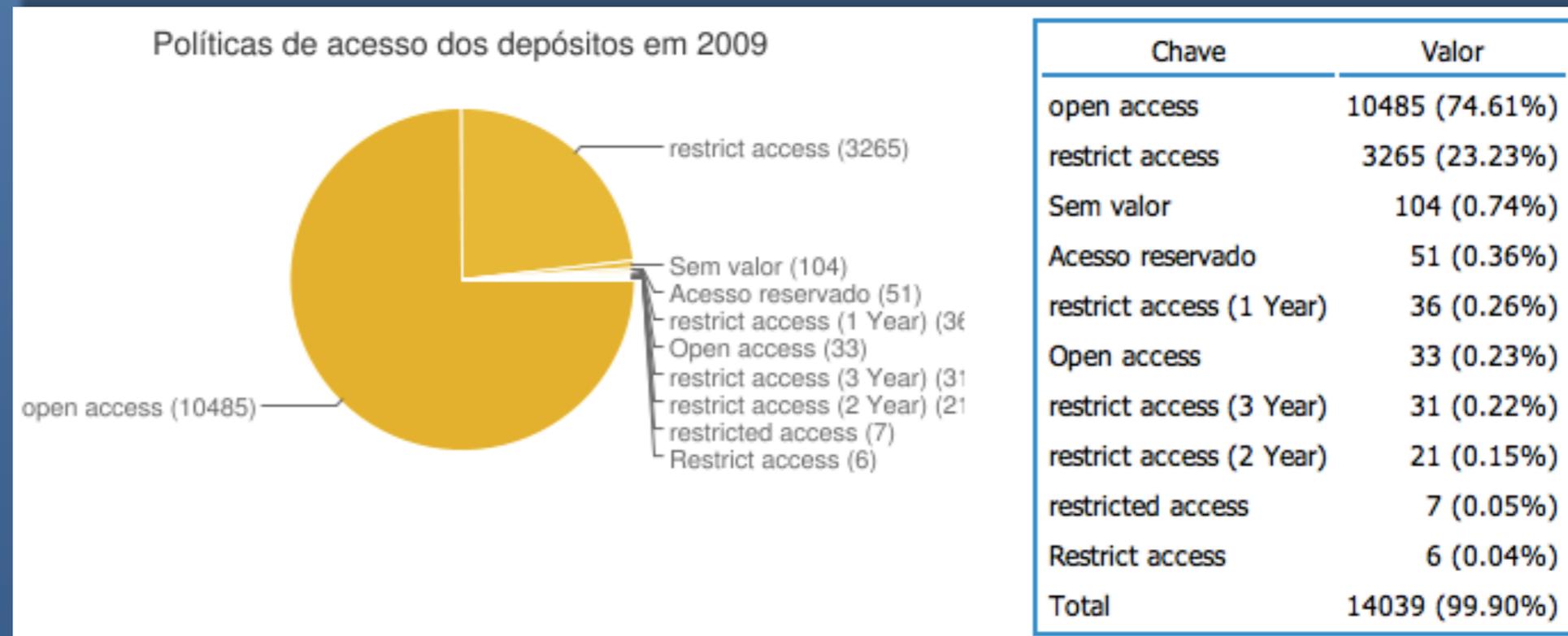
Opções

Voltar e editar

Voltar

Resultados

Depósitos



Resultados

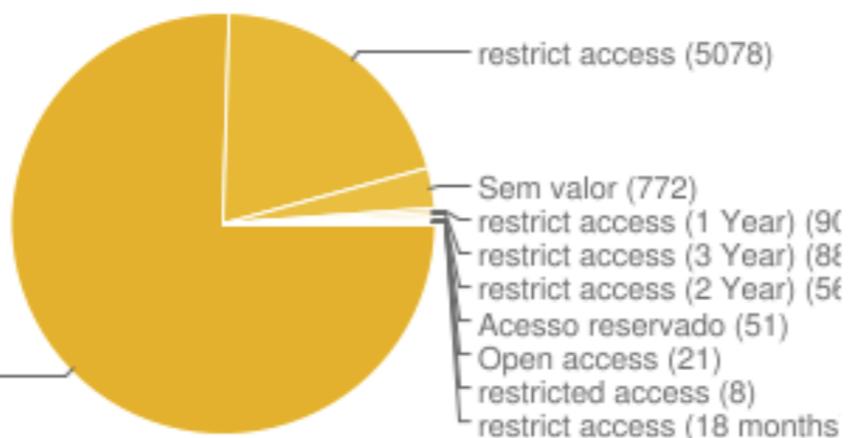
Depósitos

Políticas de acesso dos depósitos em 2009



Chave	Valor
open access	10485 (74.61%)
restrict access	3265 (23.23%)
Sem valor	104 (0.74%)
Acesso reservado	51 (0.36%)
restrict access (1 Year)	36 (0.26%)
Open access	33 (0.23%)
restrict access (3 Year)	31 (0.22%)
restrict access (2 Year)	21 (0.15%)
restricted access	7 (0.05%)
Restrict access	6 (0.04%)
Total	14039 (99.90%)

Políticas de acesso dos depósitos em 2010



Chave	Valor
open access	18879 (75.30%)
restrict access	5078 (20.25%)
Sem valor	772 (3.08%)
restrict access (1 Year)	90 (0.36%)
restrict access (3 Year)	88 (0.35%)
restrict access (2 Year)	56 (0.22%)
Acesso reservado	51 (0.20%)
Open access	21 (0.08%)
