

PTCRIS

Connected Research

<REFERENCIAL NORMATIVO DE INTEROPERABILIDADE PTCRIS>

<Anexo Técnico>

Ficha Técnica

Título

Referencial Normativo de Interoperabilidade PTCRIS

Autoria

Equipa PTCRIS

Unidade para a Computação Científica Nacional da Fundação para a Ciência e a Tecnologia (FCT|FCCN)

Histórico de versões

v0.1 - Agosto 2019

v0.2 - Julho 2020

Índice

Índice	1
Enquadramento	2
Quadro Normativo PTCRIS	4
Identificadores Únicos (PIDs - Persistent Identifiers)	4
Identificador de Pessoas: Ciência ID	4
Identificador de Organizações/Instituições: ISNI+	5
Identificador de Publicações: DOI	5
Identificador de Projetos/Financiamento: DOI	6
Semântica	6
Publicações: Vocabulário controlado COAR	6
Projetos / Financiamento: CERIF	6
Curriculum Vitae: semântica CIÊNCIAVITAE/ORCID/CASRAI	6
Modelo de Dados	7
Perfis de troca de dados	7
Projetos /Financiamento: OpenAIRE	7
CIÊNCIAVITAE: XML e API	7
Infraestruturas PTCRIS	8
Serviço de Autenticação CIÊNCIA ID	8
Base de dados de Organizações	9
SciPROJ	10
APIs CIÊNCIAVITAE	10
Outras Implementações	11
Requisitos PTCRIS - Tabela Resumo	13

<Quadro Normativo PTCRIS>

Enquadramento

O [PTCRIS](#) é um programa desenvolvido pela Unidade de Computação Científica Nacional (UCCN) da FCT, que tem como objetivo o desenvolvimento sustentado de um ecossistema integrado de informação de suporte à atividade científica nacional. O desenvolvimento de tal ecossistema é crucial para garantir, por um lado, a desburocratização e a simplificação dos processos administrativos, por outro, a produção, gestão e acesso a dados fidedignos, completos e atualizados, dos quais depende a efetividade e eficiência dos processos de tomada de decisão.

Para dar cumprimento ao seu objetivo, o PTCRIS dedica-se à definição de um quadro normativo e à criação de infra estruturas que viabilizem a adoção desse mesmo normativo (Figura 1). O quadro normativo PTCRIS, em alinhamento com as melhores práticas internacionais, descreve um conjunto de princípios e regras que regem a interoperabilidade entre os componentes do ecossistema, nomeadamente, no que respeita a formatos de dados e interoperabilidade; semântica; gestão de identificadores únicos e fluxos de dados.

