

Curadoria Digital de Dados no Arca - Repositório Institucional da Fiocruz: Relato de Experiência

Digital Data Curation in the Arca - Fiocruz Institutional Repository: Experience Report

Comisaría Digital de Datos en el Arca - Repositorio Institucional de Fiocruz: Relato de Experiencia

Ana Maria Neves Maranhão | ana.maranhao@icict.fiocruz.br

Fundação Oswaldo Cruz. Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT). Rio de Janeiro, Brasil.

Claudete Fernandes de Queiroz | claudete.queiroz@icict.fiocruz.br

Fundação Oswaldo Cruz. Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT). Rio de Janeiro, Brasil.

Raphael Belchior Rodrigues | raphael.rodrigues@icict.fiocruz.br

Fundação Oswaldo Cruz. Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT). Rio de Janeiro, Brasil.

Resumo

A curadoria digital envolve atividades de gestão de dados e de informações que são muito importantes para a manutenção da qualidade de um repositório. Pensando nesse propósito, a equipe do Arca, Repositório Institucional da Fiocruz, estabeleceu um plano de ação para dar início ao trabalho de curadoria no repositório. O trabalho teve início em 2015 em caráter permanente, e procurou estabelecer diretrizes e padrões para o processo de gestão e de organização das informações e objetos digitais. Alguns aspectos relevantes foram levados em consideração durante o processo: alimentação descentralizada, crescimento acentuado do repositório, diversidade de conteúdo, formatos de armazenamento, validação dos metadados, utilização de padrões abertos para formatos de arquivo, procedimentos adotados, uniformização dos dados, monitoramento das comunidades e gestores, acessibilidade aos documentos, garantia de qualidade das informações, capacitação do profissional de informação/curador, inconformidades detectadas e resolução dos problemas abordados. Os resultados esperados com a realização da curadoria são: assegurar um padrão de qualidade dos registros e dos objetos digitais inseridos; fomentar ações para a preservação digital; e promover a organização de forma adequada da produção científica da Fiocruz inserida no repositório.

Palavras-chave: Arca – Repositório Institucional da Fiocruz; Curadoria de dados; Curadoria digital; Gestão da Informação; Gestores de repositórios; Bibliotecários; Profissionais de Informação.

Abstract

Digital curation involves data and information management activities that are very important in maintaining the quality of a repository. With such thought in mind, the Arca team, Institutional Repository of Fiocruz, established a plan of action to begin the curatorial work in the repository. The work began in 2015 on a permanent basis, and sought to establish guidelines and standards for managing and organizing digital information and objects. Some relevant aspects were taken into account during the process: decentralized data entry, repository acute growth, content diversity, data storage formats, validation of metadata, use of open standards for file formats, adopted procedures, standardization of data, supervising communities and managers, accessibility to documents, quality assurance of information, training of the information professional / curator, detecting nonstandard data and resolution of the arisen problems. The expected results of the curatorship are to ensure a quality standard of the records and digital objects inserted; Foster (Foment) actions for digital preservation; and promote the proper organization of the scientific production of Fiocruz inserted in the repository.

Keywords: Arca - Institutional Repository of Fiocruz; Data curation; Digital curation; Information management; Repository managers; Librarians; Information Professionals.

Resumen

La curaduría digital abarca actividades de gestión de datos y de informaciones que son muy importantes para el mantenimiento de la calidad de un repositorio. Con este objetivo, el equipo del Arca, Repositorio Institucional de la Fiocruz, estableció un plan de acción para dar inicio al trabajo de curaduría en el repositorio. El trabajo se ha iniciado en el año 2015 con carácter permanente, y ha buscado establecer directrices y padrones para el proceso de gestión y de organización de las informaciones y objetos digitales. Se consideraron algunos aspectos relevantes durante el proceso: alimentación descentralizada, crecimiento acentuado del repositorio, diversidad de contenido, formatos de almacenamiento, validación de los metadatos, utilización de padrones abiertos para formatos de archivo, procedimientos adoptados, uniformización de los datos, monitorización de las comunidades y gestores, accesibilidad a los documentos, garantía de calidad de las informaciones, capacitación del profesional de información/curador, inconformidades detectadas y solución de los problemas abordados. Los resultados esperados con la realización da curaduría son: asegurar un padrón de calidad de los registros y de los objetos digitales insertos; fomentar acciones para la preservación digital; y promover la organización de forma adecuada de la producción científica de la Fiocruz inserta en el repositorio.

