

PTCRIS

Connected Research

**<IMPLEMENTAÇÃO DE
IDENTIFICADORES DE ORGANIZAÇÕES
NO ECOSISTEMA PTCRIS>**

Ficha Técnica

Título

Implementação de Identificadores de Organizações no ecossistema PTCRIS

Autoria

Cátia Laranjeira

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

Edição

João Moreira

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

Histórico de versões

Versão	Data
v0.1	Jun 2020

Índice

Enquadramento	3
Macroplano de trabalho - ORGs@PTCRIS	3
Fase I - Trabalho realizado/Resultados	4
API de pesquisa	8
Acesso FTP a uma cópia atualizada dos registos Ringgold	8
Serviço de reconciliação de dados	8
Serviço de atualização da tabela de autoridade Ringgold	9
Casos de uso	9
Fase II	9
Iniciativas recentes	9
Anexo A	11
Princípios orientadores do SGO	11

<Implementação de ORG IDs>

Enquadramento

O programa [PTCRIS](#) 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 de interoperabilidade, um conjunto de princípios e regras que regem a interoperabilidade entre os componentes do ecossistema. Um desses princípios diz respeito a **identificadores únicos**, já que 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, nomeadamente Investigadores, **Organizações**, Projetos/Financiamentos, Produções e Infraestruturas científicas.

Macroplano de trabalho - ORGs@PTCRIS

Em 2015, no âmbito do programa PTCRIS, foi constituído um grupo de trabalho (GT) dedicado à gestão das organizações que operam no ecossistema científico-tecnológico nacional. Este GT, constituído por representantes da Universidade de Évora, ISCTE-IUL e FCTIUCCN, definiu um plano de trabalhos faseado, com os seguintes objetivos:

Fase I

- Definir o sistema de identificadores de organizações (OrgIDs) **único e persistente** a adoptar de acordo com as melhores práticas nacionais e internacionais
- Promover o uso de OrgIDs através do uso de uma **tabela de autoridade**
- Garantir a **precisão** dos dados, no que diz respeito a organizações portuguesas
- Desenvolvimento de serviços para gestão de organizações

<Implementação de ORG IDs>

Fase II

- Desenvolvimento de um sistema de registo e gestão de organizações

Fase I - Trabalho realizado/Resultados

Desde 2015 até à data foram realizadas várias tarefas tendo em vista o cumprimento dos objetivos da Fase I. Estas tarefas são descritas abaixo:

Objetivo 1

Definir o sistema de identificadores de organizações (OrgIDs) único e persistente a adoptar de acordo com as melhores práticas nacionais e internacionais

Dando cumprimento a este objetivo foi realizado um [estudo](#) do estado d'arte no país e a definição dos [princípios orientadores](#) da Base de Dados Nacional de Organizações. Na sequência destes e em conformidade com as recomendações do [estudo](#) produzido no âmbito do CASRAI-UK-Org-Id, o PTCRIS adotou o sistema de identificadores organizacionais ISNI+. Este sistema combina a infraestrutura [ISNI](#) (International Standard Name Identifier)¹ com a infraestrutura e os serviços de valor acrescentado assegurados pela [Ringgold](#)². De acordo com este sistema, cada organização/instituição é representada por um par de identificadores ISNI/Ringgold único na base de dados Ringgold.

Objetivo 2

Promover o uso de OrgIDs através do uso de uma tabela de autoridade

O sistema ISNI+ assenta na tabela de autoridade gerida pela [Ringgold](#), uma organização especializada em identificadores organizacionais que opera como agência de registo de identificadores ISNI. A tabela de autoridade inclui mais de 500.000 registos de identificadores organizacionais e metadados associados e é amplamente utilizada por diversos sistemas do ecossistema científico-tecnológico internacional, nomeadamente, o ORCID e sistemas de submissão de manuscritos como o [ScholarONE](#) e [Aries](#). No contexto nacional, à data, a tabela de autoridade Ringgold é utilizada na plataforma de gestão curricular [CIÊNCIAVITAE](#) e no portal [Study & Research in Portugal](#) (Figura 1).