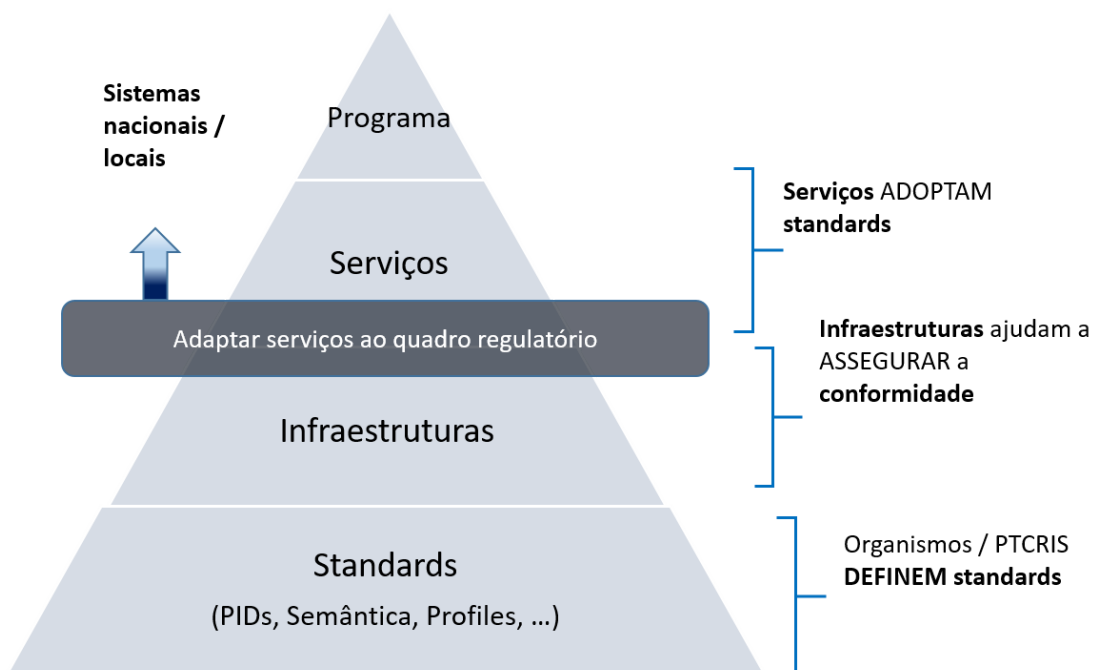


Figure 1. Representação esquemática do programa PTCRIS

<Quadro Normativo PTCRIS>

A adoção do quadro normativo PTCRIS pelos diversos sistemas que operam no ecossistema acadêmico-científico nacional é fundamental para garantir uma integração ampla e eficaz e, conseqüentemente, a consecução do princípio basilar do PTCRIS: *Preencha uma vez, reutilize múltiplas*.

Este documento descreve as normas e as infraestruturas de suporte a adoptar pelos componentes do ecossistema científico nacional, bem como a metodologia para implementação das mesmas. De um modo geral, cada uma das normas PTCRIS é suportada por infraestruturas que viabilizam a adoção dessas mesmas normas (*vide* tabela 1).

Tabela 1. Normas e Infraestruturas PTCRIS

Tipologia		Normas	Infraestruturas
Identificadores	Pessoas	CIÊNCIA ID	Sistema de Autenticação
	Organizações	ISNI+	Tabela de Autoridade Ringgold
	Publicações	DOI	
	Financiamento	DOI	
Semântica/Dicionário de dados	Publicações	COAR	
	Projetos/Financiamento	CERIF	SciPROJ
	CV	CASRAI/ORCID	
Modelo de Dados		CERIF	
Perfis de troca de dados	Financiamento	OpenAIRE	API SciPROJ
Perfis de troca de dados	CV	CV-XML	API CV

<Quadro Normativo PTCRIS>

Quadro Normativo PTCRIS

1. Identificadores Únicos (PIDs - Persistent Identifiers)

O fluxo de dados num ecossistema de informação aberto e integrado depende de identificadores únicos e persistentes que sejam reconhecidos por todos os componentes que constituem o ecossistema. Tais identificadores devem permitir distinguir de forma unívoca os diferentes intervenientes do ecossistema académico-científico: Indivíduos, Organizações, Publicações e Projetos .

1.1. Identificador de Pessoas: Ciência ID

Ao abrigo do quadro normativo PTCRIS todos os componentes do ecossistema académico ou científico devem adotar o CIÊNCIA ID como identificador.

O CIÊNCIA ID é um meio de identificação e autenticação individual e permanente para os cidadãos que desenvolvem atividade científica.

Este identificador integra a AUTENTICAÇÃO.GOV o que significa que o CIÊNCIA ID é atribuído com base na informação existente no cartão do cidadão e que nos permite uma validação forte da informação que está associada ao identificador (Figura 2).



Figura 2. CIÊNCIA ID funciona como a chave de acesso a diversos sistemas do ecossistema académico-científico

O CIÊNCIA ID fornece também um mecanismo de autenticação single sign-on comum aos vários serviços de ciência que integram o CIÊNCIA ID. O mecanismo de autenticação permite também que a informação registada

<Quadro Normativo PTCRIS>

no CIÊNCIA ID seja partilhada com os vários serviços. Desta forma sempre que o investigador acede a um serviço de ciência através da sua conta CIÊNCIA ID, a sua informação registada no CIÊNCIA ID é partilhada com o serviço, deixando de ser necessário o processo de registo de informação pessoal sempre que acedemos a um novo serviço.

1.2. *Identificador de Organizações/Instituições: ISNI+*

Ao abrigo do quadro normativo PTCRIS deverá ser adoptado o sistema de identificadores de organizações ISNI+, em conformidade com recomendações internacionais¹. O sistema ISNI+ combina o identificador [ISNI](#) (International Standard Name Identifier)² com os serviços e infraestrutura [Ringgold](#)³.

