

# Lo Science Gateway del progetto agINFRA per l'accesso a una data infrastructure per le Scienze Agrarie



Riccardo Bruno<sup>1</sup>, Giovanni Allegri<sup>2</sup>, Giuseppe Andronico<sup>1</sup>, Roberto Barbera<sup>1,3</sup>, Federico Bitelli<sup>1,4</sup>, Antonio Budano<sup>1</sup>, Antonio Calanducci<sup>1</sup>, Edoardo A. C. Costantini<sup>5</sup>, Marco Fargetta<sup>1</sup>, Adrea Fornaia<sup>6</sup>, Giovanni L'Abate<sup>5</sup>, Salvatore Monforte<sup>1</sup>, Antonio Puliafito<sup>7</sup>, Rita Ricceri<sup>1</sup>, Federico Ruggieri<sup>1</sup>, Davide Saitta<sup>6</sup>, Massimo Villari<sup>7</sup>

<sup>1</sup>INFN, <sup>2</sup>GIS3W s.a.s., <sup>3</sup>Dipartimento di Fisica e Astronomia dell'Università di Catania, <sup>4</sup>Dipartimento di Fisica dell'Università di Roma Tre, <sup>5</sup>Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Centro di ricerca per l'agrobiologia e la pedologia (CRA-APB), <sup>6</sup>Consortium GARR, <sup>7</sup>Facoltà di Ingegneria, Università di Messina

**Abstract.** agINFRA è un progetto che ha lo scopo di sviluppare strumenti per supportare la comunità delle scienze agrarie basati sull'uso di infrastrutture dati distribuite a livello europeo e intercontinentale, insistenti sulle reti nazionali ed internazionali della ricerca.

Quest'articolo descrive l'architettura dello *Science Gateway* che è stato sviluppato per agINFRA basandosi sul Catania *Science Gateway Framework* e illustra le sue caratteristiche di ambiente integrato per l'accesso e la gestione dei dati, offerte mediante l'adozione dei paradigmi sia di Grid computing che di cloud computing. In particolare vengono descritti degli standard adottati, quali SAGA e SAML, e la loro importanza per la sostenibilità dell'infrastruttura e dei suoi servizi.

A titolo di esempio, saranno mostrate alcune delle applicazioni disponibili agli utenti attraverso lo Science Gateway, tra cui l'*Italian Soil Information System* (ISIS).

## 1. Introduzione

agINFRA [1] è un progetto co-finanziato dalla Commissione Europea nell'ambito del Settimo Programma Quadro che ha lo scopo di sviluppare strumenti per supportare la comunità delle scienze agrarie, basati sull'uso di infrastrutture dati distribuite a livello europeo e intercontinentale, insistenti sulle reti nazionali ed internazionali della ricerca. Scopo di agINFRA è progettare e sviluppare una grande infrastruttura digitale per dati scientifici relativi all'agricoltura, su cui creare e offrire servizi dedicati alla comunità scientifica di riferimento. Tali servizi dovranno essere fruibili in maniera tale da non violare le regole di sicurezza imposte dalle infrastrutture di tipo Grid e Cloud coinvolte.

Nell'ambito delle scienze agrarie esiste una crescente quantità di argomenti multidisciplinari, i cui soggetti possono variare dallo studio delle piante all'orticoltura e all'ingegneria agraria, dall'economia agraria agli studi di tipo ambientale. Un numero sempre crescente di studi si sta focalizzando sulle interconnessioni tra questi temi, come ad esempio il collegamento tra cambiamento climatico e produzione o perdita di biodiversità, oppure la pressione esercitata da una specie sull'altra. Questi studi richiedono l'accesso trasparente e ubiquo a enormi volumi di dati da parte di una comunità globale di ricercatori.