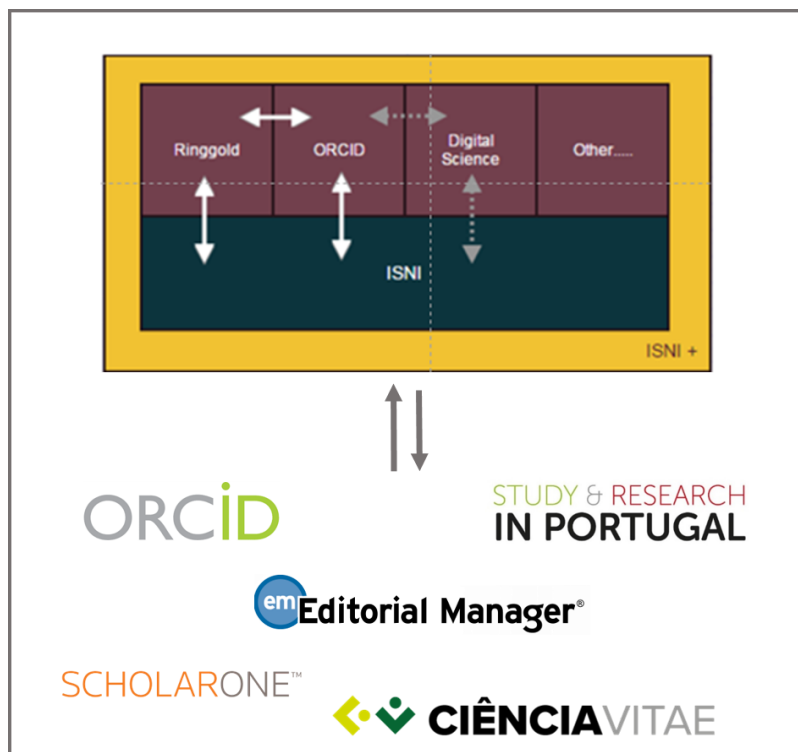
¹ Norma ISO amplamente utilizada para distinguir pessoas e organizações

² A Ringgold é uma organização especializada em identificadores organizacionais que opera como agência de registo de identificadores ISNI.

<Implementação de ORG IDs>

O PTCRIS disponibiliza à comunidade académico-científica acesso à tabela de autoridade Ringgold (*vide* [Objetivo 4 - Acesso FTP a uma cópia atualizada dos registos Ringgold](#)).

Figura 2. Sistema ISNI+ e a sua integração com sistemas de informação nacionais e internacionais



Objetivo 3

Garantir a precisão dos dados, no que diz respeito a organizações portuguesas (ORGs PT)

Ao abrigo do quadro normativo PTCRIS, todas as organizações que operam no ecossistema académico-científico devem ser representadas por um par de identificadores ISNI/Ringgold único na base de dados Ringgold. No cumprimento deste objetivo é necessário assegurar:

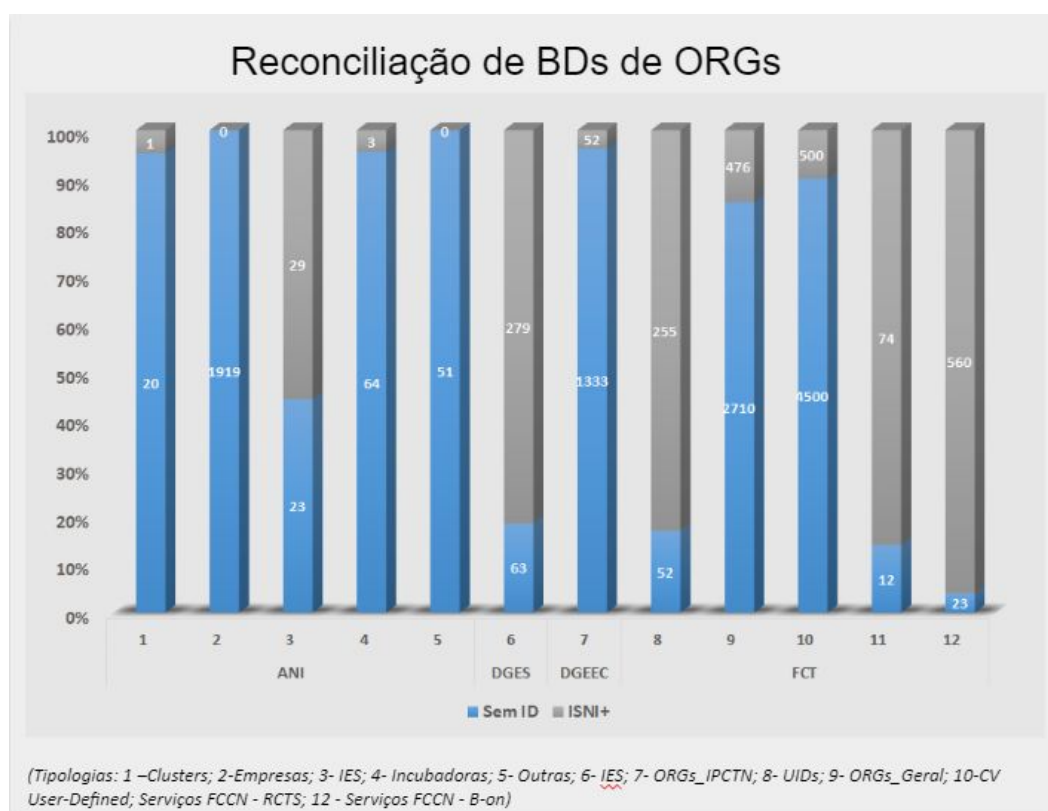
- a) O mapeamento das ORGs PT na tabela de autoridade Ringgold

<Implementação de ORG IDs>

No âmbito do trabalho previsto na Fase I, foi realizado o mapeamento de várias bases de dados de organizações de múltiplas entidades, nomeadamente, a DGGECC, FCT e ANI. O resultado deste mapeamento está ilustrado na figura 3.

