

Especificação de um Repositório de Dados Confiável para Experimentos de Engenharia de Software

Pedro Henrique de Melo Costa (PIBIC/CNPq/FA/UEM), Edson A. Oliveira Junior (Orientador), e-mail: ra112653@uem.br, edson@din.uem.br

UEM/CTC/DIN

**Área: Ciências exatas e da Terra,
Subárea: Ciência da Computação**

Palavras-chave: Repositório, Open Data, Open Science

Resumo

Open Science representa uma nova abordagem ao processo científico baseado em trabalho cooperativo e várias formas de difundir o conhecimento usando tecnologias digitais e novas ferramentas colaborativas. Uma possível definição da European Commission é “tornar resultados primários de pesquisas com financiamento público acessível em formato digital mínima ou nenhuma restrição”. Para isso, é necessário estender os princípios de openness a todo ciclo de pesquisa, buscando compartilhar e colaborar o mais cedo possível para permitir mudanças sistemáticas na forma como a ciência e a pesquisa são realizadas. Dentre as facetas de Open Science, destacam-se: Open Notebooks, Open Data, Open Peer Review, Open Access, Open Source e Open Educational Resources. A faceta de open Data engloba diversos conceitos como ciclo de vida de dados e repositórios. Os repositórios de dados abertos são essenciais para que as novas tecnologias possam ser usadas em benefício dos pesquisadores e profissionais da indústria. Vários repositórios, de cunho geral, têm se destacado para que pesquisadores possam compartilhar seus dados, como: Zenodo, arXiv e figshare. Este projeto tem como objetivo especificar um repositório de dados para experimentos de Engenharia de Software.

Introdução

Inicialmente, após análise dos TDRs existentes e as suas estruturas e mecanismos de compartilhamento de dados abertos considerando as características de dados de experimentos de engenharia de software, é selecionado e especificado um repositório digital confiável e então feita sua análise comparando sua estrutura com as normas existentes para especificação de repositórios de dados abertos.

Materiais e Métodos

Como metodologia, foi realizada uma revisão de literatura sobre os temas open science e open data, além dos conceitos de experimentação em engenharia de software. Também foram analisados os resultados de um projeto PIBIC, “Uma análise comparativa de repositórios para compartilhamento de Open Data de experimentos de engenharia de software”, que compara vários repositórios de dados abertos. Feito isso, foi realizado um estudo sobre normas de especificação de TDRS.

Após os estudos, foi especificado o repositório Zenodo para uma análise detalhada, comparando sua estrutura com os requisitos de um repositório digital confiável.

Resultados e Discussão

A implantação de um repositório digital confiável tem o propósito de garantir a preservação, acesso e autenticidade em longo prazo dos dados. Para isso, ao longo dos anos foram criadas diversas diretrizes para padronizar a implantação dos repositórios digitais. As primeiras iniciativas de padronizar a modelagem e implementação de repositórios digitais datam da década de 1990, onde em 1996 foi publicado o relatório *Task Force on Archiving of Digital Information*, pela RLG e pela *Commission on Preservation and Access*, visando a apreensão com a confiabilidade dos repositórios digitais. Já em 2003, foi elaborada a norma ISO 14721:2003 que descreve as funções de um repositório digital e os metadados usados para preservar e acessar os arquivos digitais gerenciados pelo repositório. Posteriormente foi elaborada a norma ISO 16363: 2012, que lista os critérios que um repositório digital confiável deve satisfazer.

Dito isto, segundo o relatório “Trusted Digital Repositories: attributes and responsibilities”, temos que os repositórios digitais confiáveis devem satisfazer os seguintes requisitos:

- Responsabilidade administrativa
- Viabilidade organizacional
- Sustentabilidade financeira
- Adequação tecnológica e processual
- Sistema de segurança
- Responsabilidade processual

Responsabilidade administrativa: A responsabilidade administrativa se estende ao cumprimento de normas nacionais e/ou internacionais quanto ao ambiente físico do repositório, e o repositório Zenodo provê evidências de comprometimento quanto a viabilidade e sustentabilidade organizacional, devido ao comprometimento com a

preservação, manutenção e recuperação dos dados nele armazenados. Contudo, o repositório Zenodo não envolve especialistas externos da comunidade científica para a validação nem certificação de seus processos e procedimentos, sendo isso uma das responsabilidades administrativas do open science.