No sistema ISNI+ cada organização/instituição está associada a um par identificadores ISNI/Ringgold único.

Em processos de migração de dados será necessário determinar os códigos ISNI/Ringgold que representam uma dada organização/instituição. Para o efeito, o PTCRIS disponibiliza um serviço de reconciliação de dados, baseado na ferramenta *open-source* [OpenRefine](#). No âmbito deste serviço, o PTCRIS disponibiliza [documentação de suporte](#) e o URL do serviço de reconciliação.

O PTCRIS disponibiliza ainda uma tabela de mapeamento que permite determinar os códigos ISNI/Ringgold que representam uma dada organização/instituição a partir do NIPC (Número de Identificação de Pessoa Colectiva) da mesma.

1.3. *Identificador de Publicações: DOI*

Os objetos digitais, nomeadamente as publicações, deverão estar associadas a um DOI - Digital Object Identifier⁴. O registo de DOIs está a cargo da Crossref⁵. No processo de registo é recolhida informação diversa sobre os objetos digitais e a sua associação aos demais intervenientes do ecossistema

¹ Review of selected organisational IDs and development of use cases for the Jisc CASRAI-UK. Organisational Identifiers Working Group. <http://repository.jisc.ac.uk/5853/>.

² Norma ISO amplamente utilizada para distinguir pessoas e organizações

³ A Ringgold é uma organização especializada em identificadores organizacionais que opera como agência de registo de identificadores ISNI.

⁴ O sistema DOI é gerido pela International DOI Foundation (IDF).

⁵ A Crossref é uma agência de registo oficial do DOI da Fundação Internacional DOI.

<Quadro Normativo PTCRIS>

académico-científico. Esta informação pode depois ser disponibilizada via API para pesquisa, descoberta e monitorização.

1.4. *Identificador de Projetos/Financiamento: DOI*

À semelhança das publicações também os projetos/financiamentos devem estar associados a um DOI que permita associar de forma unívoca projetos/financiamento às pessoas e à sua informação.

2. Semântica

Um dos pilares da interoperabilidade entre os sistemas é a capacidade de mapear os termos usados em cada sistema com a semântica partilhada. A semântica definida no âmbito do quadro normativo PTCRIS incide sobre publicações, projetos/financiamento e informação curricular.

2.1. *Publicações: Vocabulário controlado COAR*

Os metadados bibliográficos devem ser expressos usando o vocabulário controlado [COAR](#). Em particular, o SGP deverá usar a componente do vocabulário controlado COAR que define os termos para identificar o tipo de recurso e que está descrito em [SKOS](#).

2.2. *Projetos / Financiamento: CERIF*

Os [metadados](#) da Base de Dados Nacional de Financiamento assentam nos atributos definidos pelo modelo de dados [CERIF](#) (Common European Research Information Format)⁶. Estes metadados englobam as dimensões: financiamento, pessoa, organização e projecto,

2.3. *Curriculum Vitae: semântica CIÊNCIAVITAE/ORCID/CASRAI*

A coleção de metadados de um *currículum vitae* foi definida no âmbito do quadro normativo PTCRIS. Esta coleção de metadados, designada semântica **CIÊNCIAVITAE**, tem por base o dicionário [CASRAI](#)⁷, também adoptado pelo ORCID⁸. A semântica CASRAI foi adaptada às especificidades do ecossistema

⁶ O CERIF é uma norma para o modelo de base de dados dos actuais sistemas de informação de investigação (CRIS) proposto pela UE e desenvolvido e mantido pela euroCRIS.

⁷ O dicionário CASRAI define a informação basilar representativa da história académica de um indivíduo. Mais informações em: https://dictionary.casrai.org/CV_-_Full_Academic

⁸ O ORCID é uma organização sem-fins-lucrativos responsável pela gestão de um registo de identificadores digitais únicos e persistentes que permitem a desambiguação dos nomes. Mais ainda, a

<Quadro Normativo PTCRIS>

académico-científico nacional de forma a garantir que a semântica **CIÊNCIAVITAE** abarca e reflete a realidade nacional.

