

Relato Técnico • Proposta de *framework* de internacionalização no contexto da evolução da plataforma Lattes

José Francisco Salm Junior, Paulo Henrique de Assis Sant’ana e Geraldo Sorte

Evolução do Lattes, Fontes de Dados Curriculares, Ontologias de Currículo, Internacionalização, *Framework* de Cooperação

INTRODUÇÃO

Em um esforço para evoluir iniciativas ligadas à gestão da informação em ciência e melhorar os procedimentos de extração de dados na Plataforma Lattes e sua interoperabilidade com fontes de dados curriculares, o Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq e a Fundação para a Ciência e a Tecnologia – FCT de Portugal estabeleceram, em novembro de 2021, um acordo para um programa de cooperação científica, tecnológica e de inovação focado no desenvolvimento conjunto de atividades de pesquisa e outras iniciativas em temas de comum interesse. Dentre as intenções do Acordo de Cooperação Internacional (ACI), previa-se o intercâmbio de pesquisadores(as) nos temas de gestão de dados científicos e repositórios de dados em pesquisas correntes.

Neste contexto, apresenta-se nesse relato técnico a proposta de intervenção, denominada *framework* CV-PTBR, que norteará o trabalho de assimilação de tecnologias e seus elementos para o uso nos projetos de aprimoramento das rotinas de importação de dados na Plataforma Lattes e a evolução das integrações de dados na Plataforma CIÊNCIAVITAE de Portugal. Tal *Framework* objetiva estabelecer um protocolo de trabalho conjunto, baseado em diretrizes, entre as equipes do CNPq e FCT. No caso do Currículo Lattes

(CV Lattes), a possibilidade de aprimoramento com a adoção do *Framework* dar-se-á pela apresentação de protótipo de interfaces e a proposição de cenários, em especial relativos à melhoria das rotinas de dados e ao uso de padronização internacional para sua manutenção e atualização.

No início de abril de 2022, iniciou-se pesquisa no âmbito do ACI, cujo escopo de atividades tratou da avaliação das ontologias de currículos e de cenários de evolução do CV Lattes, concentrando esforços na identificação e no estudo de modelos internacionais (i.e. VIVO *Ontology*¹, CERIF² e ORCID³) como potenciais referências para o projeto de evolução do modelo ontológico do referido CV Lattes. A pesquisa também contemplou o estudo de propostas de modelos narrativos de *Curricula Vitae* como um elemento de evolução do CV Lattes para aprimoramento de seu uso em procedimentos de avaliação ad hoc e do uso de recursos de extração de dados para o apoio à criação das narrativas, tanto para quem preenche seu CV quanto para as pessoas responsáveis por sua avaliação. Complementar ao plano de trabalho da pesquisa, elaborou-se um estudo com base no reaproveitamento de ativos colocados à disposição dos envolvidos no ACI. Esse estudo produziu recomendações e a proposição do *framework* CV-PTBR para trabalho conjunto, vi-

1. Conforme o site da VIVO, a Ontologia VIVO fornece um conjunto de tipos (classes) e relações (propriedades) para representar os dados sobre pesquisadores e o contexto em que trabalham. A Ontologia VIVO está vinculada a outras Ontologias, que fornecem um contexto complementar e significado em suas relações e tipos. Informações adicionais estão em: <https://wiki.lyrasis.org/display/VIVODOC111X/Ontology+Reference>.

2. Conforme o site do CERIF, o Formato Europeu Comum de Informação sobre a Pesquisa (CERIF) é o modelo de informação abrangente para o domínio da investigação científica. Destina-se a apoiar o intercâmbio de informações de pesquisa entre e com os Sistemas de Informação Sobre Pesquisas Correntes (CRIS). Mais informações estão disponíveis em: <https://eurocris.org/services/main-features-cerif>.

3. Conforme o site do ORCID, esse serviço serve como um centro de informações, permitindo que autores e revisores se conectem de forma confiável às suas contribuições e compartilhem informações de seus currículos registrados enquanto eles interagem com seu sistema. Mais informações estão disponíveis em: <https://info.orcid.org/what-is-orcid/>.

sando a sustentabilidade e a coprodução de ativos de software que serão adotados pelas organizações envolvidas no ACI.

Os elementos utilizados no diagnóstico incluíram o modelo de fontes e bases importadas ou sincronizadas pela Plataforma CIÊNCIAVITAE, a forma de estruturação dessas operações de dados e o modelo de resolução de conflitos ou inconsistências após a execução dessas rotinas. A proposta do integrador e sincronizador da Plataforma CIÊNCIAVITAE é semelhante a um pool de serviços de integração. Conceitualmente, se estabelece um barramento de dados, cujo objetivo é permitir interoperabilidade técnica, não tendo necessariamente base em um *data lake*.

Os objetivos esperados ao se adotar o *framework* CV-PTBR proposto nesse relatório são, além dos benefícios do modelo de internacionalização, o aumento e a simplificação do reaproveitamento dos ativos de interoperabilidade técnica. Também se espera, com a adoção do *framework* CV-PTBR, que se descentralizem as iniciativas de evolução das tecnologias ligadas ao currículo e a melhoria do compartilhamento dos dados associados aos currículos nos respectivos repositórios, tanto no contexto nacional, quanto no contexto do Acordo Binacional Brasil-Portugal.

Caso seja uma decisão dos representantes do país adotante do *Framework*, esse poderá apoiar o envolvimento de técnicos nas universidades para que contribuam com a evolução da tecnologia, provendo também apoio nas atividades de documentação técnica e no material de gestão de dados para os recursos comuns derivados do *framework* CV-PTBR.

Entende-se que os beneficiários deste *framework* terão diferentes expectativas em relação ao valor produzido por esse modelo. Inicialmente, foram considerados beneficiários todos aqueles representantes ligados à gestão da informação em ciência, tecnologia e inovação. De forma mais ampla, consideraram-se também os representantes da sociedade que desejem maior transparência das informações relativas aos resultados obtidos em decorrência do investimento público em formação e na realização de pesquisa. Dependendo da política de inovação estabelecida pelo país adotante do *Framework*, podem ser considerados beneficiários os representantes do segundo setor que desejarem prover serviços com os dados disponibilizados pela adoção do futuro *framework* CV-PTBR.

Diferentes representantes do governo em suas variadas esferas (i.e. nacional, regional ou estadual) terão a oportunidade de desenvolver iniciativas próprias de uso dos ativos de dados derivados do *framework* CV-PTBR. No caso do Brasil, esses ativos poderão ser disponibilizados aos agentes estaduais de CT&I, como as Fundações de Apoio à Pesquisa, ou no contexto Federal, bem como a outras agências de fomento que desejarem os dados para uso em seus processos internos. Aos agentes públicos será possível coordenar aplicações entre órgãos no uso desses ativos ou serviços, em um ciclo continuado de melhoria desses ativos decorrentes da aplicação do *Framework* proposto.