Per soddisfare questa esigenza, l'Istituto Nazionale di Fisica Nucleare ha realizzato, nell'ambito di agINFRA, uno *Science Gateway* dedicato [2], basato sul Catania *Science Gateway Framework* [3,4], che permette ai membri della comunità di riferimento di accedere ad applicazio-

ni, i cui soggetti possono variare dallo studio delle piante all'orticoltura e all'ingegneria agraria, dall'economia agraria agli studi di tipo ambientale. Un numero sempre crescente di studi si sta focalizzando sulle interconnessioni tra questi temi, come ad esempio il collegamento tra cambiamento climatico e produzione o perdita di biodiversità, oppure la pressione esercitata da una specie sull'altra. Questi studi richiedono l'accesso trasparente e ubiquo a enormi volumi di dati da parte di una comunità globale di ricercatori.

ni e servizi che vengono eseguiti trasparentemente su infrastrutture sia di tipo Grid che di tipo Cloud.

La descrizione dello Science Gateway di agINFRA è data nel paragrafo seguente, insieme alla presentazione di alcuni delle applicazioni e servizi disponibili al suo interno, mentre nel paragrafo 3 sono riportate le conclusioni.

## 2. L'agINFRA Science Gateway

Lo Science Gateway di agINFRA è un portale web che offre allo stesso tempo la possibilità di accedere ad applicazioni e servizi che girano su ambienti Grid o anche Cloud. Il portale consente di autenticare gli utenti tramite le credenziali possedute all'interno delle Federazioni d'identità [5,6] alle quali appartengono e gestisce in modo separato l'autenticazione dall'autorizzazione, in modo da offrire a ogni utente la possibilità di accedere ai diversi servizi offerti dal portale in maniera coerente e rispettosa dei suoi ruoli all'interno della comunità di riferimento e, quindi, dello Science Gateway. La figura 1 mostra l'architettura dell'agINFRA Science Gateway la quale si compone di 3 componenti principali.

Il modulo AAI si occupa dell'autenticazione e dell'autorizzazione degli utenti. La prima si basa sull'uso delle cosiddette identità federa-

te. Ciò è reso possibile dall'adozione dello standard SAML 2.0 [7] e della sua implementazione Shibboleth [8]. Al fine di massimizzare il numero di potenziali utenti, lo Science Gateway di agINFRA è stato abilitato quale *Service Provider* della federazione IDEM [6] e, attraverso questa, dell'inter-federazione internazionale eduGAIN [9]. Inoltre, coloro che non appartengono a nessuna federazione possono utilizzare la federazione *catch-all* GrIDP [10] e l'IDP Open [11] che sono co-gestiti dal GARR e dalla Sezione di Catania dell'INFN.

La parte del modulo AAI relativa all'autorizzazione è invece gestita tramite un server LDAP, il quale registra per ciascun utente i ruoli ricoperti all'interno delle varie organizzazioni facenti parte della comunità e le relative autorizzazioni che questi deve avere all'interno dello Science Gateway. Per soddisfare un preciso requisito di sicurezza, la registrazione delle autorizzazioni è gestita manualmente, garantendo o negando in modo preciso l'accesso alle risorse offerte dal portale.

Il secondo elemento dell'architettura dello Science Gateway di agINFRA è l'insieme delle diverse interfacce utente offerte dal portale per le varie applicazioni e servizi che sono integrati in esso. Queste sono implementate attraverso

