



15th International Symposium on Electronic Theses and Dissertations

RPTD and AMAUTA:



**Repositories that promotes more
visibility of and easier access to open
access scientific information in Peru**

Libio Huaroto
Biblioteca Central, UNMSM
lhuaroto@unmsm.edu.pe
Lima, Perú



Agenda

- Aspectos generales
- Características técnicas
- Estadísticas
- Conclusiones



RPTD y AMAUTA

- Repositorios (services providers) que recolectan los metadatos de tesis, revistas, libros y otros tipos de contenidos de portales peruanos.
- ARC como plataforma tecnológica
Old Dominion University, primer servicio federado de búsqueda de información basado en el protocolo OAI-PMH (E-ciencia, Recolecta)
- Gestionado por la UNMSM
- Actualización semanal (*cada viernes*)



Características

- Software ARC Service Provider (Old Dominion University)
- OAI-PMH 2.0 (protocolo para recolección)
- Dublin Core (metadatos)
- MySQL Server
- Java serlets
- Windows 2003 (Sistema operativo)
- AWStats (estadísticas)
- RSS



Advanced Search

Search all bibliographic fields

Author

Title

Abstract

Search for Records Containing

All the specified terms

Any of the specified terms

Filter Options

Archive

Archive Set

Subject [Type in or Select Interactively](#)

Date Stamp (yyyy/mm/dd)

Discovery Date (yyyy/mm/dd)

Display Options

Group By

Sort By



15th International Symposium on Electronic Theses and Dissertations

RPTD

Red Peruana de Tesis Digitales



PERÚ

Inicio Acerca Directorio Eventos Recursos Contacto

SHARE

Buscar

Búsqueda Avanzada

Listado por Autor

Listado por Título

Listado por Año de Publicación

Listado por Universidad

Universidades Participantes

- Pont. Univ. Católica del Perú
- Univ. Peruana de Ciencias Aplicadas
- Univ. Nac. de Ingeniería
- Univ. Nac. de Piura
- Univ. Nac. Mayor de San Marcos
- Univ. Nacional del Callao
- Univ. Privada del Norte
- Univ. San Martín de Porres
- Universidad Ricardo Palma
- Univ. Católica Santo Toribio de Mogrovejo

Otras Universidades

- Univ. Católica de Santa María
- Universidad ESAN



Networked Digital Library of Theses and Dissertations



Tesis Digitales Brasil



Repositorio Digital Peruano



ETD 2012
12 - 14 Septiembre
Lima, Perú

2011 Universidad Nacional Mayor de San Marcos

cybertesis@unmsm.edu.pe

Estadísticas del Sitio

Cantidad de Registros : 5608

Última Actualización : 03-09-2012

Próxima Actualización : 15-10-2012

Estadísticas del repositorio

Cantidad de registros



15th International Symposium on Electronic Theses and Dissertations

Inicio Instituciones
Acerca Enlaces de Interés

AMAUTA Repositorio Digital Peruano

Listado por Autor Listado por Título Listado por Año de Publicación Listado por Institución

SHARE [Facebook] [Twitter] [Email]

Búsqueda Avanzada

Networked Digital Library of Theses and Dissertations

Red Peruana de Tesis Digitales

ETD 2012
12 - 14 Septiembre
Lima, Perú

© 2011 AMAUTA - Repositorio Digital Peruano
Oficina General del Sistema de Bibliotecas y Biblioteca Central - UNMSM

Estadísticas del Sitio

Cantidad de Registros	: 18842
Última Actualización	: 03-09-2012
Próxima Actualización	: 10-09-2012

Estadísticas del repositorio

Cantidad de registros

[Nueva Búsqueda](#)Registros de Archivo: **Pont. Univ. Catolica del Peru**[Resumido](#) [Detallado](#)Pregunta: **Buscar educacion****Listado por Institución**Pont. Univ. Catolica del Peru(73)
USAT(3)

Univ. Nac. Mayor de San Marcos(174)

Univ. Nac. de Ingenieria(6)

Univ. Nac. de Piura(1)

Univ. Privada del Norte(2)

Univ. San Martin de Porres(1)

Univ. de Ciencias Aplicadas(28)



Página: 1 2 3 4

Título : Impacto de un proyecto de educación ambiental en estudiantes de un colegio en una zona marginal de Lima

Autor : Zeballos Velarde, Mauricio

Publicación : 2011-05-09T05:17:51Z | 2011...

