

CONTRIBUIÇÕES DO CKAN PARA A CONSTRUÇÃO DO REPOSITÓRIO DIGITAL DO OBSERVATÓRIO DO ENSINO MÉDIO INTEGRADO DO INSTITUTO FEDERAL DE SERGIPE

Contributions of CKAN for the construction of the digital repository of the Integrated High School Observatory from the Federal Institute of Sergipe
Contribuciones de CKAN a la construcción del repositorio digital del Observatorio Integrado de la Enseñanza Media del Instituto Federal de Sergipe



Carla da Conceição Andrade
Mestra em Educação, Universidade Federal de Sergipe (UFS), São Cristóvão, Sergipe, Brasil.
Técnica em Assuntos Educacionais, Instituto Federal de Sergipe (IFS), Aracaju, Sergipe, Brasil.
Lattes: <http://lattes.cnpq.br/8750728449633224>
ORCID: <https://orcid.org/0000-0003-1920-2190>



Edson Barbosa Lisboa
Doutor em Ciência da Computação, Universidade Federal de Pernambuco (UFPE), Recife, Pernambuco, Brasil.
Professor de Ensino Básico, Técnico e Tecnológico, Instituto Federal de Sergipe (IFS), Aracaju, Sergipe, Brasil.
Lattes: <http://lattes.cnpq.br/0390918018254837>



Mirer Balbino de Andrade Júnior
Graduando em Análise e Desenvolvimento de Sistemas, Instituto Federal de Sergipe (IFS), Aracaju, Sergipe, Brasil.
Lattes: <http://lattes.cnpq.br/2175100817559726>



Leila Buarque Couto de Matos
Mestra em Ciência da Computação, Universidade Federal de Sergipe (UFS), São Cristóvão, Brasil.
Professora de ensino técnico e tecnológico, Instituto Federal de Sergipe (IFS), Aracaju, Sergipe, Brasil.
Lattes: <http://lattes.cnpq.br/8856865945516149>



Roberto Julião Vieira Santos
Estudante do Curso Técnico Integrado em Informática, Instituto Federal de Sergipe (IFS),
Aracaju, Sergipe, Brasil.
Lattes: <http://lattes.cnpq.br/3250195385748501>

Resumo

Introdução: No bojo das discussões sobre os pressupostos da Ciência Aberta e do gerenciamento para a abertura de dados, apresentamos este estudo sobre a criação do repositório digital do Observatório do Ensino Médio Integrado do Instituto Federal de Sergipe (OBEMI-IFS). **Objetivo:** Para tanto, buscamos investigar as contribuições de uma ferramenta tecnológica para a criação de repositórios digitais de dados abertos, o CKAN. Este é um software livre capaz de armazenar coleções de dados, além de fornecer a infraestrutura necessária para gestão e publicação. **Metodologia:** Este estudo é de caráter descritivo e de natureza aplicada. Consiste no relato da experiência destes pesquisadores no emprego da tecnologia CKAN durante a criação da plataforma virtual do OBEMI-IFS. **Resultados:** Como resultado, percebemos que, apesar de a ferramenta ser estruturada para disponibilizar dados em forma bruta, foi possível implementar customizações de forma a adequar o CKAN para atender às nossas necessidades.

Palavras-chave: gerenciamento de dados; repositórios digitais; CKAN.

Abstract

Introduction: During discussions on the assumptions of Open Science and management for opening data, we present this study on the creation of the digital repository of the Integrated Secondary Education Observatory of the Federal Institute of Sergipe (OBEMI-IFS). **Objective:** To do so, we seek to investigate the contributions of a technological tool for the creation of digital open data repositories, the CKAN. This is free software capable of storing data collections, in addition to providing the necessary infrastructure for management and publication. **Methodology:** This study is descriptive and applied in nature. It consists of reporting the experience of these researchers in the use of CKAN technology during the creation of the OBEMI-IFS virtual platform. **Results:** As a result, we realized that, despite the tool being structured to provide raw data, it was possible to implement customizations to adapt the CKAN to meet our needs.

Keywords: data management; digital repositories; CKAN.