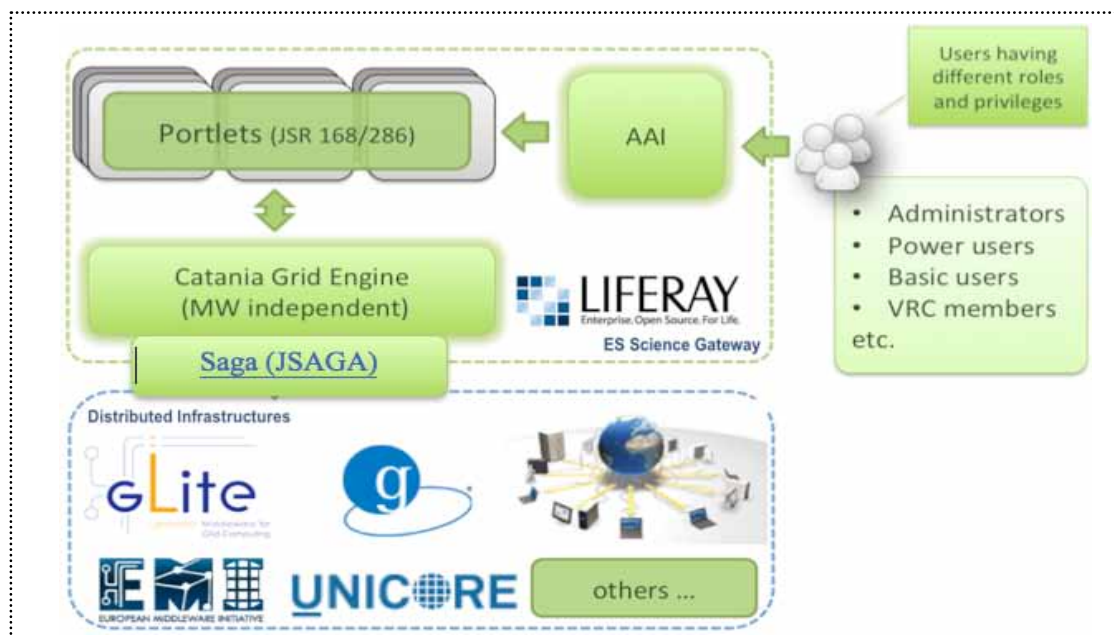


Fig 1 - Architettura dello Science Gateway di agINFRA

so *portlet* conformi allo standard JSR286 [12] e rese disponibili all'utente attraverso il *portlet framework Liferay* [13]. Le *portlet* fanno da tramite tra gli utenti del portale e le diverse infrastrutture distribuite attraverso il terzo e ultimo modulo; il Catania *GridEngine* [3]. Quest'ultimo elemento si occupa dell'accesso fisico alle infrastrutture distribuite, attraverso l'adozione dello standard SAGA [14] e della sua implementazione in Java chiamata JSAGA

[15], e del tracciamento delle attività degli utenti, in conformità alle politiche che regolano l'accesso all'*European Grid Infrastructure* attraverso portali [16,17]. JSAGA, garantisce la possibilità di accedere a infrastrutture Grid sulle quali insistono middleware differenti, mediante l'utilizzo di librerie specifiche chiamate *adaptors*, che nascondono le specificità dei vari *middleware* allo sviluppatore di *portlet*. Sono queste ultime librerie che gestiscono l'accesso fisico alle risorse per una specifica infrastruttura, interfacciandosi poi con un unico insieme di chiamate API di alto livello.

## 2.1 Applicazioni e servizi

Lo Science Gateway di agINFRA, la cui pagina d'accesso è riportata in figura 2, raggruppa l'accesso alle diverse applicazioni e servizi attraverso le voci di menu: "Applications" e "Cloud Services".

Sotto la voce "Applications" sono raggruppate le applicazioni installate sull'infrastruttura Grid del progetto, che accedono a risorse Grid utilizzabili attraverso il middleware EMI-gLite [18] e sono brevemente descritte di seguito:



Fig 2 - Home page dello Science Gateway di agINFRA

1. AGRIS AP XML2RDF: converte file dal formato AGRIS AP XML sviluppato dalla FAO nel formato RDF, standard del web semantico, lanciando una serie di esecuzioni parallele su Grid.
2. AGROVOC *Tagging*: elabora i file di output provenienti da un servizio di *web crawling* e restituisce in output un file di triple in formato RDF.
3. *R Statistical Analysis*: R è uno strumento per l'analisi statistica di dati e la loro visualizzazione grafica.
4. *RiceInfo*: è un servizio di ricerca sviluppata dall'*Indian Statistical Institute* che offre risultati riguardanti tutte le tipologie di parassiti del riso quali: virus, vermi, insetti, ecc. Queste informazioni sono memorizzate in una *knowledge base* che contiene anche altre tipologie di dati riguardanti i difetti di nutrizione, la tossicità, i meccanismi di controllo per gli agenti contaminanti. I risultati delle ricerche sono riportati come insiemi di triple in formato N3. Il servizio integra anche risultati di ricerche semantiche effettuate su

DBpedia e GoogleScholar.