Visitas : 432

Nro de visitas

Título : La enseñanza de las funciones trigonométricas en el quinto grado de educación secundaria

Autor : Vilchez Guizado, Jesús

Publicación : 2011-05-09T05:40:58Z | 2011...

Visitas : 240

Nro de visitas

Aplicación de un diseño metodológico basado en el aprendizaje

Estadísticas del sitio web Amauta - UNMSM

Estadísticas de:
www.amauta.edu.pe

Resumen

Cuándo:
[Histórico Mensual](#)
[Días del mes](#)
[Días de la semana](#)
[Visitas por Horas](#)

Quién:

[Países](#)
 Lista completa

[Servidores](#)
 Lista completa
 Última visita

[Dirección IP no identificada](#)
[Visitas de Robots/Spiders](#)

Lista completa
 Última visita

Navegación:

[Duración de las visitas](#)
[Tipos de ficheros](#)
[Downloads](#)

Lista completa

[Accesos](#)
 Lista completa
 Página de entrada
 Salida

[Sistemas Operativos](#)
 Versiones

Desconocido

[Navegadores](#)

Versiones
 Desconocido

Enlaces:

[Origen de la conexión](#)
 Enlaces desde buscadores
 Sitios de enlace

[Búsquedas](#)

Búsquedas por frases clave
 Búsquedas por palabras clave

Otros:

[Misceláneos](#)

[Códigos de error HTTP](#)

Páginas no encontradas

Última actualización: 13 Sep 2012 - 02:33

Periodo mostrado: Sep 2012

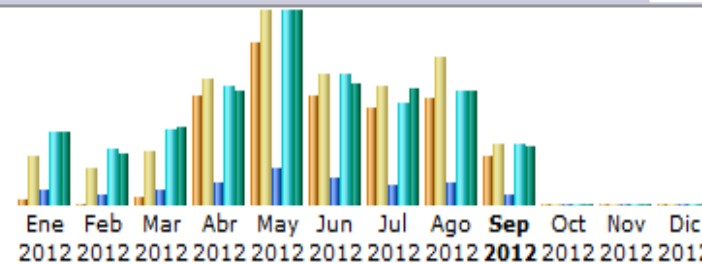


Resumen

Periodo mostrado	Mes Sep 2012				
Primera visita	01 Sep 2012 - 09:25				
Última visita	12 Sep 2012 - 23:55				
	Visitantes distintos	Número de visitas	Páginas	Solicitudes	Tráfico
Tráfico visto *	171	216 (1.26 visitas/visitante)	959 (4.43 Páginas/Visita)	4,676 (21.64 Solicitudes/Visita)	76.30 MB (361.7 KB/Visita)
Tráfico no visto *			12,350	13,556	139.99 MB

