



Meeting of Repository Networks COAR May 14 & 15, 2018 - Hamburg, Germany

Current network services: enhancement and curation, text and data mining, and heuristics

- **Lautaro Matas: Gerente Técnico. LA Referencia**
- **Washington Segundo. Coordinador do Laboratório de Metodologias de Tratamento e Disseminação da Informação. IBICT Brasil.**
- **Alberto Cabezas B. Secretario Ejecutivo LA Referencia.**

<http://www.lareferencia.info/>

LA Referencia

- Mission:

To give visibility to the **publicly funded scientific production** in Latin America, through the cooperation and articulation of a **federated network** of institutional repositories, based on **regional agreements and national open access strategies**.

Associates



MINCYT - Ministerio de Ciencia, Tecnología e Innovación Productiva. Argentina



Ibict - Instituto Brasileiro de Informação em Ciência e Tecnologia. Brasil



Colciencias - Ministerio de Educación, Renata. Colombia



CONICYT - Comisión Nacional de Investigación Científica y Tecnológica. Chile



CONARE - Ministerio de Ciencia, Tecnología y Telecomunicaciones. Costa Rica



SENESCYT - Secretaría de Educación Superior, Ciencia y Tecnología. Ecuador



Ministerio de Educación
Viceministerio de Ciencia y Tecnología. El Salvador



CONACYT - Consejo Nacional de Ciencia y Tecnología. México



CONCYTEC - Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica



RedCLARA -
Cooperación
Latinoamericana de
Redes Avanzadas

Service

AGREEMENTS



Interoperability
Regional
Alliances
Projects

GUIDELINES



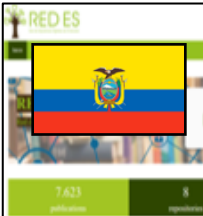
OpenAIRE
LR Application Profile
Information quality
Scientific Data
Guidelines
Training

TECHNOLOGY



Harvesters
Transfer
Pilots
Developments
Community

Transfer



Beta



3.2 version

- Reporting errors for IRs.
- Simplified Administration.
- LR version is the same as nodes.
- It supports current and future guidelines.
- End of Transfer: 2018,
- Potential data harvester.

Roadmap 2016: Common Guidelines Challenges 2018

OpenAIRE

PARTICIPATE SEARCH MONITOR SUPPORT OPEN ACCESS

Find out what OpenAIRE is about

RESEARCHERS
Why Open Access. How to comply. What services to use.

DATA PROVIDERS
How to make your content more visible. What to do to increase quality. How to join.

RESEARCH ADMINS
How to monitor research results. What services to use for your needs.

FUNDERS
Why align policies and practices. How to monitor and analyze results.

EU and Latin America working together towards a common Open Access implementation

OA repositories are connected through regional and thematic networks. Two of the largest regional networks are *OpenAIRE in Europe* and *LA Referencia in Latin America*. Given the truly international and collaborative nature of research, *these networks must also be connected and aligned around issues such as policy, technologies and services*.
LA Referencia and OpenAIRE are collaborating to expand the adoption of guidelines developed through OpenAIRE in Latin America in order to demonstrate of the value of repository interoperability.

- ## Challenges
- New Project OpenAIRE Advance 2018-2021
 - Common statistics
 - Notifications (broker)
 - V. 3.0 to 4.0

OpenAIRE

LAReferencia – Red Federada de Repositorios Institucionales de Publicaciones Científicas Latinoamericanas
PUBLICATION REPOSITORY AGGREGATOR

