

Accessing biodiversity databases: a domain specific query language and a unifying data model

Leonardo Candela^{1,*}, Federico De Faveri², Lucio Lelii¹, Francesco Mangiacrapa¹,
Valentina Marioli¹, Pasquale Pagano¹

Abstract

Species data are scattered among several databases and information systems. During the last years, considerable progresses have been made in developing on-line species databases. However, there is no single database that can claim to host, and make available in a seamless way, all the species data needed by the communities willing to have access to such typology of data. In this report we present a domain specific query language and a unifying data model for species data that characterise a mediator service specifically conceived to act as a single access point to the plethora of existing species databases.

Keywords

Species Data Integration — Distributed Information Retrieval — Domain Specific Query Language — Species Data Model

¹ Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", Consiglio Nazionale delle Ricerche, Via G. Moruzzi 1, 56124, Pisa, Italy

² Food and Agriculture Organisation (FAO) of the United Nations, 00100, Rome, Italy

*Corresponding author: leonardo.candela@isti.cnr.it

Contents

1	Introduction	1
2	SPD Reference Architecture	2
3	SPeCies Query Language	2
3.1	Query Evaluation	3
3.2	Query Examples	4
4	Unifying Data Model	4
4.1	Occurrence Object	5
4.2	Taxonomy Object	5
5	Mappings	6
6	Conclusion	7
	Acknowledgments	7
	References	7

1. Introduction

When dealing with species data, one of the major issues is to provide users with a single access point to the plethora of databases and information systems giving access to such typology of data. A number of initiatives (e.g., GBIF) aggregate large amount of data from different databases and publish integrated versions of them through a single uniform interface. In order to implement such services they ask to the database providers to adhere to established publication guidelines, formats and protocols. Moreover, during the aggregation phase they apply specific transformations in order to generate the required unified view. Usually, these transformations are not only limited to the syntactic format. They often

implement harmonisation and quality enhancement practices that are decided by the service provider and are not explicitly made known to the data consumers.

In this report, it is presented the approach promoted by *Species Products Discovery* (SPD) [1, 2], a gCube¹ service for discovering and accessing species data, both occurrence data and nomenclature data. SPD is conceived as a sort of mediator service [13] over a number of databases. In particular, the service is equipped with a number of plug-ins interfacing with major information systems, including *GBIF* [14], *Catalogue of Life* [15], *speciesLink* [16], *List of Species of the Brazilian Flora* [17]. Differently from the other solutions provided so far in the biodiversity domain, SPD does not impose any specific guideline or protocol/format to the databases and information systems it aggregates. Rather, it is conceived to deal with the heterogeneity and challenges resulting from a scenario where the providers are neither expected to be collaborative nor to modify their strategies for data publication. Moreover, SPD does not build an aggregated database. Rather, it realises data aggregation dynamically, at query time.

The remainder of this report is organised as follows. Section 2 briefly describes the reference architecture characteris-

¹gCube [3, 4] is a software system specifically conceived to enable the creation and operation of an innovative typology of infrastructure, i.e., an *Hybrid Data Infrastructure* [5], that by leveraging Grid [6], Cloud [7], Digital Library [8] and Service-orientation [9] principles and approaches is delivering a number of data management facilities *as-a-Service*. One of its distinguishing feature is the orientation to serve the needs of diverse Communities of Practice [10] by providing each of them with a dedicated, flexible, ready-to-use, web-based working environment, i.e., a *Virtual Research Environment* [11, 12].

ing the service. Section 3 presents a domain specific query language, i.e., SPecies Query Language (SPQL), devised to express typical queries. Section 4 discusses the unifying data model for representing in an homogeneous way occurrence objects and taxonomy objects resulting when querying the aggregated databases. Section 5 describes the mappings put in place to transform objects coming from the various databases into objects in the unifying data model. Finally, Section 6 concludes the report.

2. SPD Reference Architecture

The Reference Architecture characterising the SPD service is depicted in Figure 1.

