D4Science Catalogue Service

Brief Overview

Mangiacrapa Francesco
(francesco.mangiacrapa@isti.cnr.it)

Istituto di Scienza e Tecnologie dell'Informazione (ISTI) "A. Faedo"
Consiglio Nazionale delle Ricerche, via G. Moruzzi, 1 – 56124, Pisa (IT).

Abstract. Questa nota tecnica documenta la piattaforma software open source "D4Science Catalogue Service" progettata e sviluppata da F. Mangiacrapa e ne dimostra l'uso da parte della comunità scientifica.

1. Introduzione

Un catalogo è un servizio chiave in molti contesti e scenari applicativi. In particolare è un servizio che permette di pubblicare e ricercare "oggetti informativi" caratterizzati da informazioni descrittive di vario tipo (metadati) e oggetti digitali che rappresentano una materializzazione dell'oggetto informativo o di una sua parte. Varie comunità scientifiche hanno necessità sempre più pressanti di creare e gestire cataloghi di dati, servizi, metodi/algoritmi, software, mappe e più in generale di oggetti informativi di interesse per il loro contesto applicativo caratterizzandoli con specifici metadati.

L'infrastruttura D4Science[1] offre servizi per l'accesso e l'analisi di dati di "natura" diversa, appartenenti a diversi domini, inclusi dati biologici ed ecologici, dati geospaziali, dati statistici e dati semi-strutturati provenienti da molteplici fonti di dati autorevoli e sistemi di informazione. Questi servizi possono essere sfruttati sia tramite web GUI che protocolli basati sul web per l'accesso programmatico, ad es. OAI-PMH, CSW, WMS, WFS, SDMX. Questa offerta è in grado di integrare le specifiche richieste provenienti dalle diverse community e/o le loro specifiche applicazioni.

I Cataloghi di D4Science contengono una vasta gamma di risorse derivanti da diverse attività, progetti e comunità tra cui BlueBRIDGE (www.bluebridge-vres.eu/), i-Marine (www.i-marine.eu), SoBigData.eu (www. sobigdata.eu) e FAO (www.fao.org). Tutti i prodotti sono accompagnati da ricche descrizioni che catturano attributi generali, ad es. titolo e creatore / i, nonché politiche di utilizzo e licenze.

[1] Infrastruttura di ricerca D4Science: *https://www.d4science.org/*

Lo scopo del servizio Catalogo di D4Science è di favorire e supportare la pubblicazione e catalogazione di “risultati scientifici” prodotti dalle diverse comunità di ricercatori e renderli accessibili a chiunque promuovendo le pratiche della Scienza Aperta (Open Science).

2. Background

Operazione preliminare alla progettazione e sviluppo del catalogo è stata quella di analizzare e confrontare soluzioni esistenti così da poterne capire le potenzialità rispetto alle necessità riscontrate e agli scenari applicativi identificati. Tra le soluzioni esistenti quella che è apparsa più vicina alle esigenze da soddisfare è stata la piattaforma CKAN[2] (ckan.org), largamente diffusa nell’ambito di iniziative per la realizzazione di portali per l’accesso ai dati. Tra le caratteristiche più interessanti la flessibilità e l’estensibilità della piattaforma grazie alla quale sono stati sviluppati una serie di “extensions” pronte all’uso.
Il catalogo di D4science è stato costruito utilizzando (ed estendendo) la piattaforma CKAN.
CKAN è un potente Data Management System (DMS) che rende i dati accessibili, fornendo strumenti per semplificare la pubblicazione, la condivisione, la ricerca e l'utilizzo dei dati. CKAN è un software open source, con una comunità attiva di collaboratori che sviluppano e mantengono la sua tecnologia di base ed altre comunità che sviluppano e mantengono le sue estensioni. La sua diffusione, la semplicità del browsing e della search fornite tramite la GUI e la sua licenza open-source ci hanno portato alla sua “adozione” nel contesto dell’infrastruttura D4Science.
Il modello di CKAN è costituito dalle entità e loro relazioni riportate nella *Figura 1*.

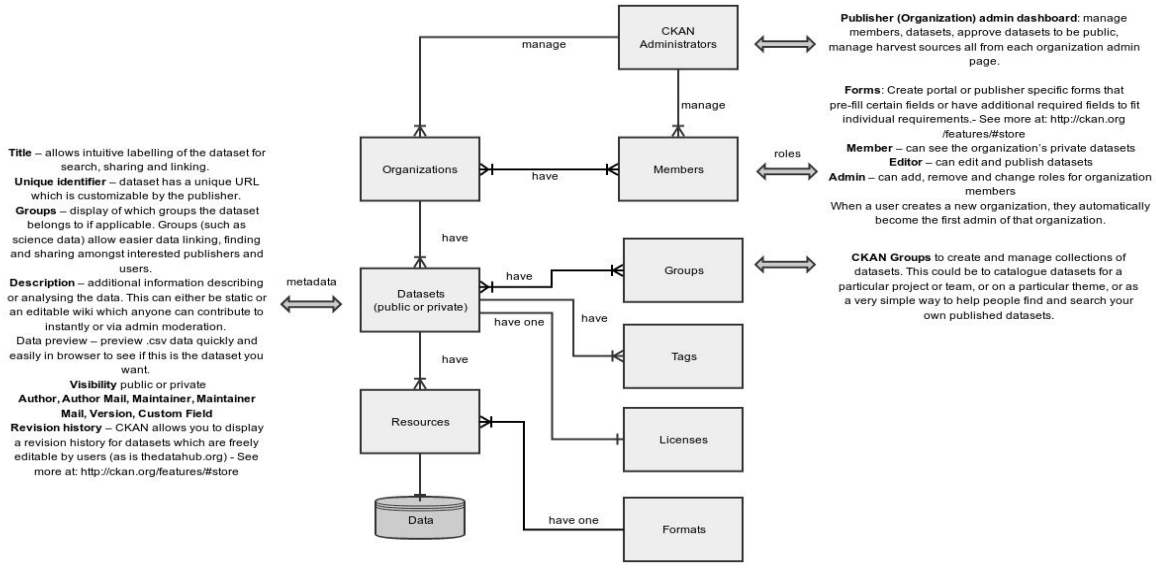

Figura 1: entità di CKAN e loro relazioni

[2] CKAN: *https://ckan.org/*

Un *dataset* in CKAN è un prodotto che appartiene (univocamente) ad *un'Organizzazione* e può essere pubblico o privato ad essa. *Pubblico* significa che è accessibile da chiunque utilizzi il Catalogo, *privato* significa che è accessibile ai soli utenti dell'organizzazione a cui appartiene il dataset. Un *dataset* è fatto da un insieme di metadati obbligatori (*Titolo, Licenza, Organizzazione di appartenenza e Autore, vedi Figura 2*) ed altri opzionali di coppie chiavi (univoche) e valori (chiamati *extra fields*). Esso può essere correlato da risorse (cioè file) e/o URL[3] che rappresentano dati di interesse del dataset.

