

ExtGASoai: OAI-PMH gateway per ExtGAS

luglio 2007

Giuseppe A. Romano
Istituto di Scienza e Tecnologie dell'Informazione,
e.mail giuseppe.romano@isti.cnr.it

Indice

0.Premessa	pag	4
1. ExtGAS OAI-PMH gateway	pag	5
2. La security di ExtGAS OAI-PMH gateway	pag	7
3. Il software ExtGAS OAI-PMH gateway	pag	8
4. Esempio di OAI System Table	pag	10
2. Bibliografia	pag	11

O.Premessa

ExtGASoi è il gateway OAI-PMH per il sistema **ExtGAS/IsisGAS**. **ExtGAS/IsisGAS** è composto dal sistema **IsisGAS** descritto in "Isis/GAS : un sistema multi piattaforma per l'accesso a basi di dati testuali via Internet - ISTI-B4-28, 2003" e **ExtGAS** che accede a sistemi catalografici memorizzati su database system diversi da CDS/ISIS; vedi : " ExtGAS un sistema multipiattaforma per l'accesso ai cataloghi di sistemi biblioteconomici eterogenei - ISTI-B4-15,2005 ". Naturalmente il gateway si applica alla parte **IsisGAS** del sistema **ExtGAS/IsisGAS**.

L'implementazione del gateway si è resa necessaria dato l'utilizzo del sistema **ExtGAS/IsisGAS** per la selezione e l'accesso alla Digital Library delle pubblicazioni degli Istituti del CNR partecipanti al progetto Metapub. Tramite il gateway è stato/è tuttora possibile esportare i metadati che descrivono le pubblicazioni verso altri servizi integrati a livello italiano e europeo.

Nella descrizione delle procedure di installazione e configurazione del gateway gli esempi fanno riferimento all'ambiente Unix, ma sono validi anche per ambiente Windows. Non esiste un ambiente preferenziale.

Si presuppone che il lettore abbia conoscenza di OAI-PMH Protocol version 2.0 del 2002-06-14 e del sistema CDS/ISIS, delle sue procedure di generazione, indicizzazione e ordinamento, e conosca, inoltre, il linguaggio di stampa/formattazione dei dati (vedi Bibliografia).

Nel seguito si parlerà indifferentemente di:

1. OAI System File
2. OAI System Table
3. Database System Table
4. Database System File
5. Configuration File

Intendendo il file di configurazione che definisce l'applicazione **ExtGASoi** verso un particolare database CDS/Isis: in sostanza viene definito un repository e le sue caratteristiche operative. Lo stesso software può gestire più di un gateway OAI/database per volta. Da ciò consegue che lo OAI System File definisce il database da utilizzare, le modalità di selezione e accesso ai dati e le modalità di restituzione secondo lo standard OAI-PMH Protocol version 2.0 del 2002-06-14.

1. ExtGAS OAI-PMH gateway

Il gateway ExtGAS OAI-PMH quando venga applicato a un database CDS/ISIS che contenga i metadati e i pointer ai documenti/oggetti o i documenti/oggetti stessi costituisce un **repository**. Mediante la OAI System Table è possibile configurare il repository in modo che si allinei ai requirements del repository stesso.

Il gateway sviluppato implementa i seguenti verbi del OAI-PMH Protocol version 2.0 del 2002-06-14

1. **Identify** che accede alle informazioni relative al repository di competenza.

Di default il repository gestito da gateway ha le seguenti caratteristiche:

```
<repositoryName>ExtGAS/ISIS Open Archive Initiative Repository </repositoryName>
<baseURL>http://leonardo.isti.cnr.it/oai</baseURL>
<protocolVersion>2.0</protocolVersion>
<adminEmail>giuseppe.romano@isti.cnr.it</adminEmail>
<earliestDatestamp>1977-08-15</earliestDatestamp>
<granularity>YYYY-MM-DD</granularity>
<deletedRecord>no</deletedRecord>
```

Fondamentalmente sono definite:

1. la data più bassa rispetto a tutti record che si necessita di esportare
 2. la più fine granularità supportata dal repository.
 3. Il modo con cui il sistema gestisce i record cancellati
2. **ListIdentifiers** per accedere al contenuto degli identificatori dei record e inviarli all'Harvester client. In fase di configurazione del gateway è possibile definire il massimo numero di identifier da inviare sia la prima che le successive volte per soddisfare la richiesta (il default è 100) che può selezionare un massimo numero di record (il default è 10000).