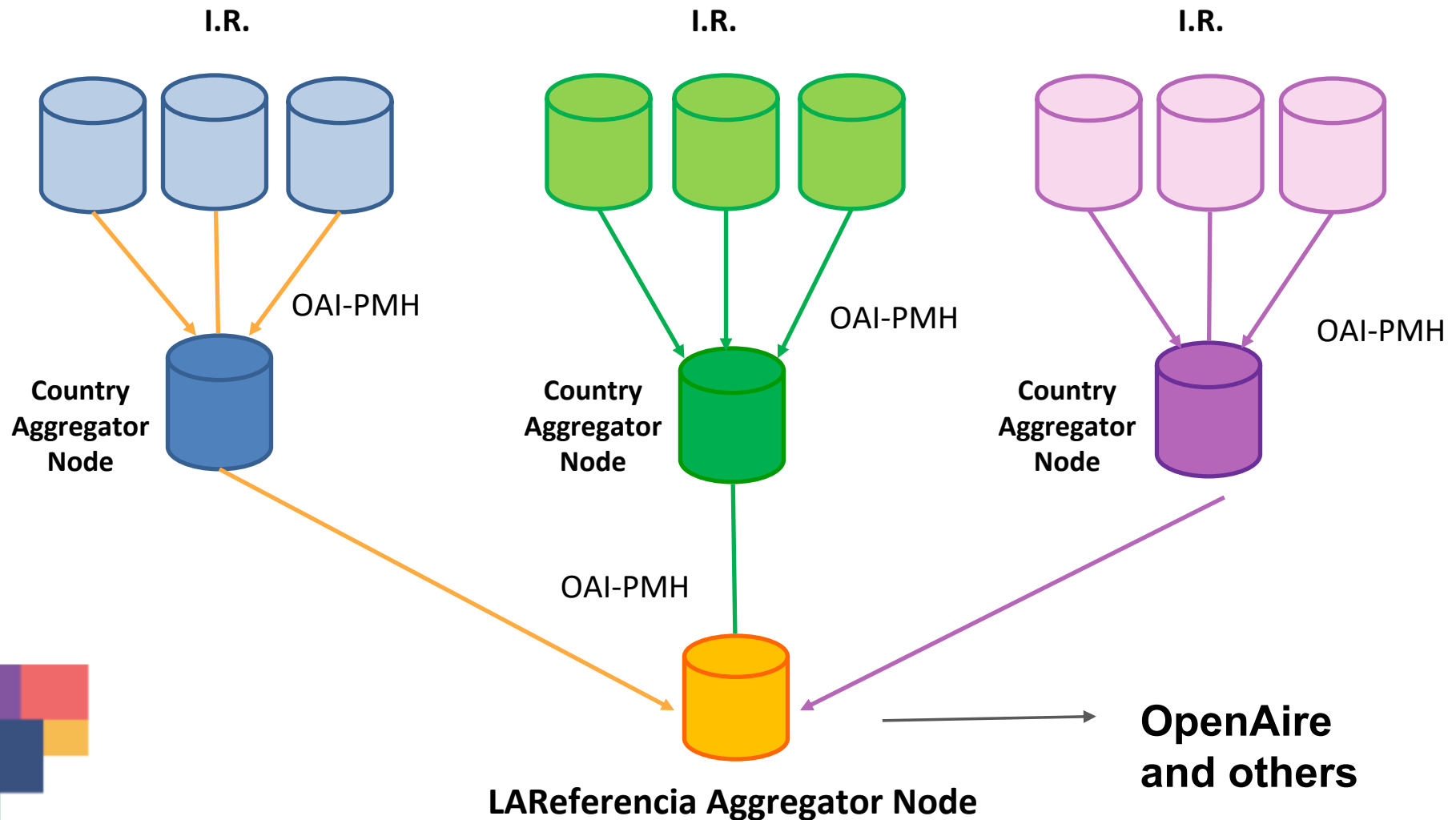
1.259.383 Open Access publications

- 267 EC (H2020/FP7) funded projects outputs identified
- 246 NSF (US) funded projects outputs identified
- 107 FCT (Portugal) funded projects outputs identified
- 56 Wellcome Trust funded projects outputs identified

<http://www.lareferencia.info>

<https://www.openaire.eu/search/dataset/provider?dataset=source&openid=...dcfe224a8b74ce03b0311526fe33e97>