Palabras clave: Arca – Repositorio Institucional de la Fiocruz; Curaduría de datos; Curaduría digital; Gestión de la Información; Gestores de repositorios; Bibliotecarios; Profesionales de la Información.

Introdução

A curadoria digitalⁱ se propõe a assegurar a pertinência das informações e preservar o objeto digital de forma que não se torne obsoleto, tendo em vista o rápido avanço tecnológico e da grande quantidade de dados armazenados nos sistemas de informação.

Atento a esse desafio, a equipe do Repositório Institucional Arcaⁱⁱ, percebeu a necessidade de iniciar um trabalho de curadoria dos dados a partir de 2015, de forma mais efetiva e sistêmica, visando promover a gestão dos dados e estabelecer diretrizes de forma a garantir um padrão de qualidade consistente.

i "A partir do ano 2003, o termo curadoria passou a ser trabalhado nas áreas de Ciência da Informação e Ciência da Computação, motivados pelo crescimento exponencial da informação digital, dando origem a noção de Curadoria Digital" (Siebra; Borba; Miranda, 2016).

ii Criado em 2007, mas lançado oficialmente em 2011 como o Repositório Institucional da Fundação Oswaldo Cruz, cuja função é "reunir, hospedar, disponibilizar e dar visibilidade à produção intelectual da Instituição, com o compromisso de proporcionar o livre acesso à informação em saúde". O Arca está sob a responsabilidade do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT) e utiliza o software livre Dspace¹.

Para o desenvolvimento do trabalho de curadoria foi preciso definir alguns critérios como: período, quantitativo de registros por comunidade e coleção, formatos de armazenamento, validação dos metadados, procedimentos adotados, uniformização dos dados, acessibilidade aos documentos, objetos digitais, inconformidades e possíveis soluções.

Metodologia

O Arca encontra-se organizado em comunidades que correspondem às unidades técnico-científicas da Fiocruz, e estão sob a gestão de profissionais de informação, previamente treinados. Cada comunidade apresenta seus documentos em diferentes coleções. Dando uma visão geral do povoamento do Arca, podemos resumir da seguinte forma: 2015 - 18 comunidades com um total de 9.427 registros; 2016 - 23 comunidades com um total de 13.038 registros; e 2017 - 25 comunidades com aproximadamente 14.000 registrosⁱⁱⁱ.

Com base nos dados levantados e de acordo com a Política de Acesso Aberto ao Conhecimento da Fiocruz^{iv}, lançada em 2014², que determina o caráter mandatório para depósitos as teses, dissertações e artigos científicos publicados, ficou definido que essas tipologias seriam as primeiras verificadas durante o processo de curadoria. A partir de 2016, outras coleções também foram verificadas.

A curadoria foi realizada nas comunidades: Casa de Oswaldo Cruz; Centro de Desenvolvimento de Tecnologia em Saúde; Escola Politécnica de Saúde Joaquim Venâncio; Fiocruz Rondônia; Fiocruz Paraná - Instituto Carlos Chagas; Fiocruz Piauí; Instituto Aggeu Magalhães; Instituto de Ciência e Tecnologia em Biomodelos; Instituto de Comunicação e Informação Científica e Tecnológica em Saúde; Instituto de Tecnologia em Fármacos; Instituto de Tecnologia em Imunobiológicos; Instituto Gonçalo Moniz; Instituto Leônidas e Maria Deane; Instituto Nacional de Controle de Qualidade em Saúde; Instituto Nacional de Infectologia Evandro Chagas; Instituto Nacional de Saúde da Mulher, da Criança e do Adolescente Fernandes Figueira; e Instituto Oswaldo Cruz.