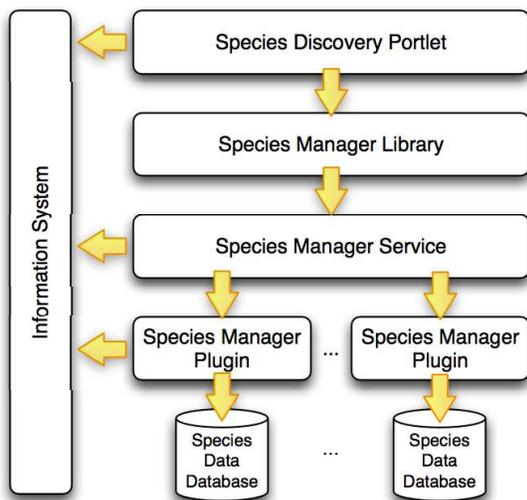


Figure 1. SPD Service Architecture

The *Species Discovery Portlet* implements the GUI of the entire service. It offers a a web-based user interface (Fig. 2) consisting of (i) a search panel (on the top) for specifying the information need, (ii) a results view panel (on the right) for browsing the list of datasets resulting from a query, and (iii) a classification panel (on the left) offering clustered views of the results, e.g., the classification, the data provider, the database, and the rank.

The *Species Manager Library* is a client library implementing a high-level façade to the remote APIs of the Species Manager Service.

The *Species Manager Service* implements the main engine of the service. It is a stateless Web Service that exposes read operations and implements it by delegation to dynamically deployable plugins for target databases and/or information systems.

The *Species Manager Plugins* implement the business logic needed to interact with specific databases and/or information systems. Each plugin take care of (i) the rewriting of queries from the unifying SPD query language (cf. Sec. 3) to the query language supported by the specific data provider,

and (ii) the transformation of results from the specific data provider format to the unifying SPD format (cf. Sec. 4).

The *Information System* is a core component of the gCube system. It acts as the registry of the distributed system and enables the discovery of the service instances forming it. In the case of the SPD service it offers information about the available plugins and their capabilities. Moreover, it provides the plugins with the base url to use to interface with the databases and information systems.

3. SPecies Query Language

In order to enable users to make queries into species databases or information systems, it was defined the query language reported in Table 1. Such a query language is characterised by the following clauses:

SEARCH BY to specify the “terms” characterising the query. Each term can be either a common name (CN) or a scientific name (SN).

In the case of a common name, the user is requested to specify the “resolve” data sources, i.e., the data sources to use to acquire the scientific names associated with the specified common name. Moreover, the user can specify the “expand” data sources, i.e., the data sources that will be used to acquire synonyms of the scientific names to “expand” the query initially specified. The entire set of scientific names resulting from the “resolve” and “expand” clauses will be actually used to execute the query.

In the case of a scientific name, the user can specify the “unfold” data sources, i.e., the data sources to use to acquire additional scientific names of “sub-species” associated with the specified scientific name. Moreover, the user can specify the “expand” data sources, i.e., the data sources that will be used to acquire synonyms of the scientific names to “expand” the query initially specified. The entire set of scientific names resulting from the “unfold” and “expand” clauses in addition to the initially specified scientific name will be actually used to execute the query.

IN to specify the list of databases / information systems target of the query, i.e., the data sources that will provide records matching the query.

WHERE to specify the criteria to filter the results. Each criterion expresses a condition on either the coordinates or the dates characterising the needed data.

RETURN to specify the typology of product to search for ranging from “any” product (PRODUCT) to occurrence datasets (OCCURRENCE) or taxonomic records (TAXON).

HAVING to specify the criteria to apply as a post-filter to the results. Each criterion can be either a “simple” condition on a property of a result item or a “complex” condition specified by an XPath or an EXL expression.