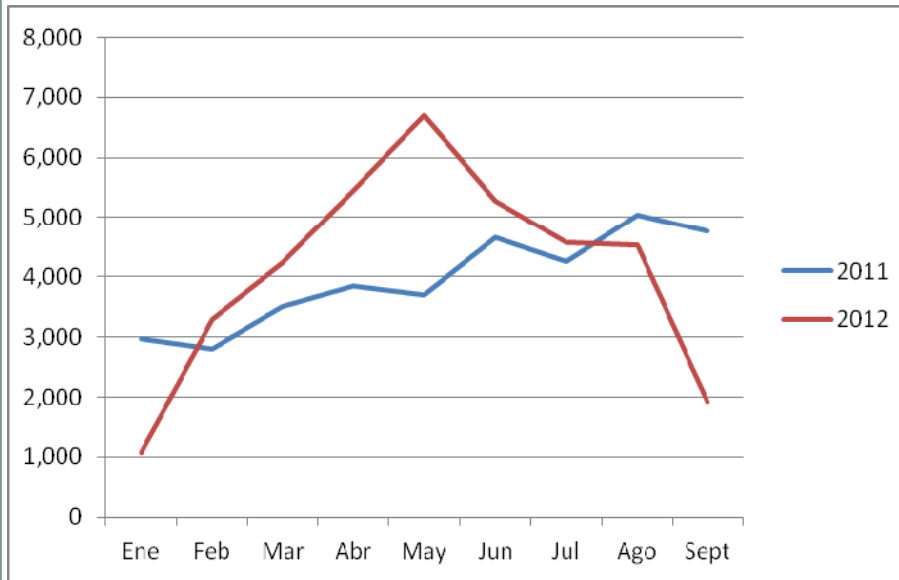
* El tráfico "no visto" es tráfico generado por robots, gusanos o respuestas de código especial de estado HTTP.

Histórico Mensual

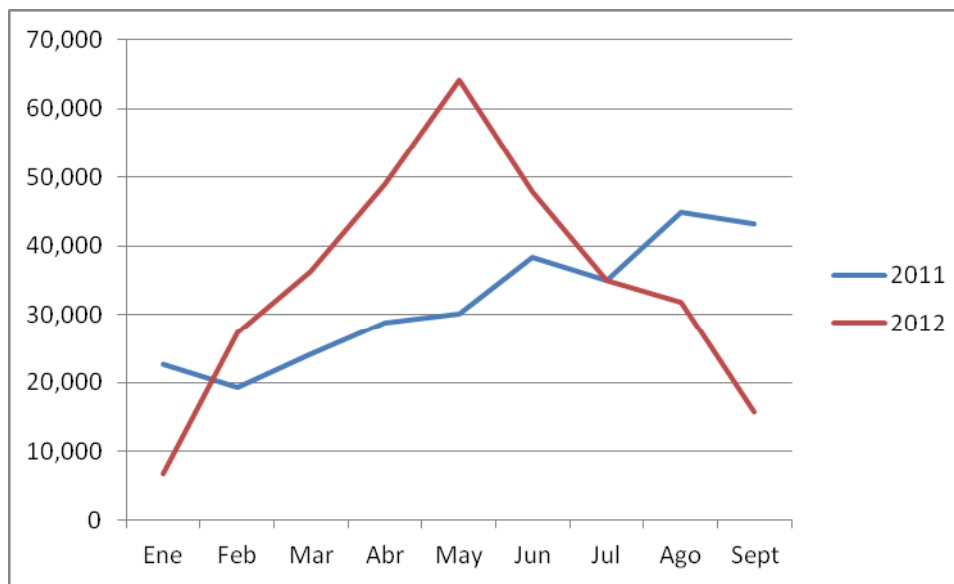
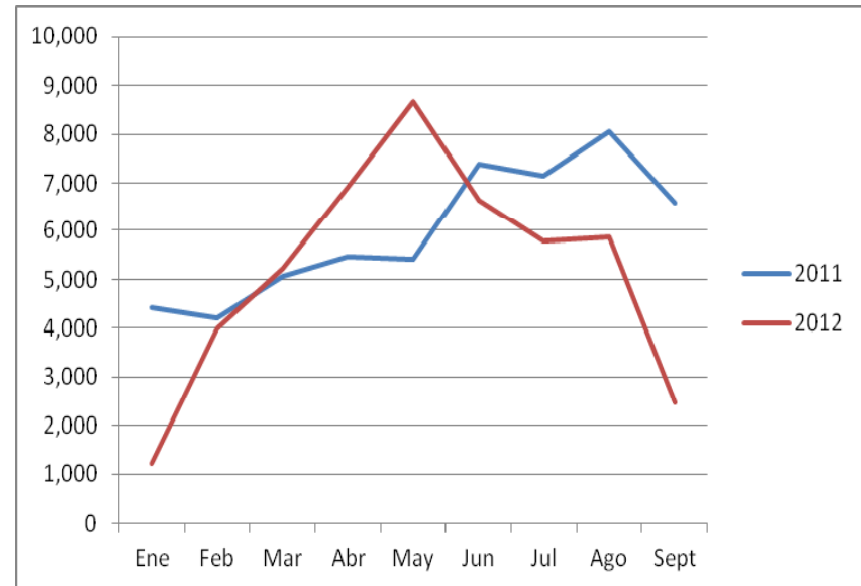