Label	Field Name (API)	Definition	Guidelines	Example
Title*	title	Name given to the dataset.	Short phrase, written in plain language. Should be sufficiently descriptive to allow for search and discovery.	Aquaculture Production and Consumption in Africa (2011)
Description	description	Short description explaining the content and its origins.	Description of a few sentences, written in plain language. Should,provide a sufficiently comprehensive overview of the resource for anyone,to understand its content, origins, and any continuing work on it. The,description can be written at the end, since it summarizes key,information from the other metadata fields.	This dataset contains attributes of aquaculture production and,consumption for each of Africa's provinces in 2011. The data was,provided by.....
Tags	tags	An array of Taxonomic terms stored as tags	Taxonomic terms	Access to education, Bamboo
License*	lincese_title	the license that applies to published dataset.		
Organization*	organization	Organization the datasets belongs to	See list of organizations on https://ckan-d-d4s.d4science.org/organization	D4Science
Version	version	Version of dataset	Increase manually after editing	1.0
Author*		Owner of dataset	The person who created the dataset in the format: Surname, Name	Bloggs, Joe
Author Contact		Contact details of owner	The email or other contact details of the person who created the dataset.	joe@example.com
Mantainer		Mantainer of the dataset	The person or the authority that maintains the dataset	A person: Bloggs, Joe. An authority: D4Science
Mantainer Contact		Contact details of mantainer	The email or other contact details of the person who maintains the dataset.	joe@example.com

mandatory fields are marked with an asterisk (*)

Figura 2: metadati obbligatori di CKAN

3. Limiti di CKAN (nel contesto di D4Science)

L'idea e la necessità di avere un Catalogo nel contesto di D4Science per pubblicare i "prodotti della ricerca" e favorire la diffusione del "sapere" nasce in un contesto di ricerca multidisciplinare il cui obiettivo era rafforzare lo sviluppo di capacità ed aumentare le conoscenze scientifiche sullo sfruttamento delle risorse ambientali ed il degrado dei suoi ecosistemi[4]. La piattaforma CKAN allo "stato dell'arte" riesce a soddisfare un sottoinsieme dei requisiti risultanti dal contesto applicativo a causa delle seguenti limitazioni.

(i) Nello "schema" ossia nella definizione di un dataset/prodotto del catalogo:

- *Metadati con chiavi univoche* - sono un vincolo troppo restrittivo in quanto un "prodotto" potrebbe avere la necessità di avere chiavi ripetibili. Es. molteplici Autori;

- *Nessun modello per la definizione dei tipi di metadati* - impossibilità di definire il set minimo di metadati che un "item" deve possedere per essere di un certo "tipo" e quindi essere

[3] Uniform Resource Locator: è una sequenza di caratteri che identifica univocamente l'indirizzo di una risorsa su una rete di computer

[4] Per ulteriori dettagli si rimanda al progetto EU *https://bluebridge.d4science.org/*

conforme ad esso. Cos'è un dataset, o un servizio, o un prodotto della ricerca, ecc.. quali metadati deve avere per essere definito tale?

(ii) In fase di pubblicazione:

- *Nessun data entry guidato per gli attori*;
- *Nessun tipo di validazione dei dati prima della loro pubblicazione*;

(iii) In fase di visualizzazione:

- *Vista univoca "flat" (a tabella) dei metadati e non categorizzata* - es. raggruppata per classi di campi aventi particolari e/o le stesse "caratteristiche";

(iv) In termine di qualità dei metadati raccolti:

- *I Metadati pubblicati "difficilmente" risulteranno armonizzati* - la possibilità di utilizzare "chiavi libere" in fase di pubblicazione di un "prodotto" non rappresenta sicuramente un "buon metodo" per pubblicare dati "armonizzati" tra i diversi attori autorizzati a pubblicare nel catalogo.

4. Il Catalogo come servizio D4Science

Il Catalogo è un servizio fornito da D4Science e fa parte della sua piattaforma di publishing. Fornisce servizi per: (i) definire e pubblicare (meta)dati secondo un certo tipo, (ii) pubblicare i prodotti risultati della ricerca, (iii) rendere disponibili pubblicamente o non i prodotti (secondo policy di accesso), (iv) validare i metadati in fase di publishing, (v) rendere ricercabili i prodotti (tramite titolo, tag, ecc..), (vi) rendere accessibili i prodotti tramite PURL[5]. Il Catalogo è fornito tramite la tecnologia CKAN che nel contesto D4Science è stata estesa per superare i limiti riportati nella sezione 3.

4.1 Un Item del Catalogo

Un dataset CKAN nel contesto di D4Science viene chiamato "item" del Catalogo e la sua organizzazione di appartenenza è rappresentata in maniera univoca da un "Virtual Research Environment" (VRE) dell'infrastruttura di ricerca "D4Science". Vi è una corrispondenza 1:1 tra una Organizzazione CKAN ed un VRE di D4Science.