The screenshot shows the Species Discovery Portlet interface. At the top, there is a search bar with 'Occurrence' selected for 'Search', 'Scientific name' for 'By', and 'Sarda sarda' for 'Term'. Below the search bar are several filter tabs: 'Advanced Option', 'Filter by Source', 'Filter by BBox', 'Filter by Date', 'Synonyms From', and 'Expand'. The main area is titled 'Filter your results' and contains a table of search results. The table has columns for 'Data S...', 'Dataset', 'Name', 'S.N. Authorship', 'Matching', 'Rank', and 'Occ...'. The results list various data sources like GBIF, OBIS, and KUBI, along with their respective datasets and matching information. At the bottom, there is a 'Count' row showing 175 results and a 'Page 1 of 3' indicator.

Data S...	Dataset	Name	S.N. Authorship	Matching	Rank	Occ...
GBIF	Fish collection, Natural Histor...	Sarda s...	not found	Fish collection, Natural Hist...	Spe...	8
GBIF	Colecci ó n Nacional de Pece...	Sarda s...	not found	Colecci ó n Nacional de Pe...	Spe...	1
OBIS	iziko South African Museum - ...	Sarda s...	(Bloch, 1793)	Intergovernmental Oceano...	Spe...	2
GBIF	Biodiversity Research and Te...	Sarda s...	not found	Biodiversity Research and ...	Spe...	1
OBIS	Pembrokeshire Marine Specie...	Sarda s...	(Bloch, 1793)	Intergovernmental Oceano...	Spe...	1
GBIF	KUBI Ichthyology Collection	Sarda s...	not found	KUBI Ichthyology Collection	Spe...	1
OBIS	Canadian Museum of Nature -...	Sarda s...	(Bloch, 1793)	Intergovernmental Oceano...	Spe...	4
GBIF	KUBI Ichthyology Tissue Colle...	Sarda s...	not found	KUBI Ichthyology Tissue C...	Spe...	1
OBIS	NOAA Southeast Fishery Scie...	Sarda s...	(Bloch, 1793)	Intergovernmental Oceano...	Spe...	75
Count						175

Figure 2. Species Discovery Portlet

```

SEARCH BY term [, ...]
  [ IN data source [, ...] ]
  [ WHERE condition [AND condition] ]
  [ RETURN PRODUCT | OCCURRENCE | TAXON ]
  [ HAVING having expression ]

Where term can be one of:
CN common name [, ...]
  RESOLVE [ WITH data source [, ...] ]
  [EXPAND [ WITH data source [, ...] ] ]
SN scientific name [, ...]
  [UNFOLD WITH data source ]
  [EXPAND [ WITH data source [, ...] ] ]

Where condition can be one of:
coordinate OP Lat, Long
eventDate OP Date
and OP: < | <= | == | >= | >

Where having expression is a boolean expression composed by those elements:
item expression: item property OP value
function call: xpath(xpath expression) | exl(exl expression)
    
```

Table 1. SPecies Query Language Syntax

3.1 Query Evaluation

The following approach is used to reply to a query.

The evaluation process starts from the terms. Terms are elaborated depending on the specified type.

In the case of a common name, the engine transforms it into the set of scientific names associated with the common name by using the specified data sources or all the available data sources if no one has been specified. If the EXPAND clause is specified the list of scientific names is further expanded by using the synonyms provided by the specified data

sources or by using all the available data sources if no one has been specified.

In the case of a scientific name, if the UNFOLD clause is specified the list of scientific names is expanded by using the descendants species provided by the specified data source. If the EXPAND clause is specified, the list of scientific names is further expanded by using the synonyms provided by the specified data sources or using all the available data sources if no one has been specified.

The resulting list of species scientific names is used to

retrieve the objects matching the typology specified in the RETURN clause. If such clause is not present, any product is returned.

If the WHERE clause is specified, all the elements that do not satisfy the condition are eliminated from the result set. This filtering task is performed by the specific plug-in. Depending on the data source capabilities, it can be implemented by sending a proper query to the data source or by programmatically filtering the records.

If the HAVING clause is specified, all the elements that do not satisfy the having expression are eliminated from the result set. This filtering task is performed by the service.

The items in the result sets resulting from the various data sources are not ranked, thus there is no need to re-rank them once aggregated in the global result set. Moreover, the global result set is dynamically populated as soon as result items are produced by the single data sources via the plug-ins mediation.