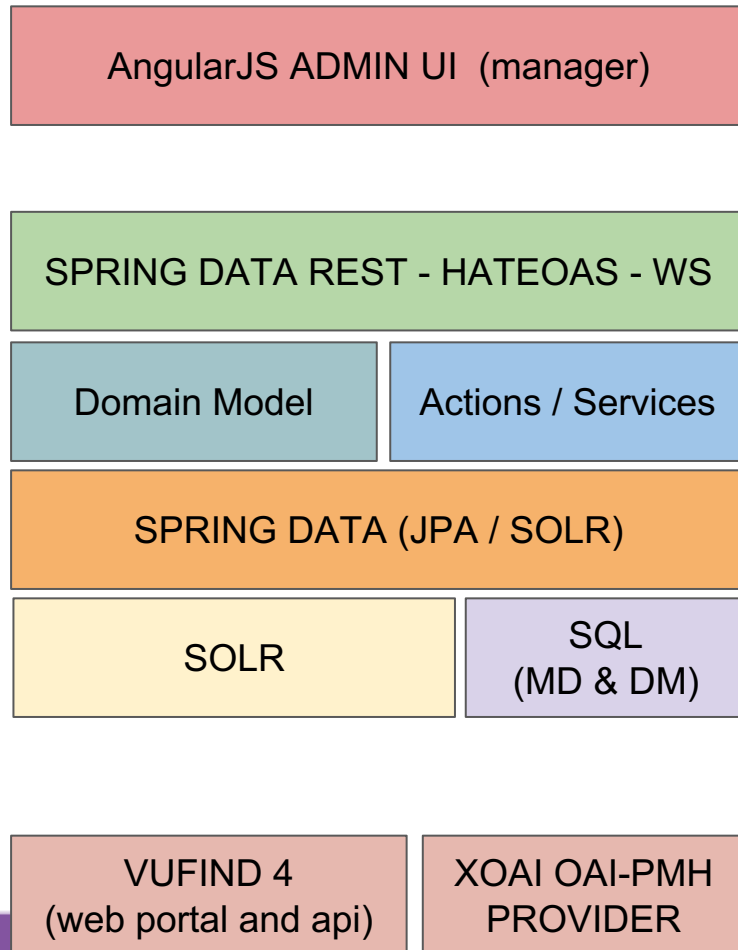
LR Aggregation Model



LA Referencia Harvester

- 5 years of development (2013-2018->)
- Easy to install / maintain
- Scalable: runs in low end laptops or across multiple servers in distributed mode
- GPL 3.0 License – Growing development community
- Currently supporting large repository networks (IBICT/BRASIL)
harvesting/validation/transformation/indexing:
 - 200+ repositories
 - 1.5+ Million records
 - Multiple metadata schemas (standard o custom)
- OpenAire 3.0 (soon 4.0) compatible:
 - Metadata harvesting / validation / Transformation
 - Distributed statistics / OpenAire Broker integration (to be developed 2018 / 2019)

LA Referencia Software Modules

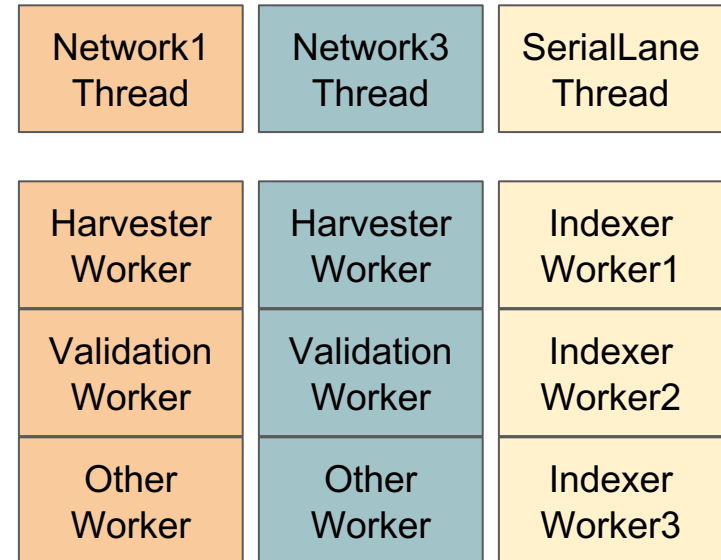


- Java 8 Web Applications
- Spring 5 based
- Spring Data JPA / SOLR
- Hibernate5
- Spring Data REST - HATEOAS - HAL

- All components are integrated with dependency injection.
- Flexible architecture with component replacement/change philosophy.
- Public Web Services
- Private Web Service for admin tasks
- Vufind 4 vanilla biblio core indexing
- Vufind4 plugins for extra field display and export (tesis md fields)
- Vufind4 search and record API

Task Execution Model - Workers

- Multithreading scheduled task model based on Spring TaskScheduler
- Custom TaskManager implemented on top of Spring TS.
- A repository or national node is called **network**
- Networks can be processed using **actions**
- Actions can be launched on demand or scheduled
- An action is accomplished by one or more workers.
- Workers examples are: HarvesterWorker, ValidationWorker, IndexingWorker.
- Base classes are provided (record batch processing)
- Inside a network actions runs sequentially, but in parallel between networks
- Some actions can run in serial mode (SerialLane) to enforce sequential tasks.