A seguir são apresentadas diretrizes de adoção do *Framework*, que deverão apoiar a tomada de decisão em relação ao seu uso nas ações de evolução dos ativos de software em plataformas de informação curricular.

Diretrizes da Adoção do *framework* CV-PTBR

As diretrizes para a adoção do *framework* representam elementos norteadores para projetos de manutenção dos ativos produzidos durante os ciclos compartilhados de engenharia das plataformas de informação curricular. As seis diretrizes reforçam as práticas de independência na execução dessas ações e a autonomia das equipes em cada país, estabelecendo acordos que apoiarão a condução dessas iniciativas conjuntas.

Assimilação e Adoção dos ativos produzidos a qualquer tempo

A criação de artefatos feita por uma organização, com base no *Framework*, deve permitir seu uso a qualquer tempo por parte das entidades participantes da cooperação, garantindo que esse artefato seja suficientemente modularizado para encaixe e agregação em ativos existentes na plataforma de informação curricular adotante.

Planejamento e Gestão de Projeto Desvinculados

A assimilação e adoção dos ativos não está associada a questões temporais ou à dependência de cronogramas entre as equipes participantes da cooperação. A adoção do *Framework* e seus elementos não obriga o alinhamento de cronogramas entre as organizações participantes.

Priorização de reaproveitamento dos ativos

No contexto da cooperação, os adotantes do *framework* poderão manifestar seu interesse em priorizar o reaproveitamento de novos ativos. Essa manifestação de interesse não gera obrigações aos demais participantes da cooperação em relação ao cronograma, como apontado pela diretriz de Planejamento e Gestão de Projeto Desvinculados.

Investimento e financiamento independentes

A organização participante da cooperação terá independência na execução de projetos para implementação de ativos compartilhados e cabe somente a essa participante a definição do volume de investimento em projetos, respeitando sua condição e autonomia de execução orçamentária.

Priorização da adoção de padrões internacionais

A organização participante da cooperação, sempre que possível, priorizará o uso de padrões internacionais adotados no contexto da informação curricular. Esses padrões podem representar ontologias, protocolos de comunicação, harvesting e padrões de identificadores persistentes entre outros.

Respeito à realidade cultural e nacional

Considerando que cada país estabelece suas normas e legislações, possuindo elementos culturais próprios, a adoção do *Framework* deve reforçar a configurabilidade e adaptabilidade dos ativos compartilhados, de modo a permitir sua configuração dentro de rotinas de operação da organização participante. Dentre esses aspectos, pode-se destacar a legislação sobre Acessibilidade Web, Lei de Proteção de Dados, Design systems locais e outros regramentos estabelecidos por iniciativas de governo eletrônico.

Estas diretrizes devem promover a redução de riscos relacionados à dependência de cronogramas, ao mesmo tempo que aumentam a autonomia de decisão na implementação das demandas regionais, sem obrigar o comprometimento prévio de orçamento por parte das organizações adotantes do *Framework*.

Na próxima sessão serão apresentados os elementos componentes do *Framework* de localização, as características do programa de construção do *framework* e suas fases.

Elementos do Framework de CV de língua Portuguesa – CV-PTBR

A internacionalização é entendida como um processo integral, onde a perspectiva universal permeia as informações curriculares, em especial aquelas ligadas às atividades acadêmicas por conta de sua natureza, abrangendo desde a pesquisa, os programas acadêmicos, a inovação e extensão, a formação continuada de professores e a evolução das práticas profissionais.

Características do Programa de Construção do framework CV-PTBR

A adoção do *framework* CV-PTBR será reavaliada a cada ciclo de tempo, permitindo assim a evolução desse modelo e a proposição de melhorias por parte das organizações que o adotarem. A possibilidade de execução em fases permitirá validar o modelo com casos práticos de criação de ativos baseados no *Framework*.

Recursos candidatos e melhorias desejáveis na adoção do framework CV-PTBR

A seguir são apresentados alguns recursos e melhorias desejáveis na adoção do novo *framework* CV-PTBR. Esses recursos são sugeridos nesse relatório, mas deverão passar pela revisão e avaliação das lideranças envolvidas no Acordo de Cooperação Internacional, podendo futuramente representar requisitos implementados em ativos de gestão de dados compartilhados.

- Fornecer uma atualização dos dados do perfil do CV, em duas vias, para outras APIs como, por exemplo, o ORCID. Ou seja, se existem dois sistemas diferentes, cada um deles com um perfil representando a mesma pessoa, que haja um mecanismo para sincronizar esses dois perfis.
- Fornecer uma atualização dos dados dos perfis nos currículos, em duas vias, para a dupla CIÊNCIA-VITAE/Lattes (dependendo da origem da requisição de atualização).
- Fornecer uma atualização de perfil unidirecional com o Oasisbr (Portal Brasileiro de Acesso Aberto à Informação Científica), LattesData (Repositório de dados abertos de pesquisa da Plataforma Lattes), PICC (Plataforma Integrada Carlos Chagas de gestão do fomento no CNPq) e Plataforma Sucupira (Plataforma

de avaliação de cursos na CAPES).