3.2 Query Examples

Here are examples of queries that can be expressed by using the SPQL.

- Q1** Search for all the occurrence data about the scientific name 'sarda sarda' within all the available data sources.

```
SEARCH BY SN 'sarda sarda'
RETURN OCCURRENCE
```

- Q2** Search for all the occurrence data about the scientific name 'sarda sarda' and its synonyms in WoRMS, within all the available data sources.

```
SEARCH BY SN 'sarda sarda'
EXPAND WITH WoRMS
RETURN OCCURRENCE
```

- Q3** Search for all the occurrence data about the scientific name 'sarda sarda' and its synonyms in WoRMS, within the GBIF data source.

```
SEARCH BY SN 'sarda sarda'
EXPAND WITH WoRMS
IN GBIF
RETURN OCCURRENCE
```

- Q4** Search for all the taxonomic records about the scientific name 'sarda sarda' and its synonyms in WoRMS, within the ITIS data source.

```
SEARCH BY SN 'sarda sarda'
EXPAND WITH WoRMS
IN ITIS
RETURN TAXON
```

- Q5** Search for all the occurrence data about the scientific names retrieved by searching the common 'shark' in the data source WoRMS and mapping it into a scientific name by using ITIS. Accept only the results with coordinate less or equals to (15.12, 16.12).

```
SEARCH BY CN 'shark'
RESOLVE WITH WoRMS
EXPAND WITH ITIS
WHERE coordinate <= 15.12, 16.12
RETURN OCCURRENCE
```

- Q6** Search for all the occurrence data about the scientific names 'sarda sarda' and 'Carcharodon carcharias' expanded with synonyms from the WoRMS data source, that are in ITIS and with an event date between 2000 and 2005.

```
SEARCH BY SN 'sarda sarda',
'Carcharodon carcharias'
EXPAND WITH WoRMS
IN ITIS
WHERE eventDate >= '2000'
AND eventDate <= '2005'
RETURN OCCURRENCE
```

- Q7** Search for all the occurrence data about the scientific names retrieved by searching the common 'shark' in the data source OBIS and mapping it into a scientific name by using ITIS. Accept only the results with coordinate less than or equals to (15.12, 16.12). Filter the results by using an xpath expression.

```
SEARCH BY CN 'shark'
RESOLVE WITH OBIS
EXPAND WITH ITIS
WHERE coordinate <= 15.12, 16.12
RETURN OCCURRENCE
HAVING xpath("//product[type='TAXON'
and counter>0]")
```

- Q8** Search for all the taxonomic records in CatalogueOfLife, OBIS, WoRDSS, WoRMS by using the scientific name 'cervidae' and its sub-species resulted by using CatalogueOfLife and expanding the results with synonyms retrieved from the WoRMS data source.

```
SEARCH BY SN 'cervidae'
UNFOLD WITH CatalogueOfLife
EXPAND WITH WoRMS
IN CatalogueOfLife, OBIS, WoRDSS, WoRMS
RETURN TAXON
```

4. Unifying Data Model

SPD aims at offering a unifying view over the data made available by the diverse data sources. In order to do that, the service makes available two typologies of objects:

- *Occurrence Objects*, i.e., characterisations of a dataset of species occurrence records;
- *Taxonomy Objects*, i.e., characterisations of species taxa;

by interacting with the diverse data sources.

4.1 Occurrence Object

An Occurrence Object is characterised by the following 20 attributes² (alphabetic order):

basisOfRecord The specific nature of the data record. Recommended best practice is to use a controlled vocabulary³.

catalogNumber An identifier (preferably unique) for the record within the data set or collection.

citation Citation information specified by the data publisher.

collectionCode The name, acronym, coden, or initialism identifying the collection or data set from which the record was derived.

country The name of the country or major administrative unit in which the Location occurs.

credits Attribution information specified according to the data publisher policy.

dateLastModified The most recent date-time on which the resource was changed.

decimalLatitude The geographic latitude.