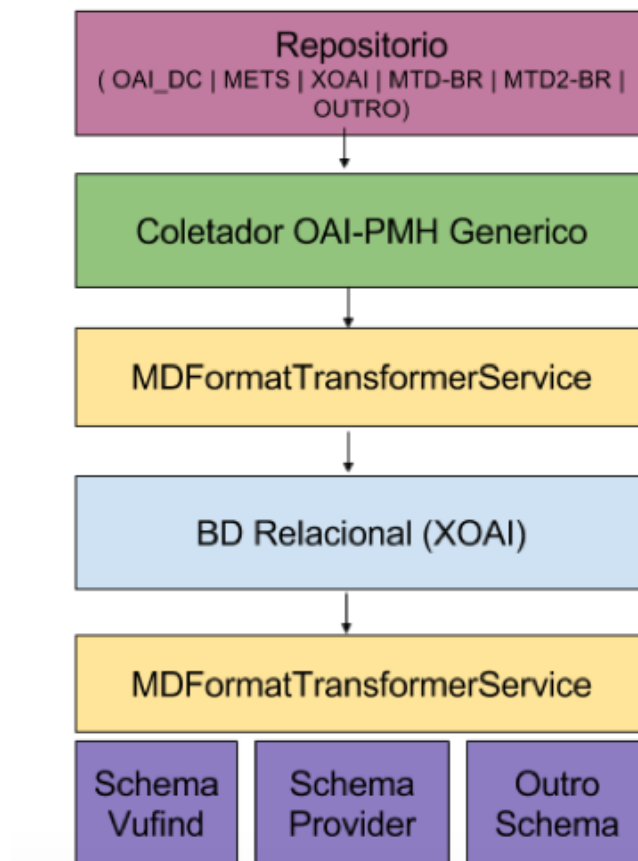


Number of concurrent tasks is configurable. Allows to adapt the software to different hardware scenarios.

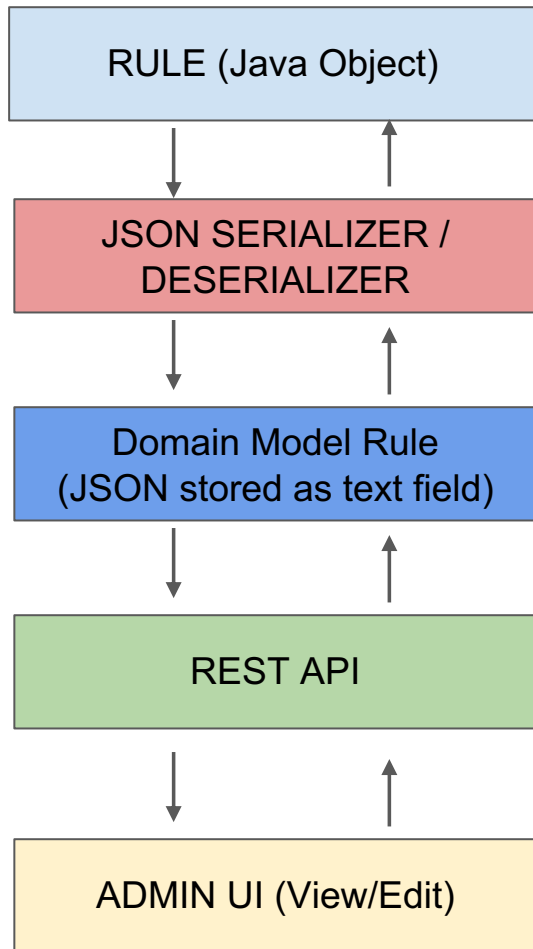
Runs in a 512mb RAM single core processor

XML Schema Transformation Service

- Records are transformed and stored in **XOAI format**.
- XOAI is a “neutral” format than can be easily transformed from/to present and future xml schemas. IE: Datacite/OpenAire4
- Transformations are made by a **configurable metadata schema transformation service**.
- Is easy to add new in/out schema format. (XSLT transformation and xml configuration)
No coding needed!
- Customs schemas can be **harvested/validated/transformed/indexed and shared via OAI-PMH**



Validation / Transformation Rules



- Validation and Transformation are rule based
- Standard rules for validation/transformation are implemented: controlled vocabularies, regexp, search/replace, etc.
- A set of base classes is provided in order to encapsulate common tasks and give developers an easy way to implement more rules.
- The rules are stored as JSON Serialized java objects. It allows to view/edit rules from the web UI.
- Non-technical users can define the rules using a friendly interface.
- Rules are deserialized at validation/transformation runtime.

Validation Rule Editing

Administración de repositorios Editar Validadores Editar Transformadores Logout

Editando: Vali

Aquí puede agreg

Nombre

Validador Nodo

Guardar

Nombre

Creator no vacío

Description no vacío

Relation no vacío

Subject no vacío

Driver:Status

Language iso-63

Identifier URL

Fecha de finaliza
embargo (dc:dat

Driver:Type

Type:Tesis

Filtros México

Edición de regla: Driver:Type

Tipo de regla: Validación por valores controlados

Nombre

Driver:Type

Description

Esta regla evalúa que haya al menos una ocurrencia del vocaburio de tipo driver

¿Es obligatoria?