Resumen

Introducción: En medio de discusiones sobre los supuestos de la Ciencia Abierta y la gestión para la apertura de datos, presentamos este estudio sobre la creación del repositorio digital del Observatorio Integrado de Educación Secundaria del Instituto Federal de Sergipe (OBEMI-IFS). **Objetivo:** Para ello, buscamos investigar los aportes de una

herramienta tecnológica para la creación de repositorios digitales de datos abiertos, el CKAN. Este es un software libre capaz de almacenar colecciones de datos, además de proporcionar la infraestructura necesaria para su gestión y publicación. **Metodología:** Este estudio es de carácter descriptivo y aplicado. Consiste en relatar la experiencia de estos investigadores en el uso de la tecnología CKAN durante la creación de la plataforma virtual OBEMI-IFS. **Resultados:** Como resultado, nos dimos cuenta de que, a pesar de que la herramienta estaba estructurada para proporcionar datos sin procesar, era posible implementar personalizaciones para adaptar el CKAN a nuestras necesidades.

Palabras clave: gestión de datos; repositorios digitales; CKAN.

1. Introdução

Os repositórios digitais têm-se tornado importantes dispositivos de salvaguarda e disseminação da produção científica em todo o mundo, contribuindo com os pressupostos universais para a Ciência Aberta como a transparência, o compartilhamento e a colaboração (ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, 2022). Entretanto, a implantação destes dispositivos demanda cuidados, dentre os quais, estão a observância aos modelos de referência e princípios estabelecidos internacionalmente, a garantia do ciclo de vida dos dados e os recursos necessários para o desenvolvimento dessas estratégias.

Neste estudo, buscamos investigar um desses aspectos, que é o da infraestrutura tecnológica. Mais especificamente, pretendemos apresentar as contribuições e os desafios encontrados com o uso de uma ferramenta livre para a criação de repositórios digitais de dados abertos: o *Comprehensive Knowledge Archive Network* (CKAN). Trata-se de um software de gerenciamento de dados abertos, desenvolvido pela *Open Knowledge Foundation* e que foi escolhido como o sistema de gerenciamento do Portal Brasileiro de Dados Abertos.

Por esse motivo, o CKAN foi escolhido para ser a tecnologia empregada na criação do Observatório do Ensino Médio Integrado do Instituto Federal de Sergipe (OBEMI-IFS), uma plataforma virtual de monitoramento e disseminação do conhecimento produzido por professores e estudantes no desenvolvimento das práticas pedagógicas de formação profissional desenvolvidas nos cursos técnicos na forma integrada dessa instituição.

Aqui, relatamos a experiência de empregar o CKAN para gerenciar os dados dispostos no OBEMI-IFS. Nosso intuito é demonstrar as potencialidades da ferramenta, mas também, os desafios que enfrentamos para implementar a plataforma. Para tanto, organizamos o trabalho em três seções, além desta introdução: a descrição dos procedimentos metodológicos empregados; o relato propriamente dito, onde descrevemos as funcionalidades do sistema que contribuíram para a sua escolha, o potencial do software e as possibilidades de adequação; além das considerações.

Esperamos que este estudo contribua para a ampliação do debate acadêmico sobre a utilização de softwares de gerenciamento de dados, especialmente, sobre o CKAN, visto que a produção científica sobre essa ferramenta ainda é bastante incipiente.

2. Procedimentos Metodológicos

Este é um estudo qualitativo e descritivo, de natureza aplicada. Refere-se ao relato da experiência destes pesquisadores no emprego da tecnologia CKAN durante a criação da plataforma virtual do OBEMI-IFS. Entendemos o relato de experiência como “uma vivência acadêmica e/ou profissional em um dos pilares da formação universitária (ensino, pesquisa e extensão), capaz de contribuir na produção do conhecimento das mais variadas temáticas” (MUSSI; FLORES; ALMEIDA, 2021), visto que representa “um momento importante para a articulação dessa experiência e registro do processo e das interpretações para aprendizagem do grupo que a protagonizou” (LÜDKE; CRUZ, 2010).