A semântica **CIÊNCIAVITAE** está disponível [aqui](#) e deverá ser utilizada para descrever informação curricular.

3. Modelo de Dados

O modelo de base de dados preconizado no âmbito do quadro normativo PTCRIS é o CERIF (*vide* 2.2). Em particular, o PTCRIS adopta o [CERIF-XML](#), um formato de dados CERIF amplamente utilizado tendo em vista a interoperabilidade entre os sistemas.

4. Perfis de troca de dados

4.1. *Projetos /Financiamento: OpenAIRE*

O quadro normativo PTCRIS definiu como perfil de troca de dados sobre projetos/financiamento o perfil [OpenAIRE](#), que pela sua maior abrangência (identidade de ~80% com os metadados da Base de Dados Nacional de Financiamento), demonstrou ser o mais adequado para dar cumprimento aos objectivos a que a Base de Dados se propõe. Este perfil de dados poderá vir a ser partilhado via API.

4.2. **CIÊNCIAVITAE: XML e API**

O PTCRIS definiu um perfil de dados **CIÊNCIAVITAE** para partilha de informação curricular. Este perfil de dados pode ser partilhado via API, através da qual é possível aceder a informação pública e autorizada dos currículos registados na plataforma **CIÊNCIAVITAE**. Esta informação é acedida em tempo real pelo que disponibiliza os dados do currículo associado a um dado CIÊNCIA ID, com as atualizações mais recentes (registadas pelos utilizadores), nos formatos XML e PDF.

integração do ORCID com vários sistemas do ecossistema académico-científico permite a associação automática entre as pessoas e a sua informação.

<Quadro Normativo PTCRIS>

Infraestruturas PTCRIS

1. Serviço de Autenticação CIÊNCIA ID

A infraestrutura CIÊNCIA ID tem por base o protocolo SAML2 (*Security Assertion Markup Language*). Este protocolo desenvolvido pelo OASIS, é uma estrutura baseada em XML para comunicar informações de autenticação, autorização e atributos do utilizador entre dois intervenientes:

- **Fornecedor de Identidade**

Componente responsável por autenticar os utilizadores e enviar informação do utilizador (ex.: email, nome do utilizador) para o serviço após uma autenticação com sucesso.

- **Fornecedor de Serviço**

Componente responsável por requisitar a autenticação do utilizador ao fornecedor de identidade e por autorizar o utilizador tendo por base a informação pessoal do utilizador enviada pelo fornecedor de identidade.

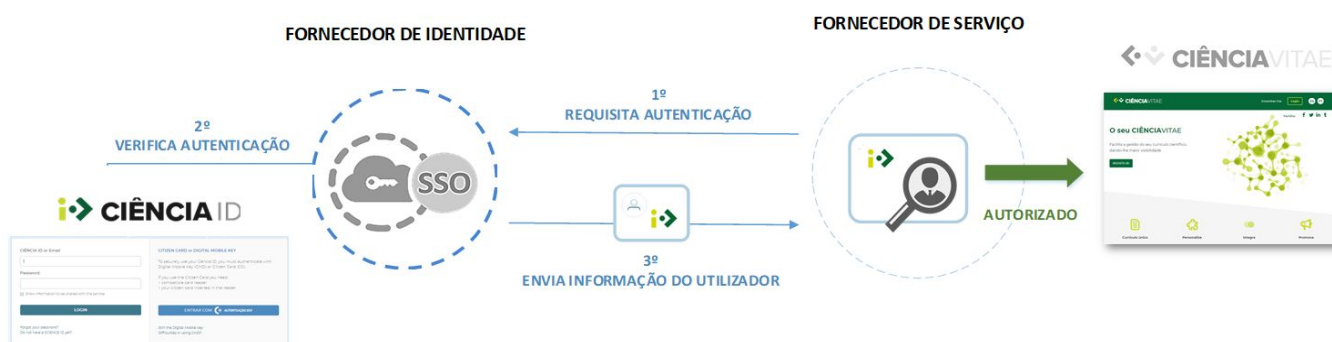


Figura 3. Serviço de autenticação CIÊNCIA ID

Quando um utilizador acede a um serviço e realiza a autenticação no CIÊNCIA ID os seguintes passos ocorrem:

1. O utilizador acede à página do serviço e clica em “Login”.

<Quadro Normativo PTCRIS>

2. O serviço redireciona o utilizador para o Login do CIÊNCIA ID.
3. O utilizador insere as suas credenciais no CIÊNCIA ID.
4. O CIÊNCIA ID verifica as credenciais do utilizador. Se a autenticação é realizada com sucesso a informação do utilizador pode ser enviada para o serviço. Cada serviço recebe apenas os dados pessoais do utilizador previamente acordados com o CIÊNCIA ID.
5. O serviço recebe a informação do utilizador e decide se autoriza ou não o seu acesso com base na informação recebida.

A FCT|FCCN é responsável por gerir o Fornecedor de Identidade do CIÊNCIA ID (login.ciencia-id.pt).

Todos os serviços que pretendem integrar a autenticação do CIÊNCIA ID tem necessariamente de instalar e configurar um Fornecedor de Serviço SAML2. Para o efeito, o CIÊNCIA ID disponibiliza o documento "[CIÊNCIA ID - SAML2 Requisitos Técnicos](#)".

2. Base de dados de Organizações

A base de dados (BD) de organizações gerida pela Ringgold é a infraestrutura subjacente ao sistema ISNI+. Esta BD inclui mais de 500.000 registos de identificadores organizacionais e metadados associados.

No âmbito da adoção do sistema ISNI+, o SGP deverá utilizar a BD de organizações Ringgold para gerir organizações. Para o efeito, o PTCRIS disponibiliza:

- a) Documentação de apoio à utilização da BD (modelo de dados, classificação e tipologias)
- b) Acesso FTP a uma cópia atualizada⁹ dos registos Ringgold referentes a organizações portuguesas.

O acesso à BD Ringgold via FTP permite o *bulk download* dos dados atualizados¹⁰, bem como a criação de listas *drop-down* (Fig.1). A coleção de

⁹ A atualização periódica da BD Ringgold ocorre semanalmente, à sextas-feira.

¹⁰ Um exemplo dos dados disponibilizados por esta via está em <https://tinyurl.com/y8v92nzw>

<Quadro Normativo PTCRIS>

dados disponibilizada inclui os códigos ISNI+, nome oficial, dados de localização e códigos locais (código DGEEC, código FCT e NIPC).

O acesso à BD Ringgold ocorre mediante criação de uma conta na plataforma Sharefile e é condicional à assinatura de um protocolo de acordo com os termos e condições da Ringgold..

3. SciPROJ

A SciPROJ é uma base de dados única e agregadora de todas as referências de financiamento (projectos, pessoas e instituições) que sustentam a investigação científica em Portugal. As fontes de financiamento podem ser nacionais ou internacionais.

À data, constam da BDFund todas as referências de financiamento a projectos atribuídas pela FCT¹¹. O enriquecimento da SciPROJ está em curso, prevendo-se que desta venham a constar informação sobre os demais instrumentos de financiamento geridos pela FCT, assim como, de outras agências de financiamento nacionais e do CORDIS.

A SciPROJ é uma BD CERIF que assenta no perfil de dados OpenAIRE (vide [Caracterização da SciPROJ](#)). A SciPROJ disponibiliza um serviço de REST API que permite partilhar os dados SciPROJ à comunidade. As especificações deste serviço, podem ser consultados no respetivo [Manual de Utilização](#).

4. APIs CIÊNCIAVITAE

O quadro normativo PTCRIS define uma API (v1.1) standard para acesso e partilha programática de informação de suporte à atividade científica, nomeadamente, no que diz respeito a pessoas, organizações, afiliações e informação curricular. A consulta da descrição pormenorizada dos serviços poderá ser efetuada com recurso ao gerador de [documentação](#) Swagger UI. Na tabela 1 estão indicados alguns dos principais métodos da API.

¹¹ Fundação para a Ciência e a Tecnologia

<Quadro Normativo PTCRIS>

As respostas retornadas pela API podem estar num de dois formatos: xml ou json.