Mes	Visitantes distintos	Número de visitas	Páginas	Solicitudes	Tráfico
Ene 2012	20	170	1,214	5,646	95.11 MB
Feb 2012	7	133	838	4,374	68.16 MB
Mar 2012	30	188	1,269	5,862	100.71 MB
Abr 2012	378	435	1,810	9,177	147.75 MB
May 2012	562	672	2,939	14,977	251.50 MB
Jun 2012	379	453	2,137	10,151	158.82 MB
Jul 2012	336	417	1,633	7,944	151.03 MB
Ago 2012	366	512	1,768	8,681	147.56 MB
Sep 2012	171	216	959	4,676	76.30 MB
Oct 2012	0	0	0	0	0
Nov 2012	0	0	0	0	0

Nro de visitantes



Nro de visitas



Nro de páginas

Estadísticas de RPTD



Facilitating trust-worthy repository use statistics

What is IRS?

A JISC project investigating and developing a joint approach to OAI statistics gathering and sharing.

Who are the Project partners?

Southampton University (UK), University of Tasmania (Australia), Long Island University (USA), Key Perspectives Ltd (UK).

What is the Timescale?

June 2005 - May 2007.

What are the Outcomes?

The [IRStats software package](#), [Evaluation report](#).

What the project is doing

Download data is being logged by every repository as a by-product of the Web requests they receive. This raw data is being and turned into useful download statistics for individual papers and users by a few institutional repositories (e.g. University of Tasmania, Southampton University), thematic repositories (e.g. RePEc) and OAI services (e.g. Citebase). However, there is no consensus over what data needs to be collected, what filtering mechanisms are appropriate, and what analyses are useful for academics in various disciplines. To create effective research statistics services, an interoperable usage statistics service will be created for all OAI-PMH-compliant repositories. This project will investigate the requirements for UK and international stakeholders and build generic collection and distribution software for all



SERVICIOS

- **Citas bibliográficas:**
 - **Citebase** (análisis de citas CC.FF.)
 - **CitEC** (análisis de citas CC.EE.)
- **CC 0 (CC Zero)** (Creative Commons licensing)
- **Repository maps**
- **Estadísticas:**
 - **Google analytics**
 - **Awstats**
 - **Interoperable Repository Statistics (IRStas)**
- **SWORD** (protocol that facilitates the deposit of materials into repositories)



15th International Symposium on Electronic Theses and Dissertations



PROCESOS DE RECOLECCION DE METADATOS EN RPTD Y AMAUTA



Proceso de recolección de metadatos (RPTD y AMAUTA)

Proceso Harvest

Escribe los
metadatos
en la tabla
MySQL
(Cosechamiento)