La regla es determinante en la validez de registro

Cuantificador

Una y sólo una

¿Cuántas ocurrencias deben cumplir la regla?

Esta regla es válida si el campo contiene ocurrencias con estos valores resulta válido

Campo

dc.type

El nombre del campo oai_dc. Ej: dc.type

Valores Controlados

valor

info:eu-repo/semantics/article

valor

info:eu-repo/semantics/masterThesis

Transformation Rule Editing

Administración

Editando: Transformador

Aquí puede agregar

Nombre

Transformador

Guardar

Orden

1	Agregado
1	Reemplazo de valores de vocabulario
2	Reemplazo de valores de vocabulario

Reemplazo de valores de vocabulario Type Driver



Description de la regla

Reemplazo de valores de vocabulario Type Driver



Orden

1



Orden en que se aplicará la transformación

Guardar cambios

Campo de búsqueda

dc.type



El nombre del campo oai_dc donde se buscare el valor. Ej: dc:type

Campo de escritura

dc.type.none



El nombre del campo oai_dc que se creará con la ocurrencia de reemplazo Ej: dc:type

¿Se reemplaza la ocurrencia encontrada?

Indica si se eliminará la ocurrencia en el campo de búsqueda

¿Evaluar como prefijo prefijo?

Indica si el valor de búsqueda se evaluará como prefijo del contenido del campo de búsqueda.

Listado de traducciones

Traducción



Buscar

article



Reemplazo

info:eu-repo/semantics/article



Agregar una nueva regla +



25 50 100

Validation/Transformation result explorer

✓ Validez de registros


Inválidos (1510) +
Válidos (987) +

⚙ Transformación de registros

Sin transformar (2124) +
Transformados (373) +


✓ Reglas Válidas

Filtros Argentina (2497) +
Filtros Brasil (2497) +
Filtros Perú (2497) +
Filtros México (2497) +
Filtros Costa Rica (2497) +
Driver:Status (2497) +
Identifier URL (2497) +
Title no vacío (2497) +
Validación Date (2497) +
Fecha de finalización de embargo (dc:date) (2496) +
Subject no vacío (2489) +
Type:Tesis (2409) +
Language iso-639-3 (2401) +
Relation no vacío (2236) +
Creator no vacío (2192) +
Nivel de acceso (dc:rights) (2134) +
Description no vacía (1970) +
Format (1851) +
Driver:Type (1397) +


 2497
R. Totales

 987
R. Válidos

 373
R. Transf.

 1510
Reg. Inválidos

📊 Nivel de cumplimiento de reglas

 Exportar CSV

Nombre ↕	Descripción	Mandatoria ↕	% Cumplimiento ↕	#R. Válidos ↕	
Type:Tesis	Si el elemento dc.type es = a tesis entonces dc.contributor es obligatorio.	No	96.48%	2409	Detalles
Fecha de finalización de embargo (dc:date)	Fecha partir a partir de la cual el OD estará en AA. Solo si el campo Nivel de Accesibilidad (dc:rights) informa OD bajo período de embargo (info:eu-repo/semantics/embargoedAccess)	No	99.96%	2496	Detalles
Driver:Type	Esta regla evalúa que haya al menos una ocurrencia del vocaburio de tipo driver	Si	55.95%	1397	Detalles
Driver:Status	Esta regla evalúa que haya al menos una ocurrencia del vocaburio de status driver	Si	100.00%	2497	Detalles
Identifier URL	Debe existir al menos una ocurrencia de dc:identifier apuntando a URL válida	Si	100.00%	2497	Detalles
Title no vacío	Verifica que exista dc:title	Si	100.00%	2497	Detalles
Creator no vacío	Verifica que exista dc:creator no vacío.	Si	87.79%	2192	Detalles
Validación Date	Debe existir al menos una fecha válida	Si	100.00%	2497	Detalles

Transformed XML record display

Administración

Language iso-639-1
Type:Tesis (88) +
Subject no vacío (8)
Fecha de finalización
(1) +

Instituciones

UNKNOWN (2497)

Subredes

UNKNOWN (2497)