5. Weka: è una suite di algoritmi di machine learning che possono essere applicati a problemi di data mining in scienze agrarie. Gli algoritmi di Weka sono sviluppati in Java e questa applicazione permette di eseguire diversi algoritmi su un singolo file di input, distribuendo l'operazione su un'infrastruttura Grid.

Oltre alle applicazioni appena esaminate, lo Science Gateway di agINFRA offre al momento altre due applicazioni di diversa tipologia. Queste sono trattate nel successivo sotto-paragrafo.

### 2.2.1 Italian Soil Information System

Il servizio *Italian Soil Information System* (ISIS), la cui interfaccia è mostrata in figura 3, utilizza lo standard WebGIS per offrire mappe del territorio sulle quali presentare dati relativi alle varie tipologie dei suoli delle regioni italiane, correlandoli poi con i dati del suolo del territorio europeo, rispettivamente alle tipologie di suolo unitarie "*Soil Typological Units*" (STUs), a livello nazionale, e di sotto-sistemi a livello regionale.

ISIS è stato sviluppato da ricercatori Centro di Ricerca per l'Agrobiologia e la Pedologia del Consiglio per la Ricerca e la Sperimentazione in Agricoltura (CRA) ed è mantenuto presso la Sezione di Catania dell'INFN. Lo Science Gateway

di agINFRA consente di accedere al servizio remoto in maniera del tutto trasparente all'utente nelle due modalità descritte di seguito.

### 2.2.2 Soil Map Annotator

Questa applicazione è in stretta relazione all'applicazione ISIS in quanto gestisce delle informazioni di tipo metadati sulle mappe del suolo utilizzate dall'applicazione ISIS. L'applicazione inoltre, permette di aggiungere delle annotazioni definite dall'utente sulle mappe e successivamente di eseguire ricerche non solo sui metadati associati alle mappe ma anche sulle annotazioni definite dagli utenti. L'utilizzo delle annotazioni è una forma di collaborazione sempre più in uso nell'ambito di progetti che coinvolgono numerosi partner o più in generale organizzazioni virtuali.

### 2.3 Cloud Services

Oltre a quella del *Grid computing*, il progetto agINFRA utilizza tecnologie di tipo Cloud, per tutti quei servizi che devono essere disponibili "24x7" e non possono quindi essere associati risorse non dedicate. L'applicazione ISIS è un esempio di servizio che gira su una Cloud il cui accesso da parte degli utenti è solamente possibile tramite lo Science Gateway. Inoltre, come mostrato in figura 4, l'interfaccia web offerta dal portale permette due diverse modalità di accesso a seconda delle autorizzazioni dell'utente. La prima permette un accesso di tipo pubblico e le funzionalità offerte dal servizio ISIS sono limitate ad un insieme minimo di opzioni. La seconda richiede che l'utente abbia eseguito il sign-in e offre ulteriori fun-



Fig 3 - Interfaccia WebGIS del servizio ISIS

ta dal portale permette due diverse modalità di accesso a seconda delle autorizzazioni dell'utente. La prima permette un accesso di tipo pubblico e le funzionalità offerte dal servizio ISIS sono limitate ad un insieme minimo di opzioni. La seconda richiede che l'utente abbia eseguito il sign-in e offre ulteriori fun-