[5] PURL: Persistent uniform resource locator

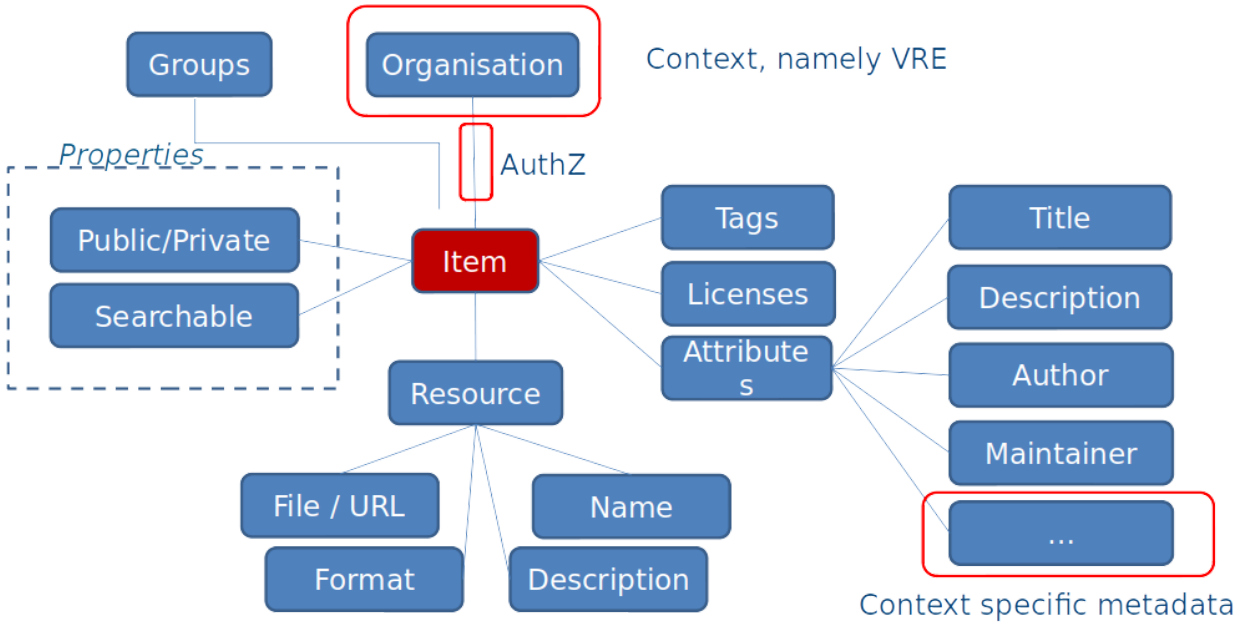

Figura 3: entità del Catalogo e loro relazioni

Nel contesto di D4Science lo schema di base (definito per l'engine Solr[6]) usato per indicizzare *gli item* del Catalogo è stato esteso per supportare chiavi multiple e superare uno dei limiti riportato nel paragrafo 3. Gli attributi di un item (properties) sono *Public* e *Private,* esclusivi tra loro ne specificano la politica di accesso (vedi *Figura 4*):

(i) Public - i metadati dell'item sono accessibili agli utenti guest, mentre le (eventuali) risorse correlate ad esso richiedono una user identity valida per essere accedute.

(ii) Private - i metadati dell'item e le (eventuali) risorse correlate ad esso sono accessibili solo tramite una user identity valida.

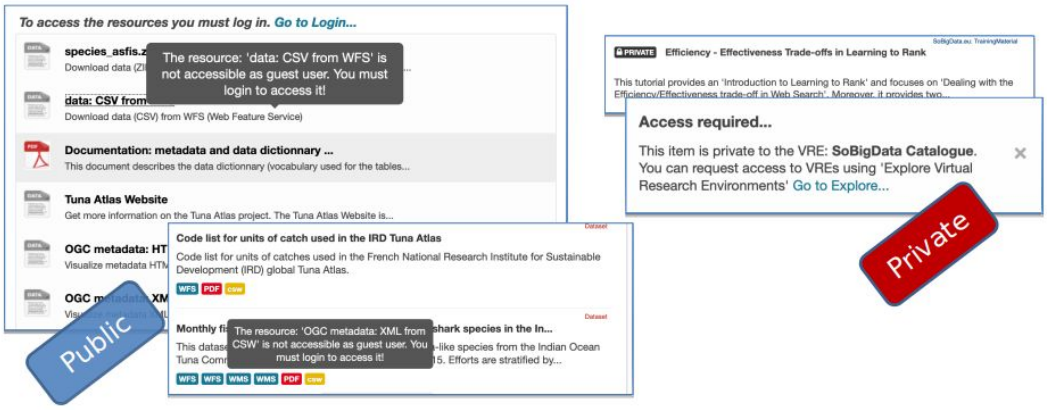

Figura 4: attributi Public e Private

[6] Apache Solr: https://lucene.apache.org/solr/

4.2 Architettura e componenti

L'architettura della piattaforma Catalogo di D4Science ed i suoi componenti sono mostrati nella *Figura 5*:

Figura 5: architettura del Catalogo

Di seguito verranno presentati i componenti principali della suddetta architettura e le loro caratteristiche peculiari.

4.2.1 Catalogue Item type: specifica del tipo di metadato

Un item del catalogo in base al contesto di ricerca (VRE) avrà dei metadati specifici conformi ad un certo "*tipo di (meta)dato"* (*catalogue item type*) definibile tramite un profilo XML-based chiamato "gCube Metadata Profile" (vedi *Figura 6*).

```
<?xml version="1.0" encoding="UTF-8"?>
<metadataformat type="YOUR TYPE HERE">
    <metadatafield categoryref="category_id_#">
        <fieldName>Name of Metadata Field</fieldName>
        <mandatory>true|false</mandatory>
        <dataType>String|Time|Time_Interval|Times_ListOf|Text|Boolean|Number|GeoJSON</dataType>
        <maxOccurs>N|*</maxOccurs>
        <defaultValue>default value</defaultValue>
        <note>[the note is shown as a suggestion in the insert/update metadata form of Catalogue Publisher Widget]
                </note>
        <vocabulary isMultiSelection="true|false">
            <vocabularyField>field1</vocabularyField>
            <vocabularyField>field2</vocabularyField>
            <vocabularyField>field3</vocabularyField>
        </vocabulary>
        <validator>
            <regularExpression>a regular expression for validating values</regularExpression>
        </validator>
        <tagging create="true|false" separator="char_to_separate">onFieldName|onValue|onFieldName_onValue|onValue_onFieldName</tagging>
        <grouping create="true|false">onFieldName|onValue|onFieldName_onValue|onValue_onFieldName</grouping>
    </metadatafield>
</metadataformat>
```

Figura 6: il gCube Metadata Profile

Un *gCube Metadata Profile* è composto da un *Metadata Format* (`<metadataformat>`) contenente una lista ordinata di (ed almeno un) *Metadata Field* (`<metadatafield>`). Un *Metadata Field* può avere un riferimento ad una *CategoryRef* (`categoryref="category_id_#"`). Aggiungere una categoria di riferimento a un metadata field significa che il campo appartiene ad una "classe"

con certe caratteristiche (*category*) che in fase di visualizzazione consente alla GUI del catalogo di presentare una vista raggruppata dei metadati per categoria di appartenenza.
Il nome del field è dichiarato tramite il *<fieldName>* la sua obbligatorietà o meno tramite il valore true/false nel tag <mandatory>.