decimalLongitude The geographic longitude.

eventDate The date-time when the event was recorded.

family The full scientific name of the family in which the taxon is classified.

id This is the Global Unique Identification (GUID) for the specimen. This identification code should be stable and uniquely identify the specimen relative to all other specimens within the world.

identifiedBy A list (concatenated and separated) of names of people, groups, or organizations who assigned the Taxon to the subject.

institutionCode The name (or acronym) in use by the institution having custody of the object(s) or information referred to in the record.

kingdom The full scientific name of the kingdom in which the taxon is classified.

maximumDepthInMeters The greater depth of a range of depth below the local surface, in meters.

minimumDepthInMeters The lesser depth of a range of depth below the local surface, in meters.

recordedBy A list (concatenated and separated) of names of people, groups, or organizations responsible for recording the original Occurrence. The primary collector or observer, especially one who applies a personal identifier (recordNumber), should be listed first.

scientificName The scientific name of taxon with or without authorship information depending on the format of the source database.

scientificNameAuthorship The authorship information for the scientificName formatted according to the conventions of the applicable nomenclaturalCode.

4.2 Taxonomy Object

A TaxonomyItem object is characterised by the following 13 attributes⁴ (alphabetic order):

author The authorship information for the scientificName represented by the TaxonomyItem. If the authority is known and can be separated from the rest of the scientific name, the authority string should also be placed in the “author” field.

citation Citation information specified by the data publisher.

commonNames A list of common or vernacular names. Every CommonName is characterised by the *language*, i.e., the language of the name, and *name*, i.e., a common or vernacular name.

credits Citation information specified by the data publisher.

id The unique identifier used for a name or taxon reference in the core data file.

lsid An identifier for the nomenclatural (not taxonomic) details of a TaxonomyItem. It is exclusively used to reference an external and resolvable identifier that returns nomenclatural (not taxonomic) details of a name. Use taxonID to refer to taxa. Use to explicitly refer to an external nomenclatural record.

modified Date when the record the TaxonomyItem is representing was last updated.

parent The TaxonomyItem of the immediate higher-rank parent taxon (in a classification).

properties Any property besides the others that is worth to associate with the TaxonomyItem. These properties might include comments or notes about the taxon or

²Names and semantics are inherited from the Darwin Core Terms.

³For instance <http://rs.tdwg.org/dwc/terms/type-vocabulary/index.htm>

⁴Names and semantics are inherited from the Darwin Core Terms.

name or data sources. An ElementProperty is characterised by *name*, i.e., the name of the property, and *value*, i.e., the value of such a property.

rank The taxonomic rank⁵ of the most specific name in the scientificName.

scientificName The scientific name of taxon with or without authorship information depending on the format of the source database.

status The status of the taxon represented by the TaxonomyItem. A TaxonomyStatus is characterised by: (i) *status*, i.e., the status of the use of the scientificName as a label for a taxon with the following allowed values “accepted”, “invalid”, “misapplied”, “provisional”, “synonym”, “unknown”, and “valid”; (ii) *refId*, i.e., a pointer to the record representing the valid (zoological) or accepted (botanical) name using the taxonID of that record as reference; and (iii) *statusAsString*, i.e., comments or notes about the taxon or name.

5. Mappings

In order to produce the Occurrence Objects and Taxonomy Objects previously described, the SPD service relies on mappings implemented by the plug-ins transforming the information retrieved by interfacing the data sources into the information represented according to the data model.

At the time of writing this report, the following plug-ins have been developed:

BrazilianFlora to interface with the List of Species of Brazilian Flora data source⁶ and give access to Taxonomy Objects.

CatalogueOfLife to interface with the Catalogue of Life data source⁷ and give access to Taxonomy Objects.

GBIF to interface with the Global Biodiversity Information Facility data source⁸ and give access to Occurrence Objects.

IRMNG to interface with the Interim Register of Marine and Nonmarine Genera data source⁹ and give access to Taxonomy Objects.

ITIS to interface with the ITIS data source¹⁰ and give access to Taxonomy Objects.