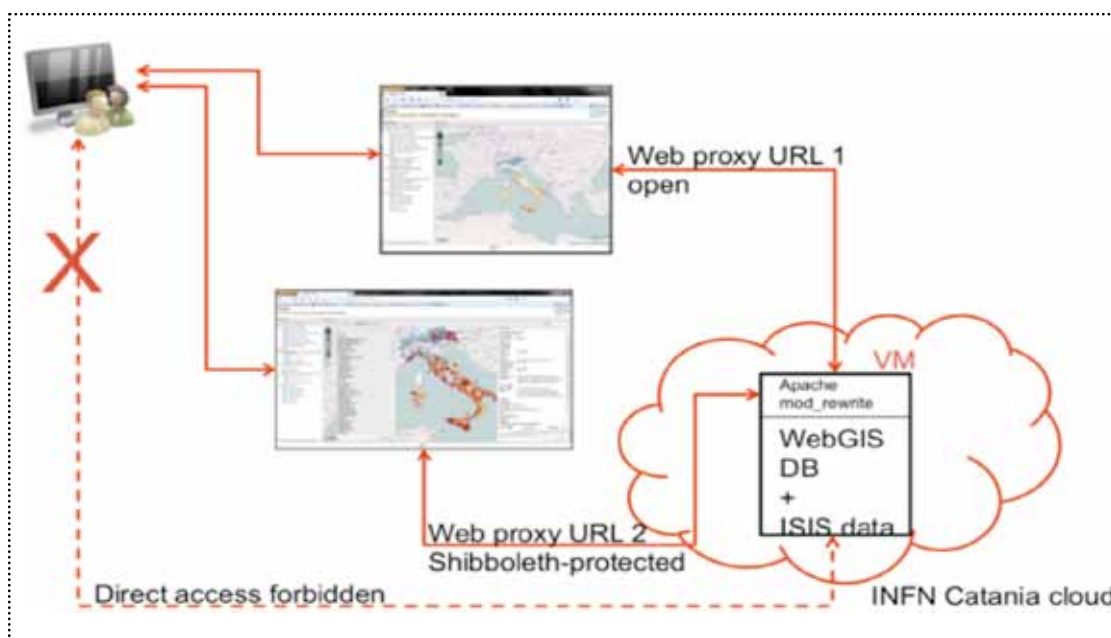


Fig 4 - Integrazione del servizio cloud ISIS nello Science Gateway di agINFRA

zionalità, come ad esempio la possibilità di eseguire interrogazioni al geo-database per ricercare informazioni sul suolo dei siti italiani in esso memorizzati.

Il progetto agINFRA utilizza come tecnologia Cloud il middleware CLEVER [19], che verrà brevemente descritto di seguito.

### 2.3.1 CLEVER

CLEVER è un innovativo middleware cloud, progettato e realizzato da ricercatori della Facoltà di Ingegneria dell'Università di Messina, che prevede la gestione di dispositivi virtuali come astrazione delle risorse fisiche. Questo facilita la gestione dei servizi cloud di tipo privato o ibri-

do. Tramite un'interfaccia grafica è possibile interagire con differenti infrastrutture di computer interconnessi via rete.

CLEVER offre la possibilità di gestire intere infrastrutture di tipo virtuale (IaaS). Il middleware è basato su un'architettura cluster di tipo distribuito, dove ogni cluster può appartenere a due diversi livelli gerarchici, come mostrato in figura 5.

Ciascun nodo CLEVER contiene un modulo di gestione host chiamato *Host manager* (HM) ed inoltre può anche includere un modulo di gestione cluster chiamato *Cluster Manager* (CM). Il Cluster manager è responsabile del trattamen-