restricted access	8 (0.03%)
restrict access (18 months)	4 (0.02%)
Total	25047 (99.90%)

open access (18879)

restrict access (5078)

Sem valor (772)

restrict access (1 Year) (90)

restrict access (3 Year) (88)

restrict access (2 Year) (56)

Acesso reservado (51)

Open access (21)

restricted access (8)

restrict access (18 months)

Resultados

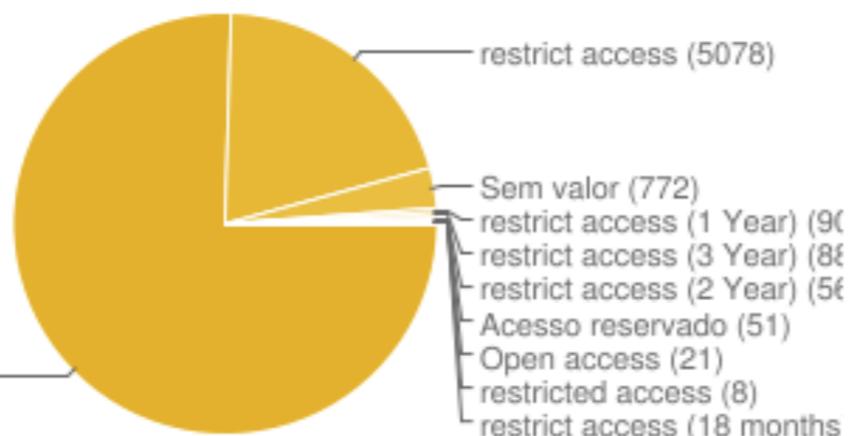
Depósitos

Políticas de acesso dos depósitos em 2009



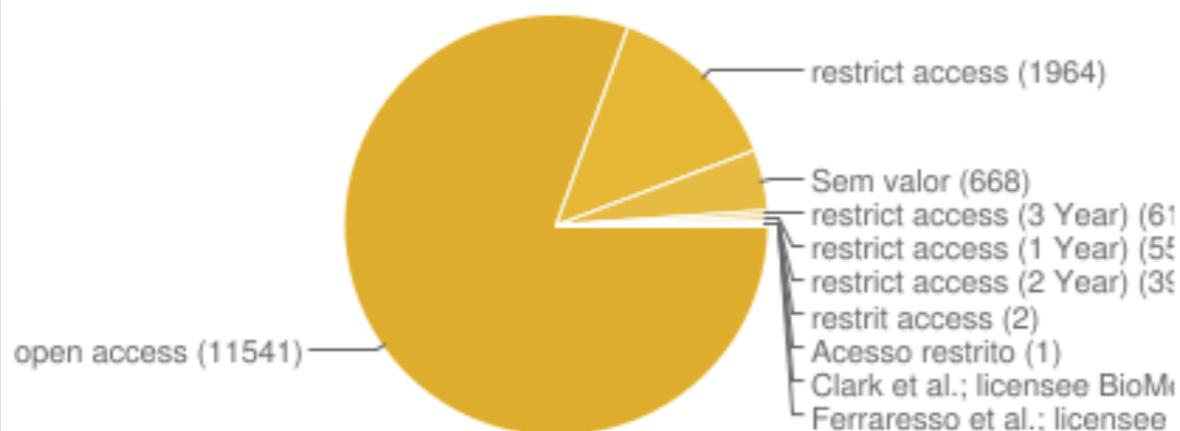
Chave	Valor
open access	10485 (74.61%)
restrict access	3265 (23.23%)
Sem valor	104 (0.74%)
Acesso reservado	51 (0.36%)
restrict access (1 Year)	36 (0.26%)
Open access	33 (0.23%)
restrict access (3 Year)	31 (0.22%)
restrict access (2 Year)	21 (0.15%)
restricted access	7 (0.05%)
Restrict access	6 (0.04%)
Total	14039 (99.90%)