Il campo *dataType* (*<dataType>)* specifica il tipo di dato. Un *dataType* valido deve essere uguale a uno dei valori[7]:

{String, Time, Time_Interval, Times_ListOf, Text, Boolean, Number, GeoJSON}

Al momento un gCube Metadata Profile per ogni suo field definito in esso è in grado di modellare e gestire:

(i) la ripetibilità del dato

<maxOccurs>N|*</maxOccurs>

(ii) il dato spaziale - può essere specificato utilizzando il valore *GeoJSON*

```
<metadatafield idref="category_id_#">
    <fieldName>spatial</fieldName>  <!--'spatial' is the reserved field name to assign a GeoSpatial dimension to metadata  -->
    <dataType>GeoJSON</dataType>
    <defaultValue>{"type": "Point","coordinates": [-20.145,74.078]}</defaultValue>
    <note>Please, insert a valid GeoJSON</note>
</metadatafield>
```

Dimensione Spaziale

(iii) il dato temporale - può essere specificato utilizzando il valore *Time* o *Time_Interval* o *Times_ListOf* (basato su *ISO_8601*[8])

```
<metadatafield idref="category_id_#">
    <fieldName>time_date</fieldName>  <!--'time_date' is the reserved field name to assign a Temporal dimension to metadata -->
    <dataType>Time</dataType>
    <defaultValue>2019-7-29</defaultValue>
    <note>Please, insert a valid ISO 8601 date</note>
</metadatafield>
```

Dimensione temporale

(v) il valore di default

<defaultValue>default value</defaultValue>

(vi) assumere valori da vocabolario controllato con o senza multiselezione

[7] In fase di data entry tramite il servizio del Catalogo se il tipo di dato non è specificato, il campo dei metadati assume il valore predefinito "String".
[8] Standard *ISO 8601:* https://www.iso.org/iso-8601-date-and-time-format.html

```
<vocabulary isMultiSelection="true|false">
    <vocabularyField>field1</vocabularyField>
    <vocabularyField>field2</vocabularyField>
    <vocabularyField>field3</vocabularyField>
</vocabulary>
```

(vii) la validazione del dato basata su regex

```
<validator>
  <regularExpression>insert your REGEX</regularExpression>
</validator>
```

(viii) direttive per la generazione automatica di tag e l'assegnazione a specifici gruppi.[9]

Lo schema XML per validare un Metadata Profile è reperibile al link:
https://wiki.gcube-system.org/images_gcube/e/e8/Gcdcmetadataprofilev3.xsd

"Il Tipo di Item" con statistiche, tipologie e facility di filtering e browsing su di esso sono informazioni e funzionalità accessibili tramite la GUI del Catalogo e riportati nella *Figura 7*

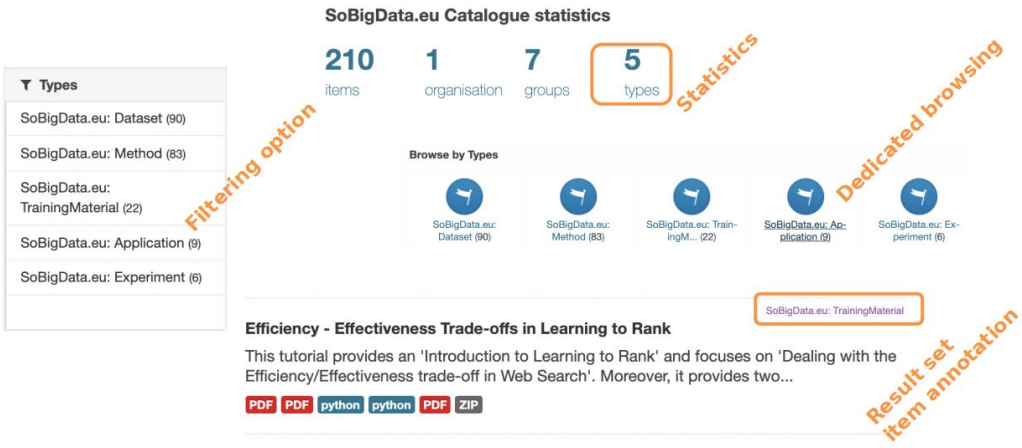

Figura 7: statistics, browsing e filtering per tipo di item

4.2.2 Il *Publishing Widget: pubblicare un item nel Catalogo*

[9] Per ulteriori dettagli si rimanda a *https://wiki.gcube-system.org/gcube/GCat_Background#gCube_Metadata_Profile*

Il Publishing Widget è il componente tramite il quale un utente autorizzato è “guidato” nella pubblicazione di un item nel Catalogo. Esso tramite il *Catalogue Service* recupera i “Metadata Profile” che sono stati creati nel VRE (e che ne definiscono i tipi di item pubblicabili) e crea un web-form con il set dei campi mandatory richiesti per un item e la richiesta di selezione di un tipo di Profile (il “Type”).

Lo scenario di pubblicazione in seguito alla selezione del “Type” è presentato nella *Figura 8*