O estudo correspondeu a vivência dos pesquisadores durante o desenvolvimento de um projeto de pesquisa denominado “Construção do Observatório do Ensino Médio Integrado do Instituto Federal de Sergipe: uma plataforma digital indutora do currículo e do desenvolvimento regional”, iniciado no segundo semestre de 2022, após a aprovação pelo Edital nº. 03/2022/PROPEX/IFS, e que ainda está em andamento. Os dados utilizados foram coletados durante as reuniões da equipe do projeto, por meio do instrumento diário de bordo, onde foram anotados os desafios e as dúvidas encontrados pela equipe de desenvolvimento.

3. Resultados

3.1 Conhecendo o CKAN

Como mencionado anteriormente, o CKAN é uma ferramenta livre de licenças de uso, criado para catalogação e gerenciamento de dados abertos. É um software programado para “armazenar coleções organizadas de dados e fornecer ferramentas de gestão, recuperação e apresentação de bases de dados” (COSTA et al., 2017, p. 14-15).

Para tanto, a ferramenta possui um arcabouço de funcionalidades que possibilitam aos seus usuários a publicação e gestão de dados com qualidade e garantia de reuso. Dentre essas funcionalidades, destacamos:

- Disposição de metadados: a unidade informacional do CKAN é o *dataset* (conjunto de dados), composto por um ou mais dados e seus conjunto de metadados, que são as informações sobre os dados. O CKAN possui um conjunto de metadados padrão (título, descrição, etiquetas, licença...) e que pode ser customizado.
- Diversidade no formato: o CKAN permite o carregamento do arquivo do recurso em diferentes formatos: JSON, CSV, XLS, PDF, TXT... O carregamento pode ser feito tanto por arquivo no computador quanto por link disponível em endereço web.
- Autocadastramento: a ferramenta possibilita que o usuário realize o seu cadastro e tenha acesso ao conjunto de informações, podendo, também criar e gerenciar dados dentro do sistema. No entanto, as configurações do que o usuário cadastrado pode fazer são controladas pelo administrador do sistema. Assim, o sistema permite liberdade e privacidade ao usuário, mas também oferece segurança à organização, através dos controles do administrador e das permissões exclusivas por usuário.
- Gestão de conteúdo: o CKAN permite as funcionalidades de adição, alteração, exclusão e reuso dos dados.
- Customização da interface: é possível alterar a aparência do sistema, personalizando-o com as cores ou logomarca institucionais, ou ainda, adicionando um texto de apresentação do sistema. Além disso, fornece recursos para personalizar a experiência de busca e publicação de dados.
- Recuperação de dados: visando o atendimento aos princípios FAIR (*Findable, Accessible, Interoperable and Reusable*), este é um dos principais serviços do CKAN. Há diversas possibilidades, como uma busca simples ou personalizada, visualização dos dados de acordo com o seu formato, ou por georreferenciamento, caso possua metadados de localização geográfica.

- Dimensão global: o CKAN é o sistema de gerenciamento de dados mais empregado no mundo, sendo utilizado por diversas organizações governamentais. No Brasil, é o software empregado no Portal de Dados Abertos do Governo Federal e também para a divulgação de dados abertos do IFS.

Este compilado de funcionalidades corroborou para a escolha da ferramenta como sistema de gerenciamento do OBEMI-IFS. Ainda que o CKAN seja um sistema para gestão de dados brutos, em sua dinâmica, favorece os princípios FAIR, o que possibilita a preservação dos dados em todo o seu ciclo de vida. Além disso, possui capilaridade em nível mundial, possibilitando a customização para atender a diferentes demandas. No próximo tópico, descreveremos como esta ferramenta vem contribuindo para a criação da plataforma virtual.

3.2 A construção do OBEMI-IFS e o suporte do CKAN

Nosso interesse em implantar um observatório sobre a formação profissional desenvolvida nos cursos técnicos surge com a necessidade de acompanhar o desenvolvimento das atividades pedagógicas e processos avaliativos voltados para a superação da fragmentação curricular nestes cursos. O OBEMI-IFS é pensado, então, como um instrumento de *accountability*, ou seja, de acompanhamento e monitoramento, com foco nas práticas profissionais e na integração curricular. Essa característica é o eixo para o desenvolvimento do software.

