

# REPOSITÓRIO INSTITUCIONAL DA FIOCRUZ - ARCA

## MANUAL DE TRATAMENTO DE OBJETOS DIGITAIS

**Patricia Henning**

Professora da Universidade Federal do Estado do Rio de Janeiro – Unirio  
Doutoranda do Instituto de Comunicação e Informação Científica e Tecnológica – ICICT

**Éder de Almeida Freyre**

Bibliotecário do Instituto de Comunicação e Informação Científica e Tecnológica - ICICT  
Mestre em Ciência da Informação

**Cléa Mara Barradas dos Reis**

Bibliotecária do Instituto de Comunicação e Informação Científica e Tecnológica – ICICT  
Especialista em Informação Científica e Tecnológica em Saúde - ICICT

**Maria Cristina Soares Guimarães**

Vice-Diretora do Instituto de Comunicação e Informação Científica e Tecnológica - ICICT  
Doutora em Ciência da Informação

O Repositório Institucional da Fiocruz – ARCA, localizado no site <<http://www.arca.fiocruz.br>>, foi criado em 2007 com o objetivo inicial de preservação da produção técnico-científica do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde - ICICT. Em 2010 se institucionalizou e ampliou seu escopo, aumentando a visibilidade da produção intelectual por meio da informação proveniente dos institutos, escolas, centros, editoras e revistas da Fiocruz.

Atualmente disponibiliza nove comunidades: Casa de Oswaldo Cruz, Centro de Pesquisa Gonçalo Moniz, Editora Fiocruz, Escola Nacional de Saúde Pública, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Instituto Fernandes Figueira, Instituto Leônidas e Maria Deane, Instituto Oswaldo Cruz e Revista Eletrônica de Comunicação, Informação e Inovação em Saúde.

As coleções de cada uma dessas comunidades são de natureza técnico-científica, tais como: artigos, livros, capítulos de livro, teses e dissertações, trabalhos de conclusão de curso, trabalhos apresentados em eventos, relatórios de pesquisa, relatórios institucionais, manuais, vídeos e a Revista RECIIS.



Página inicial do site do Repositório Institucional da FIOCRUZ



Página das Comunidades & Coleções do Repositório Institucional da FIOCRUZ

Utiliza a tecnologia de software livre DSpace, mantido pelo MIT Libraries e pela Hewlett-Packard Cia, desenvolvido para facilitar a criação dos Repositórios Institucionais – RI's, distribuído, no Brasil, pelo Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT.

Para o tratamento técnico das coleções foi utilizado o conjunto de elementos de metadados da Dublin Core Metadata Element Set, tomando como referência o modelo conceitual definido pelo Open Archival Information System (OAIS) e a inclusão de elementos adicionais para atender às particularidades da Fiocruz.

Tipologias	Metadados
Artigos de Periódicos	(dc.title), (dc.contributor.author), (dc.creator.affiliation), (dc.data.issues), (dc.identifier.issn), (dc.identifier.doi), (dc.type), (dc.language.iso), (dc.identifier.citation), (dc.rights), (dc.subject.other), (dc.subject.desc), (dc.subject.en), (dc.subject.es), (dc.subject.fr), (dc.description.abstract), (dc.description.abstract.en), (dc.description.abstract.es), (dc.description.abstract.fr), (dc.description.sponsorship), (dc.description), (dc.relation.isbason), (dc.description.abstract.en), (dc.publisher), (dc.identifier.isbn)
Capítulo de Livro	(dc.publisher), (dc.identifier.isbn)
Dissertações de Mestrado	(dc.contributor.advisor), (dc.contributor.advisor.en), (dc.contributor.advisor.es), (dc.degree.department), (dc.degree.program), (dc.degree.level), (dc.degree.date), (dc.degree.local)
Livros Publicados	*Nenhum específico
Relatórios de Pesquisa	(dc.rights.holder), (dc.contributor.coeditor)
Relatórios Institucionais	(dc.title.original)
Teses de Doutorado	Metadados comum - dissertação de mestrado e Trabalho de conclusão de curso (TCC)
Trabalhos Apresentados em Eventos	*Nenhum específico
Trabalhos de Conclusão de Curso	Metadados comum - dissertação de mestrado e Trabalho de conclusão de curso (TCC)
Vídeos	(dc.title.original), (dc.creator.director), (dc.creator.screenplay), (dc.creator.producer), (dc.coverage.temporal), (dc.type.color), (dc.format.material), (dc.contributor.coeditor), (dc.format.medium)

Como podemos observar na tabela acima, a tipologia que mais se distingue das outras é a de "Vídeos", sendo a que demandou criar metadados específicos.

Em função das peculiaridades de cada coleção foram definidos 'metadados' que descrevem e estruturam a informação de forma particular para cada uma das coleções do repositório.