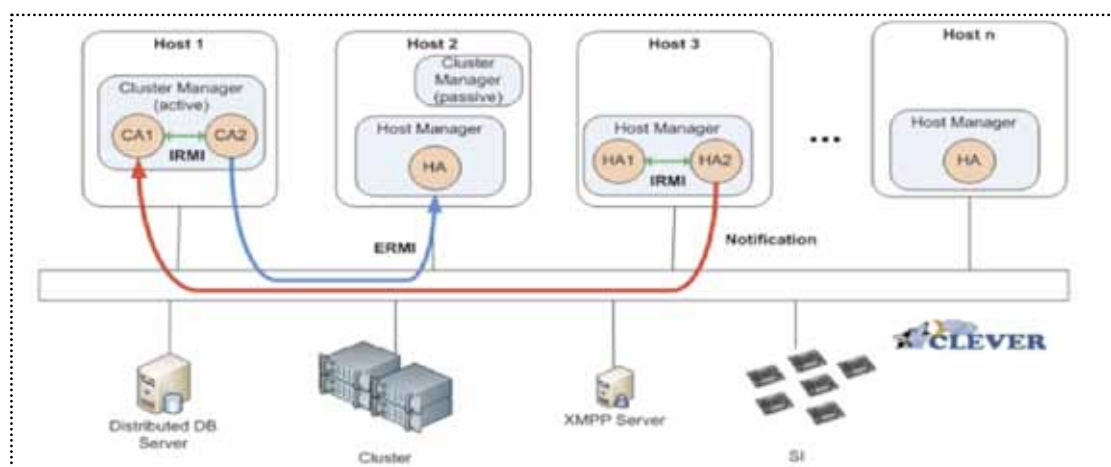


Fig 5 - Architettura del middleware CLEVER

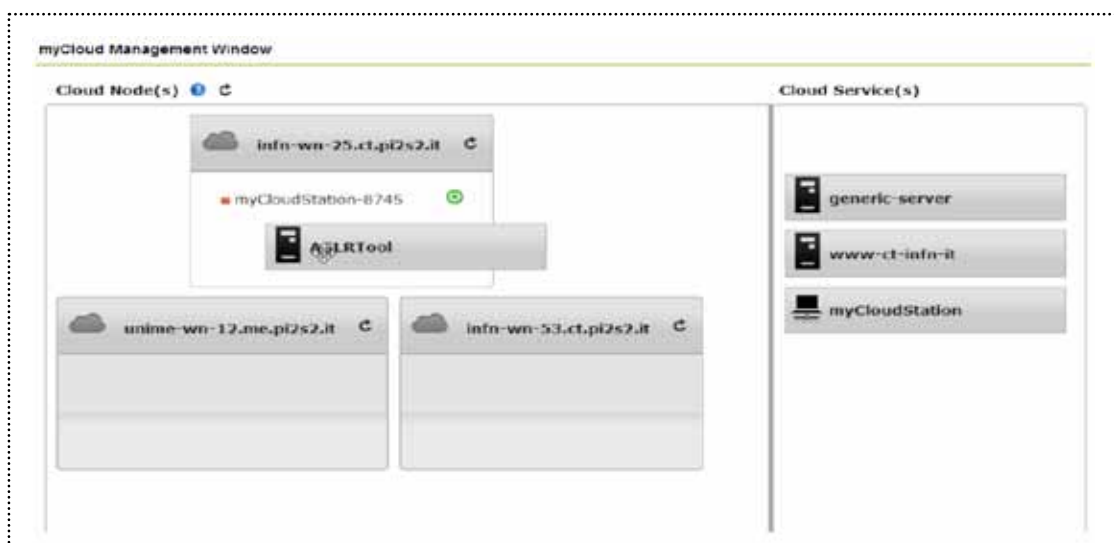


Fig 6 - Vista della portlet MyCloud

to e dell'analisi del flusso dei dati mentre l'HM, a più basso livello, agisce come agente remoto per il CM. In questo modo nel Cluster si avrà almeno un CM attivo di alto livello e diversi HM di basso livello dipendenti da esso. I nodi CLEVER sono gestiti, all'interno dello Science Gateway di agINFRA, dalla portlet MyCloud, accessibile solo agli utenti del portale identificati come Cloud Manager sul server LDAP.