Sono accettati i seguenti parametri:

1. **metadataPrefix** che identifica il formato di restituzione dei record: è **obbligatorio**
 2. **set** che identifica, se esiste, un set o una gerarchia di set che permettono di selezionare un particolare sottoinsieme dei record
 3. **resumptionToken** che definisce il restart-download pointer: esso è esclusivo cioè non accetta alcun parametro aggiuntivo in quanto si riferisce a una selezione già di per se definita.
 4. **from** che identifica il from-date da cui partire a selezionare i record
 5. **until** che identifica il to-date fino a cui selezionare i record
3. **ListRecords** per accedere al contenuto dei record e inviarli all'Harvester client nel

metadataPrefix specificato. In fase di configurazione del gateway è possibile definire il massimo numero di record da inviare sia la prima che le successive volte per soddisfare la richiesta (il default è 100) che può selezionare un massimo numero di record (il default è 10000).

Sono accettati i seguenti parametri:

1. **metadataPrefix** che identifica il formato di restituzione dei record: è **obbligatorio**
 2. **set** che identifica, se esiste, un set o una gerarchia di set che permettono di selezionare un particolare sottoinsieme dei record
 3. **resumptionToken** che definisce restart-download pointer: esso è esclusivo cioè non accetta alcun parametro aggiuntivo in quanto di riferisce a una selezione già di per se definita.
 4. **from** che identifica il from-date da cui partire a selezionare i record
 5. **until** che identifica il to-date fino a cui selezionare i record
4. **GetRecord** per accedere selettivamente al contenuto del record e inviarlo all'Harvester client nel **metadataPrefix** specificato.

Sono accettati i seguenti parametri:

1. **identifier** che rappresenta l'identificatore del record: è **obbligatorio**
 2. **metadataPrefix** che identifica il formato di restituzione dei record: è **obbligatorio**
5. **ListSets** per accedere alla descrizione dei **Set** e **Subset** cui possono appartenere i record.

Di default possono essere definiti un massimo di 100 set, ma il valore può essere modificato in fase di configurazione del repository. Si tenga presente che il **resumptionToken** che definisce il restart-download pointer **non** è supportato per il verbo ListSet

6. **ListMetadataFormats** che descrive le modalità di restituzione dei metadati (formati di restituzione)

Sono accettati i seguenti parametri:

1. **identifier** che rappresenta l'identificatore del record: è **opzionale**

2. La security di ExtGAS OAI-PMH gateway

Nonostante la definizione del nome del protocollo stesso "Open Access Initiative" per motivi di sicurezza e riservatezza l'accesso al repository gestito dall'ExtGAS OAI-PMH gateway è controllato e di default non è aperto a tutti ma solo agli autorizzati.

Mediante la OAI System Table è possibile definire lo status generale della security che di default è impostata a **restricted**.

I livelli di sicurezza vengono associati all'indirizzo IP dell'Harvester client definendo liste di HostRestricted, HostAllowed e HostDenied.

I livelli di sicurezza sono:

1. **allowed**: l'Harvester client può eseguire tutti i verbi implementati, accedere alle liste dei record e, ove ne abbia ulteriormente diritto, scaricare dalla Digital Library il documento/oggetto.
2. **denied**: l'Harvester client pur accedendo al repository non può visualizzare né record identifier né record. Può eseguire i verbi ListSets, ListMetadataFormats e Identify
3. **restricted**: l'Harvester client pur accedendo al repository può visualizzare al massimo i primi MinObjectToBeDisplayed (default 5) identifier/record che soddisfano la richiesta e vengono successivamente ordinati in ordine crescente di data, ma non può eseguire il verbo GetRecord. Può eseguire invece, i verbi ListSets, ListMetadataFormats e Identify

Il livello di sicurezza **restricted** permette di rendere pubblico il gateway e le sue funzionalità garantendo una certa sicurezza negli accessi. Si tenga presente che per default viene tracciato un log degli accessi con i diritti associati agli indirizzi IP e le operazioni eseguite.

Genericamente potrebbe essere utile implementare un repository completamente controllato definendo la default security **denied** e aprendo solo a coloro che sono menzionati nella allowed list. Contemporaneamente si potrebbe definire un repository logico implementato sullo stesso database del repository operativo che abbia (mediante al definizione di set appropriati) una versione molto parziale delle informazioni.

