



Conference: Congreso Interdisciplinario de Energías Renovables -  
Mantenimiento Industrial - Mecatrónica e Informática

*Booklets*



**RENIECYT**

Registro Nacional de Instituciones  
y Empresas Científicas y Tecnológicas

2015-20795

**CONACYT**

LATINDEX - Research Gate - DULCINEA - CLASE - Sudoc - HISPANA - SHERPA UNIVERSIA - E-Revistas - Google Scholar  
DOI - REBID - Mendeley - DIALNET - ROAD - ORCID

**Title:** Componentes de acoplamiento para la infraestructura de soporte de minería de datos de desarrollo de software

**Author:** Edgardo Pérez-Luna

**Editorial label ECORFAN:** 607-8324  
**BCIERMIMI Control Number:** 2016-01  
**BCIERMIMI Classification(2016):** 191016-0101

**Pages:** 21

**Mail:** [edgardo@cenidet.edu.mx](mailto:edgardo@cenidet.edu.mx)  
**RNA:** 03-2010-032610115700-14

**ECORFAN-México, S.C.**

244 – 2 Itzopan Street  
La Florida, Ecatepec Municipality  
Mexico State, 55120 Zipcode  
Phone: +52 1 55 6159 2296  
Skype: ecorfan-mexico.s.c.  
E-mail: [contacto@ecorfan.org](mailto:contacto@ecorfan.org)  
Facebook: ECORFAN-México S. C.

**Twitter:** @EcorfanC

[www.ecorfan.org](http://www.ecorfan.org)

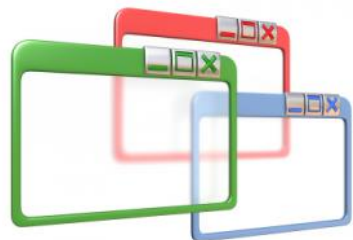
**Holdings**

Bolivia	Honduras	China	Nicaragua
Cameroon	Guatemala	France	Republic of the Congo
El Salvador	Colombia	Ecuador	Dominica
<b>Peru</b>	<b>Spain</b>	<b>Cuba</b>	<b>Haití</b>
Argentina	Paraguay	Costa Rica	Venezuela
Czech Republic			

# Agenda

- Introducción
- Problemática
- Objetivo
- Solución
- Selección de herramientas
- Integración de herramientas
- OSGi
- Implementación
- Pruebas y resultados
- Conclusiones
- Trabajos futuros

# Introducción



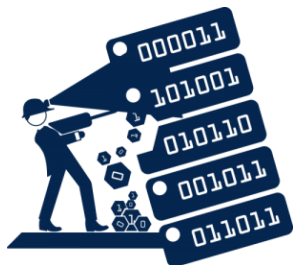
Software



Datos



Estudio



Minería de datos



Mejorar productividad y calidad

# Problemática

- La aplicación del proceso de minería de datos implica el uso de diferentes herramientas que generan archivos con formatos incompatibles, también se realizan actividades o tareas repetitivas que generan trabajo extra.

# Objetivo

- Integrar herramientas mediante el uso de sus bases de datos, con el fin de evitar la captura de datos repetitiva y eliminar la incompatibilidad entre formatos de entrada y salida.

# Solución

- Desarrollo componentes (módulos) Java que monitoreen cada cierto tiempo la base de datos en espera de la inserción de un nuevo registro para posteriormente replicar dicho registro en las demás herramientas.

# Selección de herramientas

- Metodología QSOS (del inglés *Qualification and Selection of Open Source Software*).
- La metodología QSOS consiste en una lista de cuatro pasos dentro de un proceso iterativo:
  - Definir - y organizar lo que será evaluado (criterios)
  - Evaluar - el software contra los criterios definidos
  - Calificar - establecer pesos relacionados al contexto
  - Seleccionar - la herramienta apropiada

# Selección de herramientas

- Herramientas seleccionadas:
- MantisBT
- MySQL
- dotProject
- OpenRefine
- RapidMiner
- LibreOffice (Calc, Writer, Impress, Draw)