Detalles de validación: oai:http://redicces.org.sv:10972/1291

registro válido

Validación

XML del registro transformado

```
<metadata
  xmlns="http://www.lyncode.com/xoai"
  xmlns:fn="http://www.w3.org/2005/xpath-functions"
  xmlns:fo="http://www.w3.org/1999/XSL/Format"
  xmlns:xdtd="http://www.w3.org/2005/xdtd-datatypes"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.lyncode.com/xoai http://www.lyncode.com/xsd/xoai.xsd">
  <element name="dc">
    <element name="title">
      <element name="none">
        <field name="value">La moral del sermón de la montaña de Mateo</field>
      </element>
    </element>
    <element name="creator">
      <element name="none">
        <field name="value">Alegre, Xavier, 1940-</field>
      </element>
    </element>
    <element name="subject">
      <element name="none">
        <field name="value">Moral bíblica</field>
        <field name="value">Teología</field>
        <field name="value">Biblia.-N.T.-San Mateo-Comentarios</field>
        <field name="value">Sermón de la montaña-Crítica, interpretación, etc.</field>
        <field name="value">Bienaventuranzas</field>
      </element>
    </element>
    <element name="publisher">
      <element name="none">
        <field name="value">Universidad Centroamericana "José Simeón Cañas"</field>
      </element>
    </element>
  </element>
```

Exportar CSV

les

les

les

les

les

les

les

les

les

les

les

les

les

les

les

les

les

LRSoftware 3.2 Highlights

(2017 - actual version)

- Incremental:
 - Harvesting
 - Validation
 - Transformation
 - Indexing (SOLR, VUFIND, XOAI Provider)
- Full text download and indexing
- Vufind 4.0 Indexing and search
- Author ID and CV metadata linking (when available)
- JSON Search/Result API
- Custom metadata fields (IE: BDTD Thesis specific fields)
- CV Data links

LRHarvester 3.3

(now in test phase)

- **LA Referencia - IBICT (Brasil) / RCAPP (Portugal) source code integration. Bitbucket code repository / Common issue reporting / Testing and Documentation**
- GPL 3.0 Software License / Open Source Community
- Custom repository attributes profiles:
 - Examples: geographic coordinates, URL, repository manager data, notes, tags
 - Custom fields and profiles can be added without via XML configurations
 - Customizable Admin UI with tabs and controls
 - Repository data is indexed in a separated solr core
- Harvesting historic data store: historic harvesting information now is stored in a separated solr core *
- Improvements in fulltext indexing and search
- OpenAire4 Harvesting (alpha): preliminary openaire4/datacite harvesting profile. Metadata transformation, validation and store. *