Políticas de acesso dos depósitos em 2010



Chave	Valor
open access	18879 (75.30%)
restrict access	5078 (20.25%)
Sem valor	772 (3.08%)
restrict access (1 Year)	90 (0.36%)
restrict access (3 Year)	88 (0.35%)
restrict access (2 Year)	56 (0.22%)
Acesso reservado	51 (0.20%)
Open access	21 (0.08%)
restricted access	8 (0.03%)
restrict access (18 months)	4 (0.02%)
Total	25047 (99.90%)

Políticas de acesso dos depósitos em 2011

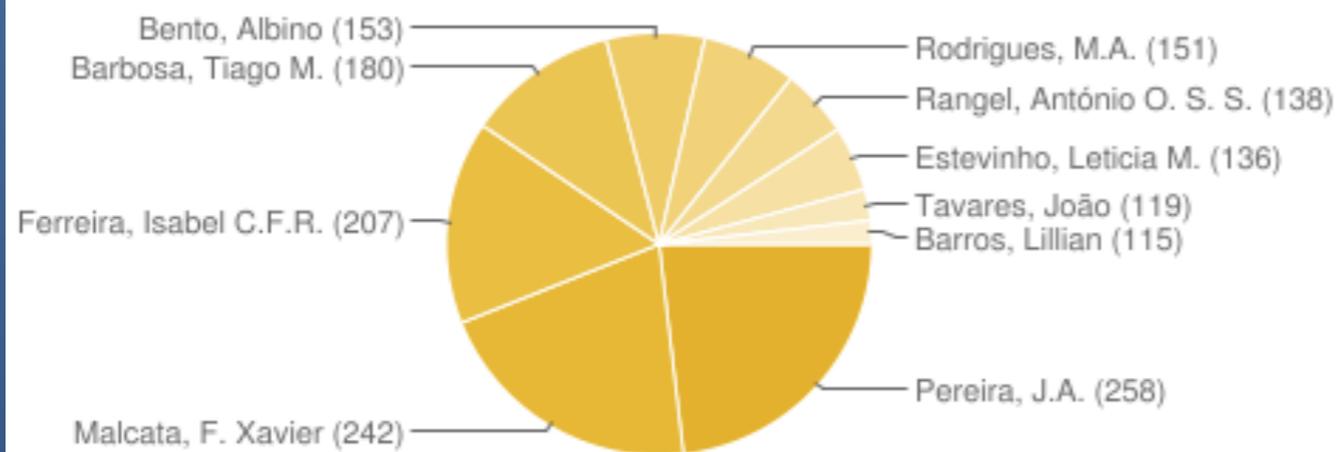


Chave	Valor
open access	11541 (80.45%)
restrict access	1964 (13.69%)
Sem valor	668 (4.66%)
restrict access (3 Year)	61 (0.43%)
restrict access (1 Year)	55 (0.38%)
restrict access (2 Year)	39 (0.27%)
restrit access	2 (0.01%)
Acesso restrito	1 (0.01%)
Clark et al.; licensee BioMed Central Ltd.	1 (0.01%)
Ferraresso et al.; licensee BioMed Central Ltd.	1 (0.01%)
Total	14333 (99.91%)