Proceso Parser

Construye la
tabla
(MySQL) y
los índices

Ejecutar el Harvest

```
C:\ Símbolo del sistema - java scheduler.Harvest
D:\Amauta\build\classes>java scheduler.Harvest
ConnectionPool: JDBCDriver = org.gjt.mm.mysql.Driver
ConnectionPool: JDBCConnectionURL = jdbc:mysql://localhost:3306/amauta
ConnectionPool: ConnectionPoolSize = 5
ConnectionPool: ConnectionPoolMax = 100
ConnectionPool: ConnectionUseCount = 5
ConnectionPool: ConnectionTimeout = 20 seconds
ConnectionPool: (JDBC Property) characterEncoding = UTF-8
ConnectionPool: (JDBC Property) user = root
ConnectionPool: (JDBC Property) useUnicode = TRUE
ConnectionPool: (JDBC Property) password = pol
ConnectionPool: Registering org.gjt.mm.mysql.Driver
Start to Harvest
data source=database
lastharvesttime=2012.08.24 08:41:11
m_lasttime in dataprovider.javaFri Aug 24 08:41:11 GM
lastharvesttime=2008.12.05 14:50:14
m_lasttime in dataprovider.javaFri Dec 05 14:50:14 GM
lastharvesttime=2012.08.07 09:41:53
m_lasttime in dataprovider.javaTue Aug 07 09:41:53 GM
lastharvesttime=2012.08.24 08:50:02
m_lasttime in dataprovider.javaFri Aug 24 08:50:02 GM
lastharvesttime=2012.08.24 08:51:33
m_lasttime in dataprovider.javaFri Aug 24 08:51:33 GM
```

Ejecutar el Parser

```
C:\ Símbolo del sistema - java parser.Parse -n oai_dc
Finish Harvesting task:2012-08-28
D:\Amauta\build\classes>java parser.Parse -n oai_dc
ConnectionPool: JDBCDriver = org.gjt.mm.mysql.Driver
ConnectionPool: JDBCConnectionURL = jdbc:mysql://localhost:3306/amauta
ConnectionPool: ConnectionPoolSize = 5
ConnectionPool: ConnectionPoolMax = 100
ConnectionPool: ConnectionUseCount = 5
ConnectionPool: ConnectionTimeout = 20 seconds
ConnectionPool: (JDBC Property) characterEncoding = UTF-8
ConnectionPool: (JDBC Property) user = root
ConnectionPool: (JDBC Property) useUnicode = TRUE
ConnectionPool: (JDBC Property) password = pol
ConnectionPool: Registering org.gjt.mm.mysql.Driver
Start Parsing (only new records)
Parsing Archive: Univ. Nac. Mayor de San Marcos ..
*
***Record sdx:www.cybertesis.edu.pe:80:sisbib/documents/sisbib.2004.ramirez_pe-principal has not got parsed value ..
<?xml version="1.0" encoding="UTF-8" ?><record><header><identifier>sdx:www.cybertesis.edu.pe:80:sisbib/documents/sisbib.2004.ramirez_pe-principal</identifier><datestamp>2008-11-05T17:30:55Z</datestamp></header><metadata></metadata></record>
*
***Record sdx:www.cybertesis.edu.pe:80:sisbib/documents/sisbib.2000.barron_ph-principal has not got parsed value ..
```



ETD
PERU 2012
EL ELECTRONICO, TESIS, TESIS, TESIS, TESIS, TESIS

15th International Symposium on Electronic Theses and Dissertations

DC, tabla donde se almacena información recopilada

tabla se almacena la información necesaria para el cosechamiento

Tablas [22]

- cache
- dc
- dcnew
- dcsubject
- deleterecord
- pdidentify
- dpinfo
- dplist
- free
- harvestlog
- main
- prueba
- querylog
- scheduler
- spidentify
- subjectmapping
- subjectschema
- subjecttext

tabla que almacena los valores necesarios para cosechar



Bases de datos con 22 tablas



Analizando datos

RPTD

- 5608 registros *(03/09/2012)*
- 3602 tiene resumen
- 4412 Palabras clave
- Diferentes tipo de identificador (Handle, otros)

AMAUTA

- 18842 registros *(03/09/2012)*
- 12476 tienen resumen
- 12994 tienen palabras clave
- Diferentes tipo de identificador (Handle, otros)