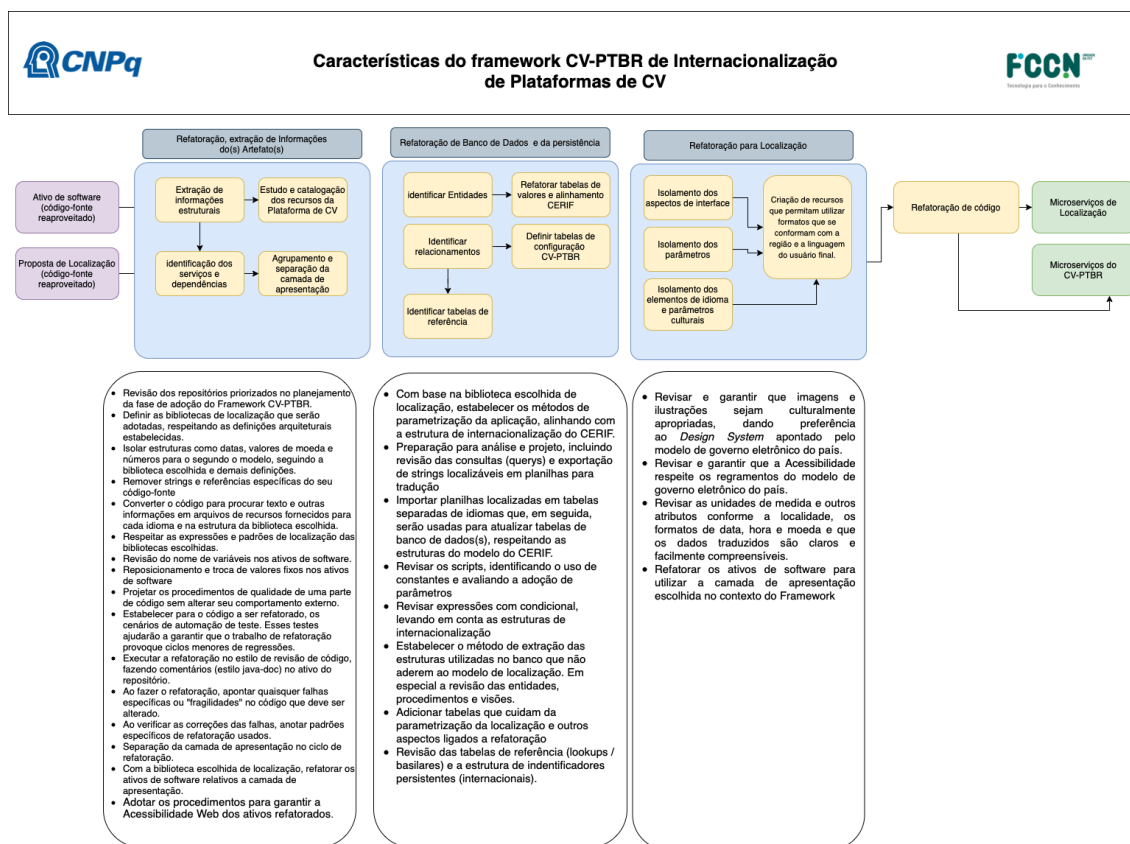
- Fornecer uma rotina de *upload* e *download* dos trabalhos de uma pessoa para/de o formato BibTex.
- Fornecer uma rotina de *upload* e *download* de registro de produções (i.e. técnicas, científicas e artísticas) e projetos de uma pessoa para/de em formato de preferência aberto.
- Fornecer uma carga de metadados do Google Scholar, Mendeley, Zotero ou aplicações semelhantes.
- Prover agrupamento de suporte de qualquer parte dos dados do CV — definindo uma ação de grupo. Esse recurso viabilizaria análise em redes de pesquisa e formaria uma estrutura de apoio para iniciativas como o Diretório de Grupos de Pesquisa do CNPq.
- Fornecer rotinas de indexação e a criação de um portal de pesquisa unificado Lattes (*Lattes Research Portal*) com uma interface de busca, permitindo que o resultado da busca utilize classes (conceitos), como por exemplo, busca por produção científica, tecnológica e artística, busca por organizações de pesquisa, fomento de projetos e permitir a criação de subdomí-

nios do portal para uso pelas universidades.

- Viabilizar suporte para análise de impacto de pesquisa. Prover extensões que utilizem a ontologia do CV e suas saídas e associá-los aos dados de impacto de pesquisa.
- Permitir a visualização dos resultados da busca com representação desses pontos de dados em um mapa. Permitir associar as informações curriculares que possuem datas em uma linha do tempo.
- Fornecer capacidade de indexação dos dados para que seja possível encontrar especialistas, incluindo buscas que selecionem especialistas com perfil semelhante ao do que caracteriza determinado projeto de pesquisa.

Como proposição da intervenção, a seguir apresenta-se as características do *framework* CV-PTBR divididas em: aspectos de refatoração na extração de informações dos artefatos existentes; aspectos de refatoração considerados para o banco de dados e persistência dos ativos existentes e aspectos de refatoração geral que promovam a localização e as mudanças necessárias na camada de apresentação dos dados (ver Figura 1).

FIGURA 1 - CARACTERÍSTICAS DO FRAMEWORK CV-PTBR



Fonte: Autores

Aspectos de refatoração considerados na adoção do Framework e na extração de informações do(s) artefato(s) existentes:

- Revisão dos repositórios priorizados no planejamento da fase de adoção do *framework* CV-PTBR.
- Definir as bibliotecas de localização que serão adotadas, respeitando as definições arquiteturais estabelecidas.
- Nas rotinas que fazem acesso ao banco de dados, isolar estruturas tais como datas, valores e tipos de software (i.e. resultado da refatoração), seguindo a biblioteca escolhida e demais definições. Cuidar da completude dos dados, realizando rotinas de verificação antes de iniciar a revisão dos objetos no banco de dados.
- Projetar os procedimentos de qualidade de dados sem alterar o comportamento da rotina.
- Estabelecer, para o código a ser utilizado, os cenários de automação e testes. Esses testes ajudarão a garantir que o trabalho de reaproveitamento de ativos

provoque ciclos menores de ajustes.

- Executar a refatoração no estilo de revisão de código-fonte, fazendo comentários que expliquem o que foi modificado (estilo java-doc) no ativo de software do repositório.
- Apontar quaisquer falhas específicas ou “fragilidades” no código que devam ser alteradas durante os procedimentos de readequação e reaproveitamento do ativo de software.
- Anotar padrões específicos de refatoração usados.
- Adotar os procedimentos para garantir a Acessibilidade Web dos ativos de software utilizados.

Aspectos de refatoração considerados na adoção do framework para o Banco de Dados e da persistência para os ativos existentes:

- Com base na biblioteca de localização escolhida, estabelecer os métodos de parametrização da aplicação, alinhando-os com a estrutura de internacionalização do CERIF.
- Preparação para análise e projeto, incluindo revisão das consultas (*queries*) e exportação de textos (*strings*)

localizáveis em planilhas para tradução.

- Importar planilhas localizadas em tabelas separadas de idiomas que, em seguida, serão usadas para atualizar tabelas de banco de dados, respeitando as estruturas do modelo do CERIF / VIVO e CIÊNCIAVITAE.
- Revisar os scripts, identificando o uso de constantes e avaliando a adoção de parâmetros.
- Estabelecer o método de extração das estruturas utilizadas no banco que não aderem ao modelo de localização. Em especial, a revisão das entidades, procedimentos e visões materializadas no sistema gerenciador de banco de dados (SGBD).
- Adicionar tabelas que cuidem da parametrização da localização e outros aspectos ligados ao ciclo de revisão dos ativos de software.
- Revisão das tabelas de referência (*lookups* / basilares) e a estrutura de identificadores persistentes (internacionais).

Aspectos de revisão considerados na adoção do *framework* para os elementos de localização na camada de apresentação:

- Revisar e garantir que imagens e ilustrações sejam do *Design System* apontado pelo modelo de governo eletrônico do país.
- Revisar e garantir que a Acessibilidade Web e a legislação de proteção de dados (e.g. LGPD e GDPR) atendam aos regramentos do modelo de governo eletrônico do país.
- Revisar as unidades de medida e outros atributos conforme a localidade, os formatos de data, hora e moeda e que os dados traduzidos sejam claros e facilmente compreensíveis.
- Refatorar os ativos de software para utilizar a camada de apresentação escolhida no contexto do *Framework* CV-PTBR.