Resultados

Depósitos

10 autores com mais depósitos em 2010



Chave	Valor
Pereira, J.A.	258 (1.03%)
Malcata, F. Xavier	242 (0.97%)
Ferreira, Isabel C.F.R.	207 (0.83%)
Barbosa, Tiago M.	180 (0.72%)
Bento, Albino	153 (0.61%)
Rodrigues, M.A.	151 (0.60%)
Rangel, António O. S. S.	138 (0.55%)
Estevinho, Leticia M.	136 (0.54%)
Tavares, João	119 (0.47%)
Barros, Lillian	115 (0.46%)
Total	1699 (6.78%)

Conclusões

Ferramenta ideal para medir impacto baseado em **utilização**

Alternativa prática e rápida a métricas baseadas em **citações**

Centraliza informação num único sítio Web

Está em **acesso-livre**

Em **constante** evolução

Trabalho futuro

Normalização dos dados

Robots

Metadados

Multiplos *clicks*

Novos eventos

Publicação

Quantidade de documentos

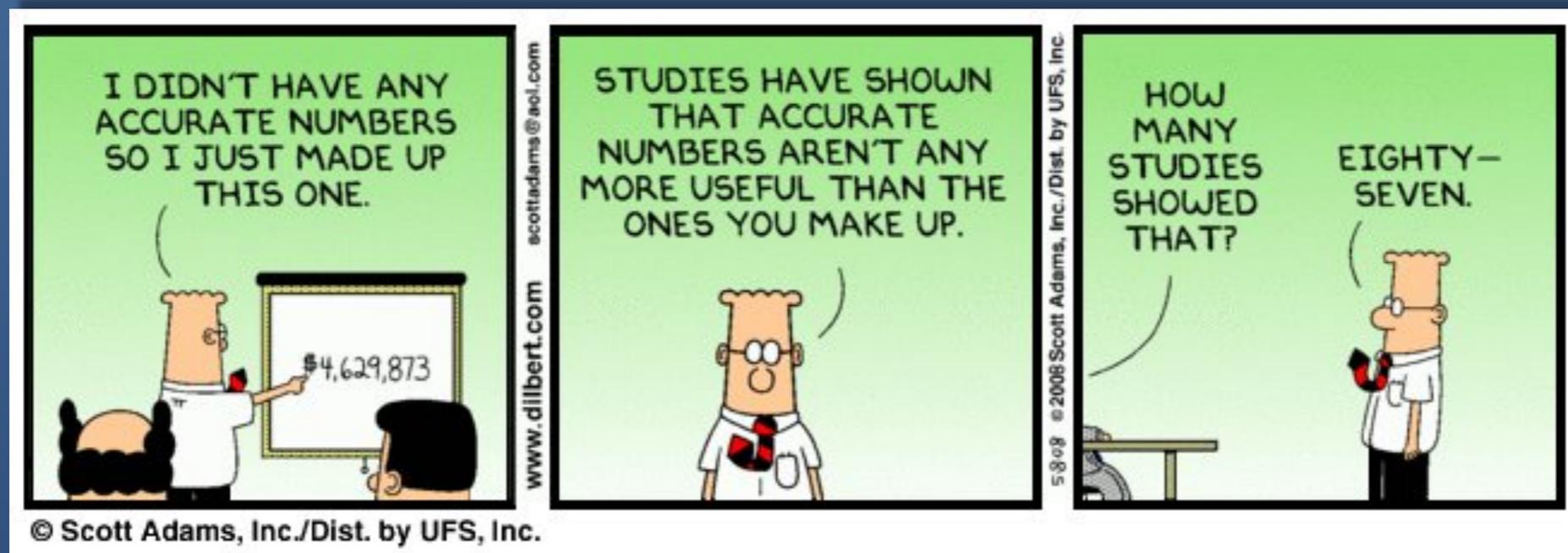
Data Provider OAI-PMH

Otimizações

Add-ons para outras plataformas

SCEUR – Serviço Centralizado de Estatísticas de Utilização de Repositórios

Construindo um Serviço de Estatísticas
(Hélder Silva, José Carlos Ramalho e Miguel Ferreira)



Hélder Silva hsilva@keep.pt
KEEP SOLUTIONS info@keep.pt
www.keep.pt