Problemas comunes

- Si los metadatos tienen un carácter extraño (no compatible con UTF-8), la recolección de información no concluye exitosamente.
(retirar registro para continuar).
- El identificador de los objetos digitales no son homogéneos
UNMSM http://www.cybertesis.edu.pe/sisbib/2006/acevedo_fa/html/index-frames.html
UNI: http://cybertesis.uni.edu.pe/uni/2007/abad_dj/html/index-frames.html
Ateneo <http://hdl.handle.net/123456789/4290>
PUCP <http://tesis.pucp.edu.pe/repositorio/handle/123456789/1399>
- El proceso de recolección no es completamente automatizado
- Complejidad en la modificación en el código fuente del ARC (ejemplo implementar índices por Grado académico)



15th International Symposium on Electronic Theses and Dissertations

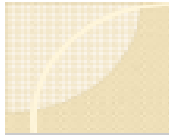


POLÍTICAS Y DIRECTRICES



Por que?

- Mayor visibilidad de la producción académica.
- Mayor impacto de las investigaciones de la institución.
- **Adopción de políticas institucionales para promover el acceso libre.**
- Protección de los derechos de autor
- Preservación de los contenidos a lo largo del tiempo.



*Ley 14/2011, de 1 de junio, de la Ciencia, la
Tecnología y la Innovación.*
JUAN CARLOS I REY DE ESPAÑA

Artículo 37. *Difusión en acceso abierto.*

1. Los agentes públicos del Sistema Español de Ciencia, Tecnología e Innovación impulsarán el desarrollo de repositorios, propios o compartidos, de acceso abierto a las publicaciones de su personal de investigación, y establecerán sistemas que permitan conectarlos con iniciativas similares de ámbito nacional e internacional.
2. El personal de investigación cuya actividad investigadora esté financiada mayoritariamente con fondos de los Presupuestos Generales del Estado hará pública una versión digital de la versión final de los contenidos que le hayan sido aceptados para publicación en publicaciones de investigación seriadas o periódicas, tan pronto como resulte posible, pero no más tarde de doce meses después de la fecha oficial de publicación.
3. La versión electrónica se hará pública en repositorios de acceso abierto reconocidos en el campo de conocimiento en el que se ha desarrollado la investigación, o en repositorios institucionales de acceso abierto.
4. La versión electrónica pública podrá ser empleada por las Administraciones Públicas en sus procesos de evaluación.
5. El Ministerio de Ciencia e Innovación facilitará el acceso centralizado a los repositorios, y su conexión con iniciativas similares nacionales e internacionales.

<http://www.boe.es/boe/dias/2011/06/02/pdfs/BOE-A-2011-9617.pdf>



15th International Symposium on Electronic Theses and Dissertations



Imprint

DINI – Deutsche Initiative für Netzwerkinformation e. V.

DINI-Geschäftsstelle
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen
Popeniek 14
D-37073 Göttingen
Tel.: +49 551 39-3857
Fax: +49 551 39-3856
E-Mail: gs@ dini.de
www.dini.de

2 Criteria

The DINI Certificate comprises eight criteria that are described in detail in this section. The criteria are:

Criterion 1 – *Visibility of the Service* (section 2.1)

Criterion 2 – *Policy* (section 2.2)

Criterion 3 – *Support of Authors and Publishers* (section 2.3)

Criterion 4 – *Legal Aspects* (section 2.4)

Criterion 5 – *Information Security* (section 2.5)

Criterion 6 – *Indexing and Interfaces* (section 2.6)

Criterion 7 – *Access Statistics* (section 2.7)

Criterion 8 – *Long-Term Availability* (section 2.8)



15th International Symposium on Electronic Theses and Dissertations



Minimum Requirements

M.1-1 The entire range of services must be available via a website.

⇒ This refers to a Document and Publication Service's main page from which both publication workflow and access to already published documents are possible.

M.1-2 The service's homepage must be referenced in a central location on the institution's homepage.

⇒ Potential users must be guided mostly intuitively from an institution's, a research facility's or a library's central website to the Document and Publication Service.

M.1-3 The service is registered and listed on the DINI website.