Come mostrato in figura 6, questa interfaccia permette di installare servizi diversi sui nodi della cloud con semplici azioni di trascinamento.

Una volta installato il servizio richiesto su uno dei nodi di CLEVER, questo sarà reso dinamicamente fruibile da parte del portale ai suoi utenti, esattamente come avviene per l'applicazione ISIS descritta sopra.

### 3. Conclusioni

Lo Science Gateway di agINFRA offre al progetto diverse funzionalità, che spaziano dall'accesso a file e ai relativi metadati all'esecuzione di applicazioni che richiedono numerose risorse di computazione su infrastrutture di calcolo distribuite geograficamente. Vi è pure l'opportunità di gestire in modo semplice e dinamico, servizi ospitati da nodi cloud e renderli poi fruibili attraverso l'interfaccia web del portale, riuscendo a separare accessi di tipo pubblico da quel-

li ristretti.

Il portale è stato sviluppato con i più moderni standard riguardanti l'autenticazione e autorizzazione, l'esecuzione di applicazioni su infrastrutture diverse, che vengono così rese interoperabili tra loro, e il trattamento di dati e metadati su infrastrutture distribuite. L'adozione di standard è un punto cruciale per la sostenibilità tecnologica del portale e di conseguenza per i servizi offerti dal progetto agINFRA. La sostenibilità tecnologica avrà sicuramente ripercussioni positive su quella economica.

### Riferimenti Bibliografici

- [1] [www.aginfra.eu](http://www.aginfra.eu)
- [2] <http://aginfra-sg.ct.infm.it>
- [3] V. Ardizzone et al. J. Grid Computing (2012) 10:689-707, DOI 10.1007/s10723-012-9242-3
- [4] [www.catania-science-gateways.it](http://www.catania-science-gateways.it)
- [5] Per maggiori informazioni sulle Federazioni d'Identità esistenti nel mondo della ricerca scientifica, si visiti il sito web <https://refeds.org>
- [6] Per maggiori informazioni sulla Federazione d'Identità italiana, si visiti il sito web [www.idem.garr.it](http://www.idem.garr.it)
- [7] <http://saml.xml.org>
- [8] <http://shibboleth.net>

- [9] [www.edugain.org](http://www.edugain.org)
- [10] <http://gridp.garr.it>
- [11] <http://idpopen.garr.it>
- [12] [www.jcp.org/en/jsr/detail?id=286](http://www.jcp.org/en/jsr/detail?id=286)
- [13] [www.liferay.com](http://www.liferay.com)
- [14] [www.gridforum.org/documents/GFD.90.pdf](http://www.gridforum.org/documents/GFD.90.pdf)
- [15] <http://grid.in2p3.fr/jsaga>
- [16] <https://documents.egi.eu/public/ShowDocument?docid=80>
- [17] <https://documents.egi.eu/public/ShowDocument?docid=81>
- [18] [www.eu-emi.eu](http://www.eu-emi.eu)
- [19] <http://clever.unime.it>



**Riccardo Bruno**

[riccardo.bruno@ct.infn.it](mailto:riccardo.bruno@ct.infn.it)

Laureato in Scienze dell'Informazione presso l'Università degli Studi di Catania, ha lavorato presso una società operante nel settore della telefonia mobile.

Dal 2006 svolge attività di ricerca presso l'INFN Sez. di Catania in relazione ad ambienti di calcolo e storage distribuito: Grid/Cloud.

### **Ringraziamenti**

Si ringraziano tutti i collaboratori dell'INFN sezione di Catania che hanno fatto del Catania Science Gateway Framework un prodotto affidabile al servizio delle comunità scientifiche italiane e internazionali nell'ambito di diversi progetti europei. Si ringrazia inoltre tutto lo staff tecnico e amministrativo del progetto agINFRA nonché il team del Consiglio delle Ricerche e Sperimentazioni in Agricoltura (CRA) per il prezioso contributo offerto durante l'integrazione dell'applicazione ISIS.