Os aspectos apresentados acima não devem ser considerados como uma lista exaustiva de elementos ou mesmo uma lista finalizada para adoção do *fra-*

mework CV-PTBR. Como já informado, o *Framework* deve ser validado e adaptado após a realização do primeiro ciclo de adoção.

Cenário de importação e sincronização dos dados na base Lattes

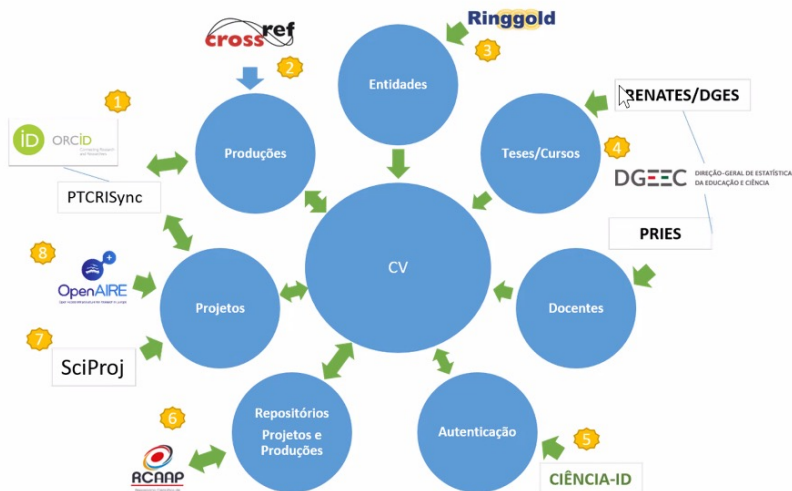
Nesse cenário, avaliaram-se três aspectos relacionados às rotinas de importação e sincronização de dados implementados no CIÊNCIAVITAE e seu modelo de serviços. Essa avaliação foi realizada por meio revisão documental, revisão simples de código-fonte e entrevistas com o engenheiro de software responsável na Fundação para a Computação Científica Nacional - FCCN de Portugal. Nessa parte do estudo, verificou-se a possibilidade de reaproveitamento dos artefatos de software mantidos pela FCCN e a possível aplicação do *framework* CV-PTBR na criação de um módulo isolado na Plataforma Lattes para importação e sincronização de dados.

Elementos utilizados na avaliação do Cenário

Os elementos utilizados no presente cenário incluíram o modelo de fontes e bases importadas ou sincronizadas pelo CIÊNCIAVITAE, a forma de estruturação dessas duas operações e o modelo de resolução de conflitos ou inconsistências de dados após a execução dessas rotinas, considerando o *framework* CV-PTBR.

Como já apresentado, a proposta do integrador e sincronizador de dados do CIÊNCIAVITAE e PTCRIS (*Current Research Information System* de Portugal) é semelhante a um pool de serviços de integração e as fontes de informação são apresentadas na Figura 2.

FIGURA 2 - MODELO DE IMPORTAÇÃO/SINCRONIZAÇÃO DO CIÊNCIAVITAE



Fonte: FCCN

O modelo de importação/sincronização do CIÊNCIAVITAE estabelece uma visão de “central de serviços” que cuidam da importação ou sincronização de conteúdos para o currículo. No caso específico do ORCID, após o primeiro ciclo de importação, é possível manter a sincronização de dados com a base CIÊNCIAVITAE. Nessa sincronização, é facultado ao operador realizar a gestão de ocorrências e/ou inconsistências dos dados entre essas fontes sincronizadas.

No modelo do CIÊNCIAVITAE, cada serviço de importação depende de um conjunto de regras/rotinas de preparação de dados. Um exemplo disso foi a preparação feita pela equipe da FCCN, em conjunto com as Instituições de Ensino Superior Portuguesas, estabelecendo uma rotina de dados cadastrais dessas Instituições. Atualmente, os dados são atualizados semanalmente por uma rotina de extração, transformação e carga (ETL) entre a base do *Ringgold* e as bases do CIÊNCIAVITAE e de outras aplicações da FCCN. O

Ringgold é uma fonte de informações sobre a estrutura organizacional de Instituições e Ensino Superior e Instituições de Pesquisa.

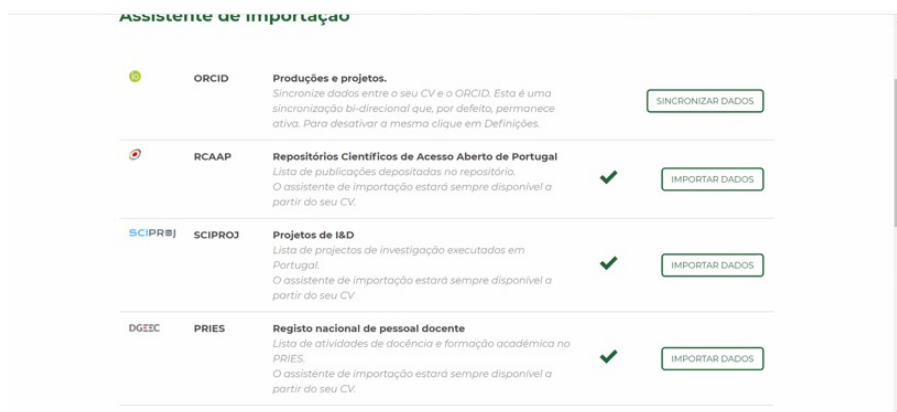
Em outra situação de importação de dados por parte do CIÊNCIAVITAE, para a sincronização de dados curriculares do sistema ORCID, a implementação do CIÊNCIAVITAE utiliza um cliente REST para as APIs que foi derivado de um pacote publicado em licença de código aberto. A rotina realiza a comparação entre a produção (output) informada no CIÊNCIAVITAE e o que está informado no ORCID e notifica quando as fontes não estão sincronizadas. Cabe ao operador do CIÊNCIAVITAE aceitar ou não a sincronização dos dados. A Figura 3 demonstra a interface do assistente de importação do CIÊNCIAVITAE, com a opção de sincronização de dados do ORCID e importação de dados do RCAAP⁴, SCIPROJ⁵ e DGEEC⁶ entre outras.

4. Conforme o site do portal RCAAP (Repositório Científico de Acesso Aberto de Portugal), esse portal tem como objetivo a coleta, agregação e indexação dos conteúdos científicos em acesso aberto (ou acesso livre) existentes nos repositórios institucionais das entidades nacionais de ensino superior, e outras organizações de I&D. Mais informações em: <https://www.rcaap.pt/about.jsp>

5. Base de dados com o registro nacional de financiamento de ciência e tecnologia de Portugal.

6. A DGEEC disponibiliza as bases de dados da área da Educação e da Ciência e Tecnologia para fins de investigação, de acordo com o Protocolo estabelecido com o Instituto Nacional de Estatística e com a Fundação para a Ciência e a Tecnologia de Portugal (FCT).