* Developed by RCAAP Team / Ibict Team

LRHarvester 3.4 / 3.5

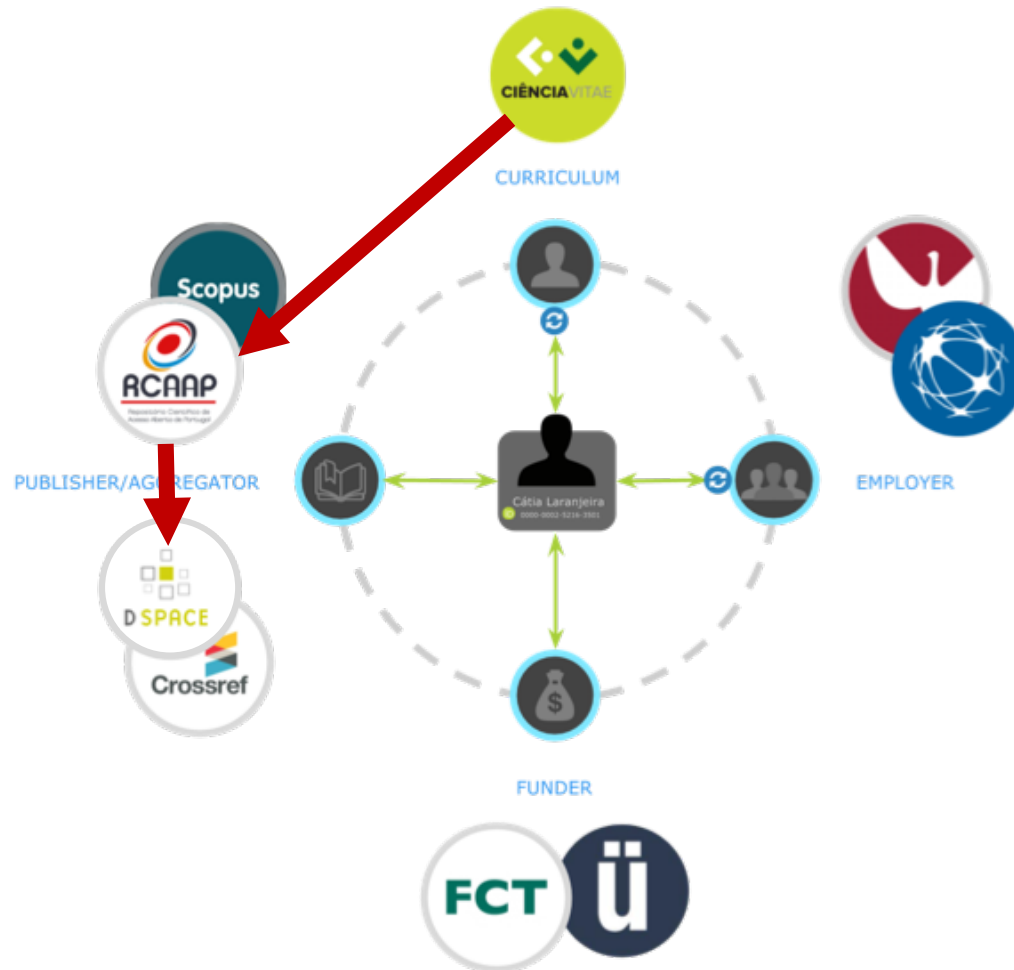
Preliminary Roadmap – 2018/2019

- Distributed statistics (OpenAire Advance Project)
- OpenAire 4.0 Harvesting (final version)
- Data Repositories Harvesting/Integration testing (end 2018)
- Author ID and Funding data link and curation
- OpenAIRE Broker integration (notifications)
- AngularJS to Angular UI migration

Addenda

- Case of Portugal
- Curation

Demonstração



F4 – Curadoria de autor

F5 – Atualização de registo

Relatório de Validação da Agregação

- Validação dos Registos

- Identificação de Melhorias



Administración de repositorios Editar Validadores Editar Transformadores Logout

- Relatório <
- Validade de registos <
Válidos (3) +
- Transformación de registros <
Sin transformar (3) +
- Reglas Válidas <
- Reglas Inválidas <
- Instituciones <
UNKNOWN (3) +
- Subredes <
UNKNOWN (3) +
- Improvement <**
Relations

R. Totales 3 R. Válidos 3 R. Transf. 0 Reg. Inválidos 0

Nivel de cumplimiento de reglas Exportar CSV

Nombre	Descripción	Mandatoria	% Cumplimiento	#R. Válidos
--------	-------------	------------	----------------	-------------

Detalle de registros Exportar CSV

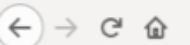
ID	Identifera	Estado	
4816	oai:dev4.rcaap.pt:123456789/1460	Válido	Detalles
4817	oai:dev4.rcaap.pt:123456789/1464	Válido	Detalles
4818	oai:dev4.rcaap.pt:123456789/1467	Válido	Detalles

20

Listagem de Melhorias Identificadas



LAReferencia - Administrac



localhost:8080/backend/home#/diagnose/relation/1



Search



Administración de repositorios Editar Validadores Editar Transformadores Logout

Relatório

Validação e transformação

Tipo de relação

Publicação - Autor

recordsSize
R. Totales

validSize
R. Válidos

transformedSize
R. Transf.

recordsSize -
validSize
Reg. Inválidos

Detalle de registros

ID	RECORD ID	AUTHOR	Estado
4bf507c7-ee45-4...cc2f2	oai:dev4.rcaap.pt:123456789/1460	Laranjeira, Cátia (330)	deleted

20

Informação Pormenorizada

LAReferencia - Administraci X +

localhost:8080/backend/home#/diagnose/relation/1

Administração de repositórios Editar Validadores Editar Transformadores Logout

Relatório

Tipo de relação

Publicação - Autor +

registrosSize - validSize g. Inválidos

Detalhes da relação: 4bf507c7-ee45-4b06-88d0-950584fcc2f2

O registo: [Dissertação da Cátia](#) tem uma relação no Ciência Vitae com o autor: Laranjeira, Cátia

Identificadores do Autor:
ORCID: 0000-0002-5216-3501
Ciência ID: 8C1E-2DFE-D8B5

Cerrar

ID	RECORD ID	AUTHOR	Estado	
4bf507c7-ee45-4b06-88d0-950584fcc2f2	oai:dev4.rcaap.pt:123456789/1460	Carvalho, José (329)	deleted	Details

20

Atualização no Repositório

- **Atualmente:** Registo Manual
- **Futuro:** Registo Automático com aprovação do gestor do repositório