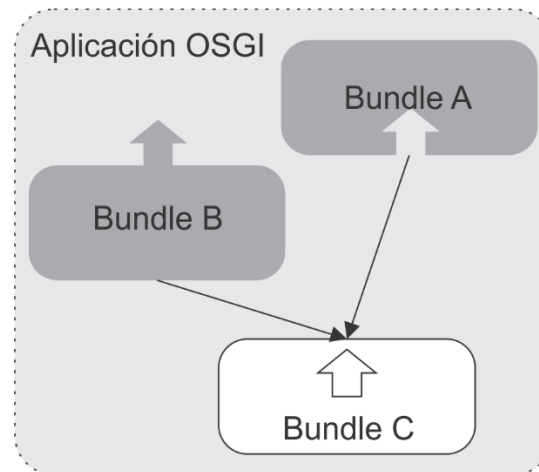
# Integración de herramientas

La integración de herramientas no es una propiedad únicamente de un sola herramienta pero sí de sus relaciones con otros elementos en el ambiente.  
(Thomas, 1992).

- Plataforma
- Presentación
- Datos
- Control (notificaciones)
- Proceso (rol)
- API's

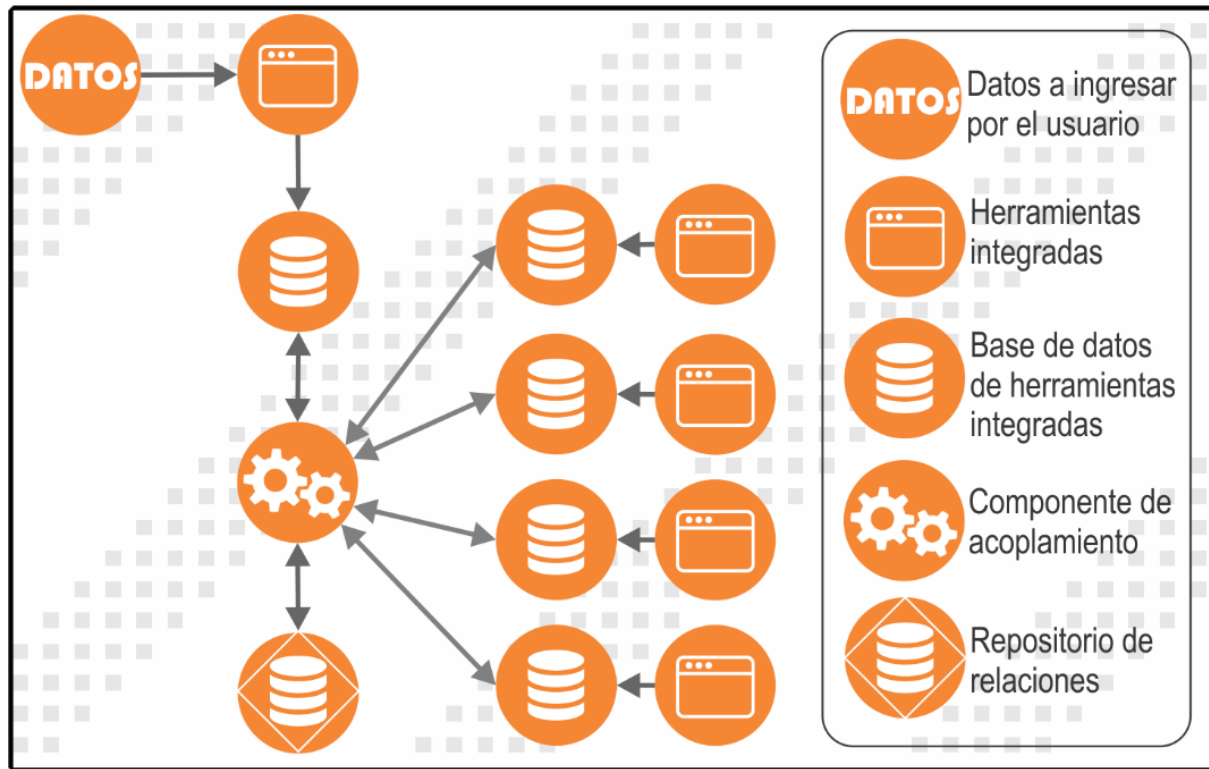
# OSGi

- *Open Services Gateway Initiative (OSGi)*, creado por la *OSGi Alliance* en 1999, es un consorcio de empresas líderes en tecnología.
- La iniciativa OSGi define un sistema modular dinámico para Java.
- *Bundle* es una unidad física de modularidad en la forma de un archivo JAR (por sus siglas en inglés *Java ARchive*) que contiene código, recursos y metadatos (Hall, 2011).