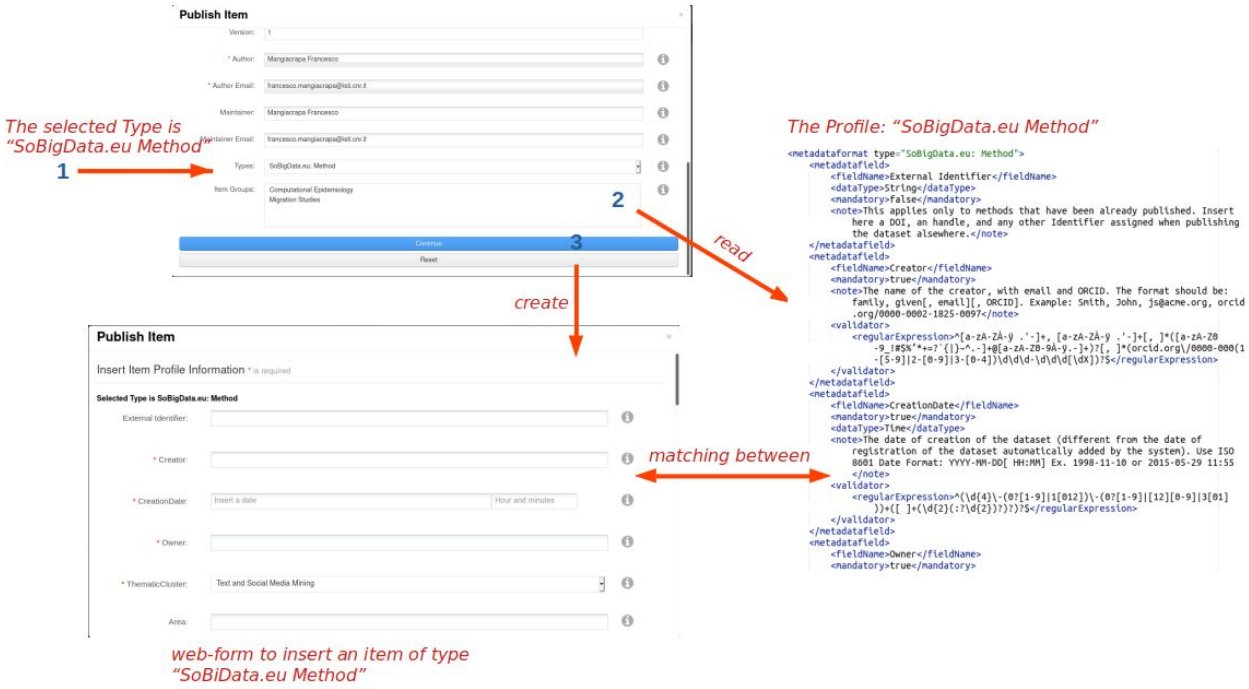

Figura 8: scenario di pubblicazione di un item

Selezionato il tipo di item (*Figura 8, step 1*), il Publishing Widget legge il Profile corrispondente (*Figura 8, step 2*) definito nella VRE e crea un web-form (*Figura 8, step 3*) per il data entry che corrisponde esattamente in termini di metadati e suoi “constraint” a quelli definiti nel profilo scelto. I campi corredati di una regex (field `<validator>`) sono validati in questa fase e richiesta la loro correzione in caso di mismatching con la regex. Quindi, in seguito alla scelta del tipo, il Publishing Widget creerà un web-form e pertanto un item conforme ai campi dichiarati nel profilo che si vuole pubblicare.

4.2.3 Il *Catalogue Portlet: discovery e browsing dei (meta)dati*

Il *Catalogue Portlet* è la GUI (vedi *Figura 9*) che permette il discovery ed il browsing dei (meta)dati inseriti nel Catalogo, di avere la lista degli item, le sue organizzazioni, i suoi gruppi ecc. Per un utente accreditato di vedere le organizzazioni ed i gruppi di cui fa parte e di recuperare gli item che ha pubblicato. Inoltre fornisce le statistiche del Catalogo.

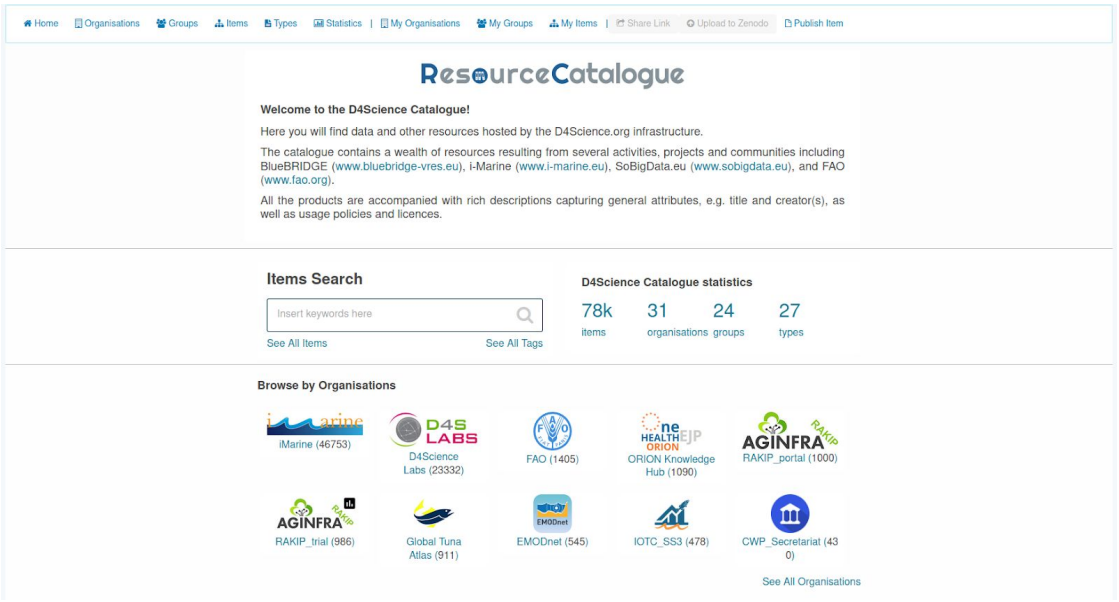

Figura 9: la GUI del Catalogue Portlet

La vista che il *Catalogue Portlet* presenta per un item è riportata nella *Figura 10*

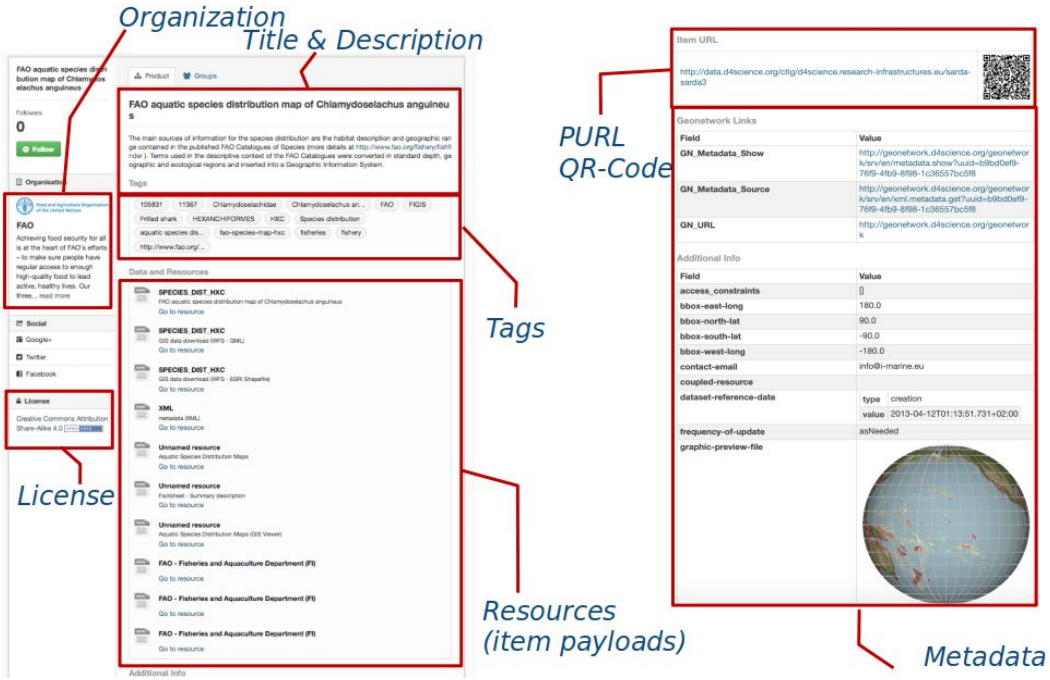

Figura 10: la vista di un item

Un esempio di vista dei metadati (con chiavi ripetibili) raggruppata per Categoria è mostrata in *Figura 11*