Template referente aos metadados da tipologia "Vídeos"



Template referente aos metadados da tipologia "Artigos de Periódicos"

O tratamento utilizado para a inserção dos objetos digitais por meio de 'metadados' específicos para cada uma das tipologias do repositório ARCA, consistiu na, a) utilização da Norma ABNT 6023 – Informação e documentação – Referências – Elaboração (Associação Brasileira de Normas Técnicas) para preenchimento dos campos: título; autoria; afiliação; referência e bibliografia; b) os Descritores em Ciência da Saúde (DeCS) em campo específico, para normalizar as palavras-chaves atribuídas pelo autor ao objeto digital que está sendo depositado e c) Consultas aos diretórios virtuais Sherpa Romeo e DOAJ, para determinar o status de acesso livre ou restrito, dos objetos digitais que estão sendo depositados no ARCA respeitando-se os direitos autorais dos produtores e editores científicos.



O vocabulário estruturado e trilingue DeCS - Descritores em Ciências da Saúde foi criado pela BIREME para servir como uma linguagem única na indexação de artigos de revistas científicas, livros, anais de congressos, relatórios técnicos, e outros tipos de materiais, assim como para ser usado na pesquisa e recuperação de assuntos da literatura científica nas fontes de informação disponíveis na Biblioteca Virtual em Saúde (BVS) como LILACS, MEDLINE e outras.



O portal SHERPA/ROMEO (<http://www.sherpa.ac.uk/romeo>) reúne as políticas de auto-arquivo de cerca de 1000 editoras internacionais. O serviço ROMEO, em parceria com o SHERPA, permite aos autores saber se podem ou não auto-arquivar as suas publicações em repositórios institucionais, através de um simples código de cores.

A Library of Congress, a NARA (National Archives and Records Administration), a Adobe e a Xerox, elegeram um novo formato para a preservação de documentos eletrônicos a longo prazo, que veio a ser homologado como norma ISO: o PDF/A.

Esse formato fornece especificações para a criação, a visualização e a impressão de documentos digitais para uma preservação a longo prazo e vem sendo utilizado pelo ARCA, para a preservação digital, assim como os metadados Dublin Core para garantir a interoperabilidade dos objetos digitais depositados com outros Repositórios Institucionais

### Vantagens:

- Autodescrição: uma vez que pressupõe a existência de metadados;
- Eliminação de mecanismos de proteção e de restrição de acesso;
- Disponibilidade no mercado: qualquer pessoa pode usar a Referência PDF e a Especificação XMP (formato de metadados do PDF) para conceber aplicações informáticas que leiam e criem arquivos PDF/A-1.

Comparison of Archive Formats	PDF/A	PDF	TIFF without OCR
Device independent	✓	✓	✗
100% self-contained	✓	✗	✗
Guaranteed WYSIWYG	✓	✗	✗
Archive standard approved by ISO	✓	✗	✗
Guaranteed color reproduction	✓	✗	✗
Contains XMP metadata	✓	✗	✗
Guaranteed accessibility across multiple platforms/devices	✓	✗	✗
Embedded fonts	✓	✗	✗
Free from proprietary constraints	✓	✗	✗
Able to capture documents' logical structure	✓	✗	✗
High quality output	✓	✓	✗
Searchable text	✓	✓	✗
Most compact file size	✓	✓	✗

O PDF/A-1 caracteriza-se por ser um formato independente de qualquer plataforma de software ou hardware que se utilize; por ser auto-suficiente, isto é, tudo quanto é necessário para visualizar e imprimir um PDF/A-1 está embutido no próprio arquivo.



Foi criado um 'Manual de tratamento de dados', com o intuito de auxiliar os bibliotecários e pesquisadores na descrição e identificação dos recursos de informação.

Futuramente, o Manual será disponibilizado no formato 'HTML' dentro do próprio Repositório Institucional da FIOCRUZ, de modo a facilitar o autoarquivamento pelos seus usuários.

Ao reunir, preservar e dar visibilidade à produção técnico-científica da Fiocruz, o ARCA representa parte significativa do esforço da divulgação e acesso à pesquisa pública em saúde no Brasil em formato aberto.

### Bibliografia:

- PORTUGUAL. Secretaria-Geral do Ministério das Finanças e da Administração Pública. **O PDF/A-1 como meio de preservação de documentos eletrônicos a longo prazo**. Disponível em: <<http://www.sgmf.pt/Arquivo/Revista/Documents/OPDF.pdf>>. Acesso em: 11 nov. 2011.
- FIOCRUZ. Instituto de Comunicação e Informação Científica e Tecnológica em Saúde. Centro de Tecnologia e Informação Científica. Seção de Informação. **ARCA – Repositório Institucional da Fiocruz: manual de tratamento de dados: preenchimento dos metadados – versão beta 2.0**. Rio de Janeiro: ICICT, set. 2011.
- NARA. **Frequently Asked Questions (FAQs) about Transferring Permanent Reservations in PDF/A-1 to NARA**. Disponível em: <<http://www.archives.gov/records-mgmt/initiatives/pdf-faq.html?template=print>>. Acesso em: 11 nov. 2011.