Tabela 1. Descrição sumária dos principais métodos da API PTCRIS (v1.1)

Entidade	Método	Tipo	Descrição
Pessoas	search-persons	GET	Pesquisa de pessoas por nome ou identificadores
Organizações ¹²	search-institutions	GET	Pesquisa de instituições por nome na base de dados Ringgold.
Afiliações	<code>curriculum/{ciencia-id}/employment</code>	GET	Função para retornar todas as afiliações de um determinado Ciência ID
Currículos	curriculum/{ciencia-id}	GET	Função para retornar o currículo de um determinado Ciência ID
	curriculum/{ciencia-id}/pdf	GET	Função para retornar o currículo de um determinado Ciência ID em formato PDF

Outras Implementações

Usando o serviço de API PTCRIS, qualquer componente do ecossistema poderá popular o currículo de um dado utilizador, mediante a autorização deste. Na tabela 2 está descrito um caso de uso e métodos da API associados. os métodos da API no âmbito de um caso de uso que se traduz no reconhecimento da atividade de avaliação.

Tabela 2. Métodos da API PTCRIS (v1.1) associados a casos de uso

Casos de Usos	Método	Tipo	Descrição
Reconhecimento da atividade de avaliador	curriculum/{ciencia-id}/service	POST	Função para criar um registo na área funcional “Atividades” do currículo de um determinado Ciência ID

¹² O acesso à API é condicional à assinatura de um protocolo de acordo com os termos e condições da Ringgold.

<Quadro Normativo PTCRIS>

curriculum/{ciencia-id}/service/{id}	PUT	Função para atualizar um dado registo na área funcional “Atividades” do currículo de um determinado Ciência ID
--	-----	--

Requisitos PTCRIS - Tabela Resumo

Entidade	Designação	Requisitos
Quadro Normativo		
Identificadores		
Pessoas	CIÊNCIA ID	Associar Ciência ID nas bases de dados com referência a investigadores. Esta associação deve ser feita através de <i>log in</i> no Ciência ID para comprovação da titularidade do identificador.
Pessoas	ORCID	Associar ORCID nas bases de dados com referência a investigadores. Esta associação deve ser feita através de <i>log in</i> no ORCID para comprovação da titularidade do identificador.
ORGs	ISNI+	Associar códigos ISNI+ nas bases de dados com referência a organizações
Publicações	DOI	Associar códigos DOI nas bases de dados com referência a publicações
Projetos/Financiamento	DOI	Associar códigos DOI nas bases de dados com referência a projetos/financiamento
Teses e Dissertações	TID	Associar códigos TID nas bases de dados com referência a teses/dissertações
Semântica		
Publicações	COAR	Utilizar o vocabulário controlado COAR para representar tipos de produtos científicos, nomeadamente publicações
Projetos/Financiamento	CERIF	Utilizar a semântica CERIF para representar a informação sobre a dimensão projetos e dimensões associadas (pessoas, organizações e financiamento)
Currículos	CIÊNCIAVITAE	Utilizar a semântica CIÊNCIAVITAE para representar informação curricular
Modelo de Dados		
	CERIF	Recomenda-se a utilização do modelo de dados CERIF-xml como modelo de dados
Perfis de Dados		
Projetos/Financiamento	OpenAIRE	A informação sobre projetos/financiamento deve ser partilhada segundo as guidelines OpenAIRE
Currículos	xml/API	A informação curricular deve ser acedida através das APIs disponibilizadas, nos formatos xml e pdf
Infraestruturas		
Pessoas	CIÊNCIA ID -	A autenticação no sistema deve ser feita via serviço de

<Quadro Normativo PTCRIS>

	Serviço de Autenticação	autenticação CIÊNCIA ID através do <i>sign in/log in</i> na plataforma
Pessoas	API	A pesquisa de pessoas pelo nome ou identificador deve fazer-se invocando a API PTCRIS disponibilizada
Organizações	Ringgold - Tabela de Autoridade API	A pesquisa de organizações deve fazer-se através do acesso à tabela de autoridade Ringgold seja via cópia local da mesma ou via API
Pessoas-ORG	API	A pesquisa de afiliações (relação Pessoa-Organização) deve fazer-se usando a API PTCRIS
Currículos	API	A pesquisa de currículos deve fazer-se usando a API PTCRIS