Durante a definição dos requisitos, estabelecemos a plataforma como um repositório digital temático, ou seja, “um ambiente de armazenamento e gerenciamento de materiais digitais” (CONARQ, 2015, p.09), “que reúne a produção científica de uma área” (WEITZEL, 2006, p. 59). A pretensão era reunir dados sobre as práticas desenvolvidas pelos cursos, além de estudos e normativos sobre o tema do ensino médio integrado. Todavia, não havíamos definido como os diversos dados ficariam dispostos na plataforma. O modelo de ciclo de vida recomendado pelo *Digital Curation Center (DCC)* foi fundamental para a condução dos trabalhos.

Este modelo possibilita o mapeamento das funcionalidades do sistema, a definição de papéis e responsabilidades, além da identificação de padrões e de tecnologias a serem utilizadas.

The DCC Curation Lifecycle Model offers a graphical high-level overview of the lifecycle stages required for successful curation. Generic in nature, the model is indicative rather than exhaustive. When used as an organizational planning tool, it is adaptable to different domains, and extensible to allow curation and preservation activities to be planned at different levels of granularity (HIGGINS, 2008, p. 135)¹.

A partir do modelo do DCC, definimos como primeiro passo, a atribuição de metadados, que são informações sobre os recursos estruturais e que possibilitam a identificação, a gestão, a recuperação e a interoperabilidade dos dados. Foi durante o processo de atribuição de metadados que tivemos contato com a iniciativa de transfor-

¹ O Modelo de ciclo de vida do DCC oferece uma visão geral gráfica de alto nível dos estágios do ciclo de vida necessários para uma curadoria bem-sucedida. De natureza genérica, o modelo é indicativo e não exaustivo. Quando usado como uma ferramenta de planejamento organizacional, é adaptável a diferentes domínios e extensível para permitir que as atividades de curadoria e preservação sejam planejadas em diferentes níveis de granularidade.

mação digital do Governo Federal, o GOV.BR, e com o CKAN, como sistema de gerenciamento dados empregado na iniciativa. Como o próprio sistema já possuía um conjunto padrão e o Governo Federal constituiu, também, um padrão de metadados, o e-PMG, decidimos que o OBEMI-IFS empregaria esses modelos e realizaria as adequações necessárias.

O CKAN permite a customização do seu conjunto padrão de metadados, mas, de forma limitada. Para resolver o impasse, utilizamos o API ckanext-scheming (<https://github.com/ckan/ckanext-scheming>).

Quadro 1 - Relação de Metadados do OBEMI-IFS, por tipo, nome e definição

| TIPO | NOME | DEFINIÇÃO |
|--------------------|---------------------|--|
| Técnico-descritivo | Título | Grupo de caracteres que nomeia um recurso. |
| | Descrição (Resumo) | Explicação do conteúdo do recurso. |
| | Identificador | URL exclusivo que identifica cada pacote de dados |
| | Autor (Criador) | Uma entidade (pessoa ou organização) originalmente responsável pela criação e manutenção do conteúdo do recurso. |
| | Etiquetas (Assunto) | Palavras-chaves ou termos que representem corretamente o conteúdo do recurso. |
| | Catálogo de origem | Página (URL) onde está publicado o conjunto de dados. |
| | Formato | Conjunto de características físicas e lógicas que define a apresentação do recurso. Ex: PDF/A; JSON; JPEG. |
| | Tipo | Natureza ou gênero do conteúdo do recurso. Ex: projeto, relatório, tese, fotografia. |
| | Data | Uma data associada a um evento no ciclo de vida do recurso. Ex: data.publicação; data.submissão. |
| | Licença | Informação relativa às permissões sobre o acesso e o uso ao recurso |

| TIPO | NOME | DEFINIÇÃO |
|------------|-------------------------|--|
| Específico | Campus | Unidade administrativa no âmbito da instituição em que o dado foi produzido. |
| | Curso | Unidade didático-pedagógica no âmbito da instituição em que o dado foi produzido. |
| | Área do conhecimento | Taxonomia elaborada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico para a organização da produção do conhecimento. |
| | Eixo Tecnológico | Taxonomia de organização dos cursos técnicos e tecnológicos conforme suas características científicas e tecnológicas. |
| | Arranjo Produtivo Local | Informação sobre qual arranjo produtivo da região em que está inserido o dado se refere. |
| | Fator de Inovação | Taxonomia elaborada pela equipe gestora do OBEMI-IFS que indica a tipologia de inovação presente no dado. |

Fonte: Dados da pesquisa (2023).