3. Il software ExtGAS OAI-PMH gateway

Il software ExtGASoai è costituito da un insieme di moduli java che implementano i "verbi" OAI e da una API CDS/ISIS per accedere ai dati e produrre i risultati secondo lo standard OAI:

1. **ExtGASoai.java** è il driver principale del sistema
2. **ExtGASoai_v_Id.java** gestisce il verbo Identify
3. **ExtGASoai_v_LI.java** gestisce il verbo ListIdentifiers
4. **ExtGASoai_v_LS.java** gestisce il verbo ListSets
5. **ExtGASoai_v_LMF.java** gestisce il verbo ListMetadataFormats
6. **ExtGASoai_v_LR.java** gestisce il verbo ListRecords
7. **ExtGASoai_v_GR.java** gestisce il verbo GetRecord
8. **ExtGASoaiUtil.java** è un insieme di Utility generali
9. **ExtGASoaiIsisQuery.java** costruisce le query e esegue gli accessi a CDS/ISIS
10. **ExtGASoaiWriteResults.java** produce i risultati da restituire via protocollo OAI-PMH
11. **ExtGASoaiSysTable.java** descrive e definisce i parametri generali del sistema in modo che sia possibile la personalizzazione del Gateway OAI-PMH. Nel seguito verranno referenziati i parametri più importanti, mentre la lista completa e la discussione relativa verranno trattati nel manuale di installazione del Gateway
12. **AccessIsis.java** accede, mediante l'API CDS/ISIS, ai dati.
13. **<wwwisis>** che è la API CDS/ISIS per accedere al database. Esso corrisponde al programma eseguibile **www-isis**, distribuito free nella versione 3.0, per piattaforme unix-linux, unix-solaris, unix-ibmaix, unix-hp e windows, da www.unesco.org e prodotto da BIREME www.bireme.br

Nel complesso il software è costituito da:

1. I driver: da 1 a 7 nella lista precedente
2. I servizi: da 9 a 10 nella lista precedente
3. Le Utility: 8 nella lista precedente
4. Le definizioni: 11 nella lista precedente
5. L'API: 13 nella lista precedente

Il software richiede l'installazione di java e di un java http server per rendere operativo il gateway

Il test del gateway è stato certificato da <http://re.cs.uct.ac.za/> : Computer Science Dpt.
- Cape Town University.

Il software può essere scaricato alla seguente URL:

<http://leonardo.isti.cnr.it/metaopac/MetaopacDirectorySaver/solaris9/ExtGASoai.src.tgz>

4. Esempio di OAI System Table

Il seguente è un esempio di OAI System Table che implementa un repository che rende disponibile i dati dell'Istituto IENI (Sezione di Genova) che si appoggia sul database del progetto Metapub citato all'inizio.

```
# OAI-pmh gateway versus CDS/ISIS: oai.system.file
#TRACE on
MaxObjectsToBeDisplayed 10
MinObjectsToBeDisplayed 12
ServerURI http://leonardo.isti.cnr.it/oai_jenige
ExtGASoaiCiparFileName /export/home/metaopac/mpisa/oai_jenigeConf/cnr.txt
GeneralQuerySuffix AND ap/(904) and IENIGE/(900)
ExtGASoaiProgramName /export/home/metaopac/mpisa/BiremePrograms/bin/wwwisis
ExtGASoaiArchiveName cnr
AllowedHosts 146.48.80-87.0-255, 137.158.128.0-255, 81.189.125.130, 155.232.128.10
#DeniedHosts 146.48.82.119
#RestrictedHosts 146.48.82.119
ExtGASoaiDataRecordDelimiter ;
ExtGASoaiDateQualifier /(5)
ExtGASoaiIdentifierQualifier /(1)
ExtGASoaiSortPftFile @Sdate.pft
ExtGASoaiGizmoFile lat2html
ResumptionTokenDir /export/home/metaopac/mpisa/OAISAVE/
```

In questa installazione, l'eventuale host **restricted** potrà accedere ai primi 12 identifier/record selezionati ordinati in ordine crescente di data verificando il funzionamento del **resumptionToken** restart.

Il repository può essere acceduto alla URL http://leonardo.isti.cnr.it/oai_jenige e eventualmente la stessa URL può essere usata in test a <http://re.cs.uct.ac.za/>