Viabilidade Organizacional: O repositório Zenodo é hospedado pelo CERN (grande colisor de hádrons da organização europeia de pesquisas nucleares), uma organização intergovernamental e pioneira no trabalho com Open Access, atualmente o Zenodo é embutido no departamento de tecnologia da informação, na seção de repositórios digitais. Zenodo é oferecido pelo CERN como parte de sua missão de disponibilizar os resultados de seu trabalho.

Sustentabilidade Financeira: O repositório Zenodo é financiado pelos órgãos:

- Comissão Europeia através dos projetos OpenAIRE; FP7: OpenAIRE, OpenAIREplus, projetos de repositórios de informação científica de apoio ao programa europeu de pesquisa FP7 e Horizon 2020
- CERN
- Fundação Alfred P. Sloan
- Arcadia Fund
- Doações via CERN & Society Foundation

Para reduzir os custos operacionais, o repositório Zenodo é hospedado na infraestrutura já existente do CERN (juntamente com seus serviços). Além disso, segundo o repositório Zenodo, o CERN consta com os melhores especialistas possíveis na execução de infraestruturas de dados de pesquisa em grande escala e repositórios digitais, dos quais fornecem uma base sólida para um repositório digital confiável.

Sistema de Segurança do ZENODO: Os centros de dados do CERN estão localizados nas instalações do CERN e o acesso físico a essa área é restrito a um número limitado de funcionários com treinamento apropriado e que tiveram autorização para o acesso de acordo com suas funções profissionais.

Os servidores são gerenciados de acordo com a linha de base de segurança CERN para servidores, isso significa que diferentes grupos atuam em diferentes áreas dos servidores, com funções fixas. Por exemplo, o acesso remoto aos servidores do CERN é restrito à equipe da Zenodo com treinamento apropriado. O sistema operacional e os aplicativos instalados são sempre mantidos atualizados com os mais recentes patches de segurança através de um sistema de gerenciamento de configuração automático Puppet.

Quanto à rede do CERN, a equipe de segurança do CERN executa sistemas de detecção de intrusão baseados em host e rede, além de monitorar o fluxo de tráfego, padrão e conteúdo nas redes do CERN. O acesso ao zenodo.org acontece inteiramente através de protocolo HTTPS, considerada atualmente como a versão de protocolo mais segura, com exceção apenas para páginas de documentação estática hospedadas nas páginas do GitHub.

Já quanto a segurança relativa às senhas dos usuários, o Zenodo armazena-as usando algoritmos criptográficos fortes de hashing (atualmente PBKDF2+SHA512). Os tokens de acesso dos usuários ao GitHub e ORCID são armazenados criptografados e só podem ser descriptografados com a chave secreta do aplicativo.

A equipe do Zenodo que possui acesso aos dados dos usuários tem uma série de protocolos de segurança para lidar com os dados, onde o acesso aos dados deve ser sempre consistente com as responsabilidades profissionais dos mesmos, além disso, o acesso aos dados pela equipe só é permitido para a resolução de problemas, detecção de questões de segurança e monitoramento de recursos.

Conclusões

Com base nos dados expostos na sessão de resultados, concluímos que o repositório Zenodo satisfaz todos os requisitos de um repositório digital confiável. Sendo eles a responsabilidade administrativa, viabilidade organizacional, sustentabilidade financeira, adequação tecnológica e processual, sistema de segurança e responsabilidade processual.

Agradecimentos

Os autores agradecem à Fundação Araucária e ao CNPq pelo suporte financeiro.

Referências

- WOHLIN, C, et al. **Experimentation in Software Engineering**. Springer ed. 2012
- FUENTE, G. B. **Foster Open Science. What is Open Science? Introduction**.
- OCLC. **Trusted Digital Repositories: Attributes and Responsibilities - An RLG-OCLC Report**.