Detalhes da relação: 4bf507c7-ee45-4b06-88d0-950584fcc2f2

O registo: [Dissertação da Cátia](#) tem uma relação no Ciência Vitae com o autor: Laranjeira, Cátia

Identificadores do Autor:
ORCID: 0000-0002-5216-3501

Cerrar

Após Alteração no Repositório

JORNADAS
COMPUTAÇÃO
CIENTÍFICA
2018
INL 11-13 ABRIL



RCAAP | Repositórios Científicos de Acesso Aberto de Portugal

HOME PESQUISA AVANÇADA DIRETÓRIO FINANCIAMENTO AJUDA

CONTACTOS English

Detalhes do Documento

Dissertação da Cátia

Autor(es): Laranjeira, Cátia

Data: 2015

Origem: Repositório de Teste

Assunto(s): Ciências da Informação e Documentação

Descrição

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo. Nemo enim ipsam voluptatem quia voluptas sit aspernatur aut odit aut fugit, sed quia consequuntur magni dolores eos qui ratione voluptatem sequi nesciunt. Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit, sed quia non numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur? Quis autem vel eum iure reprehenderit qui in ea voluptate velit esse quam nihil molestiae consequatur, vel illum qui dolorem eum fugiat quo voluptas nulla pariatur.

Tipo de Documento
Dissertação de Mestrado

Idioma
Português

Documentos Relacionados

Integrar uma rede de repositórios no ecossistema de gestão de ciência nacional

— Carvalho, José; Moreira, João Mendes; Laranjeira, Cátia; Ribeiro, Fernando; Graça, Paulo; Lopes, ...

[mais](#)

Data: 2016

Integrating a national network of institutional repositories into the national/international rese...

— Moreira, João Mendes; Laranjeira, Cátia; Carvalho, José; Ribeiro, Fernando; Lopes, Paulo; Graça, ...

[mais](#)

Data: 2017

Caso de Uso para 2 Cenários

RCAAP | Repositórios Científicos de Acesso Aberto de Portugal

HOME PESQUISA AVANÇADA DIRETÓRIO FINANCIAMENTO AJUDA

Detalhes do Documento

Dissertação da Cátia

Autor(es): Laranjeira, Cátia

Data: 2015

Origem: Repositório de Teste

Assunto(s): Ciências da Informação e Documentação

Descrição

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, ut voluptatem quia voluptas sit aspernatur aut odit aut fugit, sed quia consequuntur magni dolores sequi necessitatibus. Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, et numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea con autem vel eum iure reprehenderit qui in ea voluptate velit esse quam nihil molestiae consequatur, fugiat quo voluptas nulla pariatur.

Tipo de Documento
Dissertação de Mestrado

Idioma
Português

RCAAP | Repositórios Científicos de Acesso Aberto de Portugal

HOME PESQUISA AVANÇADA DIRETÓRIO FINANCIAMENTO AJUDA

Detalhes do Documento

Tese da Cátia

Autor(es): Laranjeira, Cátia

Data: 2018

Origem: Repositório de Teste

Assunto(s): Ciências da Informação e Documentação

Descrição

PTCRIS (Portugal Current Research Information System) é um programa que visa a criação e desenvolvimento sustentado de um ecossistema integrado de informação nacional para apoiar a gestão da investigação, de acordo com os melhores padrões e práticas internacionais. Para alcançar essa visão, o PTCRIS tem dois objetivos principais. O primeiro é definir um quadro regulamentar com base nas melhores normas e práticas internacionais. O segundo é promover a adoção de tal quadro nos diversos sistemas de informação, tanto a nível nacional (incluindo RCAAP) e local (ao nível institucional). Este artigo relata o contexto, estratégia e trabalho desenvolvido até agora para tornar a rede repositórios portugueses (RCAAP), com 50 repositórios, compatíveis com o ecossistema de investigação nacional (PTCRIS), mediante a adoção do seu quadro normativo.

Tipo de Documento
Tese de Doutoramento

Idioma
Português

Documentos Relacionados

Integrar uma rede de repositórios no ecossistema de gestão de ciência nacional

Carvalho, José; Moreira, João Mendes; Laranjeira, Cátia; Ribeiro, Fernando; Graça, Paulo; Lopes, ...

Data: 2016

Integrating a national network of institutional repositories into the national/international rese...

Moreira, João Mendes; Laranjeira, Cátia; Carvalho, José; Ribeiro, Fernando; Lopes, Paulo; Graça, ...

Data: 2017