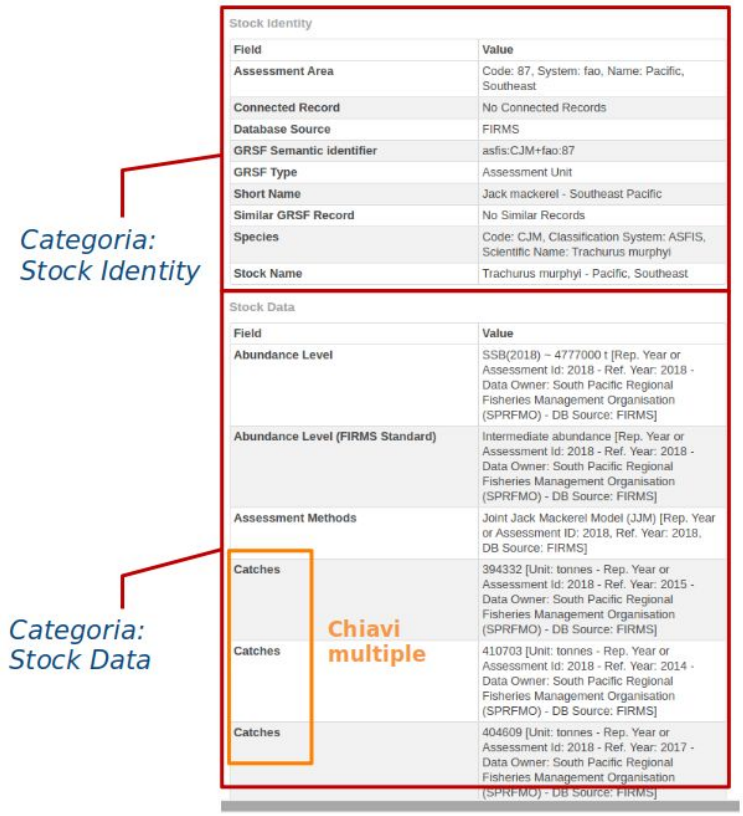

Figura 11: vista per categoria dei metadati

Il *Catalogue Portlet* usa il *Publishing Widget* per pubblicare un item nel Catalogo ed è integrato con una serie di servizi:

(i) *il Workspace* - per pubblicare nel Catalogo un file oppure un folder salvato nell'ambiente di lavoro comune agli utenti dell'infrastruttura D4Science (vedi *Figura 12*);

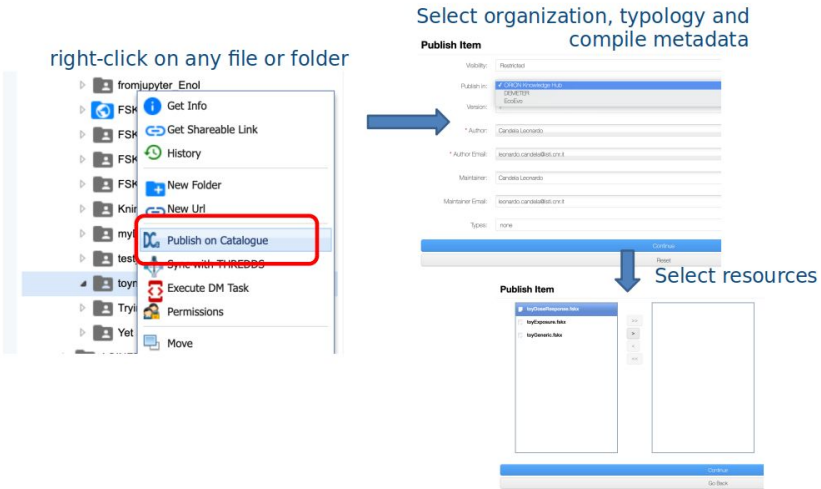

Figura 12: pubblicazione di un item dal Workspace

(ii) *il Social Networking* - per la notifica via social post agli utenti della VRE che un nuovo item del catalogo è stato pubblicato (vedi *Figura 13*);

Figura 13: post di notifica di un item pubblicato

(iii) URI-Resolver: tramite la GUI Share Link è possibile ottenere lo PURL di un item per la sua condivisione (vedi *Figura 14*);

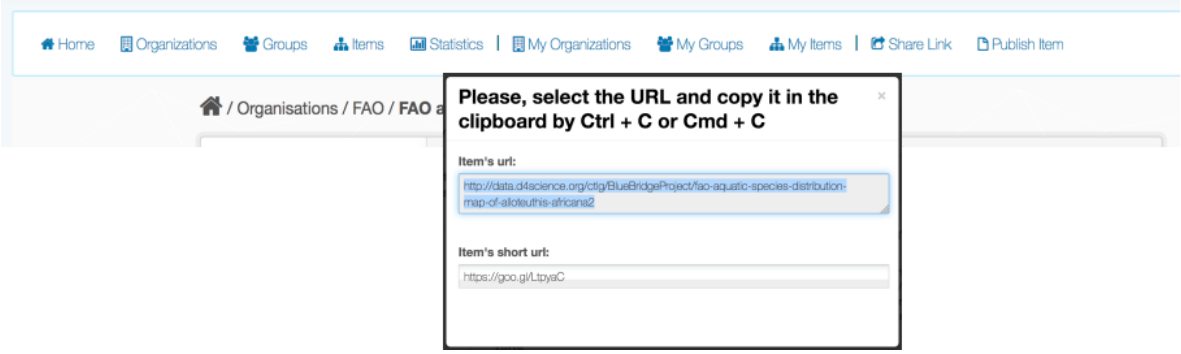

Figura 14: lo share link di un item

(iv) *Zenodo*[10] - per la pubblicazione di un item del Catalogo nel repository Zenodo (vedi *Figura 15*).