FIGURA 3 - INTERFACE DO ASSISTENTE DE IMPORTAÇÃO DO CIÊNCIAVITAE PARA ESCOLHA DAS FONTES



Fonte: FCCN

Como já mencionado, em um esforço para a adoção de identificador persistente, a equipe da FCCN fez a preparação e a atualização dos dados das IES na fonte do *Ringgold*. O *Ringgold Identifier* é um identificador persistente para organizações que inclui cerca de quinhentos mil IDs representando organizações e consórcios que utilizam conteúdos de publicações e outras fontes acadêmicas. Para manter as informações atualizadas no CIÊNCIAVITAE e outros projetos que utilizam esses dados, a equipe da FCCN realiza periodicamente o *download* e a atualização dos dados do *Ringgold*. O processo utiliza-se de uma rotina em conjunto com a ferramenta de ETL da Talend e inclui o *download* do arquivo com os dados via FTP. O processo está estruturado na ferramenta de ETL onde o script cuida da descompactação do arquivo e das rotinas de atualização dos dados na base operacional. A rotina é executada a cada ciclo de tempo para garantir que a atualização feita pelas organizações no *Ringgold* possa estar refletida nas bases da FCCN.

A rotina do assistente de importação também se utiliza da interface do *OpenAIRE*⁷, sendo nesse caso uma iniciativa de projeto em andamento. A identificação de publicações é feita com 28 repositórios portugueses de dados de pesquisa (repositórios da Universidade do Minho, entre outros, hospedados pela rede

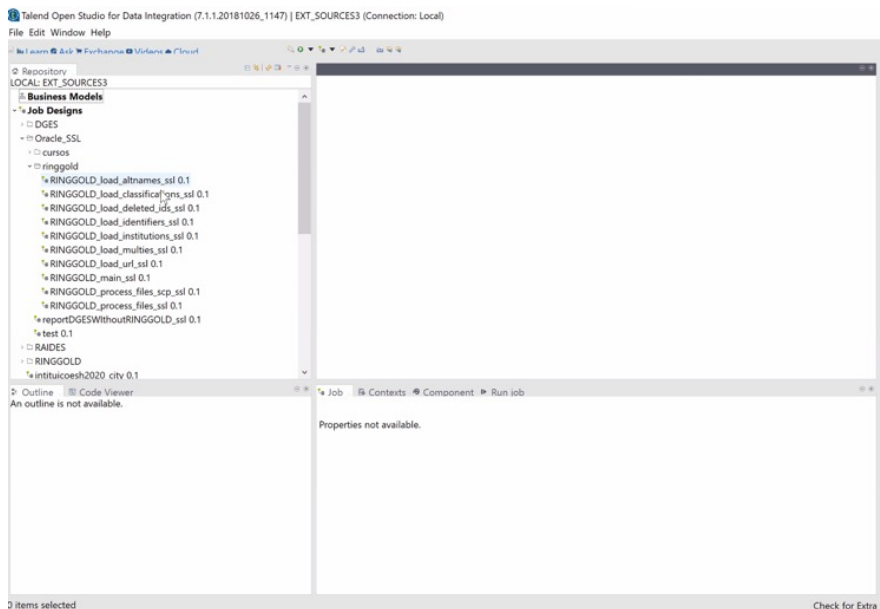
RCAAP) a partir de um conjunto de 10 mil projetos financiados pela FCT. Essa abordagem sugere um modelo descentralizado de dados para alimentar a informação curricular nacional.

Outra integração é com a base do SCIPROJ, um repositório de dados sobre projetos financiados pela FCT e desenvolvido seguindo os padrões estabelecidos pela OpenAIRE CERIF, tendo sido implementado uma interface cliente para realizar sua importação de dados. Outras fontes podem ser consideradas candidatas ao projeto conjunto com o currículo Lattes, em especial a integração com o RCAAP. Para aproveitamento desses ativos criados pela FCT, a intervenção proposta neste relatório novamente sugere a adoção do *framework* CV-PTBR.

A seguir são apresentadas imagens das rotinas e ferramentas utilizadas pelos scripts de atualização dos dados da base do *Ringgold*. Para tal, a Figura 4 demonstra o ambiente do *software* Talend e a edição dos scripts do *Ringgold*; a Figura 5 apresenta a rotina de sincronização do *Ringgold* com notificação via e-mail e a Figura 6 demonstra uma parte do procedimento (*Script Oracle*) para atualização da visão materializada do *Oracle* para o *Ringgold*. Essas figuras representam diferentes aspectos e elementos da rotina de atualização da base.

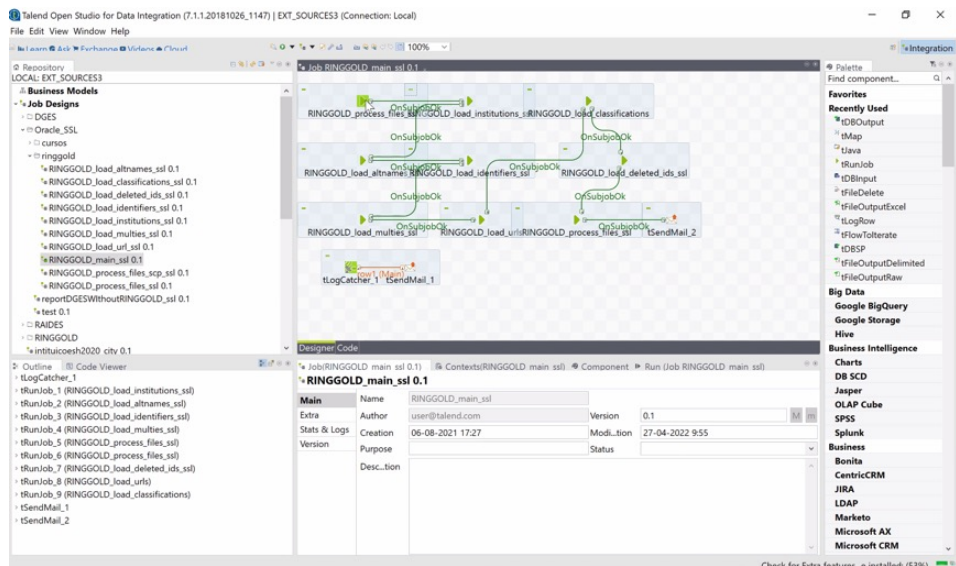
7. Conforme informado pelo site, o OpenAIRE (Open Access Infrastructure for Research in Europe) tem como missão mudar a comunicação científica para a abertura e a transparência e facilitar formas inovadoras de comunicar e monitorizar a investigação. Informações adicionais estão disponíveis em: <https://www.openaire.eu/mission-and-vision>.