NCBI to interface with the National Center for Biotechnology Information data source¹¹ and give access to Taxonomy Objects.

OBIS to interface with the Ocean Biogeographic Information System data source¹² and give access to Occurrence and Taxonomy Objects.

SpeciesLink to interface with the speciesLink data source¹³ and give access to Occurrence Objects.

WoRDS to interface with the World Register of Deep-Sea Species data source¹⁴ and give access to Taxonomy Objects.

WoRMS to interface with the World Register of Marine Species data source¹⁵ and give access to Taxonomy Objects.

Depending on the information made available by the various data sources it is not always possible to compile all the attributes devised by the data model previously described.

Information on the level of coverage supported when compiling the SPD objects for each data source is given in Table 2 for Occurrence Objects and Table 3 for Taxonomy Objects. Details on how the specified attributes are compiled are maintained in a dedicated Wiki page¹⁶.

Table 2. Occurrence Object Compilation by Data Source

Attribute	GBIF	OBIS	speciesLink
basisOfRecord	✓	✓	✓
catalogNumber	✓	✓	✓
citation	✓	✓	✓
country	✓	✓	✓
collectionCode	✓	✓	✓
credits	✓	✓	✓
dateLastModified		✓	✓
decimalLatitude	✓	✓	✓
decimalLongitude	✓	✓	✓
eventDate	✓	✓	✓
kingdom		✓	✓
family		✓	✓
id	✓	✓	✓
identifiedBy			✓
institutionCode	✓	✓	✓
maximumDepthInMeters	✓	✓	✓
minimumDepthInMeters	✓	✓	✓
recordedBy	✓	✓	✓
scientificName	✓	✓	✓
scientificNameAuthorship	✓	✓	✓

⁵Recommended vocabulary: <http://rs.gbif.org/vocabulary/gbif/rank.xml>

⁶<http://checklist.florabrasil.net/>

⁷<http://www.catalogueoflife.org/>

⁸<http://www.gbif.org/>

⁹<http://www.cmar.csiro.au/datacentre/irmng/>

¹⁰<http://www.itis.gov/>

¹¹<http://www.ncbi.nlm.nih.gov/>

¹²<http://iobis.org/>

¹³<http://splink.cria.org.br/>

¹⁴<http://www.marinespecies.org/deepsea/>

¹⁵<http://www.marinespecies.org>

¹⁶Species Products Discovery Objects Wiki page http://gcube.wiki.gcube-system.org/gcube/index.php/Species_Products_Discovery_Objects

Table 3. Taxonomy Object Compilation by Data Source

Attribute	BrazilianFlora	CoL	IRMNG	ITIS	OBIS	NCBI	WoRDSS	WoRMS
author	✓	✓	✓	✓	✓		✓	✓
citation			✓		✓		✓	✓
commonNames	✓	✓	✓		✓		✓	✓
credits	✓	✓	✓	✓	✓	✓	✓	✓
id	✓	✓	✓	✓	✓	✓	✓	✓
lsid				✓			✓	✓
modified		✓	✓	✓				
parent		✓	✓	✓	✓	✓	✓	✓
properties		✓		✓	✓		✓	✓
rank	✓	✓	✓	✓	✓	✓	✓	✓
scientificName	✓	✓	✓	✓	✓	✓	✓	✓
status	✓	✓	✓	✓	✓	✓	✓	✓

6. Conclusion

Species data are scattered among several databases and information systems. In this report it is presented the approach implemented by the gCube Species Products Discovery service to give seamless access to such scattered data. In particular, the report discusses the reference architecture characterising the service and the overall approach based on a domain specific query language and an unifying data model. This effectiveness of the approach is demonstrated by its successful use to serve the needs of various communities [1, 2].

Acknowledgments

The work reported has been partially supported by the *iMarine* project (FP7 of the European Commission, FP7-INFRA-STRUCTURES-2011-2, Contract No. 283644) and the *EU-BrazilOpenBio* project (FP7 of the European Commission, FP7-ICT-2011.EU-Brazil, Contract No. 288754).

