

Tratamento de dados FAIR no Projeto Laguna

Washington Luís Ribeiro de Carvalho Segundo, Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)
Adilson Luiz Pinto, Universidade Federal de Santa Catarina (UFSC)
Fabio Lorensi do Canto, Universidade Federal de Santa Catarina (UFSC)
Patricia da Silva Neubert, Universidade Federal de Santa Catarina (UFSC)

Introdução

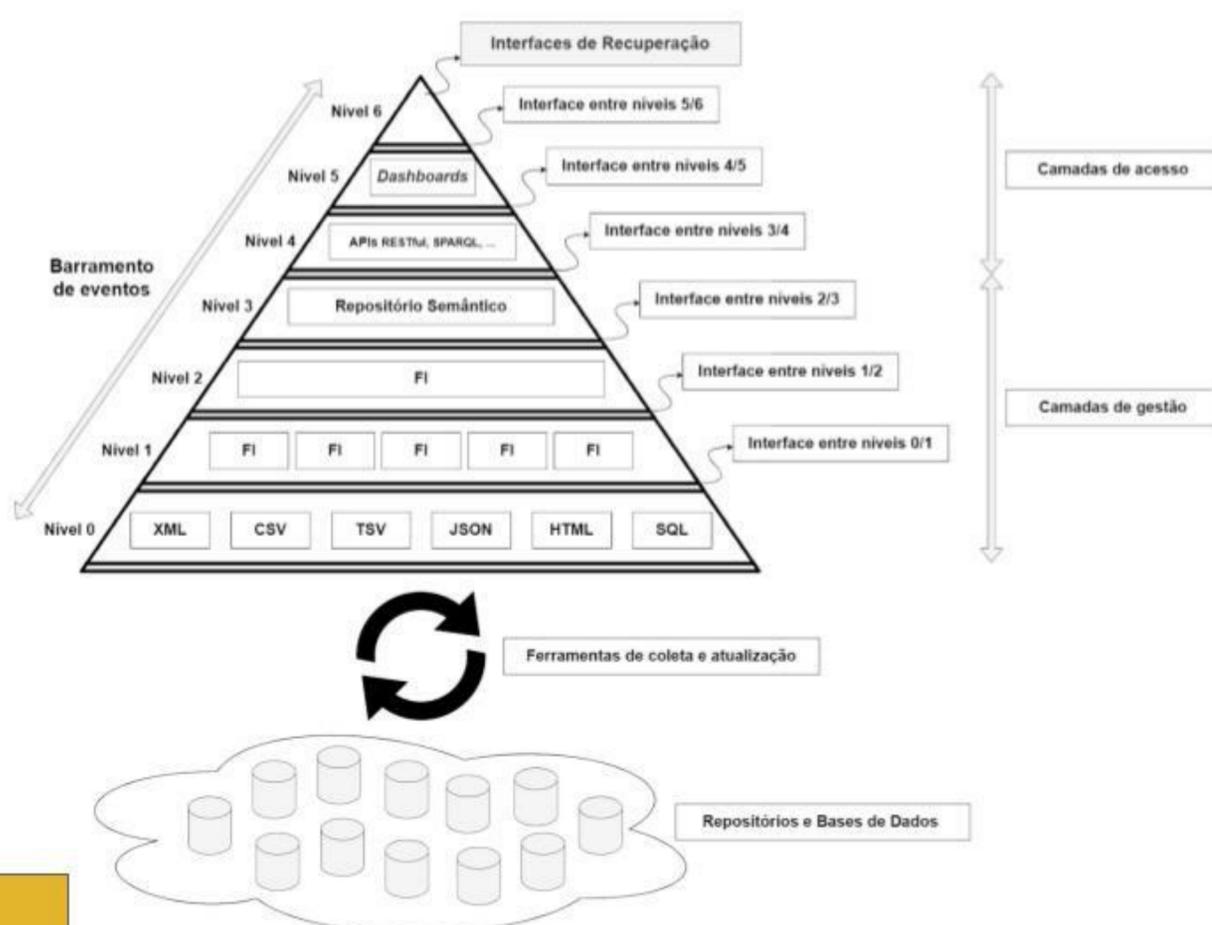
O Laguna é um projeto em andamento que visa a criação de uma Infraestrutura Informacional Aberta (IIA) de dados científicos que estejam de acordo com os princípios FAIR - Findable, Accessible, Interoperable, Reusable. Sua infraestrutura é organizada na forma de um lago de dados (*data lake*).

Metodologia

O conjunto de rotinas computacionais, representado na Figura 1, mostra uma estrutura alimentada por dados coletados no “mundo exterior”, ou seja, os repositórios e as bases de dados, composta por interfaces de processamento de dados em seis níveis. O processamento de dados ocorre tanto em uma infraestrutura servidores *on premise* quanto em nuvem (*Laguna Cloud*).

A metodologia segue as seguintes etapas de processamento:

- Coleta dos dados (Nível 0);
- seleção e separação dos conteúdos capturados (Nível 1 e 2);
- operações de transformação, vinculação e enriquecimento dos dados (Nível 2 e 3);
- Organização e indexação dos registros (Nível 3);
- recuperação e visualização da informação indexada (Nível 4,5 e 6).



Resultados esperados

Ao promover a integração de dados científicos de variada fonte o Laguna torna possível o desenvolvimento de uma variedade de serviços de informação científica a partir da integração, relacionamento e tratamento dos dados disponibilizados em outras fontes de informação.

Apoio