FIGURA 4 - IMAGEM DO TALEND NA EDIÇÃO DOS SCRIPTS DO RINGGOLD



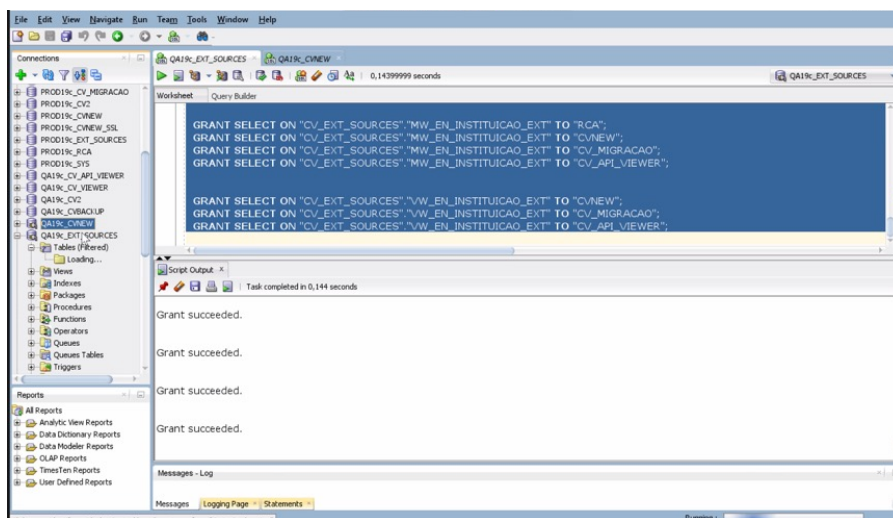
Fonte: FCCN

FIGURA 5 - IMAGEM DA ROTINA DE SINCRONIZAÇÃO DO RINGGOLD COM NOTIFICAÇÃO VIA E-MAIL



Fonte: FCCN

FIGURA 6 - SCRIPT ORACLE PARA ATUALIZAÇÃO DA VISÃO MATERIALIZADA RINGGOLD

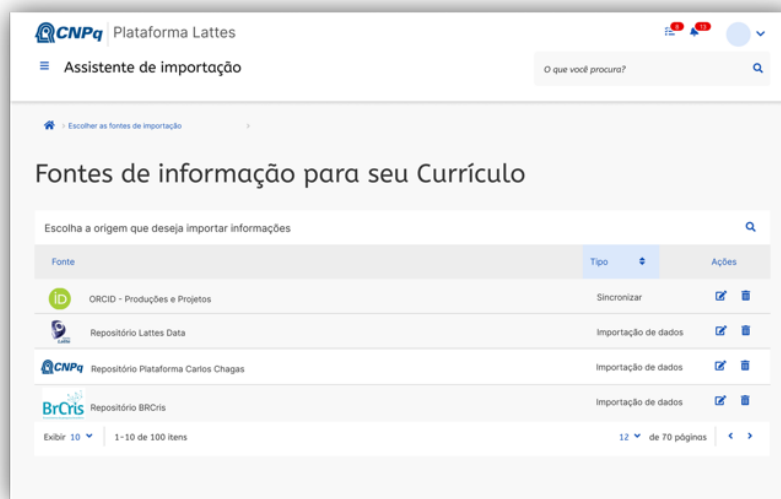


Fonte: FCCN

Os elementos utilizados no presente cenário tratam da escolha das fontes de informação participantes das rotinas de importação e sincronização. Também está demonstrado a visualização dos objetos resultantes da importação e sincronização e o seu respectivo status.

As figuras 7, 8, 9, 10 e 11 ainda demonstram a prototipação da complementação de conteúdo em registros de produção científica, a complementação de conteúdo em projeto lançado nas Plataformas PICC e na Sucupira⁸ e a complementação de conteúdo em projeto lançado no Lattes Data⁹.

FIGURA 7 - PROTÓTIPO DO ASSISTENTE DE IMPORTAÇÃO (MÓDULO INDEPENDENTE NA PLATAFORMA LATTES)



Fonte: Proposta do serviço de importação feito por José Francisco Salm Jr. (2022)

8. A Plataforma Sucupira é um sistema de coleta de informações, análises e avaliações a serem utilizadas como base padronizadora do Sistema Nacional de Pós-Graduação brasileira. A Plataforma disponibiliza informações, processos e procedimentos que a CAPES realiza no SNPG para toda a comunidade acadêmica.

9. O Lattes Data é o repositório de dados de pesquisa, que foi criado em parceria entre Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

FIGURA 8 - RESULTADOS E PENDÊNCIAS NO MÓDULO DE IMPORTAÇÃO

The screenshot displays the 'Assistente de importação' (Import Assistant) interface. At the top, there's a search bar and a 'Pendências de Todos' (All Pending) button. Below, a legend defines the status icons: a green checkmark for 'Importado' (Imported), a yellow warning for 'Incompleto' (Incomplete), and a red error for 'Falha' (Failure). The main section, 'Minhas pendências' (My Pending), features a table with columns for Status, Pendência (with a 'Selecionar' dropdown), Permissões (with a 'Selecionar' dropdown), Tipo de produção (with a 'Selecionar' dropdown), No. do PID (with a 'Selecionar' dropdown), Data (with a 'Selecionar' dropdown and calendar icon), and Ação. The table lists five items, all of which are 'Produção bibliográfica' (Bibliographic production). The first two are 'Importado', the third is 'Incompleto', and the last two are 'Falha'. At the bottom, there are pagination controls showing '1-10 de 100 itens' and '12 de 70 páginas'.