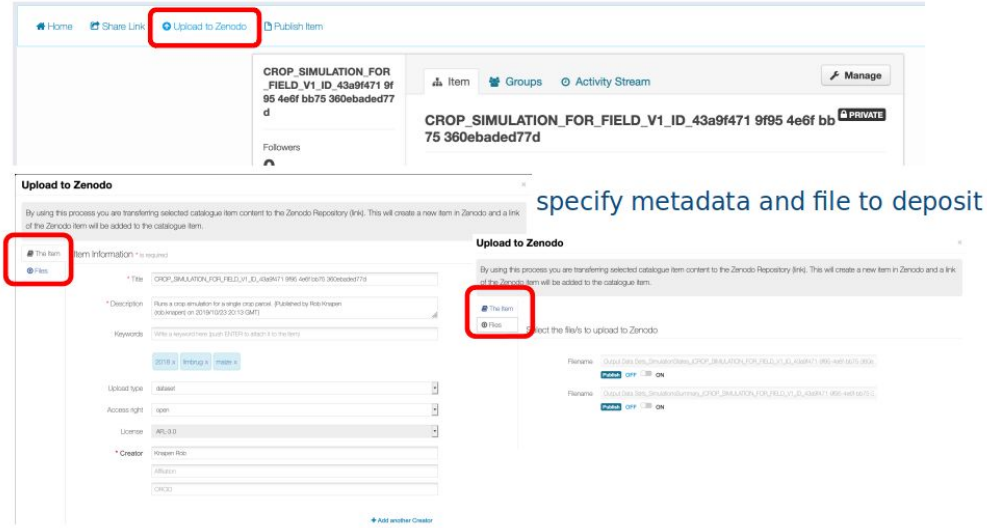

Figura 15: upload di un item su Zenodo

[10] Zenodo: https://zenodo.org/

4.4 (Altre) Interfacce di accesso e raccolta dati

Nel paragrafo *4.2.3* è stato presentato il componente "*Catalogue Portlet*" che rappresenta la GUI per gli utenti del Catalogo interessati ad accedere ai contenuti pubblicati.
Dal punto di vista "programmatico", il Catalogo, offre un API tramite il servizio *gCat*[11]. *gCat* è un servizio RESTful che consente a qualsiasi client HTTP di interagire in modo programmatico con il Catalogo.
Altro aspetto importante riguarda le interfacce di accesso per la raccolta dei dati (harvesting) *da* "fonti" esterne *verso* il Catalogo, quindi la capacità del Catalogo di raccogliere meta(dati) da servizi esterni che espongono dei protocolli di harvesting dei loro dati. Al momento il Catalogo offre un sistema per collezionare contenuti da server che espongono i loro metadati mediante lo standard OGC[12] "*Catalog Service for the Web*" (CSW[13]) ed ottimizza la raccolta delle informazioni. L'ottimizzazione è frutto di un lavoro di miglioramento di un *plugin CKAN* (*GeoNetwork harvester for CKAN*[14]), ed in particolare nel mapping tra i metadati *ISO19139*[15] e dataset di CKAN, lavoro fatto in collaborazione con la FAO[16] (usando come "campione" i metadati forniti tramite il server CSW *GeoNetwork* della FAO), onde evitare la perdita di metadati nel passaggio tra i due sistemi in seguito al processo di harvesting.

5. Il Catalogo in azione

Diverse istanze del servizio Catalogo sono state create per gestire le necessità di varie comunità e casi d'uso. Ad oggi esistono 13 istanze (di produzione) della piattaforma che servono altrettante comunità scientifiche. Nella tabella seguente vengono mostrate le comunità servite dal Catalogo con il numero di profili creati ed item pubblicati per ognuna di essa.

Nome della Comunità (numero di utenti)[17]	Istanza del Catalogo	Numero di Profili creati	Numero di Item pubblicati
Aginfra (589)	http://ckan-aginfra.d4science.org	17	3.364

[11] gCat: https://wiki.gcube-system.org/gcube/GCat_Service *sviluppato da Luca Frosini presso ISTI-CNR (luca.frosini@isti.cnr.it)*
[12] OGC: Open Geospatial Consortium https://www.ogc.org
[13] CSW: uno standard del OGC che definisce un'interfaccia per servizi di ricerca, di navigazione e di interrogazione di metadati su dati e servizi che si riferiscono e vogliono pubblicare metadati geo-spaziali.
[14] *Geonetwork Harvester:* https://github.com/geosolutions-it/ckanext-geonetwork
[15] ISO 19139: fornisce lo schema di implementazione XML per ISO 19115 che specifica il formato del record di metadati e può essere utilizzato per descrivere, convalidare e scambiare metadati geospaziali preparati in XML.
[16] FAO: http://www.fao.org/home/en/
[17] Il numero di utenti riportato è quello degli utenti registrati nel portale. Gli utenti che usano i servizi offerti dal portale in modalità "guest" non sono documentati (in alcuni casi eccedono largamente il numero di utenti registrati, e.g. in GRSF).

Blue-Cloud (428)	https://ckan-bluecloud.d4science.org *(Catalogo in fase di validazione e rilascio)*	2	2
Desira (342)	https://ckan-desira.d4science.org	1	12
EcoEvo (22)	https://ckan-ecoevo.d4science.org	4	23
EOSC_Stakeholder-Registry (155)	https://ckan-eosc.d4science.org	2	5
EOSC-Pillar (573)	https://ckan-eoscpillar.d4science.org	14	81.035
GRSF (70)	https://ckan-grsf.d4science.org	3	1.522
GRSF for Admins (41)	https://ckan-grsf-admin2.d4science.org	5	23.570
i-Marine (665)	https://ckan-imarine.d4science.org	9	74.380
RISIS2 (38)	https://ckan-risis2.d4science.org	3	122
SoBigData (7140)	https://ckan-sobigdata.d4science.org	7	233
Territori Aperti (61)	https://ckan-territoriaperti.d4science.org	8	206
D4Science Catalogue (3062)	https://catalogue.d4science.org	27	78.089

dati aggiornati al 29 Novembre 2020

Dalla tabella sopra riportata si evince che il numero totale di Metadata Profile creati è *102*, il numero totale di item pubblicati nei Cataloghi è *262.563*.

Link alla piattaforma open-source Catalogue:

- *Software*: https://code-repo.d4science.org/gCubeSystem/gcube-ckan-datacatalog
- *Versioni rilasciate*[18]: https://code-repo.d4science.org/gCubeSystem/gcube-ckan-datacatalog/releases
- *Documentazione*: https://wiki.gcube-system.org/GCat_Background

6. Conclusioni e Sviluppi futuri

In questo documento è stato presentato il "Catalogo" come piattaforma di publishing open-source e servizio per le comunità che l'infrastruttura di ricerca *D4Science* gestisce presso ISTI-CNR. Il contesto applicativo, fortemente eterogeneo e multidisciplinare, pone una serie di sfide e esigenze per la realizzazione e la gestione di un servizio "catalogo" che sia efficace ed effettivo. Le soluzioni studiate ed implementate introducendo la nozione di Metadata Profile rappresentano un valido approccio che ha permesso di gestire casi di uso fortemente diversi tra loro senza dover ricorrere ad implementazioni ad-hoc. L'utility Publishing Widget che costruisce

[18] Le versioni precedenti alla *v.1.7.0* sono disponibili sul repository SVN dell'infrastruttura D4Science

il web-form per il data entry in base al tipo che si vuole pubblicare aggiunge valore e semplicità di uso all'intera offerta. Infatti quest'ultima utility "guida" gli utenti chiamati all'inserimento dei meta(dati), garantendo "armonizzazione" dei metadati pubblicati nel Catalogo e conformità rispetto ai tipi definiti. Il risultato è pertanto una migliore "qualità" nei metadati pubblicati. Inoltre, sono state realizzate e presentate le facility di integrazione del Catalogo con altri servizi di D4Science (i) il "*Workspace*": per pubblicare i prodotti di interesse *dal* Workspace *nel* Catalogo, (ii) "*Zenodo*": per pubblicare un item del Catalogo su Zenodo, e con quest'ultimo si è migliorata l'integrazione verso un diffuso repository utilizzato per le pubblicazioni in ambito scientifico. Infine, sono state presentate alcune statistiche e l'utilizzo della piattaforma Catalogo da parte delle comunità scientifiche di D4Science gestite presso ISTI-CNR.