La successiva OAI System Table definisce un repository operativo da cui il progetto Driver carica i dati

```
# OAI-pmh gateway versus CDS/ISIS: oai.system.file
#TRACE on
#ResumptionTokenTTL 1000
ExtGASoaiCiparFileName /export/home/metaopac/mpisa/oai_cnrpubdriverConf/cnr.txt
#GeneralQuerySuffix AND ap/(904) and IENIGE/(900)
ExtGASoaiProgramName /export/home/metaopac/mpisa/BiremePrograms/bin/wwwisis
ExtGASoaiArchiveName cnr
AllowedHosts 146.48.80-87.0-255, 129.70.12.42, 137.158.128.0-255, 155.232.250.0-255,
137.158.96.152, 81.189.125.130, 155.232.128.0-255
#DeniedHosts 146.48.82.119
ServerURI http://leonardo.isti.cnr.it/oai_cnrpubdriver
#MaxObjectsToBeDisplayed 100
MaxObjectsToBeDisplayed 10
ExtGASoaiLowerDate 19770815
ExtGASoaiDataRecordDelimiter ;
ExtGASoaiDateQualifier /(5)
ExtGASoaiIdentifierQualifier /(1)
ExtGASoaiSortPftFile @Sdate.pft
ExtGASoaiGizmoFile lat2html
SetDefinition IAMCTA+x_iamcta+cnr IAMCTA publications+AND ap/(904) and IAMCTA/(900);
```

```

SetDefinition IBFGE+x_ibfge+cnr IBFGE publications+AND ap/(904) and IBFGE/(900);
SetDefinition IBFPI+x_ibfge+cnr IBFPI publications+AND ap/(904) and IBFPI/(900);
SetDefinition IENIGE+x_ienige+cnr IENIGE publications+AND ap/(904) and IENIGE/(900);
SetDefinition IFC+x_ifc+cnr IFC publications+AND ap/(904) and IFC/(900);
SetDefinition IGGPI+x_igppi+cnr IGGPI publications+AND ap/(904) and IGGPI/(900);
SetDefinition IIT+x_iit+cnr IIT publications+AND ap/(904) and (IIT/(900) OR IAT/(900));
SetDefinition IMATIGE+x_imatige+cnr IMATIGE publications+AND ap/(904) and IMATIGE/(900);
SetDefinition ISEMGE+x_isemge+cnr ISEMGE publications+AND ap/(904) and ISEMGE/(900);
SetDefinition ISIBPD+x_isibpd+cnr ISIBPD publications+AND ap/(904) and ISIBPD/(900);
SetDefinition ISTI+x_isti+cnr ISTI publications+AND ap/(904) and ISTI/(900);
SetDefinition AP+x_ap+publications+AND ap/(904)
IdentifierPrefix Metapub:
ResumptionTokenDir /export/home/metaopac/mpisa/OAISAVE/
# the Identify data
Identify <Identify>
Identify <repositoryName>ExtGAS/ISIS Open Archive Initiative Repository </repositoryName>
Identify <baseURL>http://leonardo.isti.cnr.it/oai_cnrpubdriver</baseURL>
Identify <protocolVersion>2.0</protocolVersion>
Identify <adminEmail>giuseppe.romano@isti.cnr.it</adminEmail>
Identify <earliestDatestamp>1977-08-15</earliestDatestamp>
Identify <granularity>YYYY-MM-DD</granularity>
Identify <deletedRecord>no</deletedRecord>
Identify <compression>deflate</compression>
Identify <description>
Identify <oai-identifier xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
Identify <scheme>oai</scheme>
Identify <repositoryIdentifier>leonardo.isti.cnr.it</repositoryIdentifier>
Identify <delimiter>:</delimiter>
Identify <sampleIdentifier>oai:leonardo.isti.cnr.it:Metapub:cnr.isti/cnr.isti/2003-B4-28</sampleIdentifier>
Identify </oai-identifier>
Identify </description>
Identify </Identify>
# ends the Identify data
# the MetadataFormats definition
ListMetadataFormats <ListMetadataFormats>
ListMetadataFormats <metadataFormat>
ListMetadataFormats <metadataPrefix>oai_dc</metadataPrefix>
ListMetadataFormats <schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd </schema>
ListMetadataFormats <metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/
</metadataNamespace>
ListMetadataFormats </metadataFormat>
ListMetadataFormats <metadataFormat>
ListMetadataFormats <metadataPrefix>oai_rfc1807</metadataPrefix>
ListMetadataFormats <schema>http://www.openarchives.org/OAI/1.1/rfc1807.xsd</schema>
ListMetadataFormats
<metadataNamespace>http://www.w3.org/2001/XMLSchemainstance</metadataNamespace>
ListMetadataFormats </metadataFormat>
ListMetadataFormats </ListMetadataFormats>
# ends the MetadataFormats definition

```

Da notare le definizioni dei **set** come metodo per costruire un repository fisico costituito da più repository logici indipendenti.