Status	Pendência	Permissões	Tipo de produção	No. do PID	Data	Ação
Importado	Sentiment analysis using support...	(r) (w) (x)	Produção bibliográfica	12.1.000000077-4	16/02/2021	BrCris
Importado	Use of support vector machines in ...	(r) (w)	Produção bibliográfica	-	15/02/2021	iD
Incompleto	An experiment in integrating sentim...	(r) (w)	Produção bibliográfica	12.1.000001117-2	10/02/2020	Lattes
Falha	De-conflated semantic representat...	(r)	Produção bibliográfica	12.1.000008888-9	12/12/2020	iD
Falha	Uncovering text mining: A survey of...	(r) (w) (x)	Produção bibliográfica	12.1.000000000-3	12/12/2020	PICRIS

Fonte: Módulo de serviço de importação proposto na pesquisa pós-doutoral de José Francisco Salm Jr. (2022)

FIGURA 9 - PROTÓTIPO DE TELA PARA COMPLEMENTAÇÃO DE CONTEÚDO EM PRODUÇÃO CIENTÍFICA

Plataforma Lattes

Assistente de importação

Complementar informações sincronizadas

Categoria (obrigatório): Publicações

Tipo (obrigatório): Livro

Identificador DOI (Identificador de Objetos Digitais) da produção:

Importar

Referência

Sentiment analysis using support vector machines with diverse information sources

Sentiment analysis using support vector machines with diverse information sources

Detalhes:

Título: Sentiment analysis using support vector machines with diverse information sources

Volume: Edição: Número da página:

Complemento:

Entre as 5 produções mais relevantes (obrigatório):

Sim Não

Cancelar Importar

Fonte: Módulo de serviço de importação proposto na pesquisa pós-doutoral de José Francisco Salm Jr. (2022)

FIGURA 10 - PROTÓTIPO DE TELA PARA COMPLEMENTAÇÃO DE CONTEÚDO EM PROJETO LANÇADO NAS PLATAFORMAS PICC E NA SUCUPIRA

Completar informações importadas de Projetos (PICC ou CAPES)

Tipo de projeto (obrigatório):
Projeto de pesquisa

Referência (Nome do projeto)

Atualização da Plataforma Lattes com foco em Ontologias internacionais de CV

Comunidades de padronização semântica de dados de bases científicas

Nome do Projeto:
Atualização da Plataforma Lattes com foco em Ontologias internacionais de CV

Descrição:
Atualmente, a Plataforma Lattes possui mecanismos heterogêneos de interoperabilidade e serviços que conectam diferentes fontes de informação da CETI. Com os avanços dos serviços e o surgimento de novas interfaces de aplicação (Science data API), faz-se necessário avaliar novas abordagens de barramentos de serviços, permitindo que os pesquisadores completem seu currículo de forma estruturada e apresentem suas contribuições mais importantes para a ciência em narrativas breves, e não apenas como listas de publicações. Ao avaliar a iniciativa da CIENCIAVITA, pode-se entender como a Plataforma Lattes poderá evoluir nos aspectos do uso de barramentos.

Situação: Em andamento Ano início: 2022 Ano fim:

É um projeto de cooperação instituição de pesquisa e empresa? Sim Não

O projeto possui potencial de inovação de produtos, processos ou serviços? Sim Não

Instituição de execução:

Órgão/Unidade:

Equipe (obrigatório):

Cancelar Importar

Fonte: Módulo de serviço de importação proposto na pesquisa pós-doutoral de José Francisco Salm Jr. (2022)

FIGURA 11 - PROTÓTIPO DE TELA PARA COMPLEMENTAÇÃO DE CONTEÚDO EM PROJETO LANÇADO NO LATTES DATA

Completar informações importadas de Projetos (PICC ou CAPES)

Tipo de projeto (obrigatório):
Projeto de pesquisa

Referência (Nome do projeto)

As Dimensões da Biodiversidade em Lepidoptera Neotropicales

As Dimensões da Biodiversidade em Lepidoptera Neotropicales

Nome do Projeto:
As Dimensões da Biodiversidade em Lepidoptera Neotropicales

Descrição:
O projeto propõe três linhas de pesquisa com Lepidoptera Neotropicales: 1) Taxonomia e filogenia; 2) Ecologia de comunidades e conservação; e 3) Estrutura filogenética e funcional de comunidades. O objetivo do projeto é que estas três linhas combinem seus resultados para que padrões mais amplos de diversificação de borboletas neotropicales sejam desenvolvidos. Os resultados obtidos não são de interesse acadêmico, pois irão contribuir para uma melhor compreensão de padrões e processos de diversificação de borboletas neotropicales, mas também devem ser úteis para subsidiar ações de conservação de espécies e habitats ameaçados nos Neotrópicos, que incluem os sistemas.

Situação: Em andamento Ano início: 2022 Ano fim:

É um projeto de cooperação instituição de pesquisa e empresa? Sim Não

O projeto possui potencial de inovação de produtos, processos ou serviços? Sim Não

Instituição de execução:
UNICAMP

Órgão/Unidade:

Equipe (obrigatório):

Dataverse
Data (2022-06-24)
Assunto: Medicina, Saúde e Ciências da Vida
Palavra-chave: Ecologia de comunidades, Borboletas, Evolução, Conservação, Estrutura filogenética, Taxonomia
Agência financiadora: CNPq
Número do processo: 303834/2015-3
Depositante: André Victor Lucci Freitas
Data do depósito: 2022-06-24

Cancelar Importar

Fonte: Módulo de serviço de importação proposto na pesquisa pós-doutoral de José Francisco Salm Jr. (2022)

RECOMENDAÇÕES

Cabe destacar que o modelo adotado pela FCT/FCCN para o importador e sincronizador atende a todas as características contemporâneas de interoperabilidade técnica e está aderente às boas práticas de reuso das fontes de informação no contexto de sistemas RIMS¹⁰ (*Research Information Management Systems*). Dessa forma, recomenda-se aplicar o *framework* CV-PTBR no modelo de assistente de importação e sincronização para a criação de um módulo complementar à Plataforma Lattes que seja capaz de incorporar a interoperabilidade técnica e os serviços de dados em fontes internacionais de C&T,I.

RECONHECIMENTO E FINANCIAMENTO

Esse trabalho é fruto do projeto de pós-doutorado, cuja bolsa de estudo foi concedida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, sendo que o termo de outorga está associado ao processo de N^o: 400286/2022-0 (Pós-Doutorado no Exterior – PDE). Os autores agradecem ao CNPq, à FCT/FCCN e à Universidade do Estado de Santa Catarina pelo apoio no desenvolvimento desse projeto.