⇒ This DINI list can be found at <http://www.dini.de/wiss-publizieren/repository/>. This is also where a service can be registered.

M.1-4 The service is registered with the Directory of Open Access Repositories (OpenDOAR).

⇒ Find OpenDOAR at <http://www.opendoar.org/>. This is also where a service can be registered.

Recommendations



2.5.2 Document Handling

Minimum Requirements

- M.5-9 Documents uploaded into the repository will not be changed.
⇒ Changes on the content of published documents will be considered additional editions that do not overwrite or render obsolete earlier editions.
- M.5-10 Every document and every edition/version is assigned a Persistent Identifier (PI).
⇒ Available PI systems are e. g. URN and DOI.
- M.5-11 Persistent Identifiers are indicated on the service's web pages and in the exported metadata as primary identifiers in the form of an operable URL.
⇒ This requires a resolving service's URL to be added to the Persistent Identifier. As for the metadata export see also criterion 6 – *Indexing and Interfaces*, section 2.6.2, minimum requirement M.6-7.



15th International Symposium on Electronic Theses and Dissertations



HOME / NEWS / BENEFITS / PREPARING FOR AN AUDIT / STANDARDS / RESOURCES / MEMBERS

> ISO 16363:2012 now published on ISO site as International Standard



ISO 16363:2012 has been published by ISO as a full International Standard.



Recommendation for Space Data System Practices

AUDIT AND
CERTIFICATION OF
TRUSTWORTHY DIGITAL
REPOSITORIES

RECOMMENDED PRACTICE

CCSDS 652.0-M-1

MAGENTA BOOK
September 2011

ISO 16363 (Audit and certification of trustworthy digital repositories) has been available from the [CCSDS site](#) for some time, and has for several months been going through the final ISO processes, available as a standard under development.

We are please to announce that it is now available from the [ISO site](#) as a full international standard.



15th International Symposium on Electronic Theses and Dissertations



A	B	C	D	E	F	G
3. ORGANIZATIONAL INFRASTRUCTURE						
3.1 GOVERNANCE & ORGANIZATIONAL VIABILITY						
	Metric	Supporting Text	Examples of Documents the Repository can use to demonstrate it is Meeting this Requirement:	Brief description of evidence (add rows if necessary to list all relevant documents for a metric) Use short titles for documents. Provide detailed description of each document on the	Explanation of how the repository addresses this metric	
3.1.1	THE REPOSITORY SHALL HAVE A MISSION STATEMENT THAT REFLECTS A COMMITMENT TO THE PRESERVATION OF, LONG TERM RETENTION OF, MANAGEMENT OF, AND ACCESS TO DIGITAL	This is necessary in order to ensure commitment to preservation and access at the repository's highest administrative level.	Mission statement or charter of the repository or its parent organization that specifically addresses or implicitly calls for the preservation of information and/or other resources under its purview; a legal, statutory, or government regulatory mandate applicable to the repository that specifically addresses or implicitly requires the preservation of information and/or other			
3.1.2	THE REPOSITORY SHALL HAVE A PRESERVATION STRATEGIC PLAN THAT DEFINES THE APPROACH THE REPOSITORY WILL TAKE IN THE LONG-TERM SUPPORT	This is necessary in order to help the repository make administrative decisions, shape policies and allocate resources in order to successfully preserve its holdings.	Preservation Strategic Plan; meeting minutes; documentation of administrative decisions which have been made.			
3.1.2.1	<i>The repository shall have an appropriate, formal succession plan, contingency plans, and/or escrow arrangements in place in case the repository</i>	This is necessary in order to preserve the information content entrusted to the repository by handing it on to another custodian in the case that the repository ceases to	Written and credible succession and contingency plan(s); explicit and specific statement documenting the intent to ensure continuity of the repository, and the steps taken and to be taken to ensure continuity; escrow of critical code, software, and			



15th International Symposium on Electronic Theses and Dissertations

Muchas gracias