# Implementación

- Modelo de integración



# Implementación

- Interfaz de usuario

The screenshot displays the AMDADS user interface. At the top left, there is a yellow header with the AMDADS logo and a hamburger menu icon. Below the header, a sidebar on the left contains a navigation menu with items: Perfil, Proyectos, Metodologías MD, Herramientas, and Ayuda. The main content area features a user profile for 'daniel' with a blue and orange avatar. Below the name, it states 'Al momento tienes 2 proyectos asociados, para ver la lista selecciona la opción "Proyectos" del menú de la izquierda.' and provides contact information: 'Datos de contacto: Daniel Aguilar, daniel@cenidet.edu.mx'. On the right side, there is a dark blue sidebar with a calendar for December 2015 and a 'Bienvenido a AMDADS' notification box with instructions on how to close it.

# Pruebas y resultados

Requerimiento	Componente de Acoplamiento	Caso de Prueba
Integrar Usuarios	CAUsuarios	CA-TC-01
Integrar Proyectos	CAProyectos	CA-TC-02
Integrar Tareas	CATareas	CA-TC-03
Integrar Archivos	CAArchivos	CA-TC-04
	CANucleo	CA-TC-05
	CAI	CA-TC-06
	CAStarter	CA-TC-07

Elemento a probar	Características	Caso de prueba
Interfaz	<ul style="list-style-type: none"> <li>Autenticación de usuarios</li> <li>Listado de proyectos asociados</li> </ul>	CA-TC-08

# Pruebas y resultados

- Reporte de pruebas

Requerimiento	Componente de Acoplamiento	Caso de Prueba	Resultado de la Prueba
Integrar Proyectos	CAProyectos	CA-TC-01	Exitoso
Integrar Usuarios	CAUsuarios	CA-TC-02	Exitoso
Integrar Tareas	CATareas	CA-TC-03	Exitoso
Integrar Archivos	CAArchivos	CA-TC-04	Exitoso
-	CANucleo	CA-TC-05	Exitoso
-	CAI	CA-TC-06	Exitoso
-	CASarter	CA-TC-07	Exitoso
-	Interfaz de usuario	CA-TC-08	Exitoso

# Conclusiones

- En esta investigación se demostró la factibilidad de la implementación del ambiente de soporte de minería de datos de desarrollo de software (AMDADS).
- El ambiente resultante, de esta investigación, no incluye modificación o eliminación de registros (proyectos, usuarios, etc.), pero se muestra el beneficio del uso de un ambiente integrado.
- Se demostraron los beneficios de implementar los componentes de acoplamiento basados en las especificaciones OSGI, entre los que están: el versionado de componentes y provee una clara definición de su modularidad.

# Conclusiones

Durante el proceso de desarrollo de AMDADS se afrontaron varios retos, entre los que están:

- El cifrado que manejan las herramientas para almacenar contraseñas.
- La integración de un módulo Python con el componente de acoplamiento CAProyectos desarrollado en Java.
- La extracción de datos acerca de los usuarios y sus proyectos, así como su preparación para hacerlos coincidir con los registros en las diferentes bases de datos.



# Trabajos futuros

- Se recomienda ampliar las funcionalidades necesarias para modificar o eliminar registros (usuarios, proyectos, tareas, archivos), con el fin de promover el uso de AMDADS.
- Determinar la manera de integrar herramientas adicionales, que cubran en su totalidad otras etapas del proceso de minería de datos de desarrollo de software.