Gli indicatori di uso del servizio nonché i feedback ricevuti dalle comunità indicano che quanto proposto incontra le esigenze di vari contesti applicativi facendone un prodotto maturo e flessibile.

Il Catalogo è in continua fase di miglioramento, (i) nella GUI, (ii) nel modello per la definizione dei tipi di metadati (i Metadata Profile), (iii) nella sue interfacce di accesso tramite servizi per la raccolta dei dati (harvesting) *da* fonti esterne *verso* il Catalogo e viceversa. Da quest'ultimo punto di vista un aspetto migliorativo già in programma come sviluppo futuro è l'implementazione e la possibilità di accedere ai dati del Catalogo tramite il protocollo OAI-PMH[19].

Pubblicazioni

Produzione scientifica in cui si parla del Catalogo e/o a cui esso ha contribuito.

2020

- *[Article] Assante, M, Boizet, A, Candela, L, Castelli, D.; Cirillo, R.; Coro, G.; Fernández, E.; Filter, M.; Frosini, L.; Georgiev, T.; Kakaletris, G.; Katsivelis, P.; Knapen, R.; Lelii, L.; Lokers, R. M.; Mangiacrapa, F.; Manouselis, N.; Pagano, P.; Panichi, G.; Penev, L.; Sinibaldi, F.* ***Realizing virtual research environments for the agri-food community: The AGINFRA PLUS experience****. Concurrency Computat Pract Exper. 2020;e6087. https://doi.org/10.1002/cpe.6087*
- *[Article] Ribeiro Duarte, A.S.; Liv Nielsen, C.; Candela, L.; Valentin, L.; Aarestrup, F. A.; Vigre, H.* ***Global Food-source Identifier (GFI): Collaborative virtual research environment and shared data catalogue for the foodborne outbreak investigation international community*** *Food Control, Volume 121, 2021, https://doi.org/10.1016/j.foodcont.2020.107623*
- *[Book chapter] Candela, L.; Stocker, M.; Häggström, I.; Enell, C.-F.; Vitale, D.; Papale, D.; Grenier, B.; Chen, Y.; Obst, M.* ***Case Study: ENVRI Science Demonstrators with D4Science****. Towards Interoperable Research Infrastructures for Environmental and Earth Sciences 2020: 307-323 https://doi.org/10.1007/978-3-030-52829-4_17*
- *[Book chapter] Jeffery, K. G.; Candela, L.; Glaves, H.* ***Virtual Research Environments for Environmental and Earth Sciences: Approaches and Experiences****. Towards Interoperable*

[19] Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH): è un insieme di sei verbi e/o servizi accessibili tramite HTTP per consentire l'interoperabilità tra repository che vogliono esporre archivi di metadati tramite il protocollo OAI-PMH https://www.openarchives.org/pmh

Research Infrastructures for Environmental and Earth Sciences 2020: 272-289 https://doi.org/10.1007/978-3-030-52829-4_15

2019

- *[Article] Assante M., Candela L., Castelli D., Cirillo R., Coro G.,Frosini L., Lelii L., Mangiacrapa F., Pagano P., Panichi G., Sinibaldi F.* ***Enacting open science by D4Science****. Future generation computer systems. http://doi.org/10.1016/j.future.2019.05.063*
- *[Conference paper] Assante M., Candela L., Castelli D., Coro G., Mangiacrapa F., Pagano P., Perciante C.* ***Enacting Open Science by gCube****. IWSG 2017 - International Workshop on Science Gateways - Proceedings of the 9th International Workshop on Science Gateways, Poznan, Poland, 19-21 June, 2017 http://ceur-ws.org/Vol-2363/paper3.pdf*
- *[Conference paper and Part of book] Tzitzikas Y., Marketakis Y., Minadakis N., Mountantonakis M., Candela L., Mangiacrapa F., Pagano P., Perciante C., Castelli D., Taconet M., Gentile A., Gorelli G.* ***Methods and tools for supporting the integration of stocks and fisheries****. Information and Communication Technologies in Modern Agricultural Development, edited by Salampasis M., Bournaris T., pp. 20–34, 2019 http://doi.org/10.1007/978-3-030-12998-9_2*

2018

- *[Article] Assante M., Candela L., Castelli D., Cirillo R., Coro G., Frosini L., Lelii L., Mangiacrapa F., Marioli V., Pagano P., Panichi G., Perciante C., Sinibaldi F.* ***The gCube system: delivering virtual research environments as-a-service****. Future generation computer systems (2018). http://doi.org/10.1016/j.future.2018.10.035*
- *[Article] Culina, A., Baglioni, M., Crowther, T.W. et al.* ***Navigating the unfolding open data landscape in ecology and evolution****. Nat Ecol Evol 2, 420–426 (2018). https://doi.org/10.1038/s41559-017-0458-2*
- *[Conference paper] A. Ballis, A. Boizet, L. Candela, D. Castelli, E. Fernández, M. Filter, T. Günther, G. Kakaletris, P. Karampiperis, D. Katris, M. J. R. Knapen, R. M. Lokers, L. Penev, G. Sipos, P. Zervas* ***Serving Scientists in Agri-Food Area by Virtual Research Environments****. eScience 2018: 405-406 https://doi.org/10.1109/eScience.2018.00124*

2017

- *[Conference paper] Tzitzikas Y., Marketakis Y., Minadakis N., Mountantonakis M., Candela L., Mangiacrapa F., Pagano P., Perciante C., Castelli D., Taconet M., Gentile A., Gorelli G.* ***Towards a global record of stocks and fisheries****. 8th International Conference on Information and Communication Technologies in Agriculture, Food and Environment, pp. 328–340, Chania, Greece, 21-24 http://ceur-ws.org/Vol-2030/HAICTA_2017_paper39.pdf*