REFERÊNCIA BIBLIOGRÁFICA

- ANDON, P. I.; REZNICHENKO, V. A.; CHISTYAKOVA, I. S.. Mapping of Description Logic to the Relational Data Model. *Cybernetics And Systems Analysis*, [S.L.], v. 53, n. 6, p. 963-977, nov. 2017. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s10559-017-9998-3>.
- CHUJAI, Pasapitch; KERDPRASOP, Nittaya; KERDPRASOP, Kittisak. On Transforming the ER Model to *Ontology* Using Protégé OWL Tool. *International Journal Of Computer Theory And Engineering*, [S.L.], v. 6, n. 6, p. 484-489, dez. 2014. IACSIT Press. <http://dx.doi.org/10.7763/ijcte.2014.v6.914>.
- GONÇALVES, Adriano de Oliveira; JACYNTHO, Mark Douglas de Azevedo. Um método para publicação semântica Linked Data de bases de dados convencionais e um estudo de caso real de artigos acadêmicos. *Transinformação*, [S.L.], v. 32, n. 0, p. 0-0, 2020. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/1678-9865202032e180051>.
- KRAFFT, Dean; CAPPADONA, Nicholas; CARUSO, Brian; CORSON-RIKERT, Jon; DEVARE, Medha; LOWE, Brian Vivo. (2010). *VIVO: Enabling National Networking of Scientists. Proc of Web Sci Conf*.
JSON-SCHEMA <https://json-schema.org/understanding-json-schema/UnderstandingJSONSchema.pdf>
- KHUZAYEM, Lama Al; MCBRIEN, Peter. OWLRel: Learning Rich Ontologies from Relational Databases. *Baltic Journal Of Modern Computing (Bjmc): ISSN 2255-8950 (Online), ISSN 2255-8942 (Print)*. Acta Universitatis Latviensis, p. 466-482. 30 abr. 2016.
- MAHRIA, Bilal Ben; CHAKER, Ilham; ZAHI, Azeddine. A novel approach for learning *Ontology* from relational database: from the construction to the evaluation. *Journal Of Big Data*, [S.L.], v. 8, n. 1, p. 2-22, 28 jan. 2020. Springer Science and Business Media LLC. <http://dx.doi.org/10.1186/s40537-021-00412-2>.
- MOGOTLANE, Kgotatso Desmond; FONOU-DOMBEU, Jean Vincent. AUTOMATIC CONVERSION OF RELATIONAL DATABASES INTO ONTOLOGIES: A COMPARATIVE ANALYSIS OF PROTÉGÉ PLUG-INS PERFORMANCES. *International Journal Of Web & Semantic Technology*. Pretoria, p. 21-40. out. 2016.
- SEQUEDA, Juan F.; TIRMIZI, Syed Hamid; CORCHO, Oscar; MIRANKER, Daniel P.. Survey of directly mapping SQL databases to the Semantic Web. *The Knowledge Engineering Review*, [S.L.], v. 26, n. 4, p. 445-486, dez. 2011. Cambridge University Press (CUP). <http://dx.doi.org/10.1017/s0269888911000208>

10. Os sistemas de informação de pesquisa (RIMSs) podem ser definidos como os tipos de sistemas de informação que permitem gerenciar e fornecer acesso às informações de conteúdo e identidade de autoria dos pesquisadores. Editores, bibliotecas, universidades, mecanismos de busca e agregadores de conteúdo criaram diferentes RIMSs, todos com diferentes modelos de dados, cobertura e qualidade.

SEGUNDO, José Eduardo Santarém; CONEGLIAN, Caio Saraiva; LUCAS, Elaine Rosangela de Oliveira. Conceitos e tecnologias da Web semântica no contexto da colaboração acadêmico-científica: um estudo da plataforma vivo. *Transinformação*, [S.L.], v. 29, n. 3, p. 297-309, dez. 2017. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/2318-08892017000300007>.

SMITH, Barry; CEUSTERS, Werner. Ontological realism: a methodology for coordinated evolution of scientific ontologies. *Applied Ontology*, [S.L.], v. 5, n. 3-4, p. 139-188, 2010. IOS Press. <http://dx.doi.org/10.3233/ao-2010-0079>.

SURE, York; BLOEHDORN, Stephan; HAASE, Peter; HARTMANN, Jens; OBERLE, Daniel. The SWRC *Ontology* – Semantic Web for Research Communities. *Progress In Artificial Intelligence*, [S.L.], p. 218-231, 2005. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/11595014_22.

**José Francisco Salm Junior**

Universidade do Estado de Santa Catarina – UDESC

jose.salmjunior@udesc.br<https://orcid.org/0000-0002-8492-1645>

Possui Doutorado em Engenharia do Conhecimento pela Universidade Federal de Santa Catarina. É professor permanente do Programa de Pós-Graduação em Gestão da Informação da Universidade do Estado de Santa Catarina e pertence ao Departamento de Administração Pública do Centro de Ciências da Administração e Socioeconômicas - ESAG. Criou a Comunidade que auxiliou o CNPq na padronização da Plataforma Lattes. Atuou como pesquisador e consultor no Programa Star Metrics do Escritório de Políticas Públicas em Ciência e Tecnologia da Casa Branca (OSTP - Executive Office of the President of the United States). É membro do Grupo Consultivo Técnico de Evidências e Inteligência para Ação em Saúde das Américas na OPAS em Washington, DC.

**Paulo Henrique de Assis Santana**

Servidor aposentado do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq

santanapha@gmail.com<https://orcid.org/0000-0002-9674-3401>

Engenheiro mecânico pela Escola de Engenharia de São Carlos – EESC/USP e Mestre em Ciências pelo Instituto Nacional de Pesquisas Espaciais - INPE. Foi gestor de TI no Ministério da Agricultura, no Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, no Ministério do Meio Ambiente – MMA e no Ministério do Desenvolvimento Social e Combate à Fome – MDS. Mais recentemente, foi Coordenador Geral de Gestão e Inovação no Ministério de Ciência, Tecnologia e Inovação – MCTI e Coordenador de Informação e Estudos Internacionais no CNPq. Hoje, aposentado, é consultor independente. Coordenou o desenvolvimento do Currículo Lattes e sua disseminação pelas universidades brasileiras e agências de fomento, o projeto da Rede ScienTI, que disseminou o padrão do currículo Lattes internacionalmente, o Acordo de Cooperação Internacional entre o CNPq e a Fundação para a Ciência e a Tecnologia de Portugal para evolução conjunta de suas plataformas curriculares e o Acordo de Cooperação Técnica entre CNPq e IBICT para integração entre o CV Lattes e o Portal OasisBR. Coordenou, até maio de 2022, as atividades em Ciência Aberta do CNPq e o desenvolvimento do Repositório Aberto de Dados de Pesquisa do CNPq – LattesData.

**Geraldo Sorte**

Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq

gsorte@cnpq.br<https://orcid.org/0000-0002-9281-8859>

Graduado em Processamento de Dados pela Universidade de Brasília (1986), possui 34 anos de experiência na área de TI, sendo 17 anos coordenando áreas de TI na Administração Pública Federal. Participou da idealização e implementação da Plataforma Lattes e Plataforma Integrada Carlos Chagas do CNPq. Coordenou projetos de implantação de modelos de gestão por serviços na área de TI no CNPq e Fiocruz. Coordenou o projeto de implantação da infraestrutura tecnológica e de serviços de TI do Centro Hospitalar da COVID-19 da Fiocruz, bem como do novo sistema de gestão acadêmica da Fundação. Atualmente é Coordenador Geral de Tecnologia da Informação e Membro do Comitê de Governança Digital do CNPq. Atua nas áreas de Ciência da Computação, Modelos de Gestão de TIC, Ciência de Dados, Data Warehouse e Análise de Redes.