# Referencias

- (Meza 2014) Meza Bazán, María del Rosario. (2014, Junio). Especificación de Requerimientos para un Ambiente de Soporte al Proceso de Minería de Datos Aplicado a Repositorios de Datos de Desarrollo de Software. Tesis de maestría en Ciencias Computacionales, CENIDET, Cuernavaca, Morelos.
- (Iturbide 2013) Iturbide Domínguez, Gregorio Emmanuel. (2013, Febrero) Metodología de Preparación de Datos Orientada a Aplicaciones de Epidemiología Basada en el Modelo CRISP-DM. Tesis de maestría en Ciencias Computacionales, CENIDET, Cuernavaca, Morelos.
- (Halkidi, et al 2011) Halkidi, M., Spinellis, D., Tsatsaronis, G., & Vazirgiannis, M. (2011). Data mining in software engineering. *Intelligent Data Analysis*, 15(3), 413-441.
- (Xie, et al 2009) Xie, T., Thummalapenta, S., Lo, D., & Liu, C. (2009). Data mining for software engineering. *Computer*, 42(8), 55-62.
- (Hassan, et al 2010) Hassan, A. E., & Xie, T. (2010, November). Software intelligence: the future of mining software engineering data. In *Proceedings of the FSE/SDP workshop on Future of software engineering research* (pp. 161-166). ACM.

# Referencias

(Xie et al 2010)

Hassan, A. E., & Xie, T. (2010, November). Software intelligence: the future of mining software engineering data. In Proceedings of the FSE/SDP workshop on Future of software engineering research (pp. 161-166). ACM.

(KDNuggets, 2014)

*Portal dedicado a Minería de datos, análisis de negocios y Big Data*  
<http://www.kdnuggets.com> Fecha de consulta: Septiembre 2014

(Chapman et al 2000)

Pete Chapman (NCR), Julian Clinton (SPSS), Randy Kerber (NCR), Thomas Khabaza (SPSS), Thomas Reinartz (DaimlerChrysler), Colin Shearer (SPSS) and Rüdiger Wirth (DaimlerChrysler). CRISP-DM 1.0 Step-by-step data mining guide. SPSS Inc. CRISPMWP-1104 2010

# Referencias

- (Zhang et al 203) Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied Artificial Intelligence*, 17(5-6), 375-381.
- (Thomas 1992) Thomas, I., & Nejme, B. A. (1992). Definitions of tool integration for environments. *Software, IEEE*, 9(2), 29-35.
- (Wasserman 1990) Wasserman, A. I. (1990, January). Tool integration in software engineering environments. In *Software Engineering Environments* (pp. 137-149). Springer Berlin Heidelberg.
- (Hall 2011) Hall, R. S., Pauls, K., McCulloch, S., & Savage, D. (2011). OSGi in action. *Creating Modular Applications in Java*.
- (QSOS, 2014) *Metodología de comparación y evaluación de herramientas FLOSS*  
<http://www.qsos.org> Fecha de consulta: Septiembre 2014
- (Amyot, 2003) Daniel Amyot Introduction to the User Requirements Notation: Learning by Example SITE, University of Ottawa 800 King Edward Ottawa (ON), Canada Revista Computer Networks pág. 285-301 2003/06/21



**ECORFAN®**

**© ECORFAN-Mexico, S.C.**

No part of this document covered by the Federal Copyright Law may be reproduced, transmitted or used in any form or medium, whether graphic, electronic or mechanical, including but not limited to the following: Citations in articles and comments Bibliographical, compilation of radio or electronic journalistic data. For the effects of articles 13, 162,163 fraction I, 164 fraction I, 168, 169,209 fraction III and other relative of the Federal Law of Copyright. Violations: Be forced to prosecute under Mexican copyright law. The use of general descriptive names, registered names, trademarks, in this publication do not imply, uniformly in the absence of a specific statement, that such names are exempt from the relevant protector in laws and regulations of Mexico and therefore free for General use of the international scientific community. BCIERMIMI is part of the media of ECORFAN-Mexico, S.C., E: 94-443.F: 008- ([www.ecorfan.org/](http://www.ecorfan.org/) booklets)