Para realização do mapeamento das ORGs PT na tabela de autoridade Ringgold, o PTCRIS disponibiliza um serviço de reconciliação de dados, baseado na ferramenta *open-source* [OpenRefine](#) (vide [Objetivo 4 - Serviço de reconciliação de dados](#)).

Figura 3. Mapeamento de múltiplas várias bases de dados de organizações nacionais com a tabela de autoridade Ringgold



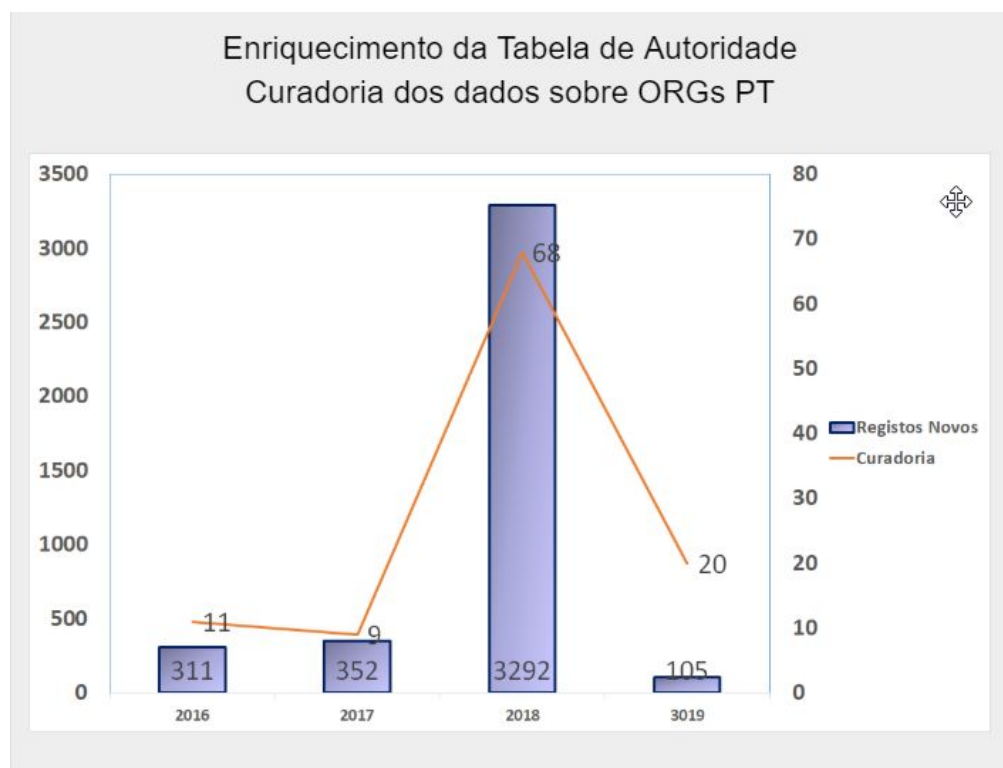
b) A curadoria dos dados referentes a ORGs PT que constam da tabela de autoridade Ringgold.

Os resultados da curadoria de dados realizada está indicada na figura 4.

No âmbito da fase I, a atualização dos dados que constam da tabela de autoridade Ringgold é mediada pela equipa PTCRIS (vide [Objetivo 4 - Serviço de atualização de dados](#)).

<Implementação de ORG IDs>

Figura 4. Atualização de dados referentes a ORGs nacionais na tabela de autoridade Ringgold



c) Enriquecimento da tabela de autoridade Ringgold com dados referentes a ORGs PT que aí não estejam representadas.

A análise quantitativa do enriquecimento de dados realizado no período 2016-2019 está representada no gráfico 4.

No âmbito da fase I, o enriquecimento de dados referentes a ORGs PT na tabela de autoridade Ringgold é mediada pela equipa PTCRIS (*vide* [Objetivo 4 - Serviço de atualização de dados](#)).