Para guiar a prototipagem da plataforma, nós desenvolvemos também um documento inicial, denominado Manual Descritivo do OBEMI-IFS. Nele, encontram-se os objetivos estratégicos do observatório, a composição do acervo e seu conjunto de metadados, bem como o detalhamento dos requisitos funcionais da plataforma. Destacamos que a construção do Manual, a partir do estabelecimento dos requisitos essenciais, foi realizada de acordo com as funcionalidades do CKAN, adaptando-as, quando necessário, para a realidade e objetivos do OBEMI-IFS.

Os próximos passos serão o desenvolvimento do projeto de interface e a aplicação dos testes de usabilidade e desempenho. Ainda que o sistema permita as customizações de fontes, cores e visualização, é de nosso interesse elaborar uma interface própria, respeitando os aspectos já existentes na ferramenta.

4. Conclusão ou Considerações Finais

Neste estudo, buscamos demonstrar as contribuições da ferramenta CKAN para a construção da plataforma virtual do Observatório do Ensino Médio Integrado do IFS. Dentre os aspectos do software que buscamos destacar, estão a possibilidade de uma gestão dos dados ao longo do seu ciclo de vida, a disposição de conjunto de metadados padrão, a facilidade de customização das funcionalidades e a sua dimensão global, com presença em diversos países e organizações, sendo a ferramenta eleita pelo Governo Brasileiro para a abertura dos seus dados.

O uso da ferramenta também apresentou desafios, visto que seu propósito é de salvaguarda de dados brutos. Mesmo assim, estamos constantemente buscando a aplicação de customizações visando adaptar o CKAN para atender às necessidades e objetivos do OBEMI-IFS.

5. Referências

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Diretrizes para a implementação de repositórios arquivísticos digitais confiáveis – RDC-Arq**. Brasília: CONARQ, 2015. Disponível em: <https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-no-43-de-04-de-setembro-de-2015>.

Acesso em: 19 setembro 2022.

COSTA, Lucas Rodrigues. SHINTAKU, Milton. SILVEIRA, Lucas Angelo. MACÊDO, Diego José. FONSECA, Ramón Martins Sodoma da. **Guia do usuário CKAN**. Brasília: IBICT, 2017.

HIGGINS, Sarah. The DCC Curation Lifecycle Model. **The International Journal of Digital Curation, Aberystwyth**, v. 03, n. 01, p. 34-140, jun., 2008. Disponível em: <http://www.ijdc.net/article/view/69>. Acesso em: 24 maio 2023.

LÜDKE, Menga. CRUZ, Gisele Barreto da. Contribuições ao debate sobre a pesquisa do professor da educação básica. **Formação Docente**, Belo Horizonte, v. 02, n. 03, p. 86-107, ago./dez., 2010. Disponível em: <http://formacaodocente.autenticaeditora.com.br>. Acesso em: 24 maio 2023.

MUSSI, Ricardo Franklin de Freitas. FLORES, Fábio Fernandes. ALMEIDA, Cláudio Bispo de. Pressupostos para a elaboração de relato de experiência como conhecimento científico. **Práxis Educacional**, Vitória da Conquista, v. 17, n. 48, p. 60-77, 2021. Disponível em: <https://doi.org/10.22481/praxisedu.v17i48.9010>. Acesso em: 24 maio 2023.

WEITZEL, Simone da Rocha. O papel dos repositórios institucionais e temáticos na estrutura da produção científica. **Em Questão**, v. 12, n. 1, p. 51-71, 2006. Disponível em: <https://www.redalyc.org/pdf/4656/465645954004.pdf>. Acesso em: 24 maio 2023.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO. **Recomendação da UNESCO sobre Ciência Aberta**. Brasília: UNESCO, 2022. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000379949_por. Acesso em: 24 maio 2023.