Após o término de verificação das coleções, os gestores das comunidades receberam por e-mail um relatório informando as inconformidades detectadas e as orientações específicas para a realização dos acertos:

- Quanto ao aspecto de validação dos metadados, os acertos deveriam estar em conformidade com o “*Manual de tratamento de dados: preenchimento de metadados^v para entrada no Arca - Repositório Institucional da Fiocruz^{vi}*”, que fornece as informações visando o preenchimento correto das tipologias no Repositório.

- Quanto ao aspecto do objeto digital, os gestores foram informados, quando necessário, sobre as opções de armazenamento; questões de embargos e restrições, tipos de arquivos e a obrigatoriedade do upload do objeto digital no registro.

Considerações Finais

Segundo Sales e Sayão³

Curadoria digital envolve a gestão atuante e a preservação de recursos digitais durante todo o ciclo de vida de interesse do mundo acadêmico e científico, tendo como perspectiva o desafio temporal de atender a gerações atuais e futuras de usuários. Torna-se claro, portanto, que subjacente às metodologias utilizadas pela curadoria digital estão os processos de arquivamento digital e de preservação digital; porém, inclui também as metodologias necessárias para a criação e gestão de dados de qualidade e a capacidade de adicionar valor a esses dados no sentido de gerar novas fontes de informação e de conhecimento.

iii Dados atualizados até abril de 2017.

iv Disponível em: https://portal.fiocruz.br/sites/portal.fiocruz.br/files/documentos/portaria_-_politica_de_acesso_aberto_ao_conhecimento_na_fiocruz.pdf

v Baseado no padrão Dublin Core.

vi Disponível em: <https://www.arca.fiocruz.br/handle/icict/13257>

Desta forma, podemos afirmar que o trabalho de curadoria em desenvolvimento no Arca – Repositório Institucional da Fiocruz, vem obtendo resultados positivos como: verificação das inconformidades visando correção; aceitação por parte das comunidades como uma ferramenta de gestão de dados; monitoramento sistêmico das comunidades e das coleções relacionadas no Repositório; padronização e gestão dos metadados; avaliação contínua dos dados; e a recuperação das informações de forma mais precisa e eficiente.

A curadoria está sendo importante também para apontar ações que precisam ser realizadas como: desenvolvimento de um plano de preservação digital específico para o Arca, visando garantir a autenticidade, confiabilidade e a integridade do repositório e dos objetos digitais; desenvolvimento de um sistema de demandas específico para a gestão das curadorias realizadas; atentar para os avanços tecnológicos, principalmente no que diz respeito às atualizações e versões dos softwares e hardwares e da obsolescência tecnológica dos diversos suportes.

Referências

1. Freyre EA. Manual de tratamento de dados: preenchimento de metadados para entrada no Arca - Repositório Institucional da Fiocruz. 2.ed. Rio de Janeiro: Fiocruz/ICICT; 2016.
2. Maranhão AMN, Veiga VSO. Política de acesso aberto da Fiocruz: plano operativo: núcleo de acesso aberto ao conhecimento. Rio de Janeiro: Fiocruz/ICICT; 2014.
3. Sales LF, Sayão LF. O impacto da curadoria digital dos dados de pesquisa na comunicação científica. Enc. Bibli: Rev. Eletr. Bibliotec. Ci. Inf. 2012; 17(2): 118-135.
4. Siebra AS, Borba VR, Miranda MKFO. Curadoria digital: um termo interdisciplinar. Anais do 17. Encontro Nacional de Pesquisa em Ciência da Informação; 20–25, novembro 2016; Salvador. Bahia: UFBA, 2016. p. 1-17.