References

- [1] R. Amaral, R. M. Badia, I. Blanquer, Ricardo Braganeto, L. Candela, D. Castelli, C. Flann, R. De Giovanni, W. A. Gray, A. Jones, D. Lezzi, P. Pagano, V. Perez-Canhos, F. Quevedo, R. Rafanell, V. Rebello, M. Sousa-Baena, and E. Torres. Supporting biodiversity studies with the EUBrazilOpenBio Hybrid Data Infrastructure. *Concurrency and Computation: Practice and Experience*, n/a:n/a, 2014.
- [2] Leonardo Candela, Donatella Castelli, Gianpaolo Coro, Lucio Lelii, Francesco Mangiacrapa, Valentina Marioli, and Pasquale Pagano. An infrastructure-oriented approach for supporting biodiversity research. *Ecological Informatics*, n/a:n/a, 2014.
- [3] gCube Development Team. gCube Website. <https://www.gcube-system.org>, 2008.
- [4] Leonardo Candela, Donatella Castelli, and Pasquale Pagano. gCube v1.0: A Software System for Hybrid Data Infrastructures. Technical Report 2008-TR-035, Istituto di Scienza e Tecnologie dell’Informazione “A. Faedo”, CNR, 2008.
- [5] Leonardo Candela, Donatella Castelli, and Pasquale Pagano. Managing big data through hybrid data infrastructures. *ERCIM News*, (89):37–38, 2012.
- [6] Ian Foster and Carl Kesselman. *The Grid: Blueprint for a Future Computing Infrastructure*. Morgan-Kaufmann, 2004.
- [7] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. A view of cloud computing. *Communications of the ACM*, 53(4):50–58, April 2010.
- [8] Leonardo Candela, Donatella Castelli, Nicola Ferro, Yannis Ioannidis, Georgia Koutrika, Carlo Meghini, Pasquale Pagano, Seamus Ross, Dagobert Soergel, Maristella Agosti, Milena Dobрева, Vivi Katifori, and Heiko Schuldt. *The DELOS Digital Library Reference Model - Foundations for Digital Libraries*. DELOS: a Network of Excellence on Digital Libraries, February 2008. ISSN 1818-8044 ISBN 2-912335-37-X.
- [9] M. N. Huhns and M. P. Singh. Service-oriented computing: key concepts and principles. *IEEE Internet Computing*, 9:75–81, 2005.
- [10] E. Wenger. *Communities of Practice: Learning, Meaning and Identity*. Cambridge University Press, 1998.
- [11] Leonardo Candela, Donatella Castelli, and Pasquale Pagano. Virtual research environments: an overview and a research agenda. *Data Science Journal*, 12:GRDI75–GRDI81, 2013.
- [12] Leonardo Candela, Donatella Castelli, and Pasquale Pagano. Making Virtual Research Environments in the

Cloud a Reality: the gCube Approach. *ERCIM News*, (83):32–33, October 2010.

- [13] Gio Wiederhold. Mediators in the Architecture of Future Information Systems. *Computer*, 25(3):38–49, 1992.
- [14] James L. Edwards, Meredith A. Lane, and Ebbe S. Nielsen. Interoperability of biodiversity databases: Biodiversity information on every desktop. *Science*, 289(5488):2312–2314, 2000.
- [15] Y Roskov, T. Kunze, L. Paglinawan, T. Orrell, D. Nicolson, A. Culham, N. Bailly, P. Kirk, T. Bourgoin, G. Bailargeon, F. Hernandez, and A. De Wever. Species 2000 & ITIS Catalogue of Life, March 2013. Digital resource at www.catalogueoflife.org/col/. Species 2000: Reading, UK.
- [16] speciesLink Consortium. speciesLink. <http://splink.cria.org.br>, 2013.
- [17] List of Species of the Brazilian Flora Consortium. List of Species of the Brazilian Flora. <http://floradobrasil.jbrj.gov.br/>, 2013. Rio de Janeiro Botanical Garden.