Objetivo 4

Desenvolvimento de serviços para gestão de organizações

De acordo com o quadro normativo PTCRIS, no decurso da fase I do macroplano, deve garantir-se que as ORGs PT que operam no ecossistema académico-científico estão associadas a um identificador ISNI+, e, conseqüentemente, representadas na tabela de autoridade Ringgold. Esta tabela poderá ser utilizada para gerir organizações. Para o efeito, o PTCRIS disponibiliza os seguintes serviços:

<Implementação de ORG IDs>

a) API de pesquisa

A API PTCRIS permite pesquisar organizações na BD Ringgold, com base no nome ou nome alternativo (por exemplo, acrónimos, nomes antigos, variantes de idioma, entre outros).

Na eventualidade da organização não se encontrar representada na BD autoritativa, deve ser dada ao utilizador a possibilidade de “Criar Instituição”.

A lista destas organizações user-defined deverá ser enviada por email para a equipa PTCRIS para validação.

b) Acesso FTP a uma cópia atualizada dos registos Ringgold

O acesso à BD Ringgold via FTP permite o *bulk download* dos dados atualizados³, bem como a criação de listas *drop-down*. A coleção de dados disponibilizada inclui a lista de organizações portuguesas e respetivos códigos ISNI+, nome oficial, dados de localização e códigos locais (código DGEEC, código FCT e NIPC). O acesso à lista de organizações completa (nacionais e estrangeiras) é condicional à assinatura de um protocolo de acordo com os termos e condições da Ringgold

O acesso à BD Ringgold ocorre mediante criação de uma conta na plataforma Sharefile. Para solicitar a criação de uma conta sharefile envie-nos um [email](#).

No âmbito deste serviço, o PTCRIS disponibiliza a seguinte documentação de apoio à utilização da BD:

- Modelo de Dados
- Tipologias de Organizações
- Classificação de Organizações
- Cópia da BD à data de 06-12-17

c) Serviço de reconciliação de dados

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

³ A atualização periódica da BD Ringgold ocorre semanalmente, ao domingo.

<Implementação de ORG IDs>
open-source [OpenRefine](#). No âmbito deste serviço, o PTCRIS disponibiliza [documentação de suporte](#) e o URL do serviço de reconciliação.

d) Serviço de atualização da tabela de autoridade Ringgold

O PTCRIS disponibiliza um serviço que assegura a compilação dos dados e envio para a equipa de curadoria de dados Ringgold. Tal ocorre com uma periodicidade de 1-2 vezes por ano. Prevê-se que na fase II este processo seja automatizado.

Casos de uso

No [documento](#) estão descritos 3 casos de uso que ilustram diferentes cenários de adoção do sistema ISNI+, ao abrigo do quadro normativo PTCRIS. No âmbito destes casos de uso, a adoção do sistema ISNI+ assenta na utilização de diferentes serviços PTCRIS.

Fase II

No âmbito da Fase II do projeto, está previsto o desenvolvimento de um sistema de gestão de organizações nacionais (SGO). Em 2017 foram iniciados os trabalhos nesse sentido. Em particular, foi realizada:

1. a definição dos princípios orientadores do SGO (*vide* [Anexo A](#))
2. a [especificação técnico-funcional](#) do SGO
3. o processo aquisitivo para contratação de serviços especializados para desenvolvimento do SGO
4. a [proposta de implementação](#) do SGO

Os trabalhos foram interrompidos de acordo com as diretrizes da SECTES.

Iniciativas recentes

Em 2019 foi lançado o [Research Organization Registry](#) (ROR), uma iniciativa internacional dedicada ao desenvolvimento de um sistema sustentável e aberto de identificadores únicos de organizações de investigação.

O ROR surge como um registo de organizações de investigação - qualquer organização que conduz, produz ou gere atividades de investigação. Este âmbito está de acordo

<Implementação de ORG IDs>

com o caso de uso que o ROR pretende endereçar - afiliações e a necessidade de associar com precisão a organização ao indivíduo e as suas informações.

Na sua génese, o ROR baseou-se na base de dados de organizações [GRID](#), gerida pela [Digital Science](#) tendo sido atribuídos identificadores ROR a cerca de 91 000 organizações. Tendencialmente, as 2 bases de dados divergirão, à medida que o ROR vai sendo enriquecido.

<Implementação de ORG IDs>

Anexo A

Princípios orientadores do SGO

- O sistema de gestão de ORG tem por objetivo suportar a BD nacional de ORG facilitando o acesso e a gestão da informação;
- É objeto de registo todas as organizações que desenvolvem atividades de ensino; formação e/ou investigação, bem como estabelecimentos prestadores de cuidados de saúde hospitalares;
- O sistema de gestão de ORG deve simplificar os processos de registo das instituições preservando a autonomia das mesmas enquanto fontes de informação;
- O sistema de gestão de ORG deve sincronizar com sistemas internacionais equivalentes;
- A responsabilidade da gestão da informação deve ser partilhada – gestão distribuída;
- O reconhecimento do sistema nacional de gestão de ORG como fonte autoritativa de informação pelos stakeholders deve ser garantido.