Per la comprensione dei parametri di configurazione si invia alla lettura del " ExtGAS OIA-PMH gateway: manuale di installazione e uso" : vedi bibliografia.

Bibliografia

1. Giuseppe A. Romano - Isis/GAS : un sistema multi piattaforma per l'accesso a basi di dati testuali via Internet - ISTI-B4-28, 2003 - URL:
<http://leonardo.isti.cnr.it/metaopac/IsisGas.pdf>
2. Giuseppe A. Romano - Isis/GAS : manuale di installazione e d'uso - ISTI-B4-31, 2003 - URL: <http://leonardo.isti.cnr.it/metaopac/IsisGasInstall.pdf>
3. Giuseppe A. Romano - Il MetaOpac Pisano: l'organizzazione logica e il mapping fisico - ISTI-B4-02, 2003 - URL:
<http://leonardo.isti.cnr.it/metaopac/MetaOpacLogic.pdf>
4. Giuseppe A. Romano - Il MetaOpac Pisano: descrizione tecnica e procedure di recovery - ISTI-B4-30, 2003 - URL:
<http://leonardo.isti.cnr.it/metaopac/MetaopacPisano.pdf>
5. Giuseppe A. Romano - Il metaopac: controllo e recovery del sistema - ISTI-B4-07, 2004 - URL: http://leonardo.isti.cnr.it/metaopac/IsisGAS_BriefMan.pdf
6. Biagioni S., Giannini S., Giuseppe A. Romano - Il Metaopac del CNR di Pisa e la sua collocazione nel Metaopac pisano - ISTI-B4-14, 2004 - URL:
<http://leonardo.isti.cnr.it/metaopac/OpacCNR.pdf>
7. Pierlugi Niccolai, Enrico Venturini - TECA: Applicazione di CDS/ISIS per i dati bibliografici - Titivillus , Firenze 1993
8. Stefania Biagioni, Carlo Carlesi, Daniela Musa - Sistema Atlas: catalogazione derivata da registrazioni in formato UNIMARC - IEI-B4-30 : Istituto di Elaborazione della Informazione, settembre 1994
9. CDS/ISIS - software, tools, descriptions - URL:
http://portal.unesco.org/ci/ev.php?URL_ID=2071&URL_DO=DO_TOPIC&URL_SECTION=201&reload=1082372136&PHPSESSID=7462f4450cf4d320877976df1ae94292
10. BIREME - CISIS reference manual - URL:
<http://productos.bvsalud.org/html/en/home.html>
11. BIREME - wwwisis version 3.0 - URL:
<http://productos.bvsalud.org/html/en/home.html>
12. LIBERO - software - URL: <http://www.infologic.it/ita/libero.htm>
13. Antonio Scolari - Unimarc, Roma, AIB, 2000 Enciclopedia Tascabile ETIS
14. Giuseppe A. Romano - Il MetaOpacPisano : organizzazione, controllo e recovery del sistema per l'anno 2005 - ISTI-B4-17,2005 URL:
http://leonardo.isti.cnr.it/metaopac/MOP_BriefMan.pdf
15. Giuseppe A. Romano - ExtGAS un sistema multipiattaforma per l'accesso ai cataloghi di sistemi biblioteconomici eterogenei - ISTI-B4-15,2005 - URL:
<http://leonardo.isti.cnr.it/metaopac/ExtGas.pdf>

16. BIREME - Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde - WWWISIS: a world-wide web server for ISIS-databases Version 3.0, URL <http://leonardo.isti.cnr.it/metaopac/wwwisis.pdf>
17. Giuseppe A. Romano - ExtGAS manuale di installazione e d'uso - ISTI-B4-16,2005 - URL: <http://leonardo.isti.cnr.it/metaopac/ExtGasInstall.pdf>
18. Giuseppe A. Romano - ExtGas/IsisGAS revisione 2.0: manuale di installazione e d'uso - ISTI-B4-16, 2006 URL: <http://leonardo.isti.cnr.it/metaopac/ExtGASOmoDbInstall.pdf>
19. Giuseppe A. Romano - ExtGASoai - OAI-PMH gateway per ExtGAS: organizzazione logica del software - ISTI-B4-13 , 2007 URL: http://leonardo.isti.cnr.it/metaopac/ExtGAS_OAI_logic.pdf
20. Giuseppe A. Romano - ExtGASoai - OAI-PMH gateway per ExtGAS: manuale di installazione e d'uso (Draft version) - ISTI-B4-12 , 2007 URL: http://leonardo.isti.cnr.it/metaopac/ExtGAS_OAI_install.pdf